Int. J. Appl. Math. Comput. Sci., 2020, Vol. 30, No. 3, 463–473
DOI: 10.34768/amcs-2020-0034

RECOGNITION OF SPECIES AND GENERA OF BACTERIA BY MEANS OF THE PRODUCT OF WEIGHTS OF THE CLASSIFIERS

ANNA PLICHTA [a]

[a]Department of Computer Science
Cracow University of Technology
Warszawska 24, 31-155 Cracow, Poland
e-mail: aplichta@pk.edu.pl

In microbiology, computer methods are applied in the analysis and recognition of laboratory-acquired microscopic images concerning, for example, bacterial cells or other microorganisms. Proper recognition of the species and genera of bacteria is a key stage in the microbiological diagnostics process, because it allows a quick start of the appropriate therapy. The original method proposed in the paper concerns the automatic recognition of selected species and genera of bacteria presented in digital images. The classification was made on the basis of the analysis of the physical characteristics of bacterial cells using the product of classifier confidence weights. The end result of the classification process is the classification list, sorted in descending order according to the weights of the classifiers. In addition to the correct classification, a list of other possible results of the analysis is obtained. The method thus allows not only the classification, but also an analysis of the confidence level of the selection made. The proposed method can be used to recognize not only bacterial cells, but also other microorganisms, for example, fungi that exhibit similar morphological characteristics. In addition, the use of the method does not require the application of specialized computer equipment, which widens the scope of applications regardless of the laboratory IT infrastructure, not only in microbiological diagnostics, but also in other diagnostic laboratories.

Keywords: pattern recognition, recognition of bacterial cells, classifiers, product of weights of the classifiers.

1. Introduction

Application of computer science in medical and microbiological sciences covers various research areas such as imaging methods dedicated to specific disorders, systems supporting the processing and analysis of the acquired data, medical software, systems supporting diagnostics, and medical data repositories. In the case of microbiology, computer methods are used in the analysis and recognition of laboratory-acquired microscopic organisms including, for example, bacterial cells or other microorganisms (Tadeusiewicz and Wajs, 1999).

Proper recognition of species and genera of bacteria is often carried out manually using specialized equipment and diagnostic tests. It requires participation of an experienced expert in the field of microbiology and is a time-consuming and costly process. At the same time, the correct and quick diagnosis is considered to be a key stage in the microbiological diagnosis process and undertaking appropriate therapy. Moreover, due to the lack of a digital data repository, the procedure of obtaining samples for analysis and the recognition process itself must be often repeated (Bulanda and Brzychczy-Włoch, 2015; Murray *et al.*, 2015).

The proposed classification method pertains to selected twenty species and genera of bacteria presented in digital images being part of DIBaS DB (Digital Image of Bacterial Species Database) resources. The database is available on the website http://misztal.edu.pl/software/databases/dibas. The classification was made on the basis of the analysis of 7 physical characteristics of bacterial cells by means of the product of weights of classifiers. The end result of the classification is the classification list, sorted in descending order according to the weights of the classifiers. In addition to the correct classification, one is also provided with a list of other possible outcomes of the analyzed images of bacteria. Thus, this method allows not only the classification of samples, but also the analysis of the confidence level of the selection made. The results of the

conducted tests confirm the effectiveness of the proposed classification method.

2. Computer methods for recognition of bacterial cells

Computer methods used to classifying bacteria are often based on artificial intelligence, statistical methods or other solutions aimed at automating the process of analyzing and classifying the obtained data. The most commonly used systems are dedicated to identifying one species and genera of bacteria, for example, bacteria of tuberculosis or a group of microorganisms, including bacteria with similar shapes or other microbiological characteristics (Blackburn *et al*., 1998; Perner, 2001; Trattner *et al*., 2004).

There are also computer methods integrated with a specialized microscope or other research equipment being a component of the entire diagnostic system. This allows the recognition of various microorganisms, but hardware and financial limitations (the type of microscope, the need to use high-quality preparations) may limit wide application of such systems. One way to identify bacteria is to recognize them on the basis of geometric the features, such as the shape or the ratio of the length to the cell width. In addition, because the shape is not a distinguishing feature (due to the same morphology shared by different types and species of bacteria), the color of bacterial cells obtained during their biochemical staining is also taken into account (Hiremath and Bannigidad, 2009). In some other approaches, pre-segmented images obtained from the scanner and various methods of extraction of features (cell size and shape) are used, on the basis of which the classification of bacteria is performed, for example, by means of a decision tree leading to the appropriate morphotype (Liu *et al*., 2001; Bruyne *et al*., 2011). Other methods of bacterial identification are based on the analysis of bacterial colony patterns (cluster of bacterial cells resulting from divisions of individual cells). This type of analysis uses, among others, Fisher's vectors, random forest algorithms, or support vector machines (SVMs) (Cortes and Vapnik, 1995; Holmberg *et al*., 1998; Ates and Gerek, 2009; Sommer and Gerlich, 2013).

