Digital Signal Processing and Artificial Intelligence Methods in Agricultural Production Technologies Using Innovative Biofertilizers

Doctoral Dissertation
(Rozprawa doktorska)

M.Sc. Eng. Aleksandra Konopka
(mgr inż. Aleksandra Konopka)

Warsaw University of Life Sciences
Institute of Information Technology
(Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
Instytut Informatyki Technicznej)

Supervisor:
D.Sc. Ryszard Kozera
(Promotor: dr hab. Ryszard Kozera)

Assistant Supervisor:
Prof. D.Sc. Lidia Sas-Paszt
(Promotor pomocniczy:
prof. dr hab. Lidia Sas-Paszt)

SZKOŁA GŁÓWNA
GOSPODARSTWA
WIEJSKIEGO

Warsaw, 2026
(Warszawa, 2026)

Oświadczenie promotora rozprawy doktorskiej

Oświadczam, że niniejsza rozprawa została przygotowana pod moim kierunkiem i stwierdzam, że spełnia warunki do przedstawienia jej w postępowaniu o nadanie stopnia naukowego doktora.

Data 04.02.2026 Czytelny podpis promotora R Kozera

Oświadczenie autora rozprawy doktorskiej

Świadom/a odpowiedzialności prawnej, w tym odpowiedzialności karnej za złożenie fałszywego oświadczenia, oświadczam, że niniejsza rozprawa doktorska została napisana przez mnie samodzielnie i nie zawiera treści uzyskanych w sposób niezgodny z obowiązującymi przepisami prawa, w szczególności z ustawą z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (tj. z dnia 28 października 2022 r., Dz.U. z 2022 r. poz. 2509 ze zm.)

Oświadczam, że przedstawiona rozprawa nie była wcześniej podstawą żadnej procedury związanej z uzyskaniem stopnia naukowego doktora.

Oświadczam ponadto, że niniejsza wersja rozprawy jest identyczna z załączoną wersją elektroniczną.

Przyjmuję do wiadomości, że rozprawa doktorska poddana zostanie procedurze antyplagiatowej.

Data 4.02.2026 Czytelny podpis autora rozprawy Aleksandra Konopka

Abstract

The transformation of modern agriculture toward a sustainable model makes the development of efficient methods for creating and evaluating biofertilizers a scientific imperative. Research on biofertilizer production focuses on various aspects. These studies can be performed independently by biologists, but their repetitive nature and the need to analyze large amounts of data make the automation essential to accelerate analysis, reduce human resources, and draw meaningful conclusions. An interdisciplinary team consisting of biologists and computer scientists conducted research on these topics, and the computer science aspect of the collaboration is presented in this dissertation.

The presented research focuses on two main problems. The first concerns automatic classification of selected genera of soil bacteria. Microorganisms are an important component of biofertilizers, so their proper classification is crucial in the production process. The second area of research involves the classification of plant ploidy based on microscopic images of blackcurrant stomata. These structures are responsible for gas exchange in plants and transpiration, therefore their function is crucial for photosynthesis and, accordingly, for proper plant growth and development. The density and distribution of stomata depend on plant hormones, which are influenced by biofertilizers. Furthermore, foliar-delivered biofertilizers can be applied through stomata.

This dissertation addresses the following issues. The effectiveness of artificial intelligence methods and image processing in the classification of soil bacteria was analyzed. The impact of soil bacteria dataset preparation on the learning process and model evaluation was examined. Experiments were conducted on data prepared by microbiologists with Nikon 80i microscope and automatically collected using KEYENCE VHX-7000N Digital Microscope. Modifications of the Extreme Learning Machine Radial Basis Function (ELM-RBF) method were developed, applying k-medoids and mean shift clustering algorithms, and were subsequently evaluated for soil bacteria classification. The next section of the work concerns the classification of plant ploidy based on stomata microscopic images, comparing the effectiveness of Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs). It is investigated whether the model learns from the stomata cells visible in the images and to what extent the color of the tested samples influences the classification. To test the robustness and generalizability of the developed methods, experiments were performed on two separate sets of stomata images, acquired in different microscopy sessions, thus simulating real-world application conditions.

The included works cover areas where previous research has been very limited, such as the classification of soil bacteria. New topics not previously described in the literature were also investigated, including: studies on an automatically created dataset of soil bacteria; computer-aided classification of plant ploidy; and the use of Vision Transformers in stomata research.

The computer methods are presented in this dissertation, taking into account their historical context. The research utilized digital image processing techniques and artificial intelligence methods. The classical approach involved extracting features related to geometry, color, and texture from the images. Features were computed from both raw images and those subjected to filtering, binarization, and segmentation. The computed features were used as input to classical machine learning methods: K Nearest Neighbors, Support Vector Machine, Random Forest, and Multilayer Perceptron. The ELM classifier, ELM-RBF, and its new modifications presented in this dissertation were also used

for classification. Furthermore, the study employed Convolutional Neural Networks for both segmentation - YOLO (You Only Look Once) algorithm and classification phase - Residual Neural Networks. Stomata were also classified applying Vision Transformers.

This dissertation, written in a form of research paper collection, consists of seven related publications. The first work presents the classification of soil bacteria with classical artificial intelligence methods based on features related to shape and texture. The next paper computes distribution, shape, and color features. The third one employs Convolutional Neural Networks as classifiers. The next paper discusses a modification of the ELM-RBF method using k-medoids and mean shift clustering algorithms, conducted on publicly available datasets. The fifth publication covers research on an automatically created set of soil bacteria, where classification was performed employing both classical methods, proposed modifications of ELM-RBF, and Convolutional Neural Networks. The sixth paper focuses on the classification of plant ploidy based on microscopic images of stomata. YOLO method was used for segmentation, and classical methods and Convolutional Neural Networks were applied for classification. In the latest work, both Convolutional Networks and Vision Transformers were used to classify stomata, and the performance of the classifiers was explained with the Grad-CAM (Gradient-weighted Class Activation Mapping) method.

The conducted research led to the formulation of important conclusions in both considered research areas. In soil bacteria classification, the classical machine learning methods (KNN, ELM-RBF) outperformed deep learning methods when applied to soil bacteria dataset prepared manually by microbiologists. On the other hand, deep learning methods (ResNet152V2) yielded higher classification performance than classical machine learning methods on an automatically captured soil bacteria dataset. In ploidy level classification based on stomata microscopic images, deep learning methods outperformed classical machine learning approach. Both ResNets and ViTs yielded satisfactory performance. The biologically interpretable behavior of ViT models indicates their great potential for application to similar stomata-related tasks.

The dissertation includes recommendations on classification methods for both considered problems and information on factors important in preparing datasets for image analysis. Furthermore, it proposes future research directions, identifies alternative methods for classification and feature extraction, and suggests the application of other advanced Interpretable Artificial Intelligence techniques to these tasks.

Streszczenie

Transformacja współczesnego rolnictwa w kierunku modelu zrównoważonego czyni opracowanie wydajnych metod tworzenia i oceny bionawozów imperatywem badawczym. Badania dotyczące produkcji bionawozów koncentrują się na różnych aspektach. Mogą być one wykonywane samodzielnie przez biologów, jednak ich powtarzalny charakter oraz potrzeba przetworzenia dużej liczby danych sprawia, że ich automatyzacja staje się istotna, aby przyspieszyć analizy, ograniczyć nakłady ludzkie oraz wyciągać znaczące wnioski. Interdyscyplinarny zespół, składający się z biologów oraz informatyków, przeprowadził badania na powyższe tematy, a aspekt informatyczny tej współpracy został przedstawiony w niniejszej rozprawie.

Zaprezentowane badania koncentrują się na dwóch głównych problemach. Pierwszy dotyczy automatycznej klasyfikacji wybranych rodzajów bakterii glebowych. Mikroorganizmy stanowią ważny składnik bionawozów, więc ich właściwa klasyfikacja jest kluczowa w procesie produkcyjnym. Drugi obszar badań obejmuje klasyfikację ploidalności roślin na podstawie mikroskopowych zdjęć aparatów szparkowych czarnej porzeczki. Aparaty szparkowe odpowiadają za wymianę gazową w roślinach oraz transpirację, dlatego ich działanie jest kluczowe dla procesu fotosyntezy, a tym samym, do właściwego wzrostu i rozwoju rośliny. Gęstość i rozmieszczenie aparatów szparkowych zależy od hormonów roślinnych, na które wpływ mają bionawozy. Ponadto, przez aparaty szparkowe mogą być aplikowane bionawozy dostarczane dolistnie.

W ramach niniejszej rozprawy podjęto badania nad następującymi problemami. Przeanalizowano skuteczność metod sztucznej inteligencji i przetwarzania obrazu w klasyfikacji bakterii glebowych. Zbadano wpływ przygotowania zbioru danych bakterii glebowych na proces uczenia i ocenę modelu. Eksperymenty przeprowadzono na danych przygotowanych przez mikrobiologów z wykorzystaniem mikroskopu Nikon 80i oraz zarejestrowanych automatycznie przy pomocy cyfrowego mikroskopu KEYENCE VHX-7000N. Opracowano modyfikacje metody Exteme Learning Machine Radial Basis Function (ELM-RBF) z metodami klasteryzacji k-medoids i mean shift, które zostały następnie przetestowane w klasyfikacji bakterii glebowych. Dalsza część prac dotyczy klasyfikacji ploidalności roślin na podstawie zdjęć mikroskopowych aparatów szparkowych, z porównaniem efektywności Konwolucyjnych Sieci Neuronowych (CNNs) oraz Transformerów Wizyjnych (ViTs). Badane jest czy model uczy się na podstawie widocznych na zdjęciach aparatów szparkowych oraz w jakim stopniu kolor badanych próbek ma wpływ na klasyfikację. Aby przetestować odporność i zdolność do generalizacji opracowanych metod, eksperymenty wykonano na dwóch odrębnych zbiorach zdjęć aparatów szparkowych, pozyskanych w różnych sesjach mikroskopowych, symulując w ten sposób realne warunki aplikacyjne.

Uwzględnione prace obejmują obszary, w których dotychczasowe badania były bardzo ograniczone, takie jak klasyfikacja bakterii glebowych. Zbadano również zagadnienia nowe, nieopisane dotychczas w literaturze, w tym: badania na automatycznie stworzonym zbiorze danych bakterii glebowych; komputerowa klasyfikacja ploidalności roślin; zastosowanie Transformerów Wizyjnych w badaniach nad aparatami szparkowymi.

Metody informatyczne zostały przestawione w niniejszej rozprawie, uwzględniając ich kontekst historyczny. W badaniach zastosowano techniki cyfrowego przetwarzania obrazów oraz metody sztucznej inteligencji. Klasyczne podejście obejmowało ekstrakcję ze zdjęć cech odnoszących się do geometrii, koloru, czy tekstury. Cechy były wyliczane zarówno na zdjęciach nieprzetworzonych, jak i tych, które zostały poddane działaniom fil-

trów, binaryzacji lub segmentacji. Wyliczone cechy zostały użyte jako dane wejściowe do klasycznych metod uczenia maszynowego: K Najbliższych Sąsiadów, Maszyny Wektorów Nośnych, Losowego Lasu Decyzyjnego i Perceptronu Wielowarstwowego. Do klasyfikacji wykorzystano również klasyfikator ELM, ELM-RBF oraz jego nowych modyfikacji zaprezentowanych w tej rozprawie. Ponadto, w badaniach zastosowano Konwolucyjne Sieci Neuronowe zarówno na etapie segmentacji - algorytm YOLO (ang. You Only Look Once) oraz klasyfikacji - Rezydualne Sieci Neuronowe. Aparaty szparkowe były również klasyfikowane przy pomocy Transformerów Wizyjnych.

Niniejsza rozprawa, napisana w formie zbioru artykułów badawczych, składa się z siedmiu powiązanych publikacji. Pierwsza prezentuje klasyfikację bakterii glebowych z zastosowaniem klasycznych metod sztucznej inteligencji na podstawie cech związanych z kształtem i teksturą. W kolejnej liczone są cechy związane z rozmieszczeniem, kształtem oraz kolorem. W trzeciej jako klasyfikatorów użyto Konwolucyjnych Sieci Neuronowych. Następna praca dotyczy modyfikacji metody ELM-RBF z zastosowaniem metod klasteryzacji k-medoids i mean shift, badania przeprowadzono na publicznie dostępnych zbiorach danych. Piąta praca obejmuje badania na automatycznie stworzonym zbiorze bakterii glebowych, gdzie do klasyfikacji wykorzystano zarówno metody klasyczne, zaproponowane modyfikacje ELM-RBF oraz Konwolucyjne Sieci Neuronowe. Szósta praca koncentruje się na klasyfikacji ploidalności roślin na podstawie zdjęć mikroskopowych aparatów szparkowych. Do segmentacji użyto metody YOLO, a do klasyfikacji wykorzystano metody klasyczne oraz Konwolucyjne Sieci Neuronowe. W ostatniej pracy do klasyfikacji aparatów szparkowych zastosowano zarówno Sieci Konwolucyjne, jak i Transformery Wizyjne, a działanie klasyfikatorów wyjaśniono za pomocą metody Grad-CAM (ang. Gradient-weighted Class Activation Mapping).

Przeprowadzone badania pozwoliły na sformułowanie ważnych wniosków w obu rozważanych obszarach badawczych. W klasyfikacji bakterii glebowych, klasyczne metody uczenia maszynowego (KNN, ELM-RBF) okazały się skuteczniejsze niż metody głębokiego uczenia w przypadku zastosowania ich do zbioru danych bakterii glebowych przygotowanego ręcznie przez mikrobiologów. Metody głębokiego uczenia (ResNet152V2) zapewniły jednak wyższą dokładność klasyfikacji niż klasyczne metody uczenia maszynowego w przypadku automatycznie pozyskanego zbioru danych bakterii glebowych. W klasyfikacji poziomu ploidalności opartej na mikroskopowych obrazach aparatów szparkowych, metody głębokiego uczenia przewyższyły klasyczne podejście uczenia maszynowego. Zarówno ResNets, jak i ViT zapewniły zadowalającą wydajność. Biologicznie interpretowalne zachowanie modeli ViT wskazuje na ich duży potencjał w zakresie zastosowania do podobnych zadań związanych z aparatami szparkowymi.

Niniejsza rozprawa zawiera rekomendację metod klasyfikacji dla obu zagadnienień oraz informacje na temat czynników ważnych przy przygotowywaniu zbiorów danych pod kątem analizy obrazu. W rozprawie zaproponowano przyszłe kierunki badań, alternatywne metody klasyfikacji i ekstrakcji cech, oraz zastosowanie innych zaawansowanych technik Interpretowalnej Sztucznej Inteligencji.

Acknowledgements

I would like to extend my most heartfelt thanks to all the individuals without whose support the completion of this doctoral dissertation would not have been possible.

First and foremost, I thank my Supervisors:

- Ryszard Kozera, Ph.D., D.Sc. for introducing me to the world of science, the invaluable guidance of my research in the computational aspects, constant support, and valuable advice during the preparation of publications.
- Lidia Sas-Paszt, Prof. for openness to interdisciplinary collaboration, coordination of activities within the research team, and the fundamental guidance and substantive support regarding the biological aspects.

I also sincerely thank:

- Karol Struniawski, Ph.D. (Eng.) for collaboration in the field of computer science and kind advice during the preparation of publications.
- Paweł Trzciński, Ph.D. for the preparation and provision of a specialized dataset of microscopic images of bacteria for research purposes.
- Agnieszka Marasek-Ciołakowska, Ph.D., D.Sc. and Aleksandra Machlańska, M.Sc. for inspiring collaboration in research on stomata, providing data, granting access to and assistance with the microscope, as well as working together in an interdisciplinary and international team.
- Luciano Ortenzi, Ph.D. for collaboration on the classification of stomata and significant guidance of the work from the computational side.

I thank all the other co-authors of the publications that formed the foundation of this dissertation for their valuable substantive oversight and contribution to the development of the biological aspects of the presented research.

The memory of my academic teachers from my student days remains particularly important to me. I thank them for inspiring classes taught with great dedication, for fascinating publications that ignited my curiosity about the world, and for instilling in me a passion for research work.

To all the individuals mentioned, I am deeply grateful for the time dedicated, kindness, expert knowledge, and unwavering support at every stage of this scientific journey.

Podziękowania

Pragnę złożyć najserdeczniejsze podziękowania wszystkim osobom, bez których wsparcia powstanie tej rozprawy doktorskiej nie byłoby możliwe.

Przede wszystkim dziękuję moim Promotorom:

- dr. hab. Ryszardowi Kozerze za wprowadzenie w świat nauki, nieocenione ukierunkowanie badań w aspektach informatycznych, stałe wsparcie oraz cenne wskazówki przy tworzeniu publikacji.
- prof. dr hab. Lidii Sas-Paszt za otwartość na interdyscyplinarną współpracę, koordynację działań w zespole badawczym oraz fundamentalne ukierunkowanie i wsparcie merytoryczne dotyczące aspektów biologicznych.

Serdecznie dziękuję również:

- dr inż. Karolowi Struniawskiemu za współpracę w zakresie informatyki oraz życzliwe rady podczas przygotowywania publikacji.
- dr Pawłowi Trzcińskiemu za wykonanie i udostępnienie na potrzeby badań specjalistycznego zbioru zdjęć mikroskopowych dotyczących bakterii.
- dr hab. Agnieszce Marasek-Ciołakowskiej oraz mgr Aleksandrze Machlańskiej za inspirującą współpracę w badaniach nad aparatami szparkowymi, dostarczenie danych, udostępnienie mikroskopu i pomoc w jego obsłudze, a także wspólną pracę w interdyscyplinarnym i międzynarodowym zespole.
- dr Luciano Ortenzi za współpracę przy klasyfikacji aparatów szparkowych oraz istotne ukierunkowanie prac od strony informatycznej.

Dziękuję wszystkim pozostałym współautorom publikacji, które stały u podstaw tej rozprawy, za ich cenny nadzór merytoryczny oraz wkład w rozwój aspektów biologicznych prezentowanych badań.

Wspomnienie moich nauczycieli akademickich z czasów studiów pozostaje dla mnie szczególnie ważne. Dziękuję za inspirujące zajęcia prowadzone z ogromnym zaangażowaniem, za fascynujące publikacje, które rozbudziły moją ciekawość świata, i za to, że zaszczepili we mnie pasję do pracy badawczej.

Wszystkim wymienionym osobom jestem głęboko wdzięczna za poświęcony czas, życzliwość, ekspercką wiedzę i nieustające wsparcie na każdym etapie tej naukowej drogi.

Badania finansowane przez Narodowe Centrum Badań i Rozwoju
w ramach programu BIOSTRATEG
numer umowy BIOSTRATEG3/344433/16/NCBR/2018

Tytuł projektu:
NOWE ROZWIĄZANIA BIOTECHNOLOGICZNE W DIAGNOSTYCE, ZWALCZANIU I MONITORINGU KLUCZOWYCH PATOGENÓW GRZYBOWYCH W EKOLOGICZNEJ UPRAWIE OWOCÓW MIĘKKICH

Contents

Chapter 1

Introduction

The dissertation presents the application of computational methods in the biological domain in the context of: automatic classification of selected genera of soil bacteria and classification of plant ploidy based on microscopic images of blackcurrant stomata. In this chapter, the research background is introduced, outlining the interdisciplinary character. An overview of artificial intelligence methods employed in the constituent works is provided within a historical context to emphasize their interrelationships and applications. The research problems addressed in this dissertation are determined and explained, a review of the current state of the art is presented, and the main contributions of the dissertation are outlined.

1.1 Research background

The artificial intelligence (AI) models have been widely developed. There are numerous theoretical works that focus on describing mathematical models and information technology solutions; however, what is that really matters for humans are their practical applications having impact on our life quality and economical growth of our societies. The computerized approaches are incorporated to automate processes in various domains.

The biology is a very broad field. It explains the processes that form our environment, interprets animal behaviors, analyses the anatomy of living organisms, etc. The biologists are also responsible for controlling amount and quality of cultivated crops. That supervision is important to maintain human health and prevent famine. The biologists are aware how to monitor plant growth and which fertilizers are crucial to ensure their proper development. Although there are many chemical fertilizers stimulating the plant growth process, they could also have a negative impact on environment. In order to replace the artificially developed and harmful substances, biologists are developing biofertilizers which utilize microorganisms [1].

The National Institute of Horticultural Research in Skierniewice, Poland specializes in creation of biofertilizers and has been developing and producing them for decades. This institution has developed numerous patents in this field and its researchers cooperate within many international research groups. Nowadays, when many processes are accelerated with computerized approaches, the idea emerged to form an interdisciplinary team contained of biology and information technology (IT) specialists aiming to automate selected research tasks related to biofertilizer creation.

The interdisciplinary group is important to perform a comprehensive and significant

research. In the team of biologists and computer scientists, different attitudes are represented by both groups. The biologists form the problem that needs to be solved, they understand the nature of the issue, understand which conditions are needed to perform experiments, understand the importance of the experiment and its impact on the environment and economy. Whereas, computer scientists have an expertise in statistics, mathematics, computer vision, and artificial intelligence, they are able to automate processes accelerating their performance so that large-scale experiments can be conducted with reduced human intervention and resources. Additionally, they understand a need to normalize technical aspects such as microscope settings or light conditions. IT specialists are capable of drawing statistical conclusions, visualizing, explaining, and verifying the significance of the yielded results. This cooperation permits not only to solve the existing problems, but also to locate the research gaps and investigate the orientation of further investigation.

The research on biofertilizers is not limited to simply testing their performance on cultivated crops. The cooperation with biologists, which yielded the works involved in this dissertation, is focused on verifying specific areas related to biofertilizer production. The biofertilizers are composed of microorganisms which need to be correctly detected and classified [1]. This work can be time consuming and tedious when done with conventional methods. Due to that reason, the aim of this research is to verify, whether the computer systems, powered by artificial intelligence, are capable of classifying selected soil bacteria genera based on microscopic images. The main focus of further research is the analysis of stomata cells in plants, especially classification of plant ploidy levels based on microscopic images including stomata cells, as these structures can be influenced by the biofertilizer application [2, 3, 4].

This dissertation summarizes the research on bacteria and stomata classification focusing on IT aspects. The included works are presented, not only as a collection of related publications, but also as a logical narrative that traces the evolution of the applied methodologies. The initial works focus on handcrafted features extraction and application of classical machine learning approaches (Section 3.1,3.2), then Convolutional Neural Networks (CNNs) [5] are applied (Section 3.3, 3.6), modifications of machine learning methods are proposed (Section 3.4) and tested (Section 3.5), Vision Transformers (ViTs) [6] are used and their performance is explained (Section 3.7).

This dissertation includes a literature review in the development of artificial intelligence to present applied methods in a wide historical context (Section 1.2). The main research problems considered in the publications are also explained (Section 1.3). In addition, an overview of state of the art solutions is presented in the field of soil bacteria classification and stomata analysis in the context of problems solved in this dissertation (Section 1.4). The contributions of this research to the field are outlined (Section 1.5). The individual contribution of each author is explained (Section 2.1), and the experiments conducted in each work together with the summarized results are provided (Section 2.2). The full texts of the publications constituting this dissertation are presented in Section 3. Finally, the obtained results are summarized, conclusions are drawn (Section 4) and recommendations for future extensions of these publications are provided (Section 5).

1.2 A review of artificial intelligence

Most of the methods applied in this dissertation represent selected artificial intelligence techniques. One could assume that AI is a concept that emerged some years ago as a mysterious humanoid form created of ones and zeros - but that is not true. This section presents the performance and application of the methods employed in this dissertation within the broader historical context of neural network development, beginning with the fundamental concept of a single artificial neuron.

1.2.1 From a single neuron to deep network

The history of artificial intelligence began in 1940's when McCulloch and Pitts proposed a model of an artificial neuron [7]. This mathematical model, inspired by the biological neuron, treats input and output data as numerical values rather than electrochemical signals. Artificial neuron takes a vector of numbers as input and process these data by multiplying input vector by a vector of weights. The weights leverage the importance of input vector elements in decision making process. Then this product is assigned as input to activation function which is a simple mathematical function with binary output defining a threshold for the decision. This biologically-inspired model became a fundamental building block of the artificial neural networks which are foundations of such solutions as famous Large Language Models (LLMs) [8] such as ChatGPT [9] or DeepSeek [10].

The single neuron weights were fixed values which could not adapt to the considered task [7]. The solution to this problem was the perceptron algorithm, created by Frank Rosenblatt in 1957 [11], which iteratively adjusted parameters of artificial neuron. In each iteration, it is verified whether the binary output of the network is as expected - if not, then an update of weights is performed. In case of two dimensional data, the weights at each iteration form a straight line which is a decision boundary. All data samples which were assigned to first class (e.g., values 0) should be separated with the boundary from the second class (e.g., values 1). As the decision boundary forms a hyperplane, the algorithm is only able to solve linearly-separable two-class problems. To overcome this constraint and address more complex problems, a new approach had to be developed.

In 1958 [12], Rosenblatt experimented with combining neurons with randomized weights all together; however, there was a need for an algorithm that could perform the training process of all parameters in such network. In 1960 [13], Widrow and Hoff proposed Adaptive Linear Neuron (ADALINE), which parameters were trained based on Gradient Descent (GD) Algorithm [14]. GD was developed in 1847 by Cauchy [15]. The GD algorithm aim is to find local minimum of a function which in neural networks is applied to minimize the error of a loss function [16]. The Gradient Descent is an iterative approach which takes steps in the opposite direction of the gradient. In 1967 [17], Amari proposed a multilayered neural network which was trained by Stochastic Gradient Descent (SGD) algorithm [18]. This neural network was capable of solving non-linearly separable problems. In GD, all the samples from the dataset are considered for a single update. In SGD, computations are applied to a single randomly selected data point, which optimizes the process is performed. SGD optimized the process comparing to GD; however, in Multilayered Neural Networks (MLNNs) many updates of weights were needed. What's more, the updates of weights which were at the beginning layers of the network were based on the updates of weights modification in deeper layers. MLNNs form hierarchical structures which learn more complex features in deeper layers. The weight adaptation

process also needed an optimization. The method for efficient automatic differentiation proposed by Linnainmaa in 1970 [19] was applied to Multilayer Neural Network in 1982 by Werbos [20] and was later termed Backpropagation [21]. This algorithm applied chain rule, a mathematical concept developed in 1676 by Leibniz [22], to GD algorithm for adjustment of the parameters in neural network. The Backpropagation updated the weights from the deepest layers back to the ones at the beginning considering the modifications that needed to be adjusted to weights, based on differences between expected and yielded output. The values of gradients for the modifications are computed without redundancy and these values are propagated from the back of the network yielding an efficient process.

1.2.2 Multilayer Perceptron

The multilayered neural networks were termed as Multilayer Perceptron (MLP) for the first time in 80's [23]. The Multilayer Perceptron was designed to solve classification and regression problems upon suitable parameter adjustments. The following architecture modifications are commonly applied:

- number of layers and number of neurons in layers - generally deeper networks are capable of solving more complex problems,
- choice of activation functions - these functions, unlike in perceptron where binary functions were applied, need to be differentiable so optimization methods can be applied, the most common activation functions employed in MLP are: sigmoid, hyperbolic tangent, ReLU, linear (identity), and SoftMax [24].

The training process configuration can also be specified with the determination of the following factors:

- batch size - is the number of samples that the network is feed at once so that the input data are not a single vector of features but a matrix including many samples which optimizes memory usages, reduces training time and stabilizes the training process [25],
- number of epochs - number of times where whole dataset of samples is applied to train the network which allows model to iteratively adjust weights and biases,
- optimization methods - Gradient Descent [14], Stochastic Gradient Descent [18], Momentum [21], Adaptive Gradient Algorithm (Adagrad) [26], Adadelta [27], Root Mean Square Propagation (RMSprop) [28], Adaptive Moment Estimation (Adam) [29], AdamW [30],
- learning rate - a value used in optimization methods to specify the step size in each iteration of these algorithms, this value can be fixed or adjusted during training,
- loss function - a function that is minimized during the training process e.g., Mean Squared Error or Mean Absolute Error for regression and Binary Cross-Entropy or Categorical Cross-Entropy for classification [16].

Regularization is a methodology applied to stabilize the learning process and prevent overfitting. Overfitting is a phenomenon when a model adjusts its performance to data it is trained on and does not generalize well. The model's performance is well on training data, but the results obtained on testing data are significantly less accurate. To overcome this issue, one could employ selected regularization methods listed below:

- dropout [31] - applied as random deactivation of some neurons in a layer [32] by setting temporarily their output values to zero,
- parameter regularization [33] - e.g., ℓ^1 (Lasso - Least Absolute Shrinkage and Selection Operator) and ℓ^2 (Weight Decay) regularization applied to loss function, in ℓ^1 by focusing on more important features or in ℓ^2 proportionally reducing weight impact of all features,
- early stopping [34] - a methodology that is based on stopping the learning process before the model overfits to the data in training set. The analysis usually covers values of loss function computed on validation set - when the values of this function do not decrease for a given number of epochs (patience) then the training is stopped,
- batch normalization [35] - the methodology that speeds up and stabilizes training process by normalizing activations of a given layer before passing them to activation functions.

MLP is a feedforward neural network [36] which means that model is fed with data in one direction from input, through hidden layers to output. The input data to MLP are vectors of values which describe given samples. The more descriptive they are the better. In classification problem these values are the basis to differentiate samples from each classes.

1.2.3 Feature computation

The features are represented as vectors or values describing samples in the dataset. A natural question arises: How to choose which features should be calculated? This decision is made by computer specialists in cooperation with experts in a given field ensuring their significance in describing the task in question. In case of the biological research presented in this dissertation, these features describe microscopic images. One can compute features related to such aspects as e.g., color, geometry, texture, object dispersion. Selected features describe a single image and need to be computed for all the images included in the examined dataset.

The image is a two dimensional signal which can be preprocessed to extract specific patterns enabling to distinguish classes, before the feature set is computed. For example, the Gaussian mean and median filters can be applied to reduce the noise and remove small unimportant details [37]. To highlight or detect edges Laplace and Sobel filters are employed [38]. Contrast in images can be corrected by application of histogram stretching or equalization techniques [39]. The images can also be transformed to frequency domain by application of Fourier, Wavelet, or Cosine Transforms in order to extract features in this alternative domain [40]. Additionally, the filtering can be applied in frequency domain and then the image is transformed back to spatial domain with reverse Transforms. The image may consist of the objects that should be measured or classified e.g., bacteria cells or stomata structures and the background which is unimportant in the classification process. To extract region of interest in images, binarization is applied by either specifying a threshold or by application of algorithms such as Otsu [41]. Once binary masks are computed morphological operations can be employed such as erosion, dilatation, opening or closing to obtain such mask modifications as reducing object size, enlarging them, removing small objects or filling holes in region of interest [42].

Features can be extracted on either preprocessed or raw images. Since images may differ in color, the computed features could consist of statistical values such as median, mean, kurtosis, standard deviation, skewness computed in various color spaces [43] (e.g., RGB, HSV, and Lab). The color features can also be extracted from color histograms. In order to calculate traits of objects which are characterized by various shapes, features based on geometry could be measured by calculating various distances and angles in extracted objects. These features include computations of object bounding boxes size, convex area, convex hull, Feret diameter [44], eccentricity, or circularity [45]. Interpolation techniques can be employed to outline contour of an object [46] enabling computation of the following statistics: perimeter, curvature, or arc lengths. Texture features could be derived, when classified objects have different patterns. Gray-Level Co-occurrence Matrix (GLCM) and Gray-Level Run-Length Matrix (GLRLM) could be employed to compute contrast, homogeneity, correlation, energy, and entropy [47]. Dispersion features can also be specified when the density and shape of the group of objects are crucial. These may involve evaluating object clustering with methods such as k-means [48] or mean shift [49].

Once many features are computed, it can be hard to draw conclusions about which of them will turn out to be significant for the algorithm or whether they include redundant information (features correlated with other features). The meaningful traits can be selected with feature selection methods such as: Correlation-Based Feature Selection (CFS) developed in 1999 by Hall [50], Fast Correlation-Based Filter (FCBF) created in 2003 by Yu and Liu [51], SBMLR (Sparse Bayesian Multinomial Logistic Regression) proposed in 2006 by Cawley et al. [52]. These methods are provided with a computed set of features, and they choose which of them are significant and should be considered to yield the highest classification performance.

1.2.4 Model performance evaluation

Selection of features in cooperation with experts and reduction of the feature set to the most descriptive ones and adjustment of parameters in feature selection methods allow to yield satisfactory results in considered tasks. The performance of the MLP models should be evaluated by objective metrics. These metrics are based on comparison of the expected values with the output yielded by the model. The model is trained on labeled data, meaning every sample includes information about the class it belongs to.

For classification tasks considered in this dissertation some of these metrics that could be computed are (Section 3.7):

- accuracy (ACC) - percentage of all correct predictions,
- precision - proportion of true positives divided by true and false positives combined,
- recall (sensitivity) - proportion of true positives divided by true positives and false negatives combined,
- specificity - proportion of the true negatives divided by true and false negatives combined,
- f1-score - harmonic mean of precision and recall,
- confusion matrix - a table summarizing the assignment of samples to classes, presenting whether the predictions were correct.

The obtained results may substantially vary depending on which part of the data is assumed to be a training set and which set constitutes the testing set. One could select the data on which the model learns well to distinguish data from testing set properly, but for the different combination the results would be much worse. In order to ensure that the considered results are the most general, objective, reproducible and have an impact in a considered domain one needs to repeat the experiments many times on different sets or subsets. The approach which is commonly applied in such situation is cross-validation proposed in 1970's independently by Stone [53] and Geisser [54]. In k-fold cross-validation dataset is divided into a number of sets - termed as folds. For example, 10-fold cross-validation divides data into 10 folds (Section 3.1). The samples are usually assigned to folds randomly. Each fold serves as testing set in a given iteration while the remaining folds form training set. The model is trained on training set and tested on testing set. Then next fold is considered the testing set and remaining 9 form training set and the procedure is repeated until all combinations were processed. Finally, all the results are summarized and one can draw conclusions about the model performance analyzing e.g., mean accuracy and standard deviations of metrics computed on all dataset combinations. This approach ensures that the final result is more general and does not rely on specific combination of samples considered in training and testing sets. To guarantee that in imbalanced datasets each training and testing set includes sufficient representations of all the considered classes, stratified cross-validation [55] is applied. Consequently, the ratio of class samples in the entire dataset is preserved across all training and testing sets.

1.2.5 Convolutional Neural Networks

The features considered as input to MLP created by programmers in cooperation with experts are termed handcrafted features. The handcrafted features in image classification are based on human sight - the expert explains how the given classes are differentiated and with this knowledge the programmer decides which features should be computed. What if an algorithm could process the images for us extracting the features on his own without neither help of an expert in a field nor a programmer effort?

In 1950s and 1960s neurophysiologist Hubel and neurobiologist Wiesel examined the mechanisms of visual cells in the brain [56]. The experiments were carried out on cats, but the conclusions can be extended to other mammals including humans. The researchers revealed that neurons in visual cortices are activated by certain patterns (e.g., shapes and edges). These neurons form hierarchical structures which are capable of distinguishing more and more complex patterns. Hubel and Wiesel specified two types of visual cells in brain: simple cells - which are responsible for detecting simple patterns in a given location and complex cells - which react to the same type of pattern but they do not precise its exact location.

The researchers proposed a hierarchical model which explained how the cells form cascade structures enabling the sight process. The scientific achievements of Hubel and Wiesel were applied in 1969 by computer scientist Fukushima to create multilayer visual feature detection network; however, the weights of proposed model were not trained [57]. In 1980 [58], Fukushima proposed Neocognitron - a hierarchical multilayer neural network - which consisted of two types of altering layers:

- S-layer - later termed as convolution layer, in which the image is processed by a number of filters responsible for emphasizing specific patterns,

- C-layer - known as pooling layer, responsible for reduction of the dimension while retaining the most important aspects of the image.

This model was applied for Japanese handwritten characters recognition. Neocognitron was trained in an unsupervised process. This means that the model learned to distinguish different classes of objects without prior labeling by experts. The first time Backpropagation algorithm was applied in Convolutional Neural Network was in 1987 by Waibel, in Time Delay Neural Network, which was applied for one dimensional data sequences and was used in speech recognition [5]. The person who applied Backpropagation to learn CNN parameters for two dimensional data was LeCun in 1989, who is considered the father of CNNs [23]. LeCun performed classification of hand-written numbers.

Convolutional Neural Networks include many hyperparameters and parameters. The parameters are computed through the training process and hyperparameters are set by a programmer manually. The hyperparamers in CNNs are:

- number of convolution and pooling layers,
- number of filters in each layer and their sizes,
- stride in convolution and pooling, which is a number of pixels by which a filter is moved when it goes through whole image,
- size of pooling window,
- padding in convolution and pooling,
- type of pooling (e.g., max pooling or average pooling [59]).

The parameters that are learned through Backpropagation are values of weights in kernels and their corresponding biases. The altering convolution and pooling layers are responsible for computing a representation of automatically generated traits (deep features). These values are then treated as input to a Multilayer Perceptron which performs the classification. MLP can also be replaced with other classifiers.

The idea of CNNs, although very promising, could not evolve due to the fact that it required a lot of computational resources and classical machine learning approaches such as K Nearest Neighbours [60], Random Forest [61] or Support Vector Machines [62] were easier to apply in research problems due to their superior computational efficiency. A breakthrough in this field happened with development of large datasets, increased GPU computational power and the application of transfer learning to CNNs. Transfer learning is an approach, when model is pretrained on some dataset and then parameters in their selected layers are fine-tuned to the new dataset to solve similar or more specific tasks.

In 2010, a Chinese-American computer scientist Li and her team, after three years of labeling, created a large database of various images - ImageNet [63]. This database consists of over 20,000 categories of images. ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [64] is an annual event which first took place in 2010 - the participants needed to classify the images minimizing the error rate. In 2010 and 2011, the winning algorithms were based on traditional computer vision and machine learning techniques; however, in 2012 a Convolutional Neural Network managed to outperform the competition. This CNN was AlexNet created by Krizhevsky, Sutskever and Hinton at the University of Toronto [65].

The idea of transfer learning was known before AlexNet, but only when it was applied to CNN architectures trained on large datasets caused a breakthrough in development of AI. It proved that CNNs can obtain higher performance than classical approaches encouraging machine learning experts to work on development of other deep learning architectures.

After the success of AlexNet, many other CNN architectures were developed. The next ILSVRC winner was OverFeat [66] proposed in 2013, which combined classification, localization, and detection. This architecture applied multi-scale classification processing the entire image at once which increased computation efficiency. In 2014, VGGNet [67] was created which consisted of smaller convolutional filters and larger amount of layers comparing to AlexNet. That same year GoogLeNet [68] was proposed, which introduced Inception module capable of performing computation of multiple convolution operations in parallel matter (the ILSVRC winner in 2014). In 2015, Residual Neural Network (ResNet) [69] was developed which introduced a concept of residual learning and skip connections addressing the vanishing and exploding gradient problem (the winner of ILSVRC in 2015). The algorithm which contains a classification component; however, its commonly applied to detection and segmentation tasks is You Only Look Once (YOLO) algorithm [70]. The first version of this algorithm was developed in 2015 combining the architecture of GoogLeNet with localization concept of the OverFeat architecture. In 2016, DenseNet [71] was proposed which applied dense connections between layers ensuring maximization of information flow and gradient propagation. In 2017, MobileNet [72] was designed as a tool applicable to mobile devices ensuring reduction of computation costs with depthwise separable convolutions. In 2019, EfficientNet [73] was proposed which applied compound scaling method balancing network depth, width, and resolution.

With the development of Convolutional Neural Networks the models craved for data to process. In CNNs it is essential to privide sufficient amount of data which can be hundreds or thousands images per class as the models consist of millions of parameters which need to be adjusted. This large amount of data can not always be available due to lack of samples or limited human resources to provide them. If one simply copied the images to enlarge their quantity this approach would lead to overfit of the model. Additionally, the number of samples in considered classes could be imbalanced leading to artificially higher or lower performance. To overcome these challenges an augmentation technique [74] is applied. This approach is based on artificially generating multiple data based on the raw dataset. In the context of images, augmentation can contain geometry and color modifications, sample mixing, replacing or removing selected image parts. The more advanced data augmentation can also be performed with Generative AI including such algorithms as Stable Difussion [75] or ControlNet [76].

1.2.6 Other selected neural network architectures

In parallel to deep Convolutional Neural Networks, other techniques were developed. Randomized Neural Networks based on random assignment of some weights and training remaining model parameters with analytical approaches. The most famous architecture of Randomized Neural Networks is a family of Extreme Learning Machines first proposed in 2004 by Huang et al. [77]. Extreme Learning Machine is a feedforward neural network consisting of a single input, hidden and output layers of neurons. The weights between input and hidden layers, and the biases are initialized by uniform distribution. The weights between hidden and output layer are computed with a single algebraic operation

Moore-Penrose Pseudoinverse which minimizes the mean residual squared error. This approach yields a very fast training process, which, however, is limited by the amount of space in Random Access Memory (RAM) when more data is considered as it leads to an operation on a high dimensional matrix.

Along with development of feedforward neural networks other approaches were also proposed. Recurrent Neural Networks (RNNs) designed to process sequential data applied cycles (recurrent connections) to their architecture. This concept was based on neurobiology - recurrent connections in brain were observed in the late 1800s and early 1900s by Cajal in cerebellar cortex [78]. McCulloch and Pitts proposed a model with considered recurrent connections in their work from 1943 [7]. In 1982, Hopfield network [79] was proposed forming auto-associative memory. The iterative minimization of energy function enabled to restore complete patterns from noisy data resulted in first application of neural networks in optimization problems. The idea of RNNs was developed through time forming two main architectures: Long short-term memory networks (LSTM) proposed in 1995 by Hochreiter and Schmidhuber [80], and Bidirectional recurrent neural networks (BRNN) developed in 1997 by Schuster and Paliwal [81]. LSTM aimed to mitigate the vanishing gradient problem in RNNs introducing concept of gates and cell states to ensure stable gradient flow. BRNN architecture consists of two hidden layers set in opposite directions forming a generative deep learning approach. BRNN are applied when the context is crucial for data analysis. The considered two architectures are usually combined yielding bidirectional LSTM. RNNs are commonly applied to tasks such as speech recognition [82] and natural language processing [83].

Another concept that was developed are Graph Neural Networks (GNN). This type of networks are applied to structures made of nodes such as: social networks, molecules [84] or transportation systems [85]. GNN were first proposed in 2009 work by Scarselli et al. [86]. The original GNN implementation used a framework based on RNN. Then graphs were also applied to signal processing [87]. This concept was further developed, e.g., in 2016 they were combined with CNNs forming Graph Convolutional Networks [88].

Generative Adversarial Networks (GAN), proposed in 2014 [89] by Goodfellow, consist of two neural networks (generator and discriminator) which compete against each other in zero-sum game. Generator aims to generate data that pretend to be the real and discriminator tries to verify, whether the generated data are authentic or artificial. These neural networks are applied in fields such as medicine [90], gaming [91], engineering [92] or natural language processing [93], allowing to create hyper-realistic media [94] such as images (including photographs of faces), videos, and audio.

In 2017, a new AI architecture called Transformer was proposed [95]. Transformers were applied to natural language processing often outperforming Recurrent Neural Networks [96]. This architecture allows to generate text based on the acquired knowledge [97]. Its core components are encoder and decoder. The encoder's role is to build a comprehensive, contextualized representation of the input sequence, while decoder generates an output sequence using the contextualized representation provided by the encoder [98]. These two can be combined, though many models use only one of these stacks. BERT [99] proposed by Google was employing only the encoder part applied to classification, sentiment analysis or answering questions. GPT [100] designed by OpenAI solely relied on decoder to generate creative meaningful texts. Transformers are based on self-attention mechanism and multi-head attention allowing parallel text analysis [101]. The Transformer models are trained on a training corpus analyzing the meaning and relations of tokens [102]. To capture these relations, a self-attention mechanism calculates attention

scores which indicate the degree of dependency between token pairs. The mechanism learns to generate these scores using three specialized matrices (Query, Key, and Value) which are trained via Backpropagation with all other parameters in the model [95].

Various Transformer architectures were developed including current state of the art: GPT-5 [103], DeepSeek [104], Gemini [105], and many others. Vision Transformers are Transformer architectures which are applied to images [6]. ViTs extract image patches and capture global image context effectively [106]. In training process the positional encoding of patches is provided as patches unlike words do not have their natural order defined.

1.3 Research problem

This dissertation is divided into two main research issues: classification of soil bacteria genera based on microscopic images and classification of ploidy levels of plants based on microscopic images of leaves with visible stomata structures. The research presented in this dissertation combines analysis of biofertilizer ingredients (bacteria) and analysis of plant structures which can be influenced by biofertilizers, and which are critical for development of biofertilizer applications (stomata).

1.3.1 Classification of bacteria microscopic images

Bacteria are microorganisms which despite their microscopic size have a significant impact in various fields including medicine [107], industry [108], and ecology [109]. These organisms have an ability to exist in various environments including water, atmosphere, soil, and interior of other living organisms [110]. Bacteria have both positive and negative impact on human well-being e.g., some species cause infections [111], while others are crucial in production of medicines [112].

The field in which bacteria are indispensable is the agriculture [113]. Presence of specific bacteria species is important in maintaining soil fertility [114]. Some bacteria are responsible for decomposition of plant residues which enables their reuse by the growing plants. These microorganisms are crucial in bioremediation process and circulation of chemical elements [115]. Bacteria are also a basis of biofertilizers, which enable sustain plant growth, without having a negative impact on environment, overcoming a problem of fertilizers consisting of artificially created chemical substances [1].

To produce biofertilizers a crucial aspect is to accurately classify the samples which are to be applied in the production. This classification can be performed based on microscopic images including bacteria cells. Microbiologists are aware of the differences in morphology between specific bacteria genera, and due to that fact they can distinguish them manually. This process is costly, time consuming, and requires qualified biological specialists. Therefore, this dissertation aims to automate this process.

In works which focus on soil bacteria classification we first focus on classification of soil bacteria genera (*Enterobacter*, *Rhizobium*, *Pantoea*, *Bradyrhizobium*, and *Pseudomonas*) performed on a set of microscopic images created manually by microbiologists. The initial part focused on application of digital signal processing methods to extract handcrafted features based on geometry (Section 3.1 and 3.2), texture (Section 3.1), dispersion (Section 3.2) and color (Section 3.2). Selected features were applied as input to classical machine learning methods (Section 3.1 and 3.2). Classification performance was also

compared with results yielded by ELM (Section 3.1). In further research (Section 3.3), Convolutional Neural Networks were employed, and majority voting was applied to classify images based on subimages including single bacteria cells.

In the next paper (Section 3.4), a modification of ELM was proposed which was successfully applied to classify bacteria images (Section 3.5). In the next work (Section 3.5), the automation went beyond the analysis of microscopic images, but the microscopic images were taken in automatic manner enabling to create larger dataset. The prepared dataset contained images of selected bacteria genera (*Enterobacter*, *Rhizobium*, *Pantoea*, and *Pseudomonas*) taken under different light conditions and using various glass types. The aim of this research was to verify whether it is possible to classify soil bacteria on automatically created dataset and whether this classification is robust to factors such as light, glass or sample preparation. In this work, both CNNs, classical machine learning approaches and proposed ELM modifications were applied.

1.3.2 Classification of stomata microscopic images

The biofertilizers aim to have a positive impact on plant growth. The plants need specific conditions including access to sunlight, carbon dioxide, microelements and sufficient amount of water to perform photosynthesis [116]. The water loss is controlled by stomata - microscopic pores located in leaves, which allow gas exchange and are responsible for water transpiration [117]. Under conditions of water insufficiency, the stomata pores close to reduce the further water loss. In contrast, when there is enough water and light to perform photosynthesis, but the carbon dioxide amount is insufficient, the stomata cells open to allow gases exchange. A consequence of this opening is increased water loss through transpiration. The biofertilizers application have impact on stomata size, density, and behavior, which is the reason why stomata-related research is crucial in biofertilizer production [3, 4].

The biofertilizers have impact on plant hormones including auxins and cytokinins [118], the growth hormones, which cause the open of stomata cells enhancing photosynthesis; absciscic acid [118], jasmonic acid and ethylene [119], stress hormones which close stomata cells making them sustain to difficult conditions and resistant against pathogen entry. Some of the biofertilizers foliar formulations penetrate plant tissues through the stomata pores [120]. It is essential to perform a research on stomata analysis to propose optimal strategies towards biofertilizer application regarding the state of stomata cells. The biofertilizers aim to manage stomata opening and closing more efficiently to adjust their performance to the changing conditions. The stomata analysis can be performed manually by analysis of microscopic images of leaves; however, this process is very time consuming, and there are thousands of stomata cells that need to be analyzed to withdraw significant conclusions. This is the reason why, there is a need to create automated computerized systems which allow fast and robust stomata analysis.

In the paper included in this dissertation (Section 3.6), the attention is focused on analysis of microscopic images of stomata structures. The research aims to verify whether it is possible to classify the ploidity level based on images of stomata cells. This research is important in biofertilizer production as ploidy level impacts the plant physiology including size, density of stomata cells [2, 121]. Understanding these relations and distinguishing these classes allows optimal biofertilizer selection to specific plant genotypes. The research is performed on microscopic images of blackcurrant plants including diploidy, triploidy and tetraploidy. The YOLO algorithm is applied to segment the images and both classical

machine learning methods and CNNs are employed for classification.

This issue is extended by the last work (Section 3.7), where additional dataset is prepared. The aim is to ensure that the classification models do not perform satisfactorily only on one dataset, drawing conclusions based on irrelevant or misleading, dataset-specific image details. The experiments are carried out on one set and model performance is verified on another set. The classification methods employed here are CNNs and Vision Transformers. In this research, it is analyzed whether ViTs applied to stomata images perform well as it was the first application of these method to such images in any context. Additionally, Grad-CAM method is applied to selected models to verify which areas of images are important in decision-making process of considered classifiers.

1.4 Overview of state of the art

This section surveys the existing research relevant to this dissertation. It identifies established contributions, highlights current limitations and research gaps, situating the present work within the broader scholarly context.

1.4.1 Classification of bacteria microscopic images

There are numerous works which focus on computerized analysis and classification of microorganisms based on microscopic images in fields such as medicine, food, or environment [122]. The domain of agricultural microorganisms includes research which focuses on soil bacteria; however, the research in this specific area of is very limited.

The literature contains works which present classification of soil bacteria as part of the considered set e.g., in [123, 124] the dataset consists of both soil bacteria and fungi, in [125, 126] soil is one of sources for the considered microorganisms. In works [123, 124, 127, 128] the selected microorganisms are explored using fluorescence microscopy, while works included in this dissertation aim to work on raw images. In [126] the classified bacteria samples were stained using the Gram method. Some works focus on classification of various morphotypes [125] and not on species or genera classification. Nevertheless, it is important to analyse which methodologies were applied in bacteria classification and analysis of microorganisms in works which consider similar problems related to bacteria living in soil and other environments. In further sections, the diversity of methodologies applied in analysis of microscopic images of bacteria, including classification of soil bacteria, is presented.

In overall bacteria classification with microscopic images various methods have been applied. The workflow of this methodology consists of two crucial steps: feature extraction and classification. Additionally, image preprocessing and image segmentation could be applied to enhance feature calculation.

In research based on microscopic bacteria analysis, features extracted to describe bacteria are based on shape, texture, and color. Features based on shape such as width, length, diameter, volume were applied in [129] to describe differences between cocci, rods and other bacteria morphotypes in the overall biomass. In [130] shape features including area, perimeter, angle, curvature, circularity, features related to distances and radii, Hu Moments were computed to classify bacteria to three groups: *Hyphomicrobium* and *Methylotrophus* which live in water, and *Heterotrophs* which describes various bacteria types. Shape features such as area, perimeter, feret diameter, major, and minor axis,

roundness, elongation, curvature, aspect ratio and Fourier Descriptors were applied in [125] CMEIAS system where 11 bacteria morphotypes (cocci, spiral, curved rods, etc.) are classified which include bacteria living in environments such as soil, animal digestive systems, and bioreactors. Texture features based on GLCM and GRLEM are computed in order to classify five selected bacteria species: *Bacillus thuringiensis*, *Escherichia coli* K12, *Lactobacillus brevis*, *Listeria innocua*, and *Staphylococcus epidermidis* (each living in different environments including living organisms or soil) with application of fluorescence microscopy [128]. The same set of bacteria was classified with different set of features including shape (e.g., area and perimeter) and color descriptors (e.g., mean gray-level intensity and selected percentile values of gray-level histogram) in fluorescent images [127]. Color features computed in RGB, CIELAB, HSV color spaces such as mean, variance, kurtosis, skewness along with geometric descriptors based on clusters were employed to classify 12 soil microorganisms including bacteria and fungi [123], and in the subsequent work [124], authors additionally considered a class with a mix of five soil bacteria.

The classification methods which are applied to bacteria can either rely on the features computed manually, which are selected by experts, or employ algorithms to extract features automatically. Classical machine learning approaches are applied in bacteria classification in various researches. In [127, 128], authors perform statistical analysis of the considered classes. In [130], Threshold-Based Classifiers were employed by applying thresholds to handcrafted features. In [125], authors used KNN classifier to classify bacteria morphotypes applying shape features. In [131], Support Vector Machines were employed for classification of heterotrofic bacteria colonies with geometric features. Artificial Neural Networks (namely Probabilistic Neural Networks) are employed in classification of five soil bacteria in [127]. KNN, Random Forest, SVM and Multilayer Perceptron are applied to classify soil microorganisms in [123, 124]. Since 2017, various architectures of deep neural networks are widely being applied to microscopic bacteria images [132]. In [126], the features of 33 bacteria genera included in DIBaS dataset were computed with CNN architectures (AlexNet, VGG-M, and VGG-VD trained on ImageNet) along with handcrafted features, and the employed classifiers were SVM and Random Forest. In [133], InceptionNet was applied with transfer learning to feature extraction and classification of five bacteria species living in various environments and which are harmful to human health. In [134], ResNet was applied in bacteria colony segmentation in medical imaging. Vision Transformers are also applied to bacteria classification. ViT model trained on ImageNet-21k dataset is employed in [135] where three image datasets: KMC, CellDirect, DIBaS are classified as: Gram-Positive Cocci, Gram-Negative Bacilli and neutrophils.

The research on soil bacteria classification is very limited. The public datasets of raw soil bacteria microscopic images are unavailable. The research on soil bacteria is performed on stained samples or with fluorescence microscopy. There are many works that focus on automatic bacteria classification (e.g., morphology, bacteria living in other environments) and based on the methodology applied in these works we decided to fill the research gaps in soil bacteria domain. We decided to perform experiments on soil bacteria classification applying various digital image processing and classification methods. We perform our experiments on manually and on automatically created datasets.

1.4.2 Classification of stomata microscopic images

In literature, there are no articles which focus on ploidy level classification based on microscopic images of stomata with application of computerized approach. Nonetheless, there are numerous works that focus on analysis of stomata microscopic images with application of AI methods in other context [136, 137]. Tasks preformed on stomata microscopic images include: object detection [138, 139], segmentation [140, 141], and classification (open or closed cells [142], species [143], presence or absence of the stomata cells [144]).

Deep learning methods are applied in all of the considered tasks. Architectures of YOLO algorithm are used in stomata detection. In [138], YOLO (versions v3, v4 and v5) are employed to detect stomata cells on soyabean leaves. Single Shot Multibox Detector with MobileNetV1 backbone was applied in [139] to stomata detection in oil palm. Detection and semantic segmentation of stomata cells on various plant species was performed in [140], with application of Mask R-CNN. Semantic segmentation of stomata cell type in wheat and poplar dataset is performed in [141] with U-Net architecture. In [143] classification of 11 mangrove and freshwater swamp forest tree species was performed based on microscopic images of stomata with application of EfficientNetV2, Xception, VGG16, VGG19, MobileNetV2, ResNet50V2, Resnet152, DenseNet201, and NasNetLarge classifiers. Classification to open or closed stomata is performed in [142], where stomata located in microscopic images of leaf epidermal peels are classified to open or closed, applying Mask R-CNN with Faster R-CNN object detection backbone.

Although there are numerous applications of deep learning methods to stomata related issues the classical machine learning classification method application in this field is very limited. Classical machine learning methods are used in stomata analysis in [144], where authors employed both handcrafted and deep learning features extracted with Convolutional Neural Networks (DenseNet121, InceptionV3, InceptionResNetV2, MobileNet, NasNetMobile, VGG16) to describe stomata microscopic images of maize leaves. These features were applied as input to classical ML algorithms: SVM and MLP. In this article, both detection of stomata and classification of stomata presence or absence was performed. To the best of our knowledge there are no articles including application of Vision Transformers to stomata images in any context.

There are no works that focus on automatic ploidy level classification based on stomata microscopic images. There are numerous works that analyze stomata structures with computerized approach, the methodology applied in these works inspired us to perform our research. The application of classical machine learning classification methods to stomata images is very limited, and this research is focused on application of CNNs; however, the Transformers architectures have not been employed to stomata images. We decided to fill the research gap and perform automatic ploidy level classification based on stomata microscopic images. We used not only methodologies applied in other stomata researches, but also employed classical machine learning classification methods and Transformers architectures.

1.5 Thesis main contributions

This section presents the principal contributions of this dissertation. The publications comprising the thesis are summarized and organized into logically interconnected thematic groups, providing a coherent synthesis of the work's contribution to the field.

1.5.1 Computerized classification of soil bacteria microscopic images

Research on the computerized classification of microscopic soil bacteria images is limited in the literature. This dissertation presents a variety of experiments performed on a dataset prepared by microbiology experts which include microscopic images of five selected soil bacteria genera: *Enterobacter*, *Rhizobium*, *Pantoea*, *Bradyrhizobium*, and *Pseudomonas* (Section 3.1, 3.2, 3.3). The experiments consider a vast diversity of computed features and classification methods. The features considered in classification focus on such aspects as: color, geometry, dispersion, and texture. The features were computed on whole images, on subimages including single bacteria cells, or on individual cells themselves. The considered classification methods are: K Nearest Neighbors, Random Forest, Support Vector Machine, Multilayer Perceptron, and Extreme Learning Machines. Deep learning methods are also applied including Residual Neural Networks architectures such as: ResNet50, ResNet50V2, ResNet101, ResNet101V2, ResNet152, and ResNet152V2. Classification was performed either on whole images or on subimages, and then the majority voting approach was applied.

The presented results computed on the considered dataset yield that the highest accuracy obtained when solely color features are applied yields 95.6% with KNN (Section 3.2). However, it is important to note that preparing this dataset it was not ensured sustainability when it comes to such aspects as light, which could impact color of samples within the dataset. Due to that fact other types of features were considered and yielded 97% ACC when texture and shape features were considered applying KNN and ELM-RBF, separately (Section 3.1). Selected geometry and dispersion features combined yielded 91.56% with Random Forest (Section 3.2). In the evaluation of ResNet architectures on this dataset, ResNet152V2 reached 92% accuracy (Section 3.3).

1.5.2 Computerized classification of automatically acquired soil bacteria microscopic images

A diverse dataset of soil bacteria microscopic images was prepared automatically with KEYENCE VHX-7000N Digital Microscope which include over 3000 images taken under different light conditions and with application of different microscopic glasses including four selected bacteria genera *Enterobacter*, *Rhizobium*, *Pantoea*, and *Pseudomonas* (Section 3.5). Classification was performed on this dataset and on various subsets of this dataset applying classification methods: Support Vector Machine, Random Forest, K Nearest Neighborhood, MultiLayer Perceptron, Extreme Learning Machine Radial Basis Function, and ResNet152V2. The features computed in this research which were employed as input to classical machine learning methods included color and texture features.

Classification performed on whole dataset yielded the highest accuracy obtained with ELM-RBF with k-medoids (Section 3.4) 95.54% when handcrafted features were considered (Section 3.5). However, the highest overall accuracy was equal to 99% upon application of ResNet152V2.

In the next part of experiment (Section 3.5), the models were trained and tested on different subsets of whole dataset which were prepared in specific conditions. It turned out that the classification performance decreases significantly - especially when different glasses including different smears are compared (e.g., change from ACC 95% to 22%).

Under different lighting conditions, the decrease of accuracy was also noticeable; however, it was not that significant (e.g., change from ACC 99% to 89%).

1.5.3 Method modification and its application

A new method which is a modification od Extreme Learning Machine Radial Basis Function is presented (Section 3.5). The idea lies in replacing the default k-means method applied for clustering of input data with mean shift and k-medoids algorithms. Experiments include different numbers of neurons, activation functions, values of k in k-means and k-medoids, different values of radii and kernel configurations in mean shift. The ELM-RBF with k-medoids yields satisfactory results when tested on benchmark datasets: Wine Quality-White (Mean Squared Error 0.619) and Ionosphere (94.6% accuracy). In further research, various activation functions are considered and various distance metrics are analyzed in k-means and k-medoids.

The ELM-RBF with k-medoids is successfully applied for classification purposes in the work where automatically taken dataset of microscopic images of bacteria is considered, and obtains the highest accuracy result among other methods which considered handcrafted features yielding 95.54% ACC.

1.5.4 Classification of ploidy levels in plants based on stomata microscopic images

The state-of-the-art research is performed on classification of ploidy levels in plants based on microscopic images of stomata. Dataset consists of three classes: diploid, triploid, and tetraploid. For the best of our knowledge the first time AI methods were employed to classify ploidy levels.

In first work (Section 3.6), experiments are performed on a single dataset of microscopic images including stomata cells. The experiments include analysis of stomata cells and whole microscopic images. YOLO algorithm is employed for ploidy level segmentation. Deep learning methods including Residual Neural Networks and methods such as KNN, SVM, RF, MLP are employed for the classification, and the classification is performed on various datasets including whole images, subimages and masks. The highest classification accuracy equal to 97.3% is obtained with ResNet152V2. However, classes in this dataset differed in shade of green, so to eliminate impact of color other sets considered of processed images were used, and then the classification accuracy decreased e.g., for set of image masks the obtained accuracy yielded 78.7% with ResNet152V2.

In order to ensure that the classification performance does not rely solely on some traits of this dataset (e.g., shade of green) in second work (Section 3.7) two separate raw datasets prepared in different times of vegetation season are considered. Then, the classification model is trained on one dataset and tested on another, ensuring that the model indeed learns to distinguish between ploidy levels and not between specific classes in one dataset. In this research, ResNets and ViTs are employed. The highest accuracy, when three classes are considered, is equal to 68% obtained with ResNet152V2, and in binary classification problem Vision Transformer architecture *vit-base-patch16-224-in21k* yielded the highest ACC equal to 88%.

1.5.5 Application of Vision Transformers to stomata images

In this dissertation (Section 3.7), Vision Transformers are applied for the first time to microscopic stomata images. Microscopic stomata images are used in various researches including stomata counting, measuring, dynamics tracking, verification whether these structures are open or closed. However, the ViTs had not yet been employed in any of the research experiments in the literature.

The Vision Transformers are applied along with Residual Neural Networks. In three class classification the ResNets outperform ViTs; however, in two class classification better results are yielded with ViTs. Grad-CAM method was applied to both ResNets and ViTs in classification task, in order to verify how these methods make classification decisions considering whole microscopic images including stomata cells. As it turns out, ResNets incorporate global image features in their classifications, while ViTs demonstrated biologically interpretable behavior, consistently focusing on stomata structures. This result shows the potential of ViTs application to stomata-related tasks.

Chapter 2

Overview of thesis and publications

This chapter provides an overview of the publications that form the foundation of this dissertation and delineates the contribution of each author. For each included work, a summary is presented, covering the methodology and key findings.