A new approach to recognizing bacteria or other medical images (e.g., X-rays, ultrasound pictures) are methods in which a texture model is used to analyze and classify images. The texture represents such image properties as the orientation direction of the pattern or porosity. As a result, it is possible to determine areas in the given image that meet specific conditions, and thus to classify the samples to a given type of texture based on the observation of certain small patterns and their regular deployment. For the mathematical description of the texture, parameters based on the properties of the digital image are calculated, for example, using statistical methods or signal processing techniques. The numeric representation of the texture property is used later for further analysis and classification. The use of this approach to the analysis and classification of selected types and species of bacteria is justified by the fact that different types and species of bacteria reproduce in a specific way. They form clusters of a characteristic shape, which can be treated as a texture. The proposed methods for the analysis and classification of such defined samples make use of, among others, Fisher's vectors, SVMs, and deep neural networks (Krizhevsky *et al*., 2012; Simonyan and Zisserman, 2014; Cimpoi *et al*., 2016; Zieliński *et al*., 2017).

An innovative group of effective methods for identifying bacteria are techniques that use sensors, i.e., devices dedicated to acquiring and processing chemical data describing bacterial cells. New solutions applied in diagnosing bacteria are so-called artificial noses or sensors based on gas identification. The need to prepare an appropriate database for device learning and the ability to detect only ten different chemicals in one sample limits the use of the artificial nose, although these types of commercial devices are used to identify bacteria in diabetic foot infections such as *Escherichia coli, Pseudomonas aeruginosa* and *Staphylococcus aureus*, which makes it possible to quickly take an effective therapy (Hasman *et al*., 2013; Abdullah *et al*., 2014; Arabestani *et al*., 2014; Green *et al*., 2014). As for the other, more frequently used methods, one should mention optical techniques based on the use of light as an information carrier. Fluorescence or spectroscopy techniques, however, are expensive, as they require the preparation of high quality samples and an appropriate database of emission spectra of all identified bacteria and a long analysis time (Alvarez-Ordonez *et al.*, 2011; Suchwałko *et al.*, 2013; 2014; Kusic *et al.*, 2014; Kim *et al.*, 2015).

Most methods are used to recognize a few selected species and genera of bacteria; sometimes it is restricted to just one species, e.g., tuberculosis. Besides, in many cases, the operation of algorithms for bacterial classification is based on morphological characteristics of their cells combined with a certain classification method. This makes them useless in recognizing polymorphic bacterial cells, i.e., those that can exhibit different shapes within the same species, for example, the dominating shape is round, but the bacterium may also have oblong or other shape. Limited possibilities of using computer methods supporting microbiological diagnostics also result from the need to use specialized equipment, both computer and diagnostic.

3. DIBaS DB database

The DIBaS DB database was created for the needs of the conducted research thanks to the cooperation with the Department of Microbiology of the Collegium Medicum of Jagiellonian University in Cracow. The images included in the repository were made by microbiology specialists (employees of the Chair of Microbiology at Jagiellonian University in Cracow) on the basis of properly prepared microbiological preparations for microscopic and biochemical analysis. Thanks to these analyzes, it is possible to classify the examined bacterial cells by an expert in the laboratory. The proposed method requires an input database of samples that are already classified by an expert in order to calculate confidence levels of the classifiers used in the method. The database also includes standard samples obtained from the American collection of the American Type Culture Collection (ATTC) bank reference strains, which are used for comparative analysis facilitating correct classification in laboratory diagnostics. These samples were included in other sample sets pertaining to each species and genera of bacteria. In this research, minimum 20 images for each of the 20 different bacteria species and genera were used. In the training during the classification process half of them were used.

Images of several samples of all species and genera of bacteria taken from the DIBaS DB database and used for the classification are presented in Fig.1.

4. Proposed method for classification

The proposed method of automatic recognition of the selected species and genera of bacteria is based on the classification using the product of weights of classifiers. It uses appropriately implemented classifiers to extract characteristics of bacterial cells, on the basis of which the classification of analyzed samples is performed. In addition, the described method broadens the classification possibilities by analyzing the error made when choosing a decision path and its modification.

The applied classifiers have been developed and implemented specifically for the analysis of samples and are based on physical characteristics of bacterial cells such as the color, the shape, the size of a single cell, the number of clusters formed, the cluster shape, the density and the distribution of cells in the image. These features were selected after consultation with microbiologists.

The method uses 7 classifiers based on physical characteristics of bacterial cells. Within each classifier, the analyzed bacterial cells have been divided into categories defined by the author for the purposes of conducted experiments:

- the colour classifier: purple for Gram-positive (G+) and pink for Gram-negative (G-),
- the classifier of the shape of a single bacterial cell: round, rod-shaped, stick, club, donut and boat,
- the size classifier: large and small,
- the classifier of the number of clusters formed: single cells, diplococci, tetrads, larger,
- the cluster shape classifier: parquet, snake and others,
- the density classifier: rare, dense, very dense,
- the classifier of the distribution of cells in the image: evenly, unevenly, very unevenly.