2.1 Publications constituting to the thesis

1. *Classification of soil bacteria based on machine learning and image processing*

   A. Konopka, K. Struniawski, R. Kozera, P. Trzciński, L. Sas-Paszt, A. Lisek, K. Górnik, E. Derkowska, S. Głuszek, B. Sumorok, and M. Frąc, "Classification of soil bacteria based on machine learning and image processing," in *Computational Science – ICCS 2022: 22nd International Conference, London, UK, June 21–23, 2022, Proceedings, Part III, Lecture Notes in Artificial Intelligence*, vol. 13352, pp. 263–277, Springer, 2022. https://doi.org/10.1007/978-3-031-08757-8_23 (140 MNiSW points)

   Contribution of Aleksandra Konopka:

   - software when it comes to segmentation of region of interest to extract specific bacteria cells, computation of features based on shape, planning end performing classification experiments with classical machine learning methods on features based on shape and texture, application of feature selection methods,
   - writing following sections or subsections of original draft: abstract, introduction, work-flow scheme, segmentation of region of interest, feature generation, feature selection, class recognition, features based on shape, KNN, experiments and results, conclusions,
   - reviewing and editing whole manuscript,
   - data curation,
   - conceptualization,
   - methodology.

   Contribution of other authors:

   - Karol Struniawski - software - Extreme Learning Machine and Extreme Learning Machine RBF, computations of features based on texture, writing sections

in original draft: features based on texture, ELM-RBF, reviewing and editing whole manuscript,

- Ryszard Kozera - supervision on the IT side, reviewing and editing whole manuscript especially in context of mathematical notations,
- Paweł Trzciński - preparation of the dataset,
- Lidia Sas-Paszt - supervision of the biological aspects, reviewing and editing the biological aspects,
- Anna Lisek, Krzysztof Górnik, Edyta Derkowska, Sławomir Głuszek, Beata Sumorok and Magdalena Frąc - reviewing and editing the biological aspects.

Percentage contribution of authors: Aleksandra Konopka - 60%, Karol Struniawski - 10%, Ryszard Kozera - 10%, Paweł Trzciński - 8%, Lidia Sas-Paszt - 6%, Anna Lisek - 1%, Krzysztof Górnik - 1%, Edyta Derkowska - 1%, Sławomir Głuszek - 1%, Beata Sumorok - 1%, and Magdalena Frąc - 1%.

2. *Identification of the selected soil bacteria genera based on their geometric and dispersion features*

   A. Konopka, R. Kozera, L. Sas-Paszt, P. Trzcinski, and A. Lisek, "Identification of the selected soil bacteria genera based on their geometric and dispersion features," *PLoS ONE*, vol. 18, no. 10, e0293362, pp. 1–11, 2023. https://doi.org/10.1371/journal.pone.0293362 (100 MNiSW points, IF – 2.9)

   Contribution of Aleksandra Konopka:

   - conceptualization,
   - data curation,
   - formal analysis,
   - methodology,
   - software,
   - visualization,
   - writing original draft, reviewing and editing.

   Contribution of other authors:

   - Ryszard Kozera - supervision of the IT part, we cooperated in conceptualization in context of interpolation methods, we cooperated on writing original draft in context of presenting the geometric features with proper mathematical formulas, writing - reviewing and editing,
   - Lidia Sas-Paszt - supervision in biological aspects, resources,
   - Paweł Trzciński - resources, validation of biological descriptions,
   - Anna Lisek - resources, validation of biological descriptions.

   Percentage contribution of authors: Aleksandra Konopka - 75%, Ryszard Kozera - 10%, Lidia Sas-Paszt - 8%, Paweł Trzciński - 5%, and Anna Lisek - 2%.

3. *Performance Analysis of Residual Neural Networks in Soil Bacteria Microscopic Image Classification*

   A. Konopka, K. Struniawski, and R. Kozera, "Performance analysis of Residual Neural Networks in soil bacteria microscopic image classification," in *Modelling and Simulation'2023: The 2023 European Simulation and Modelling Conference, October 24–26, 2023, Toulouse, France*, pp. 144–149, EUROSIS-ETI, 2023. (70 MNiSW points)

   Contribution of Aleksandra Konopka:

   - conceptualization,
   - data curation,
   - formal analysis,
   - methodology,
   - software,
   - visualization,
   - writing - original draft, reviewing and editing.

   Contribution of other authors:

   - Karol Struniawski - software - supported the application of ResNet models, writing - reviewing and editing,
   - Ryszard Kozera - supervision, writing - reviewing and editing.

   Percentage contribution of authors: Aleksandra Konopka - 80%, Karol Struniawski - 10%, and Ryszard Kozera - 10%.

4. *Classification performance of Extreme Learning Machine Radial Basis Function with k-means, k-medoids and mean shift clustering algorithms*

   A. Konopka, K. Struniawski, and R. Kozera, "Classification performance of Extreme Learning Machine Radial Basis Function with k-means, k-medoids and mean shift clustering algorithms," in *Computational Science – ICCS 2023: 23rd International Conference, Prague, Czech Republic, July 3–5, 2023, Proceedings, Part IV, Lecture Notes in Artificial Intelligence*, vol. 10476, pp. 171–186, Springer, 2023. https://doi.org/10.1007/978-3-031-36027-5_13 (140 MNiSW points)

   Contribution of Aleksandra Konopka:

   - conceptualization,
   - data curation,
   - formal analysis,
   - methodology,
   - software,
   - visualization,
   - writing - original draft, reviewing and editing.

   Contribution of other authors:

- Karol Struniawski - software - supported implementation of raw Extreme Learning Machine Radial Basis Function, writing - reviewing and editing,
- Ryszard Kozera - supervision, writing - reviewing and editing.

Percentage contribution of authors: Aleksandra Konopka - 80%, Karol Struniawski - 10%, and Ryszard Kozera - 10%.

5. *Automated imaging and machine learning for soil bacteria classification: Challenges and insights*

   A. Konopka, R. Kozera, L. Sas-Paszt, and P. Trzciński, "Automated imaging and machine learning for soil bacteria classification: Challenges and insights," *Engineering Applications of Artificial Intelligence*, vol. 159C, pp. 1–9, 2025. https://doi.org/10.1016/j.engappai.2025.111369 (140 MNiSW points, IF – 8.0)

   Contribution of Aleksandra Konopka:

   - conceptualization,
   - data curation,
   - formal analysis,
   - methodology,
   - software,
   - visualization,
   - writing - original draft, reviewing and editing,
   - prepared the microscopic specimens and acquired the images.

   Contribution of other authors:

   - Ryszard Kozera - supervision, writing - reviewing and editing, formal analysis,
   - Lidia Sas-Paszt - writing - reviewing and editing, supervision, conceptualization,
   - Paweł Trzciński - writing – review and editing, resources - provided bacteria strains representing various genera.

   Percentage contribution of authors: Aleksandra Konopka - 80%, Ryszard Kozera - 8% and Lidia Sas-Paszt - 7%, and Paweł Trzciński - 5%.

6. *Deep Learning Classification of Blackcurrant Genotypes by Ploidy Levels on Stomata Microscopic Images*

   A. Konopka, K. Struniawski, R. Kozera, L. Ortenzi, A. Marasek-Ciołakowska, and A. Machlańska, "Deep learning classification of blackcurrant genotypes by ploidy levels on stomata microscopic images," in *Computational Science – ICCS 2025 Workshops: 25th International Conference, Singapore, Singapore, July 7–9, 2025, Proceedings, Part III, Lecture Notes in Artificial Intelligence*, pp. 135–148, Springer, 2025. https://doi.org/10.1007/978-3-031-97564-6_11 (140 MNiSW points)

Contribution of Aleksandra Konopka:

- conceptualization,
- data curation,
- formal analysis,
- methodology,
- software,
- visualization,
- writing - original draft, reviewing and editing.

Contribution of other authors:

- Karol Struniawski - writing - reviewing and editing, help with image labeling,
- Ryszard Kozera - writing - reviewing and editing, supervision,
- Luciano Ortenzi - conceptualization, writing - reviewing and editing,
- Agnieszka Marasek-Ciołakowska - conceptualization, writing - original draft - plant materials description, reviewing and editing, resources, supervision, microscopic image dataset preparation,
- Aleksandra Machlańska - writing - reviewing and editing, resources, microscopic image dataset preparation.

Percentage contribution of authors: Aleksandra Konopka - 60%, Karol Struniawski - 8%, Ryszard Kozera - 8 %, Luciano Ortenzi - 8%, Agnieszka Marasek-Ciołakowska - 8%, and Aleksandra Machlańska - 8%.

7. *Classification of Blackcurrant Genotypes by Ploidy Levels on Stomata Microscopic Images with Deep Learning: Convolutional Neural Networks and Vision Transformers*

   A. Konopka, R. Kozera, A. Marasek-Ciołakowska, and A. Machlańska, "Classification of blackcurrant genotypes by ploidy levels on stomata microscopic images with deep learning: Convolutional Neural Networks and Vision Transformers," *Applied Sciences-Basel*, vol. 15, no. 19, 10735, pp. 1–14, 2025. https://doi.org/10.3390/app151910735 (100 MNiSW points, IF – 2.5)

   Contribution of Aleksandra Konopka:

   - conceptualization,
   - data curation,
   - formal analysis,
   - methodology,
   - software,
   - validation,
   - investigation,
   - visualization,
   - writing - original draft, reviewing and editing,

- project administration.

Contribution of other authors:

- Ryszard Kozera - validation, writing - review and editing, supervision,
- Agnieszka Marasek-Ciołakowska - conteptualization, investigation, resources, writing - review and editing,
- Aleksandra Machlańska - resources, writing - review and editing.

Percentage contribution of authors: Aleksandra Konopka - 70%, Ryszard Kozera - 10%, Agnieszka Marasek-Ciołakowska - 10%, and Aleksandra Machlańska - 10%.

2.2 Synopsis of research articles constituting the thesis

1. *Classification of soil bacteria based on machine learning and image processing*

   A. Konopka, K. Struniawski, R. Kozera, P. Trzciński, L. Sas-Paszt, A. Lisek, K. Górnik, E. Derkowska, S. Głuszek, B. Sumorok, and M. Frąc, "Classification of soil bacteria based on machine learning and image processing," in *Computational Science – ICCS 2022: 22nd International Conference, London, UK, June 21–23, 2022, Proceedings, Part III, Lecture Notes in Artificial Intelligence*, vol. 13352, pp. 263–277, Springer, 2022. https://doi.org/10.1007/978-3-031-08757-8_23 (140 MNiSW points)

   In this paper, the research problem was to verify how classical machine learning methods and ELM perform classification when handcrafted features based on shape and texture are considered. The image dataset was prepared manually by microbiologists and contained 128 images of five soil bacteria genera: *Enterobacter*, *Rhizobium*, *Pantoea*, *Bradyrhizobium*, and *Pseudomonas*.

   The classifiers considered in this research are: SVM, RF, KNN, MLP, ELM, and ELM-RBF. Feature selection methods applied in this research are: FCBF, SBMLR, and CFS. The features set consisted of 6 features based on shape calculated with application of interpolation presenting values of angles, curvature and arc-length computed on selected bacteria instances extracted from images and 19 texture-based features computed based on GLCM and GLRLM matrices. The experiments were repeated 50 times with 10-fold cross-validation was applied. The accuracy results were computed on texture and geometry features combined, and were calculated separately on each of these sets. For classifications based solely on texture features, the highest ACC - 82.27% was obtained with KNN with SBMLR. On the other hand, when classification was performed on shape features the highest ACC was equal to 78.92% when SVM was applied with CFS. After combining both sets of features the highest yielded ACC was 97.07% for KNN with FCBF and second highest result was 97.03% for ELM-RBF with SBMLR.

   The conclusion from this research is that combining shape and texture features yielded the highest classification accuracy, and application of feature selection methods increased the accuracy on the considered set of microscopic images.

2. *Identification of the selected soil bacteria genera based on their geometric and dispersion features*

   A. Konopka, R. Kozera, L. Sas-Paszt, P. Trzcinski, and A. Lisek, "Identification of the selected soil bacteria genera based on their geometric and dispersion features," *PLoS ONE*, vol. 18, no. 10, e0293362, pp. 1–11, 2023. https://doi.org/10.1371/journal.pone.0293362 (100 MNiSW points, IF – 2.9)

   In this work, the research problem was to verify how classical machine learning methods perform classification when handcrafted features based on shape and dispersion are considered, and whether application of these features outperform the results obtained when color features are applied. The image dataset was prepared manually by microbiologists and contained 128 images of five soil bacteria genera: *Enterobacter*, *Rhizobium*, *Pantoea*, *Bradyrhizobium*, and *Pseudomonas*.

The classifiers considered in this research are: SVM, RF, KNN, and MLP. Feature selection methods applied in this research are: FCBF, SBMLR, and CFS. The feature set consisted of 10 geometric features computed based on angles, distances in bacteria instances, their amount and area, and 17 dispersion features computed with mean shift algorithm and k-means combined with regression. Additionally, 32 color features were computed as a reference point. The classification was performed 50 times with 10-fold cross-validation and different sets of features and their combinations were used. The computations were performed on all 5 bacteria genera and on 4 selected bacteria genera (all possible 5 combinations) to verify if there is a specially problematic class causing the decrease of accuracy. For all five bacteria genera, the results show that application of feature selection methods did not increase accuracy when geometry and dispersion features were combined. When geometry and dispersion features were considered together, the highest accuracy was equal to 85.14% employing Random Forest classifier.

In four-genera classification, the highest accuracy obtained with geometry and dispersion features combined was equal to 91.55% yielded with Random Forest, when the *Bradyrhizobium* genera was not considered - this class turned out to be the most problematic to classify. The second best classification result was also obtained with Random Forest and was equal to 88.43%.

For the classification of all five bacteria genera, different combinations of feature sets were tested. Applying Random Forest, the geometry and dispersion features combined (85.14% accuracy) did not yield better classification results than color features (94.3%). However, combining dispersion, geometry, and color features together increased the accuracy to 94.82% for Random Forest. The highest accuracy obtained with color features alone was 95.6% yielded with KNN.

3. *Performance analysis of Residual Neural Networks in soil bacteria microscopic image classification*

   A. Konopka, K. Struniawski, and R. Kozera, "Performance analysis of Residual Neural Networks in soil bacteria microscopic image classification," in *Modelling and Simulation'2023: The 2023 European Simulation and Modelling Conference, October 24–26, 2023, Toulouse, France*, pp. 144–149, EUROSIS-ETI, 2023. (70 MNiSW points)

   In this paper, it was decided to verify the performance of Convolutional Neural Networks when applied to classify microscopic images of selected soil bacteria genera. The image dataset was prepared manually by microbiologists and contained 128 images of five soil bacteria genera: *Enterobacter*, *Rhizobium*, *Pantoea*, *Bradyrhizobium*, and *Pseudomonas*. The dataset was partitioned into three distinct groups: training, validation and testing sets in a $6:3:1$ ratio. From each of the microscopic images 50 subimages with bacteria instances were extracted. Classification of each of the subimages was perfomed with six different ResNet architectures (ResNet50, ResNet50V2, ResNet101, ResNet101V2, ResNet152, and ResNet152V2). Transfer learning was employed and the parameters were fine-tuned in selected layers. Then majority voting was applied to subimages deciding to which of the class each image (from which the subimages were derived) should be assigned. The highest obtained accuracy results was achieved when majority voting was applied with ResNet152V2 and ResNet101V2, and was equal to 92% while the worst performance was achieved

by ResNet152 yielding 42%. Application of majority voting increased the overall accuracy by 16 percentage points.

Residual Neural Networks perform well in classifying soil bacteria genera. Higher accuracy is obtained with the V2 architecture variant and the application of majority voting to subimages increases classification performance.

4. *Classification performance of Extreme Learning Machine Radial Basis Function with k-means, k-medoids and mean shift clustering algorithms*

   A. Konopka, K. Struniawski, and R. Kozera, "Classification performance of Extreme Learning Machine Radial Basis Function with k-means, k-medoids and mean shift clustering algorithms," in *Computational Science – ICCS 2023: 23rd International Conference, Prague, Czech Republic, July 3–5, 2023, Proceedings, Part IV, Lecture Notes in Artificial Intelligence*, vol. 10476, pp. 171–186, Springer, 2023. https://doi.org/10.1007/978-3-031-36027-5_13 (140 MNiSW points)

   In this work, a new modification of ELM-RBF is compared with the traditional approach and it is verified whether it yields better results then the standard algorithm. In ELM-RBF the input dataset is clustered and then transformed into a new feature space with RBF kernel function. Then the modified input data are passed to standard ELM network. In this work, a new modification of ELM-RBF is proposed replacing default clustering method - k-means - with alternative algorithms k-medoids and mean shift. The performance proposed methodology is tested on Wine Quality-White (Wine) (regression) and Ionosphere (IS) (classification) benchmark datasets. The performance is measured with Mean Squared Error and accuracy, for the datasets, respectively.

   All the computations applied 5-fold cross-validation repeated 10 times. The computations were performed for selected values of k from 10 to 100 for k-means and k-medoids. Different numbers of neurons were considered between 100 and 5000. In mean shift method, different values of window radious were considered from 0.01 up to 1.5 and flat or Gaussian kernels were used. Softplus and linear activation functions in ELM-RBF were tested. The lowest MSE on Wine for k-medoids was equal to 0.619 ($n = 700$, $k = 100$, linear), 0.62 for k-means ($n = 700$, $k = 100$, linear) and 0.686 for mean shift ($n = 700$, $r = 0.6$, Softplus, flat). The highest accuracy on IS dataset was equal to 94.6% for k-medoids ($n = 100$, $k = 96$, linear), 94.2% for k-means ($n = 200$, $k = 80$, linear), and 0.892 for mean shift ($n = 100$, $r = 0.51$, linear, flat).

   In second part of experiment, additional computations were performed on the best two results yielded on IS dataset obtained with k-means ($n = 100$, $k = 96$) and k-medoids ($n = 200$, $k = 80$) employing different distances in k-means and k-medoids and various activation functions in the network. The computations revealed that the best performance in both k-means (94.2%) and k-medoids (94%) is achieved with linear activation function and cityblock algorithm for distance computations.

   This work revealed that new proposed method ELM-RBF with k-medoids yields similar and better results than ELM-RBF with traditional k-means implementation. It is important to note, that these algorithms are stochastic which impacts the performance (e.g., random values of weights in ELM) that is why in second part of experiment the highest ACC was obtained by k-means and in the first part of

experiment ELM-RBF with k-medoids outperformed ELM-RBF with k-means on both datasets.

5. *Automated imaging and machine learning for soil bacteria classification: Challenges and insights*

   A. Konopka, R. Kozera, L. Sas-Paszt, and P. Trzciński, "Automated imaging and machine learning for soil bacteria classification: Challenges and insights," *Engineering Applications of Artificial Intelligence*, vol. 159C, pp. 1–9, 2025. https://doi.org/10.1016/j.engappai.2025.111369 (140 MNiSW points, IF – 8.0)

   In this work, the classification was performed on a new dataset. This dataset was prepared automatically with KEYENCE VHX-7000N Digital Microscope which enabled to prepare a large dataset consisting of over 3,000 high-resolution images. The goal of this research was to verify classification performance of bacteria genera when dataset of images is not prepared manually by experts, but automatically by IT specialists (after being instructed by experts on microscope usage). This research did not focus on a homogeneous dataset, but the samples were photographed under two different light conditions and three different microscopic glass slides were applied. The considered soil bacteria genera were: *Enterobacter*, *Rhizobium*, *Pantoea*, and *Pseudomonas*. Considered classification methods were classical machine learning algorithms: Support Vector Machine, Random Forest, KNN, MLP, and more advanced neural networks: ELM-RBF (including the modification proposed in previous work) and ResNet152V2. For classical machine learning methods and ELM-RBF the set of features consisted of 30 color and texture features. ELM-RBF with k-medoids obtained the highest accuracy on whole dataset 95.54% comparing to other methods. However, when ResNet152V2 was applied and whole images were used as input to this CNN the obtained accuracy was 99%, outperforming other methods.

   In the second part of experiment, ResNet152V2 was trained and tested on various subsets of whole dataset. It turned out that classification performance is decreasing to a great extent when model is trained in one set of conditions and is tested on another one. This difference is very large when we compare results yielded when in training and testing sets different glasses are considered (e.g., change from ACC 95% to 22%). However, when model is trained and tested in different light conditions the difference in accuracy is not that significant (e.g., change from ACC 99% to 89%).

   The conclusions that could be drawn from these results are: the same samples were photographed under two light conditions so ResNet learns to distinguish patterns in these samples well. To prevent artificially better classification results, the samples were moved after each light change so that the images would not cover exactly the same area. As the classification ACC is relatively good when the same samples are considered (only light changes), it means that either the glasses have a great impact on classification or the way samples were prepared (smear) was created impacts classification performance. We could not 'copy and paste' smear from one glass to another to compare results.

   In the next part of work, experiments were performed on subimages. Majority voting was applied to classify whole images based on how subimages, derived from these images, were classified. In most of the scenarios (different combinations of

training and testing sets), application of majority voting increased ACC comparing to results obtained on subimages without application o MV (up to 11 percentage points increase).

In the last part of experiment, two sets of samples were prepared in the same light conditions. Two glasses were considered yielding 16 samples: 4 bacteria genera, on 2 glasses, in one light condition and repeated twice. The aim of this part of research was to verify if we prepare two sets of samples in identical conditions will ResNet model be able to distinguish well both samples. It turned out that the classification results are much worse when considering one set of samples created in one set conditions for testing and another set of samples created in the same set of conditions for training (e.g., accuracy decrease from 90% to 21%). This experiment suggests that there is a need to develop an automatic procedure to create samples to ensure that aspects such as sample thickness and smear shape will not have an impact on classification accuracy.

6. *Deep learning classification of blackcurrant genotypes by ploidy levels on stomata microscopic images*

   A. Konopka, K. Struniawski, R. Kozera, L. Ortenzi, A. Marasek-Ciołakowska, and A. Machlańska, "Deep learning classification of blackcurrant genotypes by ploidy levels on stomata microscopic images," in *Computational Science – ICCS 2025 Workshops: 25th International Conference, Singapore, Singapore, July 7–9, 2025, Proceedings, Part III, Lecture Notes in Artificial Intelligence*, pp. 135–148, Springer, 2025. https://doi.org/10.1007/978-3-031-97564-6_11 (140 MNiSW points)

   In this paper, the goal was to verify whether artificial intelligence is capable of distinguishing ploidy levels in plants based on microscopic leaf images. The dataset consisted of microscopic images of blackcurrant plants on three ploidity levels: diploidy, triploidy, and tetraploidy. A set of microscopic images including stomata cells was first labeled manually, then this process was continued with YOLO algorithm segmenting the images (locating stomata structures).

   The problem that needed to be solved was the difference in color among the classes (different shades of green) due to that reason many modifications of the dataset aiming to reduce impact of color were applied (e.g., grayscale or binary masks). Applying ResNets to subimages including single stomata cells and whole images, when color was not modified, the highest ACC was equal to 97.3% for whole images with ResNet152V2.

   Classical machine learning methods (KNN, SVM, RF, MLP employing 10-fold cross-validation) relied on geometry features of stomata instances ensuring the color will have no impact on classification performance. The highest accuracy among classical methods was obtained on *Image masks* dataset yielding 75.8% when Random Forest was used, for the same set ResNet152V2 obtained 78.7%.

   The ploidity level can be successfully distinguished based on microscopic images of stomata cells on this dataset even when color is not considered. This experiment needed to be extended to other datasets, to ensure the performance is not only successful on this set, because of specific traits of the photographed sample, but it also is reproducible on other datasets. This idea is implemented in the next work included in this dissertation.

7. *Classification of blackcurrant genotypes by ploidy levels on stomata microscopic images with deep learning: Convolutional Neural Networks and Vision Transformers*

   A. Konopka, R. Kozera, A. Marasek-Ciołakowska, and A. Machlańska, "Classification of blackcurrant genotypes by ploidy levels on stomata microscopic images with deep learning: Convolutional Neural Networks and Vision Transformers," *Applied Sciences-Basel*, vol. 15, no. 19, 10735, pp. 1–14, 2025. https://doi.org/10.3390/app151910735 (100 MNiSW points, IF – 2.5)

   In this work, Convolutional Neural Networks and Vision Transformers were applied to solve the ploidy classification with microscopic images issue. The experiments are carried out on two datasets of blackcurrant including diploidy, triploidy, and tetraploidy. These sets are prepared in different time of vegetation seasons. The models are trained on one set and tested on another to ensure that the results are not reliant on the specific characteristics of the training set. The aim to verify if models learn to distinguish ploidy levels or other traits of images, which rely for example on color of samples, as it varies between these sets due to differences in timing when they were captured. The considered classifiers were: ResNet152V2 and Vision Transformes: *vit-large-patch16-224-in21k* (vit-l) and *vit-base-patch16-224-in21k* (vit-b). The highest accuracy was obtained with ResNet152V2 - 68%. The most problematic class to classify turned out to be the class of triploidy. In two-class classification (diploid and tetraploid), the highest accuracy was obtained with vit-b yielding 88%. Application of augmentation techniques increased accuracy values of some models and helped with classification of the triploid class; however, it did not cause increase of maximal accuracy obtained on both three and two classes. Grad-CAM method was applied to verify which areas of image ResNets and ViTs find the most important in decision making process. It turned out that ViTs demonstrated biologically interpretable behavior, consistently focusing on stomata structures. In contrast, ResNets exhibited progressively broader attention patterns in deeper layers, incorporating more global image features in their classifications.

   The application of ViTs is a promising direction in stomata images analysis. This work was the first to apply ViTs to stomata images in any context (such as stomata counting, measuring or ploidy level classification). This work can be a crucial basis for other researches to employ ViTs to other tasks in stomata analysis such as e.g., analysis of impact on biofertilizer application.

Chapter 3

Included publications

3.1 Classification of soil bacteria based on machine learning and image processing

Publication: A. Konopka, K. Struniawski, R. Kozera, P. Trzciński, L. Sas-Paszt, A. Lisek, K. Górnik, E. Derkowska, S. Głuszek, B. Sumorok, and M. Frąc, "Classification of soil bacteria based on machine learning and image processing," in *Computational Science – ICCS 2022: 22nd International Conference, London, UK, June 21–23, 2022, Proceedings, Part III, Lecture Notes in Artificial Intelligence*, vol. 13352, pp. 263–277, Springer, 2022. https://doi.org/10.1007/978-3-031-08757-8_23 (140 MNiSW points)

Abstract: Soil bacteria play a fundamental role in plant growth. This paper focuses on developing and testing some techniques designed to identify automatically such microorganisms. More specifically, the recognition performed here deals with the specific five genera of soil bacteria. Their microscopic images are classified with machine learning methods using shape and image texture descriptors. Feature determination based on shape relies on interpolation and curvature estimation whereas feature recognition based on image texture resorts to the spatial relationships between chrominance and luminance of pixels using co-occurrence matrices. From the variety of modeling methods applied here the best reported result amounts to 97% of accuracy. This outcome is obtained upon incorporating the set of features from both groups and subsequently merging classification and feature selection methods: Extreme Learning Machine - Radial Basis Function with Sparse Multinomial Logistic Regression with Bayesian Regularization and also k-Nearest Neighbors classifier with Fast Correlation Based Filter. The optimal parameters involved in merged classifiers are obtained upon computational testing and simulation.

Classification of Soil Bacteria Based on Machine Learning and Image Processing

Aleksandra Konopka[1], Karol Struniawski[1(✉)], Ryszard Kozera[1,2], Paweł Trzciński[3], Lidia Sas-Paszt[3], Anna Lisek[3], Krzysztof Górnik[3], Edyta Derkowska[3], Sławomir Głuszek[3], Beata Sumorok[3], and Magdalena Frąc[4]

[1] Institute of Information Technology, Warsaw University of Life Sciences - SGGW, ul. Nowoursynowska 159, 02-776 Warsaw, Poland
{aleksandra_konopka,karol_struniawski,ryszard_kozera}@sggw.edu.pl
[2] School of Physics, Mathematics and Computing, The University of Western Australia, 35 Stirling Highway, Perth, Crawley, WA 6009, Australia
ryszard.kozera@uwa.edu.au
[3] Department of Microbiology and Rhizosphere, The National Institute of Horticultural Research, ul. Pomologiczna 18, 96-100 Skierniewice, Poland
{pawel.trzcinski,lidia.sas,anna.lisek,krzysztof.gornik,edyta.derkowska,slawomir.gluszek,beata.sumorok}@inhort.pl
[4] Institute of Agrophysics, Polish Academy of Sciences, ul. Doświadczalna 4, 20-290 Lublin, Poland
m.frac@ipan.lublin.pl

Abstract. Soil bacteria play a fundamental role in plant growth. This paper focuses on developing and testing some techniques designed to identify automatically such microorganisms. More specifically, the recognition performed here deals with the specific five genera of soil bacteria. Their microscopic images are classified with machine learning methods using shape and image texture descriptors. Feature determination based on shape relies on interpolation and curvature estimation whereas feature recognition based on image texture resorts to the spatial relationships between chrominance and luminance of pixels using co-occurrence matrices. From the variety of modelling methods applied here the best reported result amounts to 97% of accuracy. This outcome is obtained upon incorporating the set of features from both groups and subsequently merging classification and feature selection methods: Extreme Learning Machine - Radial Basis Function with Sparse Multinomial Logistic Regression with Bayesian Regularization and also k-Nearest Neighbors classifier with Fast Correlation Based Filter. The optimal parameters involved in merged classifiers are obtained upon computational testing and simulation.

Keywords: Soil bacteria · Machine learning · Image analysis · Shape and image texture extraction · Spline interpolation · Modelling and simulation · Computational optimization

D. Groen et al. (Eds.): ICCS 2022, LNCS 13352, pp. 263–277, 2022.
https://doi.org/10.1007/978-3-031-08757-8_23

1 Introduction

Soil bacteria despite of their small size may have a large impact on plant growth. Some of them are beneficial to agricultural sector, while the others are either harmless or pathogenic causing a vast diversity of plant diseases. Consequently, bacteria recognition becomes an important task for scientists equally as a research and agricultural problem. Bacteria identification is usually carried out using specific markers changing their color as a reaction to specific chemical compounds. The morphology of the bacteria colony is also usually analyzed by examining its shape, edges, color, colony distribution, consistency and surface structure [7]. This approach is usually laborious and depends on the subjective perceptiveness of the scientist. A natural step accelerating and facilitating the latter is to automate the process of microscopic image analysis. This paper[1,2] resorts to machine learning and image processing methods applied to soil bacteria recognition. In general, comparing images of bacteria belonging to certain species is difficult since they adopt similar morphologies [22]. Due to this reason, it is decided to distinguish here the input bacteria on the genera level. The microscopic images of soil bacteria examined in this paper (which are part of our data-set available in full resolution under the URL link: https://bit.ly/3qdDuHo) include pictures of *Enterobacter*, *Rhizobium*, *Pantoea*, *Bradyrhizobium* and *Pseudomonas* (see Fig. 1). The pictures of investigated bacteria are obtained from Symbio-Bank - the collection of microorganisms of The National Institute of Horticultural Research in Skierniewice. Some of *Enterobacter* are considered plant pathogens, whereas the others are conducive for plant growth [13]. The bacteria of the genus *Rhizobium* have a positive effect on increasing the yield of grains and the protein content in pea grains [23]. *Rhizobium* and *Bradyrhizobium* are nitrogen-fixing soil bacteria that live in symbiosis with legumes [3]. On the other hand, *Pantoea* causing plant infections [24] is also used in the production of antibiotics [2]. Some *Pseudomonas* are plant pathogens, while the others are used to stimulate plant growth and to remediate contaminated soil [30]. This paper discusses the identification of bacteria genera based on their morphological features. The calculated traits refer to bacteria shape and image texture. In order to automatize the entire recognition process a variety of feature selection and class recognition methods adapting the concept of machine learning are applied. On the basis of the supplied training data-set a classification model is built permitting to automatically categorize soil microorganisms.

2 Work-flow Scheme

The work-flow scheme adopted in this work consists of the following four consecutive steps: *segmentation of the Region of Interest, feature generation, feature selection* and *class recognition.*

[1] This research is financed by The National Centre for Research and Development of the BIOSTRATEG Project (Eco-Fruits) BIOSTRATEG3/344433/16/NCBR/2018.
[2] This work is a part of Polish National Centre of Research and Development research project POIR.01.02.00-00-0160/20.

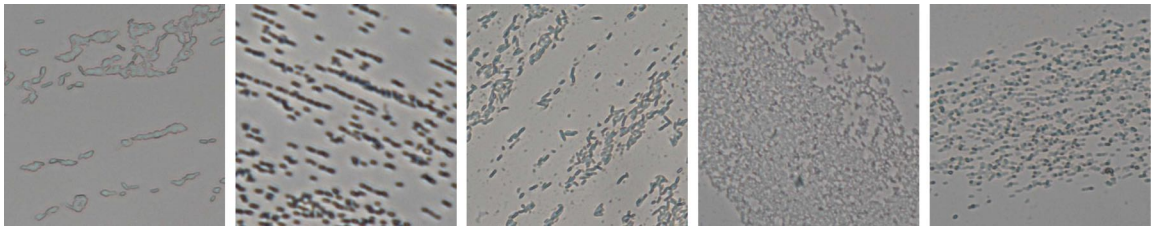


Fig. 1. Microscopic images of (from left): *Enterobacter*, *Rhizobium*, *Pantoea*, *Bradyrhizobium* and *Pseudomonas*. For more pictures see URL link: https://bit.ly/3qdDuHo.

2.1 Segmentation of the Region of Interest

The aim of this step is to extract bacteria and background image regions. Binary mask filter is applied yielding white pixels representing bacteria zones and black pixels corresponding to the background. To achieve the latter the image is first converted into gray-scale and then Otsu automatic image thresholding [18] with open and close morphological operations [28] is applied (see Fig. 2).

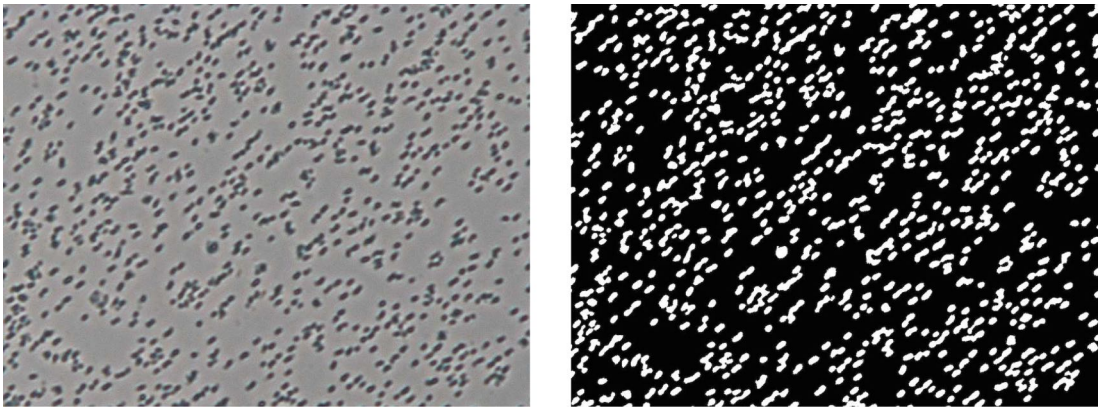


Fig. 2. Microscopic image of *Rhizobium* and its binary mask.

2.2 Feature Generation

Features that are considered in this paper refer to bacteria shape and texture of the input image. The determination of bacteria shape relies on estimating its boundary with the aid of cubic spline interpolation [5]. The latter permits to estimate the curvature of bacteria's boundary and to extract some correlation between selected distances and angles concerning the shape of bacteria in question. On the other hand, image texture features contain information about spatial relations between chrominance and luminance of the image pixels. To exploit such information, a statistical approach based on computation of the co-occurrence matrices is used [27]. The latter permits to estimate an image texture as a quantitative measure of luminance over the entire input image.

2.3 Feature Selection

It is common that processing large set of generated features may yield some of them highly correlated with one another. Such potential redundancy usually impacts on the classification accuracy. In contrast, the other group of extracted features can be poorly correlated to the dependent variable affiliated to the respective class. Consequently, the reduced set of selected features (sifted from the full set of initially determined features) should consist of those which are strongly correlated to the image class and weakly associated with the remaining features. In order to accomplish the latter the following methods for improving feature selection are used here: FCBF (*Fast Correlation Based Filter*), SBMLR (*Sparse Multinomial Logistic Regression with Bayesian Regularization*) and CFS (*Correlation-based Feature Selection*) - see e.g. [21].

2.4 Class Recognition

This paper resorts to the machine learning classifiers such as RF (*Random Forest*), SVM (*Support Vector Machine*), kNN (*k-Nearest Neighbors*), MLP (*Multilayer Perceptron*) [17], ELM (*Extreme Learning Machine*) [26] and ELM-RBF (*Extreme Learning Machine - Radial Basis Function*) [10].

3 Features Based on Shape

3.1 Planar Cubic Spline Interpolation

Consider now the ordered set of $m+1$ planar points $\mathcal{Q}_m = \{q_k\}_{k=0}^m$ i.e. sequence of points $q_k = (x_k, y_k)$ contained in 2D-Euclidean space $\mathbb{E}^2$. In the context of this work $\mathcal{Q}_m$ represents selected points of bacteria's boundary $\partial\Gamma$. In a quest to extract some shape information of $\partial\Gamma$ (or its estimate) an interpolation based approach is applied here [5]. In the classical setting of fitting input data, $\mathcal{Q}_m$ is also supplemented with the associated parameters, called interpolation knots $\mathcal{T}_m = \{t_k\}_{k=0}^m$ subject to $t_k < t_{k+1}$, $t_0 = 0$, $t_m = T$ and $t_k \in [0, T]$. Here the unknown function $\gamma : [0, T] \to \mathbb{E}^2$ meeting the constraints $\gamma(t_k) = q_k$ is assumed to satisfy $graph(\gamma) = \partial\Gamma$. For a given pair $(\mathcal{Q}_m, \mathcal{T}_m)$ there is a variety of interpolation schemes $\gamma_I : [0, T] \to \mathbb{E}^2$ fulfilling $\gamma_I(t_k) = q_k$ - see e.g. [5] or [9]. Since the selected interpolant γ_I to fit $(\mathcal{Q}_m, \mathcal{T}_m)$ should be both twice-differentiable (for curvature calculation) and should not render too excessive variations of $graph(\gamma_I)$ (for arbitrary m) a cubic spline $\gamma_I = \gamma^{cs}$ is a natural choice [5]. The interpolant γ^{cs} is defined as a track-sum of cubics $\{\gamma_k^{cs}\}_{k=0}^{m-1}$ with each cubic $\gamma_k^{cs} : [t_k, t_{k+1}] \to \mathbb{E}^2$ depending on four 2D-parameters (as $a_k, b_k, c_k, d_k \in \mathbb{R}^2$)

$$\gamma_k^{cs}(t) = a_k + b_k(t - t_k) + c_k(t - t_k)^2 + d_k(t - t_k)^3. \tag{1}$$

Here $4 \times m$ coefficients $\{(a_k, b_k, c_k, d_k)\}_{k=0}^{m-1}$ are calculable from $4 \times m$ constraints:

1. $2 \times m$ interpolation conditions for $k \in \{0, 1, \ldots, m-1\}$:

$$\gamma_k^{cs}(t_k) = q_k \quad \text{and} \quad \gamma_k^{cs}(t_{k+1}) = q_{k+1}. \tag{2}$$

2. $m-1$ internal points' first-order smoothness for $k \in \{0, 1, \ldots, m-2\}$:

$$\dot{\gamma}_k^{cs}(t_{k+1}) = \dot{\gamma}_{k+1}^{cs}(t_{k+1}). \tag{3}$$

3. $m-1$ internal points' second-order smoothness for $k \in \{0, 1, \ldots, m-2\}$:

$$\ddot{\gamma}_k^{cs}(t_{k+1}) = \ddot{\gamma}_{k+1}^{cs}(t_{k+1}). \tag{4}$$

4. 2 boundary conditions complementing (2), (3), (4) to yield $4 \times m$ equations.

Usually, the last two equations are obtainable from extra conditions such as e.g. $\dot{\gamma}(0) = v_0$ and $\dot{\gamma}(T) = v_m$. Indeed, the latter yields two missing equations:

$$\dot{\gamma}_0^{cs}(t_0 = 0) = v_0 \quad \text{and} \quad \dot{\gamma}_{m-1}^{cs}(t_m = T) = v_m. \tag{5}$$

Although, in our setting both velocities v_0 and v_m are not *a priori* given, they can be still estimated from $(\mathcal{Q}_m, \mathcal{T}_m)$ following the concept of modified Hermite interpolation [14]. Indeed, a unique Lagrange cubic $\gamma_0^{L(3)} : [0, t_3] \to \mathbb{E}^2$ interpolating the first four points $\{q_k\}_{k=0}^3$ at $\{t_k\}_{k=0}^3$ (see [5]) yields some estimate of $v_0 \approx \hat{v}_0 = \dot{\gamma}_3^{L(3)}(0)$. Similarly, a Lagrange cubic $\gamma_{m-3}^{L(3)} : [t_{m-3}, t_m] \to \mathbb{E}^2$ interpolating the last four points $\{q_k\}_{k=m-3}^m$ at $\{t_k\}_{k=m-3}^m$ renders some approximation of terminal velocity $v_m \approx \hat{v}_m = \dot{\gamma}_{m-3}^{L(3)}(t_m)$. Consequently, taking into account the latter, condition (5) modifies into:

$$\dot{\gamma}_0^{cs}(t_0 = 0) = \hat{v}_0 \quad \text{and} \quad \dot{\gamma}_{m-1}^{cs}(t_m = T) = \hat{v}_m. \tag{6}$$

The scheme for selection $\mathcal{Q}_m$ from $\partial\Gamma$ is described in Subsect. 3.3. Note that in our setting to approximate $\partial\Gamma$ with the closed curve as $q_0 \neq q_m$ we extend $\mathcal{Q}_m$ to $\hat{\mathcal{Q}}_{m+1} = \{\hat{q}_k\}_{k=0}^{m+1}$ so that $\hat{q}_k = q_k$ (for $k = 0, 1, \ldots, m$) and $\hat{q}_{m+1} = q_0$.

Upon selecting the interpolation points $\mathcal{Q}_m$ (and thus $\hat{\mathcal{Q}}_{m+1}$) from the bacteria's boundary $\partial\Gamma$ the next step is to estimate the accompanying knots $\mathcal{T}_{m+1} \approx \hat{\mathcal{T}}_{m+1} = \{\hat{t}_k\}_{k=0}^{m+1}$ (as $\mathcal{T}_{m+1}$ is not available out of input images) from the distribution of $\hat{\mathcal{Q}}_{m+1}$. This permits to construct the interpolant $\hat{\gamma}^{cs} : [0, \hat{T}] \to \mathbb{E}^2$ as a track-sum $\hat{\gamma}^{cs} = \{\hat{\gamma}_k^{cs}\}_{k=0}^{m+1}$, with $\hat{\gamma}_k^{cs} : [\hat{t}_k, \hat{t}_{k+1}] \to \mathbb{E}^2$ satisfying (1), (2), (3), (4) and (6) along $\hat{\mathcal{Q}}_{m+1}$ with somehow estimated knots $\hat{\mathcal{T}}_{m+1}$. Addressing the latter, we resort here to the so-called *exponential parameterization* commonly used in computer graphics [19] and defined in accordance with:

$$\hat{t}_0 = 0, \qquad \hat{t}_{k+1} = \hat{t}_k + \|q_{k+1} - q_k\|^\lambda, \quad k = 0, 1, \ldots, m, \tag{7}$$

for some parameter $\lambda \in [0, 1]$, where $\|\cdot\|$ is a standard Euclidean norm. This paper selects $\lambda = 0.5$ in (7) yielding the so-called *centripetal parameterization* with $\hat{T} = \sum_{k=0}^m \|q_{k+1} - q_k\|^{1/2}$ (see [19]). Note that in order to preserve $t_k < t_{k+1}$ it is also assumed that $q_k \neq q_{k+1}$. More information on exponential parameterization (7) and other knots selection schemes can be found e.g. in [14–16,19].

3.2 Curvature Calculation

Having found a cubic spline $\hat{\gamma}^{cs}$ approximating bacteria's boundary $\partial\Gamma$ one may extract some shape information of $\partial\Gamma$ by analyzing the geometry of $graph(\hat{\gamma}^{cs})$ forming the planar curve assumed also to estimate $\partial\Gamma$ (for m sufficiently big). In this work a curvature of $\hat{\gamma}^{cs}$ is computed to form a geometrical marker of $\partial\Gamma$ used later as one of the differentiating ingredients in classification process. Recall, that the curvature $\kappa(t)$ of a planar curve $\gamma : [a, b] \to \mathbb{E}^2$ at a given point $t \in [a, b]$ measures the amount by which such curve deviates from a tangent line at point $\gamma(t)$ - see [29]. The respective formula for the curvature $\kappa(t)$ of regular curve γ (i.e. for γ for which $\dot{\gamma}(t) \neq \mathbf{0}$ over $t \in [a, b]$) reads as:

$$\kappa(t) = \frac{\left\|\boldsymbol{T}'(t)\right\|}{\|\boldsymbol{r}\,'(t)\|}, \tag{8}$$

where $\boldsymbol{r}(t) = \dot{\gamma}(t)$ is a tangent vector to γ at t with its normalized vector $\boldsymbol{T}(t) = \boldsymbol{r}(t)/\|\boldsymbol{r}(t)\|$. In particular, for arc-length parameterization expressed as $s = \phi(t) = \int_a^t \|\boldsymbol{r}\,'(u)\|\, du$, for which reparameterized curve $\bar{\gamma}(s) = (\gamma \circ \phi^{-1})(s)$ satisfies $\|\dot{\bar{\gamma}}(s)\| = 1$ (yielding $\|\boldsymbol{T}(s)\| = 1$ - see [8]), the equation (8) reformulates into (with the respective derivative calculated for s-variable) $\kappa(s) = \|\boldsymbol{T}'(s)\|$.

3.3 Features Calculation

To estimate bacteria's shape (assumed here to be "more or less" convex), Region of Interest (ROI) mask is applied. In doing so, the following Matlab functions are exploited: *rgb2gray, imbinarize, imfill, bwareaopen* and *multithresh*. Upon localizing a single bacteria with ROI mask, the Laplacian filter is used to extract $\partial\Gamma$ of the analyzed object [4]. Next all computed boundary points $\mathcal{Q}_{\hat{m}} = \{q_j\}_{j=0}^{\hat{m}}$ are sorted out clock-wisely. To achieve the latter, we calculate and compare the angle between a given point $q_j \in \mathcal{Q}_{\hat{m}}$ and mean location of $\mathcal{Q}_{\hat{m}}$ i.e. the point $(\bar{x} = (1/(\hat{m}+1))\sum_{k=0}^{\hat{m}} x_k, \bar{y} = (1/(\hat{m}+1))\sum_{k=0}^{\hat{m}} y_k)$. It is assumed here that (x_k, y_k) represent Cartesian coordinates of the centers of bacterial boundary pixels (the center of the coordinate system is set in the upper left corner of an image). As a result the boundary of each single bacteria $\partial\Gamma$ is represented by a large set of points $\mathcal{Q}_{\hat{m}}$ which in turn is reduced to terser set $\mathcal{Q}_m$ with $m+1 = 10$ or $m+1 = 20$ points (and thus to $\hat{\mathcal{Q}}_{m+1}$ - see Subsect. 3.1) to be fitted with $\hat{\gamma}^{cs}$ and $\hat{\mathcal{T}}$ governed by (7). Such reduction is carried out upon selecting from $\mathcal{Q}_{\hat{m}}$ "more or less" equally spaced points with respect to their index distribution taken in clockwise order (e.g. for $\hat{m} = 54$ a possible reduction leads to $\{q_0, q_8, q_{17}, q_{26}, q_{35}, q_{44}, q_{54}\}$).

The feature extraction process aimed to determine some bacteria's shape information relies on curvature calculation from the estimated bacteria's boundary $\partial\Gamma$ - see [1]. In doing so, formula (8) is applied to cubic spline $\hat{\gamma}_k^{cs}$ (see Subsect. 3.1). More precisely, with the aid of (8) for each $\hat{\gamma}_k^{cs}$ we compute over $[\hat{t}_k, \hat{t}_{k+1}]$ the maximal and minimal values of the curvature function $\kappa(\hat{t})$ (i.e. κ^{max} and κ^{min}) yielding the corresponding knots $\hat{t}_k^{max}, \hat{t}_k^{min} \in [\hat{t}_k, \hat{t}_{k+1}]$ obtained

from $\kappa^{max} = \kappa(\hat{t}_k^{max})$ and $\kappa^{min} = \kappa(\hat{t}_k^{min})$, respectively. Note that if the pair of knots $(\hat{t}_k^{max}, \hat{t}_k^{min})$ is not uniquely determined one can choose e.g. the smallest two knots $t_k^{max,min} \in [\hat{t}_k, \hat{t}_{k+1}]$, respectively. This in turn, permits to determine two points $q_k^{max} = \hat{\gamma}_k^{cs}(\hat{t}_k^{max})$ and $q_k^{min} = \hat{\gamma}_k^{cs}(\hat{t}_k^{min})$ having maximal and minimal curvature $\hat{\gamma}_k^{cs}$ over the segment $[\hat{t}_k, \hat{t}_{k+1}]$. According to the order of all knots $\hat{t}_k^{max}$, $\hat{t}_k^{min}$ we re-index points q_k^{min} and q_k^{max} placing them into one sequence formula $\{q_k^{ext}\}_{k=0}^{2m+1}$ (where either $ext = max$ or $ext = min$). In the next step we determine the center of mass $q_c = (1/2(m+1)) \sum_{k=0}^{2m+1} q_k^{ext}$ needed as a reference point to compute the $2(m+1)$ distance values $a_k = \|q_k^{ext} - q_c\|$. In sequel, a family of triangles $\Delta_k(q_k^{ext}, q_c, q_{k+1}^{ext})$ with common apex at q_c (with the respective lengths of Δ_k sides: a_k, $b_k = \left\|q_{k+1}^{ext} - q_k^{ext}\right\|$ and a_{k+1}) form a polygonal approximation of $\partial\Gamma$. Given the lengths of all sides of triangle Δ_k, its respective angles $\alpha_k, \beta_k, \gamma_k = \pi - (\alpha_k + \beta_k)$ are easily computable from the cosine theorem - here $\alpha_k = \measuredangle(q_k^{ext}, q_c, q_{k+1}^{ext})$, $\beta_k = \measuredangle(q_k^{ext}, q_{k+1}^{ext}, q_c)$ and $\gamma_k = \measuredangle(q_{k+1}^{ext}, q_k^{ext}, q_c)$. Having determined the above distances and angles a given bacteria can be represented by the following four vectors (forming de facto its *polygonal shape descriptors*): $\boldsymbol{a} = (a_0, \ldots, a_{2m+1})$, $\boldsymbol{b} = (b_0, \ldots, b_{2m+1})$, $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_{2m+1})$ and $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_{2m+1})$. At this point, we assume that we are given a reference bacteria (a kind of "geodetic benchmark" not necessarily belonging to any investigated herein soil microorganisms' classes) to which different five classes of examined bacteria are compared accordingly. Experiments carried out so far based on 5 generic representatives - one for each bacteria class - did not improve the results over selecting one reference bacteria. To juxtapose vectors representing an examined bacteria with the reference bacteria vectors $\boldsymbol{a}^{ref}$, $\boldsymbol{b}^{ref}$, $\boldsymbol{\alpha}^{ref}$, $\boldsymbol{\beta}^{ref}$ we calculate the cross correlation coefficient [6] between the respective pairs (i.e. $xcorr(\boldsymbol{a}, \boldsymbol{a}^{ref})$). For four cross correlation vectors one chooses their respective greatest values $a^{max}, b^{max}, \alpha^{max}$ and β^{max}. In each picture, we select from 5 to 50 bacteria whose surface area is the closest to the median surface area of all the bacteria in the input picture. The latter permits to select bacteria characterized by the average size and stage of growth. The less bacteria we select the less likely we qualify a group of overlapping bacteria as a single object. We considered 6 features (listed below) based on shape calculated for a fixed amount of points on one bacteria and the number of bacteria analyzed in a single image. We estimated the edge of bacteria using $m+1 = 10$ or $m+1 = 20$ points on one bacteria and compared $l = 5$, 10, 20, 25, 30, 40 and 50 bacteria on one image. The following 6 features based on shape information are considered:

1. *Mean bacteria arc-length* - which is a sum of all arc-lengths representing the perimeters of all selected bacteria divided by l.
2. *Mean curvature of l bacteria* - is a sum $(1/l) \sum_{k=1}^{l} \sum_{j=0}^{m} \int_{\hat{t}_j}^{\hat{t}_{j+1}} \kappa_j^k(\hat{t}) d\hat{t}$, where κ_j^k represents the curvature of k-th bacteria along j-th segment (see (8)).
3. *Mean maximal first distance correlation* - $(1/l) \sum_{k=1}^{l} a_k^{max}$.
4. *Mean maximal second distance correlation* - $(1/l) \sum_{k=1}^{l} b_k^{max}$.
5. *Mean maximal first angle correlation* - $(1/l) \sum_{k=1}^{l} \alpha_k^{max}$.
6. *Mean maximal second angle correlation* - $(1/l) \sum_{k=1}^{l} \beta_k^{max}$.

4 Features Based on Texture

The second group of examined features relies on image texture analysis. In [12] Haralick introduced statistical measures resorting to the second order image histogram called GLCM (*Grey-Level Co-Occurrence Matrix*). In this paper we also used GLRLM (*Gray-Level Run-Length Matrix*) measures - see [11].

4.1 GLCM Features

GLCM is calculated for the following directional angles α: 0°, 45°, 90°, 135° and distance d on quantized image Ω to n levels that are represented in gray-scale. The co-occurrence matrix M of size $n \times n$ is initialized with all its coefficients set to zero. Assume the image Ω is represented by the pixel table $\bar{M}$ having m_1 rows and n_1 columns. Note that here pixel $(1,1)$ represents the top-left pixel in Ω, whereas pixel (m_1, n_1) corresponds to the bottom-right image pixel. In addition, let matrix $W^{k,l}$ have m_w rows and n_w columns. $W^{k,l}$ is used iteratively to extract the following pixels of $\bar{M}[i,j]$: $m_w(k-1) < i \leq km_w$ and $n_w(l-1) < j \leq ln_w$. The latter can be geometrically viewed as positioning top-left corner of $W^{k,l}$ at (k,l) pixel of Ω. The coverage of Ω with $W^{k,l}$ abides the following pattern. First Ω is horizontally covered by windows $W^{1,1}, W^{1,n_w+1}, W^{1,2n_w+1}, \ldots, W^{1,n_1-n_w+1}$, respectively. Next after vertical shift to $W^{m_w+1,1}$ we move horizontally up to W^{m_w+1,n_1-n_w+1}. This procedure of disjoint coverage of Ω is continued up until reaching $W^{m_1-m_w+1,n_1-n_w+1}$ window. Note that if either $m_1(mod\ m_w) \neq 0$ or $n_1(mod\ n_w) \neq 0$ (since in practice $n_w \ll n_1$ and $m_w \ll m_1$) one can supply extra missing pixels for the most right or bottom part of Ω by extrapolation techniques. Additionally for each of $W^{k,l}$ we iterate over the pixels in that window incrementing values in M based on the correlations between pixels for direction α and distance d that is explained below (for more details see [12]).

Assume we use a window of size $m_w = 5$ and $n_w = 5$. For every $W^{k,l}$ we go through each pixel in that window. Let $n = 5$, $d = 2$, $\alpha = 45°$ and $w_{11}^{k,l}, \ldots, w_{55}^{k,l}$ be certain pixel values in $W^{k,l}$ (see Fig. 3). Here we are in the iterative step that analyzes pixel $w_{42}^{k,l}$ (that is marked as dark gray), its value is equal to 0. Then we know that we increment co-occurrences in GLCM matrix in first row that is responsible for relationships between values of pixels 0 and $n_i = 0, \ldots, 4$. Next step is to check values of pixels located in direction $\alpha = 45°$ and maximum distance of $d = 2$. There are two pixels meeting these requirements: $w_{33}^{k,l}$ and $w_{24}^{k,l}$. Since $w_{33}^{k,l} = 0$ we have to increment value in first row and in first column, and since $w_{24}^{k,l} = 3$ we need to increment value in first row and in fourth column in GLCM. Finally after we moved our window through the entire image registering co-occurrences of the pixels we divide each of the values in $W^{k,l}$ by n^2 that gives us probabilities of co-occurrences between gray levels of pixels for direction α and maximum distance d. Based on GLCM computation, the following 8 statistical measures are calculated [32]: *Contrast, Correlation, Energy, Homogeneity, Autocorrelation, Cluster Prominence, Inverse Difference and Dissimilarity*. Note that as GLCM is calculated for 4 different directions, we

obtain measures such as Contrast 0°, Contrast 45°, Contrast 90° and Contrast 135°. The final value of Contrast is taken as mean of these four values. The 8 statistical measures from above determine 8 texture features based on GLCM.

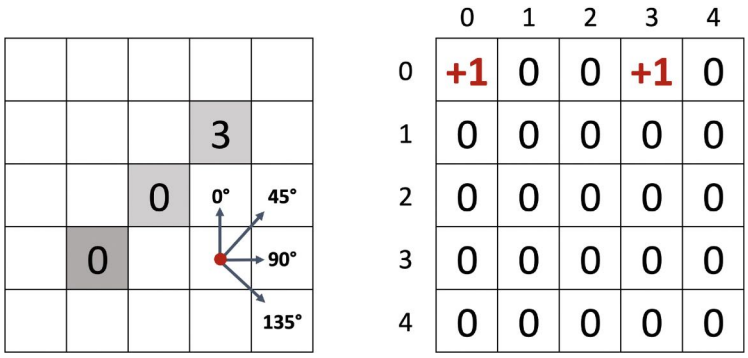


Fig. 3. Increment of the values in GLCM (right) based on the window $W^{k,l}$ (left) of size 5×5, pixel $w_{42}^{k,l}$ for $n = 5$, $d = 2$, $\alpha = 45°$ and a presentation of all possible directions α from the pixel $w_{44}^{k,l}$ marked with a red dot. (Color figure online)

4.2 GLRLM Features

The calculations of GLRLM based features are very similar to determining GLCM. The computations involved are also carried out for same directional angles α: 0°, 45°, 90°, 135° and for the maximum distance d on quantized image to n levels that is represented in gray-scale. We initialize GLRLM of size $n \times d$ with zeros. Using window $m_w \times n_w$ we move through the image in the same manner as in GLCM and increment values in matrix according to the run-length of co-occurrence pixels for direction α and length that is equal to $d_i = 1, \ldots, d$. In the next step each of the values in GLRLM is divided by dn. The latter introduces matrix with probabilities of the respective co-occurrences of the gray level $n_i = 0, \ldots, n-1$ and run-length d_i.

As an example assume an input image is quantized to 5 gray levels, maximum run-length $d = 3$, window's size is 5×5, $\alpha = 90°$ and that we analyze pixels with values equal to zero. Let $W^{k,l}$ have sample entries as shown in left window of Fig. 4. The algorithm represents searching sequences of length d_i of pixels that have values equal to zero in $W^{k,l}$ e.g. two sequences of length 3 (marked as green - see Fig. 4) are found filling GLRLM for $n = 0$ with value $d_3 = 2$. For more details on GLRLM see also [11].

Again based on probabilities stored in GLRLM the following statistical measures are calculated [11]: *Short Run Emphasis, Long Run Emphasis, Grey Level Non-uniformity, Run Length Non-uniformity, Run Percentage, Low Grey Level Run Emphasis, High Grey Level Run Emphasis Short Run Low Gray Level Emphasis, Short Run High Gray Level Emphasis, Long Run Low Gray Level Emphasis* and *Long Run High Gray Level Emphasis*. As previously, these 11 measures are computed along 4 different directions and analogously to the case of GLCM the respective mean values of each measures are determined. Ultimately,

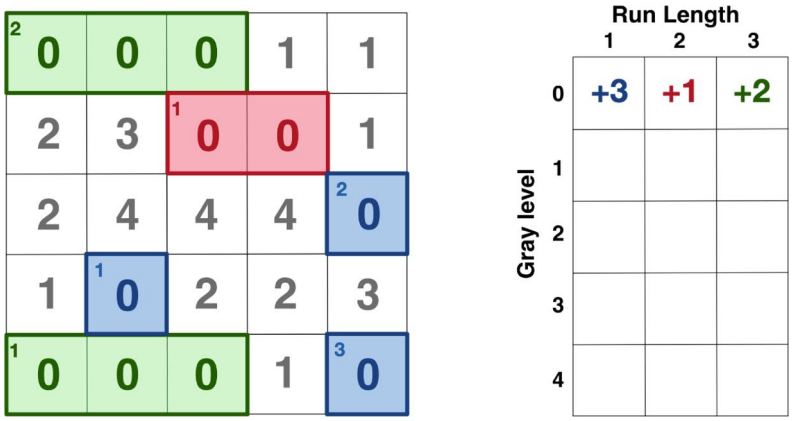


Fig. 4. Incrementing values in GLRLM (right) according to pixels in window $W^{k,l}$ (left) for $\alpha = 90°$ and the gray level equal to zero. (Color figure online)

this approach determines additional group of 11 features based on texture information. The latter together with the previously introduced GLCM based group yields 19 texture based features considered in this paper. They complement previously discussed 6 shape based features introduced to accomplish bacteria classification.

5 Selected Class Recognition Methods

5.1 KNN

K-Nearest Neighbors is a classification method [17] that permits to assign a new object to one of the constructed classes. We are supplied here with a data-set attributed with the existing features and its membership to the respective class. Based on the latter, a new object needs to be classified with respect to the same set of attached features. In doing so, the calculation of Euclidean distances between a new object and every single object from the data-set is performed. We sort these objects by distance in ascending order. Then we choose k objects (k value is set arbitrarily) whose distances are the smallest and conduct majority voting (see [17]) to decide to which class the new object should be attached.

5.2 ELM-RBF

Extreme Learning Machine - Radial Basis Function is feed-forward neural network with two hidden layers [10]. First one is responsible for the input vectors conversion to the distances based on Gaussian radial function to the closest centroid. Their amount is selected arbitrarily and they are computed using k-means method, where k defines number of centroids to be calculated. This procedure is very similar to RBN (*Radial Basis Network*) and brings much more robustness to the prediction [20]. Upon converting the input vectors into distances in the first hidden layer, we treat them as input vector in ELM network that contains one

hidden layer with experimentally chosen number of neurons and their activation function. The bias and weights between hidden layers are assigned randomly. Weights between hidden and output layer are calculated using Moore-Penrose matrix pseudo-inverse operation [25].

6 Experiments and Results

We report now on generated experimental results based either on combined set of shape and texture features or solely relying on shape or texture information.

First, the experiments incorporating a full set of shape and texture attributes juxtapose different feature selection and classification methods to reach satisfactory classification accuracy. Additionally, the tests f determining the optimal number of interpolation points $\hat{\mathcal{Q}}_{m+1}$ along $\partial\Gamma$ together with gauging the amount of selected bacteria are carried out. The best classification results obtained yield $m+1=10$ and 50 bacteria. We present now the list of implemented methods (see also [18,31]) with the experimentally picked up optimal parameters guaranteeing the highest possible classification accuracy: SVM (using *radial basis function*), RF (using 200 trees), kNN (using $k=1$), MLP (using back-propagation learning method, topology of the net $22-20-22$ and *tanh* as an activation function on all hidden layers), ELM (with 2800 neurons in hidden layer units with *tanh* activation function) and ELM-RBF (900 neurons with linear activation function in hidden layer units and 40 centroids).

Table 1. Mean accuracy percentage of 50 tests using 10% cross validation and feature selection with classification method performed on shape and texture features.

Feature selection method	SVM	RF	kNN	ELM	MLP	ELM-RBF
None	93.84	92.76	95.30	92.50	92.03	94.95
FCBF	95.69	93.62	97.07	78.80	89.73	94.05
SBMLR	96.61	93.92	96.00	91.61	92.65	97.03
CFS	93.61	91.23	95.00	85.34	91.03	94.74

As shown in Table 1, the best result in bacteria classification amounts to 97% in accuracy recognition which is obtained upon either applying ELM-RBF with SBMLR or using kNN with FCBF. In this case ELM-RBF shows superiority over ELM method increasing accuracy by over 5% and due to a smaller number of neurons in hidden layer has a vastly shorter training and testing time. Here SBLMR selects 5 shape and 15 texture features whereas FCBF relies on using 3 shape and 4 texture traits. In order to justify merging features from both classes of examined attributes (i.e. shape and texture), we subsequently tested the bacteria classification accuracy when either only the set of shape or the set of texture features is admitted, respectively.

Table 2. Mean accuracy percentage of 50 tests using 10% cross validation and feature selection with classification method performed on features based on shape.

Feature selection method	SVM	RF	kNN	ELM	MLP	ELM-RBF
None	75.69	72.61	77.00	49.84	73.84	76.92
FCBF	78.07	74.07	74.76	29.38	70.84	76.00
SBMLR	76.06	76.40	77.56	48.00	73.47	77.96
CFS	78.92	73.30	76.69	42.92	74.38	77.69

Table 3. Mean accuracy percentage of 50 tests using 10% cross validation and feature selection with classification method performed on features based on texture.

Feature selection method	SVM	RF	kNN	ELM	MLP	ELM-RBF
None	80.07	76.30	81.61	64.84	67.46	68.23
FCBF	75.76	79.76	76.61	41.61	68.30	74.76
SBMLR	78.93	77.06	82.27	63.46	67.81	69.64
CFS	79.00	78.84	78.53	55.46	70.53	74.38

The best experimental result relying exclusively on shape features amounts to 78.92% in classification accuracy (see Table 2). It is derived with the aid of SVM coupled with CFS. In contrast, the best accuracy using solely texture based features equals 82.27% (see Table 3) and is achieved upon combining kNN with SBMLR. Having juxtaposed results from Table 1, 2 and 3 it is transparent that merging shape and texture features improves classification accuracy by 15%. As shown in Table 1 the mean accuracy from 50 tests using 10% cross validation amounts to 97% matching state of the art results.

7 Conclusions

Experiments based on 6 shape features render (for our data) the best accuracy reaching 78.92% correct classification for SVM combined with CFS. On the other hand class recognition based on 19 image texture traits yields up to 82.27% for kNN and SBMLR. In contrast, gathering together both shape and texture information (totalling 25 conjugated features) leads up to 97% correct classification upon coupling either kNN with FCBF or ELM-RBF with SBMLR. The iterative optimization of the classification model parameters including selection of the number of knots and the amount of bacteria analyzed in one picture, improves accuracy and reduces time execution of the implemented congregated classifier. These results seem to be unexpectedly satisfactory for our proposed *aggregated bacteria classifier* in the absence of incorporating color information. Still, within the setting of this work, there is a natural scope for further improvements. In particular, any method selecting characteristic benchmark bacteria for a given genus permitting to compare bacteria's curvature with the reference bacteria

would be desirable. In this work originally, such five exemplary bacteria were selected arbitrarily but the results obtained did not improve significantly the case of fixing one reference bacteria for five considered genera. Another related issue refers to the task of selecting all significant points (and knots) on the bacteria's boundary (see e.g. [5,14–16,19]). This work assumes "more or less" equally spaced points Q_m. The impact of convexity or non-convexity of the bacteria should also be analyzed with respect to ordering $\mathcal{Q}_{\hat{m}}$. Furthermore, the comparison of standard classification methods with deep learning methods and extending admissible set of features incorporating color and dispersion information forms potential research topics within the field of soil bacteria classification. Lastly, the robustness of all examined methods may also be tested against the varying number of bacteria genera (or their respective representatives) and possibly in regard to other dynamic factors such as time aspect impacting on shape, size or a color of the examined bacteria and/or its bacterial colony distribution.

References

1. Amirani, M.C., Gol, Z.S., Shirazi, A.A.B.: Efficient feature extraction for shape-based image retrieval. J. Appl. Sci. **8**, 2378–2386 (2008). https://doi.org/10.3923/jas.2008.2378.2386
2. Anderson, L.M., Stockwell, V.O., Loper, J.E.: An extracellular protease of Pseudomonas fluorescens inactivates antibiotics of Pantoea agglomerans. Phytopathology **94**(11), 1228–1234 (2004). https://doi.org/10.1094/PHYTO.2004.94.11.1228
3. Beeckmans, S., Xie, J.: Glyoxylate cycle. In: Reference Module in Biomedical Sciences. Elsevier (2015). https://doi.org/10.1016/B978-0-12-801238-3.02440-5
4. Bhairannawar, S.S.: Chapter 4 - efficient medical image enhancement technique using transform HSV space and adaptive histogram equalization. In: Soft Computing Based Medical Image Analysis, pp. 51–60. Academic Press (2018). https://doi.org/10.1016/B978-0-12-813087-2.00003-8
5. de Boor, C.: A Practical Guide to Splines. Springer (2001)
6. Buck, J.R., Daniel, M.M., Singer, A.: Computer Explorations in Signals and Systems Using MATLAB. Prentice Hall (2002)
7. Caprette, D.R.: Describing colony morphology. https://bit.ly/324cqkA
8. do Carmo, M.P.: Differential Geometry of Curves and Surfaces. Prentice Hall (1976)
9. Das, B., Chakrabarty, D.: Lagrange's interpolation formula: representation of numerical data by a polynomial curve. Internat. J. Math. Trends Technol. 34, 64–72 (2016). https://doi.org/10.14445/22315373/IJMTT-V34P514
10. Dhini, A., Surjandari, I., Kusumoputro, B., Kusiak, A.: Extreme Learning Machine-Radial Basis Function (ELM-RBF) networks for diagnosing faults in a steam turbine. J. Ind. Prod. (2021). https://doi.org/10.1080/21681015.2021.1887948
11. Ferro-Flores, G., et al.: Prediction of overall survival and progression-free survival by the 18F-FDG PET/CT radiomic features in patients with primary gastric diffuse large B-Cell Lymphoma. Contrast Media Mol. Imaging (2019). https://doi.org/10.1155/2019/5963607
12. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. IEEE Trans. Syst. Man. SMC-**3**(6), 610–621 (1973). https://doi.org/10.1109/TSMC.1973.4309314

13. Iversen, C.: Encyclopedia of Food Microbiology: Enterobacter. Academic Press, 2nd edn. (2014)
14. Kozera, R.: Curve modeling via interpolation based on multidimensional reduced data. Studia Inform. **25**(4B), 1–140 (2004)
15. Kozera, R., Noakes, L., Wiliński, A.: Generic case of Leap-Frog Algorithm for optimal knots selection in fitting reduced data. In: ICCS 2021. LNCS, vol. 12745, pp. 337–350. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-77970-2_26
16. Kozera, R., Noakes, L., Wilkołazka, M.: Parameterizations and Lagrange cubics for fitting multidimensional data. In: ICCS 2020. LNCS, vol. 12138, pp. 124–140. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-50417-5_10
17. Kramer, O.: K-Nearest Neighbors, pp. 13–23. Springer, Berlin Heidelberg (2013). https://doi.org/10.1007/978-3-642-38652-7_2
18. Kruk, M., Kozera, R., Osowski, S., Trzciński, P., Paszt, L.S., Sumorok, B., Borkowski, B.: Computerized classification system for the identification of soil microorganisms. Appl. Math. Inf. Sci. **10**(1), 21–31 (2016). https://doi.org/10.18576/amis/100103
19. Kvasov, B.: Methods of Shape-Preserving Spline Approximation. World Scientific (2000)
20. Lee, C.C., Chung, P.C., Tsai, J.R., Chang, C.I.: Robust radial basis function neural networks. IEEE Trans. Syst. Man Cybern. **29**(6), 674–685 (1999). https://doi.org/10.1109/3477.809023
21. Lefakis, L., Fleuret, F.: Jointly informative feature selection made tractable by Gaussian modeling. J. Mach. Learn. Res. **17**(182), 1–39 (2016). http://jmlr.org/papers/v17/15-026.html
22. Lim, Y., et al.: Mechanically resolved imaging of bacteria using expansion microscopy. PLOS Biol. **17**(10), 1–19 (2019). https://doi.org/10.1371/journal.pbio.3000268
23. Malhi, S.S., Sahota, T.S., Gill, K.S.: Chapter 5 - potential of management practices and amendments for preventing nutrient deficiencies in field crops under organic cropping systems. In: Agricultural Sustainability, pp. 77–101. Academic Press (2013). https://doi.org/10.1016/B978-0-12-404560-6.00005-8
24. Morin, A.: Encyclopedia of Food Microbiology: Pantoea. Academic Press, 2nd edn. (2014)
25. Rao, C.R.: Generalized Inverse of a Matrix and its Applications, pp. 601–620. University of California Press (1972). https://doi.org/10.1525/9780520325883-032
26. Satoto, B.D., Utoyo, M.I., Rulaningtyas, R., Koendhori, E.B.: Classification of features shape of Gram-negative bacterial using an Extreme Learning Machine. IOP Conf. Ser. Earth Environ. Sci. **524**(1), 012005 (2020). https://doi.org/10.1088/1755-1315/524/1/012005
27. Shapiro, L.G., Stockman, G.: Computer Vision. Pearson, 1st edn. (2001)
28. Soille, P.: Morphological Image Analysis: Principles and Applications. Springer-Verlag (1999)
29. Sokolov, D.D.: Encyclopedia of Mathematics. EMS Press (2001)
30. Sorensen, J., Nybroe, O.: Pseudomonas: Volume 1 Genomics, Life Style and Molecular Architecture, chap. Pseudomonas in the Soil Environment, pp. 369–401. Springer, US (2004). https://doi.org/10.1007/978-1-4419-9086-0_12
31. Toprak, A.: Extreme Learning Machine (ELM)-based classification of benign and malignant cells in breast cancer. Med. Sci. Monit. **24**, 6537–6543 (2018). https://doi.org/10.12659/MSM.910520

32. Yang, X., et al.: Ultrasound GLCM texture analysis of radiation-induced parotid-gland injury in head-and-neck cancer radiotherapy: an in vivo study of late toxicity. Med. Phys. **39**(9), 5732–5739 (2012). https://doi.org/10.1118/1.4747526

3.2 Identification of the selected soil bacteria genera based on their geometric and dispersion features

Publication: A. Konopka, R. Kozera, L. Sas-Paszt, P. Trzcinski, and A. Lisek, "Identification of the selected soil bacteria genera based on their geometric and dispersion features," *PLoS ONE*, vol. 18, no. 10, e0293362, pp. 1–11, 2023. https://doi.org/10.1371/journal.pone.0293362 (100 MNiSW points, IF – 2.9)

Abstract: The visual analysis of microscopic images is often used for soil bacteria recognition in microbiology. Such task can be automated with the aid of machine learning and digital image processing techniques. The best results for soil microorganism identification usually rely on extracting features based on color. However, accommodating in the latter an extra impact of lighting conditions or sample's preparation on classification accuracy is often omitted. In contrast, this research examines features which are insensitive to the above two factors by focusing rather on bacteria shape and their specific group dispersion. In doing so, the calculation of layout features resorts to k-means and mean shift methods. Additionally, the dependencies between specific distances determined from bacteria cells and the curvature of interpolated bacteria boundary are computed to extract vital geometric shape information. The proposed bacteria recognition tool involves testing four different classification methods for which the parameters are iteratively adjusted. The results obtained here for five selected soil bacteria genera: *Enterobacter*, *Rhizobium*, *Pantoea*, *Bradyrhizobium* and *Pseudomonas* reach 85.14% classification accuracy upon combining both geometric and dispersion features. The latter forms a promising result as a substitutive tool for color-based feature classification.

PLOS ONE

RESEARCH ARTICLE

Identification of the selected soil bacteria genera based on their geometric and dispersion features

Aleksandra Konopka[1]*, **Ryszard Kozera**[1,2], **Lidia Sas-Paszt**[3], **Pawel Trzcinski**[3], **Anna Lisek**[3]

1 Institute of Information Technology, Warsaw University of Life Sciences - SGGW, Warsaw, Poland, **2** School of Physics, Mathematics and Computing, The University of Western Australia, Perth, Australia, **3** Department of Microbiology and Rizosphere, The National Institute of Horticultural Research, Skierniewice, Poland

* aleksandra_konopka@sggw.edu.pl

Citation: Konopka A, Kozera R, Sas-Paszt L, Trzcinski P, Lisek A (2023) Identification of the selected soil bacteria genera based on their geometric and dispersion features. PLoS ONE 18(10): e0293362. https://doi.org/10.1371/journal.pone.0293362

Editor: Carlos Fernandez-Lozano, University of A Coruña, SPAIN

Received: April 25, 2023

Accepted: October 10, 2023

Published: October 27, 2023

Data Availability Statement: The microscopic image dataset and the code written in support of this publication is publicly available at 10.5281/zenodo.7789436.

Funding: This research was supported by The National Centre for Research and Development within the framework of the project BIOSTRATEG, grant number BIOSTRATEG3/344433/16/NCBR/2018. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

The visual analysis of microscopic images is often used for soil bacteria recognition in microbiology. Such task can be automated with the aid of machine learning and digital image processing techniques. The best results for soil microorganism identification usually rely on extracting features based on color. However, accommodating in the latter an extra impact of lighting conditions or sample's preparation on classification accuracy is often omitted. In contrast, this research examines features which are insensitive to the above two factors by focusing rather on bacteria shape and their specific group dispersion. In doing so, the calculation of layout features resorts to *k*-means and mean shift methods. Additionally, the dependencies between specific distances determined from bacteria cells and the curvature of interpolated bacteria boundary are computed to extract vital geometric shape information. The proposed bacteria recognition tool involves testing four different classification methods for which the parameters are iteratively adjusted. The results obtained here for five selected soil bacteria genera: *Enterobacter*, *Rhizobium*, *Pantoea*, *Bradyrhizobium* and *Pseudomonas* reach 85.14% classification accuracy upon combining both geometric and dispersion features. The latter forms a promising result as a substitutive tool for color-based feature classification.

Introduction

Identification of bacteria can be realized with the use of many molecular techniques, including ribotyping, repetitive extragenic palindromic PCR (Rep-PCR), denaturing gradient gel electrophoresis (DGGE), terminal (T)-restriction fragment length polymorphism (T-RFLP), multilocus sequence typing (MLST) and whole-genome sequencing (WGS) [1]. MLST uses DNA sequencing of internal fragments of the housekeeping gene loci (seven in number) of bacterial strains to characterize alleles [2]. In practice, a common stance for bacteria identification is based on sequence analysis of 16 SrRNA gene [3, 4] and MLST unveiling the same intraspecific genetic structure patterns as genomes [5]. In bacteria recognition process, the morphological

Competing interests: The authors have declared that no competing interests exist.

features can also be considered while analyzing the microscopic images. However, sometimes it is hard to distinguish between different bacteria species due to their morphological similarities within a genera [6]. The image-based identification can be tedious and laborious.

The aim of this research is to create a system that automates the process of microscopic image classification. Incorporating the computerized approach facilitates the identification process replacing or supporting human expertise and eyesight assessment with the modern computer vision image processing techniques. Machine learning methods used in this paper have already been applied to solve pattern recognition, prediction and classification problems in various fields of biology [7] and, in particular, to identify the microorganisms [8]. Some bacteria can be easily discerned from others due to their specific morphological features e.g. *Mycobacterium tuberculosis* [9] and *Escherichia Coli* [10] both having characteristic shapes. Here a fast and robust recognition scheme is in demand as these bacteria may inflict serious human illnesses. Some works perform classification not on the genera or species level but defining each class as a shape type [11]. The features relying on shape, texture or on pixel-based measures are applied in bacteria classification [12–15]. In this paper, the classification task is accomplished on the genera level via differentiating microscopic images of five selected soil bacteria genera: *Enterobacter*, *Rhizobium*, *Pantoea*, *Bradyrhizobium* and *Pseudomonas* (see Fig 1) grown in specific conditions on selected medium. Some of these bacteria genera have a positive impact on plant growth while the others are pathogenic. For this reason it is important to accurately classify their character [16, 17].

Identification of microorganisms with machine learning methods is widely applied for recognition of pathogens causing human infections (see e.g. [12]). In contrast, the topic of soil microorganism classification has not been so-far extensively investigated. In case of image-based soil microorganism identification discussed in [18], the analysis of color features used for bacteria recognition yields up to 97% of classification accuracy (ACC). In the latter work, the goal was to create a system enabling automatic recognition of samples that are preprocessed by the microbiologists. The introduced chemical reactions result in the color change of samples depending on the species of the microorganism which ultimately facilitates the efficiency of the classifier in achieving more accurate results.

In our research a different approach is adopted. The microscopic images can be taken with various microscopes and under different lighting conditions. In addition, the photographed samples can also be processed by the microbiologists upon administering a contrast or initiating a chemical reaction. Furthermore, the analyzed samples are usually colored with dyes to improve visibility of the objects examined under the microscope. In order not to rely on these factors, the different types of features based on bacteria geometry and their group dispersion are considered in this work which yields an alternative for the color-based traits classification. Developing such a set of features can help to create an automatic program that performs an accurate classification on both raw images and on those subjected to chemical reactions. The computations are performed here on images of bacteria samples that are not earlier processed by the microbiologists. In the prior research, the combination of geometric and texture features [19] calculated on the same image dataset resulted in up to 97% classification accuracy. However, in this research features based on texture are excluded as they rely on luminance (i.e. pixels intensity values which in turn may depend on lighting conditions). Instead, only features related directly to the geometry and dispersion of the analyzed objects are considered. The highest classification accuracy obtained here for such a set of features equals 85.14%. The present findings suggest that alternative feature types have the potential to supplant chrominance and luminance features in the realm of bacterial classification. Such an approach would enable classification with comparable precision for images captured under diverse illumination conditions, amalgamating preprocessed and raw images, as the outcome remains impervious to

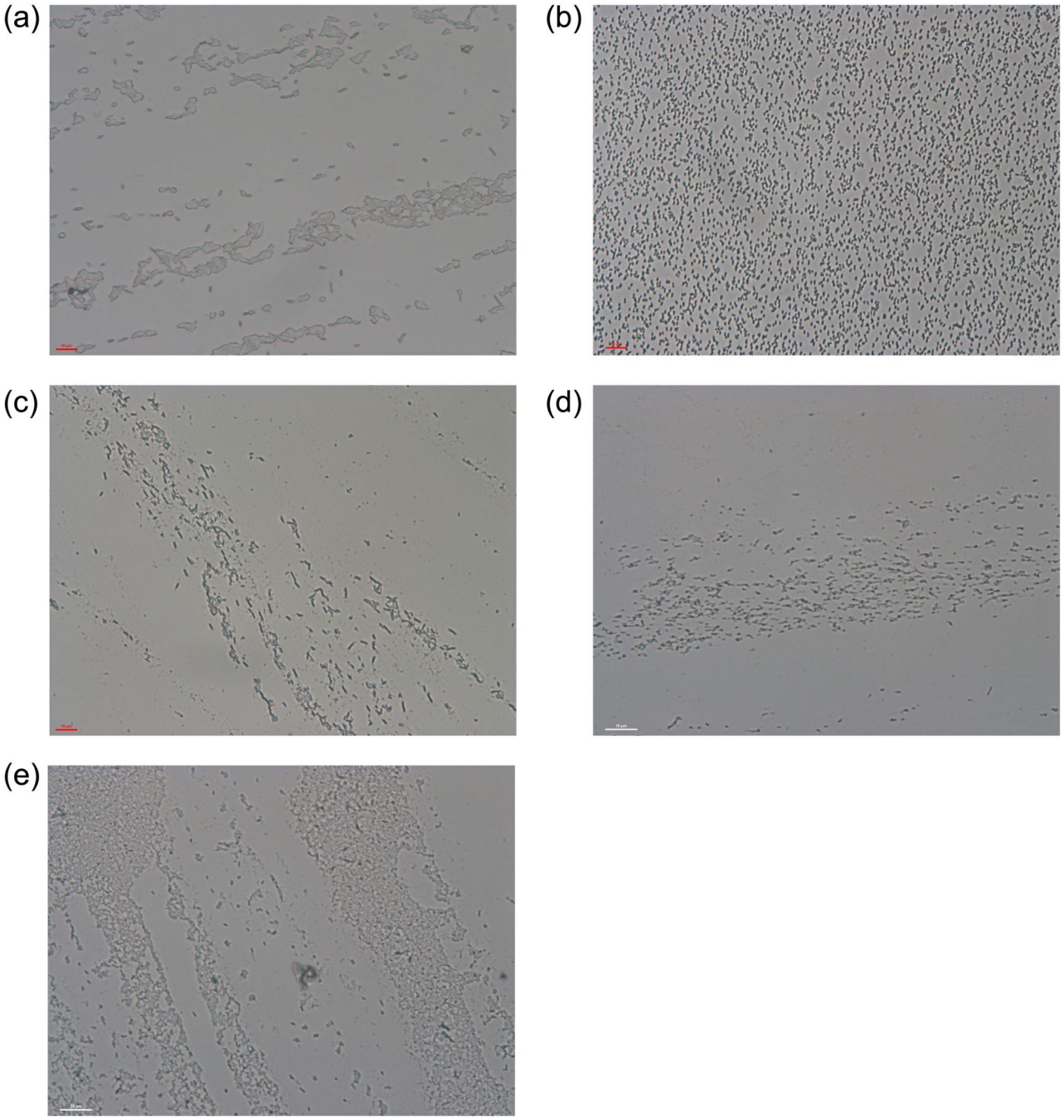


Fig 1. Examples of bacteria images: (a) *Enterobacter*, (b) *Rhizobium*, (c) *Pantoea*, (d) *Bradyrhizobium* and (e) *Pseudomonas*. For more pictures see URL link: bit.ly/3TwOgFB.

https://doi.org/10.1371/journal.pone.0293362.g001

color and light influences. Notably, the investigated set of microscopic images demonstrated an accuracy of 95.6% for exemplary color features, indicating a remaining deviation of 10% in classification accuracy. Nonetheless, given the multifaceted nature of this issue, further exploration of various factors is warranted, and the classification results are anticipated to be

enhanced upon adjustment of pertinent variables as explicated in the ensuing discussion section.

Material

Strains X58AD (*Pantoea sp*), Pi72ED (*Enterobacter sp*), Ps118AA (*Pseudomonas sp*) were grown for 48 hours in 26 Celsius degrees on Plate Count Agar (BTL P-0037) medium. E77AO (*Rhizobium*) bacteria strain and a strain that was not present in Symbio-Bank (*Bradyrhizobium*) were grown on Yeast Mannitol Agar medium for 96 hours in 26 Celsius degrees. Bacteria of each strain were collected from a single colony and transferred on the surface of glass plate. In the next step, a drop of sterile water was added and mixed with the bacteria. The resulting smear was covered with microscope slide. The analyzed images were taken with a Nikon 80i microscope.

Methods

Work-flow scheme

The work-flow scheme applied in this work consists of the following steps:

1. Segmentation of the Region of Interest.
2. Feature Calculation.
3. Feature Selection.
4. Class Recognition.

Segmentation of the Region of Interest

The aim of image segmentation is to separate the Region of Interest (ROI) from the background by creating a binary mask. In our case, ROI is the area where bacteria are located and this subarea of the mask is set to be white, while the background remains black. At first the image is converted to grayscale, then the Otsu method [18] combined with open and close morphological operations [20] is applied. These computations are performed with MATLAB functions: *rgb2gray*, *multithresh*, *imbinarize*, *imfill* and *bwareaopen*.

Calculation of geometric features

The shape of bacteria depends e.g. on their genera and growth phase. The geometric features are measured here on typical bacteria instances selected from each microscopic image and applied later for the classification purposes.

Dependencies between vectors. Let $\mathcal{Q}_m = \{q_k\}_{k=0}^m$ be a set of $m + 1$ planar points $q_k = (x_k, y_k)$ on a single bacteria's boundary in 2D-Euclidean space. These points are set in a clockwise order according to the following procedure. Recall that in MATLAB function $atan2(\tilde{y}, \tilde{x})$ calculates the angle between x-axis and a line joining point $\tilde{p} = (\tilde{x}, \tilde{y})$ with the origin of the coordinate system i.e. a point (0, 0). Upon shifting the origin to the point $c = (x_c, y_c)$ where $x_c = (1/(m+1)) \sum_{k=0}^m x_k$ and $y_c = (1/(m+1)) \sum_{k=0}^m y_k$ we applied here $atan2(x_k - x_c, y_k - y_c)$ —note that we also flipped variables in *atan2* to guarantee a clockwise order in $\mathcal{Q}_m$. The points are thus indexed in ascending order based on the *atan2* values. We pick now a point $q_{md} \in \mathcal{Q}_m$ whose Euclidean distance towards the point c is the smallest and then reorder points. If we have a sequence of elements $q_0, q_2, \ldots, q_m$ and we choose one of them as q_{md} it becomes the first element of the new sequence $\tilde{\mathcal{Q}}_m$. Then all elements following q_{md} are shifted after q_{md}, and finally

we append the elements that preceded q_{md} at the end of the sequence (so if we had q_0, q_1, q_2, q_3, q_4, q_5, q_6 and $q_{md} = q_3$ the new order reads as q_3, q_4, q_5, q_6, q_0, q_1, q_2). Next the set $\tilde{\mathcal{Q}}_m$ is reduced to $\hat{\mathcal{Q}}_n = \{\hat{q}_i\}_{i=0}^n$ upon picking $n + 1$ points. In this work $n + 1 = 10$ is arbitrarily selected for all bacteria. The points forming $\hat{\mathcal{Q}}_n$ are selected applying the following formula *fix*(*linspace*(0, m, n + 2)). Employing these functions provides a guarantee that the distances between the selected points are equal in terms of their indices, while minimizing the differences between these distances. Assume m is equal to 108 and $n + 2$ to 11, applying the *linspace* function results in the following values: 0, 10.8, 21.6, 32.4, 43.2, 54, 64.8, 75.6, 86.4, 97.2, 108. After processing by the *fix* function and omitting the first element, we obtain the indices of the points in $\tilde{\mathcal{Q}}_m$—10, 21, 32, 43, 54, 64, 75, 86, 97, 108, which form the set of points $\hat{\mathcal{Q}}_n$. In the next step we calculate distances between each $\hat{q}_i$ and $\hat{q}_{i+1}$ (and the distance between $\hat{q}_n$ and $\hat{q}_0$), and between each $\hat{q}_i$ and c.

The latter approach is illustrated in Fig 2. Note that no matter how the figure is rotated we always pick q_{md} placed in the corresponding similar position on bacteria's boundary resulting in a similar order of vector elements (starting with its q_{md}).

In addition, for a selected k-th bacteria based on its boundary points $\hat{\mathcal{Q}}_n^{(k)}$, a set of $n + 1$ triangles $\{\Delta_i^{(k)}\}_{i=0}^n$ is formed each determined by the vertices $\{\hat{q}_i^{(k)}, \hat{q}_{i+1}^{(k)}, c^{(k)}\}$ (the last triangle $\Delta_n^{(k)}$ is defined by $\{\hat{q}_n^{(k)}, \hat{q}_0^{(k)}, c^{(k)}\}$)—see Fig 2. Recalling that $\rho(x, y) = \sqrt{\sum_{i=0}^n (x_i - y_i)^2}$ defines the Euclidean distance for $x, y \in \mathbb{E}^n$, define now for each $\hat{\mathcal{Q}}_n^{(k)}$ (with $\vec{v}_{k,i} = \hat{q}_i^{(k)} - c^{(k)}$, $\vec{r}_{k,i} = c^{(k)} - \hat{q}_i^{(k)}$ and $\vec{w}_{k,i} = \hat{q}_i^{(k)} - \hat{q}_{i+1}^{(k)}$, for $i = 0, \ldots, n$; note $\vec{w}_{k,n} = \hat{q}_0^{(k)} - \hat{q}_n^{(k)}$) the following measures (see Fig 2):

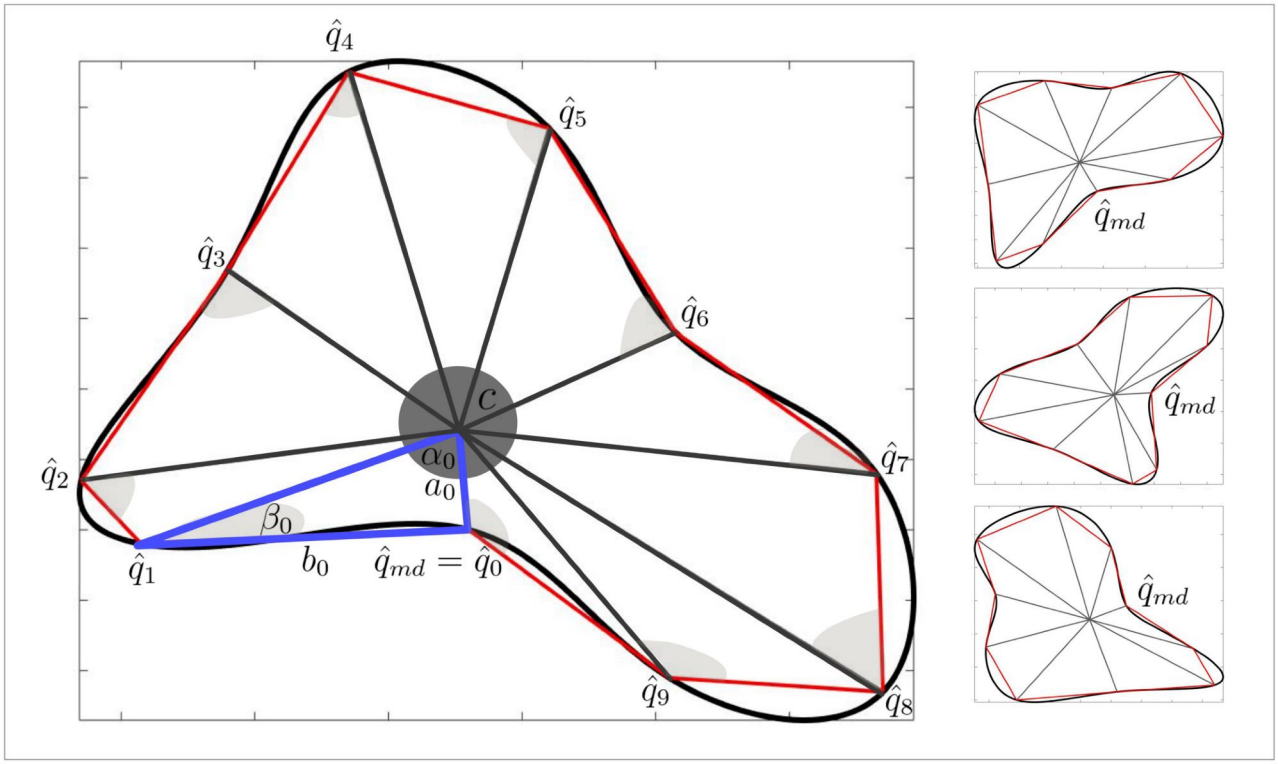


Fig 2. Distances and angles on exemplary shape, and exemplary rotations of a shape with applied method. Vectors: $\vec{a}_k$—dark gray lines, $\vec{b}_k$—red lines, $\vec{\alpha}_k$—dark gray angles, $\vec{\beta}_k$—light gray angles.

https://doi.org/10.1371/journal.pone.0293362.g002

- $\vec{a}_k$—vector of $n + 1$ distances $\rho(\hat{q}_i^{(k)}, c^{(k)})$,
- $\vec{b}_k$—vector of $n + 1$ distances $\rho(\hat{q}_i^{(k)}, \hat{q}_{i+1}^{(k)})$ (with $b_{k,n} = \rho(\hat{q}_n^{(k)}, \hat{q}_0^{(k)})$),
- $\vec{\alpha}_k$—vector of of $n + 1$ angles $\arccos(\vec{v}_{k,i+1}, \vec{v}_{k,i})$ (with $\alpha_{k,n} = \arccos(\vec{v}_{k,0}, \vec{v}_{k,n})$),
- $\vec{\beta}_k$—vector of $n + 1$ angles $\arccos(\vec{r}_{k,i+1}, \vec{w}_{k,i})$ (with $\beta_{k,n} = \arccos(\vec{r}_{k,0}, \vec{w}_{k,n})$).

In one microscopic image there might be hundreds of bacteria instances. Some of them are grouped together with overlaps which results in being identified as a single object once Region of Interest mask is applied. Another impeding factor comes from the fact that bacteria image can be taken at various stage of growth potentially related to its varying shape. Burying in mind the above concerns, only several b bacteria instances from the ROI mask are considered. These bacteria are selected based on the area value of the objects. All objects are sorted in ascending order and b items are selected with the area value closest to the median of all area values of the objects in a single image. Such approach ensures selection of objects representing single bacteria cells rather than groups of overlapped cells. In this research, for the calculation of features {1} and {2} we set $b = 50$ and for {3}, {4}, {5} and {6} $b = 10$ (for enumeration of features see subsection—All geometric features).

Each of the selected b bacteria on a given image represented by vector measures $(F_b = \{F_k\}_{k=1}^{b}$ where $F_k = (\vec{a}_k, \vec{b}_k, \vec{\alpha}_k, \vec{\beta}_k) \in \mathbb{R}^{4(n+1)})$ is compared with the exemplary bacteria measure selected by experts which is represented by $F_e = (\vec{a}_e, \vec{b}_e, \vec{\alpha}_e, \vec{\beta}_e) \in \mathbb{R}^{4(n+1)}$.

To illustrate the vector comparison procedure and to prove its credibility on more distinctive shapes an example of shape comparison between $F_{bs} = (\vec{a}_{bs}, \vec{b}_{bs}, \vec{\alpha}_{bs}, \vec{\beta}_{bs}) \in \mathbb{R}^{4(n+1)}$ with other vectors is presented (see Fig 3). F_{bs} is a set of vectors of values calculated for bacteria-shaped object. This object is an irregular oval shape that represents a bacteria cell. F_{bs} is compared with: F_{b2}—vectors calculated for bacteria-shaped object with double magnified size, F_h—vectors calculated for horseshoe shape, F_r—vectors calculated for a round shape and lastly, F_o—vectors calculated for oval shape. All shapes in question are artificially created with a slight irregularity applied. The latter corresponds to the objects selected by the ROI mask as they are also irregular and not the symmetric round or oval shapes.

For a given bacteria-like shape represented by F_{bs} we calculate the Pearson coefficient value [21] for all corresponding pairs of vectors in (F_{bs}, F_n), where $F_n \in \{F_{bs}, F_{b2}, F_h, F_r, F_o\}$, to verify how its value corresponds to the object shape Fig 3. Table 1 shows the correlation values between bacteria-shaped object and the same object resized, whereas Table 2 reports on correlation between bacteria-shaped object and other shapes. These calculations are conducted on 200 × 200 pixel images rotated by angles 0˚, 45˚, 90˚ and 135˚ in a counterclockwise direction around the center of the image. In this experiment, all F_n are calculated for each of the four selected angles yielding: $F_{n^{0°}}, F_{n^{45°}}, F_{n^{90°}}, F_{n^{135°}}$ (e.g. for round shape we have $F_{r^{0°}}, F_{r^{45°}}, F_{r^{90°}}, F_{r^{135°}}$). Note that F_n is equal to $F_{n^{0°}}$. Calculated data show significant impact of the object shape on the coefficient value for vectors of a and α, and that coefficient value is almost independent from the object size and rotation.

Despite the fact that Pearson coefficient properly reflects the relationship between shapes (expressed in vector forms and compared respectively), in the case of bacteria comparison, better classification results can be obtained upon replacing this coefficient with a slightly different approach presented in the following example.

To compare two vectors (e.g. representing some abstract feature), assume that the similarity between two vectors $w_1 = [1, 2, 3, 4]$ and $w_0 = [4, 1, 2, 3]$ is to be established. In doing so, the normalization of both vectors renders $w_{1n} = [0, 0.33, 0.66, 1]$ and $w_{0n} = [1, 0, 0.33, 0.66]$. Then

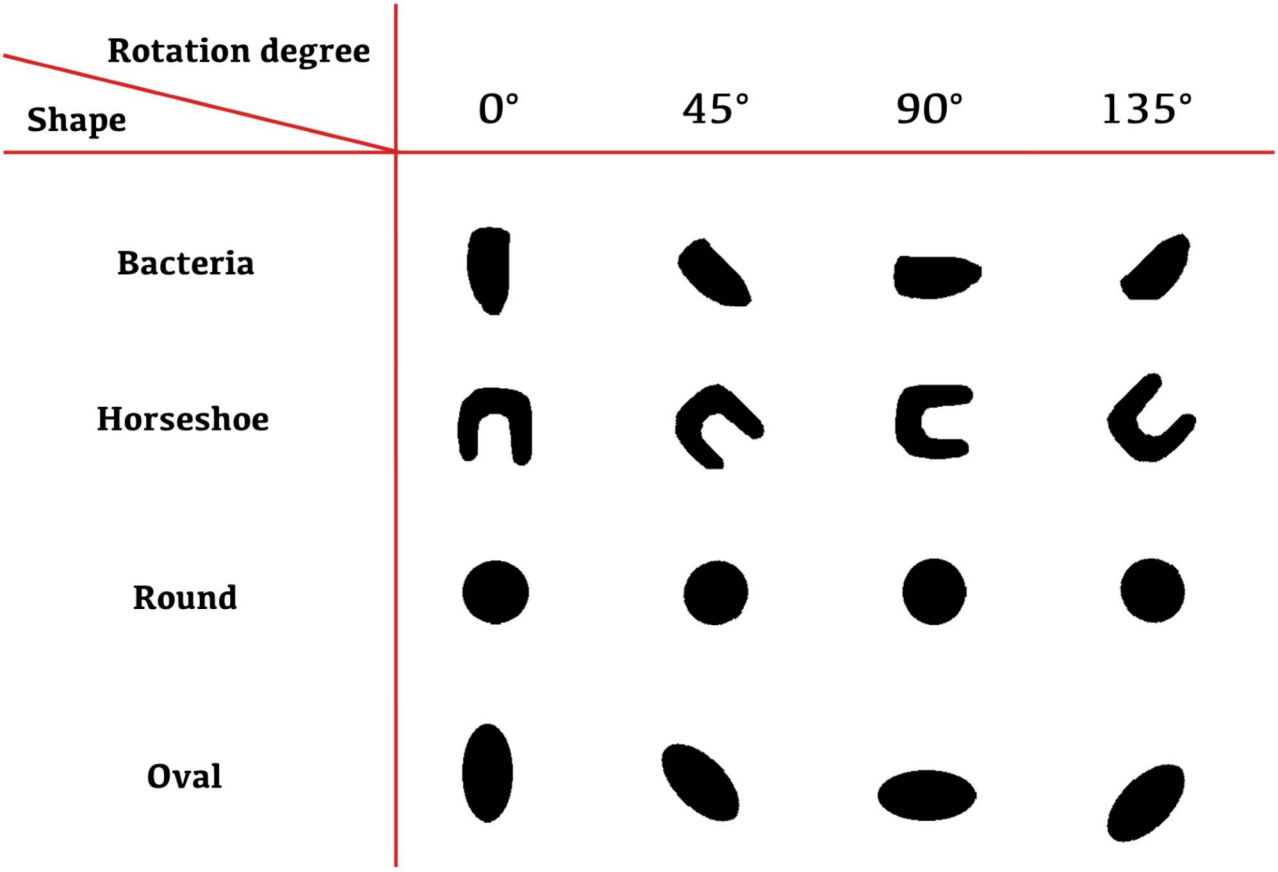


Fig 3. Shapes that were compared with a bacteria-like shape.

https://doi.org/10.1371/journal.pone.0293362.g003

three more vectors are created as they correspond to different positions of the object (on which w_1 was calculated): w_{2n} = [1, 0, 0.33, 0.66], w_{3n} = [0.66, 1, 0, 0.33] and w_{4n} = [0.33, 0.66, 1, 0]. In the next step, we calculate mean squared error [22] between each w_i (for $i = 1, \ldots, 4$) and w_0 which is equal to: 1.33, 0, 1.33 and 1.78, respectively. Then the smallest value is chosen, which here reads as 0 meaning that both w_{0n} and w_{in} are equal.

Vector w_1 can represent a certain vector calculated on currently analyzed shape (e.g. a_n from F_n) and w_0 stands for a corresponding vector calculated on a bacteria-like shape (e.g. a_{bs}

Table 1. Table presents Pearson coefficient between vectors calculated on bacteria-shaped object F_{bs} and the same object rotated ($F_{bs^{0°}}$, $F_{bs^{45°}}$, $F_{bs^{90°}}$, $F_{bs^{135°}}$) and F_{bs} with vectors calculated for bacteria-shaped object twice magnified rotated ($F_{b2^{0°}}$, $F_{b2^{45°}}$, $F_{b2^{90°}}$, $F_{b2^{135°}}$).

SV	$F_{bs^{0°}}$	$F_{bs^{45°}}$	$F_{bs^{90°}}$	$F_{bs^{135°}}$	$F_{b2^{0°}}$	$F_{b2^{45°}}$	$F_{b2^{90°}}$	$F_{b2^{135°}}$
$\vec{a}$	**1.00**	**0.98**	**1.00**	**0.98**	**0.99**	**0.98**	**0.99**	**0.98**
$\vec{b}$	1.00	0.30	1.00	0.30	0.03	0.36	0.03	0.32
$\vec{\alpha}$	**1.00**	**0.97**	**1.00**	**0.97**	**0.99**	**0.97**	**0.99**	**0.97**
$\vec{\beta}$	1.00	0.92	1.00	0.92	0.97	0.93	0.97	0.93

SV stands here for the Set of Vectors which is a set that consists of vectors corresponding to $\vec{a}$, $\vec{b}$, $\vec{\alpha}$ and $\vec{\beta}$ that are compared with the corresponding vectors in F_{bs} with the Pearson Coefficient.

https://doi.org/10.1371/journal.pone.0293362.t001

Table 2. Table presents pearson coefficient between vectors calculated on bacteria-shaped object F_{bs} with vectors calculated for horseshoe F_h, round F_r and oval F_o shapes rotated by 0˚, 45˚, 90˚ and 135˚.

SV	$F_{h0°}$	$F_{h45°}$	$F_{h90°}$	$F_{h135°}$	$F_{r0°}$	$F_{r45°}$	$F_{r90°}$	$F_{r135°}$	$F_{o0°}$	$F_{o45°}$	$F_{o90°}$	$F_{o135°}$
$\vec{a}$	**0.25**	**0.23**	**0.25**	**0.23**	**0.51**	**0.56**	**0.51**	**0.56**	**0.97**	**0.98**	**0.97**	**0.98**
$\vec{b}$	0.28	0.33	0.28	0.33	0.03	0.05	0.03	0.05	0.29	0.23	0.29	0.23
$\vec{\alpha}$	**0.07**	**0.16**	**0.07**	**0.16**	**0.46**	**0.60**	**0.46**	**0.60**	**0.91**	**0.92**	**0.91**	**0.92**
$\vec{\beta}$	0.19	0.19	0.19	0.19	0.20	0.35	0.20	0.35	0.88	0.90	0.88	0.90

SV stands here for the Set of Vectors which is a set that consists of vectors corresponding to $\vec{a}$, $\vec{b}$, $\vec{\alpha}$ and $\vec{\beta}$ that are compared with the corresponding vectors in F_{bs} with the Pearson Coefficient.

https://doi.org/10.1371/journal.pone.0293362.t002

from F_{bs}). Here all the shapes from Fig 3 are compared with the bacteria-like shape. Vectors w_1 and w_0 can also represent a vector calculated on currently analyzed bacteria (e.g. a_k from F_k) and on the exemplary one (e.g. a_e from F_e). These dependencies are calculated for every selected bacteria in the analyzed image.

Subsequently, the minimum mean square error value is computed for specific vectors corresponding to each of the b chosen bacteria in the image, across all four vectors. The corresponding results are denoted by $MSE\vec{a}_k^{min}$, $MSE\vec{b}_k^{min}$, $MSE\vec{\alpha}_k^{min}$, $MSE\vec{\beta}_k^{min}$, where $k = 1, 2, \ldots, b$. Then respective mean values of the minimum values for vectors $\vec{a}, \vec{b}, \vec{\alpha}, \vec{\beta}$ are calculated for all the selected bacteria from an analyzed image rendering four features based on geometry: $\overline{MSE\vec{a}^{min}}$, $\overline{MSE\vec{b}^{min}}$, $\overline{MSE\vec{\alpha}^{min}}$, $\overline{MSE\vec{\beta}^{min}}$.

Curvature and arc-length. Having selected $\hat{\mathcal{Q}}_n$ points (described in previous subsection) one can estimate the object's boundary with the aid of interpolation [23]. In order to define any interpolant γ which graph forms a closed curve the set $\hat{\mathcal{Q}}_n$ is augmented with an extra point $\hat{q}_{n+1} = \hat{q}_0$. The missing interpolation knots $\{\hat{t}_i\}_{i=0}^{n+1}$ for which $\hat{q}_i = \gamma(\hat{t}_i)$ are estimated from *exponential parameterization* [24, 25]:

$$\hat{t}_i = 0, \quad \hat{t}_{i+1} = \hat{t}_i + \|q_{i+1} - q_i\|^{\lambda}, \quad i = 0, \cdots, n$$

with $\lambda \in [0, 1]$. Here a special case of $\lambda = 0.5$ (the so-called *centripetal parameterization*) is used. Next a cubic spline $\gamma = \hat{\gamma}^{cs}$ with clamped boundary conditions [26] is applied (a complete spline). The latter requires an a priori information on $\hat{\gamma}'(\hat{t}_0) = v_0$ and $\hat{\gamma}'(\hat{t}_{n+1}) = v_{n+1}$ which is originally unavailable. In order to extract somehow v_0 and v_{n+1} an approach based on Modified Hermite scheme is used [27], where both v_0 and v_{n+1} are estimated from Lagrange Cubics $\hat{\gamma}_0^C$, $\hat{\gamma}_{n-2}^C$ fitting $\{\hat{q}_0, \hat{q}_1, \hat{q}_2, \hat{q}_3\}$ and $\{\hat{q}_{n-2}, \hat{q}_{n-1}, \hat{q}_n, \hat{q}_{n+1}\}$ yielding $v_0 = \hat{\gamma}_0^{C'}(\hat{t}_0)$ and $v_{n+1} = \hat{\gamma}_{n-2}^{C'}(\hat{t}_{n+1})$, respectively.

Having constructed a complete spline $\gamma = \hat{\gamma}^{cs}$ one can compute its curvature:

$$\kappa(t) = \frac{\|\vec{T}'(t)\|}{\|\vec{r}\,'(t)\|},$$

where $\vec{r}(t) = \dot{\gamma}(t)$ is a tangent vector to γ at t with its normalized vector $\vec{T}(t) = \vec{r}(t)/\|\vec{r}(t)\|$ or arc-length of the curve γ on interval $[a, t]$:

$$s = \int_a^t \|\vec{r}\,'(u)\|\, du.$$

All geometric features. The final set of features based on size and geometry of the selected objects reads as:

- {1} *Mean bacteria arc-length*—which is a sum of all $n + 1$ arc-lengths representing the perimeters of all selected bacteria divided by b,
- {2} *Mean curvature of b bacteria in one image*—a sum of all integrals of a curvature $\kappa(t)$ on each of the $[t_i, t_{i+1}] \ni t$ intervals calculated for each bacteria where $i = 0, 1, \ldots, n$. Then the sum of integrals is divided by b,
- {3} *Minimal mean square error first distance*—$\overline{MSE\vec{a}^{min}} = (1/b)\sum_{k=1}^{b} MSE\vec{a}_k^{min}$,
- {4} *Minimal mean square error second distance*—$\overline{MSE\vec{b}^{min}} = (1/b)\sum_{k=1}^{b} MSE\vec{b}_k^{min}$,
- {5} *Minimal mean square error first angle*—$\overline{MSE\vec{\alpha}^{min}} = (1/b)\sum_{k=1}^{b} MSE\vec{\alpha}_k^{min}$,
- {6} *Minimal mean square error second angle*—$\overline{MSE\vec{\beta}^{min}} = (1/b)\sum_{k=1}^{b} MSE\vec{\beta}_k^{min}$,
- {7} *Median of the object area in the image*,
- {8} *Percent of the bacteria area in the image*,
- {9} *Amount of objects in the image*—calculated sum of objects within the ROI mask,
- {10} *Amount of bacteria in the image*—calculated sum of the areas of objects within a ROI mask divided by the median of the object size in the current image.

Calculation of dispersion features

The dataset analyzed in this research consists of the images with bacteria monocultures. Each soil bacteria genera has a different colony dispersion. For some genera the bacteria cells are located close to each other in a non-uniform fashion, whereas the others are equally distributed. This section outlines the possible tools which measure the impact of such irregularities on classification in terms of mean shift [28], k-means [29] and regression [30].

Mean shift. Mean shift [28] is a scheme that allocates points through an iterative procedure to their average in a specified neighborhood (the local maxima of a density function) [31]. The output of this method consists of sets of points assigned to disjoint clusters determined by the distribution of input points. The resulting number of clusters in a clustering algorithm is determined by the algorithm itself. However, there are several input parameters that can be adjusted to customize the clustering process. These parameters include the window size, the distance metric used to evaluate the proximity of points to the cluster center and the stopping criteria for the algorithm. The mean shift algorithm flowchart is illustrated in Fig 4.

It is assumed here that the input data points of the mean shift algorithm are the centroids of the objects on the ROI mask captured with *props* MATLAB function. The generated features are the numbers of clusters to which the points were attached applying different values of r which is the radius of the window. The implementation of mean shift algorithm applied in this research can be found in MathWorks [32].

K-means. K-means [29] is a method that assigns points into k clusters. The algorithm is an iterative procedure of calculating distances between points and centroids, and shifting the centroids to new locations. The value of k is set arbitrarily. The flowchart of this algorithm is presented in Fig 5.

In order to determine features based on dispersion k-means method is firstly applied to cluster bacteria centroids. The latter incorporates their location in (x, y) coordinate system or both

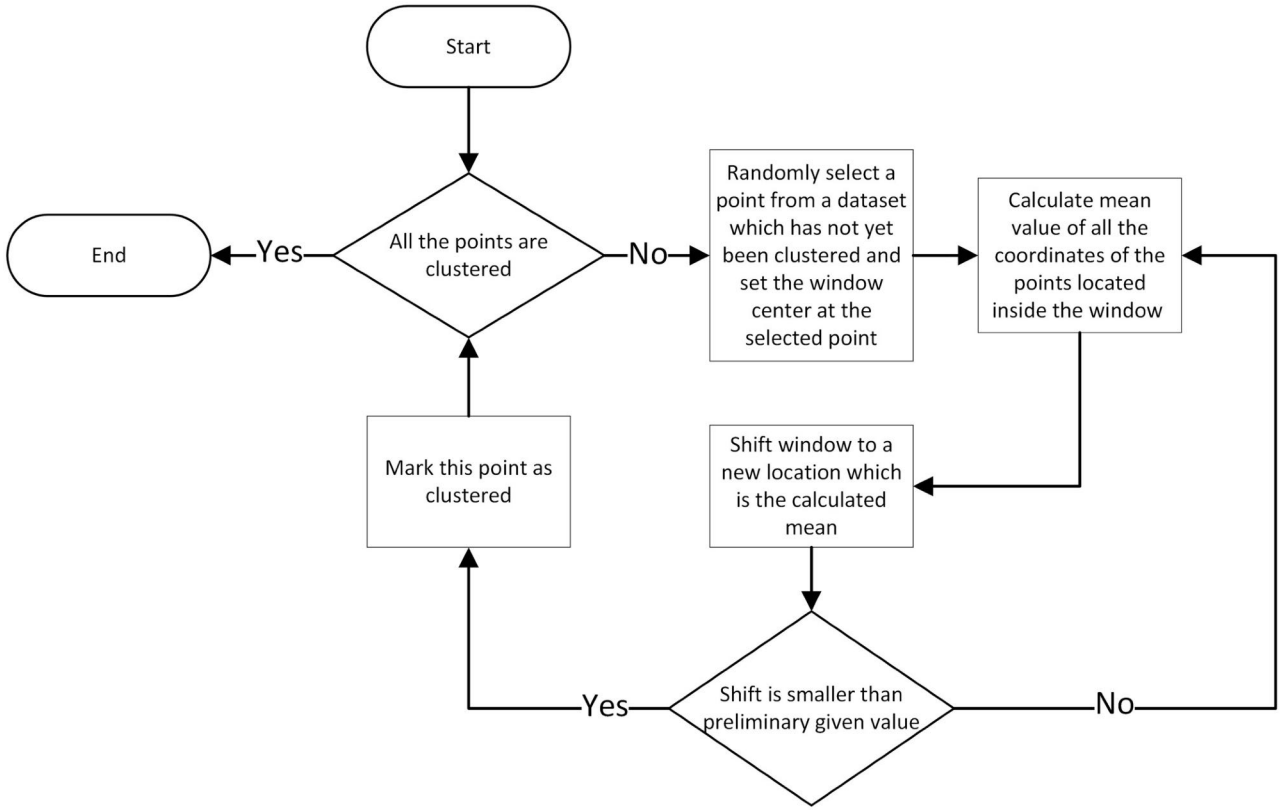


Fig 4. Flowchart of the mean shift algorithm.

https://doi.org/10.1371/journal.pone.0293362.g004

the Cartesian location combined with the area of the bacteria cell represented by (x, y, s). Assume the points $\mathcal{P}_{z_j} = \{p_i\}_{i=0}^{z_j}$ are given, where $z_j + 1$ defines the amount of points associated with the centroid c. Then a linear regression line is fitted to all points from $\mathcal{P}_{z_j}$. Let $\mathcal{Q}_{z_j} = \{q_i\}_{i=0}^{z_j}$ be the points on the fitted linear regression line such that $q_i = (x_{q_i}, y_{q_i})$ and each $x_{q_i} = x_i$

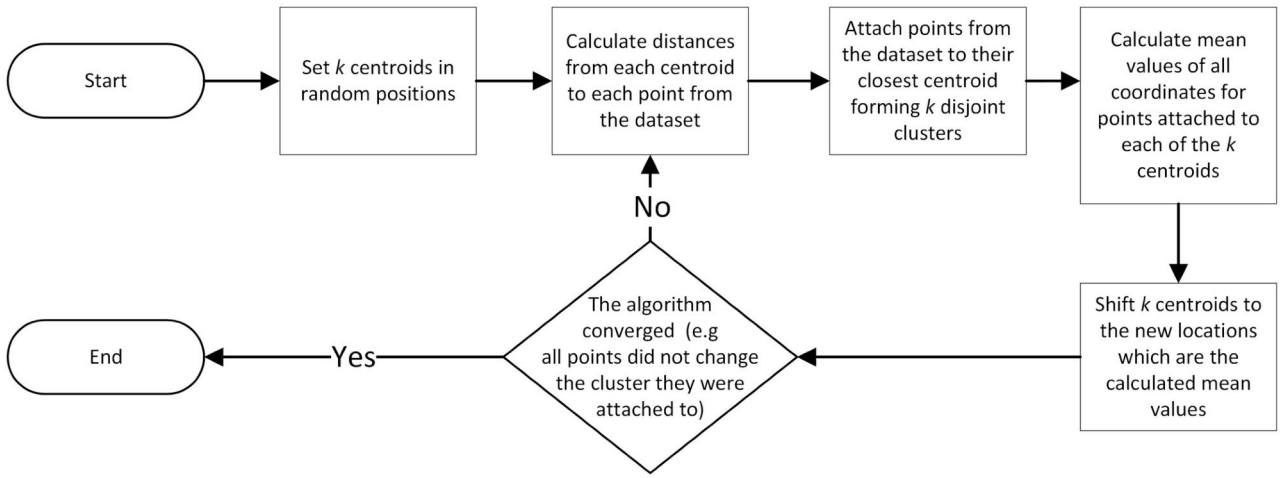


Fig 5. Flowchart of the K-means algorithm.

https://doi.org/10.1371/journal.pone.0293362.g005

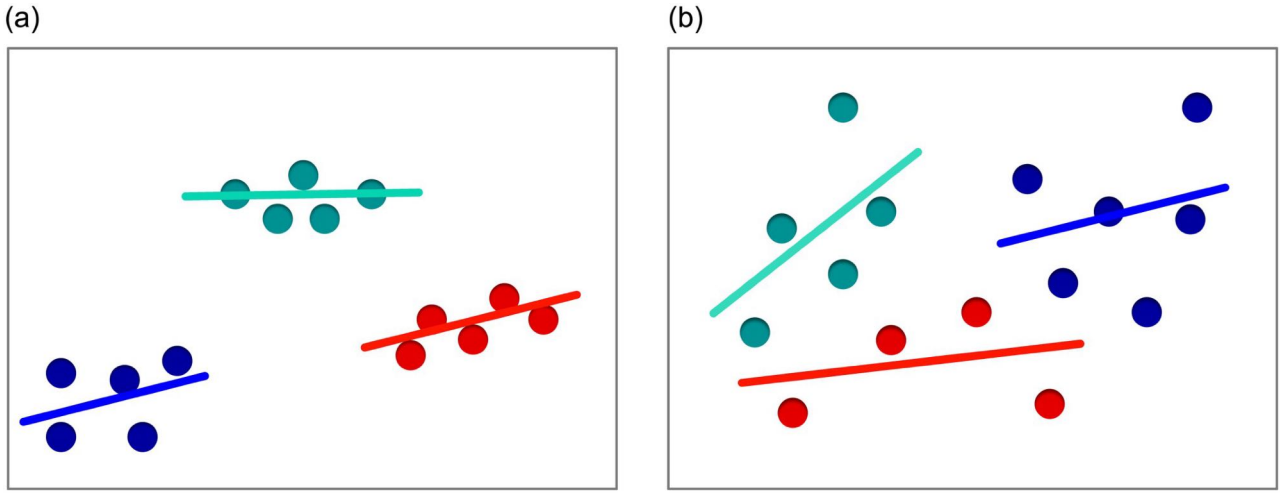


Fig 6. K-means algorithm and linear regression image with m + 1 = 15 for k = 3 put in sets (a) and evenly spaced (b).

https://doi.org/10.1371/journal.pone.0293362.g006

for x_i being first coordinate of the point $p_i \in \mathcal{P}_{z_j}$. Next $\bar{d}_j = (1/(z_j + 1)) \sum_{i=0}^{z_j} \|y_i - y_{q_i}\|$, which is the mean distance between each of the corresponding points p_i and q_i for j'th centroid, is calculated. Note that $\bar{d}_j$ can also be weighted by the values of the normalized vector of bacteria surface areas s_i computed as $\bar{d}_j = (1/\sum_{i=0}^{z_j} s_i) \sum_{i=0}^{z_j} \|y_i - y_{q_i}\| s_i$. Such procedure is repeated for each of the k clusters. The resulting sum $\bar{D} = \sum_{j=1}^{k} \bar{d}_j$ becomes the feature value for currently analyzed image.

To provide a clear example, consider two sets of $m + 1 = 15$ points. The first set is composed of points grouped into three subsets, while the second set contains evenly spaced samples. The points in both sets are attached to $k = 3$ clusters by k-means algorithm. Next, one calculates the values of $\bar{D}$ for both datasets as specified in the preceding paragraph. The computed values of $\bar{D}$ for images from Fig 6 are equal to $\bar{D} = 68$ for Fig 6a and $\bar{D} = 1058$ for Fig 6b. A marked discrepancy is observed in the results depending on the level of data dispersion. Fig 7 illustrates the latter used for the exemplary microscopic images.

All dispersion features. The final set of features based on the size and geometry consists of:

- {11-18} *Mean shift*—for different r values equal to = 25, 50, 75, 100, 125, 150, 175, 200, respectively,
- {19-27} *K-means and regression*—for (x, y) with $k = 2, 6, 10$, for (x, y, s) with $k = 2, 6, 10$ and for (x, y) with $k = 2, 6, 10$ weighted.

Calculation of luminance and chrominance features

In this work, the classification results obtained applying features based on geometry and dispersion is compared with an accuracy rendered by features based on chrominance and luminance. In doing so, statistical measures of the pixel values, i.e. colors defined by RGB (red, green and blue) color space on the whole image or only on the area covered with the ROI mask are computed. Such features are calculated either on the image converted to grayscale or within a selected color channel. The four statistical measures employed here to analyze the

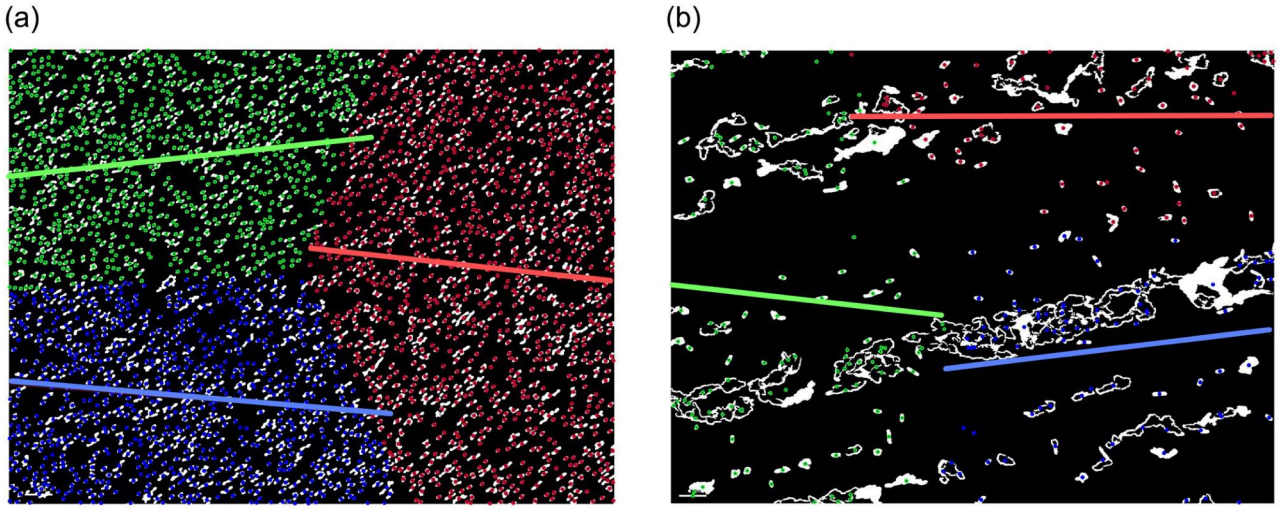


Fig 7. K-means algorithm and linear regression for microscopic image of *Rhizobium* (a) and *Enterobacter* (b).

https://doi.org/10.1371/journal.pone.0293362.g007

data are: variance [33], mean, kurtosis and skewness [34]. The resulting set of features based on color consists of:

- {28-35} *Variance*—calculated on mask {28}, on whole image {29}, on whole image red {30}, green {31}, blue channel {32}, calculated on mask for red {33}, green {34}, blue channel {35},
- {36-43} *Mean*—calculated on mask {36}, on whole image {37}, on whole image red {38}, green {39}, blue channel {40}, calculated on mask for red {41}, green {42}, blue channel {43},
- {44-51} *Kurtosis*—calculated on whole image {44}, on the mask {45}, on whole image red {46}, green {47}, blue channel {48}, on the mask for red {49}, green {50}, blue channel {51},
- {52-59} *Skewness*—calculated on whole image {52}, on the mask {53}, on whole image red {54}, green {55}, blue channel {56}, on the mask for red {57}, green {58}, blue channel {59}.

Feature selection

Noticeably, not all calculated features are appropriate for the classification. Some of them are not highly correlated with the affiliation to the class or their correlation with other features is too high which might cause redundancy. Such features should not be considered in the stage of class recognition. The feature selection methods solve this problem by picking appropriate features. In this work, we decided to verify the results given by the following methods:

- Fast Correlation Based Filter (FCBF) [35],
- Sparse Multinomial Logistic Regression with Bayesian Regularization (SBMLR) [36],
- Correlation-based Feature Selection (CFS) [37].

Class recognition

Class recognition methods are used here to assign input images to certain classes representing different bacteria genera. These methods are trained on the training set and their classification

performance is measured upon applying the testing set. Such sets contain selected features calculated for each of the images. Class recognition methods considered here include:

- Support Vector Machine (SVM) [38],
- Random Forest (RF) [39],
- K-Nearest Neighbors (KNN) [40],
- Multi-Layer Perceptron (MLP) [41].

These methods representing classical machine learning techniques rely on admitting features a priori determined by hand. Such class recognition methods continue to be widely used across a diverse range of applications [42]. In particular, these AI tools are also studied in the context of soil microorganism classification with high accuracy results reported [18].

Random forest. Random forest [39] is a group learning method whose task is to generate a set of models—trees, and then to classify the tested object into one of the classes taking into consideration the results from individual models. The trees are built based on the features table with known class assignment (supervised learning). Each node of the tree has conditions for numeric or non-numeric data. Satisfying these conditions determines object affiliation to one of the classes by the current model. In order to create a decision tree for RF (based on a table of features) one firstly randomly selects a subset of samples (table rows) with repetitions and places them into a so-called bootstrap dataset (it has as many rows as the input table of features) [43]. Having created the new dataset we draw from it x features (table columns) and verify which one will be the best for building the model (correctly separates the samples). The decision on which of the x features is to be used at a given tree node is made on the basis of methods such as e.g. Gini Impurity or Entropy [43]. The same measures allow us to set a threshold for condition concerning numeric data for a given feature. For the classification purposes hundreds of trees are generated. Upon creation of n trees one verifies to which class a new instance is assigned by each of the models. The final decision on the classification is made according to the majority voting rule. The effectiveness of this method is examined by comparing the achieved affiliation to a class by means of the algorithm with the actual instance assignment. One can arbitrarily select the value of n, however, with n increasing, the computational complexity of the algorithm explodes, resulting in a longer computation time. In this work the *TreeBagger* MATLAB function was applied.

Results

The dataset considered here [44] consists of 128 microscopic images of soil bacteria from the five selected genera: *Enterobacter*—22 images, *Rhizobium*—25 images, *Pantoea*—26 images, *Bradyrhizobium*—34 images and *Pseudomonas*—21 images. These images have not been preprocessed either by the microbiologists (no chemical reactions conducted) or by any computerized system. In the experimental section the concept of cross validation [45] is applied. More specifically, 10% ratio cross validation is used, in which the set of images is randomly shuffled and divided into ten subsets. Next, nine of these sets are selected to form the training set on which our model learns how to distinguish input objects among different classes. The remaining set (called the testing set) is used to verify how good the result of classification is by calculating its accuracy. The model accuracy represents the amount of correctly classified bacteria images divided by the amount of the whole set of images (in the testing set). Then another of the ten sets becomes the testing set so that we have ten iterations (ten different training and testing sets) and calculate the mean accuracy value of ten iterations. The tables in this section display the mean accuracy resulting from 50 iterations of 10% cross-validation.

Table 3. The accuracy obtained with different feature selection and classification methods performed on features based on geometry and dispersion for the five bacteria genera.

FSM	SVM	RF (n = 200)	KNN (k = 1)	MLP	Selected features
None	78.3438	**85.1406**	80.5938	50.9375	{1-27}
FCBF	75.0312	82.2656	79.0469	63.5	{7, 11, 2, 5, 8}
SBMLR	40.8906	36.2344	36.2344	28.1094	{1}
CFS	78.3906	83.875	79.8594	60.1719	{1, 2, 5-7, 10, 11, 14}

FSM stands here for the Feature Selection Method.

https://doi.org/10.1371/journal.pone.0293362.t003

The final results include calculations based on iteratively selected parameters of class recognition methods which are: Support Vector Machine (with default parameters in *fitcsvm* MATLAB function), Random Forest with 200 trees, K-Nearest Neighbors with $k = 1$ and Multi-Layer Perceptron with network topology 15 – 15 – 15 trained with backpropagation algorithm based on gradient descent. The parameters were selected to maximize the ACC.

The accuracy for the whole set of features consisting of geometry and dispersion traits shown in Table 3 reached 85.14% for Random Forest for the whole set of five different bacteria genera. Applying feature selection methods does not increase the achieved result. The results for four different bacteria genera presented in Table 4 are also the highest for Random Forest ranging from 81.7% to 91.6%.

Table 5 presents accuracy for different sets and subsets of features. Features based on dispersion obtained the best accuracy of 63.72% for KNN, whereas features based on geometry reached 82.59% for Random Forest. Combining these sets increases the result by 2.55

Table 4. The accuracy computed with different classification methods performed on the whole set of features based on geometry and dispersion for four selected bacteria genera (subsets of five bacteria genera).

Selected bacteria genera	SVM	RF (n = 200)	KNN (k = 1)	MLP
1, 2, 3, 4	81.7944	86.2243	85.9252	58.8785
2, 3, 4, 5	76.1132	81.6792	80.3208	54.9057
3, 4, 5, 1	75.8641	84.1748	78.6408	51.9223
4, 5, 1, 2	88.3922	88.4314	86.8235	64.8627
5, 1, 2, 3	86.0638	**91.5532**	87.8298	69.617

1, 2, 3, 4 and 5 stand for the following bacteria genera respectively: *Enterobacter*, *Rhizobium*, *Pantoea*, *Bradyrhizobium* and *Pseudomonas*.

https://doi.org/10.1371/journal.pone.0293362.t004

Table 5. The accuracy obtained with different classification methods performed on different sets of features based on color, geometry and dispersion (and their subsets).

Set of features based on. . .	SVM	RF (n = 200)	KNN (k = 1)	MLP
geometry and dispersion {1-27}	78.3438	**85.1406**	80.5938	50.9375
geometry {1-10}	74.7969	**82.5938**	70.9062	53.0312
vectors (from geometry) {3-6}	47.5469	**48.5469**	41.9844	32.0625
dispersion {11-27}	60.4531	63.4219	**63.7188**	42.5312
k-means (from dispersion) {19-27}	39.8281	**46.8594**	43.7188	24.7976
mean shift (from dispersion) {11-18}	53.7188	**63.9375**	63.2812	49.9219
color {28-59}	86.2188	94.2969	**95.5938**	88.3906
color, geometry and dispersion {1-59}	89.7969	**94.8281**	91.2031	75.4688

https://doi.org/10.1371/journal.pone.0293362.t005

Table 6. The accuracy with different classification methods for the five bacteria genera performed on features based on k-means and regression. The set consists of 60 features, for k = 1, . . ., 20 and 2 dimensional vector for k-means, 2 dimensional vector for k-means with weighted variance and 3 dimensional vector for k-means.

Quantile	SVM	RF (n = 200)	KNN (k = 1)	MLP
0	48.0469	58.9062	48.2031	28.2031
0.2	45.7812	60.625	41.875	26.5625
0.4	50.8594	**61.7969**	44.5312	26.0938
0.6	40	50.3125	36.6406	23.5938

The value in Quantile column informs that the bacteria were taken into account if their area value was greater then certain quantile of area value of the bacteria on a given image.

https://doi.org/10.1371/journal.pone.0293362.t006

percentage points and amounts to 85.14% for Random Forest. One can also analyze the results of the selected subsets of features based on dispersion and geometry. As an example, features extracted based on mean shift yields up to 63.94% accuracy, whereas features based on *k*-means render 46.86%. Applying features based on vectors reached only 46.86% accuracy; however, one of them—feature number 5—is accepted by both FCBF and CFS feature selection methods which is shown in Table 3 what proves its significant impact on increasing the classification accuracy.

The features calculated with *k*-means may seem insignificant. For this reason the classification results are presented depending on the amount of bacteria analyzed on a single image based on their area value. For each of the calculations, as shown in Table 6, a different quantile value is selected which means that analyzed bacteria are ones which area exceeds or is equal to that quantile area value in a single image. The 60 features were calculated for *k*-means for each of the quantiles: 0, 0.2, 0.4 and 0.6. For example for quantile equal to 0.2 the calculated features are: for k = 1, . . ., 20 and 2 dimensional vector for k-means, 2 dimensional vector for k-means with weighted variance and 3 dimensional vector for k-means—yielding 60 features for this quantile. The highest results are reached for the quantile equal to 0.4 amounting to 61.8%. It is remarked here that extending the final set of features by these 60 *k*-means features does not improve the final result. For that sheer reason only nine previously calculated features based on *k*-means are chosen.

In this research the highest classification accuracy for a set of geometry and dispersion features yields 85.14%. In previous work [19], based on the same image data set, the accuracy obtained amounts to 97%. The latter analyses different set of features, involving geometric and texture characteristics. The texture features rely on luminance and chrominance, which may artificially improve the accuracy of the results. For example, this may occur when microscopic images from each genera are taken under different lighting conditions. Thus, the obtained accuracy 85.14% forms a promising result as the examined features are not based on color information.

Discussion

The classification based on extracting features from bacteria geometry and dispersion yields a promising 85.14% ACC. The latter is reached for the Random Forest classifier to identify five selected soil bacteria genera. The experiments conducted on features based on geometry and dispersion separately rendered 82.59% in case of Random Forest and 63.72% for K-Nearest Neighbors. These results illustrate that applying a proper set of features with no color traits enables classification of soil bacteria. The latter permits to bypass the impact of lighting conditions and coloring of samples on classification. In contrast the geometry and dispersion based

classification is insensitive to the last two factors. However, the difference between the classification accuracy based on geometry and dispersion traits versus this one based on color traits is significant (around 10 percentage points) and there are some issues requiring future research investigation.

Indeed, one needs to apply a different method of selecting points on bacteria boundary to highlight the characteristic elements of its shape. In addition, various parameterizations to estimate the unknown interpolation knots combined with different interpolation fitting schemes might also be considered [46]. Other methods for object dispersion in the image should also be examined. In this work we compared the results given by the four classification methods: Support Vector Machine, Random Forest, K-Nearest Neighbors and Multi-Layer Perceptron. Other classifiers such as Extreme Learning Machines or Deep Learning Methods may provide more effective recognition tools. The features in the future research can be also computed applying Convolutional Neural Networks [47].

The generated results are calculated on the dataset with a single bacteria genera on an input image. These organisms were grown under laboratory conditions, with no contamination involved (as they are all immersed in uniform medium). In future research, the testing should also be performed on images taken from the natural environment (e.g. from the genuine rizosphere sample). The ultimate goal is to classify different bacteria genera mixed with extra organic or non-organic objects as they cohabit in a real soil sample. More importantly, the classification results on images that contain different bacteria genera (for example mixes of two or three genera on one image) should also be examined. In particular, the final recognition tool should allow to assess the quantity of bacteria cells affiliated to a certain genera on the currently analyzed microscopic image.

The classification system created in this work can be applied in practice. However, further research is needed for samples containing strains of different species of bacteria representing the same genus. These species differ in phenotypic features (morphological, biochemical and physiological). The number of analyzed strains of bacteria has an important meaning. We are unable to draw a conclusion from a single photograph of cells or bacterial colonies known to be of some type of bacteria. As an example, the genus *Pseudomonas* includes both fluorescent and non-fluorescent bacterial species. Problems with identifying bacteria based on their morphology result from reasons such as: (i) the influence of the environment, i.e. the composition of the medium and incubation time on the cell morphology, (ii) the phase of the bacterial cell cycle, (iii) the common morphology of cells of different types of bacteria. It is worth mentioning that so far there are over 10 thousand species of culturable bacteria, with a huge number of species that cannot be cultured in the laboratory. It is very important to accurately classify the bacteria as a representative of the appropriate species. The latter permits to decide whether to use it for utilitarian purpose e.g. in biological protection of plants against diseases or to apply suitable control against a given organism if it causes diseases (pathogen) or is harmful in any other respect.

Author Contributions

Conceptualization: Aleksandra Konopka, Ryszard Kozera.

Data curation: Aleksandra Konopka.

Formal analysis: Aleksandra Konopka.

Methodology: Aleksandra Konopka.

Resources: Lidia Sas-Paszt, Pawel Trzcinski, Anna Lisek.

Software: Aleksandra Konopka.

Visualization: Aleksandra Konopka.

Writing – original draft: Aleksandra Konopka, Ryszard Kozera.

Writing – review & editing: Aleksandra Konopka, Ryszard Kozera.

References

1. Sharma A, Lee S, Park YS. Molecular typing tools for identifying and characterizing lactic acid bacteria: a review. Food Sci Biotechnol. 2020; 29(10):1301–1318. https://doi.org/10.1007/s10068-020-00802-x PMID: 32995049
2. Maiden MC, Van Rensburg MJJ, Bray JE, Earle SG, Ford SA, Jolley KA, et al. MLST revisited: the gene-by-gene approach to bacterial genomics. Nat Rev Microbiol. 2013; 11(10):728–736. https://doi.org/10.1038/nrmicro3093 PMID: 23979428
3. Numberger D, Ganzert L, Zoccarato L, Mühldorfer K, Sauer S, Grossart HP, et al. Characterization of bacterial communities in wastewater with enhanced taxonomic resolution by full-length 16S rRNA sequencing. Sci Rep. 2019; 9. https://doi.org/10.1038/s41598-019-46015-z PMID: 31273307
4. Church DL, Cerutti L, Gürtler A, Griener T, Zelazny A, Emler S. Performance and application of 16S rRNA gene cycle sequencing for routine identification of bacteria in the clinical microbiology laboratory. Clin Microbiol Rev. 2020; 33(4). https://doi.org/10.1128/CMR.00053-19 PMID: 32907806
5. Floridia-Yapur N, Rusman F, Diosque P, Tomasini N. Genome data vs MLST for exploring intraspecific evolutionary history in bacteria: much is not always better. Infect Genet Evol. 2021; 93. https://doi.org/10.1016/j.meegid.2021.104990 PMID: 34224899
6. Caprette DR. Describing Colony Morphology [Internet]; 2022 [cited 2022 Sep 13]. Available from: https://bit.ly/324cqkA.
7. Tarca AL, Carey VJ, Chen X, Romero R, Drăghici S. Machine learning and its applications to biology. PLoS Comput Biol. 2007; 3(6):e116. https://doi.org/10.1371/journal.pcbi.0030116 PMID: 17604446
8. Rani P, Kotwal S, Manhas J, Sharma V, Sharma S. Machine learning and deep learning based computational approaches in automatic microorganisms image recognition: methodologies, challenges, and developments. Arch Comput Methods Eng. 2021. https://doi.org/10.1007/s11831-021-09639-x PMID: 34483651
9. Khutlang R, Krishnan S, Dendere R, Whitelaw A, Veropoulos K, Learmonth G, et al. Classification of mycobacterium tuberculosis in images of ZN-stained sputum smears. IEEE trans inf technol. 2010; 14 (4):949–957. https://doi.org/10.1109/TITB.2009.2028339 PMID: 19726269
10. Kang R, Park B, Eady M, Ouyang Q, Chen K. Single-cell classification of foodborne pathogens using hyperspectral microscope imaging coupled with deep learning frameworks. Sensors and Actuators B: Chemical. 2020; 309:127789. https://doi.org/10.1016/j.snb.2020.127789
11. Hiremath PS, Bannigidad P. Identification and classification of cocci bacterial cells in digital microscopic images. Int J Comput Biol. 2011; 4(3):262. https://doi.org/10.1504/IJCBDD.2011.041414 PMID: 21778559
12. Kotwal S, Rani P, Arif T, Manhas J, Sharma S. Automated bacterial classifications using machine learning based computational techniques: architectures, challenges and open research issues. Arch Computat Methods Eng. 2022; 29:2469–2490. https://doi.org/10.1007/s11831-021-09660-0 PMID: 34658617
13. Liu J, Dazzo F, Glagoleva O, Yu B, Jain A. CMEIAS: a computer-aided system for the image analysis of bacterial morphotypes in microbial communities. Microb Ecol. 2001. https://doi.org/10.1007/s002480000004 PMID: 11391457
14. Ruusuvuori P, Seppälä J, Erkkilä T, Lehmussola A, Puhakka JA, Yli-Harja OP. Efficient automated method for image-based classification of microbial cells. 19th Int Conf Pattern Recognit. 2008.
15. Ducret A, Quardokus E, Brun Y. MicrobeJ, a tool for high throughput bacterial cell detection and quantitative analysis. Nat Microbiol. 2016. https://doi.org/10.1038/nmicrobiol.2016.77 PMID: 27572972
16. Batt CA, Tortorello ML. Encyclopedia of Food Microbiology. Academic Press; 2014.
17. Beeckmans S, Xie J. Glyoxylate cycle. Reference Module in Biomedical Sciences. 2015. https://doi.org/10.1016/B978-0-12-801238-3.02440-5

18. Kruk M, Kozera R, Osowski S, Trzciński P, Sas-Paszt L, Sumorok B, et al. Computerized classification system for the identification of soil microorganisms. Appl Math Inf. 2016; 10:21–31. https://doi.org/10.18576/amis/100103

19. Konopka A, Struniawski K, Kozera R, Trzciński P, Sas-Paszt L, Lisek A, et al. Classification of soil bacteria based on machine learning and image processing. In: ICCS 2022. Springer International Publishing; 2022. p. 263–277.

20. Soille P. Morphological image analysis: principles and applications. Springer-Verlag; 1999.

21. Okwonu FZ, Asaju BL, Irimisose AF. Breakdown analysis of pearson correlation coefficient and robust correlation methods. IOP Conference Series: Materials Science and Engineering. 2020;917(1).

22. Hodson TO, Over TM, Foks SS. Mean squared error, deconstructed. J Adv Model Earth Syst. 2021; 13 (12). https://doi.org/10.1029/2021MS002681

23. Caruso C, Quarta F. Interpolation methods comparison. Comput Math Appl. 1998; 35(12):109–126.

24. Kvasov B. Methods of shape-preserving spline approximation. World Scientific; 2000.

25. Kozera R, Noakes L, Wilkołazka M. Exponential parameterization to fit reduced data. Appl Math Comput. 2021; 391:1–19.

26. McClarren RG. Computational nuclear engineering and radiological science using python; 2018.

27. Kozera R. Curve modeling via interpolation based on multidimensional reduced data. Studia Informatica. 2004; 25(4B):1–140.

28. Fukunaga K, Hostetler L. The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Trans Inf Theory. 1975; 21(1):32–40. https://doi.org/10.1109/TIT.1975.1055330

29. Pérez-Ortega J, Almanza-Ortega NN, Vega-Villalobos A, Pazos-Rangel R, Zavala-Díaz C, Martínez-Rebollar A. The k-means algorithm evolution, in introduction to data science and machine learning. IntechOpen. 2019.

30. Schneider A, Hommel G, Blettner M. Linear regression analysis: part 14 of a series on evaluation of scientific publications. Dtsch Arztebl Int. 2010; 107:776–82. https://doi.org/10.3238/arztebl.2010.0776 PMID: 21116397

31. Cheng Y. Mean shift, mode seeking, and clustering. IEEE PAMI. 1995; 17(8):790–799. https://doi.org/10.1109/34.400568

32. Finkston B. Mean Shift Clustering, MATLAB Central File Exchange.; 2023 [cited 2023 Jan 4]. Available from: https://www.mathworks.com/matlabcentral/fileexchange/10161-mean-shift-clustering.

33. Wasserman L. All of statistics: a concise course in statistical inference. Springer texts in statistics; 2005.

34. Joanes DN, Gill CA. Comparing measures of sample skewness and kurtosis. J R Stat Soc. 1998; 47 (1):183–189.

35. Yu L, Liu H. Feature selection for high-dimensional data: a Fast Correlation-Based Filter solution. In: Proceedings, Twentieth International Conference on Machine Learning. vol. 2; 2003. p. 856–863.

36. Cawley G, Talbot N, Girolami M. Sparse Multinomial Logistic Regression via Bayesian L1 Regularisation. In: Advances in Neural Information Processing Systems. vol. 19. MIT Press; 2006. p. 209–216. Available from: https://proceedings.neurips.cc/paper/2006/file/b22b257ad0519d4500539da3c8bcf4dd-Paper.pdf.

37. Hall M. Correlation-based feature selection for machine learning [Ph.D. thesis]. The University od Waikato, Hamilton, New Zealand; 2000.

38. Cervantes J, Garcia-Lamont F, Rodríguez-Mazahua L, Lopez A. A comprehensive survey on support vector machine classification: applications, challenges and trends. Neurocomputing. 2020; 408:189–215. https://doi.org/10.1016/j.neucom.2019.10.118

39. Ho TK. Random decision forests. In: Proceedings of 3rd International Conference on Document Analysis and Recognition. vol. 1; 1995. p. 278–282.

40. Fix E, Hodges JL. Discriminatory analysis. Nonparametric discrimination: consistency properties. USAF School of Aviation Medicine. 1951;.

41. Popescu MC, Balas V, Perescu-Popescu L, Mastorakis N. Multilayer perceptron and neural networks. WSEAS Transactions on Circuits and Systems. 2009; 8.

42. Halim Z, Hussain S, Hashim Ali R. Identifying content unaware features influencing popularity of videos on YouTube: A study based on seven regions. Expert Syst Appl. 2022. https://doi.org/10.1016/j.eswa.2022.117836

43. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning: with applications in R. Springer; 2013.

44. Konopka A. Zenodo Repository: scientistit/PLOS-2023: v1.0.0; 2023 [cited 2023 Aug 18]. Available from: https://zenodo.org/record/7789436.

45. Stone M. Cross-validatory choice and assessment of statistical predictions. J R Stat Soc Series B Stat Methodol. 1974; 36(2):111–133.

46. de Boor C. A practical guide to splines. Springer; 2001.

47. Zieliński B, Plichta A, Misztal K, Spurek P, Brzychczy-Włoch M, Ochońska D. Deep learning approach to bacterial colony classification. PloS ONE. 2017. https://doi.org/10.1371/journal.pone.0184554 PMID: 28910352

3.3 Performance analysis of Residual Neural Networks in soil bacteria microscopic image classification

Publication: A. Konopka, K. Struniawski, and R. Kozera, "Performance analysis of Residual Neural Networks in soil bacteria microscopic image classification," in *Modelling and Simulation'2023: The 2023 European Simulation and Modelling Conference, October 24–26, 2023, Toulouse, France*, pp. 144–149, EUROSIS-ETI, 2023. (70 MNiSW points)

Abstract: Soil bacteria play a crucial role in various biotechnological fields and the food industry. Promising results have been shown in the efficient identification of specific bacteria genera through automated microscopic image analysis using Machine Learning and Image Processing Techniques. In this study, the emphasis is placed on the classification of five distinct soil bacteria genera: *Enterobacter*, *Rhizobium*, *Pantoea*, *Bradyrhizobium* and *Pseudomonas*. The objective is to investigate the application of Residual Convolutional Neural Networks for automating the bacteria classification. The performance of ResNet50, ResNet50V2, ResNet101, ResNet101V2, ResNet152 and ResNet152V2 models is evaluated in the classification of microscopic sub-images including single bacteria instances. Additionally, the effectiveness of the majority voting rule in enhancing image classification accuracy based on sub-images is explored. The results obtained demonstrate superior performance of ResNetV2 over standard ResNet models, with ResNet101V2 and ResNet152V2 achieving the highest precision and recall values. Furthermore, the utilization of the majority voting rule leads to improved classification results compared to sub-image-based classification solitary. These findings underscore the effectiveness of ResNetV2 and highlight the importance of aggregating information from multiple sub-images for the accurate classification of bacteria.

PERFORMANCE ANALYSIS OF RESIDUAL NEURAL NETWORKS IN SOIL BACTERIA MICROSCOPIC IMAGE CLASSIFICATION

Aleksandra Konopka[1], Karol Struniawski[1] and Ryszard Kozera[1,2]

[1]Institute of Information Technology
Warsaw University of Life Sciences - SGGW
ul. Nowoursynowska 166, 02-776, Warsaw, Poland
email: {aleksandra_konopka, karol_struniawski, ryszard_kozera}@sggw.edu.pl
[2]The School of Physics, Mathematics and Computing
The University of Western Australia
35 Stirling Highway, Perth, Australia
email: ryszard.kozera@uwa.edu.au

KEYWORDS

Soil bacteria, Microscopic image classification, Machine Learning, Image Processing, Residual Convolutional Neural Networks

ABSTRACT

Soil bacteria play a crucial role in various biotechnological fields and the food industry. Promising results have been shown in the efficient identification of specific bacteria genera through automated microscopic image analysis using Machine Learning and Image Processing Techniques. In this study, the emphasis is placed on the classification of five distinct soil bacteria genera: *Enterobacter, Rhizobium, Pantoea, Bradyrhizobium* and *Pseudomonas*. The objective is to investigate the application of Residual Convolutional Neural Networks for automating the bacteria classification. The performance of ResNet50, ResNet50V2, ResNet101, ResNet101V2, ResNet152 and ResNet152V2 models is evaluated in the classification of microscopic sub-images including single bacteria instances. Additionally, the effectiveness of the majority voting rule in enhancing image classification accuracy based on sub-images is explored. The results obtained demonstrate superior performance of ResNetV2 over standard ResNet models, with ResNet101V2 and ResNet152V2 achieving the highest precision and recall values. Furthermore, the utilization of the majority voting rule leads to improved classification results compared to sub-image-based classification solitary. These findings underscore the effectiveness of ResNetV2 and highlight the importance of aggregating information from multiple sub-images for the accurate classification of bacteria.

INTRODUCTION

Soil bacteria hold significant value in the food industry and various biotechnological fields (Enez 2019). The automation of microscopic image analysis using Machine Learning (ML) and Image Processing Techniques has emerged as a promising approach for identifying specific bacteria genera, offering advantages in terms of time and resource efficiency. In this research, our focus lies on classifying five soil bacteria genera: *Enterobacter, Rhizobium, Pantoea, Bradyrhizobium* and *Pseudomonas*, known for their substantial impact on agricultural and horticultural crops (Konopka et al. 2022). Each microscopic image in our dataset contains only one specific bacteria genera. The samples utilized in this study were provided by The National Institute of Horticultural Research in Skierniewice. Sample images from the dataset can be accessed through the following URL: https://bit.ly/3rBxlb7. In this work, the concept of selecting individual instances of bacteria and subsequently classifying given images based on extracted sub-images was applied. However, in contrast to our previous study (Struniawski et al. 2022), Residual Convolutional Neural Networks (ResNet) (Kaiming et al. 2015) are employed rather than classical ML approaches. The objective of this article is to present the research findings on the application of Residual Networks, particularly ResNet models, for the automated classification of soil bacteria. The performance of ResNet50, ResNet50V2, ResNet101, ResNet101V2, ResNet152 and ResNet152V2 models in classifying microscopic sub-images of bacteria is investigated. Furthermore, the effectiveness of utilizing the majority voting rule to enhance the classification accuracy of entire images based on their constituent sub-images is explored.

DATASET PREPARATION

The dataset utilized in this study comprises a collection of 128 raw soil bacteria microscopic images, herein denoted as $Im = \{I_i\}_{i=1}^{128}$. Each image is a representation of a distinct bacteria genera, with the following distribution: class 0 with 34 images of *Bradyrhizobium*, class 1 with 25 images of *Rhizobium*, class 2 with 22 images of *Enterobacter*, class 3 with 21 images of *Pseudomonas* and class 4 with 26 images of *Pantoea*.

To facilitate further analysis, binary masks are derived for each image. This process involves converting the selected image to grayscale and applying the Otsu method (Vala and Baxi 2013) followed by morphological operations such as opening and closing (Dougherty and Lotufo 2003). The resulting binary masks assign white color to regions where bacteria are located, which corresponds to the Region of Interest. Conversely, black color is assigned to the background area.

In the analysis of the created masks, the extraction of single instances of bacteria was performed. For each of the 128 images, a total of $m = 50$ sub-images, each containing a single bacteria instance, were extracted. The resulting dataset consisted of 6400 sub-images, with dimensions of $3 \times 224 \times 224$, which were subsequently used as input data for the analyzed Convolutional Neural Networks (CNNs). It should be noted that all sub-images present a single bacteria in the RGB color space, positioned in the center, while the surrounding background remain black (see Fig. 1).

In the extraction of individual bacteria objects, which resulted in a set of sub-images, careful consideration was given to the area value and solidity of the objects. The area value represents the count of non-zero pixels within the object's mask. A convex image, which is a binary mask depicting the smallest convex polygon enclosing the object, was generated. The convex area corresponds to the number of pixels encompassed by the convex hull of the convex image. Furthermore, the solidity of an object is calculated by dividing the number of non-zero pixels in the convex image by the number of non-zero pixels in the analyzed object. These two characteristics were meticulously considered to ensure that the selected sub-images exclusively portray authentic single instances of bacteria, while preventing the inclusion of overlapped objects or contaminated regions. The primary objective of this approach was to uphold the integrity and accuracy of the dataset, thereby providing a reliable foundation for subsequent analysis and classification tasks.

To generate the sub-images, the identification and selection of objects representing individual bacteria instances are essential. These objects correspond to the white areas within the mask M associated with the analyzed image $I \in Im$. The extraction process of these objects is carried out applying the following algorithm:

1. Let $O = \{o_i\}_{i=1}^{k}$ represent the set of all k objects extracted from the mask M of the given image.
2. The objects in O with solidity values less than 0.5 are filtered out, resulting in a new set $\hat{O} = \{\hat{o}\}_{i=1}^{n}$.
3. The objects in $\hat{O}$ are sorted based on their area values in ascending order, creating a set $S = \{s\}_{i=1}^{n}$.
4. Next, $m = 50$ objects are selected from set S, assuming that $n/2 \geq m$. The median value of S is computed, which corresponds to the value at index $n/2$. Subsequently, the first m objects are selected from S, where their area values are greater than or equal to $s_{n/2}$. This selection process forms a new set denoted as $\hat{S} = \{\hat{s}\}_{i=1}^{m} = \{s_i\}_{i=n/2}^{(n/2)+m}$.

Once the objects in $\hat{S}$ are selected, their corresponding masks are applied to the original image I through Hadamard Product (Horn and Johnson 2012). Consequently, a set of sub-images Se is generated. To ensure consistency in the dimensions of all sub-images, each sub-image in Se is placed within the center of a $3\times224\times224$ black image. This positioning is achieved by utilizing a bounding box - the smallest rectangle that an object fits within. As a result, a final set $Sb_I = \{sb_i\}_{i=1}^{m}$ of sub-images is obtained for the analyzed image I. Each sub-image within this set represents a distinct single bacteria instance, surrounded by a black background (see Fig. 1).

Figure 1: Sub-image of the selected bacteria object *Rhizobium* (left) and *Enterobacter* (right).

To facilitate the training of a CNN, the set of 6400 images needed to be partitioned into three distinct groups: training (Tr), validation (Val) and testing sets (Te) in a $0.6 : 0.3 : 0.1$ ratio commonly used in ML applications. In this study, the classification task extended beyond analyzing individual sub-images. An additional scope was to classify the entire image $I \in Im$ based on the collective results of classifying its constituent sub-images.
To ensure proper distribution and representation of the different bacteria genera images in each of the three sets, a randomized shuffling approach was adopted. Correct ratios of images from the five genera were selected and the images were assigned to the three distinct groups

(Tr, Val and Te). This approach guaranteed that each set contained a comparable ratio of images from each of the five genera. Additionally, it ensured that no sub-images from a single image were spread across multiple groups. By adhering to these guidelines, the integrity and diversity of the dataset were preserved enabling robust training, validation and testing of the CNN models. To illustrate how the images are shuffled in the correct ratio, let's consider an example with a total of 100 images from five different bacteria genera: G_1 (20 images), G_2 (30 images), G_3 (15 images), G_4 (25 images) and G_5 (10 images). The images are to be partitioned into training, validation and testing sets according to the desired ratio of $0.6 : 0.3 : 0.1$. Next, we randomly shuffle and assign the images from each genera to the respective sets resulting in the following distribution. Training set: G_1 (12 images), G_2 (18 images), G_3 (9 images), G_4 (15 images) and G_5 (6 images). Validation set: G_1 (6 images), G_2 (9 images), G_3 (4 images), G_4 (7 images) and G_5 (3 images). Testing set: G_1 (2 images), G_2 (3 images), G_3 (2 images), G_4 (3 images) and G_5 (1 image). A total of $m = 50$ sub-images is extracted from each of the Im so the number of sub-images in the training, validation and testing set is equal to: 3000, 1450 and 550, respectively.

CONVOLUTIONAL NEURAL NETWORK

In the context of our research, which primarily revolves around the classification process, we delve into the application of Convolutional Neural Networks. Unlike traditional Machine Learning approaches, where features of the dataset samples are manually computed, CNNs have the ability to automatically extract relevant traits. In classical ML, scientists typically hand-craft features based on their judgment of the most suitable characteristics for the classification task. This approach is highly effective when there is a clear and easily definable distinction between different classes. However, in cases where the differences between objects belonging to analyzed classes are difficult to define, it becomes challenging to specify appropriate hand-crafted features.
Convolutional Neural Networks provide a powerful automated approach for computing features. The output of a CNN is a matrix of features, where each vector captures the intricate relationships within an analyzed sample. Describing each individual feature becomes challenging due its complex nature. The output of the CNN serves as input to a classifier, enabling the assignment of the analyzed sample to one of the considered classes.
The CNN architecture is comprised of two primary types of layers: convolution and pooling. The convolution layer applies selected filters (kernels) (Gonzalez and Woods 2008) to traverse across the image, capturing the element-wise product of the filter and the corresponding image patch. These kernels are applied for finding different patterns of images e.g. to highlight their horizontal or vertical edges. Multiple kernels can be employed within a single layer and by introducing more convolution layers, the network becomes capable of detecting increasingly complex geometric shapes (Scholz and Scholz 2022). On the other hand, the pooling layer serves the purpose of reducing the spatial dimensions of the input image while retaining the most essential information necessary for classification. This operation is vital for enhancing computational efficiency. The pooling layer operates by sliding a window across the image and selecting a specific value within the window to be saved (Kaiming et al. 2015).

RESIDUAL NETWORK

The parameters of neural networks are typically trained using the backpropagation algorithm (Hassoun 2003) which may yield two common challenges depending on the characteristics of the dataset: the vanishing and the exploding gradient problem (Hanin 2018). The vanishing gradient problem occurs when the gradient becomes extremely small, leading to minimal changes in the tuned parameters during training. Consequently, the network becomes stuck and is unable to effectively continue learning. On the other hand, the exploding gradient problem arises when the gradient becomes excessively large, causing difficulties in reaching the optimal parameter values and converging to a solution. This issue compounds as the gradient is propagated deeper into the network.

Residual Network is a type of Convolutional Neural Network that tackles the challenges of vanishing and exploding gradient through the implementation of skip connections. These skip connections allow information to bypass one or more layers, enabling the network to effectively propagate gradients and address the gradient-related issues. As a result, ResNet architectures can be constructed with remarkable depth, even consisting of up to 152 convolutional blocks. Different ResNet variants, such as ResNet18, ResNet34, ResNet50, ResNet101 and ResNet152, are available, varying in the number of convolutional blocks and the sizes of applied filters (Kaiming et al. 2015). A standard ResNet block comprises several distinct layers arranged in a specific order, including a convolution layer, batch normalization layer, ReLU activation function, concatenation and pooling. The number of convolution blocks varies between different ResNet variants. ResNet18 and ResNet34 consist of two convolution blocks, while ResNet50, ResNet101 and ResNet152 contain three.

ResNetV2 is an alternative version of ResNet that introduces a significant modification to the original architecture. One of the key distinctions lies in the altered layer ordering within the ResNet block (Kaiming et al. 2016). In ResNetV2, the batch normalization and ReLU activation function are applied before the convolution.

TRANSFER LEARNING

In Deep Neural Networks and especially in CNN there are many variables describing net's architecture that need to be specified such as the size of filters, number of filters, size of a padding and stride. Moreover, there can be millions of parameters that have to be tuned including: weights, biases and values of filters. Due to that fact training our model from scratch requires a lot of time and computational complexity which can rarely be handled by widely accessible personal computers. Because of that reason scientists usually decide to make use of pretrained models which are trained for solving a similar task and adjust it to their problem which is called a Transfer Learning (TL) (Zhang et al. 2020).
ResNet is a neural network that has been pre-trained using the ImageNet dataset. ImageNet is a benchmark dataset for image classification, consisting of over 14 million images belonging to more than 20,000 categories (Deng et al. 2009). The pre-trained model can be used as a starting point for various image classification tasks. In Transfer Learning, different approaches can be applied. One option is to freeze all the pre-trained layers, meaning their parameters remain fixed and add new trainable layers on top of them. These new layers are then trained on a specific dataset to adapt the model to the desired classification task. Another approach, known as fine-tuning, involves unfreezing selected layers or the entire model and re-training their parameters using our dataset. This allows the model to learn task-specific features leveraging the pre-trained knowledge.

EXPERIMENTS

For the classification task in this research, various Residual Networks were employed, namely ResNet50, ResNet50V2, ResNet101, ResNet101V2, ResNet152 and ResNet152V2. To determine the optimal network's architecture, different values of parameters were explored and the following configuration yielded the best results. To improve computational efficiency, the sub-image dataset was loaded in batches of size 32. The maximum number of epochs was set to 500, indicating that the model underwent training up to 500 iterations. To mitigate overfitting, early stopping was implemented by monitoring the loss function with a patience of 10. If the loss value did not decrease within 10 consecutive epochs, the training process was halted. Average pooling was chosen as the pooling technique. The Adam optimizer with a learning rate of 0.001 was employed for training the network (Kingma and Ba 2014). All ResNet models were pre-trained on the ImageNet dataset. The last ResNet block, specifically *conv5_block1_1_conv* in ResNet or *conv5_block1_preact_bn* in ResNetV2, was unfrozen and its parameters were retrained. The output of the ResNet served as input to the subsequent neural network responsible for classification.

The classification neural network consisted of the following layers:

1. Dense layer with 512 neurons and ReLU activation function.
2. 30% dropout.
3. Dense layer with 512 neurons and ReLU activation function.
4. 30% dropout.
5. Dense layer with 5 neurons, corresponding to the 5 classes of sub-images. The last layer utilized a SoftMax activation function, which produced a probability vector indicating the likelihood of an analyzed image belonging to each of the considered classes.

RESULTS AND CONCLUSIONS

The results of the classification task are presented in two tables: Table 1 illustrates the classification based on sub-images, while Table 2 represents the classification of images using the majority voting rule. Both tables provide precision and recall values for different ResNet models, including ResNet50, ResNet50V2, ResNet101, ResNet101V2, ResNet152 and ResNet152V2. Each model's performance is evaluated for different classes $(0-4)$, and the micro and macro averages are also calculated. The precision is the ratio of true positives to the sum of true positives and false positives, while the recall is the ratio of true positives to the sum of true positives and false negatives. The micro average assumes each samples are equal, while the macro average treats equally each class. Upon analyzing the results, it can be observed that ResNetV2 consistently outperforms ResNet in terms of precision and recall for most classes. Among the ResNetV2 models, ResNet101V2 and ResNet152V2 achieve the highest precision and recall. Furthermore, applying the majority voting rule improves the classification results compared to the classification based solely on sub-images. The precision and recall values increase significantly for all classes, indicating better classification accuracy when considering the collective decision of sub-images.

In conclusion, the experimental results demonstrate the superior performance of ResNetV2 over ResNet in classifying soil bacteria sub-images. The application of the majority voting rule further enhances the classification accuracy, leading to improved precision and recall for the overall image classification task. These findings highlight the effectiveness of ResNetV2 and the importance of aggregating information from multiple sub-images in achieving better classification results.

Table 1: Classification results (precision and recall) for the set of sub-images.

	ResNet50		ResNet50V2		ResNet101		ResNet101V2		ResNet152		ResNet152V2	
class	prec	rec	prec	rec	prec	rec	prec	rec	prec	rec	prec	rec
0	0.53	0.60	0.64	0.68	0.41	0.68	0.74	0.67	0.71	0.24	0.74	0.69
1	0.76	0.13	0.86	0.80	0.90	0.08	0.92	0.83	0.67	0.13	0.84	0.95
2	0.80	0.99	0.87	0.94	0.97	0.50	0.92	0.94	0.99	0.69	0.93	0.87
3	0.56	0.64	0.64	0.71	0.39	0.97	0.63	0.81	0.07	0.12	0.61	0.89
4	0.26	0.33	0.69	0.59	0.30	0.07	0.63	0.62	0.30	0.66	0.74	0.51
micro avg	0.52	0.52	0.73	0.73	0.45	0.45	0.75	0.75	0.36	0.36	0.76	0.76
macro avg	0.58	0.54	0.74	0.74	0.60	0.46	0.77	0.77	0.55	0.37	0.77	0.78

Table 2: Classification results (precision and recall) for images based on their sub-images applying majority voting.

	ResNet50		ResNet50V2		ResNet101		ResNet101V2		ResNet152		ResNet152V2	
class	prec	rec	prec	rec	prec	rec	prec	rec	prec	rec	prec	rec
0	0.62	0.80	1.00	0.80	0.44	0.80	1.00	0.90	1.00	0.30	1.00	0.90
1	1.00	0.14	0.88	1.00	0.00	0.00	1.00	1.00	1.00	0.14	0.88	1.00
2	1.00	1.00	0.86	1.00	1.00	0.67	1.00	1.00	1.00	0.83	1.00	1.00
3	0.57	0.67	0.75	1.00	0.43	1.00	0.75	1.00	0.00	0.00	0.75	1.00
4	0.22	0.29	1.00	0.71	0.00	0.00	0.83	0.71	0.32	0.86	1.00	0.71
micro avg	0.58	0.58	0.89	0.89	0.50	0.50	0.92	0.92	0.42	0.42	0.92	0.92
macro avg	0.68	0.58	0.90	0.90	0.37	0.49	0.92	0.92	0.66	0.43	0.93	0.92

REFERENCES

Deng J.; Dong W.; Socher R.; Li L.J.; Li K.; and Fei-Fei L., 2009. *ImageNet: A large-scale hierarchical image database*. In *2009 IEEE CVPR*. 248–255. doi:10.1109/CVPR.2009.5206848.

Dougherty E.R. and Lotufo R.A., 2003. *Hands-on Morphological Image Processing*. ISBN 9780819447203. doi:10.1117/3.501104.

Enez B., 2019. *Identification of bacteria species isolated from soil and investigation of optimum conditions: application in food industry and biotechnological fields*. *Prog Nutr*, 21, no. 2, 467–472. doi:10.23751/pn.v21i2.7985.

Gonzalez R.C. and Woods R.E., 2008. *Digital Image Processing*. Prentice Hall. ISBN 978-0133356724.

Hanin B., 2018. *Which Neural Net Architectures Give Rise To Exploding and Vanishing Gradients?* In *Neural Information Processing Systems*. doi:10.48550/arXiv.1801.03744.

Hassoun M.H., 2003. *Fundamentals of Artificial Neural Networks*. Bradford Books. ISBN 9780262514675.

Horn R.A. and Johnson C.R., 2012. *Matrix analysis*. Cambridge University Press. ISBN 9780511810817. doi:10.1017/CBO9780511810817.

Kaiming H.; Xiangyu Z.; Shaoqing R.; and Sun J., 2015. *Deep Residual Learning for Image Recognition*. *Microsoft Research*. doi:10.48550/arXiv.1512.03385.

Kaiming H.; Xiangyu Z.; Shaoqing R.; and Sun J., 2016. *Identity Mappings in Deep Residual Networks*. *Microsoft Research*. doi:10.48550/arXiv.1603.05027.

Kingma D.P. and Ba J., 2014. *Adam: A Method for Stochastic Optimization*. *CoRR*. doi:10.48550/arXiv.1412.6980.

Konopka A. et al., 2022. *Classication of Soil Bacteria Based on Machine Learning and Image Processing*. In *ICCS 2022*. Springer International Publishing, 263–277. doi:10.1007/978-3-031-08757-8_23.

Scholz C. and Scholz S., 2022. *Exploring complex pattern formation with convolutional neural networks*. *Am J Phys*, 90, no. 2, 141–151. doi:10.1119/5.0065458.

Struniawski K.; Konopka A.; and Kozera R., 2022. *Identification of Soil Bacteria with Machine Learning and Image Processing Techniques applying Single Cells' Region Isolation*. In *ESM'2022*. ISBN 9789492859242, 76–81.

Vala H.J. and Baxi A., 2013. *A Review on Otsu Image Segmentation Algorithm*. *Adv Mat Res*, 1959–1961. doi:10.4028/www.scientific.net/AMR.989-994.1959.

Zhang D.; Liu Z.; and Shi X., 2020. *Transfer Learning on EfficientNet for Remote Sensing image Classification*. *ICMCCE*, 2255–2258. doi:10.1109/ICMCCE51767.2020.00489.

3.4 Classification performance of Extreme Learning Machine Radial Basis Function with k-means, k-medoids and mean shift clustering algorithms

Publication: A. Konopka, K. Struniawski, and R. Kozera, "Classification performance of Extreme Learning Machine Radial Basis Function with k-means, k-medoids and mean shift clustering algorithms," in *Computational Science – ICCS 2023: 23rd International Conference, Prague, Czech Republic, July 3–5, 2023, Proceedings, Part IV, Lecture Notes in Artificial Intelligence*, vol. 10476, pp. 171–186, Springer, 2023. https://doi.org/10.1007/978-3-031-36027-5_13 (140 MNiSW points)

Abstract: Extreme Learning Machine (ELM) is a feed-forward neural network with one hidden layer. In its modification called ELM Radial Basis Function the input data is a priori clustered into a number of sets represented by their centroids. The matrix of distances between each sample and centroid is calculated and applied as input data to the neural network. This work conducts a comparison study of the ELM Radial Basis Function classification performance upon applying either k-means, k-medoids or mean shift clustering methods. Generated results are obtained from two datasets i.e. Wine Quality-White and Ionosphere. The computations are based on full datasets or on the same both sets reduced by a feature selection algorithm. The parameters of the classifiers such as number of neurons in hidden layer, the value of k in k-means and k-medoids, the value of radius in mean shift are optimized through an iterative procedure upon maximizing an accuracy or minimizing Mean Square Error and computation time. The different distance metrics for k-means and k-medoids, and mean shift with Gaussian or flat kernel function are also compared. The results obtained with Softplus and linear activation function (applied in most of the computations in this work) are juxtaposed with the results generated by other activation functions.

Classification Performance of Extreme Learning Machine Radial Basis Function with K-means, K-medoids and Mean Shift Clustering Algorithms

Aleksandra Konopka[1], Karol Struniawski[1(✉)], and Ryszard Kozera[1,2]

[1] Institute of Information Technology, Warsaw University of Life Sciences - SGGW, ul. Nowoursynowska 159, 02-776 Warsaw, Poland
{aleksandra_konopka,karol_struniawski,ryszard_kozera}@sggw.edu.pl
[2] School of Physics, Mathematics and Computing, The University of Western Australia, 35 Stirling Highway, Crawley, Perth, WA 6009, Australia
ryszard.kozera@uwa.edu.au

Abstract. Extreme Learning Machine (ELM) is a feed-forward neural network with one hidden layer. In its modification called ELM Radial Basis Function the input data is a priori clustered into a number of sets represented by their centroids. The matrix of distances between each sample and centroid is calculated and applied as input data to the neural network. This work conducts a comparison study of the ELM Radial Basis Function classification performance upon applying either k-means, k-medoids or mean shift clustering methods. Generated results are obtained from two datasets i.e. Wine Quality-White and Ionosphere. The computations are based on full datasets or on the same both sets reduced by a feature selection algorithm. The parameters of the classifiers such as number of neurons in hidden layer, the value of k in k-means and k-medoids, the value of radius in mean shift are optimized through an iterative procedure upon maximizing an accuracy or minimizing Mean Square Error and computation time. The different distance metrics for k-means and k-medoids, and mean shift with Gaussian or flat kernel function are also compared. The results obtained with Softplus and linear activation function (applied in most of the computations in this work) are juxtaposed with the results generated by other activation functions.

Keywords: Neural Networks · Machine Learning · Extreme Learning Machine · Radial Basis Function · Clustering Algorithms

1 Introduction

The Backpropagation Algorithm (BA) introduced in 1986 by Rumelhart et al. [21] represents an important component of machine learning. The main problem of BA stems from the fact that it usually yields local minima of associated network's residual error function. In addition, BA computational cost and training

J. Mikyška et al. (Eds.): ICCS 2023, LNCS 14076, pp. 171–186, 2023.
https://doi.org/10.1007/978-3-031-36027-5_13

time, especially for the large datasets, may preclude its practical application. The new concept of neural network called Extreme Learning Machine (ELM) was introduced by Huang et al. in 2004 [7]. ELM converges much faster than traditional learning schemes as relieved from the time-consuming iterations and is also more likely to reach a global optimum [10]. ELM is successfully adapted to various machine learning applications such as classification and regression in medicine, chemistry, transportation, economy, agriculture, robotics etc. It also outperforms other methods in training time and approximation ability [9,18]. ELM is characterized by yielding extremely fast training time in comparison to other machine learning methods like e.g. Multilayer Perceptron trained with BA [11]. Currently, ELM still evolves to further improve generalization capacity in case of special applications. One of its variants called Extreme Learning Machine Radial Basis Function (ELM-RBF), weaving core principles of ELM with feature space mapping using RBF kernels, yields comparable results to BA with considerably faster computation time [2]. In the field of ELM-RBF, the most commonly used clustering method is k-means, although application of k-medoids is also found in the literature.

In this paper, the performance of ELM-RBF combined with k-means, k-medoids or mean shift clustering methods is thoroughly investigated. The comparison involves manipulating with multiple variables, such as parameter values of clustering methods, number of neurons or different activation functions in ELM-RBF. In order to obtain the most significant results a comparison is conducted on two datasets - Wine Quality-White and Ionosphere [3]. The characteristics of the selected benchmark sets allow to obtain significant results and compare different algorithms, such as those used for solving k-medoids problem, as their application relies on the size of the input dataset.

2 Extreme Learning Machine

Extreme Learning Machine consists of input, hidden and output layer aimed to solve classification and regression tasks by supervised learning [7]. Assume input data is described as pairs of values $\eta = \{(x_i, t_i)\}_{i=1}^{N}$, where $x_i = \{x_{ij}\}_{j=1}^{d}$ forms matrix $X_{d\times N}$ of d features. Here t_i is recognized as affiliation to the given class establishing T. For the classification need $t_i \in [0, M] \subseteq \mathbb{N}$, whereas for regression the target value $t_i \in \mathbb{R}^M$. Here M represents either the number of classes or dimension of target values. The number of neurons in input, output and hidden layer is assumed here to be equal to d, M and L, respectively. Here L is given a priori as there is no universal method for L optimization. Neurons in ELM are McCulloch-Pitts neurons [14] with identity function on input and output layers and any activation function $f : \mathbb{R} \to \mathbb{R}$ on hidden layer units. Matrix of weights $W_{d\times L}$ with coefficients $w_{ij} \in (-1, 1)$ connecting input neurons with L hidden layer neurons and bias values b_i of $b = \{b_i\}_{i=1}^{N}$ are randomly selected with the aid of uniform distribution function, where $i = 1, \ldots, d$ and $j = 1, \ldots, L$. Thus, $H = f(X^T W + b)$ is obtained as an output of the hidden layer. Wages β between hidden and output layer are computed once using algebraic

transformations. To calculate β the following matrix equation $H\beta = T$ should be solved. The matrix H is non-invertible and therefore solving $H\beta = T$ can be reformulated into the optimization task estimating $\hat{\beta}$ based on minimization of a Mean Square Error (MSE) between H and T [7]. The corresponding solution reads as $\hat{\beta} = H^{\dagger}T$, where $H^{\dagger}$ is the Moore-Penrose pseudo-inverse operation [20]. Various methods for $H^{\dagger}$ evaluation can be applied including e.g. Cholesky factorization of a singular matrix [12].

3 Extreme Learning Machine Radial Basis Function

Extreme Learning Machine Radial Basis Function is a method with training similar to its archetype which is based on random generation of W, b and calculation of β using generalized inverse of matrix H. The extra component involves input data transformation with the aid of Radial Basis Function [15]. Specifically, let $X_{d\times N}$ be a set of d features for N observations. First, a vector quantization technique is applied as a clustering algorithm. The aim of the latter is to partition N samples into a certain number (given a priori or designated automatically during algorithm's run) of k clusters using e.g. k-means clustering algorithm. Each sample x_i is assigned to exactly one cluster $\{c_j\}_{j=1}^{k}$, which in turn is recognized as the closest centroid to x_i in terms of a considered metric. Next, x_i is transformed to the new feature space based on a chosen kernel function. Note here that when $k > d$ then X is mapped to a higher dimension. The definition of the RBF kernel for ELM-RBF is given as $K(x_i, c_j) = exp\{\frac{-\|x_i - c_j\|^2}{2\sigma^2}\}$. The matrix $K_{N\times k}$ is computed as a measure of distance between each x_i and c_j. The σ_j value is determined as a $\sigma_j = \frac{max\{d_j\}}{\sqrt{2k}}$. Finally, K is treated as an input matrix to the typical ELM network.

4 Clustering Methods

Clustering methods divide samples into disjoint groups. In k-means, k-medoids and mean shift each cluster is represented by a centroid which is calculated through an iterative procedure.

4.1 Mean Shift

Mean shift is an unsupervised learning algorithm [5] commonly applied in clustering, tracking and smoothing. This algorithm locates maxima of a density function with the aid of an iterative procedure upon updating candidates for centroids. Mean shift requires to specify the bandwidth of a window, which is shifted until the algorithm converges. The number of clusters is a priori unknown and depends on the density of input data samples. The points are assigned to the corresponding local maxima computed by the algorithm.

The input data of the algorithm is a set of n data points $\mathcal{Q} = \{q_i\}_{i=1}^{n}$, where $q_i = (q_{i_1}, q_{i_2}, \ldots, q_{i_d}) \in \mathbb{R}^d$. The value of bandwidth b is specified

arbitrarily. Let $C = \{\{c_{ij}\}_{i=1}^{n}\}_{j=1}^{s_i}$ be the set of all locations of the window shifted in the algorithm, where $c_{ij} = (c_{ij_1}, c_{ij_2}, \ldots, c_{ij_d})$, $i = 1, \ldots, n$ is a number of point that is currently considered and $j = 1, \ldots, s_i$ is a number of the shift for i'th point. Then first point $q_1 \in \mathcal{Q}$ is selected and it is set as the first location of window's center ($c_{11} = q_1$). Then Euclidean distances $d(c_{ij}, q_i) = \sqrt{(c_{ij_1} - q_{ij_1})^2 + \cdots + (c_{ij_d} - q_{ij_d})^2}$ between current c_{ij} and each q_i are computed. All q_i with $d(c_{ij}, q_i) \leq b$ (i.e. inside the window centered at c_{ij}) are selected to the set $\hat{\mathcal{Q}}_{ij} = \{\hat{q}_m\}_{m=1}^{w_{ij}}$. Subsequently, the new location of the window is calculated as a mean value of all $\hat{\mathcal{Q}}_{ij}$ for each of their d dimensions $c_{i(j+1)} = (\frac{1}{w_{ij}} \sum_{m=1}^{w_{ij}} \hat{q}_{m_1}, \ldots, \frac{1}{w_{ij}} \sum_{m=1}^{w_{ij}} \hat{q}_{m_d})$. Such procedure is repeated for all $\mathcal{Q}$ until they all converge to their corresponding local maxima. An output of this method is the set of points assigned to each of the generated disjoint clusters.

The above procedure iterates over each of the points from the dataset resulting in high computation time for large input data. Despite the fact that shifting a selected point does not influence other iterations and the process can be parallelized it is still very slow. The elapsing computation time is highly correlated with the size of input matrix of data making this approach inefficient. The mean shift algorithm can be optimized to render the results with less computation. The Bart Finkston's implementation (available on Mathworks [4]) presents an approach which highly reduces the computational complexity of the algorithm. The idea is to mark the points which were inside a shifting window (in any of s_i iterations) as visited to disregard them in upcoming iterations when new starting points c_{i1} are to be selected. We note how many times each of the points is positioned inside a bandwidth considering all s_i iterations for i'th starting point until q_i converges to c_{is_i}. The procedure is repeated selecting random points from $\mathcal{Q}$ not so-far visited. Subsequently, the matrix with votes for all generated centroids is used to attach $\mathcal{Q}$ to the appropriate clusters applying majority voting.

4.2 K-means

K-means is a common clustering method for which each of the samples is assigned to one of k disjoint clusters [19]. The number of k is selected arbitrarily. Each of the clusters is represented by a centroid which is equal to mean value of all observations within a group. The samples in a given iteration are assigned to the closest centroid in accordance with a distance metric i.e Euclidean distance.

Let a set of n data points $\mathcal{Q} = \{q_i\}_{i=1}^{n}$, where $q_i = (q_{i_1}, q_{i_2}, \ldots, q_{i_d}) \in \mathbb{R}^d$ be given. Assume $C = \{\{c_{lj}\}_{l=1}^{k}\}_{j=1}^{m}$ is the set of all locations of k centroids-to-be in the algorithm, where $c_{lj} = (c_{lj_1}, c_{lj_2}, \ldots, c_{lj_d})$, $l = 1, \ldots, k$ and $j = 1, \ldots, m$ is a number of the algorithm's iteration. Let $C_m = \{c_{lm}\}_{l=1}^{k}$ be the set of centroids yielded by the algorithm after reaching termination condition (after m iterations). The value of m is set a posteriori once the stopping conditions are met. We start by selecting starting locations of all the k centroids $C_1 = \{c_{l1}\}_{l=1}^{k}$, they can be set randomly or upon applying an optimization algorithm (e.g. k-means++ [1]). Then, the samples are assigned to the closest centroid by means of selected distance metric. In case of Euclidean distance ρ_E, the respective values

are equal to $d_{i,l,j} = \rho_E(q_i, c_{lj})$. Subsequently, new locations of the centroids $C_{j+1} = \{c_{l(j+1)}\}_{l=1}^{k}$ are computed, where the new coordinates of each $c_{l(j+1)}$ are mean values of all the respective coordinates for data points q_i assigned to a given l'th centroid in j'th iteration. Here $c_{l(j+1)} = (\bar{c}_{lj_1}, \bar{c}_{lj_2}, \ldots, \bar{c}_{lj_d})$ with $\bar{c}_{lj_{dm}} = (1/a_{lj}) \sum_{i=1}^{a_{lj}} \hat{q}_{lj_i}$, where $dm = 1, \ldots, d$, a_{lj} is a number of points $\hat{q}_{lj_i}$ linked to l'th centroid in j'th iteration. The algorithm iterates until one of the stopping conditions is met i.e. either when computed centroids no longer switch their location or the distance between c_{lj} and $c_{l(j+1)}$ is smaller than prescribed ε or lastly, if the preselected maximal number of iterations is exceeded.

4.3 K-medoids

K-medoids is a method applied for clustering [8]. The data is assigned to one of k medoids, where each medoid is a specific point from the dataset and represents its cluster. The value of k and the dissimilarity measure are arbitrally selected. In the first step of the algorithm, k points are chosen as starting medoids. The latter is achieved either in a random manner or upon applying a specific method (e.g. k-means++ [1]) to solve an optimization problem leading to a reduction of computational time of the algorithm. The k-medoids problem can be solved with numerous algorithms such as: Partitioning Around Medoids (PAM), Voronoi-iteration k-medoids, Reynolds' improvements, FastPAM, FasterPAM algorithms, CLARA, CLARANS, FastCLARA and FastCLARANS, Lloyd's iterations [8,17].

In Partitioning Around Medoids algorithm, once a starting set of medoids is selected the values of a cost function are evaluated for all possible swaps of a medoid in a given cluster with another point belonging to the same cluster. When all the combinations are computed we finally apply only this one which has the minimal value of evaluated function. Subsequently, the affiliation of all points to points being set as current medoids is recalculated. This procedure is repeated for all the medoids as long as the value of the cost function decreases, otherwise the algorithm terminates. PAM algorithm has a high computation complexity as it calculates all possible swaps of all the medoids. In such setting, this algorithm is in practice predominantly applicable to a small amount of input data. For more complex computations one can apply variants of PAM (such as FastPAM or FasterPAM).

5 Experiments and Results

The computations were conducted on two sample datasets, namely Wine Quality-White and Ionosphere [3]. Wine Quality-White comprises of 4898 wine samples described by eleven features. Each of the wine samples was categorized by experts to the quality measure (ranging within $C = \{3, 4, \ldots, 9\}$). In order to perform computations on this set of features, a Mean Square Error is calculated to verify how a prediction deviates from an actual class as wine quality is represented by scale measure from C. We remark here that a misclassification of a wine from the class 3 to the predicted class 4 is a minor concern as

opposite to assigning it to class 9. The Ionosphere is a dataset describing signals that pass through the ionosphere. These signals are classified into two disjoint groups: "good" signals having evidence of some type of structure and "bad" signals deprived of such a feature. This dataset consists of 351 samples described by 34 features. In this binary classification task, the rendered results are compared calculating accuracy (ACC) representing the percentage of correctly classified samples in the whole classification process. These datasets are used here for classification with the aid of ELM-RBF applying a selected clustering method: k-means, k-medoids or mean shift. All computations described in this work are performed in Matlab.

In preliminary computations, which were carried on Wine Quality-White dataset, some of the parameters were selected and they are fixed in this work. The estimation of generalized inverse for ELM-RBF is based on Cholesky factorization of a singular matrix [12] yielding fast computation time and low MSE. To compare k-means, k-medoids and mean shift, different activation functions are chosen to obtain the best clustering results. In doing so, a linear activation function $f_L(x) = ax$ is used as it yields prominent classification results while applying k-means and k-medoids. On the other hand, Softplus activation function $f_{SP}(x) = log(1+\exp(x))$ is also applied as it renders the best categorization results for mean shift.

Our computations exploit Matlab implementations of k-means, k-medoids and mean shift. Most of the conducted experiments admit in these methods default parameters unless specified otherwise. More specifically, the k-means and k-medoids clustering algorithms in Matlab have default distance metric set to Squared Euclidean distance. In addition, a default method for choosing initial cluster centroid positions is k-means++ algorithm [1]. In case of k-medoids, an algorithm to find medoids is available in three variants which are applied by default depending on number of rows (samples) in the input data. More specifically, for the number of rows less than 3000 PAM algorithm is applied. For the number of rows between 3000 and 10000 a variant of the Lloyd's iterations is selected. In case of larger datasets, a default algorithm is CLARA [8]. Thus, the algorithm applied for Wine Quality-White and Ionosphere datasets are Lloyd's iterations and PAM, respectively. In this work, the applied mean shift algorithm [6] allows the implementation of Gaussian or flat kernel for distance calculations. The classification results were generated for both datasets applying 10 times a 20% cross-validation. Note that the computation results obtained for same values of parameters can still vary as all three clustering methods rely on randomness. Such difference is still noticeable even upon applying multiple cross-validations. The computations are conducted on three computers: K1 - Ryzen 5600G CPU, 16 GB DDR4 3600 MHz RAM, K2 - Ryzen 3900X CPU, 64 GB DDR4 3600 MHz and K3 - Dell 7750 Xeon W10885M CPU, 128 GB DDR4 2933 MHz MHz.

A classification task is performed on Wine Quality-White dataset for ELM-RBF with k-means and k-medoids for k ranging from 10 to 100 (with step-size 10) (see Fig. 1). These computations are applied on computer K1. The number of hidden-layer neurons n varies from 100 to 1000 (with step-size 100). Lastly,

a linear or Softplus activation function is applied. For both k-means and k-medoids MSE is the highest for $k \in \{10, 20\}$. The lower value of k gets, the worse classification result is rendered. Indeed, in the extreme case MSE attains the value 4.780 for k-medoids with $k = 10$ and $n = 500$ combined with Softplus activation function. For $k > 20$ the value of MSE for k-means and k-medoids stabilizes within the interval [0.619,0.793]. The best MSE result equal to 0.619 is achieved by k-medoids with linear activation function, $k = 100$ and $n = 700$ rendering the computation time equal to 409 s. Slightly worse result MSE = 0.628 is achieved for k-medoids with $k = 40$, $n = 100$ and linear activation function. Nevertheless, selecting the last set of parameters reduces computation time to 78 s. The best result obtained applying k-means yields MSE = 0.620 for $k = 100$, $n = 700$ and linear activation function, for which the computation time is equal to 402 s.

The mean shift clustering method is combined with ELM-RBF and the results obtained for Wine Quality-White dataset are based on the following choice of parameters: the values of the boundary width (radius) r are attuned from 0.3 to 1.3 (with step-size 0.1 - as it generated amounts of clusters from around 2 to 614) with the number of neurons n varying from 100 to 1000 (with step-size 100) combined with a linear or Softplus activation functions and Gaussian or flat kernel (see Fig. 2). The best result achieved for ELM-RBF with mean shift for the tested parameters is MSE = 0.686 for Softplus activation function, $n = 700$, $r = 0.6$ (which renders an average of 22 clusters in 10 cross-validations) with flat kernel function applied. This result is worse then the best result for k-medoids (and k-means) for this dataset for same tested number of neurons. The computation time is equal to 323 s. The best MSE (the lowest) are achieved for $r \in \{0.6, 0.7\}$ rendering between 10 and 22 clusters. ELM-RBF with mean shift obtains the worst MSE results for $r \leq 0.5$ (when $r = 0.5$ around 52 clusteres are rendered) applying linear kernel function and reaches even 2.02.

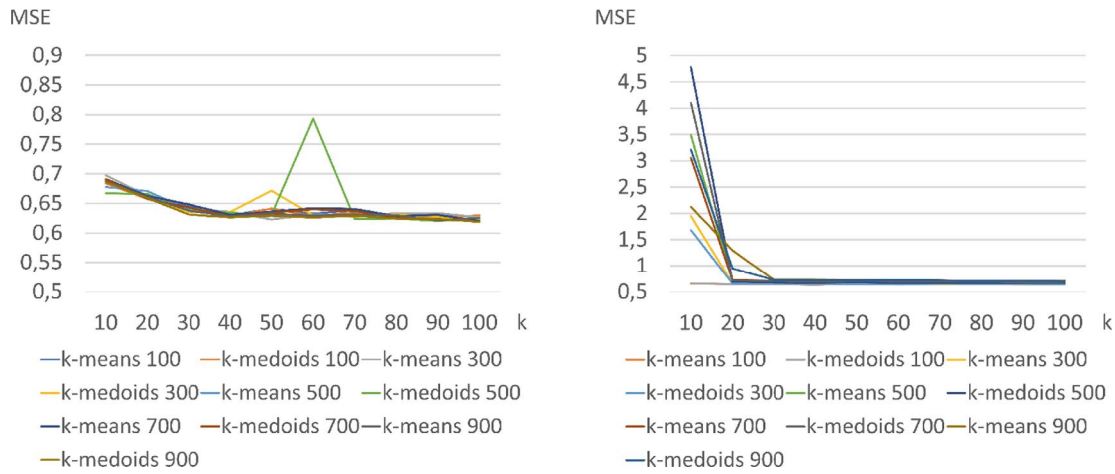


Fig. 1. MSE calculated on classification result for Wine Quality-White dataset for ELM-RBF with k-means or k-medoids , k varying between 10 and 100, 100–900 neurons applying linear (left) or Softplus activation function (right).

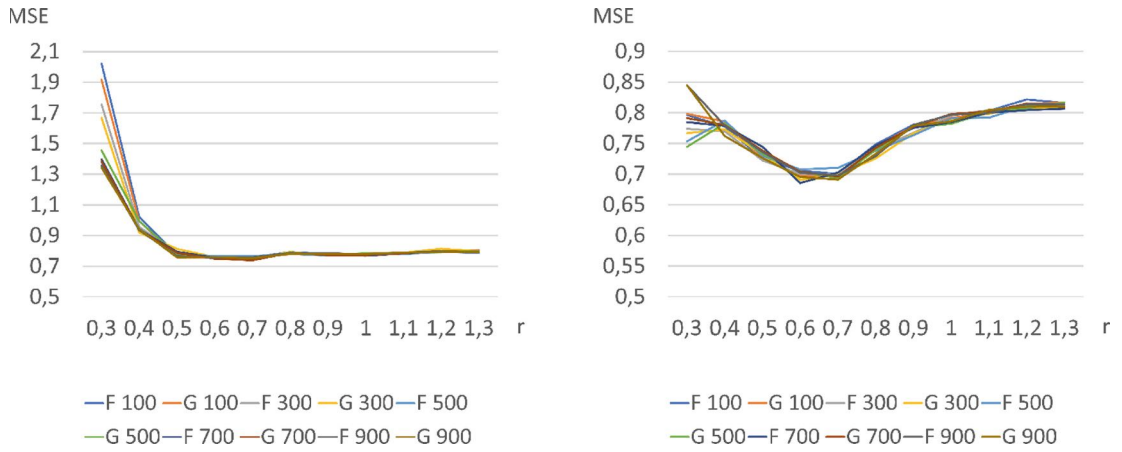


Fig. 2. MSE calculated on classification result for Wine Quality-White dataset for ELM-RBF with mean shift for r varying between 0.3 and 1.3, 100–900 neurons applying linear (left) or Softplus activation function (right), and Gaussian (G) or flat kernel (F).

The results generated by ELM-RBF on Wine Quality-White dataset are also analyzed on higher number of neurons n ranging from 1500 to 5000 (with step-size 500), for k varying from 10 to 100 (with step-size 10) in k-means and k-medoids applying linear or Softplus activation function (see Fig. 3). These computations are performed on the computer K2. Top three results are attained for k-medoids with linear activation function for $n = 2000$, $k = 100$, MSE $= 0.618$, for $n = 2000$, $k = 70$, MSE $= 0.618$ and for $n = 2500$, $k = 70$, MSE $= 0.619$. These results are equal to the best result for n ranging between 100 and 1000, but their computation time is much longer amounting to 1241, 1239 and 1676 s, respectively. Even if these computations were performed on faster computer, a huge increase in time is noticeable due to the enlarged number of neurons. The difference in results generated by Softplus and linear activation function is significant. Indeed, the values of MSE for linear function are in the interval $[0.618, 0.788]$, whereas for Softplus in $[0.732, 3.338]$. ELM-RBF with mean shift is also tested on higher values of n admitted to vary between 1500 and 5000 (with step-size 500) (see Fig. 4). The selected values of r range from 0.6 to 1 (with step-size 0.1) yielding the number of clusters between 4 and 22. The best MSE is equal to 0.688 for $n = 2000$, Gaussian kernel, Softplus activation function and $r = 0.7$ (11 clusters) computed in 1303 s.

The classification is also conducted on Ionosphere dataset applying ELM-RBF combined with k-means and k-medoids as clustering methods. The calculations were performed on computer K1. The computations were executed for k ranging from 10 to 100 for a number of neurons n varying from 100 up to 1000 (with step-size 100) (see Fig. 5). The results generated with linear activation function outperform those rendered with Softplus. The highest accuracy equal to 0.946 is reached for k-medoids with linear function, $n = 100$, $k = 96$ taking 13 s of execution time. The best result for k-means ACC $= 0.941$ is obtained for linear function, $n = 200$, $k = 80$ in 9 s. The accuracy for linear activation function rapidly increases with k running over $k = \{1, 2, \ldots, 10\}$. Once $k > 10$

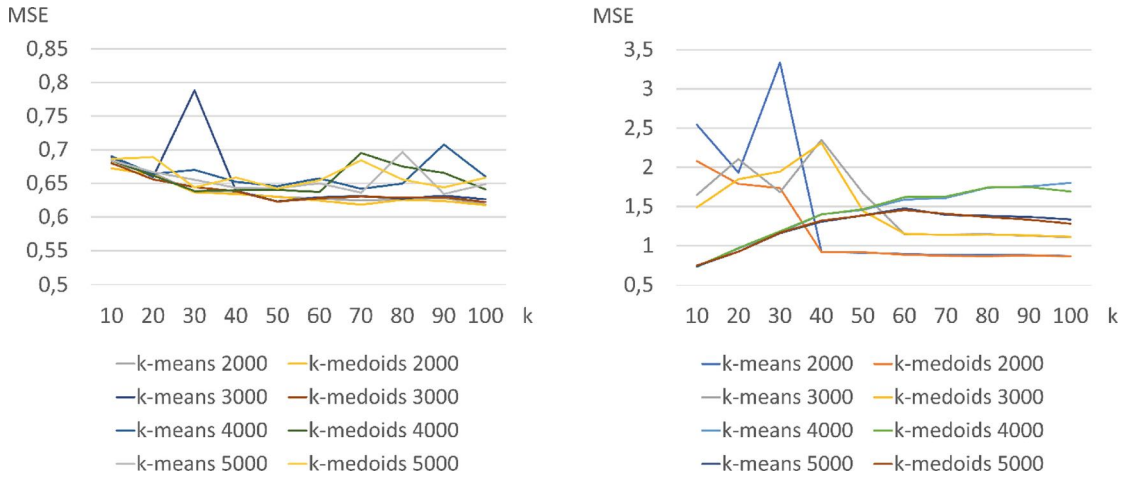


Fig. 3. MSE calculated on classification result for Wine Quality-White dataset for ELM-RBF with k-means or k-medoids for k varying between 10 and 100, 2000–5000 neurons applying linear (left) or Softplus activation function (right).

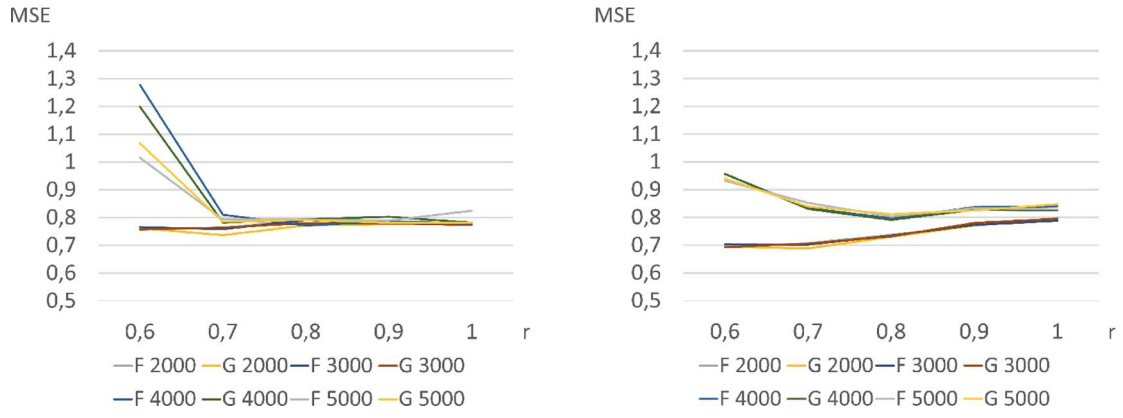


Fig. 4. MSE calculated on classification result for Wine Quality-White dataset for ELM-RBF with mean shift for r varying between 0.6 and 1, 2000–5000 neurons applying linear (left) or Softplus activation function (right), and Gaussian (G) or flat kernel (F).

the rate of improvement in classification results decelerates. The best 39 results are generated for $n \leq 200$. The lowest ACC for linear function equal to 0.599 is attained for k-medoids $k = 3$, $n = 900$ and the worst overall result for the considered methods and parameters reads as ACC = 0.587, and is achieved for k-medoids with Softplus activation function combined with $n = 200$ and $k = 11$. ELM-RBF with mean shift is also used as a classification method on Ionosphere dataset. The considered parameters' values are: n ranging from 100 to 1000 (with step-size 100), r from 0.5 to 7 (step-size 0.5) rendering from 1 to 216 clusters (see Fig. 6). The highest ACC = 0.808 is registered for Softplus activation function, $n = 100$, $r = 4.5$ rendering around 14 clusters. The 40 best results calculated with Softplus function are generated for $r = 4.5$ or $r = 5$ yielding around 14 and 4 clusters, respectively.

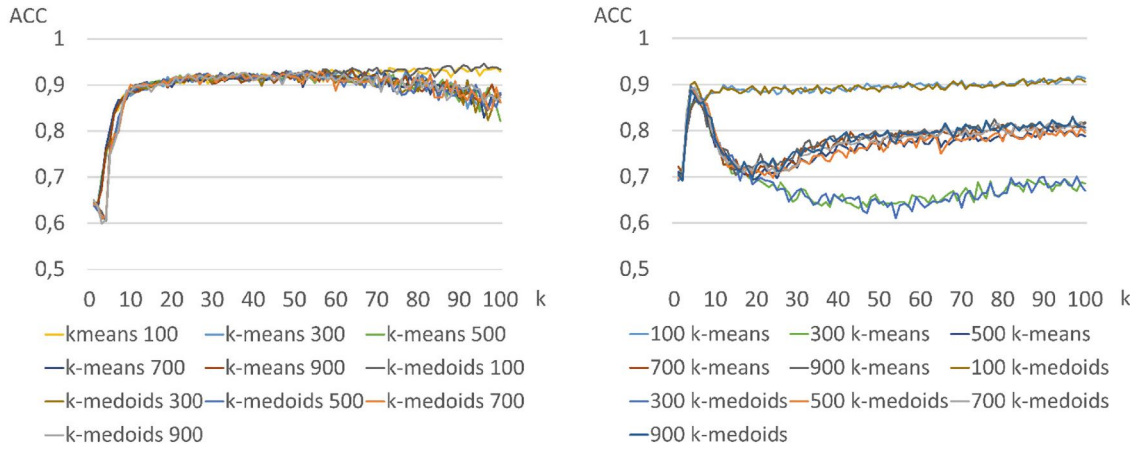


Fig. 5. ACC calculated on classification result for Ionosphere dataset for ELM-RBF with k-means or k-medoids for k varying between 1 and 100, 100–900 neurons applying linear (left) or Softplus activation function (right).

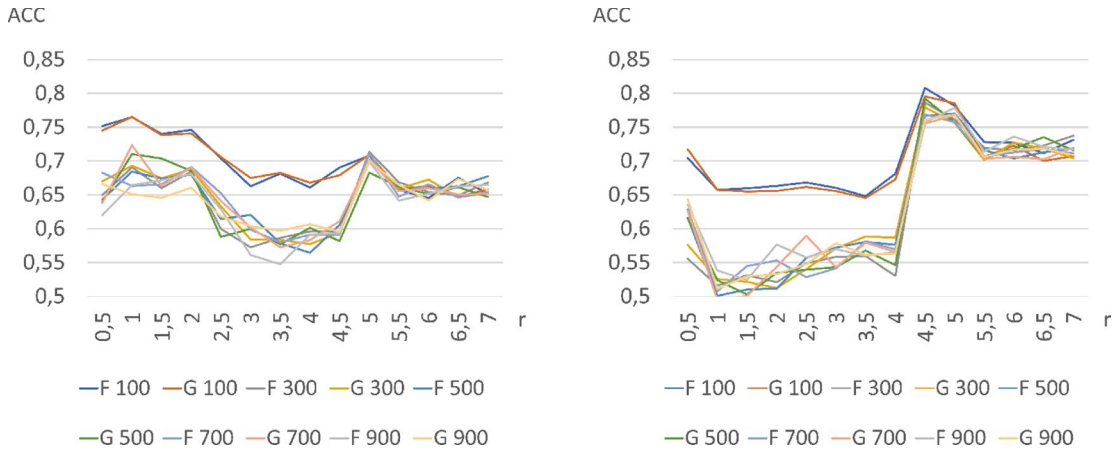


Fig. 6. ACC calculated on classification result for Ionosphere dataset for ELM-RBF with mean shift for r varying between 0.5 and 7, 100–900 neurons applying linear (left) or Softplus activation function (right), and Gaussian (G) or flat kernel (F).

K-means and k-medoids applied to ELM-RBF for clustering are tested on Ionosphere for larger number of neurons in the hidden layer. Tested parameters are n ranging from 1500 to 5000 (with step-size 500) and k from 1 to 100 (see Fig. 7). These computations are performed on the computer K3. The best ACC $= 0.937$ is attained for k-means with linear function, $n = 1500$ and $k = 49$. In contrast, the worst ACC equal to 0.602 is obtained for k-medoids with linear function, $n = 3000$ and $k = 4$. The best result for $1500 \leq n \leq 5000$ equal to 0.937 is lower then the best result for $100 \leq n \leq 1000$ reading as ACC $= 0.946$.

The classification process is performed on Ionosphere with ELM-RBF combined with mean shift testing the values of n ranging from 1500 to 5000 (with step-size 500) (see Fig. 8). The admitted values of r are taken from 0.5 to 7 (with step-size 0.5). The top 32 ACC, which are higher than 0.74, are obtained for $r = 4.5$ or $r = 5$ (around 13 or 4 centroids, respectively). The best ACC = 0.792 is achieved with Softplus function, $n = 3500$, $r = 4.5$ and Gaussian kernel. This result is worse than the best one calculated for $100 \leq n \leq 1000$ - ACC = 0.808.

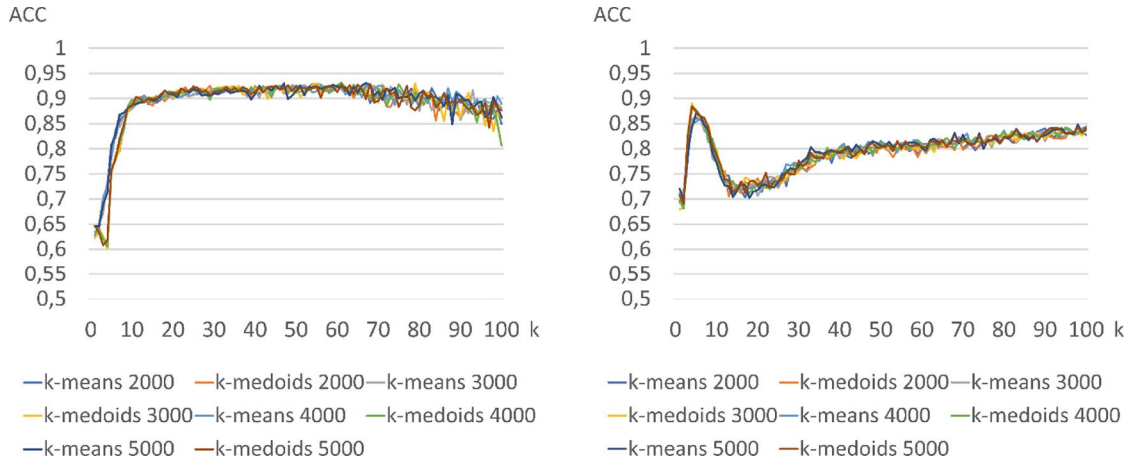


Fig. 7. ACC calculated on classification result for Ionosphere dataset for ELM-RBF with k-means or k-medoids for k varying between 1 and 100, 2000–5000 neurons applying linear (left) or Softplus activation function (right).

Feature selection method - Fast Correlation Based Filter (FCBF) [22] - is applied on Ionosphere to reduce the set of features leaving those that are highly correlated with affiliation to the class and their correlation between other features is low. The initial set of features is reduced from 34 to 4 represented by numbers: 5, 6, 28 and 33. The real aim of applying feature selection filtering is the hope to improve classification result and to reduce computation time. In the next step, ELM-RBF combined with k-means and k-medoids is tested on the set of features selected from Ionosphere. These computations are performed on the computer K2. The parameters involved are n varying between 100 and 1000 (with step-size 100) and k ranging from 1 to 100 (with step-size 1). The best ACC equal to 0.910 is obtained for k-means, linear activation function, for $n = 200$ and $k = 34$ and is computed in 11 s. This result is worse then the best result for the whole set of features which is equal to 0.946. The classification is also conducted on the selected Ionosphere features applying ELM-RBF with mean shift clustering method for n running from 100 to 1000 (with step-size 100), $r \in [0.01, 1.5]$ (with step-size 0.01) rendering from 1 up to 255 clusters. The best achieved result ACC = 0.892 is obtained for linear activation function (on the whole set of features Softplus activation function gave the best results for mean shift), flat kernel, $n = 100$, $r = 0.51$ rendering 40 clusters and is calculated in 6 s outperforming the best result obtained for the whole set of features applying mean shift clustering

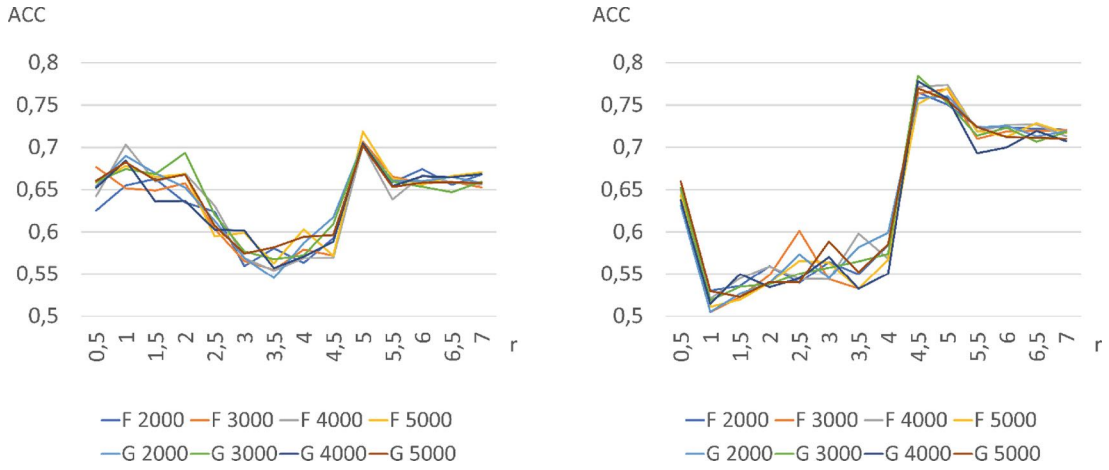


Fig. 8. ACC calculated on classification result for Ionosphere dataset for ELM-RBF with mean shift for r varying between 0.5 and 7, 2000–5000 neurons applying linear (left) or Softplus activation function (right), and Gaussian (G) or flat kernel (F).

method which is equal to 0.808 for Softplus, flat kernel, $n = 100$ and $r = 4.5$ (14 clusters).

A juxtaposition of the best ACC or MSE generated for all the selected ranges of parameters for ELM-RBF with k-means, k-medoids and mean shift for Wine Quality-White and Ionosphere is presented in Table 1 and Table 2.

In previous computations, the default distances for k-means and k-medoids were applied and the activation functions (Softplus and linear) were selected a priori as they gave the best classification results for Wine Quality-White dataset. Subsequently, the parameters that yielded the best ACC for the whole set of features in Ionosphere for the three considered clustering methods were selected and other distance functions for k-means and k-medoids were considered and combined with various activation functions tested for all three clustering methods. In case of k-means, the tested distances are: *cityblock*, *correlation*, *cosine* and *sqeuclidean*. In case of k-medoids: *chebychev*, *cityblock*, *correlation*, *cosine*, *euclidean*,

Table 1. The best classification results (measured with MSE) on Wine Quality-White dataset applying ELM-RBF for each of the considered clusterization methods (k-means, k-medoids and mean shift) on analyzed ranges of parameters. *Act fun* column stands here for the activation function, t for computation time in seconds and *rng* for range.

n rng	k rng	r rng	Method	act fun	n	r	k	Kernel	MSE	t
100–1000	10–100	-	k-medoids	linear	700	-	100.0	-	0.6188	409
100–1000	10–100	-	k-means	linear	700	-	100.0	-	0.6202	402
100–1000	-	0.3–1.3	mean shift	Softplus	700	0.6	21.5	flat	0.6856	323
1500–5000	10–100	-	k-means	Softplus	4000	-	10.0	-	0.7315	2357
1500–5000	10–100	-	k-medoids	Softplus	4500	-	10.0	-	0.7368	2566
1500–5000	-	0.6–1	mean shift	Softplus	2000	0.7	10.8	Gaussian	0.6882	1303

Table 2. The best classification results (measured with ACC) on Ionosphere dataset applying ELM-RBF for each of the considered clustering methods (k-means, k-medoids and mean shift) on analyzed ranges of parameters. *Act fun* stands here for the activation function, *rng* for range, *krn* for kernel (f - flat or g - Gaussian) and *t* for computation time in seconds. In features column there is an information about whether the computations were performed on the whole set of features or only on ones selected by FCBF.

n rng	k rng	r rng	features	Method	act fun	n	r	k	krn	ACC	t
100–1000	1–100	-	all	k-medoids	linear	100	-	96	-	0.946	13
100–1000	1–100	-	all	k-means	linear	200	-	80	-	0.941	9
100–1000	-	0.5–7	all	mean shift	Softplus	100	4.50	14	f	0.808	5
1500–5000	1–100	-	all	k-means	linear	1500	-	49	-	0.938	60
1500–5000	1–100	-	all	k-medoids	linear	4500	-	48	-	0.933	169
1500–5000	-	0.5–7	all	mean shift	Softplus	3500	4.50	13	g	0.792	155
100–1000	1–100	-	selected	k-means	linear	200	-	34	-	0.910	11
100–1000	1–100	-	selected	k-medoids	linear	200	-	41	-	0.907	14
100–1000	-	0.01–1.5	selected	mean shift	linear	100	0.51	40	f	0.892	6

hamming, *jaccard*, *minkowski*, *spearman* and *sqeuclidean* were analyzed [13]. For k-means, k-medoids and mean shift the tested activation functions for ELM-RBF are: sigmoid, tanh, relu, rbf, linear, swish, ELiSH, HardTanH, TanhRe, ELUs, Softplus, LReLU and BinaryStep [16]. The best five results obtained for k-means and k-medoids are presented in Table 3. For both k-means and k-medoids the best classification results are rendered for linear activation function. The highest ACC is obtained for *cityblock* (ACC = 0.94) and *sqeuclidean* distance (ACC = 0.93). The best five results for mean shift with $n = 100$, $r = 4.5$ (rendering 14

Table 3. The top 5 ACC on the Ionosphere dataset for k-means, $n = 100$, $k = 96$ and k-medoids, $n = 200$, $k = 80$ analyzed for different distances and activation functions.

Distance	Method	activation function	n	k	ACC	time
cityblock	k-medoids	linear	100	96	0.940	11
sqeuclidean	k-medoids	linear	100	96	0.933	11
cityblock	k-medoids	TanhRe	100	96	0.929	13
cityblock	k-medoids	ELUs	100	96	0.927	13
cityblock	k-medoids	Softplus	100	96	0.927	12
cityblock	k-means	linear	200	80	0.942	9
sqeuclidean	k-means	linear	200	80	0.927	9
sqeuclidean	k-means	relu	200	80	0.850	10
sqeuclidean	k-means	HardTanH	200	80	0.833	11
cityblock	k-means	TanhRe	200	80	0.828	11

or 13 clusters) and flat kernel, which are computed in $t \in [4.1, 5.1]$ seconds are rendered for: Softplus, tanh, swish, TanhRe and sigmoid activation functions, and their ACC are equal to: 0.794, 0.783, 0.780, 0.769 and 0.763, respectively.

6 Conclusions

The best classification results for Wine Quality-White with ELM-RBF are obtained for k-medoids and k-means. The latter is attained for $n \in N_{1000} = \{100, 200, \ldots, 1000\}$ with MSE = 0.62 which is 0.07 better than the best result applying mean shift. As it turns out, further enlargement of n (from 1500 to 5000) does not improve the MSE results. The best classification result is achieved for $n = 700$. Similarly, the best classification results for Ionosphere dataset are derived for k-means and k-medoids with $n \in N_{1000}$ for which ACC $\in [0.94, 0.95]$. Again, admitting a higher number of neurons on this dataset does not ameliorate the classification results - a fenomenon also manifested on Wine Quality-White data. The best classification result is achieved for $n = 100$, which is the lower bound of the analyzed numbers of neurons. For both datasets and lower values of n the best results for k-means and k-medoids are attained for high values of $k \in \{80, 81, \ldots, 100\}$. Again, admitting a higher number of neurons n the best results are obtained for lower values of k, i.e. $k = 10$ for Wine Quality-White and $k \in \{48, 49\}$ for Ionosphere. Furthermore, classification conducted on features selected by FCBF from Ionosphere dataset does not improve the overall best classification result. Nevertheless, the ACC obtained by ELM-RBF with mean shift on reduced set of data increases accuracy rate from 0.81 to 0.89 ACC. For k-means and k-medoids the best results are obtained with the aid of linear activation function (and *cityblock* or *sqeuclidean* distance metrics). Mean shift rendered in most computations the best results on Softplus activation function; however, the best outcome achieved with this clustering method on reduced set of Ionosphere features is attained with linear activation function. In further research, one should verify results rendered applying other parameters especially when the best classification in this work is observed on their boundary values as it is expected to obtain in those cases better results. The *cityblock* metric should be further analyzed for k-means and k-medoids in ELM-RBF.

References

1. Arthur, D., Vassilvitskii, S.: K-means++: the advantages of careful seeding. In: Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 1027–1035 (2007)
2. Dhini, A., Surjandari, I., Kusumoputro, B., Kusiak, A.: Extreme learning machine - radial basis function (ELM-RBF) networks for diagnosing faults in a steam turbine. J. Ind. Prod. Eng. **39**(7), 572–580 (2022). https://doi.org/10.1080/21681015.2021.1887948
3. Dua, D., Graff, C.: UCI machine learning repository (2017). http://archive.ics.uci.edu/ml

4. Finkston, B.: Mean shift clustering (2023). http://bit.ly/3wVVngu
5. Fukunaga, K., Hostetler, L.: The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Trans. Inf. Theory **21**(1), 32–40 (1975). https://doi.org/10.1109/TIT.1975.1055330
6. Gong, H.: An open-source implementation of meanshift clustering for matlab/octave (2015). https://github.com/hangong/meanshift_matlab
7. Huang, G.B., Zhu, Q.Y., Siew, C.K.: Extreme learning machine: a new learning scheme of feedforward neural networks. In: IEEE Proceedings of International Joint Conference on Neural Networks, vol. 2, pp. 985–990 (2004). https://doi.org/10.1109/IJCNN.2004.1380068
8. Kaufman, L., Rousseeuw, P.: Finding Groups in Data: An Introduction to Cluster Analysis. Wiley, Hoboken (1990). https://doi.org/10.2307/2532178
9. Konopka, A., et al.: Classification of soil bacteria based on machine learning and image processing. In: Groen, D., de Mulatier, C., Paszynski, M., Krzhizhanovskaya, V.V., Dongarra, J.J., Sloot, P.M.A. (eds.) ICCS 2022. LNCS, vol. 13352, pp. 263–277. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-08757-8_23
10. Leung, H.C., Leung, C.S., Wong, E.W.M.: Fault and noise tolerance in the incremental extreme learning machine. IEEE Access **7**, 155171–155183 (2019). https://doi.org/10.1109/ACCESS.2019.2948059
11. Li, H.T., Chou, C.Y., Chen, Y.T., Wang, S.H., Wu, A.Y.: Robust and lightweight ensemble extreme learning machine engine based on eigenspace domain for compressed learning. IEEE TCAS-I **66**(12), 4699–4712 (2019). https://doi.org/10.1109/TCSI.2019.2940642
12. Lu, S., Wang, X., Zhang, G., Zhou, X.: Effective algorithms of the Moore-Penrose inverse matrices for extreme learning machine. Intell. Data Anal. **19**(4), 743–760 (2015). https://doi.org/10.3233/IDA-150743
13. MathWorks: k-medoids clustering - Matlab k-medoids (2023). http://bit.ly/3RwXlNR
14. McCulloch, W.S., Pitts, W.: A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. **5**(4), 115–133 (1943). https://doi.org/10.1007/BF02478259
15. Mojrian, S., et al.: Hybrid machine learning model of extreme learning machine radial basis function for breast cancer detection and diagnosis; a multilayer fuzzy expert system. In: RIVF, pp. 1–7 (2020). https://doi.org/10.1109/RIVF48685.2020.9140744
16. Nader, A., Azar, D.: Evolution of activation functions: an empirical investigation. ACM TELO **1**(2), 1–36 (2021). https://doi.org/10.1145/3464384
17. Park, H.S., Jun, C.H.: A simple and fast algorithm for k-medoids clustering. Expert Syst. Appl. **36**(2), 3336–3341 (2009). https://doi.org/10.1016/j.eswa.2008.01.039
18. Peng, X., Lin, P., Zhang, T., Wang, J.: Extreme learning machine-based classification of ADHD using brain structural MRI data. PLoS ONE **8**(11), 1–12 (2013). https://doi.org/10.1371/journal.pone.0079476
19. Pérez-Ortega, J., Almanza-Ortega, N.N., Vega-Villalobos, A., Pazos-Rangel, R., Zavala-Díaz, C., Martínez-Rebollar, A.: The k-means algorithm evolution. In: Introduction to Data Science and Machine Learning, chap. 5. IntechOpen, Rijeka (2019). https://doi.org/10.5772/intechopen.85447
20. Rao, C.R., Mitra, S.K.: Generalized Inverse of Matrices and its Applications. Wiley, Hoboken (1971)

21. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by back-propagating errors. Nature **323**(6088), 533–536 (1986). https://doi.org/10.1038/323533a0
22. Yu, L., Liu, H.: Feature selection for high-dimensional data: a fast correlation-based filter solution. In: ICML, pp. 856–863 (2003)

3.5 Automated imaging and machine learning for soil bacteria classification: Challenges and insights

Publication: A. Konopka, R. Kozera, L. Sas-Paszt, and P. Trzciński, "Automated imaging and machine learning for soil bacteria classification: Challenges and insights," *Engineering Applications of Artificial Intelligence*, vol. 159C, pp. 1–9, 2025. https://doi.org/10.1016/j.engappai.2025.111369 (140 MNiSW points, IF – 8.0)

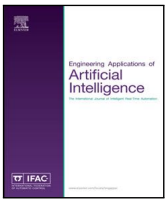
Abstract: Soil bacteria play a vital role in biofertilizer production, making accurate classification of bacteria genera essential for optimizing industrial processes. While previous studies report high classification accuracy, they often overlook the challenges of deploying such systems in real-world production environments, where image acquisition must be automated rather than conducted by the microbiologists manually. This study addresses these practical issues by investigating the automated classification of bacteria images under varying conditions of lighting and glass type. Both classical machine learning approaches and advanced neural networks, including modifications of Extreme Learning Machines - Radial Basis Function and deep convolutional architectures of Residual Neural Networks, were employed. Unforeseen challenges impacting the reliability of automated classification were identified, highlighting novel factors that could significantly affect classification accuracy. This work not only emphasizes the limitations of existing approaches in production settings but also proposes possible directions for future research in automating bacteria classification for industrial biofertilizer development. Importantly, it brings attention to a broader issue: the critical need for rigorous dataset preparation and validation when applying artificial intelligence to microscopic image analysis across biological domains.

Engineering Applications of Artificial Intelligence 159 (2025) 111369

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



Research paper

Automated imaging and machine learning for soil bacteria classification: Challenges and insights☆

Aleksandra Konopka [a,*], Ryszard Kozera [a,b], Lidia Sas-Paszt [c], Paweł Trzciński [c]

[a] Institute of Information Technology, Warsaw University of Life Sciences - SGGW, ul. Nowoursynowska 166, Warsaw, 02-776, Poland
[b] The School of Physics, Mathematics and Computing, The University of Western Australia, 35 Stirling Highway, Perth, Australia
[c] The National Institute of Horticultural Research, ul. Konstytucji 3 Maja 1/3, Skierniewice, 96-100, Poland

ARTICLE INFO

Keywords:
Soil bacteria
Microscopic image classification
Machine learning
Extreme Learning Machine
Residual Neural Network

ABSTRACT

Soil bacteria play a vital role in biofertilizer production, making accurate classification of bacteria genera essential for optimizing industrial processes. While previous studies report high classification accuracy, they often overlook the challenges of deploying such systems in real-world production environments, where image acquisition must be automated rather than conducted by the microbiologists manually. This study addresses these practical issues by investigating the automated classification of bacteria images under varying conditions of lighting and glass type. Both classical machine learning approaches and advanced neural networks, including modifications of Extreme Learning Machines - Radial Basis Function and deep convolutional architectures of Residual Neural Networks, were employed. Unforeseen challenges impacting the reliability of automated classification were identified, highlighting novel factors that could significantly affect classification accuracy. This work not only emphasizes the limitations of existing approaches in production settings but also proposes possible directions for future research in automating bacteria classification for industrial biofertilizer development. Importantly, it brings attention to a broader issue: the critical need for rigorous dataset preparation and validation when applying artificial intelligence to microscopic image analysis across biological domains.

1. Introduction

Soil bacteria play a vital role in the production of biofertilizers (Lalitha, 2017), contributing significantly to sustainable agriculture (Khmelevtsova et al., 2022). This underscores the importance of accurately identifying and distinguishing between different soil bacteria. Among the various approaches available (Bonnichsen et al., 2019; Ferrari et al., 2006), one of the methods involves analyzing microscopic images of the bacteria (Konopka et al., 2022). The microbiology experts distinguish between various bacteria genera by manually examining images, a process that, though accurate, is still time-intensive and laborious. To expedite this task, computational methods were introduced. Initial efforts utilized classical machine learning techniques that relied on handcrafted features for classification (Konopka et al., 2023a). More recently, deep learning approaches were employed, enabling algorithms to autonomously extract features critical for accurate classification (Konopka et al., 2023c). These computerized methods have demonstrated impressive accuracy (ACC), reaching up to 0.97 on soil bacteria datasets (Konopka et al., 2022), depending on the specific algorithm employed and the considered image dataset.

However, these results are largely based on datasets meticulously prepared by microbiological experts (Konopka et al., 2022, 2023a,c), who ensure high-quality images with sharp focus and minimal variability. Each image necessitates detailed analysis and a tailored approach, requiring considerable time, expertise and qualified human resources. Moreover, previous studies have not thoroughly investigated the impact of external variables such as varying lighting conditions and different types of glass slides, which could significantly influence classification outcomes (Konopka et al., 2023a). Variations in lighting can have a significant impact on the features extracted from images, subsequently affecting classification accuracy (Eady and Park, 2016). In many cases, these parameters are either normalized during dataset creation or omitted from consideration. Nevertheless, such limitations present challenges when models trained under controlled conditions are deployed in diverse real-world environments.

☆ This paper was financed by The National Centre for Research and Development, Poland within the framework of the BIOSTRATEG project, contract number BIOSTRATEG3/344433/16/NCBR/2018.

* Corresponding author.
E-mail addresses: aleksandra_konopka@sggw.edu.pl (A. Konopka), ryszard_kozera@sggw.edu.pl (R. Kozera), lidia.sas@inhort.pl (L. Sas-Paszt), pawel.trzcinski@inhort.pl (P. Trzciński).

https://doi.org/10.1016/j.engappai.2025.111369
Received 7 February 2025; Received in revised form 13 April 2025; Accepted 31 May 2025
Available online 19 July 2025

To address these gaps, this research adopts a fully automated approach, extending automation beyond classification to the data acquisition stage. This strategy aims to evaluate its feasibility for large-scale applications, where efficiency, scalability and reproducibility are paramount. To minimize the time and human resources required for preparing a large dataset, the images in this study are automatically captured with a specialized microscope. Automatically prepared dataset inevitably contains images of varying quality, including those with suboptimal sharpness. While this introduces technical obstacles, such imperfections reflect realistic conditions in automated systems and must be accounted for to ensure robustness.

Furthermore, the dataset for this study was designed to include images captured under varying combinations of glass types (three types) and lighting conditions (two conditions). These factors may influence the color of the images, potentially affecting the classification performance of machine learning methods. The latter ensures a diverse dataset, better simulating real-world variability. The objective of this research is to determine whether a fully automated workflow can achieve satisfactory classification performance.

This research explores the application of a broad spectrum of classical methods such as: Support Vector Machine, Multilayer Perceptron, K Nearest Neighborhood or Random Forest (Konopka et al., 2023a); and advanced machine learning algorithms including Extreme Learning Machines and Residual Neural Networks (Konopka et al., 2023c), and their state-of-the-art modifications (Konopka et al., 2023b), to classify soil bacteria. The utilized dataset comprises over 3,000 high-resolution images of four selected soil bacteria genera, captured through an automated process with the KEYENCE VHX-7000N Digital Microscope. These bacteria genera were selected by microbiology experts for their significance in biofertilizer production.

The classification is conducted on various sets and subsets of images captured under different lighting conditions and using different glass types. These results are thoroughly analyzed and compared with outcomes obtained using alternative methods and approaches to the classification problem - for instance, employing majority voting (MV) approach on subimages extracted from full images (Struniawski et al., 2022). The parameters of the methods used in this study are optimized for the specific problem, enabling improved classification performance.

The presented study offers a comprehensive analysis of the challenges in automating the classification of bacteria samples, emphasizing critical variables that affect such classification process. Although the current model has been trained on four bacteria types, which represents its limitation, it can be further fine-tuned with additional datasets. This study suggests also possible future advancements in the area in question. More importantly, it highlights broader issues in applying artificial intelligence to microbiology, extending beyond bacteria classification to address universal issues to overcome faced by both information technology experts and microbiologists. By identifying these obstacles, the research provides valuable insights for developing more robust and reliable classification systems, fostering collaboration between disciplines to ensure the accuracy and scalability of automated solutions.

2. Dataset preparation

Four bacteria genera were grown on Plate Count Agar medium. The bacteria genera were: *B1* - *Pseudomonas*, *B2* - *Pantoea*, *B3* - *Enterobacter* and *B4* - *Rhizobium*. The samples were manually collected from petri dishes and placed on microscope slide glassess. The considered glasses were: *G1* - LabGlass, *G2* - Profilab and *G3* - Chemland. Only a single bacteria genus was placed on each glass slide, with no mixing of genera. The prepared samples are the following combinations of glass and bacteria genera: *G1B1*, *G1B2*, *G1B3*, *G1B4*, *G2B1*, *G2B2*, *G2B3*, *G2B4*, *G3B1*, *G3B2*, *G3B3* and *G3B4*.

The samples prepared in this manner were photographed using KEYENCE VHX-7000N Digital Microscope. Each sample was photographed under two different lighting conditions, *L1* and *L2*. The *L1* light was directed from above, causing the sample to appear blue, while the *L2* light was directed from below, resulting in the sample appearing gray. *L1* - blue-tinted top lighting enhanced the natural iridescence and surface texture of bacteria colonies, with the cool hue arising from the microscope's LED spectral properties and its interaction with biofilm reflectivity. *L2* - neutral gray bottom lighting maintained the true color of translucent bacteria structures, as the diffused, spectrally balanced transmission illumination prevented optical artifacts that could distort pigmentation or cell wall features. The microscope was manually adjusted to ensure the bacteria were in sharp focus. Next, the area to be photographed was manually selected. This area was photographed in a tiled manner, with each image captured side by side. Once the area was defined, the microscope moved automatically, taking approximately 150 images across the selected region. The imaged areas of the sample under the two different lighting conditions did not overlap; instead, each lighting condition captured a distinct area containing bacteria. The images that did not include any bacteria, only the background were removed from the dataset.

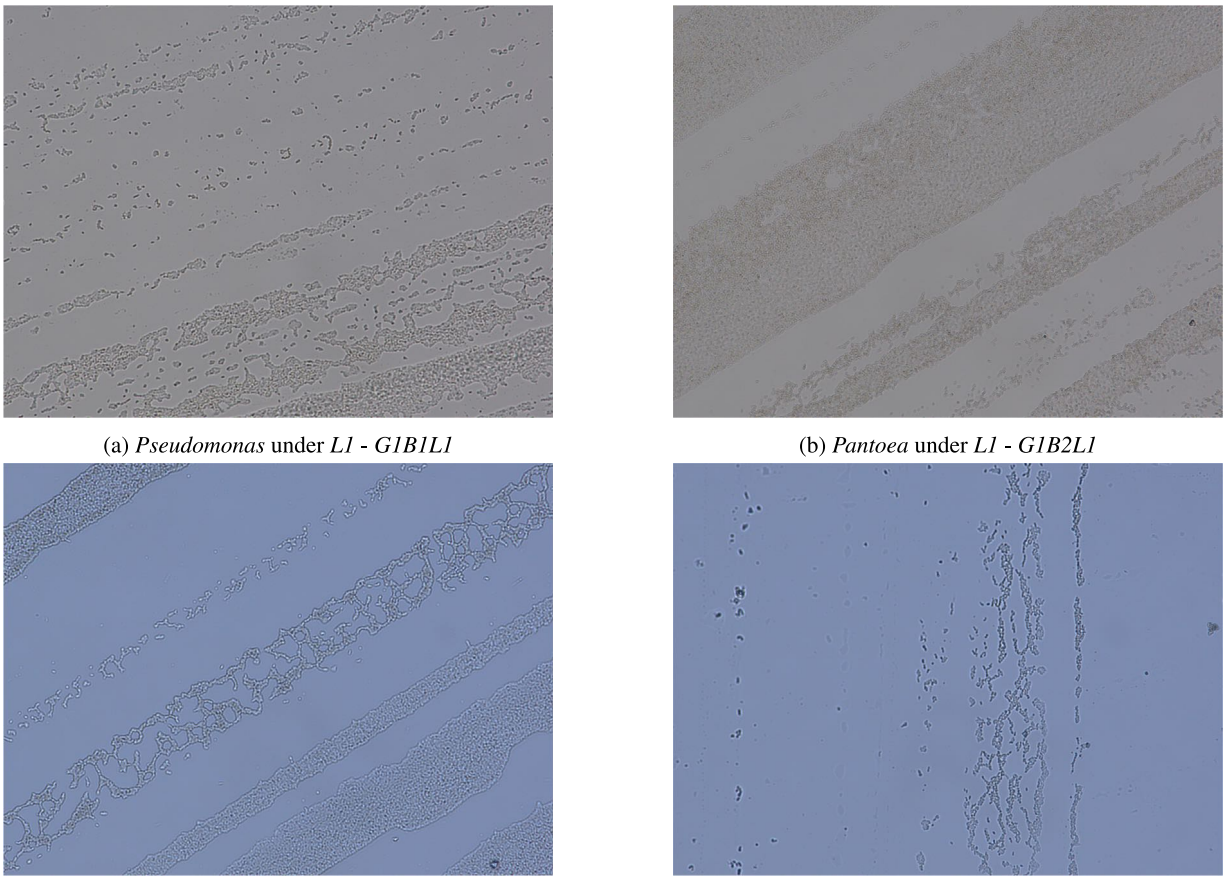
The full image dataset consists of 3366 images each in 2880×2160 resolution. The images were taken on three different glasses *G1*, *G2* and *G3*, applying different lighting conditions *L1* and *L2*. Each image contains only one bacteria genera *B1*, *B2*, *B3* or *B4*. This yields following combinations of subsets: *G1B1L1* - 165 images, *G1B1L2* - 165, *G1B2L1* - 164, *G1B2L2* - 158, *G1B3L1* - 143, *G1B3L2* - 165, *G1B4L1* - 107, *G1B4L2* - 107, *G2B1L1* - 154, *G2B1L2* - 123, *G2B2L1* - 144, *G2B2L2* - 144, *G2B3L1* - 108, *G2B3L2* - 126, *G2B4L1* - 121, *G2B4L2* - 154, *G3B1L1* - 142, *G3B1L2* - 140, *G3B2L1* - 132, *G3B2L2* - 132, *G3B3L1* - 143, *G3B3L2* - 143, *G3B4L1* - 143, *G3B4L2* - 143. Exemplary images are presented (See Fig. 1).

The experiments were conducted using an automatically generated dataset. While the focus was manually adjusted at the initial region of the dataset, no further adjustments were made for images captured from the selected areas. As a result, some images are clear, while others lack proper focus on the bacteria. Despite this, all images were retained in the dataset to prevent overfitting of the model in a training process and to reflect the types of images that would need to be classified in a production environment, should this approach be applied in practice.

3. Methods

The concept of neural networks was introduced in 1943 by Warren McCulloch and Walter Pitts (McCulloch and Pitts, 1943), inspired by a simplified model of biological neurons. The first artificial neural network, known as the perceptron, was developed in 1958 by Frank Rosenblatt (Rosenblatt, 1958). Designed to mimic the human brain's ability to process data and learn to distinguish between objects, the perceptron, however, was limited to solving only linearly separable problems. To overcome this limitation, researchers combined neurons into multilayer networks (Alsmadi et al., 2009). As more layers of neurons were added, the computations became increasingly complex and difficult for humans to interpret, leading to the development of what is now known as deep learning (Luo, 2022). These networks rely on the Backpropagation Algorithm (Aggarwal et al., 2024) (or other similar methods) for training and require the use of nonlinear, differentiable activation functions. An efficient alternative to complex deep learning models is the Extreme Learning Machine (ELM) (Huang et al., 2004), which has gained popularity for its rapid training process and ability to achieve high performance across various applications.

Initially, these networks processed input as vectors of hand-crafted features and selecting relevant features was often a complex, uncertain process. Over time, Convolutional Neural Networks (CNNs) (Deng, 2022) emerged, addressing some of these challenges. CNNs accept raw images as input and automatically extract features using alternating convolutional and pooling layers. While this approach enhanced the ability to solve more complex problems and automate feature extraction, issues like the vanishing and exploding gradient problems persisted. Residual Neural Networks (ResNets) emerged as a solution to overcome these issues (He et al., 2016a).



(a) *Pseudomonas* under *L1* - *G1B1L1*

(b) *Pantoea* under *L1* - *G1B2L1*

(c) *Enterobacter* under *L2* - *G1B3L2*

(d) *Rhizobium* under *L2* - *G1B4L2*

Fig. 1. Images taken with KEYENCE VHX-7000N Digital Microscope of different bacteria genera on glass *G1* - LabGlass under two lighting conditions *L1* - light directed from above and *L2* - light directed from below.

3.1. Extreme Learning Machine

Extreme Learning Machine is a highly computationally efficient dense feed-forward neural network introduced by Huang et al. (2004). ELM is extensively utilized for both classification and regression tasks (Huang et al., 2015) finding broad applications across various domains, including biomedical engineering (Ismaeel et al., 2015), computer vision (Safaei et al., 2019), robotics (Huang et al., 2015), big data processing (Cao and Lin, 2015) and real-time applications (Pande et al., 2024). ELM consists of input, hidden and output layer of neurons. The parameters of the model that need to be specified are: number of neurons in hidden layer, values of weights and formulas of activation functions. The number of neurons in the input layer is specified by the task in question and corresponds to the number of considered N vectors of d features, whereas number of neurons in output layer relies on number of classes to which these samples are to be assigned. Number of neurons in the hidden layer needs to be adjusted for the specific task. The weights between input and hidden layer and the biases are initialized by uniform distribution ($U(0,1)$ or $U(-1,1)$). The weights β between hidden and output layer are calculated as a minimizer of the mean residual square error. To compute them $H\beta = Y$ equation is solved for β with single algebraic operation - applying Moore–Penrose pseudo inverse $H^{\dagger}$ (Fampa et al., 2021), since H is non-invertible - thereby accelerating the training process. H is the output of the hidden layer and Y stands for the expected output of the network. Activation functions that were considered in this research are *linear* - $f_L(x) = ax$ and *Softplus* - $f_{SP}(x) = log(1 + exp(x))$.

Extreme Learning Machine - Radial Basis Function (ELM-RBF) (Mojrian et al., 2020) is a variant of ELM which transforms the input data with the aid of a clustering method. The input data $\{x_i\}_{i=0}^{N}$, which are vectors of features, are grouped applying k-means clustering algorithm into k clusters represented by $\{c_j\}_{j=1}^{k}$ centroids, which are d dimensional vectors. The value of k is selected a priori. Each sample is assigned to only one cluster which centroid is the closest in therms of selected metric. Then each sample is mapped into a new feature space applying Radial Basis Function Kernel defined as $K(x_i, c_j) = exp\frac{\|-x_i - c_j\|^2}{2\sigma_j^2}$, where $\sigma_j = \frac{max\{di_j\}}{\sqrt{2k}}$, di_j is a vector of distances to the centroid in j'th cluster. The transformed dataset is applied as input to ELM.

The modifications of ELM-RBF were proposed in Konopka et al. (2023b). It considers choosing k-medoids and mean shift as the alternatives for k-means clustering method. In k-means (Pérez-Ortega et al., 2020) the initial centroids are selected and then their new locations are iteratively computed by the algorithm placing each of them in new a location. This location is the mean of the coordinates of points belonging to a considered cluster. The points are assigned to the centroid which is the closest to them among all the centroids in means of a distance metric. K-medoids algorithm (Kaufman and Rousseeuw, 1990) is based on comparing all the possible swaps of position of a medoid with other samples within a given cluster and choosing the one that decreases the cost function to the greatest extent. In this research in both k-means and k-medoids for distance computations *sqeuclidean* and *cityblock* distance metrics were considered. The *sqeuclidean* metric is computed with $d_{SQ}(a,b) = \sqrt{\sum_{i=1}^{n}(a_i - b_i)^2}$ and *cityblock* with $d_{CB}(a,b) = \sum_{i=1}^{n}|a_i - b_i|$, where a_i, b_i are the coordinates of data points in i'th dimension and n is the number of dimensions. Mean shift algorithm (Fukunaga and Hostetler, 1975) is based on an iterative shifts of a window until it converges to local maximum. The window is shifted to new location

which is the mean of coordinates of samples within the window. This mean can be computed in two ways: either by applying the *flat* kernel, which gives the standard mean, or by using the *gaussian* kernel, where a weighted mean is applied, giving closer points greater influence on the result. The samples are then assigned to the nearest local maxima forming the clusters. In mean shift algorithm the number of clusters is a priori unknown but it is yielded by the algorithm depending on selected range of the window and distribution of the dataset (Konopka et al., 2023b).

3.2. Residual Neural Networks

Residual Neural Networks (ResNets), introduced by He et al. in 2016 (He et al., 2016a), represent a major breakthrough in deep learning. ResNets are widely used in signal processing, particularly in image classification tasks, such as medical imaging (Sýkora et al., 2023; Padmaja et al., 2024) and microorganism classification (Konopka et al., 2023c).

ResNet, as a type of Convolutional Neural Network, consists of two fundamental components:

1. *Feature Extraction*: This component generates feature representations from input images through two key types of layers i.e. convolutional layers and pooling layers. Convolutional layers apply filters to input images, highlighting distinctive features such as edges, texture, or patterns (Aishwarya and Minu, 2021). Pooling layers, in contrast, reduce the spatial dimensions of the images by retaining only the most important information, thereby decreasing computational complexity (Nikzad et al., 2019). The result of this stage is a set of automatically extracted feature vectors.
2. *Classification*: Second component is responsible for classifying the extracted features into specific categories.

ResNets are specifically designed to address the vanishing and exploding gradient problem. This is achieved through the use of residual blocks including skip connections, which allow certain layers to be bypassed (Ebrahimi and Abadi, 2021). Rather than learning a direct transformation from input to output, ResNet learns the residuals which are differences between the input and the expected output, making corrections during training. This residual learning approach helps improve the training process, by facilitating stable gradient flow and enabling more effective learning (Sabba et al., 2022).

To further stabilize and enhance the generalization of the learning process, batch normalization is applied, normalizing the input data to each layer (Peerthum and Stamp, 2023). Additionally, to mitigate the risk of overfitting, dropout is employed, randomly deactivating a subset of neurons in selected layer during training.

Various ResNet architectures can be distinguished based on the number of layers, such as ResNet-18, ResNet-50 and ResNet-152 (He et al., 2016a). These architectures consist of blocks, each containing layers arranged in a specific order. A typical ResNet architecture includes a convolutional layer, batch normalization layer, concatenation and pooling layer. Numerous modifications have been introduced to ResNet models. For example, ResNet-152v2 differs from the original by altering the sequence of layers within the convolutional block (He et al., 2016b).

ResNets can consist of hundreds of layers, resulting in millions of trainable parameters. Training such models from scratch is both time-consuming and resource-intensive. Therefore, transfer learning is commonly used when applying ResNets. This method leverages a set of pretrained parameters from a large dataset, such as ImageNet (Deng et al., 2009) and fine-tunes them to adapt to a specific task (Bal-Ghaoui et al., 2023).

4. Experiments and results

This section is divided into several subsections. Subsequent decisions regarding the next stages were made based on the outcomes of the preceding parts of the experiment. The computations were performed in Python on AMD Ryzen 5 5650 with 32 GB DDR4 and RTX 3060 GPU.

4.1. Classification on whole dataset

The subsets were divided into 4 classes representing bacteria genera:

1. *B1*: *G1B1L1*, *G1B1L2*, *G2B1L1*, *G2B1L2*, *G3B1L1*, *G3B1L2*.
2. *B2*: *G1B2L1*, *G1B2L2*, *G2B2L1*, *G2B2L2*, *G3B2L1*, *G3B2L2*.
3. *B3*: *G1B3L1*, *G1B3L2*, *G2B3L1*, *G2B3L2*, *G3B3L1*, *G3B3L2*.
4. *B4*: *G1B4L1*, *G1B4L2*, *G2B4L1*, *G2B4L2*, *G3B4L1*, *G3B4L2*.

On each of the images from the dataset a total of 30 color and texture features were computed, as these characteristics yielded high accuracies in previous research (Konopka et al., 2022). The considered color features were: mean of red color, mean of green color, mean of blue color, mean of grayscale, standard deviation of red color, standard deviation of green color, standard deviation of blue color, standard deviation of grayscale, kurtosis of red color, kurtosis of green color, kurtosis of blue color and kurtosis of grayscale. Texture features were computed as Gray-Level Co-occurrence Matrix (GLCM) properties across four angular orientations: 0, $\pi/4$, $\pi/2$, $3\pi/4$ at three different distances: contrast for distance 1, dissimilarity for distance 1, homogeneity for distance 1, energy for distance 1, correlation for distance 1, Angular Second Moment (ASM) for distance 1, contrast for distance 20, dissimilarity for distance 20, homogeneity for distance 20, energy for distance 20, correlation for distance 20, ASM for distance 20, contrast for distance 30, dissimilarity for distance 30, homogeneity for distance 30, energy for distance 30, correlation for distance 30 and ASM for distance 30.

The computed features were applied for classification task among the classes *B1*, *B2*, *B3* and *B4*. The classification was performed on the following classical machine learning methods: Support Vector Machine, Random Forest, K Nearest Neighborhood and MultiLayer Perceptron. These methods take feature vectors as input, allowing scientists to choose which features should be considered. To ensure the significance of the results the experiments involved 20% cross validation repeated 10 times. The selected methods had previously adjusted parameters which yielded best accuracy results in previous microorganism classification research (Konopka et al., 2023a): Support Vector Machine (with linear kernel function), Random Forest (200 trees), K Nearest Neighborhood ($k = 1$) and MultiLayer Perceptron (net topology: $15 \times 15 \times 15$). The classification accuracy results yielded for the following methods on the bacteria dataset were: Support Vector Machine - 71.87, Random Forest - 89.42, K Nearest Neighborhood - 83.81 and MultiLayer Perceptron - 88.15.

The classification task has also been performed with Extreme Learning Machine Radial Basis Function with different clustering methods. These computations also involved full set of color and texture features and 20% cross validation approach repeated 10 times. The different clustering methods were: k-means, k-medoids and mean shift. As these modifications of ELM-RBF and their state-of-art modifications were not applied for microorganisms classification tasks before the adjustment of parameters needed to be performed. For k-means and k-medoids clustering methods *sqeuclidean* and *cityblock* metrics were considered. The selected activation function was *linear*. The analyzed values of k were $\{10, 25, 50, 75, 100, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500\}$. For mean shift the considered kernels were *flat* and *Gaussian*; the tested activation functions were *linear* and *Softplus*, and the considered values of $r = \{0.01, 0.02, 0.04, 0.06, 0.08\}$ and from 0.1 to 1.5 by 0.1. For all three clustering methods the considered numbers of neurons were $n = \{100, 250, 500, 1000, 1500\}$. The highest accuracy results for the considered ranged of parameters are presented

Table 1
The best classification results (measured with ACC) on set of color and texture features applying ELM-RBF for each of the clustering methods (k-means, k-medoids and mean shift) on analyzed ranges of parameters. *Act fun* stands for activation function, *rng* for range and *krn* for kernel (*f* - *flat* or *g* - *Gaussian*).

Method	*n rng*	*k rng*	*r rng*	*act fun*	distance	*krn*	*n*	*r*	*k*	ACC
k-means	100–1500	10–1500	–	*linear*	*sqeuclidean*	–	1500	-	1100	0.9525
k-means	100–1500	10–1500	–	*linear*	*cityblock*	–	1000	-	900	0.9487
k-medoids	100–1500	10–1500	–	*linear*	*sqeuclidean*	–	1000	-	1200	0.9554
k-medoids	100–1500	10–1500	–	*linear*	*cityblock*	–	1000	-	1000	0.9444
mean shift	100–1500	–	0.01–1.5	*Softplus*	–	*f*	1500	0.1	-	0.7850
mean shift	100–1500	–	0.01–1.5	*Softplus*	–	*g*	1000	0.2	-	0.7858
mean shift	100–1500	–	0.01–1.5	*linear*	–	*f*	1500	0.1	-	0.8066
mean shift	100–1500	–	0.01–1.5	*linear*	–	*g*	1500	0.1	-	0.8080

in Table 1. The best classification results for ELM-RBF with mean shift clustering method yields 0.808 accuracy obtained for $n = 1500$, *Gaussian* kernel, $r = 0.1$ which is lower than most of the considered classical machine learning methods. The best ACC for k-means yields 0.9525 $n = 1500$, $k = 1100$, *sqeuclidean* distance and the best ACC for k-medoids 0.9554 is obtained for $n = 1000$, $k = 1200$, *sqeuclidean* distance. ELM-RBF with k-medoids achieved the best classification results which outperformed classical machine learning methods on the considered dataset.

To compare results obtained with classical machine learning approach with deep learning a Convolutional Neural Network - Residual Neural Network - ResNet152v2 which yielded best classification results (comparing with other ResNets) on other microorganism datasets (Konopka et al., 2023c) was applied. The comparison was intentionally limited to this one CNN in order to maintain clarity and focus in the manuscript, which already includes a diverse set of experiments. The input data to ResNet152v2 are images, not the vectors of features as in previous methods. The images from the dataset (scaled down by 4 to reduce computational complexity and computation time) were applied as input. To perform classification first the dataset was divided into training, testing and validation sets in 7:1:2 ratio. The applied batch size for the computation purposes was 16, number of epochs 1000 and early stopping with patience 50. Transfer learning was applied the parameters were pretrained on ImageNet dataset and the parameters in convolution blocks were finetuned starting from *conv5_block1_1_conv*. The average pooling method was selected. The classification layers consisted of: dense layer with 512 neurons and *ReLU* activation function, 30% dropout layer, dense layer with 512 neurons and *ReLU* activation function, 30% dropout layer, dense layer with number of neurons equal to number of classes (4) and *Softmax* activation function. The structure of the classification component was previously selected and adjusted based on other microorganism datasets (Konopka et al., 2023c). The applied optimizer was *Adam* with 0.01 learning rate. The yielded accuracy applying ResNet152v2 is equal to 0.99 which outperforms not only the classical machine learning methods but also ELM-RBF with different clustering methods.

4.2. Classification on selected subsets: different combinations of glass and light

ResNet152v2 yielded the highest classification results, which was the reason it was considered for further analysis in this research. The accuracy 0.99 shows that classification of microorganisms yields very promising results. Then it was decided to check the results of classification for different subsets of the dataset. These subsets differ in the type of glass and lighting conditions. In these computations the following subsets were formed:

1. *G1* - all images on glass *G1*: *G1B1L1*, *G1B1L2*, *G1B2L1*, *G1B2L2*, *G1B3L1*, *G1B3L2*, *G1B4L1*, *G1B4L2*.
2. *G2* - all images on glass *G2*: *G2B1L1*, *G2B1L2*, *G2B2L1*, *G2B2L2*, *G2B3L1*, *G2B3L2*, *G2B4L1*, *G2B4L2*.
3. *G3* - all images on glass *G3*: *G3B1L1*, *G3B1L2*, *G3B2L1*, *G3B2L2*, *G3B3L1*, *G3B3L2*, *G3B4L1*, *G3B4L2*.
4. *G1G2* - all images on glass *G1* and *G2*: *G1B1L1*, *G1B1L2*, *G1B2L1*, *G1B2L2*, *G1B3L1*, *G1B3L2*, *G1B4L1*, *G1B4L2*, *G2B1L1*, *G2B1L2*, *G2B2L1*, *G2B2L2*, *G2B3L1*, *G2B3L2*, *G2B4L1*, *G2B4L2*.
5. *G1G3* - all images on glass *G1* and *G3*: *G1B1L1*, *G1B1L2*, *G1B2L1*, *G1B2L2*, *G1B3L1*, *G1B3L2*, *G1B4L1*, *G1B4L2*, *G3B1L1*, *G3B1L2*, *G3B2L1*, *G3B2L2*, *G3B3L1*, *G3B3L2*, *G3B4L1*, *G3B4L2*.
6. *G2G3* - all images on glass *G2* and *G3*: *G2B1L1*, *G2B1L2*, *G2B2L1*, *G2B2L2*, *G2B3L1*, *G2B3L2*, *G2B4L1*, *G2B4L2*, *G3B1L1*, *G3B1L2*, *G3B2L1*, *G3B2L2*, *G3B3L1*, *G3B3L2*, *G3B4L1*, *G3B4L2*.
7. *L1* - all images with light *L1*: *G1B1L1*, *G1B2L1*, *G1B3L1*, *G1B4L1*, *G2B1L1*, *G2B2L1*, *G2B3L1*, *G2B4L1*, *G3B1L1*, *G3B2L1*, *G3B3L1*, *G3B4L1*.
8. *L2* - all images with light *L2*: *G1B1L2*, *G1B2L2*, *G1B3L2*, *G1B4L2*, *G2B1L2*, *G2B2L2*, *G2B3L2*, *G2B4L2*, *G3B1L2*, *G3B2L2*, *G3B3L2*, *G3B4L2*.
9. *G1G2G3/L1L2* - all images: *G1B1L1*, *G1B2L1*, *G1B3L1*, *G1B4L1*, *G2B1L1*, *G2B2L1*, *G2B3L1*, *G2B4L1*, *G3B1L1*, *G3B2L1*, *G3B3L1*, *G3B4L1*, *G1B1L2*, *G1B2L2*, *G1B3L2*, *G1B4L2*, *G2B1L2*, *G2B2L2*, *G2B3L2*, *G2B4L2*, *G3B1L2*, *G3B2L2*, *G3B3L2*, *G3B4L2*.

Each of the sets: *G1B1L1*, *G1B2L1*, *G1B3L1*, *G1B4L1*, *G2B1L1*, *G2B2L1*, *G2B3L1*, *G2B4L1*, *G3B1L1*, *G3B2L1*, *G3B3L1*, *G3B4L1*, *G1B1L2*, *G1B2L2*, *G1B3L2*, *G1B4L2*, *G2B1L2*, *G2B2L2*, *G2B3L2*, *G2B4L2*, *G3B1L2*, *G3B2L2*, *G3B3L2*, *G3B4L2* was divided into training, testing and validation sets in 7:1:2 ratio. For all the subsets: *G1* to *G1G2G3/L1L2* they are divided into training, testing and validation sets as well. As an example let us consider a subset *G2G3* - its training set is a combination of training sets derived from: *G1B1L1*, *G1B1L2*, *G1B2L1*, *G1B2L2*, *G1B3L1*, *G1B3L2*, *G1B4L1*, *G1B4L2*, *G3B1L1*, *G3B1L2*, *G3B2L1*, *G3B2L2*, *G3B3L1*, *G3B3L2*, *G3B4L1*, *G3B4L2*, the testing and validation sets are created in a similar approach.

The classification process applying ResNet152v2 was performed on all combinations of training and testing sets. The results are presented in Table 2. The columns represent sets on which the model was trained (and validated) and the rows represent sets on which the model was tested. It is important to note, that images in testing and training sets are not the same images - they were prepared in the same conditions but then they were divided into separate training, testing and validation sets. On diagonal of Table 2 there are the combinations where the model was trained and tested on corresponding sets. The obtained accuracy results here are between 0.95 and 1. However, when a model is trained on one set and tested on a set prepared in different condition the classification results are much worse. As an example let us consider a training set *G1*. The accuracy results when tested on *G1* testing set it yields 1. When the accuracy is computed on *G2* and *G3* datasets the results are: 0.39 and 0.56, respectively. These results show that the model did learn to distinguish between 4 bacteria genera, if the choice was random the result would be 0.25 and both 0.39 and 0.56 are greater than that value, although this result is much worse when comparing to training and testing on datasets derived from *G1*. Another example that presents this phenomena is: the model trained on *G2G3* dataset and tested on it corresponding testing set the accuracy was equal to 0.98. When model trained on *G2G3* was tested on *G1* the obtained accuracy also is relatively low and equals 0.53.

Table 2
The classification results (measured with ACC) on whole images applying ResNet152v2 on all combinations of training and testing subsets of images. *G1*, *G2* and *G3* stand for different glass types, *L1* and *L2* stand for different lighting conditions.

test\train	*G1*	*G2*	*G3*	*G1G2*	*G2G3*	*G1G3*	*L1*	*L2*	*G1G2G3*
G1	1.00	0.22	0.54	0.97	0.53	1.00	0.91	0.94	1.00
G2	0.39	0.95	0.38	0.98	0.98	0.47	0.86	0.92	0.98
G3	0.56	0.44	1.00	0.50	0.98	1.00	0.99	0.95	0.99
G1G2	0.71	0.57	0.46	0.98	0.75	0.75	0.88	0.93	0.99
G2G3	0.47	0.69	0.70	0.73	0.98	0.74	0.93	0.94	0.99
G1G3	0.78	0.33	0.77	0.74	0.75	1.00	0.95	0.95	1.00
L1	0.64	0.50	0.67	0.79	0.84	0.84	0.95	0.88	0.99
L2	0.67	0.55	0.62	0.85	0.82	0.83	0.89	0.99	0.99
G1G2G3	0.66	0.53	0.64	0.82	0.83	0.83	0.92	0.94	0.99

Table 3
The classification results (measured with ACC) on subimages derived from images applying ResNet152v2 on all combinations of training and testing subsets of images. *G1*, *G2* and *G3* stand for different glass types, *L1* and *L2* stand for different lighting conditions.

test\train	*G1*	*G2*	*G3*	*G1G2*	*G2G3*	*G1G3*	L1	*L2*	*G1G2G3*
G1	0.95	0.24	0.51	0.91	0.44	0.94	0.82	0.81	0.90
G2	0.28	0.93	0.45	0.91	0.92	0.47	0.83	0.87	0.92
G3	0.31	0.63	0.97	0.52	0.94	0.95	0.91	0.87	0.93
G1G2	0.62	0.58	0.48	0.91	0.68	0.71	0.83	0.84	0.91
G2G3	0.29	0.77	0.72	0.71	0.93	0.72	0.87	0.87	0.93
G1G3	0.61	0.44	0.75	0.70	0.70	0.94	0.87	0.84	0.91
L1	0.56	0.61	0.70	0.82	0.77	0.81	0.90	0.78	0.90
L2	0.45	0.59	0.61	0.72	0.78	0.77	0.81	0.92	0.93
G1G2G3	0.51	0.60	0.65	0.77	0.77	0.79	0.86	0.85	0.92

When the model was trained on one set and its corresponding testing set was a part of another set then the accuracy results were in most cases equal to around 0.6–0.85. For example when the model was trained on *G1* and tested on *G1G2*, *G1G3*, *L1*, *L2* and *G1G2G3* the results were equal to: 0.71, 0.78, 0.64, 0.67, 0.66, respectively.

It is also important to note that when the model was trained on *L*1 dataset and tested on *L*2 dataset the results are equal to 0.89 and when the training test is derived from *L*2 and model is tested on *L*1 the classification result is 0.88. These are the highest results observed when the test set includes lighting and glass conditions that differ from, or are not present in, the training set.

4.3. Classification of subimages

In further part of the research from each of the images their subimages of size 224 × 224 (non-overlapping patches) have been extracted. Patches containing only the background, without any bacteria instances, were removed based on a threshold of less than 5 for the standard deviation of the grayscale-converted image's color values. This approach subimages falling below this threshold were excluded. The ResNet152v2 model was trained on subimages derived from subsets: *G1*, *G2*, *G3*, *G1G2*, *G1G3*, *G2G3*, *L1*, *L2* and *G1G2G3/L1L2*. The subimages were divided into training, testing and validation sets in 7:1:2 ratio. The subimages derived from a single image were assigned exclusively to one of three sets: training, testing or validation. Subimages from the same original image were not distributed across multiple sets. The model was trained on selected training set and then the subimages in the selected testing set were classified. The classification results of the patches are presented in Table 3.

4.4. Classification of subimages with majority voting applied

In the next step it was decided to apply majority voting on patches classified with ResNet152v2. Majority voting enhances robustness by compensating for variability and noise in individual patches through aggregated predictions. While subimages often capture only partial bacteria colony structures - leading to ambiguous or less informative features - the majority voting mechanism mitigates this issue by averaging predictions and reducing the impact of misclassifications. The image which contained the specified patches was classified based on to which classes its subimages were assigned. The Table 4 presents the increase of accuracy applying this approach comparing with accuracy obtained on subimages without majority voting applied and accuracy obtained on whole images, for all the combinations of considered subsets. Comparing the ACC results between whole images and subimages with majority voting the classification result increased in 35 out of 81 combinations. The mean increase among these 35 combinations is equal to 4 percentage points. The mean decrease among the remaining 46 combinations is equal to 4 percentage points. However, when the results obtained by applying majority voting or not on subimages are compared, an increase in ACC was observed in 70 out of 81 combinations when the MV approach was applied. The mean increase among these 70 combinations is equal to 5 percentage points. The mean decrease among the remaining 11 combinations is equal to 2 percentage points.

4.5. Classification of two different samples prepared under same glass type and lighting condition

To assess the impact of light and glass type, a new dataset was prepared under conditions similar to those used in the previous experiments. The initial samples, which consisted of combinations of glass type and bacteria genera - *G1B1*, *G1B2*, *G1B3*, *G1B4*, *G2B1*, *G2B2*, *G2B3* and *G2B4* - were designated as set of samples 1 (*S1*) and labeled as *G1B1S1*, *G1B2S1*, *G1B3S1*, *G1B4S1*, *G2B1S1*, *G2B2S1*, *G2B3S1* and *G2B4S1*. A subsequent dataset was created by preparing a second set of samples (*S2*) under the same glass conditions, and labeled as *G1B1S2*, *G1B2S2*, *G1B3S2*, *G1B4S2*, *G2B1S2*, *G2B2S2*, *G2B3S2*, *G2B4S2*. As these samples were manually prepared, the swabs used for each set differ. Then these new samples *S2* were also photographed under two different lighting conditions *L1* and *L2*. Each of subsets prepared in different conditions contained approximately 150 images. Then the set was divided into following subsets:

1. *G1S1 - G1B1S1L1*, *G1B2S1L1*, *G1B3S1L1*, *G1B4S1L1*, *G1B1S1L2*, *G1B2S1L2*, *G1B3S1L2*, *G1B4S1L2*.
2. *G1S2 - G1B1S2L1*, *G1B2S2L1 G1B3S2L1*, *G1B4S2L1*, *G1B1S2L2*, *G1B2S2L2*, *G1B3S2L2*, *G1B4S2L2*.
3. *G2S1 - G2B1S1L1*, *G2B2S1L1*, *G2B3S1L1*, *G2B4S1L1*, *G2B1S1L2*, *G2B2S1L2*, *G2B3S1L2*, *G2B4S1L2*.
4. *G2S2 - G2B1S2L1*, *G2B2S2L1*, *G2B3S2L1*, *G2B4S2L1*, *G2B1S2L2*, *G2B2S2L2*, *G2B3S2L2*, *G2B4S2L2*.

Table 4
The classification results (measured with ACC) on subimages with majority voting compared to the result obtained on subimages without majority voting applied - *s* and compared to results obtained on whole images *w*; applying ResNet152v2 on all combinations of training and testing subsets of images. *G1*, *G2* and *G3* stand for different glass types, *L1* and *L2* stand for different lighting conditions.

	G1		*G2*		*G3*		*G1G2*		*G2G3*		*G1G3*		*L1*		*L2*		*G1G2G3*	
test\train	*s*	*w*	*s*	*w*	*s*	*w*	*s*	*w*	*s*	*w*	*s*	*w*	*s*	*w*	*s*	*w*	*s*	*w*
G1	0.05	0.00	-0.02	-0.01	0.00	-0.03	0.09	0.02	-0.02	-0.11	0.05	-0.01	0.09	0.01	0.04	-0.10	0.08	-0.03
G2	0.00	-0.10	0.07	0.05	-0.05	0.02	0.04	-0.03	0.07	0.01	0.04	0.03	0.11	0.08	0.07	0.02	0.07	0.01
G3	0.06	-0.19	-0.01	0.18	0.03	0.00	0.05	0.08	0.05	0.01	0.04	-0.01	0.07	-0.01	0.04	-0.05	0.05	-0.01
G1G2	0.04	-0.05	0.01	0.02	-0.02	0.00	0.06	0.00	0.01	-0.06	0.05	0.01	0.10	0.04	0.05	-0.04	0.07	-0.01
G2G3	0.03	-0.15	0.03	0.11	-0.01	0.01	0.05	0.03	0.06	0.01	0.03	0.01	0.09	0.04	0.06	-0.01	0.06	0.00
G1G3	0.08	-0.09	-0.03	0.08	0.00	-0.01	0.08	0.05	0.00	-0.05	0.05	-0.01	0.08	0.00	0.04	-0.07	0.06	-0.02
L1	0.07	-0.01	-0.01	0.10	-0.04	-0.01	0.07	0.10	0.03	-0.04	0.03	0.01	0.09	0.04	0.03	-0.08	0.07	-0.02
L2	0.04	-0.18	0.00	0.04	0.01	0.00	0.07	-0.05	0.01	-0.03	0.06	0.00	0.09	0.02	0.06	-0.01	0.06	0.00
G1G2G3	0.05	-0.10	0.00	0.07	-0.01	0.00	0.07	0.02	0.02	-0.03	0.05	0.01	0.09	0.03	0.05	-0.04	0.07	-0.01

Table 5
The classification results (measured with ACC) on subimages with majority voting applying ResNet152v2 on combinations of training and testing subsets of images. *G1* and *G2* stand for different glass types, *S1* and *S2* stand for two different samples which come from separate swabs.

test\train	*G1S1*	*G1S2*	*G2S1*	*G2S2*	*G1S1 G1S2*	*G2S1 G2S2*
G1S1	1.00	0.19	0.22	0.38	1.00	0.55
G1S2	0.21	0.90	0.39	0.25	0.95	0.29
G2S1	0.29	0.56	1.00	0.44	0.31	0.97
G2S2	0.42	0.36	0.55	0.98	0.49	0.95
G1S1 G1S2	0.67	0.50	0.29	0.33	0.98	0.44
G2S1 G2S2	0.35	0.46	0.76	0.72	0.41	0.96

5. *G1S1 G1S2* - *G1B1S1L1*, *G1B2S1L1*, *G1B3S1L1*, *G1B4S1L1*, *G1B1S1L2*, *G1B2S1L2*, *G1B3S1L2*, *G1B4S1L2*, *G1B1S2L1*, *G1B2S2L1*, *G1B3S2L1*, *G1B4S2L1*, *G1B1S2L2*, *G1B2S2L2*, *G1B3S2L2*, *G1B4S2L2*.
6. *G2S1 G2S2* - *G2B1S1L1*, *G2B2S1L1*, *G2B3S1L1*, *G2B4S1L1*, *G2B1S1L2*, *G2B2S1L2*, *G2B3S1L2*, *G2B4S1L2*, *G2B1S2L1*, *G2B2S2L1*, *G2B3S2L1*, *G2B4S2L1*, *G2B1S2L2*, *G2B2S2L2*, *G2B3S2L2*, *G2B4S2L2*.

Subimages were then extracted from each set in the same manner as previously described, and majority voting was applied to the results obtained using ResNet152v2. The results, presented in Table 5, show that when the dataset is prepared under identical lighting conditions and with the same type of glass but with different samples, the accuracy drops significantly, reaching as low as 0.19. However, high accuracy (between 0.9–1) is achieved when the model is trained and tested on corresponding training and testing sets derived from the same sample (swab).

5. Discussion and conclusions

The classification task presented here is challenging. The images were not curated by microbiologists, who would typically ensure high-quality image acquisition. Instead, this research focused on analyzing an automatically generated dataset to assess the feasibility of applying the classification algorithm in a production environment, where there is limited time for careful image preparation and some images may not be perfectly clear.

Despite the challenges, the classification algorithms achieved high accuracy. The dataset used for analysis consisted of color and texture features. Classical classification methods, such as the Random Forest algorithm, yielded an accuracy of up to 0.89. Among the tested ELM-RBF methods, the best results were achieved with the ELM-RBF combined with the k-medoids clustering method, which reached an accuracy of 0.96 (see Table 1). However, the highest accuracy was achieved using the deep learning method, ResNet152v2, which delivered a remarkable 0.99 accuracy. It is important to note that, unlike traditional methods, ResNet takes entire images as input rather than feature vectors and generates its own features automatically.

High accuracy results were initially obtained on the entire dataset. Following this, the analysis focused on evaluating the accuracy when the model was trained and tested on different subsets of images, each prepared under varying glass and lighting conditions. The ResNet152v2 model was used for this analysis. The accuracy decreased when the model was tested on sets created under different conditions (see Table 2).

The images were divided into subimages, which were then used as input to the Residual Network (see Table 3). It was then decided to apply a majority voting approach to determine if it would improve the classification results. The results showed that, in most cases, there was an increase in accuracy compared to the results obtained from individual subimages (see Table 4). This suggests that the majority voting approach can be successfully applied when considering multiple subimages of a selected area, particularly for classifying the bacteria genera present. However, when comparing this to the results from using whole images, the accuracy increased in almost half of the tested scenarios (see Table 4).

To ensure consistency in the results when repeating the experiment, a new dataset was prepared with consistent lighting and glass conditions. The experiment used ResNet, with the model trained on images derived from one sample under these same conditions and tested on another set of images from a different sample, also under the same conditions. Surprisingly, the accuracy results were much worse, despite the lighting and glass conditions being identical (see Table 5). Furthermore, higher accuracy results were sometimes achieved when a model trained on images from one sample with one type of glass was tested on a different type of glass, compared to when it was tested on the same glass type and different sample.

This research suggests that there may be other variables influencing the microscopic images. Based on the experiments, it can be inferred that the preparation of the swab from each sample has a significant impact. Since the swabs were manually prepared, this could lead to variations in factors such as the shape of the bacteria colonies and the density of bacteria (bacteria layer thickness).

An important aspect to note is how the sets *L1* and *L2* were created. While the swab was the same (as it came from the same sample), different areas of the sample were photographed under two distinct lighting conditions. When the model is trained and tested on *L1*, the accuracy is 0.95, and when trained and tested on *L2*, the accuracy is 0.99. However, when the model is trained on *L1* and tested on *L2*, the accuracy does not decrease significantly, resulting in 0.89. Similarly, when trained on *L2* and tested on *L1*, the accuracy is 0.88 (see Table 1). This clearly demonstrates the impact of lighting conditions, as the swabs themselves remained unchanged.

The glasses cannot be directly compared with each other, as each glass contains a different swab. It is not possible to simply 'copy and paste' the same swab from one glass to another for comparison. To enable meaningful comparisons between glasses, an automated and standarized approach for creating swabs would be necessary.

Classification results can be very high when tested on a testing set prepared under the same conditions. However, even a slight change - such as a different swab, lighting or glass - can significantly affect the accuracy. In cases where the images are perfectly prepared by microbiologists, ensuring each image is accurately focused on bacteria instances, this issue is not observed. However, creating a large set of perfectly prepared images is time-consuming and not practical for larger-scale analysis in a production environment.

This research highlights the need for automating the swab preparation process to ensure reproducibility. Developing such a procedure would inevitably require close collaboration among microbiologists, information technology specialists and mechatronics engineers. Successfully addressing this challenge could enable the implementation of automated microorganism classification systems in biofertilizer production, enhancing real-time monitoring and quality control throughout the manufacturing process.

Despite the proficiency of machine learning methods in accurately distinguishing between bacteria classes, they encounter significant challenges when classifying datasets prepared under varying conditions, particularly when the model is trained on one set of conditions and tested on another.

Importantly, this work highlights a more general concern - the necessity of careful and consistent dataset preparation and thorough validation when deploying artificial intelligence in the analysis of microscopic images across various biological domains. The issues uncovered in this study go beyond bacteria classification and reflect fundamental challenges in using artificial intelligence for microscopy-driven research in broader areas of the life sciences.

CRediT authorship contribution statement

Aleksandra Konopka: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Ryszard Kozera:** Writing – review & editing, Supervision, Formal analysis. **Lidia Sas-Paszt:** Writing – review & editing, Supervision, Conceptualization. **Paweł Trzciński:** Writing – review & editing, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We extend our heartfelt gratitude to Prof. Agnieszka Marasek-Ciołakowska of the National Institute of Horticultural Research in Skierniewice for her insightful guidance and invaluable advice on microscope imaging.

Data availability

Data will be made available on request.

References

Aggarwal, A., Singh, P., Gupta, S., 2024. An analysis of case study on application of backpropagation network using procedural: Object-oriented approach. In: 2024 International Conference on Emerging Innovations and Advanced Computing. INNOCOMP, IEEE, pp. 525–531. http://dx.doi.org/10.1109/innocomp63224.2024.00092.

Aishwarya, D., Minu, R., 2021. Convolutional neural network in computer vision. In: Applied Learning Algorithms for Intelligent IoT. Auerbach Publications, pp. 1–32. http://dx.doi.org/10.1201/9781003119838-1.

Alsmadi, M.k., Omar, K.B., Noah, S.A., Almarashdah, I., 2009. Performance comparison of multi-layer perceptron (back propagation, delta rule and perceptron) algorithms in neural networks. In: 2009 IEEE International Advance Computing Conference. IEEE, pp. 296–299. http://dx.doi.org/10.1109/iadcc.2009.4809024.

Bal-Ghaoui, M., El Yousfi Alaoui, M.H., Jilbab, A., Bourouhou, A., 2023. Optimizing ultrasound image classification through transfer learning: fine-tuning strategies and classifier impact on pre-trained inner-layers. Inform. Control. Meas. Econ. Environ. Prot. 13 (4), 27–33. http://dx.doi.org/10.35784/iapgos.4464.

Bonnichsen, L., Bygvraa Svenningsen, N., Haubjerg Nicolaisen, M., Nybroe, O., 2019. Methods to determine bacterial abundance, localization, and general metabolic activity in soil. In: Modern Soil Microbiology. CRC Press, pp. 195–213. http://dx.doi.org/10.1201/9780429059186-12.

Cao, J., Lin, Z., 2015. Extreme Learning Machines on high dimensional and large data applications: A survey. Math. Probl. Eng. 2015, 1–13. http://dx.doi.org/10.1155/2015/103796.

Deng, T., 2022. A survey of convolutional neural networks for image classification: Models and datasets. In: 2022 International Conference on Big Data, Information and Computer Network. BDICN, IEEE, pp. 746–749. http://dx.doi.org/10.1109/bdicn55575.2022.00145.

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei, L.F., 2009. ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, http://dx.doi.org/10.1109/cvpr.2009.5206848.

Eady, M., Park, B., 2016. Rapid identification of salmonella serotypes through hyperspectral microscopy with different lighting sources. J. Spectr. Imaging http://dx.doi.org/10.1255/jsi.2016.a4.

Ebrahimi, M.S., Abadi, H.K., 2021. Study of Residual Networks for image recognition. In: Intelligent Computing. Springer International Publishing, pp. 754–763. http://dx.doi.org/10.1007/978-3-030-80126-7_53.

Fampa, M., Lee, J., Ponte, G., Xu, L., 2021. Experimental analysis of local searches for sparse reflexive generalized inverses. J. Global Optim. 81 (4), 1057–1093. http://dx.doi.org/10.1007/s10898-021-01087-y.

Ferrari, B.C., Tujula, N., Stoner, K., Kjelleberg, S., 2006. Catalyzed reporter deposition-fluorescence in situ hybridization allows for enrichment-independent detection of microcolony-forming soil bacteria. Appl. Environ. Microbiol. 72 (1), 918–922. http://dx.doi.org/10.1128/aem.72.1.918-922.2006.

Fukunaga, K., Hostetler, L., 1975. The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Trans. Inform. Theory 21 (1), 32–40. http://dx.doi.org/10.1109/tit.1975.1055330.

He, K., Zhang, X., Ren, S., Sun, J., 2016a. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, IEEE, pp. 770–778. http://dx.doi.org/10.1109/cvpr.2016.90.

He, K., Zhang, X., Ren, S., Sun, J., 2016b. Identity mappings in deep residual networks. http://dx.doi.org/10.48550/ARXIV.1603.05027.

Huang, G., Huang, G.B., Song, S., You, K., 2015. Trends in extreme learning machines: a review. Neural Netw. 61, 32–48. http://dx.doi.org/10.1016/j.neunet.2014.10.001.

Huang, G.B., Zhu, Q.Y., Siew, C.K., 2004. Extreme learning machine: a new learning scheme of feedforward neural networks. In: 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541). IJCNN-04, 2, IEEE, pp. 985–990. http://dx.doi.org/10.1109/ijcnn.2004.1380068.

Ismaeel, S., Miri, A., Chourishi, D., 2015. Using the extreme learning machine (ELM) technique for heart disease diagnosis. In: 2015 IEEE Canada International Humanitarian Technology Conference (IHTC2015). IEEE, pp. 1–3. http://dx.doi.org/10.1109/ihtc.2015.7238043.

Kaufman, L., Rousseeuw, P., 1990. Finding groups in data: An introduction to cluster analysis. Wiley, New York. ISBN 0-471-87876-6. http://dx.doi.org/10.2307/2532178.

Khmelevtsova, L.E., Sazykin, I.S., Azhogina, T.N., Sazykina, M.A., 2022. Influence of agricultural practices on bacterial community of cultivated soils. Agriculture 12 (3), 371. http://dx.doi.org/10.3390/agriculture12030371.

Konopka, A., Kozera, R., Sas-Paszt, L., Trzcinski, P., Lisek, A., 2023a. Identification of the selected soil bacteria genera based on their geometric and dispersion features. PLoS One 18 (10), e0293362. http://dx.doi.org/10.1371/journal.pone.0293362.

Konopka, A., Struniawski, K., Kozera, R., 2023b. Classification performance of Extreme Learning Machine Radial Basis Function with K-means, K-medoids and mean shift clustering algorithms. In: International Conference on Computational Science – ICCS 2023. Springer Nature Switzerland, pp. 171–186. http://dx.doi.org/10.1007/978-3-031-36027-5_13.

Konopka, A., Struniawski, K., Kozera, R., 2023c. Performance analysis of residual neural networks in soil bacteria microscopic image classification. In: 37th Annual European Simulation and Modelling Conference. ESM 2023, pp. 144–148.

Konopka, A., Struniawski, K., Kozera, R., Trzciński, P., Sas-Paszt, L., Lisek, A., Górnik, K., Derkowska, E., Głuszek, S., Sumorok, B., Frąc, M., 2022. Classification of soil bacteria based on machine learning and image processing. In: International Conference on Computational Science – ICCS 2022. Springer International Publishing, pp. 263–277. http://dx.doi.org/10.1007/978-3-031-08757-8_23.

Lalitha, S., 2017. Plant growth–promoting microbes: A boon for sustainable agriculture. In: Sustainable Agriculture Towards Food Security. Springer Singapore, pp. 125–158. http://dx.doi.org/10.1007/978-981-10-6647-4_8.

Luo, Y., 2022. Deep learning evolution from shallow to deep: a case study based on language recognition. In: 2nd International Conference on Artificial Intelligence, Automation, and High-Performance Computing. AIAHPC 2022, SPIE, p. 85. http://dx.doi.org/10.1117/12.2641495.

McCulloch, W.S., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. 5 (4), 115–133. http://dx.doi.org/10.1007/bf02478259.

Mojrian, S., Pinter, G., Joloudari, J.H., Felde, I., Szabo-Gali, A., Nadai, L., Mosavi, A., 2020. Hybrid Machine Learning Model of Extreme Learning Machine Radial Basis Function for breast cancer detection and diagnosis; a multilayer fuzzy expert system. In: 2020 RIVF International Conference on Computing and Communication Technologies. RIVF, IEEE, pp. 1–7. http://dx.doi.org/10.1109/rivf48685.2020.9140744.

Nikzad, M., Gao, Y., Zhou, J., 2019. Gradient-based pooling for convolutional neural networks. In: 2019 IEEE Visual Communications and Image Processing. VCIP, IEEE, pp. 1–4. http://dx.doi.org/10.1109/vcip47243.2019.8966049.

Padmaja, D.L., Nikhil, B., Akshaya, B.S., Surya Deepak, G., 2024. A comparative analysis of ResNet and MobileNet for classifying MRI images. In: Proceedings of the 5th International Conference on Data Science, Machine Learning and Applications; Volume 1. Springer Nature Singapore, pp. 22–30. http://dx.doi.org/10.1007/978-981-97-8031-0_3.

Pande, S., Rathore, N.K., Purohit, A., 2024. Unveiling the impact of extreme learning machine in the defence and military sector. Def. Sci. J. 74 (4), 517–525. http://dx.doi.org/10.14429/dsj.74.19174.

Peerthum, Y., Stamp, M., 2023. An empirical analysis of the shift and scale parameters in BatchNorm. Inform. Sci. 637, 118951. http://dx.doi.org/10.1016/j.ins.2023.118951.

Pérez-Ortega, J., Nely Almanza-Ortega, N., Vega-Villalobos, A., Pazos-Rangel, R., Zavala-Díaz, C., Martínez-Rebollar, A., 2020. The K-means algorithm evolution. In: Introduction To Data Science and Machine Learning. IntechOpen, http://dx.doi.org/10.5772/intechopen.85447.

Rosenblatt, F., 1958. The perceptron: A probabilistic model for information storage and organization in the brain. Psychol Rev 65 (6), 386–408. http://dx.doi.org/10.1037/h0042519.

Sabba, S., Smara, M., Benhacine, M., Terra, L., Eddine Terra, Z., 2022. Residual neural network in genomics. Comput. Sci. J. Moldova 30 (3(90)), 308–334. http://dx.doi.org/10.56415/csjm.v30.17.

Safaei, A., Wu, Q.M.J., Akilan, T., Yang, Y., 2019. System-on-a-Chip (SoC)-based hardware acceleration for an online sequential extreme learning machine (OS-ELM). IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst. 38 (11), 2127–2138. http://dx.doi.org/10.1109/tcad.2018.2878162.

Struniawski, K., Konopka, A., Kozera, R., 2022. Identification of soil bacteria with machine learning and image processing techniques applying single cells' region isolation. In: The European Simulation and Modelling Conference 2022. EUROSIS-ETI, pp. 76–81.

Sýkora, P., Hlavata, R., Kamencay, P., Benco, M., Hudec, R., 2023. Body part classification from medical single channel images using deep neural networks. In: 2023 International Symposium ELMAR. IEEE, pp. 79–82. http://dx.doi.org/10.1109/elmar59410.2023.10253938.

3.6 Deep learning classification of blackcurrant genotypes by ploidy levels on stomata microscopic images

Publication: A. Konopka, K. Struniawski, R. Kozera, L. Ortenzi, A. Marasek-Ciołakowska, and A. Machlańska, "Deep learning classification of blackcurrant genotypes by ploidy levels on stomata microscopic images," in *Computational Science – ICCS 2025 Workshops: 25th International Conference, Singapore, Singapore, July 7–9, 2025, Proceedings, Part III, Lecture Notes in Artificial Intelligence*, pp. 135–148, Springer, 2025. https://doi.org/10.1007/978-3-031-97564-6_11 (140 MNiSW points)

Abstract: Polyploidy, the variation in chromosome sets within plants, influences stomata size and density, making stomata analysis a valuable method for determining ploidy levels. Traditional microscopic analysis, nevertheless, is often labor-intensive and complex. This study explores the use of artificial intelligence for the automated classification of plant ploidy levels from stomata images, presenting a novel approach in this field. Experiments were conducted on three blackcurrant genotypes: diploid, triploid and tetraploid. Deep learning techniques were employed for stomata segmentation and classification, with performance compared to traditional machine learning algorithms, including K-Nearest Neighbors, Support Vector Machine, Random Forest and Multi-Layer Perceptron. To mitigate the impact of color variations that could lead to inflated accuracy, multiple datasets were processed to reduce the influence of color. Classification was performed not only on whole images but also on subimages containing individual stomata instances, detected using the YOLOv8 algorithm. A majority voting approach was applied to classify the entire image based on subimage classifications. ResNet152V2 achieved the highest accuracy of 0.973 on color images, although accuracy declined when the influence of color was minimized. These results underscore the significant role of color in model performance and highlight the challenges associated with achieving reliable and robust classification.

Deep Learning Classification of Blackcurrant Genotypes by Ploidy Levels on Stomata Microscopic Images

Aleksandra Konopka[1(✉)], Karol Struniawski[1], Ryszard Kozera[1,2], Luciano Ortenzi[3], Agnieszka Marasek-Ciołakowska[4], and Aleksandra Machlańska[4]

[1] Institute of Information Technology, Warsaw University of Life Sciences - SGGW, ul. Nowoursynowska 159, 02-776 Warsaw, Poland
{aleksandra_konopka,karol_struniawski,ryszard_kozera}@sggw.edu.pl
[2] School of Physics, Mathematics and Computing, The University of Western Australia, 35 Stirling Highway, Crawley, Perth, WA 6009, Australia
ryszard.kozera@uwa.edu.au
[3] Department of Computer Science, Sapienza University of Rome, Viale Regina Elena, 295, 00161, Rome, Italy
luciano.ortenzi@unitus.it
[4] Department of Applied Biology, The National Institute of Horticultural Research, Konstytucji 3 Maja 1/3, 96-100 Skierniewice, Poland
{agnieszka.marasek,aleksandra.machlanska}@inhort.pl

Abstract. Polyploidy, the variation in chromosome sets within plants, influences stomata size and density, making stomata analysis a valuable method for determining ploidy levels. Traditional microscopic analysis, nevertheless, is often labor-intensive and complex. This study explores the use of artificial intelligence for the automated classification of plant ploidy levels from stomata images, presenting a novel approach in this field. Experiments were conducted on three blackcurrant genotypes: diploid, triploid and tetraploid. Deep learning techniques were employed for stomata segmentation and classification, with performance compared to traditional machine learning algorithms, including K-Nearest Neighbors, Support Vector Machine, Random Forest and Multi-Layer Perceptron. To mitigate the impact of color variations that could lead to inflated accuracy, multiple datasets were processed to reduce the influence of color. Classification was performed not only on whole images but also on subimages containing individual stomata instances, detected using the YOLOv8 algorithm. A majority voting approach was applied to classify the entire image based on subimage classifications. ResNet152v2 achieved the highest accuracy of 0.973 on color images, although accuracy declined when the influence of color was minimized. These results underscore the significant role of color in model performance and highlight the challenges associated with achieving reliable and robust classification.

Keywords: stomata · microscopic image classification · machine learning · deep learning · Residual Neural Networks · YOLO algorithm

M. Paszynski et al. (Eds.): ICCS 2025 Workshops, LNCS 15909, pp. 135–148, 2025.
https://doi.org/10.1007/978-3-031-97564-6_11

1 Introduction

Stomata are microscopic pores found in the epidermal layers of plant leaves and other aerial tissues [21]. These pores are regulated by two specialized cells that control their opening and closing. The stomata play a crucial role in regulating the exchange of gases, including CO_2 and in controlling factors such as light intensity and humidity [5]. These factors can influence the visual characteristics of stomata. Additionally, stomata traits may also be affected by the application of biofertilizers, making the analysis of stomata cells a valuable method for assessing the effectiveness of biofertilizers [2,32]. Structurally, stomata form an elliptical shape. Polyploidy, which refers to the number of chromosome sets in a plant, can be determined through the microscopic analysis of stomata characteristics. Both the size and density of stomata are influenced by polyploidy [35]. However, the process of analyzing stomata images to assess polyploidy can be time-consuming and complex. To address this challenge, the application of artificial intelligence (AI) has been proposed in this paper. Although AI is successfully applied to microscopic images of stomata for tasks such as segmentation [19], measurement [27] and counting [9], to the best of our knowledge, it has not yet been used for the automated classification of plant ploidy level based on stomata images.

In this study, experiments were conducted on three classes of blackcurrant genotypes, each distinguished by their ploidy level: diploid, triploid and tetraploid. The research faced several challenges, primarily because stomata characteristics are influenced by environmental factors [6]. Additionally, the color of the plant, as observed in microscopic images, is also affected by environmental conditions, further complicating the analysis. As a result, it is essential to ensure that each ploidy class is represented under similar environmental conditions. One challenge that arose was the variation in image color, which had the potential to artificially inflate accuracy (ACC) rates. To mitigate this, attention was concentrated on the shape of the stomata, and the impact of color was minimized to the greatest extent possible.

Due to subtle variations in brightness and color across the images, multiple datasets were generated as processed versions of the raw dataset, enabling the evaluation of various analytical approaches. The analysis aimed to determine whether better classification results could be obtained by analyzing the entire image, considering all components and details of the leaf structure for each ploidy. Alternatively, it was explored whether focusing solely on the stomata cells would yield better outcomes. Furthermore, to reduce the influence of color on the classification results, several datasets characterized by different color spaces and channel configurations were examined in this paper.

In the latter, deep learning techniques were employed for the segmentation task to accurately identify stomata objects. For the classification task, deep learning models were applied and their performance was compared against traditional machine learning algorithms. Finally, majority voting (MV) [29] was employed on subimages to investigate whether classifying individual stomata objects, followed by MV, could improve the overall classification accuracy.

2 Materials

The experiment was carried out in the Experimental Pomological Orchard in Skierniewice, central Poland, belonging to the National Institute of Horticultural Research. Plants of blackcurrant (*Ribes nigrum* L.) were grown under natural light in 50 L pots filled with a mixture of peat substrate and soil at a ratio of 1:1. Water potential in the growing medium was maintained at a level (−) 10 kPa. The moisture content of the growing medium were monitored with dielectric probes (TEROS-12, METER, USA). Plants were maintained according to standard agrotechnical measures for the blackcurrant. Leaves were sampled for photo documentation at the end of a vegetation season from the 5th to the 25th of September 2024.

The dataset utilized in this research comprises three classes of blackcurrant genotypes distinguished by their chromosome counts: diploid ($2n = 2x = 16$) cultivar (class 0), triploid ($2n = 3x = 24$) cultivar (class 1) and tetraploid ($2n = 4x = 32$) clone (class 2) (see Fig. 1). Class 0 includes 252 images of the 'Gofert' cultivar and 252 images of the 'Polares' cultivar. Class 1 contains 502 images of the triploid 'Dlinnokistnaja' cultivar, while class 2 consists of 251 images of the autotetraploid clone of 'Gofert' cultivar and 250 images of the 'Polares' cultivar obtained by in vitro poliyploidyzation [23]. All images were captured using the VHX-7000N KEYENCE digital microscope, ensuring high-quality imaging suitable for detailed analysis.

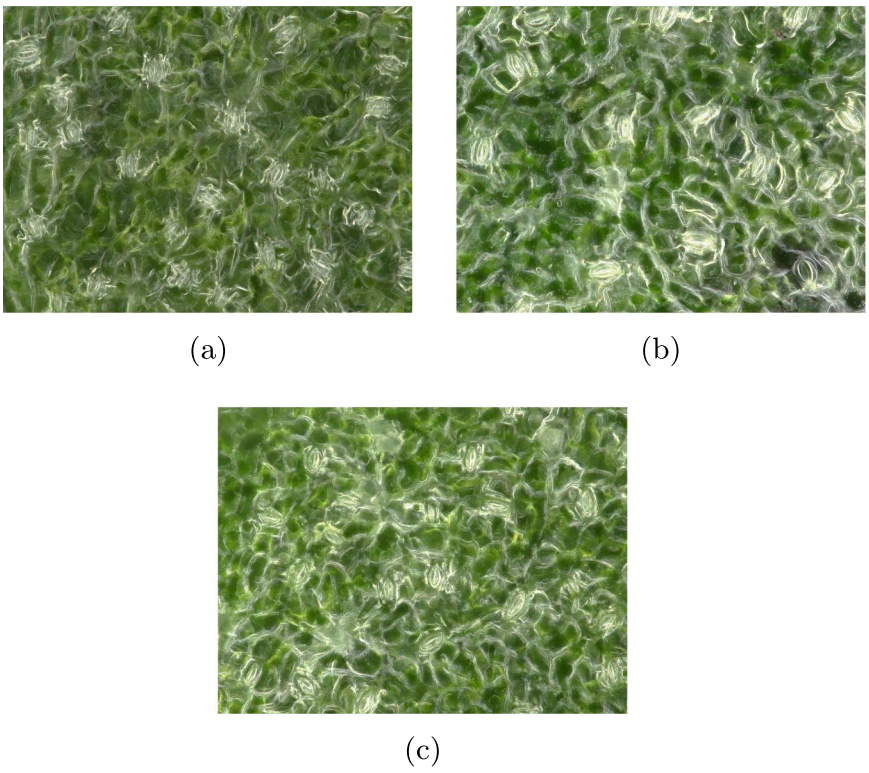


Fig. 1. Exemplary images of diploid (a), triploid (b) and tetraploid (c) genotypes from the raw dataset.

3 Methods

This work focuses on comparing classical artificial intelligence methods with approaches based on deep neural networks. Deep learning techniques were applied in this research to both image classification and segmentation tasks. The following subsections provide a detailed description of the selected deep learning methods.

3.1 Convolutional Neural Networks

This research leverages Convolutional Neural Networks (CNNs) [31] a class of deep learning models applied to tasks related with image processing and computer vision. CNNs, are widely recognized not only for their ability to perform classification, regression, object detection, segmentation task on input data [7] but also for their capacity to automatically extract relevant features. The image, which serves as the input to the CNN, undergoes processing through various network layers that iteratively extract features [10].

The term convolution in CNN refers to the convolution layer, a fundamental component responsible for performing convolution operation [8]. This procedure involves processing the input image with a filter, which is designed to emphasize specific aspects of the image, such as edges or patterns [15]. These filters are learned during training, allowing the network to identify features critical for the task at hand.

Another essential component of CNN is the pooling layer, which serves to distill the most significant information while discarding irrelevant details. The pooling operation reduces the spatial dimensions of the data, resulting in computational efficiency and mitigating resource constraints without sacrificing performance [33].

The convolution and pooling, along with other specialized layers, process the input image, transforming it into a feature vector. This vector serves as the input to the classification stage, typically implemented as a fully connected neural network - commonly referred to as a Multi-Layer Perceptron (MLP) [1]. This final stage performs classification or regression, depending on the specific task under consideration.

Residual Neural Networks (ResNets) [26] are a specialized type of Convolutional Neural Networks created in 2016 [12] that were employed in this research for the classification task. A key architectural feature of ResNets is their ability to learn based on residual errors, utilizing skip connections to mitigate the vanishing gradient problem [28]. These networks have various versions, differing in the number of convolutional blocks and the arrangement of layers [17]. In this study, the ResNet50v2 and ResNet152v2 architectures [13] are considered.

3.2 YOLO

Convolutional Neural Networks form a core component of the YOLO (You Only Look Once) algorithm [24], a highly efficient and versatile deep learning framework. YOLO is widely utilized for tasks such as classification, object detection

and instance segmentation. In this study, YOLO is employed to segment stomata in images.

YOLO is renowned for its speed and capability to detect multiple objects in a single pass [3]. The algorithm processes input data, typically images, by first scaling them to a predefined size. The scaled image is then divided into a grid structure, where each grid cell is responsible for detecting objects whose centers fall within that specific cell [4]. A CNN is then applied to extract features and detect objects within each grid cell.

The output of the CNN provides essential information, including the probability of an object's presence, the x and y coordinates of the object's center, its width and height, and a one-hot encoded vector representing the object's class [24]. Predicted bounding boxes are evaluated using the Intersection over Union metric, which measures the overlap between predefined bounding boxes for the target objects and those predicted by the model [34].

In the post-processing stage, the algorithm refines the detection results. First, the most probable object detections are retained and overlapping bounding boxes identifying the same object are eliminated using the Non-Maximum Suppression algorithm [25]. This step ensures that redundant boxes corresponding to multiple anchors in a single grid cell are removed. The final output of the YOLO algorithm is a list of detected objects.

The YOLO algorithm was originally introduced in 2015 and subsequent versions are developed to enhance its performance.

4 Experiments and Results

The experiments were conducted in multiple stages using both the raw and processed versions of the dataset. Subimages containing individual stomata were extracted for analysis. The primary objective of this study was to evaluate and compare the classification accuracy of deep learning algorithms against traditional machine learning techniques. The latter presents classification outcomes for scenarios where the raw dataset, including color information, is used and other where the influence of color is minimized to focus exclusively on the shape of the stomata objects. The computations were performed using the Python programming language.

4.1 Image Segmentation

To identify stomata cells, the YOLOv8 algorithm was applied. This version of YOLO was chosen because its output is a precise segmentation mask that covers the area of the searched object, rather than a bounding box. For this segmentation task, the algorithm was trained on 70 images and validated on 30 images of tetraploid 'Polares' cultivar, which were manually segmented. The dataset used for training included images of blackcurrant subjected to varying water conditions (50 images under high irrigation and 50 images under low irrigation). This

variation in water availability was introduced to ensure that the YOLO algorithm could learn to detect stomata cells of all sizes, as the size of the stomata changes depending on water availability.

The YOLOv8 model, using the *yolov8n-seg.pt* [14] configuration, was trained for 500 epochs. The image size was set of 2000 × 2000 pixels and a batch size of 5. To reduce computation time, the number of points in the labels used as input to the YOLO algorithm was reduced by a factor of 4 compared to those created by manual segmentation. The output of YOLO algorithm is a list of stomata object labels. Based on the coordinates of labels the stomata instances were identified which allowed creation of binary masks - white area represents stomata objects and black area is the remaining part of the image.

4.2 Preparation of Various Datasets

Due to variations in sample color, computations were performed on multiple datasets to reduce impact of color on classification accuracy. The first analyzed dataset was included raw images curated by biology experts. Additional 9 datasets were derived from this original dataset, incorporating modifications to minimize the influence of sample color. Figure 2 demonstrates a noticeable reduction in color variation across the selected color channels among the three classes. This is demonstrated through a comparative analysis of the mean green component in the RGB and grayscale image with the mean hue (H) component in the HSV color space and the mean value (V) component in the histogram-equalized HSV. The following datasets were considered (see Fig. 3):

1. *Images* - unmodified raw images in the RGB color space.
2. *Grayscale images* - the *Images* converted to grayscale.
3. *HSV images - H* - the hue component of *Images* in the HSV color space.
4. *HSV images - V equalized* - the value component of *Images* in the HSV color space, with histogram equalization applied.
5. *Image masks* - binary masks of images.
6. *Subimages* - containing a single stoma extracted from the original images.
7. *Subimages black background* - *Subimages* with the background set to black (only the stoma object retains color).
8. *Subimages grayscale* - *Subimages* converted to grayscale.
9. *Subimages grayscale black background* - *Subimages* converted to grayscale with the background set to black.
10. *Subimage masks* - binary masks of *Subimages*.

Datasets 1–5 were derived from entire microscopic images, while datasets 6–10 were based on subimages. These subimages, measuring 224×224 pixels, were extracted with a single stoma centrally positioned within each image.

4.3 Classification with Residual Neural Networks

All 10 datasets were used as input to ResNet50v2 and ResNet152v2 models which were configured with the following parameter values. Each dataset was

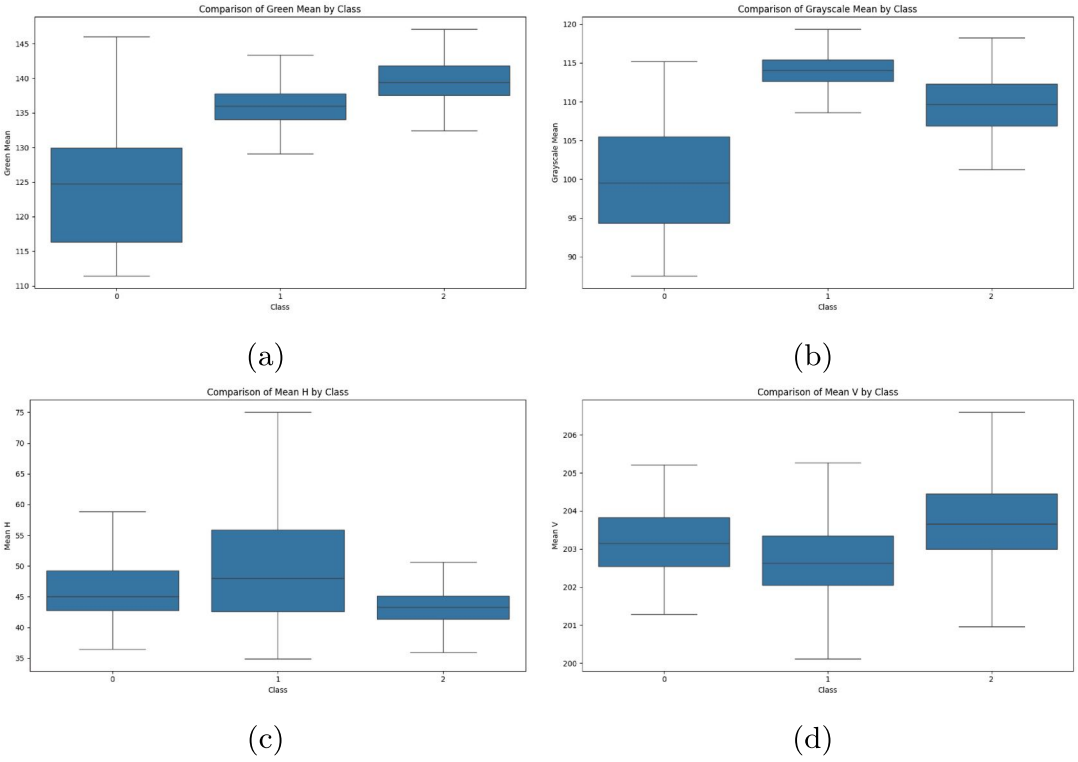


Fig. 2. The boxplots of mean values for three classes (0 - diploid, 1 - triploid, 2 - tetraploid) across selected channels: green from RGB (a), grayscale (b), hue from HSV (c) and equalized value from HSV (d). (Color figure online)

split into training, validation and test sets in a 7:2:1 ratio. The models were fine-tuned from ImageNet [17], starting from the layer *conv5_ block1_ 1_ conv*. The applied pooling method was average pooling. The classification part of the network included the following layers: a dense layer with 512 neurons and ReLU activation function, followed by a dropout layer with a rate of 0.3, another dense layer with 512 neurons and ReLU activation function, an additional dropout layer and a final dense layer with a softmax activation function. The Adam optimizer was used with a learning rate of 0.01. An early stopping mechanism was implemented with a patience of 50 epochs, with a maximum of 1000 epochs and a batch size of 32. Table 1 presents accuracy values ranging from 0.606 for the *HSV images - H* dataset using ResNet50v2 to 0.973 for the *Images* dataset with ResNet152v2. For the *Image masks* dataset, ResNet50v2 achieved an accuracy of 0.787, while ResNet152v2 yielded an accuracy of 0.618 on the *Subimage masks* dataset.

Furthermore, datasets 6–10 were also subjected to a majority voting approach. Each subimage was classified into one of three classes. Subsequently, the image from which the subimage was extracted was assigned to the class that received the majority of classifications from its constituent subimages. The application of MV on subimages achieved a maximum accuracy of 0.96 with ResNet152v2 on the *Subimages* dataset. In contrast, for the *Subimage masks* dataset, MV with ResNet152v2 yielded an accuracy of 0.624 (see Table 1).

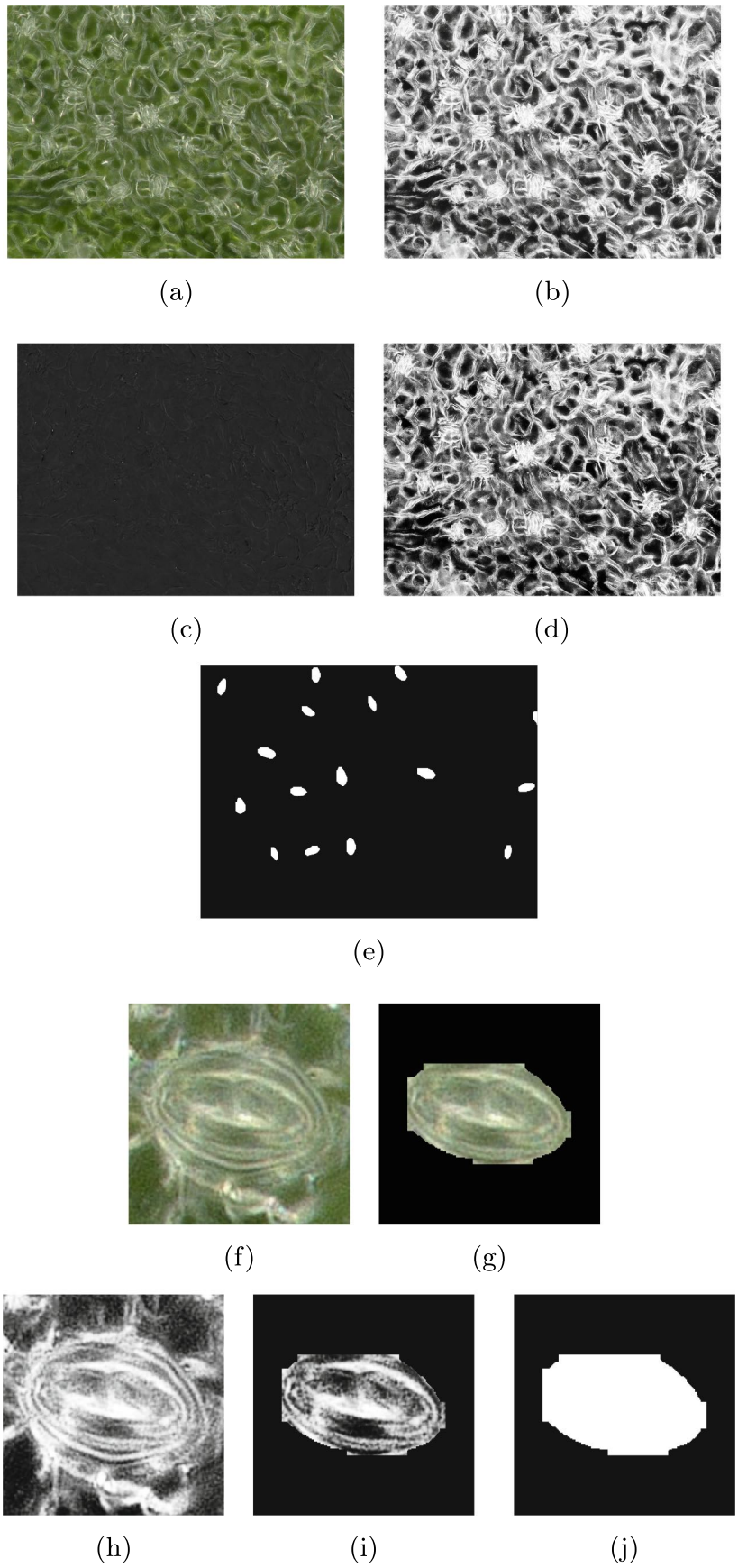


Fig. 3. Exemplary images from datasets 1–10, corresponding to subfigures (a–j), representing class 0 - diploid.

Table 1. Classification accuracy of stomata images on 10 different datasets applying Residual Neural Networks with majority voting applied - *MV* or without majority voting - *no MV*.

	Dataset	ResNet50v2		ResNet152v2	
		no MV	*MV*	*no MV*	*MV*
1	*Images*	0.953	-	0.973	-
2	*Grayscale images*	0.953	-	0.88	-
3	*HSV images - H*	0.606	-	0.636	-
4	*HSV images - V equalized*	0.62	-	0.63	-
5	*Image masks*	0.787	-	0.68	-
6	*Subimages*	0.856	0.913	0.866	0.96
7	*Subimages black background*	0.68	0.765	0.759	0.765
8	*Subimages grayscale*	0.78	0.839	0.815	0.9
9	*Subimages grayscale black background*	0.783	0.859	0.782	0.9
10	*Subimage masks*	0.607	0.577	0.618	0.624

4.4 Classification with Classical Machine Learning Methods

In this study, the performance of ResNets was compared to classical machine learning techniques [16]. The dataset utilized in this comparison consisted of *Image masks*, which represent the shapes of stomata instances without relying on color information. For each image, features were manually extracted for each stomata individually and the mean value of each feature was subsequently calculated. The features included in the dataset were: convex area [30], contour area [30], perimeter, solidity [30], elongation [11], circularity [11], object width and object height.

The classical machine learning algorithms selected for this analysis were K-Nearest Neighbors (KNN) [18], Random Forest (RF) [20], Support Vector Machine (SVM) [22] and Multi-Layer Perceptron [1]. The hyperparameters of these models were tuned to optimize performance. Specifically: for KNN, the number of neighbors k was varied from 1 to 10; in RF, the number of trees was tested from 100 to 500; the regularization parameter C in SVM was evaluated over the range from 0.8 to 2.6. The architecture of the MLP was explored using configurations with three hidden layers, ranging from $10 \times 10 \times 10$ neurons to $20 \times 20 \times 20$ neurons. All experiments were conducted using 10-fold cross-validation to ensure the robustness and reliability of the results.

The computations were conducted on the *Image masks* and *Subimage masks* with and without the application of majority voting. The highest accuracy results, corresponding to the evaluated parameter values, are summarized in Table 2. Although MV generally enhances classification accuracy compared to the performance on the *Subimage masks* dataset without MV, it only improves the classification accuracy of whole images in the case of MLP (0.627 with MV on *Subimage masks* vs. 0.548 on *Image masks*). For the other methods, classification

on *Image masks* dataset yields higher accuracy compared to its application to *Subimage masks* both with and without MV. Among classical machine learning methods, the highest accuracy obtained on binary masks is achieved with Random Forest, reaching 0.758. However, this is lower than the accuracy obtained on binary masks with deep learning, where ResNet50v2 without MV achieved 0.787, outperforming the classical approaches.

Table 2. The highest classification accuracy for stomata images on the *Image masks* and *Subimage masks* datasets achieved using classical machine learning methods, with or without the application of majority voting, across the evaluated set of adjusted hyperparameters (*adj. param.*).

	Image masks		*Subimage masks*		*Subimage masks* MV	
method	ACC	*adj. param.*	ACC	*adj. param.*	ACC	*adj. param.*
KNN	0.706	$k = 2$	0.63	$k = 10$	0.643	$k = 10$
SVM	0.647	$C = 1.8$	0.62	$C = 2.6$	0.569	$C = 2.4$
RF	0.758	$n = 400$	0.646	$n = 350$	0.692	$n = 300$
MLP	0.548	$14 \times 14 \times 14$	0.585	$16 \times 16 \times 16$	0.627	$14 \times 14 \times 14$

5 Conclusions and Discussion

The application of deep learning for dataset classification demonstrated remarkable accuracy, reaching values as high as 0.973 for *Images* dataset with ResNet152v2. It is important to emphasize that the classes under analysis exhibited distinct color variations. This characteristic may lead the model to make overly optimistic predictions, resulting in inflated accuracy metrics.

To investigate the extent to which color influences classification performance, the dataset was transformed into other color spaces designed to minimize interclass color variation. This approach aimed to challenge the model's reliance on color-based features. Consequently, the classification accuracy notably decreased, with ResNet50v2 achieving a reduced accuracy of just 0.606 on *HSV images - H* dataset. These findings underscore the importance of evaluating model performance under varying color conditions to ensure robust and generalizable classification results.

To further refine the analysis, the focus was narrowed to stomata objects by extracting subimages containing individual stomata instances. In most cases, this approach resulted in decreased classification accuracy compared to using whole images. For instance, while ResNet50v2 achieved an accuracy of 0.953 on the *Image* dataset, the accuracy dropped to 0.856 when applied to the *Subimages*.

To enhance performance on subimages, a majority voting approach was implemented, which involved aggregating predictions from multiple subimages.

This strategy generally led to improved accuracy compared to classifying individual subimages alone. However, despite the gains, the MV approach did not outperform classification on whole images. For instance, *Subimages* with ResNet50v2 and MV achieved an accuracy of 0.913, which remained lower than the 0.953 accuracy obtained with the *Image*.

To eliminate the influence of color entirely, the analysis was conducted on image masks, which preserve only the shape of the stomata. On the *Image masks* dataset, classification accuracy was 0.787 and 0.68 using ResNet50v2 and ResNet152v2, respectively. Further experimentation involved applying the same approach to *Subimage masks*, focusing on individual stomata objects. In this scenario, the highest accuracy achieved was 0.624 using ResNet152v2 with the majority voting strategy. These results highlight a noticeable decrease in performance compared to color images, emphasizing the significant role color plays in model accuracy and the challenges of relying solely on structural features for classification.

The classification results obtained using deep learning on mask datasets were compared with those achieved using a classical approach based on handcrafted shape features. After hyperparameter optimization, the highest accuracy for *Image masks* was 0.758 using Random Forest, while for *Subimage masks*, the best result was 0.692, also with Random Forest combined with majority voting. These results demonstrate that better outcomes are achieved when the entire mask is classified, as not only the shape of the stomata is important, but also the dependencies between different objects and their spatial arrangement in the image.

Deep learning models outperformed the classical approach when applied to *Image masks* containing multiple stomata instances, achieving an accuracy of 0.787 with ResNet50v2 compared to 0.758 with Random Forest on the same dataset. However, when focused on *Subimage masks*, ResNet models were outperformed by handcrafted shape features combined with majority voting. Specifically, ResNet152v2 with majority voting achieved an accuracy of 0.624, whereas Random Forest with majority voting reached 0.692. These results suggest that deep learning effectively utilizes spatial context, while handcrafted features are more effective when spatial dependencies are absent.

These findings underscore the importance of considering environmental factors in ploidy classification. A key challenge for researchers is to avoid relying on color-related features, as subtle differences in leaf shade can lead to inflated classification accuracy. The studied cultivars differed in ripening time. For instance, 'Gofert' is an early ripening variety - the fruit ripens in the first decade of July, while 'Polares' ripens in the last decade of July, which can influence the condition of the leaves collected at the end of the vegetation season.

This preliminary research highlights the need to prepare datasets under diverse conditions to introduce greater variation in stomata size and leaf color within each class. By doing so, the resulting models will be more robust to environmental variations, ensuring more reliable and generalizable classification outcomes. In addition, maintaining uniform conditions for plant growth may

facilitate visual differentiation between classes by scientists. However, this consistency can lead models to learn spurious correlations that should not be the basis for classification. Therefore, to achieve reliable and meaningful results, it is crucial to design datasets that challenge the model to focus on essential biological features rather than environmental cues.

Acknowledgements. The authors are grateful to Prof. Stanisław Pluta and Prof. Małgorzata Podwyszyńska for providing access to the plant materials used in the experiments.

This work was partly supported by Polish Ministry of Science and Higher Education through statutory funds of The National Institute of Horticultural Research, Skierniewice, Poland (Grant ZBS/7/2021).

References

1. Alsmadi, M.k., Omar, K.B., Noah, S.A., Almarashdah, I.: Performance comparison of Multi-layer Perceptron (Back Propagation, delta rule and Perceptron) algorithms in neural networks. In: 2009 IEEE International Advance Computing Conference, pp. 296–299. IEEE (2009). https://doi.org/10.1109/iadcc.2009.4809024
2. Anli, M., et al.: Improving lettuce yield and quality of an agricultural soil using a combination of arbuscular mycorrhizal fungus and phosphate-green wastes compost. Gesunde Pflanzen **74**(1), 205–217 (2021). https://doi.org/10.1007/s10343-021-00603-0
3. Avaneesh, J., Thangakumar, J., Sudalaimuthu, T., Ranjana, P., Prakash, N.S., Teja, N.S.: Accurate object detection with YOLO. Int. J. Pharm. Res. **12**(01) (2020). https://doi.org/10.31838/ijpr/2020.12.01.235
4. Balakrishnan, D., Manideep Kumar Reddy, S., Lakshmi Venkatesh, R., Aadith, K., Jebi Nalatharaj, R., Arshath, M.: Object detection on traffic data using YOLO. In: 2023 International Conference on Data Science and Network Security (ICDSNS), pp. 1–5. IEEE (2023). https://doi.org/10.1109/icdsns58469.2023.10245691
5. Casson, S.A., Hetherington, A.M.: Environmental regulation of stomatal development. Curr. Opin. Plant Biol. **13**(1), 90–95 (2010). https://doi.org/10.1016/j.pbi.2009.08.005
6. Chua, L.C., Lau, O.S.: Stomatal development in the changing climate. Development **151**(20) (2024). https://doi.org/10.1242/dev.202681
7. Deng, T.: A survey of Convolutional Neural Networks for image classification: models and datasets. In: 2022 International Conference on Big Data, Information and Computer Network (BDICN), pp. 746–749. IEEE (2022). https://doi.org/10.1109/bdicn55575.2022.00145
8. Farag, M.M.: Design and analysis of convolutional neural layers: a signal processing perspective. IEEE Access **11**, 27641–27661 (2023). https://doi.org/10.1109/access.2023.3258399
9. Fetter, K.C., Eberhardt, S., Barclay, R.S., Wing, S., Keller, S.R.: Stomatacounter: a neural network for automatic stomata identification and counting. New Phytol. **223**(3), 1671–1681 (2019). https://doi.org/10.1111/nph.15892
10. Ghosh, A., Sufian, A., Sultana, F., Chakrabarti, A., De, D.: Fundamental concepts of convolutional neural network, pp. 519–567. Springer International Publishing (2019). https://doi.org/10.1007/978-3-030-32644-9_36

11. Guan, H., Antani, S., Long, L.R., Thoma, G.R.: Comparative study of spine vertebra shape retrieval using learning-based feature selection. In: 2009 22nd IEEE International Symposium on Computer-Based Medical Systems, pp. 1–7. IEEE (2009). https://doi.org/10.1109/cbms.2009.5255384
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (2016). https://doi.org/10.1109/cvpr.2016.90
13. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks (2016). https://doi.org/10.48550/ARXIV.1603.05027
14. Jocher, G., Chaurasia, A., Qiu, J.: Ultralytics YOLOv8 (2023). https://github.com/ultralytics/ultralytics
15. Kaur, S., Agrawal, R.: Convolutional networks (2021). https://doi.org/10.1002/9781119760542.ch8
16. Konopka, A., Kozera, R., Sas-Paszt, L., Trzcinski, P., Lisek, A.: Identification of the selected soil bacteria genera based on their geometric and dispersion features. PLoS ONE **18**(10), e0293362 (2023). https://doi.org/10.1371/journal.pone.0293362
17. Konopka, A., Struniawski, K., Kozera, R.: Performance analysis of residual neural networks in soil bacteria microscopic image classification. In: 37th Annual European Simulation and Modelling Conference, ESM 2023, pp. 144–148 (2023)
18. Konopka, A., et al.: Classification of soil bacteria based on machine learning and image processing. In: Groen, D., de Mulatier, C., Paszynski, M., Krzhizhanovskaya, V.V., Dongarra, J.J., Sloot, P. (eds.) ICCS 2022. LNCS, vol. 13352, pp. 263–277. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-08757-8_23
19. Ku, K.B., et al.: New dimension in leaf stomatal behavior analysis: a robust method with machine learning approach. Plant Biotechnol. Rep. **18**(3), 361–373 (2024). https://doi.org/10.1007/s11816-024-00902-8
20. Liu, Y., Wang, Y., Zhang, J.: New machine learning algorithm: random forest. In: Liu, B., Ma, M., Chang, J. (eds.) ICICA 2012. LNCS, vol. 7473, pp. 246–252. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-34062-8_32
21. Ma, Y., et al.: A stomata-inspired superhydrophobic portable filter system. RSC Adv. **11**(31), 18783–18786 (2021). https://doi.org/10.1039/d1ra03297f
22. Mammone, A., Turchi, M., Cristianini, N.: Support vector machines. WIREs Comput. Stat. **1**(3), 283–289 (2009). https://doi.org/10.1002/wics.49
23. Podwyszyńska, M., Pluta, S.: In vitro tetraploid induction of the blackcurrant (Ribes nigrum L.) and preliminary phenotypic observations. Zemdirbyste-Agriculture **106**(2), 151–158 (2019). https://doi.org/10.13080/z-a.2019.106.020
24. Prakash, I.V., Palanivelan, M.: A study of YOLO (You Only Look Once) to YOLOv8, pp. 257–266. CRC Press (2024). https://doi.org/10.1201/9781003529231-40
25. Rothe, R., Guillaumin, M., Van Gool, L.: Non-maximum suppression for object detection by passing messages between windows. In: Cremers, D., Reid, I., Saito, H., Yang, M.-H. (eds.) ACCV 2014. LNCS, vol. 9003, pp. 290–306. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16865-4_19
26. Sabba, S., Smara, M., Benhacine, M., Terra, L., Eddine Terra, Z.: Residual neural network in genomics. Comput. Sci. J. Moldova **30**(3(90)), 308–334 (2022). https://doi.org/10.56415/csjm.v30.17
27. Sai, N., et al.: StomaAI: an efficient and user-friendly tool for measurement of stomatal pores and density using deep computer vision. New Phytol. **238**(2), 904–915 (2023). https://doi.org/10.1111/nph.18765

28. Shen, Z., Liu, Y.: A novel connectivity of deep convolutional neural networks. In: 2017 Chinese Automation Congress (CAC), pp. 7779–7783. IEEE (2017). https://doi.org/10.1109/cac.2017.8244187
29. Struniawski, K., Konopka, A., Kozera, R.: Identification of soil bacteria with machine learning and image processing techniques applying single cells' region isolation. In: The European Simulation and Modelling Conference 2022, pp. 76–81. EUROSIS-ETI (2022)
30. Wang, X., Zhang, J.: An improved automatic shape feature extraction method based on template matching. J. Phys: Conf. Ser. **2095**(1), 012053 (2021). https://doi.org/10.1088/1742-6596/2095/1/012053
31. Ye, J.C.: Convolutional Neural Networks, pp. 113–134. Springer Nature, Singapore (2021). https://doi.org/10.1007/978-981-16-6046-7_7
32. Zeiditoolabi, N., Khammari, I., Sirousmehr, A., Daneshvar, M., Galavi, M., Dahmardeh, M.: Evaluation of stomata in vetch-barley intercropping and its relationship with forage production in rainfed conditions, under the influence of biofertilizer and superabsorbent. Gesunde Pflanzen **75**(5), 2045–2073 (2023). https://doi.org/10.1007/s10343-023-00838-z
33. Zhao, L., Zhang, Z.: A improved pooling method for convolutional neural networks. Sci. Rep. **14**(1) (2024). https://doi.org/10.1038/s41598-024-51258-6
34. Zulunov, R., Soliyev, B., Kayumov, A., Asraev, M., Musayev, K., Abdurasulova, D.: Detecting mobile objects with AI using edge detection and background subtraction techniques. In: E3S Web of Conferences, vol. 508, pp. 03004 (2024). https://doi.org/10.1051/e3sconf/202450803004
35. Šmarda, P., et al.: Growth, physiology, and stomatal parameters of plant polyploids grown under ice age, present day, and future CO_2 concentrations. New Phytol. **239**(1), 399–414 (2023). https://doi.org/10.1111/nph.18955

3.7 Classification of blackcurrant genotypes by ploidy levels on stomata microscopic images with deep learning: Convolutional Neural Networks and Vision Transformers

Publication: A. Konopka, R. Kozera, A. Marasek-Ciołakowska, and A. Machlańska, "Classification of blackcurrant genotypes by ploidy levels on stomata microscopic images with deep learning: Convolutional Neural Networks and Vision Transformers," *Applied Sciences-Basel*, vol. 15, no. 19, 10735, pp. 1–14, 2025. https://doi.org/10.3390/app151910735 (100 MNiSW points, IF – 2.5)

Abstract: Plants vary in number of chromosomes (ploidy levels), which can influence morphological traits, including the size and density of stomata cells. Although biologists can detect these differences under a microscope, the process is often time-consuming and tedious. This study aims to automate the classification of blackcurrant (*Ribes nigrum* L.) ploidy levels—diploid, triploid, and tetraploid—by leveraging deep learning techniques. Convolutional Neural Networks and Vision Transformers are employed to perform microscopic image classification across two distinct blackcurrant datasets. Initial experiments demonstrate that these models can effectively classify ploidy levels when trained and tested on subsets derived from the same dataset. However, the primary challenge lies in proposing a model capable of yielding satisfactory classification results across different datasets ensuring robustness and generalization, which is a critical step toward developing a universal ploidy classification system. In this research, a variety of experiments is performed including application of augmentation technique. Model efficacy is evaluated with standard metrics and its interpretability is ensured through Gradient-weighted Class Activation Mapping visualizations. Finally, future research directions are outlined with application of other advanced state-of-the-art machine learning methods to further refine ploidy level prediction in botanical studies.

applied sciences

MDPI

Article

Classification of Blackcurrant Genotypes by Ploidy Levels on Stomata Microscopic Images with Deep Learning: Convolutional Neural Networks and Vision Transformers

Aleksandra Konopka [1,*], Ryszard Kozera [1,2], Agnieszka Marasek-Ciołakowska [3] and Aleksandra Machlańska [3]

[1] Institute of Information Technology, Warsaw University of Life Sciences—SGGW, ul. Nowoursynowska 159, 02-776 Warsaw, Poland; ryszard_kozera@sggw.edu.pl
[2] School of Physics, Mathematics and Computing, The University of Western Australia, 35 Stirling Highway, Crawley, Perth, WA 6009, Australia
[3] Department of Applied Biology, The National Institute of Horticultural Research, ul. Konstytucji 3 Maja 1/3, 96-100 Skierniewice, Poland; agnieszka.marasek@inhort.pl (A.M.-C.); aleksandra.machlanska@inhort.pl (A.M.)
* Correspondence: aleksandra_konopka@sggw.edu.pl

Abstract

Plants vary in number of chromosomes (ploidy levels), which can influence morphological traits, including the size and density of stomata cells. Although biologists can detect these differences under a microscope, the process is often time-consuming and tedious. This study aims to automate the classification of blackcurrant (*Ribes nigrum* L.) ploidy levels—diploid, triploid, and tetraploid—by leveraging deep learning techniques. Convolutional Neural Networks and Vision Transformers are employed to perform microscopic image classification across two distinct blackcurrant datasets. Initial experiments demonstrate that these models can effectively classify ploidy levels when trained and tested on subsets derived from the same dataset. However, the primary challenge lies in proposing a model capable of yielding satisfactory classification results across different datasets ensuring robustness and generalization, which is a critical step toward developing a universal ploidy classification system. In this research, a variety of experiments is performed including application of augmentation technique. Model efficacy is evaluated with standard metrics and its interpretability is ensured through Gradient-weighted Class Activation Mapping visualizations. Finally, future research directions are outlined with application of other advanced state-of-the-art machine learning methods to further refine ploidy level prediction in botanical studies.

Keywords: microscopic images; blackcurrant; stomata; deep learning; Convolutional Neural Networks; Vision Transformers

Academic Editor: Donghai Guan

Received: 22 August 2025
Revised: 30 September 2025
Accepted: 1 October 2025
Published: 5 October 2025

Citation: Konopka, A.; Kozera, R.; Marasek-Ciołakowska, A.; Machlańska, A. Classification of Blackcurrant Genotypes by Ploidy Levels on Stomata Microscopic Images with Deep Learning: Convolutional Neural Networks and Vision Transformers. *Appl. Sci.* **2025**, *15*, 10735. https://doi.org/10.3390/app151910735

1. Introduction

Plants differ in numbers of chromosomes (ploidy levels) and these variations can impact morphological features of plants at the microscopic level. The differences are evident in characteristics of stomata which are pores in plants responsible for gas exchange, controlling water transpiration and regulating light intake [1]. The objective of this research is to verify whether machine learning algorithms are capable of distinguishing ploidy levels in blackcurrant (*Ribes nigrum* L.) based on microscopic images of leaves with visible

stomata structures. In this study, the considered datasets consist of diploids, triploids, and tetraploids which refer to the number of chromosome sets.

Existing literature suggests that stomata size increases and its density decreases with higher ploidy levels [2,3]. The biologists can manually measure these differences on microscopic images; however, this approach is often time-consuming and tedious. When studying such aspects as distinguishing different ploidy levels or analyzing impact of biofertilizers on stomata characteristics, the manual measures must be repeated many times to draw significant conclusions. This is the reason why these processes need to be automated with a computerized approach, powered by artificial intelligence (AI), ultimately reducing the amount of time needed for the analysis. Although there are numerous papers on computerized stomata counting and measuring [4–6], our research that applies AI for classification of ploidy levels is a novel approach.

Historically, the stomata measurements were performed manually. To enhance the manual measuring softwares such as ImageJ [7] were commonly applied. With the development of computer vision and machine learning techniques the stomata instances were initially described with handcrafted features. As an example, in a work from 2017 [8], authors measured features such as area, major axis length, minor axis length, and eccentricity. The field of stomata analysis was transformed by deep learning [9], which became the primary tool for detection, segmentation and feature extraction. In 2020, Ref. [10] applied YOLO algorithm for stomata detection of common beans (*Phaseolus vulgaris* L.), barley plants (*Hordeum vulgare* L. cv. Henley), and two soybean (*Glycine max.* L.) cultivars (PI 398223 and PI 567201). In 2021, Ref. [11] proposed a system for tracking and monitoring stomata dynamics in wheat leaf where MobileNet was used for feature extraction. In 2025, Ref. [12] applied U-Net for image segmentation of root images on dataset comprising of monocotyledon spring and winter wheat (*Triticum aestivum* L.) and dicotyledonous faba bean (*Vicia faba* L.). Stomata microscopic images are rarely employed to solve classification problems. Existing research has addressed specific tasks: classification of certain species [9] (deep learning algorithms) and verifying open or closed stomata structures [13] (Support Vector Machine).

The segmentation of stomata in microscopic images forms a challenge due to the presence of surrounding leaf structures which exhibit high contrast comparing to the background. Due to that fact, standard traditional image analysis approaches cannot be directly applied to most datasets. In our previous work [14], the YOLO algorithm was successfully utilized for segmentation task to specify the region of interest which is the area including stomata cells. Despite the satisfactory results, this approach had the following drawbacks: a time-consuming labeling process was required and it was needed to apply an additional method for classification (yielding a two-step process). In this work, it is investigated whether a single-step classification of whole images—without prior segmentation of individual stomata—can yield satisfactory results. This study compares the performance of two leading deep learning architectures: Convolutional Neural Networks (CNNs) [15] and Vision Transformers (ViTs) [16]. Although CNNs have been frequently applied to stomata image analysis, ViTs represent a state-of-the-art approach that, to the best of our knowledge, has not yet been tested on such images.

In our latest work [14], the models learned on training dataset and were tested on testing dataset. Both datasets were derived from one set of blackcurrant microscopic images prepared under certain conditions and including selected cultivars. In contrast to prior work, this study performs computations on two distinct datasets which include three classes: diploid, triploid, and tetraploid. These datasets differ in selected cultivars and time of capturing the samples (beginning or end of different vegetation seasons). These two datasets are split into train, validation, and test subsets. The models are

trained on training subset from one set and tested iteratively on distinct subsets from the other set. This approach allows us to evaluate model performance not only on a set with the same characteristics, but also on another separate dataset. Consequently, the proposed evaluation approach enables us to verify whether the model performance is reproducible on other datasets or if the results are satisfactory only for the set prepared in initial training conditions.

The objective of this research is to verify whether AI models can successfully classify diploidy, triploidy, and tetraploidy. The models should be applicable to diverse datasets not to be limited to a constrained sample pool. In this work, performance of models is evaluated with metrics, and model interpretability is analyzed to understand decision-making criteria of considered architectures. Gradient-weighted Class Activation Mapping (Grad-CAM) method applied in this research was used in another paper on stomata detection [17] where it was applied to interpret the YOLO model. Augmentation technique is also tested to enhance generalization of the achieved results [18].

The key novel contributions of this research involve:

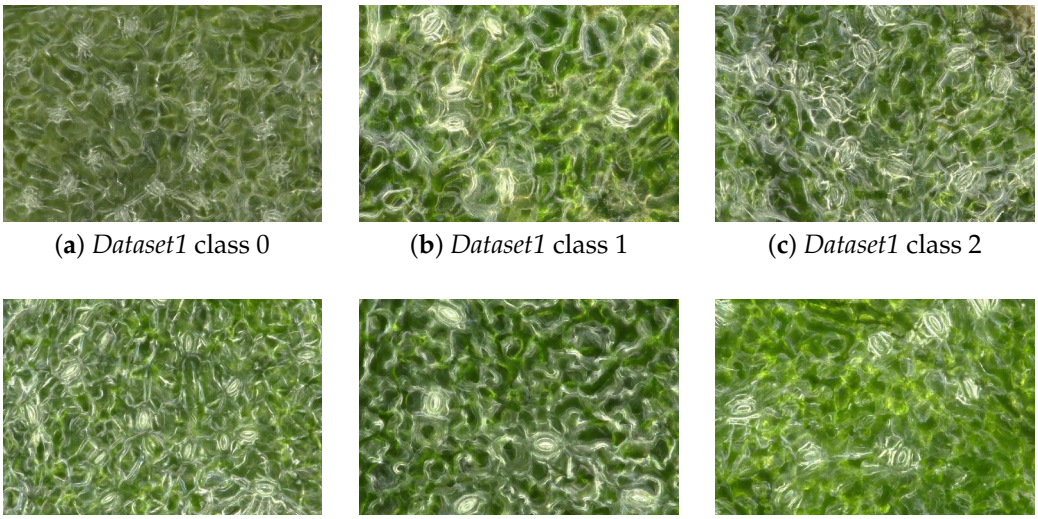
- Computerized classification of ploidy levels based on microscopic images of stomata, whereas other researchers focus on detection, segmentation and measuring stomata. To the best of our knowledge, no other research group has yet applied artificial intelligence to ploidy-level classification.
- The approach of training a model on one set and testing it on a second, different dataset for ploidy-level classification, as opposed to our previous research [14], where both training and testing was performed on subsets of a single dataset.
- Application of Vision Transformers to stomata microscopic images, as other works do not apply ViTs to any stomata-related issues.

2. Datasets

The experiments were conducted in the Experimental Pomological Orchard in Skierniewice, central Poland, which is a part of the National Institute of Horticultural Research. Blackcurrant plants (*Ribes nigrum* L.) were cultivated under natural light in 50 L pots filled with a 1:1 mixture of peat substrate and soil supplemented with slow-release fertilizers. A drip system was employed to manage the irrigation regimes, and the moisture content of the growing medium was monitored using a dielectric probe (TEROS-12; METER, San Francisco, CA, USA). The water potential in the growth medium was maintained at $(-)10$ kPa to ensure optimal irrigation. The plants were protected against pests and diseases according to the Integrated Pest Management guidelines.

The datasets comprise three classes of blackcurrant genotypes distinguished by their chromosome counts: diploid $(2n = 2x = 16)$ cultivar (class 0), triploid $(2n = 3x = 24)$ cultivar (class 1), and autotetraploid $(2n = 4x = 32)$ clone (class 2). All images were captured using a VHX-7000N KEYENCE digital microscope, to ensure high-quality images suitable for detailed analysis (see Figure 1).

In this research, computations are preformed on two distinct datasets.



(**a**) *Dataset1* class 0 (**b**) *Dataset1* class 1 (**c**) *Dataset1* class 2

(**d**) *Dataset2* class 0 (**e**) *Dataset2* class 1 (**f**) *Dataset2* class 2

Figure 1. Microscopic images of blackcurrant (*Ribes nigrum* L.) leaves showing stomata cells across different ploidy levels: (**a**–**c**) diploid (class 0), triploid (class 1), and tetraploid (class 2) samples from *Dataset1*; (**d**–**f**) corresponding ploidy classes from *Dataset2*. All images were captured using a VHX-7000N KEYENCE digital microscope.

2.1. Dataset1

Class 0 contains 252 images of the diploid 'Gofert' cultivar and 252 images of the diploid 'Polares' cultivar. Class 1 includes 502 images of the triploid 'Dlinnokistnaja' cultivar, while class 2 consists of 251 images of the autotetraploid clone of 'Gofert' cultivar and 250 images of the 'Polares' cultivar obtained by in vitro polyploidization [19]. Leaves were sampled for photo documentation at the end of a vegetation season between 5 and 25 September 2024.

2.2. Dataset2

Class 0 contains 500 images of diploid 'Gofert' cultivar. Class 1 includes 500 images of the triploid 'Dlinnokistnaja' cultivar and class 2 contains 500 images of an autotetraploid clone of 'Gofert' [19]. Leaves were sampled for photo documentation at the beginning of the vegetation season between 14 and 30 April 2025.

3. Methods

Neural networks [20] are machine learning models which evolved from a single artificial neuron. This neuron was trained with perceptron algorithm which is capable of solving linearly separable binary problems [21]. Neural networks comprise of neurons which form layers. The number of trainable parameters in neural network increases along with complexity of its architecture. When a network consists of many layers then it is referred to as a deep neural network and its training process is termed deep learning. Neural networks are applied in many fields including computer vision [22], signal processing [23], and natural language processing [24].

In the classical approach, the input to the neural network is a matrix of features. These features are handcrafted by IT specialists in collaboration with domain experts, who determine which traits are relevant for a specific task [25]. For instance, in the classification of stomata images with handcrafted features, one could consider the computing number of stomata objects on each image, as well as their size and shape eccentricity. The surge of deep learning field results in new algorithms leveraging automatic feature extraction [22,26]. This process is performed with two mechanisms employed in this study—convolution [27] in Convolutional Neural Networks and self-attention [16] in Transformers.

3.1. Convolutional Neural Networks

Convolutional Neural Networks [28] are deep neural networks which are commonly applied in image analysis for tasks involving classification, detection, or segmentation. CNNs were first studied in 1980s with models proposed by Yann LeCun [15]. Further development of these architectures accelerated in early 21st century. CNNs comprise two modules: feature extraction and task-specific part (responsible, e.g., for classification or regression).

The feature extraction module processes input image data and generates a matrix representation through automatically learned features. This module is composed of alternating convolutional and pooling layers. The yielded feature matrix is subsequently fed into the second module, which is usually implemented as Multilayer Perceptron [29], where task-specific prediction is performed.

In convolution layers [27], the filters are applied to extract features with a sliding window operation. This operation is a cross-correlation between a filter and image patches which is also referred to as convolution. These filters are matrices with values (weights), trained along with other network parameters. The weights are adjusted to detect spatial features such as edges or textures.

The pooling layers [30] are applied to reduce spatial dimensions of the image. Despite this reduction, each layer is designed in such a way that the most important details are retained. The pooling algorithm can be performed differently, e.g., as a max pooling (only the maximal value within a patch preserves) or average pooling (mean value of all elements within the window is calculated). Pooling operation is applied to reduce computational complexity and migitate overfitting.

Residual Neural Network (ResNet) is an architecture of CNN introduced by He et al. in 2016 [31]. In neural networks trained by backpropagation algorithm, a problem with unstable gradients occurs in deeper layers of CNNs leading to exploding or vanishing gradient phenomena. This issues are addressed by ResNets with application of skip connections [32,33] which are the connections that bypass one or more layers. The gradient can efficiently propagate through such network which improves its stability and performance.

ResNet architectures are modified to enhance the accuracy (ACC) and performance. In ResNet152v2 [34], the order of altering layers is changed from: convolution, batch normalization, Rectified Linear Unit (ReLU) [35] to batch normalization, ReLU and convolution. This adjustment enhances training stability and further reduces gradient vanishing.

3.2. Transformers

Transformers are deep neural networks introduced in 2017 [16] which were originally created for natural language processing tasks.

The core of this methodology is the tokenization technique sampling the training corpus into smaller units. This approach is applied to reduce the dimensionality of data (not to process each character separately) and then to model dependencies between phrases in subsequent steps. Each token is assigned an individual ID which is selected based on its frequency of appearance in training corpus. To capture dependencies between tokens, an embedding mechanism is applied, designed to explain relations between words as vectors in high dimensional space [36]. Preserving the order of words in a text is essential for learning their context. In addition, the distance between specific words is represented by positional encoding which adds positional information to embedding input [37].

Then, the self-attention mechanism is applied to weight relevance between tokens in input data. An attention score is computed between pairs of tokens, where a high attention score represents high dependency. To obtain these relations, specific matrices W_Q (query),

W_K (key), and W_V (value) are trained with backpropagation algorithm alongside the rest of the model's parameters.

Vision Transformers, proposed in 2020 [38], extend the Transformer architecture to computer vision tasks by processing images as input. Both ViTs and traditional Transformers employ the same self-attention mechanism; however, ViTs differ in their input data processing. ViTs decompose images into patches, which are then linearly embedded into tokens. These tokens are served as the input sequence for the Transformer encoder [39].

4. Experiments and Results

The experiments were performed on two sets of images *DataSet1* and *DataSet2*. Each dataset contained three classes of microscopic images with different ploidy levels (class 0: diploid, class 1: triploid, and class 2: tetraploid). The images were processed before applying them as input to classification methods. Some of the images in the *Dataset2* contained a label with information about the magnification. For this purpose, it was decided to reduce the size of all images removing the labels, yielding a final image size of $2880 \times 1950 \times 3$. Each of these two datasets was split into three subsets: train, valid, and test in 7:2:1 ratio. All computations were performed in Python programming language on AMD Ryzen 5 5650 with 32 GB DDR4 and RTX 3060 GPU. ResNet152v2 and ViTs were applied for classification purposes.

It is worth emphasizing that both ResNets and ViTs demonstrated the ability to accurately classify three classes in testing set corresponding to training set. For example, when a model was trained on *Dataset1* train subset and tested on *Dataset1* test subset, the results were higher than 0.97 accuracy—reaching even 100% for some models. However, our aim was to train a model that could distinguish any diploid, triploid, and tetraploid, rather than only recognizing samples contained within a specific dataset. Due to that reason, this work performs experiments with the following approach—each model is trained on training set derived from one set and tested on another dataset. The testing was conducted on whole datasets, with performance results presented in Tables 1–3.

Additionally, to ensure the stability of the obtained results, the validation was performed multiple times. *Dataset1* and *Dataset2* were split into five distinct subset each. These subsets contained balanced number of images from three classes (0, 1, 2) yielding sets of 100–101 samples of a given class (around 300 samples per each of 5 sets). Then, each of these five sets was applied to verify the performance of a model, e.g., when the model was trained on the training set from *Dataset1*, then it was tasted on all five subsets from *Dataset2*, and when the model was trained on the training set from *Dataset2*, then it was tasted on all five subsets from *Dataset1*. This approach was applied in all the experiments in this work, and the mean values and standard deviations of five results computed for each of the models are available in Supporting Information.

4.1. ResNet Experimental Setup

The Residual Neural Network applied in this research was ResNet152v2 implementation from *Keras* library [40]. The images from the datasets were applied as input to ResNet with size reduced by 4 (our experiments showed that applying images resized to default 224×224 yields significantly lower accuracy results). The ResNet152v2 was the first part of a sequential model which also consisted of other layers: dense layer with 512 neurons and ReLU activation function, dropout layer with 30% drop, again a dense layer with 512 neurons, and ReLU activation function, dropout layer with 30% drop, and a dense layer with 3 neurons with SoftMax activation function which outputs probabilities that a given sample is assigned to each of these three classes. The optimizer applied in this research was Adam [41] with learning rate 0.01 and *categorical_crossentropy* loss function

was used (from *Keras*). The model was pretrained on ImageNet dataset and finetuned from *conv5_block1_1_conv* layer. The model was trained in 1000 epochs with batch size 8 and patience 50. The training usually stopped around 100 epochs.

4.2. ViT Experimental Setup

The ViT applied in this research were two pretrained models: *vit-large-patch16-224-in21k* and *vit-base-patch16-224-in21k* from Hugging Face's *Transformers* library [42]. These two models differ in size (~86M parameters base and ~307M parameters large) leading to differences in computational cost and performance. The models are pretrained on ImageNet-21k dataset and all its layers are finetuned. The models resize input images to 224×224 and classify data into three categories. The batch size was set to 16 and number of epochs 10 (in most cases, the model converged around the sixth epoch). The learning optimizer is AdamW [43] with learning rate of 2×10^{-5}, the loss function is *cross-entropy* (from *PyTorch* library [44]).

4.3. Model Evaluation Metrics

The models were evaluated with computed metrics: accuracy, precision, recall, and F1-score. Accuracy measures percentage of all correct predictions. Precision computes proportion of true positives divided by true and false positives combined. Recall measures proportion of true positives divided by true positives and false negatives combined. F1-score is a harmonic mean of precision (p) and recall (r): $F1 = 2 \times \frac{p \times r}{p+r}$.

Although this research focuses on maximizing the accuracy, the other tested measures are also helpful for understanding how the model manages to discern instances within each specific class.

4.4. Experiments on Raw Datasets

The models under consideration: ResNet152v2, *vit-large-patch16-224-in21k*, and *vit-base-patch16-224-in21k*, were trained on *Dataset1* training set, tested on *Dataset2* and then trained on *Dataset2* training set and tested on *Dataset1* (see Table 1). It is shown that better accuracy results are achieved when model learns on *Dataset1* and is tested on *Dataset2* in two out of three tested scenarios. The best accuracy results are obtained with Residual Network which is equal to 0.68. Large ViT model obtained higher ACC results than the base one. Most of the models have problems with correct classification of class 1 which is illustrated with low values of precision or recall for this class. The model that performed the best in classifying class 1 is ResNet trained on *Dataset1*.

Table 1. Comparative performance analysis of Vision Transformer and ResNet architectures in datasets *Dataset1* (1) and *Dataset2* (2) for all 3 classes. Models were evaluated on whole datasets and the values of metrics—accuracy (acc), precision (p), recall (r), and F1-score (f1) are presented. Vit-b: *vit-base-patch16-224-in21k*; Vit-l: *vit-large-patch16-224-in21k*; RN: ResNet152v2. Training (Tr) and testing (Te) sets are indicated for each experiment.

Tr	Te	mod	acc	0_p	0_r	0_f1	1_p	1_r	1_f1	2_p	2_r	2_f1
1	2	vit-b	0.57	0.83	0.72	0.77	0.82	0.12	0.21	0.44	0.88	0.59
2	1	vit-b	0.52	0.49	0.99	0.66	1	0	0.01	0.57	0.56	0.57
1	2	vit-l	0.56	0.90	0.59	0.71	0.96	0.16	0.28	0.43	0.93	0.59
2	1	vit-l	0.59	0.72	0.96	0.82	1	0	0.01	0.48	0.79	0.60
1	2	RN	0.68	0.72	0.66	0.69	0.63	0.91	0.75	0.72	0.47	0.57
2	1	RN	0.54	0.91	0.64	0.75	0	0	0	0.42	0.97	0.59

4.4.1. Experiments on Two Classes

Separate experiments were performed on raw datasets: *Dataset1* and *Dataset2* considering only class 0 and class 2. These two classes differ in most extent in number of chromosomes so it was decided to verify the classification results obtained by analyzed ViT and ResNet models for this combination (see Table 2). Both base and large ViT architectures outperform ResNet152v2 on both tested datasets. The highest yielded accuracy is 0.88 for *vit-base-patch16-224-in21k* model and the lowest 0.79 for ResNet152v2.

Table 2. Comparative performance analysis of Vision Transformer and ResNet architectures in datasets *Dataset1* (1_02) and *Dataset2* (2_02) for two classes: 0—diploid and 2—tetraploid. Models were evaluated on whole datasets and the values of metrics—accuracy (acc), precision (p), recall (r), and F1-score (f1) are presented. Vit-b: *vit-base-patch16-224-in21k*; Vit-l: *vit-large-patch16-224-in21k*; RN: ResNet152v2. Training (Tr) and testing (Te) sets are indicated for each experiment.

Tr	Te	mod	acc	0_p	0_r	0_f1	2_p	2_r	2_f1
1_02	2_02	vit-b	0.81	0.88	0.71	0.78	0.76	0.90	0.82
2_02	1_02	vit-b	0.88	0.84	0.93	0.88	0.92	0.82	0.87
1_02	2_02	vit-l	0.82	0.93	0.69	0.79	0.75	0.95	0.84
2_02	1_02	vit-l	0.82	0.74	0.99	0.85	0.98	0.66	0.79
1_02	2_02	RN	0.79	0.88	0.68	0.76	0.74	0.90	0.81
2_02	1_02	RN	0.80	0.94	0.63	0.76	0.72	0.96	0.83

4.4.2. Experiments on Augmented Datasets

The images in classes 0, 1, and 2 in each of two considered datasets differed in shades of green color. Despite that fact models could successfully learn to distinguish classes in different sets. It was decided to apply augmentation [18] to prevent overfitting and enhance generalization of the model. Each of the images in the dataset was augmented to three new images. The augmentation process contained such modifications as follows: random (+/− 10–40%) brightness modification, (+/− 10–40%) contrast, (+/− 10–40%) saturation, and (+/− 15%) hue. Additionally, auto-contrast correction was applied with a 30–70% probability, while sharpness enhancement (sharpness factor: 1.5–2.5) was introduced with a 20–40% probability. These random values were generated from uniform distributions. The implementation was carried out using *Pillow* [45] and *TorchVision* [46] libraries.

The classification was performed on augmented datasets—*Dataset1* and *Dataset2*, (yielding 12000 images in both sets combined) and on classes 0 and 2 extracted from these two datasets. The classification performance was evaluated on raw datasets (not augmented ones). To verify if augmentation increased classification accuracy let us first compare results from Table 1 with Table 3. In case of *vit-base-patch16-224-in21k*, the augmentation increased accuracy results for models tested on *Dataset1* by 2 percentage points (pp) and on *Dataset2* by 8 pp. In case of *vit-large-patch16-224-in21k*, the accuracy decreased for model trained on *Dataset1* augmented by 3 pp. The use of augmentation had a mostly positive impact on the classification of class 1 (which was the most problematic class to classify) when employing Vision Transformers, as evidenced by the increase in both precision and recall. Before augmentation, class 1 showed particularly poor performance with recall often at 0%, indicating the model completely failed to detect these instances in many cases. The results yielded for ResNet152v2 trained on augmented *Dataset1* decreased classification results by 8 pp and for *Dataset2* ACC increased by 5 pp.

Comparing the results obtained in Table 2 with Table 3, one can notice that augmentation did not increase accuracy in most scenarios for Vision Transformers when two classes were classified. The decrease in ACC was between 1 and 8 pp, and the only increase (by 6 pp)

was observed for *vit-large-patch16-224-in21k* trained on *Dataset2*. In case or Residual Neural Network, the accuracy decreased by 1 pp on model trained on *Dataset1* and increased by 5 pp for model trained on *Dataset2*.

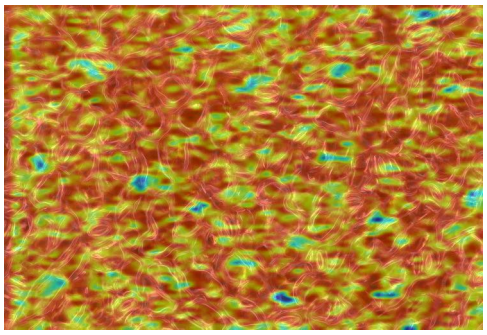
Despite the fact that augmentation increased accuracy results for many models, they did not outperform models without augmentation. For three classes, the highest ACC was 0.68 for ResNet152v2 on *Dataset1* and for two classes 0.88 for *vit-base-patch16-224-in21k* on *Dataset2* and for *vit-large-patch16-224-in21k* on *Dataset2* when augmentation was applied.

Table 3. Comparative performance analysis of Vision Transformer and ResNet architectures in datasets *Dataset1* (1) and *Dataset2* (2) for all three classes and for two classes 0—diploid and 2—tetraploid (1_02 and 2_02, respectively). Models are trained on augmented training datasets (Tr) (1_a, 2_a, 1_a_02, 2_a_02) and tested on raw testing datasets (Te). Models were evaluated on whole datasets and the values of metrics accuracy (acc), precision (p), recall (r), and F1-score (f1) metrics are presented. Vit-b: *vit-base-patch16-224-in21k*; Vit-l: *vit-large-patch16-224-in21k*; RN: ResNet152v2.

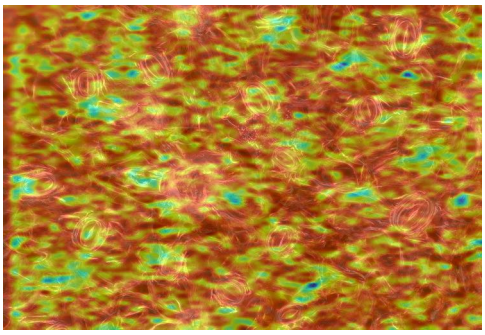
Tr	Te	mod	acc	0_p	0_r	0_f1	1_p	1_r	1_f1	2_p	2_r	2_f1
1_a	2	vit-b	0.59	0.91	0.71	0.80	0.91	0.14	0.25	0.45	0.92	0.60
2_a	1	vit-b	0.60	0.69	0.91	0.78	1	0.07	0.13	0.52	0.83	0.64
1_a_02	2_02	vit-b	0.73	0.93	0.49	0.64	-	-	-	0.65	0.97	0.78
2_a_02	1_02	vit-b	0.80	0.72	0.97	0.83	-	-	-	0.95	0.63	0.76
1_a	2	vit-l	0.53	0.94	0.44	0.60	0.90	0.19	0.31	0.42	0.97	0.58
2_a	1	vit-l	0.59	0.69	0.89	0.78	1	0.08	0.14	0.50	0.82	0.62
1_a_02	2_02	vit-l	0.78	0.91	0.62	0.73	-	-	-	0.71	0.94	0.81
2_a_02	1_02	vit-l	0.88	0.85	0.91	0.88	-	-	-	0.90	0.84	0.87
1_a	2	RN	0.60	0.52	0.90	0.66	0.68	0.44	0.54	0.73	0.47	0.57
2_a	1	RN	0.59	0.72	0.82	0.77	1	0.07	0.14	0.49	0.87	0.62
1_a_02	2_02	RN	0.78	0.91	0.63	0.74	-	-	-	0.72	0.94	0.81
2_a_02	1_02	RN	0.85	0.87	0.83	0.85	-	-	-	0.84	0.88	0.86

4.4.3. Gradient-Weighted Class Activation Mapping

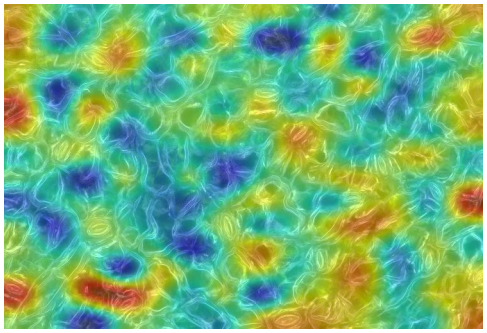
To visualize the areas on images which were the most important in the classification process, Gradient-weighted Class Activation Mapping method [47] was employed on both ResNets and Vision Transformers. In both Grad-CAMs for ResNet and ViT, the gradient weights were computed with global average pooling. The resulting activation map was normalized with min-max scaling and then resampled to original image size with bicubic interpolation. Then the heatmap was applied to image with opacity of 0.5. The results were tested on many images, models, and their layers. Figure 2 presents Grad-CAM activation maps of two models with the highest accuracy result for ViT and for ResNet in classification of two classes. The first considered model was ResNet152v2 trained on *Dataset2* class 0 and 2 with augmented data, which obtained 0.85 ACC when tested on *Dataset1*. The second model was *vit-base-patch16-224-in21k* trained on *Dataset2* class 0 and 2 with only raw data, which yielded 0.88 ACC when tested on *Dataset1*. In case of ResNet, the Grad-CAM heatmap presents results obtained on *conv5_block1_1_conv* layer. In ViT, Grad-CAM visualizes the last dense layer in the final encoder block. Figure 2 presets evaluation of the latter models on two images: image from class 0 testing set of *Dataset2* and image from class 2 testing set of *Dataset1*. In the case of ViT, the heatmaps consistently highlight stomata cells, as indicated by warm colors (red and yellow) in the relevant regions. In contrast, ResNet's heatmaps exhibit diffuse attention, focusing on irregular and less interpretable areas of the image that do not correlate with stomata positions. As deeper layers of ResNet were tested, the coverage of red areas expanded, further reducing localization accuracy.



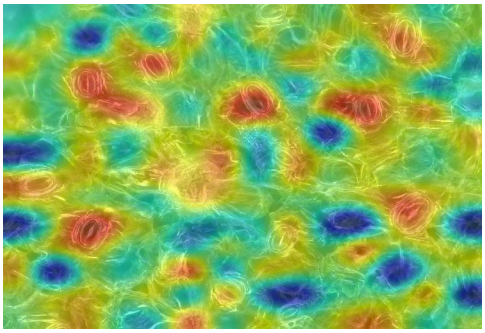
(**a**) ResNet152v2 model tested on image from *Dataset2* class 0 testing set



(**b**) ResNet152v2 model tested on image from *Dataset1* class 2 testing set



(**c**) ViT model tested on image from *Dataset2* class 0 testing set



(**d**) ViT model tested on image from *Dataset1* class 2 testing set

Figure 2. GradCAM activation maps comparing ViT (*vit-base-patch16-224-in21k*) trained on *Dataset2* of raw images with ResNet152v2 model trained on *Dataset2* with augmentation applied for ploidy level classification (Class 0: diploid, Class 2: tetraploid) tested on selected images.

5. Conclusions

In this work, a classification of three ploidy levels was performed (diploidy, triploidy, and tetraploidy) on microscopic images of blackcurrant. All the models were trained on one dataset and tested on an entirely independent datasets collected under various conditions and containing a different sets of blackcurrant genotypes. The ViT and ResNet models could achieve exceptional classification performance on three classes when one dataset was split into training, validation, and testing sets, and the model learned on training and was tested on testing set. These subsets, even if contained different images, included blackcurrant images of specific species and were taken in identical conditions. The models obtained accuracy above 0.97, confirming that distinguishing between ploidy levels within homogeneous data presents no significant challenge.

The primary scientific contribution of this work lies in evaluating model generalizability across distinct datasets. In our previous research on ploidy-level classification with AI [14], the computations were performed on training set and testing was performed on distinct testing set both derived from a single dataset. That dataset consisted of samples prepared in the same conditions. This research focuses on training models and testing them on distinct datasets which include different sets of blackcurrant species and were prepared under varying conditions. In previous work [14], the stomata instances were manually labeled, and then this process was continued with YOLO algorithm. The images with extracted stomata instances were served as input to classification methods. In this work, the classification is reduced to a single-step approach. In this research, tested Residual Neural Networks and Vision Transformers take whole images as input. The results reveal that these models can successfully distinguish blackcurrant reaching up to 0.69 ACC on three classes with ResNet152v2 and 0.88 for *vit-base-patch16-224-in21k* in binary

classification of the most distinct ploidy levels. In three-class classification, the maximal ACC obtained with ViT was equal to 0.6 on *vit-base-patch16-224-in21k* with augmentation applied. The augmentation approach increased accuracy in most cases when three classes were classified. This technique increased the availability of the model to classify the most challenging class 1 (triploidy). This class is the most problematic to distinguish as the size of stomata and its density increases along with ploidity level and the class 1 is a triploid while 0 includes diploidy and 2—tetraploidy. Comparing the results obtained in Table 1, where no augmentation was used with Table 3 with augmentation applied one can notice an increase of values of precision and recall when the model is trained on *Dataset2* their values are no longer equal to 0 when the augmentation technique is employed.

Through Grad-CAM visualization, we uncovered fundamental differences in decision-making processes of the models. ViTs demonstrated biologically interpretable behavior, consistently focusing on stomata structures. In contrast, ResNets exhibited progressively broader attention patterns in deeper layers, incorporating more global image features in their classifications.

6. Discussion

These findings demonstrate that deep learning can effectively differentiate ploidy levels in blackcurrant plants, though several opportunities for improvement remain. In further research, we suggest development of more diverse, multi-species datasets to enhance generalizability. Deeper analysis of Grad-CAM results with different classification methods and various modifications of rigor parameters can be presented and its evaluation with metrics is to be applied. Other augmentation techniques should be employed including geometric modifications such as rotation or flipping.

It is also worth considering different Convolutional Neural Networks (e.g., EfficientNet [48], ConvNeXt [49]) and different Transformer architectures (e.g., Swin Transformer [50]). A possible direction would be to evaluate hybrid approaches (e.g., Mamba [51]), which leverage both the detailed feature extraction of CNNs and the broader contextual understanding of self-attention mechanisms from Transformers.

This analysis shows that ViTs are a promising direction in analysis of stomata images. Transformers should also be employed to stomata microscopic images in other contexts such as stomata detection and measuring. The methodologies and research directions proposed in this work contribute to several scientific domains including analysis of stomata behaviour under different environmental conditions, studies involving biofertilizer production, and cross-species stomata comparisons.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/app151910735/s1.

Author Contributions: Conceptualization, A.K. and A.M.-C.; methodology, A.K.; software, A.K.; validation, A.K. and R.K.; formal analysis, A.K.; investigation, A.K. and A.M.-C.; resources, A.M.-C. and A.M.; data curation, A.K.; writing-original draft preparation, A.K.; writing-review and editing R.K., A.M.-C., and A.M.; visualization, A.K.; supervision, R.K.; project administration, A.K.; funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data and code supporting the findings of this study are available from the corresponding author upon request.

Acknowledgments: The authors are grateful to Stanisław Pluta and Małgorzata Podwyszyńska for providing access to the plant materials used in the experiments.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Casson, S.A.; Hetherington, A.M. Environmental regulation of stomatal development. *Curr. Opin. Plant Biol.* **2010**, *13*, 90–95. [CrossRef]
2. Mtileni, M.; Venter, N.; Glennon, K. Ploidy differences affect leaf functional traits, but not water stress responses in a mountain endemic plant population. *S. Afr. J. Bot.* **2021**, *138*, 76–83. [CrossRef]
3. Van Laere, K.; França, S.C.; Vansteenkiste, H.; Van Huylenbroeck, J.; Steppe, K.; Van Labeke, M.C. Influence of ploidy level on morphology, growth and drought susceptibility in Spathiphyllum wallisii. *Acta Physiol. Plant.* **2010**, *33*, 1149–1156. [CrossRef]
4. Sai, N.; Bockman, J.P.; Chen, H.; Watson-Haigh, N.; Xu, B.; Feng, X.; Piechatzek, A.; Shen, C.; Gilliham, M. StomaAI: An efficient and user-friendly tool for measurement of stomatal pores and density using deep computer vision. *New Phytol.* **2023**, *238*, 904–915. [CrossRef]
5. Wu, T.L.; Chen, P.Y.; Du, X.; Wu, H.; Ou, J.Y.; Zheng, P.X.; Wu, Y.L.; Wang, R.S.; Hsu, T.C.; Lin, C.Y.; et al. StomaVision: Stomatal trait analysis through deep learning. *bioRxiv* **2024**. [CrossRef]
6. Fetter, K.C.; Eberhardt, S.; Barclay, R.S.; Wing, S.; Keller, S.R. StomataCounter: A neural network for automatic stomata identification and counting. *New Phytol.* **2019**, *223*, 1671–1681. [CrossRef]
7. Schneider, C.A.; Rasband, W.S.; Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012**, *9*, 671–675. [CrossRef]
8. Jayakody, H.; Liu, S.; Whitty, M.; Petrie, P. Microscope image based fully automated stomata detection and pore measurement method for grapevines. *Plant Methods* **2017**, *13*, 94. [CrossRef]
9. Andayani, U.; Sumantri, I.B.; Pahala, A.; Muchtar, M.A. The implementation of deep learning using Convolutional Neural Network to classify based on stomata microscopic image of curcuma herbal plants. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *851*, 012035. [CrossRef]
10. Casado-García, A.; del Canto, A.; Sanz-Saez, A.; Pérez-López, U.; Bilbao-Kareaga, A.; Fritschi, F.B.; Miranda-Apodaca, J.; Muñoz-Rueda, A.; Sillero-Martínez, A.; Yoldi-Achalandabaso, A.; et al. LabelStoma: A tool for stomata detection based on the YOLO algorithm. *Comput. Electron. Agric.* **2020**, *178*, 105751. [CrossRef]
11. Sun, Z.; Song, Y.; Li, Q.; Cai, J.; Wang, X.; Zhou, Q.; Huang, M.; Jiang, D. An integrated method for tracking and monitoring stomata dynamics from microscope videos. *Plant Phenomics* **2021**, *2021*, 9835961. [CrossRef]
12. Wacker, T.S.; Smith, A.G.; Jensen, S.M.; Pflüger, T.; Hertz, V.G.; Rosenqvist, E.; Liu, F.; Dresbøll, D.B. Stomata morphology measurement with interactive machine learning: Accuracy, speed, and biological relevance? *Plant Methods* **2025**, *21*, 95. [CrossRef]
13. Razzaq, A.; Shahid, S.; Akram, M.; Ashraf, M.; Iqbal, S.; Hussain, A.; Azam Zia, M.; Qadri, S.; Saher, N.; Shahzad, F.; et al. Stomatal state identification and classification in quinoa microscopic imprints through deep learning. *Complexity* **2021**, *2021*, 9938013. [CrossRef]
14. Konopka, A.; Struniawski, K.; Kozera, R.; Ortenzi, L.; Marasek-Ciołakowska, A.; Machlańska, A. Deep Learning Classification of Blackcurrant Genotypes by Ploidy Levels on Stomata Microscopic Images. In *Computational Science—ICCS 2025 Workshops*; Springer Nature: Cham, Switzerland , 2025; pp. 135–148. [CrossRef]
15. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* **1989**, *1*, 541–551. [CrossRef]
16. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. *arXiv* **2017**, arXiv:1706.03762. [CrossRef]
17. Li, X.; Guo, S.; Gong, L.; Lan, Y. An automatic plant leaf stoma detection method based on YOLOv5. *IET Image Process.* **2022**, *17*, 67–76. [CrossRef]
18. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. *J. Big Data* **2019**, *6*, 60. [CrossRef]
19. Podwyszyńska, M.; Pluta, S. In vitro tetraploid induction of the blackcurrant (*Ribes nigrum* L.) and preliminary phenotypic observations. *Zemdirb. Agric.* **2019**, *106*, 151–158. [CrossRef]
20. Haykin, S. *Neural Networks and Learning Machines*, 3rd ed.; Pearson: Bloomington, MN, USA, 2008.
21. Rosenblatt, F. *The Perceptron—A Perceiving and Recognizing Automaton*; Report 85-460-1; Cornell Aeronautical Laboratory: Buffalo, NY, USA, 1957.
22. Konopka, A.; Struniawski, K.; Kozera, R. Performance analysis of Residual Neural Networks in soil bacteria microscopic image classification. In Proceedings of the 37th Annual European Simulation and Modelling Conference, ESM 2023, Toulouse, France, 24–26 October 2023; pp. 144–148.
23. Slavutskii, L.; Lazareva, N.; Portnov, M.; Slavutskaya, E. Neural net without deep learning: Signal approximation by multilayer perceptron. In Proceedings of the 2nd International Conference on Computer Applications for Management and Sustainable Development of Production and Industry (CMSD-II-2022), SPIE, Dushanbe, Tajikistan, 21–23 December 2022; p. 11. [CrossRef]
24. Wang, L.; Meng, Z. Multichannel two-dimensional Convolutional Neural Network based on interactive features and group strategy for chinese sentiment analysis. *Sensors* **2022**, *22*, 714. [CrossRef]

25. Konopka, A.; Kozera, R.; Sas-Paszt, L.; Trzcinski, P.; Lisek, A. Identification of the selected soil bacteria genera based on their geometric and dispersion features. *PLoS ONE* **2023**, *18*, e0293362. [CrossRef]
26. Konopka, A.; Kozera, R.; Sas-Paszt, L.; Trzciński, P. Automated imaging and machine learning for soil bacteria classification: Challenges and insights. *Eng. Appl. Artif. Intell.* **2025**, *159*, 111369. [CrossRef]
27. Hsu, C.Y.; Tseng, C.C.; Lee, S.L.; Xiao, B.Y. Image classification using Convolutional Neural Networks with different convolution operations. In Proceedings of the 2020 IEEE International Conference on Consumer Electronics—Taiwan (ICCE-Taiwan), Taoyuan, Taiwan, 28–30 September 2020; pp. 1–2. [CrossRef]
28. LeCun, Y.; Jackel, L.D.; Bottou, L.; Cortes, C.; Denker, J.S.; Drucker, H.; Guyon, I.; Muller, U.A.; Sackinger, E.; Simard, P.; et al. Learning algorithms for classification: A comparison on handwritten digit recognition. In *Neural Networks*; World Scientific: Pohang, Republic of Korea, 1995; pp. 261–276. [CrossRef]
29. Vang-Mata, R. *Multilayer Perceptrons: Theory and Applications*; Nova Science Publishers, Inc.: Hauppauge, NY, USA, 2020.
30. Zhao, L.; Zhang, Z. A improved pooling method for Convolutional Neural Networks. *Sci. Rep.* **2024**, *14* , 1589. [CrossRef]
31. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA , 27–30 June 2016; pp. 770–778. [CrossRef]
32. Borawar, L.; Kaur, R. ResNet: Solving Vanishing Gradient in Deep Networks. In *Proceedings of International Conference on Recent Trends in Computing*; Springer Nature: Singapore, 2023; pp. 235–247. [CrossRef]
33. Philipp, G.; Song, D.; Carbonell, J.G. Gradients explode—Deep networks are shallow—ResNet explained. In Proceedings of the 6th International Conference on Learning Representations, ICLR 2018—Workshop Track Proceedings, Vancouver, BC, Canada, 30 April–3 May 2018.
34. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity mappings in Deep Residual Networks. *arXiv* **2016**, arXiv:1603.05027. [CrossRef]
35. Varshney, M.; Singh, P. Optimizing nonlinear activation function for convolutional neural networks. *Signal Image Video Process* **2021**, *15*, 1323–1330. [CrossRef]
36. Azar, G.A.; Emami, M.; Fletcher, A.; Rangan, S. Learning embedding representations in high dimensions. In Proceedings of the 2024 58th Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ, USA, 13–15 March 2024; pp. 1–6. [CrossRef]
37. Zheng, C.; Gao, Y.; Shi, H.; Huang, M.; Li, J.; Xiong, J.; Ren, X.; Ng, M.; Jiang, X.; Li, Z.; et al. DAPE: Data-Adaptive Positional Encoding for length extrapolation. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 26659–26700.
38. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929. [CrossRef]
39. Shah, S.M.A.H.; Khan, M.Q.; Ghadi, Y.Y.; Jan, S.U.; Mzoughi, O.; Hamdi, M. A hybrid neuro-fuzzy approach for heterogeneous patch encoding in ViTs using contrastive embeddings and deep knowledge dispersion. *IEEE Access* **2023**, *11*, 83171–83186. [CrossRef]
40. Keras. 2015. Available online: https://keras.io (accessed on 30 September 2025).
41. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980. [CrossRef]
42. Wolf, T.; Debut, L.; Sanh, V.; Chaumond, J.; Delangue, C.; Moi, A.; Cistac, P.; Rault, T.; Louf, R.; Funtowicz, M.; et al. Transformers: State-of-the-art natural language processing. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Online, 16–20 November 2020; pp. 38–45.
43. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. *arXiv* **2017**, arXiv:1711.05101. [CrossRef]
44. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*; Curran Associates, Inc.: Red Hook, NY, USA, 2019; pp. 8024–8035.
45. Clark, A. Pillow (PIL Fork) Documentation. 2015. Available online: https://buildmedia.readthedocs.org/media/pdf/pillow/latest/pillow.pdf (accessed on 30 September 2025).
46. Marcel, S.; Rodriguez, Y. Torchvision the machine-vision package of torch. In Proceedings of the 18th ACM International Conference on Multimedia, Firenze, Italy, 25–29 October 2010; MM '10; ACM: New York, NY, USA; pp. 1485–1488. [CrossRef]
47. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. *arXiv* **2016**, arXiv:1610.02391. [CrossRef]
48. Tan, M.; Le, Q.V. EfficientNet: Rethinking model scaling for Convolutional Neural Networks. *arXiv* **2019**, arXiv:1905.11946. [CrossRef]
49. Liu, Z.; Mao, H.; Wu, C.Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A ConvNet for the 2020s. *arXiv* **2022**, arXiv:2201.03545. [CrossRef]

50. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer using shifted windows. *arXiv* **2021**, arXiv:2103.14030. [CrossRef]
51. Zhu, L.; Liao, B.; Zhang, Q.; Wang, X.; Liu, W.; Wang, X. Vision Mamba: Efficient visual representation learning with bidirectional state space model. *arXiv* **2024**, arXiv:2401.09417. [CrossRef]

Chapter 4

Conclusions

This chapter presents the summary and conclusions of the performed research. The key findings from the analysis of bacteria and stomata image datasets are consolidated in this section.

4.1 Bacteria research summary

The research on classification of soil bacteria was performed on two balanced datasets. First dataset consisted of 128 images of *Enterobacter*, *Rhizobium*, *Pantoea*, *Bradyrhizobium*, and *Pseudomonas* soil bacteria genera prepared thoroughly by the microbiologists. Second set considered of over 3,000 images of *Enterobacter*, *Rhizobium*, *Pantoea*, and *Pseudomonas* soil bacteria genera prepared automatically in a tiled-manner approach by IT specialists which were instructed on microscope usage by the biology expert. The first set seems to be very limited comparing to the second one; however, the considered images were in a high resolution and included thousands of bacteria instances in it. The computations on the first set were not only performed on whole images, but also the features were derived from subimages or computed on bacteria instances which justifies that dataset is sufficient for drawing significant conclusions.

The research performed on the microbiologists' dataset (first dataset) combined computation of various types of handcrafted features including texture, geometry, color, and dispersion. Other computations included deep features generated by Convolutional Neural Networks. The classification methods included ones which required handcrafted features as input: SVM, RF, KNN, ELM, MLP, and ELM-RBF. ResNet architectures (which are CNNs) were also used to classify this dataset computing deep features automatically.

The highest classification performance obtained on first dataset was equal to 97.07% ACC yielded with KNN, along with accuracy 97.03% obtained by ELM-RBF on a set of texture and shape features combined. Accuracy equal to 95.6 % was gained with KNN when color features were employed; however, due to the fact that preparing the first dataset we did not ensure consistency in light these computation were performed solely as a reference point. Combining geometry and dispersion features together, the highest yielded accuracy was equal to 85.14 % with Random Forest. Residual Neural Networks were also applied to classify this dataset. The highest accuracy was obtained with ResNet152V2 and ResNet101V2 when subimages were applied and yielded 92% ACC. These results present that the best performance is not necessarily obtained by more advances machine learning methods. The classical approach which required computation

of handcrafted features outperformed advanced deep learning architectures on the considered dataset. One could assume that this was caused due to insufficient number of training data for deep learning method; however, subimages extraction process yielded a set consisting of 6400 subimages which is adequate for a proper analysis.

In the research performed on second dataset, which was acquired automatically, the images were taken under different sets of conditions including various glass and light types. The considered image features were both handcrafted and automatically computed. The handcrafted features considered the color and texture. These features were applied as input to selected classifiers: SVM, RF, KNN, MLP, and our proposed modifications of ELM-RBF. The highest classification result was obtained with the ELM-RBF with k-medoids and yielded 95.54% on whole dataset. However, this result was outperformed with ResNet152V2 which yielded 99% accuracy.

Various experiments were performed when model was trained on one subset of images and tested on another subset. These experiments showed that testing the model on another set yields in significant decrease of accuracy. The problem in the automatically created dataset is that some images are out of focus which is caused by different thickness of the bacteria smear which could not be solved with autofocus available in the microscope. Additionally, when the second set of samples was prepared in the same conditions - the same light and glass - the accuracy also decreased. This suggests a need for a standardized preparation methodology to ensure normalized variable values including uniform sample thickness. The experiments revealed an impact of light condition on classification accuracy.

Conclusions regarding classification of soil bacteria:

- Comprehensive experiments were conducted on dedicated, expertly prepared datasets of microscopic soil bacteria images (five or four genera), addressing a gap in the literature which has primarily focused on other types of bacteria (e.g., pathogenic). It was experimentally demonstrated that computer classification of soil bacteria is feasible with high accuracy (up to 99%).

- Classical machine learning algorithms (e.g., KNN, ELM-RBF) based on handcrafted features (texture, shape) achieved higher accuracy (up to 97.07%) than advanced neural network architectures (ResNet), when applied to the first dataset (128 images). This demonstrates the efficiency of traditional methods when dealing with a dataset prepared solely by microbiologists.

- Deep learning (ResNet152V2) achieved significantly higher accuracy (99%) than classical methods (95.54% for ELM-RBF) on a large, automatically generated dataset (>3,000 images).

- The research revealed that bacteria classification model performance relies on considered datasets, their preparation and methodology of image acquisition:

  - The dataset was either prepared manually by the microbiology experts or acquired in an automated manner. The automatically prepared dataset included out-of-focus images due to variations in smear thickness, which were retained as a key factor reducing model robustness.
  - Variations in lighting and glass type, which lead to a drop in model accuracy.

  - Models trained on one automatically acquired dataset lost effectiveness when tested on another, even under identical light conditions, indicating the necessity for developing an automated and standardized sample preparation methodology.

- A modification of the ELM-RBF method was proposed and validated, involving the replacement of the default k-means algorithm with k-medoids, which improved its performance on benchmark datasets. The modified method was successfully applied to the classification of automatically acquired bacteria images, achieving the best result (95.54% ACC) among methods based on handcrafted features, constituting its practical validation.

4.2 Stomata research summary

In ploidy level classification, the computations were performed on two balanced sets of 1500 images. Both datasets contained three classes of blackcurrant cultivars. The first dataset consisted of: diploid 'Gofert' and 'Polares' cultivars, triploid 'Dlinnokistnaja', autotetraploid clone of 'Gofert' cultivar, and 'Polares' cultivar - leaves were sampled for photo documentation at the end of a vegetation season. The second dataset contained diploid 'Gofert' cultivar, triploid 'Dlinnokistnaja' cultivar, and autotetraploid clone of 'Gofert', where leaves were sampled for photo documentation at the beginning of the vegetation season.

The computations performed on first dataset included dataset modifications applied to reduce impact of color on classification. The images were segmented with YOLOv8 algorithm. Classical machine learning methods, KNN, SVM, RF, and MLP, were applied to perform classification with geometric features. The highest accuracy was obtained with RF 75.8%. The impact of color could be responsible for high classification accuracy equal to 97.3% when ResNet152V2 was employed on raw dataset. Notably, when the images were processed to reduce impact of color, the accuracy decreased. ResNet learned to distinguish ploidy levels based on binary masks with 78.7% accuracy.

Further computations were performed on both datasets together. The classification was conducted on whole images which were not processed. The employed classification methods were Convolutional Neural Networks (ResNet152V2) and Vision Transformers (*vit-large-patch16-224-in21k* (vit-l) and *vit-base-patch16-224-in21k* (vit-b) architectures). The models were trained on one dataset and tested on the second dataset to verify their performance on separate data. Although, in classification of three classes, ResNet outperformed ViT yielding 68%, in two class classification (diploid vs tetraploid) the highest accuracy equal to 88% was obtained with vit-b. Explainable AI method Grad-CAM was applied to both ResNet and ViT. Vision Transformers demonstrated biologically interpretable behavior, focusing on stomata structures which shows its potential application in similar stomata-related problems.

In conclusion, the research on stomata classification yielded the following principal findings:

- For the first time in the literature, artificial intelligence methods were applied to classify plant ploidy levels based on microscopic stomata images and yielded satisfactory results, opening a new research direction in plant phenotyping.

- Vision Transformers were employed for the first time to stomata-related tasks obtaining promising results, paving the way for the broader application of this architecture in various computer vision tasks on stomata images.
- Both Vision Transformer and Convolutional Neural Network architectures proved effective in classifying plant ploidy levels based on stomata images.
- While the ResNet network achieved higher accuracy (68%) in three-class classification, the Vision Transformer performed better (88% accuracy) in binary classification (diploid vs. tetraploid). A key advantage of the ViT model, revealed by the XAI method (Grad-CAM), is its biologically interpretable behavior - the model focused its attention on stomata structures, indicating greater potential for tasks based on specific morphological features.
- The results clearly demonstrate that classification effectiveness is strongly dependent on input data characteristics:
  - The ResNet network, trained on unprocessed images, achieved very high accuracy (97.3%), which dropped (to 78.7%) when the influence of color was reduced (binary mask). This suggests that the model may have initially relied on color artifacts rather than stomata shape.
  - While testing models trained on one dataset (e.g., end of season) on another (beginning of season) led to an accuracy drop, the performance remained high (e.g., up to 88% in binary classification). This demonstrates the model's robustness while simultaneously highlighting the challenge of generalization and the undeniable impact of the vegetation stage on data characteristics.

Chapter 5

Future work and extensions

The bacteria classification research can be extended by application of similar methodology to a more diverse dataset. The access to datasets including bacteria and other microorganisms is very limited. There is a high research potential in developing a dataset with thousands of bacteria images which could be then processed with AI methods proposed in this dissertation. The automatically created dataset considered in this research contains many images which are out of focus due to the problem with thickness of the smear layer, which could not be solved with autofocus available in the microscope. Due to this fact, the approach to automatically prepare and then photograph samples in uniform matter should be developed which requires close cooperation of microbiologists, IT specialists and mechatronics. It is also essential to ensure sustain conditions (including selected medium), when preparing bacteria samples as the differences could influence their morphology. The research can also be extended to mixed datasets consisting of various bacteria genera on single images.

The research performed on both considered bacteria datasets included application of classical machine learning methods and Convolutional Neural Networks. Vision Transformers could also be applied to these datasets to compare the obtained results. The research on application of ViTs to microscopic images of bacteria is very limited in the literature.

In research focusing on stomata images, there are multiple issues to be solved. In none of these issues were the ViTs applied. In this research, it is the first application of ViTs to stomata microscopic images in any context. The potential of ViTs in stomata images analysis is confirmed with the biologically interpretable behavior in decision making-process presented by Grad-CAM. We suggest to apply ViTs to tasks such as species classification, classification of presence or absence of a biofertilizer application, classification of open and closed stomata pores. The ploidy level classification research can be also extended with further application of ViTs on other data sets including various blackcurrant genera and various plant species.

The research on classification of ploidy level in this dissertation was considered only analyzing images of blackcurrant. This research should be extended to other species to verify, whether AI methods are capable to distinguish ploidy levels in various species and not solely on blackcurrant.

It is important to note, that the color of plant samples changes according to amount of chlorophyll, which changes overtime. At the beginning of vegetation season leaves tend to be green and healthy, while at the end of this season they become brownish. Plant sheds its leaves to prepare itself for periods where access to sunlight is reduced, the leaves

should be removed to reduce the lose of water. Important differences are also observed in appearance of stomata cells. These cells tend to become reduced in size and damaged at the end of vegetation season. It is important to take these factors into account while preparing the dataset. It should be ensured that samples are prepared at the same time of vegetation season. Another approach could be to purposefully build dataset consisting of various vegetation periods to create a mixed set. The models trained on this set could make predictions in any period of vegetation season.

The research on soil bacteria classification with microscopic images and on ploidy level classification based on stomata images could also be extended by application of other state-of-the-art AI methods.

Post-Transformer architectures could be employed e.g., Mamba or ConvNext. Mamba is an alternative architecture based on State Space Models and is employed in domains such as medicine [145] for tasks including microscopic image segmentation of organ cells or image classification of tissues microscopic images. ConvNext architecture emerged as a modification of ResNet applying selected concepts of Transformers. In [146], ConvNext is employed for classification of weed on macroscopic images. In [147], the microscopic blood images this architecture is employed to classify microscopic blood smear images as either containing malaria parasites or being uninfected.

Self-supervised learning methods including DINO could also be applied. DINO is based on Transformer architecture. This method is employed to biological domain including such examples as classification of microscopic images of parasitic eggs in [148] or detection bacteria growth on sequential data in [149].

The data augmentation can be performed with Generative AI including such algorithms as Stable Difussion [75] or ControlNet [76]. An example of Stable Diffusion application is presented in [150], where it is applied to generate microscopic images of microtopographies. ControlNet is employed in [151] for generation of medical images.

Masked Autoencoders could also be employed to reconstruct microscopic images enlarging the dataset by new artificially created images. In [152], Mask Autoencoders were applied to learn the representation of human cells on microscopic images in self-supervised learning approach. Mask Autoencoders were applied in [153] for augmentation of datasets: ImageNet, STL, and miniImageNet.

Explainable AI methods could also be applied in further research. The Grad-CAM [154], employed to stomata microscopic images, could be performed on bacteria classification models. In further research, other methods based on Saliency Maps including: Grad-CAM++, Score-CAM, or LayerCAM could be applied. Grad-CAM++ is a method that extends Grad-CAM algorithm to ensure better localization of whole objects along with explaining occurrences of multiple objects of a class in a single image [155]. Score-CAM analysis is based on class activation mapping [156]. Layer-CAM is a gradient-based method which applies ReLU function to gradients pixel-wise gradients and uses them to weight activation maps, rather than of computing mean value for channels like Grad-CAM. This method is applied when image precise localization is significant and can be applied to both deep and shallow network layers [157]. Explainable AI methods, which are based on image modifications and verification how the model estimation changes, could also be applied. The LIME [158] method is a simple linear model which generates various perturbed images and analyses how these actions impact the prediction on super-pixel level. RISE algorithm [159] generates thousands of random binary masks which cover the considered image. Each mask is analyzed by the model and output map is a weighted sum of all the analyzed masks. The pixel-level output yields a precise map.

Appendix: Research curriculum vitae

ORCID: 0000-0003-1730-5866

Education

- Master of Science in Engineering (MSc Eng) in Computer Science, graduated with honors from the Warsaw University of Life Sciences (WULS–SGGW) in 2021,
- Bachelor's degree (BSc) in Computer Science and Econometrics from the Warsaw University of Life Sciences in 2019.

Professional career

- Research and Teaching Assistant (a primary place of work), Institute of Information Technology, Warsaw University of Life Sciences, since 2021.

Publications included in dissertation

1. **A. Konopka**, K. Struniawski, R. Kozera, P. Trzciński, L. Sas-Paszt, A. Lisek, K. Górnik, E. Derkowska, S. Głuszek, B. Sumorok, and M. Frąc, "Classification of soil bacteria based on machine learning and image processing," in *Computational Science – ICCS 2022: 22nd International Conference, London, UK, June 21–23, 2022, Proceedings, Part III, Lecture Notes in Artificial Intelligence*, vol. 13352, pp. 263–277, Springer, 2022. https://doi.org/10.1007/978-3-031-08757-8_23 (140 MNiSW points)
2. **A. Konopka**, R. Kozera, L. Sas-Paszt, P. Trzcinski, and A. Lisek, "Identification of the selected soil bacteria genera based on their geometric and dispersion features," *PLoS ONE*, vol. 18, no. 10, e0293362, pp. 1–11, 2023. https://doi.org/10.1371/journal.pone.0293362 (100 MNiSW points, IF – 2.9)
3. **A. Konopka**, K. Struniawski, and R. Kozera, "Performance analysis of Residual Neural Networks in soil bacteria microscopic image classification," in *Modelling and Simulation'2023: The 2023 European Simulation and Modelling Conference, October 24–26, 2023, Toulouse, France*, pp. 144–149, EUROSIS-ETI, 2023. (70 MNiSW points)
4. **A. Konopka**, K. Struniawski, and R. Kozera, "Classification performance of Extreme Learning Machine Radial Basis Function with k-means, k-medoids and mean

shift clustering algorithms," in *Computational Science – ICCS 2023: 23rd International Conference, Prague, Czech Republic, July 3–5, 2023, Proceedings, Part IV, Lecture Notes in Artificial Intelligence*, vol. 10476, pp. 171–186, Springer, 2023. https://doi.org/10.1007/978-3-031-36027-5_13 (140 MNiSW points)

5. **A. Konopka**, R. Kozera, L. Sas-Paszt, and P. Trzciński, "Automated imaging and machine learning for soil bacteria classification: Challenges and insights," *Engineering Applications of Artificial Intelligence*, vol. 159C, pp. 1–9, 2025. https://doi.org/10.1016/j.engappai.2025.111369 (140 MNiSW points, IF – 8.0)

6. **A. Konopka**, K. Struniawski, R. Kozera, L. Ortenzi, A. Marasek-Ciołakowska, and A. Machlańska, "Deep learning classification of blackcurrant genotypes by ploidy levels on stomata microscopic images," in *Computational Science – ICCS 2025 Workshops: 25th International Conference, Singapore, Singapore, July 7–9, 2025, Proceedings, Part III, Lecture Notes in Artificial Intelligence*, pp. 135–148, Springer, 2025. https://doi.org/10.1007/978-3-031-97564-6_11 (140 MNiSW points)

7. **A. Konopka**, R. Kozera, A. Marasek-Ciołakowska, and A. Machlańska, "Classification of blackcurrant genotypes by ploidy levels on stomata microscopic images with deep learning: Convolutional Neural Networks and Vision Transformers," *Applied Sciences-Basel*, vol. 15, no. 19, 10735, pp. 1–14, 2025. https://doi.org/10.3390/app151910735 (100 MNiSW points, IF – 2.5)

Other publications

1. K. Struniawski, **A. Konopka**, and R. Kozera, "Identification of soil bacteria with machine learning and image processing techniques applying single cells' region isolation," in *Modelling and Simulation 2022: The European Simulation and Modelling Conference 2022*, pp. 76–81, EUROSIS-ETI, 2022. (70 MNiSW points)

2. R. Wyszyński, K. Struniawski, and **A. Konopka**, "Augmented single instance-driven identification of fungal pathogens through the Convolutional Neural Networks," in *Modelling and Simulation'2024: The European Simulation and Modelling Conference ESM'2024*, pp. 79–83, EUROSIS-ETI, 2024. (70 MNiSW points)

3. K. Struniawski, R. Kozera, **A. Konopka**, L. Sas-Paszt, and A. Marasek-Ciolakowska, "Machine learning-driven soil fungi identification using automated imaging techniques," *Applied Sciences-Basel*, vol. 16, no. 2, pp. 1–28, 855, 2026, doi: 10.3390/ app16020855. (100 MNiSW points, IF – 2.5)

4. K. Struniawski, A. Machlańska, A. Marasek-Ciolakowska, and **A. Konopka**, "Automated pollen classification via subinstance recognition: A comprehensive comparison of classical and deep learning architectures," *Applied Sciences-Basel*, vol. 16, no. 2, 720, pp. 1–39, 2026. https://doi.org/10.3390/app16020720 (100 MNiSW points, IF – 2.5)

5. G. Szwachta[1], D. Sulskis[1], **A. Konopka**[1], E. Jalonicka, K. Struniawski, K. Mikalauskaite, A. Sakalauskas, R. Kozera, V. Smirnovas, V. Stsiapura, M. Ziaunys, and

P. Hańczyc, "Collective methodological emission assay of Thioflavin T for qualitative α-synuclein fibril structures discrimination," *International Journal of Biological Macromolecules*, vol. 335, 148994, pp. 1-9, 2025. https://doi.org/10.1016/j.ijbiomac.2025.148994 (100 MNiSW points, IF - 8.5)
[1]These authors contributed equally to this work.

6. J. Sawicka, P. Bollin, A. Sylla, M. Panasiuk, M. Wilkowska, L. Ciołek, M. Leśniewski, **A. Konopka**, K. Struniawski, G. Całka-Kuc, A. Liwo, P. Hańczyc, M. Kozak, B. Gromadzka, M. Biernat, and S. Rodziewicz-Motowidło, "Design and characterization of antibacterial peptide nanofibrils as components of composites for biomaterial applications," *Current Protein & Peptide Science*, vol. 26, no. 10, pp. 875-895, 2025. https://doi.org/10.2174/0113892037353453241219185311 (70 MNiSW points, IF - 2)

7. K. Struniawski, **A. Konopka**, and R. Kozera, "Credibility of randomness in Extreme Learning Machine," in *Trust and Artificial Intelligence: Development and Application of AI Technology*, Ch. 9, pp. 92-106, Routledge, 2025. https://doi.org/10.4324/9781032627236-10 (50 MNiSW points)

8. K. Struniawski, **A. Konopka**, and R. Kozera, "Performance evaluation of activation functions in Extreme Learning Machine," in *ESANN 2023 - European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, pp. 351-356, Ciaco, 2023. https://doi.org/10.14428/esann/2023.ES2023-31 (70 MNiSW points)

9. K. Struniawski, **A. Konopka**, and R. Kozera, "Metaheuristic algorithms in Extreme Learning Machine for selection of parameters in activation function," in *Modelling and Simulation'2023: The 2023 European Simulation and Modelling Conference*, pp. 239–244, EUROSIS-ETI, 2023. (70 MNiSW points)

10. K. Struniawski, **A. Konopka**, and R. Kozera, "Exploring apple silicon's potential from simulation and optimization perspective," in *Computational Science – ICCS 2024: 24th International Conference, Malaga, Spain, July 2–4, 2024, Proceedings, Part V, Lecture Notes in Artificial Intelligence*, vol. 14836, pp. 35-42, Springer, 2024. https://doi.org/10.1007/978-3-031-63775-9_3 (140 MNiSW points)

11. K. Struniawski, R. Kozera, and **A. Konopka**, "Performance of selected nature-inspired metaheuristic algorithms used for Extreme Learning Machine," in *Computational Science – ICCS 2023: 23rd International Conference, Prague, Czech Republic, July 3–5, 2023, Proceedings, Part III, Lecture Notes in Artificial Intelligence*, vol. 10475, pp. 498–512, Springer, 2023. https://doi.org/10.1007/978-3-031-36024-4_38 (140 MNiSW points)

12. K. Struniawski, R. Kozera, and **A. Konopka**, "Bias or justice? Analyzing LLM sentencing variability in theft indictments across gender, ethnicity, and education factors," in *Computational Science – ICCS 2025 Workshops: 25th International Conference, Singapore, Singapore, July 7–9, 2025, Proceedings, Part III, Lecture Notes in Artificial Intelligence*, vol. 15909, pp. 19-32, Springer, 2025. https://doi.org/10.1007/978-3-031-97564-6_2 (140 MNiSW points)

13. K. Struniawski, R. Kozera, and **A. Konopka**, "Analyzing LLM sentencing variability in theft indictments across gender, family status and the value of the stolen item," *Applied Sciences-Basel*, vol. 15, no. 16, 8860, pp. 1-17, 2025. https://doi.org/10.3390/app15168860 (100 MNiSW points, IF – 2.5)

    Another two manuscripts are currently under review.

Conference and seminar presentations

1. A. Konopka, K. Struniawski, and R. Kozera, "Performance analysis of Residual Neural Networks in soil bacteria microscopic image classification," in *Modelling and Simulation'2023: The 2023 European Simulation and Modelling Conference, October 24–26, 2023, Toulouse, France*, pp. 144–149, EUROSIS-ETI, 2023. (70 MNiSW points)

2. A. Konopka, K. Struniawski, and R. Kozera, "Classification performance of Extreme Learning Machine Radial Basis Function with k-means, k-medoids and mean shift clustering algorithms," in *Computational Science – ICCS 2023: 23rd International Conference, Prague, Czech Republic, July 3–5, 2023, Proceedings, Part IV, Lecture Notes in Artificial Intelligence*, vol. 10476, pp. 171–186, Springer, 2023. https://doi.org/10.1007/978-3-031-36027-5_13 (140 MNiSW points)

3. A. Konopka, K. Struniawski, R. Kozera, P. Trzciński, L. Sas-Paszt, A. Lisek, K. Górnik, E. Derkowska, S. Głuszek, B. Sumorok, and M. Frąc, "Classification of soil bacteria based on machine learning and image processing," in *Computational Science – ICCS 2022: 22nd International Conference, London, UK, June 21–23, 2022, Proceedings, Part III, Lecture Notes in Artificial Intelligence*, vol. 13352, pp. 263–277, Springer, 2022. https://doi.org/10. 1007/978-3-031-08757-8_23 (140 MNiSW points)

4. A. Konopka, "Identyfikacja bakterii glebowych z wykorzystaniem uczenia maszynowego i metod przetwarzania obrazu," *XI Konferencja Symbioza Techniki i Informatyki, Kiry, Poland, June 19–22*, 2023.

5. A. Konopka, "Klasyfikacja i identyfikacja wybranych bakterii glebowych," *Seminar of The National Institute of Horticultural Research*, Skierniewice, Poland, November 2, 2021.

6. A. Konopka, "Klasyfikacja i identyfikacja wybranych bakterii glebowych," *Seminar of the Department of Information Systems, Warsaw University of Life Sciences*, Warsaw, Poland, September 8, 2021.

Quantitative summary of scientific publications

Metric	Dissertation	Total
Number of Publications	7	20
Cumulative MNiSW Points	830	2050
Cumulative Impact Factor (IF)	13.4	31.4
Hirsch index (Scopus)		4

Table 5.1: Quantitative summary of scientific publications for 3.02.2026.

Cooperation with institutions

National collaboration

- Department of Biomedical Chemistry and Department of Theoretical Chemistry, Faculty of Chemistry, University of Gdańsk, Gdańsk, Poland
- Laboratory of Molecular and Cellular Nephrology, Department of Molecular Biotechnology, Faculty of Chemistry, Mossakowski Medical Research Institute Polish Academy of Sciences, Gdańsk, Poland
- Department of in vitro Studies, Institute of Biotechnology and Molecular Medicine, Gdańsk, Poland
- Faculty of Chemistry and Institute of Experimental Physics, Faculty of Physics, University of Warsaw, Warsaw, Poland
- Center of Cellular Immunotherapies, Warsaw University of Life Sciences, Warsaw, Poland
- Department of Macromolecular Physics, Faculty of Physics, Adam Mickiewicz University in Poznań, Poznań, Poland
- Department of Applied Biology and Department of Microbiology and Rhizosphere, The National Institute of Horticultural Research, Skierniewice, Poland
- Biomaterials Research Group, Łukasiewicz Research Network-Institute of Ceramics and Building Materials, Cracow, Poland
- Institute of Agrophysics, Polish Academy of Sciences, Lublin, Poland

International collaboration

- Department of Computer Science, Sapienza University of Rome, Italy
- Institute of Biotechnology, Life Sciences Center, Vilnius University, Lithuania
- Institute of Information Technology, Warsaw University of Life Sciences, Poland
- University of Pennsylvania, Philadelphia, USA

Research projects

- Principal Investigator, Preludium 2023, National Science Centre (NCN): *"Automated system for analysing the growth and development dynamics of germinating seeds"*. Status: passed the merit-based evaluation stage, not awarded.
- Principal Investigator, Preludium 2024, National Science Centre (NCN): *"Automated system for analysing the growth and development dynamics of germinating seeds"*. Status: advanced to the last stage, not awarded.
- Researcher, Interreg Brandenburg–Poland *"Oder Together": "Joint analyses and solutions for transboundary water quality management, biodiversity, fisheries and other ecosystem services"*. Status: awarded. Duration: 2025–2028.
- Team Member, NCN Sonata grant (2021/43/D/ST4/01741): *"Application of Thioflavin T laser emission and its analogues for detecting early symptoms of neurodegenerative diseases"*. Status: awarded.
- Team Member, Grant No. ZBS/7/2021, National Institute of Horticultural Research in Skierniewice. Status: awarded.
- Team Member, NCBiR BIOSTRATEG (BIOSTRATEG3/344433/16/NCBR/2018): *"New biotechnological solutions in the diagnosis, control and monitoring of key fungal pathogens in the organic cultivation of soft fruits"*. Status: awarded.

Awards

- Individual Rector's Award of WULS, 3rd degree, for scientific achievements (2024).

Promotion of science among students

- Co-authored a scientific publication (70 MNiSW points) in collaboration with WULS-SGGW computer science student, Rafał Wyszyński. Supervised and mentored Rafał Wyszyński, who received an NCN scholarship as part of the SONATA grant for his research contributions.

Peer review activities

- Peer reviewer for the journal *New Biotechnology* (ISSN: 1871-6784).
- Peer reviewer for the journal *Engineering Applications of Artificial Intelligence* (ISSN: 0952-1976) – two reviews.

Organizational activities

- Support in organizing *The 28th International Conference on Applications of Computer Algebra (ACA'2023)*, July 17–21, 2023, Warsaw, Poland and support in organizing *WULS–SGGW Days*.

Bibliography

[1] M. Suhag, "Potential of biofertilizers to replace chemical fertilizers," *International Advanced Research Journal in Science, Engineering and Technology*, vol. 3, no. 5, pp. 163–167, 2016.

[2] M. K. Mishra, "Stomatal characteristics at different ploidy levels in coffea l.," *Annals of Botany*, vol. 80, no. 5, pp. 689–692, 1997.

[3] W. Estrada-Prado, L. Chávez-Suárez, Y. C. Maceo-Ramos, E. Jerez-Mompie, and M. C. Nápoles-García, "Effect of Azofert-F on the stomatal response of beans to water deficit," *Agronomia Mesoamericana*, vol. 32, no. 2, pp. 442–451, 2021.

[4] N. Zeiditoolabi, I. Khammari, A. Sirousmehr, M. Daneshvar, M. Galavi, and M. Dahmardeh, "Evaluation of stomata in vetch-barley intercropping and its relationship with forage production in rainfed conditions, under the influence of biofertilizer and superabsorbent," *Gesunde Pflanzen*, vol. 75, no. 5, pp. 2045–2073, 2023.

[5] A. Waibel, "Phoneme recognition using time-delay neural networks," in *Meeting of the Institute of Electrical, Information and Communication Engineers (IEICE)*, 1987.

[6] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv*, 2021.

[7] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The Bulletin of Mathematical Biophysics*, vol. 5, no. 4, pp. 115–133, 1943.

[8] S. Sreerakuvandana, P. Pappachan, and V. Arya, *Understanding Large Language Models*, pp. 1–24. IGI Global, 2024.

[9] D. Naik, I. Naik, and N. Naik, *Leveraging the Use of ChatGPT: Exploring Its Real-World Applications Including Their Related Ethical and Regulatory Considerations*, pp. 649–667. Springer Nature Switzerland, 2024.

[10] W. Yan, J. Liu, and W. Liang, "Deepseek empowers general medicine: Potential application and prospect," *Chinese General Practice*, vol. 28, no. 17, pp. 2065–2069, 2025.

[11] F. Rosenblatt, "The perceptron: A perceiving and recognizing automaton," report, Project PARA, Cornell Aeronautical Laboratory, 1957.

[12] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychological Review*, vol. 65, no. 6, pp. 386–408, 1958.

[13] B. Widrow and M. E. Hoff, "Adaptive switching circuits," in *1960 IRE WESCON Convention Record*, vol. 4, pp. 96–104, IRE, 1960.

[14] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.

[15] C. Lemaréchal, *Cauchy and the gradient method*, pp. 251–254. EMS Press, 2012.

[16] J. Terven, D.-M. Cordova-Esparza, J.-A. Romero-González, A. Ramírez-Pedraza, and E. A. Chávez-Urbiola, "A comprehensive survey of loss functions and metrics in deep learning," *Artificial Intelligence Review*, vol. 58, no. 7, 2025.

[17] S. Amari, "A theory of adaptive pattern classifiers," *IEEE Transactions on Electronic Computers*, vol. EC-16, no. 3, pp. 299–307, 1967.

[18] H. Robbins and S. Monro, "A stochastic approximation method," *The Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400–407, 1951.

[19] S. Linnainmaa, "The representation of the cumulative rounding error of an algorithm as a taylor expansion of the local rounding errors," master's thesis, University of Helsinki, 1970.

[20] P. J. Werbos, "Applications of advances in nonlinear sensitivity analysis," in *System modeling and optimization*, vol. 38 of *Lecture Notes in Control and Information Sciences*, pp. 762–770, Springer, 1982.

[21] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[22] O. H. Rodriguez and J. M. Lopez Fernandez, "A semiotic reflection on the didactics of the chain rule," *The Mathematics Enthusiast*, vol. 7, no. 2–3, pp. 321–332, 2010.

[23] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.

[24] S. Bouraya and A. Belangour, "A comparative analysis of activation functions in neural networks: unveiling categories," *Bulletin of Electrical Engineering and Informatics*, vol. 13, pp. 3301–3308, 10 2024.

[25] W. H. Mubark and M. Y. Sarwar Uddin, "Analyzing the effects of input batching on the inference time of vision transformer models," in *2025 4th International Conference on Computing and Information Technology (ICCIT)*, pp. 203–209, IEEE, 2025.

[26] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *Journal of Machine Learning Research*, vol. 12, pp. 2121–2159, 2011.

[27] M. D. Zeiler, "Adadelta: An adaptive learning rate method," *arXiv*, 2012.

[28] T. Tieleman and G. Hinton, "Lecture 6.5 - rmsprop: Divide the gradient by a running average of its recent magnitude." Coursera: Neural Networks for Machine Learning, 2012. [Online; accessed 03-02-2026].

[29] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv*, 2014.

[30] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv*, 2017.

[31] A. Ozgur and F. Nar, "Effect of dropout layer on classical regression problems," in *2020 28th Signal Processing and Communications Applications Conference (SIU)*, pp. 1–4, IEEE, 2020.

[32] X. Ying, "An overview of overfitting and its solutions," *Journal of Physics: Conference Series*, vol. 1168, 2019.

[33] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, Massachusetts: MIT Press.

[34] L. Prechelt, *Early Stopping — But When?*, pp. 53–67. Springer Berlin Heidelberg, 2012.

[35] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv*, 2015.

[36] C. A. R. de Sousa, "An overview on weight initialization methods for feedforward neural networks," in *2016 International Joint Conference on Neural Networks (IJCNN)*, p. 52–59, IEEE, July 2016.

[37] A. Kumar and S. S. Sodhi, "Comparative analysis of gaussian filter, median filter and denoise autoenocoder," in *2020 7th International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 45–51, 2020.

[38] S. Abdul-Kader and A. I. Khalil, "Comparative study for edge detection of noisy image using sobel and laplace operators," *Journal of the College of Education for Women*, vol. 23, no. 11, 2012.

[39] W. R. Oliver, "Histogram stretching or histogram equalization in image processing," *Microscopy Today*, vol. 6, no. 3, pp. 20–24, 1998.

[40] A. S. Samra, S. E. T. G. Allah, and R. M. Ibrahim, "Face recognition using wavelet transform, fast fourier transform and discrete cosine transform," in *2003 46th Midwest Symposium on Circuits and Systems*, vol. 1, pp. 272–275, IEEE, 2003.

[41] X. Xu, S. Xu, L. Jin, and E. Song, "Characteristic analysis of otsu threshold and its applications," *Pattern Recognition Letters*, vol. 32, no. 7, pp. 956–961, 2011.

[42] D. Sundararajan, "Morphological image processing," in *Digital Image Processing: A Signal Processing and Algorithmic Approach*, pp. 217–256, Springer, 2017.

[43] C. Di Ruberto, G. Fodde, and L. Putzu, "On different colour spaces for medical colour image classification," in *International Conference on Computer Analysis of Images and Patterns*, pp. 477–488, Springer, 2015.

[44] M.-N. Pons and J. Dodds, "Particle shape characterization by image analysis," in *Progress in filtration and separation*, pp. 609–636, Elsevier, 2015.

[45] F. Park, "Shape descriptor/feature extraction techniques." Presentation at UCI iCAMP 2011, University of California, Irvine, slides, 42 pages, 2011.

[46] I. Grain, "Computer interpolation and contouring of two-dimensional data: A review," *Geoexploration*, vol. 8, no. 2, pp. 71–86, 1970.

[47] S. Gupta and D. S. Sisodia, "Automated detection of diabetic retinopathy from gray-scale fundus images using glcm and glrlm-based textural features—a comparative study," in *Intelligent Computing Techniques in Biomedical Imaging*, pp. 251–259, Elsevier, 2025.

[48] J. Burkardt, "K-means clustering," *Virginia Tech, Advanced Research Computing, Interdisciplinary Center for Applied Mathematics*, vol. 5, 2009.

[49] K. G. Derpanis, "Mean shift clustering." Lecture Notes, 2005. [Online; accessed 03-02-2026].

[50] M. A. Hall, *Correlation-based Feature Selection for Machine Learning*. Phd thesis, University of Waikato, 1999.

[51] L. Yu and H. Liu, "Feature selection for high-dimensional data: a fast correlation-based filter solution," in *Proceedings of the Twentieth International Conference on Machine Learning*, pp. 856–863, 2003.

[52] G. Cawley, N. Talbot, and M. Girolami, "Sparse multinomial logistic regression via bayesian l1 regularisation," vol. 19, pp. 209–216, 01 2006.

[53] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 36, no. 2, pp. 111–133, 1974.

[54] S. Geisser, "The predictive sample reuse method with applications," *Journal of the American Statistical Association*, vol. 70, no. 350, pp. 320–328, 1975.

[55] X. Zeng and T. R. Martinez, "Distribution-balanced stratified cross-validation for accuracy estimation," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 12, no. 1, pp. 1–12, 2000.

[56] D. H. Hubel and T. N. Wiesel, "Receptive fields of single neurones in the cat's striate cortex," *The Journal of Physiology*, vol. 148, no. 3, pp. 574–591, 1959.

[57] K. Fukushima, "Visual feature extraction by a multilayered network of analog threshold elements," *IEEE Transactions on Systems Science and Cybernetics*, vol. 5, no. 4, pp. 322–333, 1969.

[58] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological Cybernetics*, vol. 36, no. 4, pp. 193–202, 1980.

[59] I. Shadoul, R. Al-Hmouz, A. Hossen, M. Mesbah, and M. Deveci, "The effect of pooling parameters on the performance of convolution neural network," *Artificial Intelligence Review*, vol. 58, no. 9, 2025.

[60] E. Fix and J. L. Hodges, "Discriminatory analysis: Nonparametric discrimination: Consistency properties," Technical Report 4, USAF School of Aviation Medicine, 1951.

[61] T. K. Ho, "Random decision forests," in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, vol. 1, pp. 278–282, 1995.

[62] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[63] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009.

[64] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[65] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, pp. 84–90, May 2017.

[66] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," *arXiv*, 2013.

[67] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv*, 2014.

[68] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *arXiv*, 2014.

[69] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv*, 2015.

[70] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *arXiv*, 2015.

[71] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *arXiv*, 2016.

[72] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv*, 2017.

[73] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning* (K. Chaudhuri and R. Salakhutdinov, eds.), vol. 97 of *Proceedings of Machine Learning Research*, pp. 6105–6114, PMLR, 2019.

[74] A. Agarwal, A. Dash, G. Galbale, and S. P. Singh, "Analyzing impact of data augmentation techniques in computer vision," in *2025 2nd International Conference on Computational Intelligence, Communication Technology and Networking (CICTN)*, pp. 733–737, IEEE, 2025.

[75] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," *arXiv*, 2021.

[76] L. Zhang, A. Rao, and M. Agrawala, "Adding conditional control to text-to-image diffusion models," *arXiv*, 2023.

[77] G.-B. Huang, Q.-Y. Zhu, and C. Siew, "Extreme learning machine: A new learning scheme of feedforward neural networks," vol. 2, pp. 985–990, 2004.

[78] S. R. y Cajal, *Cajal's Histology of the Nervous System of Man and Vertebrates*. Oxford: Oxford University Press, 1995. Originally published in Spanish (1904). English translation.

[79] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the National Academy of Sciences*, vol. 79, no. 8, pp. 2554–2558, 1982.

[80] S. Hochreiter and J. Schmidhuber, "Long short-term memory," tech. rep., Technische Universität München, 1995.

[81] M. Schuster and K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.

[82] Y. Zhang, D. Yu, and G. Chen, *Advanced Recurrent Neural Networks for Automatic Speech Recognition*, pp. 261–279. Springer International Publishing, 2017.

[83] K. Baktha and B. K. Tripathy, "Investigation of recurrent neural networks in the field of sentiment analysis," in *2017 International Conference on Communication and Signal Processing (ICCSP)*, pp. 2047–2050, IEEE, 2017.

[84] G. Gkarmpounis, C. Vranis, N. Vretos, and P. Daras, "Survey on graph neural networks," *IEEE Access*, vol. 12, pp. 128816–128832, 2024.

[85] S. R. Raja, G. Natarajan, E. Elango, and S. Bose, *Exploring the Versatile Applications of Graph Neural Networks*, pp. 399–413. Springer Nature Switzerland, 2025.

[86] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 61–80, 2009.

[87] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, 2013.

[88] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv*, 2016.

[89] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *arXiv*, 2014.

[90] J. Zhou, "Research on generative adversarial networks and their applications in image generation," in *2022 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*, pp. 1144–1147, IEEE, 2022.

[91] T. Mukhiddin, W. Lee, S. Lee, and T. Rashid, "Research issues on generative adversarial networks and applications," in *2020 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pp. 487–488, IEEE, 2020.

[92] K. Jangde and R. Pandey, *Advancements in Real-World Applications of Generative Adversarial Networks (GANs)*, pp. 87–104. IGI Global, 2025.

[93] R. Razavi-Far, A. Ruiz-Garcia, and V. Palade, *An Introduction to Generative Adversarial Learning: Architectures and Applications*, pp. 1–6. Springer International Publishing, 2022.

[94] P. Wang, "This person does not exist," 2019.

[95] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *31st Conference on Neural Information Processing Systems (NIPS)*, vol. 30, 2017.

[96] B. R. K, B. E. Babu, S. P. Racharla, P. Pavani, Y. Balagoni, and M. Ajmeera, "A performance evaluation of transformer models and recurrent neural networks models in efficient text classification tasks," in *2025 8th International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 1171–1177, IEEE, 2025.

[97] Taufikin, M. Orosoo, M. Rengarajan, G. Fatma, D. Yuldashev, and I. I. Raj, "Applying transformer-based neural networks to corpus linguistic data for predictive text generation in multilingual environments," in *2024 International Conference on Emerging Smart Computing and Informatics (ESCI)*, pp. 1–6, IEEE, 2024.

[98] S. Ahmed, I. E. Nielsen, A. Tripathi, S. Siddiqui, R. P. Ramachandran, and G. Rasool, "Transformers in time-series analysis: A tutorial," *Circuits, Systems, and Signal Processing*, vol. 42, no. 12, pp. 7433–7466, 2023.

[99] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, Association for Computational Linguistics, 2019.

[100] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," technical report, OpenAI, 2018.

[101] S. Islam, H. Elmekki, A. Elsebai, J. Bentahar, N. Drawel, G. Rjoub, and W. Pedrycz, "A comprehensive survey on applications of transformers for deep learning tasks," *Expert Systems with Applications*, vol. 241, p. 122666, 2024.

[102] R. Courant, M. Edberg, N. Dufour, and V. Kalogeiton, *Transformers and Visual Transformers*, pp. 193–229. Springer US, 2023.

[103] "Gpt-5 new params and tools," 2025. [Online; accessed 18-11-2025].

[104] "Deepseek," 2025. [Online; accessed 18-11-2025].

[105] "Gemini," 2025. [Online; accessed 18-11-2025].

[106] T. Ronen, O. Levy, and A. Golbert, "Vision transformers with mixed-resolution tokenization," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 4613–4622, 2023.

[107] K. Sasaki, M. Watanabe, Y. Suda, A. Ishizuka, and N. Noparatnaraporn, "Applications of photosynthetic bacteria for medical fields," *Journal of Bioscience and Bioengineering*, vol. 100, no. 5, pp. 481–488, 2005.

[108] N. Akoijam, D. Kalita, and S. R. Joshi, *Bacteria and Their Industrial Importance*, pp. 63–79. Singapore: Springer Singapore, 2022.

[109] X. Raynaud and N. Nunan, "Spatial ecology of bacteria at the microscale in soil," *PLoS ONE*, vol. 9, no. 1, p. e87217, 2014.

[110] D. Roszak and R. Colwell, "Survival strategies of bacteria in the natural environment," *Microbiological Reviews*, vol. 51, no. 3, pp. 365–379, 1987.

[111] S. Doron and S. Gorbach, "Bacterial infections: Overview," in *International Encyclopedia of Public Health*, pp. 273–282, Academic Press, 2008.

[112] S. S. Sawant, S. M. Patil, V. Gupta, and N. K. Kunda, "Microbes as medicines: harnessing the power of bacteria in advancing cancer treatment," *International Journal of Molecular Sciences*, vol. 21, no. 20, p. 7575, 2020.

[113] O. O. Babalola, "Beneficial bacteria of agricultural importance," *Biotechnology letters*, vol. 32, no. 11, pp. 1559–1570, 2010.

[114] G. Kaur and M. S. Reddy, "Influence of p-solubilizing bacteria on crop yield and soil fertility at multilocational sites," *European Journal of Soil Biology*, vol. 61, pp. 35–40, 2014.

[115] J. Rui, J. Peng, and Y. Lu, "Succession of bacterial populations during plant residue decomposition in rice field soil," *Applied and Environmental Microbiology*, vol. 75, no. 14, pp. 4879–4886, 2009.

[116] E. Kaiser, A. Morales, J. Harbinson, J. Kromdijk, E. Heuvelink, and L. F. Marcelis, "Dynamic photosynthesis in different environmental conditions," *Journal of Experimental Botany*, vol. 66, no. 9, pp. 2415–2426, 2015.

[117] I. Zelitch and P. E. Waggoner, "Effect of chemical control of stomata on transpiration and photosynthesis," *Proceedings of the National Academy of Sciences*, vol. 48, no. 7, pp. 1101–1108, 1962.

[118] H. M. Naser, E.-H. Hanan, N. I. Elsheery, and H. M. Kalaji, "Effect of biofertilizers and putrescine amine on the physiological features and productivity of date palm (phoenix dactylifera, l.) grown on reclaimed-salinized soil," *Trees - Structure and Function*, vol. 30, no. 4, pp. 1149–1161, 2016.

[119] P. Chaudhary, S. Singh, A. Chaudhary, A. Sharma, and G. Kumar, "Overview of biofertilizers in crop production and stress management for sustainable agriculture," *Frontiers in Plant Science*, vol. 13, 2022.

[120] V. Fernández and T. Eichert, "Uptake of hydrophilic solutes through plant leaves: Current state of knowledge and perspectives of foliar fertilization," *Critical Reviews in Plant Sciences*, vol. 28, no. 1-2, pp. 36–68, 2009.

[121] J. Zhang, C. Cheng, F. Xiao, X. Zhang, C. Zhang, Y. Zhao, J. Xu, S. Zhang, and X. Wang, "Effects of ploidy level on leaf morphology, stomata, and anatomical structure of hibiscus syriacus l.," *BMC Plant Biology*, vol. 24, no. 1, 2024.

[122] C. Li, K. Wang, and N. Xu, "A survey for the applications of content-based microscopic image analysis in microorganism classification domains," *Artificial Intelligence Review*, vol. 51, pp. 577–646, 04 2019.

[123] M. Kruk, R. Kozera, S. Osowski, P. Trzciński, L. S. Paszt, B. Sumorok, and B. Borkowski, "Computerized classification system for the identification of soil microorganisms," in *AIP Conference Proceedings*, vol. 1648, p. 660018, AIP Publishing LLC, 2015.

[124] M. Kruk, R. Kozera, S. Osowski, P. Trzciński, L. Sas-Paszt, B. Sumorok, and B. Borkowski, "Computerized classification systemfor the identification of soil microorganisms," *Applied Mathematics & Information Sciences*, vol. 10, no. 1, pp. 21––31, 2016.

[125] J. Liu, F. Dazzo, O. Glagoleva, B. Yu, and A. Jain, "CMEIAS: A computer-aided system for the image analysis of bacterial morphotypes in microbial communities," *Microbial Ecology*, vol. 41, no. 3, pp. 173–194, 2001.

[126] B. Zieliński, A. Plichta, K. Misztal, P. Spurek, M. Brzychczy-Włoch, and D. Ochońska, "Deep learning approach to bacterial colony classification," *PLOS ONE*, vol. 12, no. 9, pp. 1–14, 2017.

[127] S. Kumar and G. Mittal, "Geometric and optical characteristics of five microorganisms for rapid detection using image processing," *Biosystems Engineering*, vol. 99, pp. 1–8, Jan. 2008.

[128] S. KUMAR and G. MITTAL, "Textural characteristics of five microorganisms for rapid detection using image processing," *Journal of Food Process Engineering*, vol. 32, no. 1, pp. 126–143, 2009.

[129] M. Bölter, R. Möller, and W. Dzomla, "Determination of bacterial biovolume with epifluorescence microscopy: Comparison of size distributions from image analysis and size classifications," *Micron*, vol. 24, no. 1, pp. 31–40, 1993.

[130] A. K. Jain and L. Hong, "Automatic classification of bacteria culture images," technical report, Michigan State University, 1996.

[131] H. Men, Y. Wu, Y. Gao, Z. Kou, Z. Xu, and S. Yang, "Application of support vector machine to heterotrophic bacteria colony recognition," in *2008 International Conference on Computer Science and Software Engineering*, vol. 1, pp. 830–833, 2008.

[132] P. Rani, S. Kotwal, J. Manhas, V. Sharma, and S. Sharma, "Machine learning and deep learning based computational approaches in automatic microorganisms image recognition: Methodologies, challenges, and developments," *Archives of Computational Methods in Engineering*, vol. 29, no. 3, pp. 1801–1837, 2021.

[133] M. F. Wahid, T. Ahmed, and M. A. Habib, "Classification of microscopic images of bacteria using deep convolutional neural network," in *2018 10th International Conference on Electrical and Computer Engineering (ICECE)*, pp. 217–220, 2018.

[134] P. Andreini, S. Bonechi, M. Bianchini, A. Mecocci, and F. Scarselli, "A deep learning approach to bacterial colony segmentation," in *Artificial Neural Networks and Machine Learning – ICANN 2018*, pp. 522–533, 2018.

[135] S. V, K. Prasad, C. Mukhopadhyay, and B. Banerjee, "Vision transformer based bacteria classification model for gram-stained direct smear images," *Multimedia Tools and Applications*, vol. 84, no. 19, pp. 20289–20309, 2025.

[136] G. D. Tan, U. Chaudhuri, S. Varela, N. Ahuja, and A. D. B. Leakey, "Machine learning-enabled computer vision for plant phenotyping: a primer on ai/ml and a case study on stomatal patterning," *Journal of Experimental Botany*, vol. 75, pp. 6683–6703, 10 2024.

[137] J. A. Gibbs and A. J. Burgess, "Application of deep learning for the analysis of stomata: a review of current methods and future directions," *Journal of Experimental Botany*, vol. 75, no. 21, pp. 6704–6718, 2024.

[138] S. N. Sultana, H. Park, S. H. Choi, H. Jo, J. T. Song, J.-D. Lee, and Y. J. Kang, "Optimizing the experimental method for stomata-profiling automation of soybean leaves based on deep learning," *Plants*, vol. 10, no. 12, 2021.

[139] Q. B. Kwong, Y. C. Wong, P. L. Lee, M. S. Sahaini, Y. T. Kon, H. Kulaveerasingam, and D. R. Appleton, "Automated stomata detection in oil palm with convolutional neural network," *Scientific Reports*, vol. 11, no. 1, 2021.

[140] H. Jayakody, P. Petrie, H. J. d. Boer, and M. Whitty, "A generalised approach for high-throughput instance segmentation of stomata in microscope images," *Plant Methods*, vol. 17, no. 1, 2021.

[141] J. A. Gibbs, L. Mcausland, C. A. Robles-Zazueta, E. H. Murchie, and A. J. Burgess, "A deep learning method for fully automatic stomatal morphometry and maximal conductance estimation," *Frontiers in Plant Science*, vol. 12, 2021.

[142] L. Costa, L. Archer, Y. Ampatzidis, L. Casteluci, G. A. P. Caurin, and U. Albrecht, "Determining leaf stomatal properties in citrus trees utilizing machine vision and artificial intelligence," *Precision Agriculture*, vol. 22, no. 4, pp. 1107–1119, 2020.

[143] B. Dey, R. Ahmed, J. Ferdous, M. M. U. Haque, R. Khatun, F. E. Hasan, and S. N. Uddin, "Automated plant species identification from the stomata images using deep neural network: A study of selected mangrove and freshwater swamp forest tree species of bangladesh," *Ecological Informatics*, vol. 75, p. 102128, 2023.

[144] A. H. Aono, J. S. Nagai, G. d. S. M. Dickel, R. C. Marinho, P. E. A. M. de Oliveira, J. P. Papa, and F. A. Faria, "A stomata classification and detection system in microscope images of maize cultivars," *PLOS ONE*, vol. 16, pp. 1–17, 10 2021.

[145] S. Bansal, S. A, M. P. J, M. S, S. Madisetty, M. Z. U. Rehman, C. S. Raghaw, G. Duggal, and N. Kumar, "A comprehensive survey of mamba architectures for medical image analysis: Classification, segmentation, restoration and beyond," *arXiv*, 2025.

[146] F. Benchallal, A. Hafiane, N. Ragot, and R. Canals, "ConvNeXt based semi-supervised approach with consistency regularization for weeds classification," *Expert Systems with Applications*, vol. 239, p. 122222, 2024.

[147] O. P. Mmileng, A. Whata, M. Olusanya, and S. Mhlongo, "Application of ConvNeXt with transfer learning and data augmentation for malaria parasite detection in resource-limited settings using microscopic images," *PLOS One*, vol. 20, no. 6, p. e0313734, 2025.

[148] N. Pinetsuksai, V. Kittichai, R. Jomtarak, K. Jaksukam, T. Tongloy, S. Boonsang, and S. Chuwongin, "Development of self-supervised learning with dinov2-distilled models for parasite classification in screening," in *2023 15th International Conference on Information Technology and Electrical Engineering (ICITEE)*, pp. 323–328, IEEE, 2023.

[149] M. H. Marino Miguélez, M. Osaid, E. Hallström, K. Kaya, J. Larsson, V. Kandavalli, C. Wählby, J. Elf, and W. van der Wijngaart, "Automatic culture-free detection of bacteria from blood for rapid sepsis diagnostics," 2024.

[150] S. Siemens, M. Kästner, and E. Reithmeier, "Synthetically generated microscope images of microtopographies using stable diffusion," in *Automated Visual Inspection and Machine Vision V*, p. 7, SPIE, 2023.

[151] K. Qiu, Z. Zhou, and Y. Guo, "Adaptively distilled controlnet: Accelerated training and superior sampling for medical image synthesis," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2025*, pp. 55–65, Springer Nature Switzerland, 2026.

[152] O. Kraus, K. Kenyon-Dean, S. Saberian, M. Fallah, P. McLean, J. Leung, V. Sharma, A. Khan, J. Balakrishnan, S. Celik, M. Sypetkowski, C. V. Cheng, K. Morse, M. Makes, B. Mabey, and B. Earnshaw, "Masked autoencoders are scalable learners of cellular morphology," *arXiv*, 2023.

[153] H. Xu, S. Ding, M. Zhao, and D. Jiang, "Masked autoencoders are robust data augmentors," *arXiv*, 2025.

[154] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," *International Journal of Computer Vision*, vol. 128, no. 2, pp. 336–359, 2019.

[155] A. Chattopadhay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, "Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 839–847, 2018.

[156] H. Wang, Z. Wang, M. Du, F. Yang, Z. Zhang, S. Ding, P. Mardziel, and X. Hu, "Score-cam: Score-weighted visual explanations for convolutional neural networks," *arXiv*, 2020.

[157] P.-T. Jiang, C.-B. Zhang, Q. Hou, M.-M. Cheng, and Y. Wei, "Layercam: Exploring hierarchical class activation maps for localization," *IEEE Transactions on Image Processing*, vol. 30, pp. 5875–5888, 2021.

[158] M. T. Ribeiro, S. Singh, and C. Guestrin, ""Why should i trust you?": Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144, Association for Computing Machinery, 2016.

[159] V. Petsiuk, A. Das, and K. Saenko, "RISE: Randomized input sampling for explanation of black-box models," *arXiv*, 2018.

Warszawa, 24.11.2025

mgr inż. Aleksandra Konopka
aleksandra_konopka@sggw.edu.pl

asystent badawczo-dydaktyczny

Szkoła Główna Gospodarstwa

Wiejskiego w Warszawie

Rada Dyscypliny Informatyka Techniczna i Telekomunikacja

Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie

Oświadczenie o współautorstwie

Niniejszym oświadczamy, że w pracy Konopka, A., Struniawski, K., Kozera, R., Trzciński, P., Sas-Paszt, L., Lisek, A., Górnik, K., Derkowska, E., Głuszek, S., Sumorok, B., Frąc, M. (2022). Classification of soil bacteria based on machine learning and image processing, Computational Science – ICCS 2022, 22nd International Conference, indywidualny udział poszczególnych autorów jest zgodny z poniższym opisem.

Wkład Aleksandry Konopki:

- oprogramowanie w zakresie segmentacji obszaru zainteresowania w celu wyodrębnienia konkretnych komórek bakterii, obliczanie cech opartych na kształcie, planowanie i przeprowadzanie eksperymentów klasyfikacyjnych z wykorzystaniem klasycznych metod uczenia maszynowego na cechach opartych na kształcie i teksturze, stosowania metod selekcji cech,
- napisanie następujących sekcji lub podsekcji oryginalnego tekstu: streszczenie, wstęp, schemat przepływu pracy, segmentacja obszaru zainteresowania, wyliczanie cech, selekcja cech, rozpoznawanie klas, cechy oparte na kształcie, KNN, eksperymenty i wyniki, wnioski,
- recenzowanie i redagowanie całego manuskryptu,
- zarządzanie danymi,
- konceptualizacja,
- opracowanie metodologii.

Wkład pozostałych autorów:

- Karol Struniawski – oprogramowanie – Extreme Learning Machine i Extreme Learning Machine RBF, obliczanie cech opartych na teksturze, napisanie sekcji oryginalnego

tekstu – cechy oparte na teksturze, ELM-RBF oraz recenzowanie i redagowanie całego manuskryptu,

- Ryszard Kozera – nadzór merytoryczny nad stroną informatyczną, recenzowanie i redagowanie całego manuskryptu, zwłaszcza w kontekście weryfikacji poprawności zapisu matematycznego,
- Paweł Trzciński – przygotowanie zbioru danych,
- Lidia Sas-Paszt – nadzór merytoryczny nad aspektami biologicznymi, recenzowanie i redagowanie aspektów biologicznych tekstu,
- Anna Lisek, Krzysztof Górnik, Edyta Derkowska, Sławomir Głuszek, Beata Sumorok i Magdalena Frąc – recenzowanie i redagowanie aspektów biologicznych tekstu.

Procentowy wkład autorów: Aleksandra Konopka – 60%, Karol Struniawski – 10%, Ryszard Kozera – 10%, Paweł Trzciński – 8%, Lidia Sas-Paszt – 6%, Anna Lisek – 1%, Krzysztof Górnik – 1%, Edyta Derkowska – 1%, Sławomir Głuszek – 1%, Beata Sumorok – 1% i Magdalena Frąc – 1%.

Aleksandra Konopka ..... Karol Struniawski ..... Ryszard Kozera

PAWEŁ TRZCIŃSKI ..... Lidia Sas-Paszt .....

Paweł Trzciński Lidia Sas-Paszt

Warszawa, 24.11.2025

mgr inż. Aleksandra Konopka
aleksandra_konopka@sggw.edu.pl

asystent badawczo-dydaktyczny

Szkoła Główna Gospodarstwa

Wiejskiego w Warszawie

Rada Dyscypliny Informatyka Techniczna i Telekomunikacja

Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie

Oświadczenie o współautorstwie

Niniejszym oświadczamy, że w pracy Konopka, A., Kozera, R., Sas-Paszt, L., Trzcinski, P., & Lisek, A. (2023). Identification of the selected soil bacteria genera based on their geometric and dispersion features. PLoS ONE, 18, 1–11. https://doi.org/10.1371/journal.pone.0293362, indywidualny udział poszczególnych autorów jest zgodny z poniższym opisem.

Wkład Aleksandry Konopki:

- konceptualizacja,
- zarządzanie danymi,
- analiza formalna,
- opracowanie metodologii,
- oprogramowanie,
- wizualizacja,
- napisanie wersji oryginalnej tekstu, redagowanie manuskryptu.

Wkład pozostałych autorów:

- Ryszard Kozera – nadzór merytoryczny nad częścią IT, współpraca przy koncepcji w kontekście metod interpolacji, współpraca przy pisaniu wstępnej wersji manuskryptu w kontekście przedstawiania cech geometrycznych z użyciem odpowiednich wzorów matematycznych, recenzowanie i redagowanie tekstu,
- Lidia Sas-Paszt – nadzór merytoryczny nad aspektami biologicznymi, zasoby,
- Paweł Trzciński – zasoby, weryfikacja opisów biologicznych,
- Anna Lisek – zasoby, weryfikacja opisów biologicznych.

Procentowy wkład autorów: Aleksandra Konopka – 75%, Ryszard Kozera – 10%, Lidia Sas-Paszt – 8%, Paweł Trzciński – 5%, Anna Lisek – 2%.

Aleksandra Konopka ..... Aleksandra Konopka

..... Ryszard Kozera

..... Lidia Sas-Paszt

PAWEŁ TRZCIŃSKI ..... Paweł Trzciński

Anna Lisek ..... Anna Lisek

Warszawa, 24.11.2025

mgr inż. Aleksandra Konopka
aleksandra_konopka@sggw.edu.pl

asystent badawczo-dydaktyczny

Szkoła Główna Gospodarstwa

Wiejskiego w Warszawie

Rada Dyscypliny Informatyka Techniczna i Telekomunikacja

Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie

Oświadczenie o współautorstwie

Niniejszym oświadczamy, że w pracy Konopka, A., Struniawski, K., & Kozera, R. (2023). Performance Analysis of Residual Neural Networks in Soil Bacteria Microscopic Image Classification. Modelling and Simulation'2023. The 2023 European Simulation and Modelling Conference (s. 144–149). EUROSIS-ETI, indywidualny udział poszczególnych autorów jest zgodny z poniższym opisem.

Wkład Aleksandry Konopki:

- konceptualizacja
- zarządzanie danymi
- analiza formalna
- opracowanie metodologii
- oprogramowanie
- wizualizacja
- przygotowanie wersji oryginalnej manuskryptu, redagowanie tekstu

Wkład pozostałych autorów:

- Karol Struniawski – oprogramowanie – wsparcie w zastosowaniu modeli ResNet, recenzowanie i redagowanie tekstu
- Ryszard Kozera – nadzór merytoryczny, recenzowanie i redagowanie tekstu

Procentowy udział autorów: Aleksandra Konopka – 80%, Karol Struniawski – 10%, Ryszard Kozera – 10%

Aleksandra Konopka Karol Struniawski Ryszard Kozera

Warszawa, 24.11.2025

mgr inż. Aleksandra Konopka
aleksandra_konopka@sggw.edu.pl

asystent badawczo-dydaktyczny

Szkoła Główna Gospodarstwa

Wiejskiego w Warszawie

Rada Dyscypliny Informatyka Techniczna i Telekomunikacja

Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie

Oświadczenie o współautorstwie

Niniejszym oświadczamy, że w pracy Konopka, A., Struniawski, K., \& Kozera, R. (2023). Classification Performance of Extreme Learning Machine Radial Basis Function with K-means, K-medoids and Mean Shift Clustering Algorithms, Computational Science – ICCS 2023, 23rd International Conference, Prague, Czech Republic, July 3–5, 2023, Proceedings, Part IV (T. 10476, s. 171–186). https://doi.org/10.1007/978-3-031-36027-5_13, indywidualny udział poszczególnych autorów jest zgodny z poniższym opisem.

Wkład Aleksandry Konopki:

- konceptualizacja
- zarządzanie danymi
- analiza formalna
- opracowanie metodologii
- oprogramowanie
- wizualizacja
- przygotowanie wersji oryginalnej manuskryptu, redagowanie

Wkład pozostałych autorów:

- Karol Struniawski - oprogramowanie - wsparcie przy implementacji podstawowej wersji Extreme Learning Machine Radial Basis Function, recenzowanie i redagowanie pracy
- Ryszard Kozera – nadzór merytoryczny, recenzowanie i redagowanie pracy

Procentowy wkład autorów: Aleksandra Konopka - 80%, Karol Struniawski - 10% i Ryszard Kozera - 10%

Aleksandra Konopka ..... Karol Struniawski ..... Ryszard Kozera

Warszawa, 24.11.2025

mgr inż. Aleksandra Konopka
aleksandra_konopka@sggw.edu.pl

asystent badawczo-dydaktyczny

Szkoła Główna Gospodarstwa

Wiejskiego w Warszawie

Rada Dyscypliny Informatyka Techniczna i Telekomunikacja

Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie

Oświadczenie o współautorstwie

Niniejszym oświadczamy, że w pracy Konopka, A., Kozera, R., Sas-Paszt, L., Trzciński, P. (2025). Automated imaging and machine learning for soil bacteria classification: Challenges and insights. Engineering Applications of Artificial Intelligence. Indywidualny udział poszczególnych autorów jest zgodny z poniższym opisem.

Wkład Aleksandry Konopki:

- konceptualizacja,
- zarządzanie danymi,
- analiza formalna,
- opracowanie metodologii,
- oprogramowanie,
- wizualizacja,
- pisanie - wersja wstępna, recenzowanie i redagowanie,
- przygotowanie preparatów i wykonywanie zdjęć mikroskopowych.

Wkład pozostałych autorów:

- Ryszard Kozera - nadzór, recenzowanie i redagowanie tekstu, analiza formalna,
- Lidia Sas-Paszt - recenzowanie i redagowanie tekstu, nadzór, koncepcja,
- Paweł Trzciński - recenzowanie i redagowanie tekstu, zasoby - dostarczenie szczepów bakterii reprezentujących różne rodzaje.

Procentowy wkład autorów: Aleksandra Konopka - 80%, Ryszard Kozera - 8%, Lidia Sas-Paszt - 7%, Paweł Trzciński - 5%

Aleksandra Konopka	Ryszard Kozera	Lidia Sas-Paszt	Paweł Trzciński
Aleksandra Konopka		Lidia Sas-Paszt	PAWEŁ TRZCIŃSKI

Warszawa, 24.11.2025

mgr inż. Aleksandra Konopka
aleksandra_konopka@sggw.edu.pl

asystent badawczo-dydaktyczny

Szkoła Główna Gospodarstwa

Wiejskiego w Warszawie

Rada Dyscypliny Informatyka Techniczna i Telekomunikacja

Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie

Oświadczenie o współautorstwie

Niniejszym oświadczamy, że w pracy Konopka, A., Struniawski, K., Kozera, R., Ortenzi, L., Marasek-Ciołakowska, A., & Machlańska, A. (2025). Deep Learning Classification of Blackcurrant Genotypes by Ploidy Levels on Stomata Microscopic Images. Computational Science – ICCS 2025 Workshops 25th International Conference, Singapore, Singapore, July 7–9, 2025, Proceedings, Part III. https://doi.org/10.1007/978-3-031-97564-6_11, indywidualny udział poszczególnych autorów jest zgodny z poniższym opisem.

Wkład Aleksandry Konopki:

- konceptualizacja,
- zarządzanie danymi,
- analiza formalna,
- opracowanie metodologii,
- oprogramowanie,
- wizualizacja,
- napisanie wersji oryginalnej manuskryptu, redagowanie tekstu.

Wkład pozostałych autorów:

- Karol Struniawski – recenzowanie i redagowanie tekstu, pomoc w etykietowaniu obrazów,
- Ryszard Kozera – recenzowanie i redagowanie tekstu, nadzór merytoryczny,
- Luciano Ortenzi – konceptualizacja, recenzowanie i redagowanie tekstu,
- Agnieszka Marasek-Ciołakowska – konceptualizacja, pisanie wersji oryginalnej manuskryptu – opis materiału roślinnego, recenzowanie i redagowanie tekstu, zasoby - przygotowanie zbioru obrazów mikroskopowych, nadzór merytoryczny,
- Aleksandra Machlańska – recenzowanie i redagowanie tekstu, zasoby - przygotowanie zbioru obrazów mikroskopowych.

Procentowy udział autorów: Aleksandra Konopka – 60%, Karol Struniawski – 8%, Ryszard Kozera – 8%, Luciano Ortenzi – 8%, Agnieszka Marasek-Ciołakowska – 8% i Aleksandra Machlańska – 8%.

Aleksandra Konopka | Karol Struniawski | Ryszard Kozera

Agnieszka Marasek-Ciołakowska | Aleksandra Machlańska

Warszawa, 24.11.2025

mgr inż. Aleksandra Konopka
aleksandra_konopka@sggw.edu.pl

asystent badawczo-dydaktyczny

Szkoła Główna Gospodarstwa

Wiejskiego w Warszawie

Rada Dyscypliny Informatyka Techniczna i Telekomunikacja

Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie

Oświadczenie o współautorstwie

Niniejszym oświadczamy, że w pracy Konopka A., Kozera R., Marasek-Ciołakowska A., & Machlańska A., Classification of Blackcurrant Genotypes by Ploidy Levels on Stomata Microscopic Images with Deep Learning: Convolutional Neural Networks and Vision Transformers (2025), Applied Sciences-Basel, 2025, t.15. https://doi.org/10.3390/app151910735, indywidualny udział poszczególnych autorów jest zgodny z poniższym opisem.

Wkład Aleksandry Konopki:

- konceptualizacja
- zarządzanie danymi
- analiza formalna
- opracowanie metodologii
- oprogramowanie
- walidacja
- prowadzenie badań
- wizualizacja
- napisanie oryginalnej wersji manuskryptu, redagowanie tekstu
- administracja projektem

Wkład pozostałych autorów:

- Ryszard Kozera - walidacja, recenzowanie i redakcja tekstu, nadzór merytoryczny
- Agnieszka Marasek-Ciołakowska - konceptualizacja, prowadzenie badań, pozyskanie zasoby, recenzowanie i redagowanie tekstu
- Aleksandra Machlańska - zasoby, recenzowanie i redagowanie tekstu

Procentowy wkład autorów: Aleksandra Konopka - 70%, Ryszard Kozera - 10%, Agnieszka Marasek-Ciołakowska - 10%, Aleksandra Machlańska - 10%

Aleksandra Konopka ..... Ryszard Kozera

Agnieszka Marasek-Ciołakowska ..... Aleksandra Machlańska