Classification is based on the product of weights of classifiers and consists in estimating the probability that a given bacterial sample may belong to each of the analyzed species and genera of bacteria based on the analysis of all the above-mentioned features. In the proposed method, classifiers are used in order from the classifier with the highest confidence level to the classifier with the lowest level of confidence.

4.1. Product of weights of classifiers. In the classification of selected species and genera of bacteria according to the product of weights of classifiers, each classifier has a confidence level that can be used as a weight in the assessment of the correctness of classification. The confidence level is a real value in range $[0, 1]$ calculated for each classifier before the analysis.

The algorithm for calculating the product of the weights of classifiers is as follows:

1. Create the list of classifiers $K_1, K_2, K_3, \ldots, K_n$.
2. Create the list of species and genera of bacteria $B_1, B_2, B_3, \ldots, B_m$.
3. For each $i \in [1, n]$:
   (a) Classify the sample with classifier K_i.
   (b) For each $j \in [1, m]$:
      i. If the classification by classifier K_i is the same as the expected classification B_j, assign the weight $w_{i,j} = Cl_{i,j}$, where $Cl_{i,j}$ stands for classifier K_i indicating the confidence level for the specific species and genera B_j.
      ii. Otherwise, assign the weight $w_{i,j} = 1 - Cl_{i,j}$.
4. For each $i \in [1, m]$ calculate the product of weights of classifiers $W_i = \prod_j w_{i,j}$.

In the proposed algorithm we can define:

- $K_i, i \in [1, n]$ as classifiers,

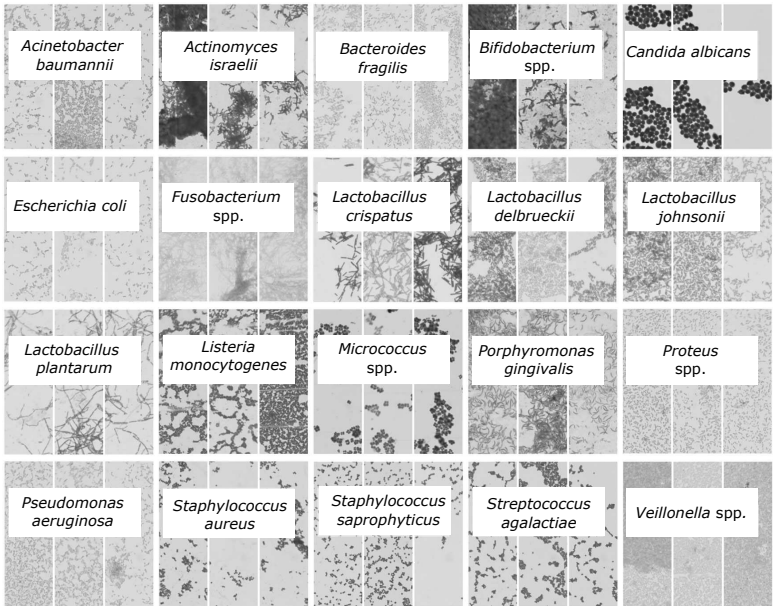

Fig. 1. Sample images of the analyzed genera and bacterial species.

- $B_j, j \in [1, m]$ as species and genera of bacteria,
- X_i as classification of a sample made by classifier K_i.

Note that X_i is independent of species and genera of bacteria of the analyzed sample as classifier K_i does not know the species and genera of bacteria of the analyzed sample.

The confidence level is calculated before the analysis based on the samples for which the correct classification is known. By comparing the correct classification with the actual classification of the classifier, we can assess the correctness of classification. The confidence level is represented as the ratio

$$Cl_{i,j} = \frac{N_{i,j}^{\text{good}}}{N_j}, \tag{1}$$

where $N_{i,j}^{\text{good}}$ is the number of test samples for this species and bacterium correctly classified using this classifier and N_j is the total number of test samples for this species and genus of bacteria.

The weight of a single classifier for a single species and genus of bacteria is calculated as

$$W_{i,j} = \begin{cases} Cl_{i,j}, & X_i = G_{i,j}, \\ 1 - Cl_{i,j}, & X_i \neq G_{i,j}, \end{cases} \tag{2}$$

where $G_{i,j}$ is the correct classification for classifier K_i for a given species and genus of bacteria B_j determined by the expert analyzing the bacteria samples.

Then, the weight determining the classification for a single species and genus of bacteria is

$$W_j = \prod_i W_{i,j}. \tag{3}$$

The weights after sorting from the largest to the smallest are a classification list, starting with the one that gives the best results to the worst performing one. The product of the weights of classifiers calculated in this way promotes correct classifications, without abandoning classifying in the case of the error for classifiers with a lower confidence level. The scheme of the classification by means of the product of weights of classifiers is presented in Fig. 2.

4.2. Classification scheme. The proposed classification method allows one to correctly classify samples even in the case when one of the classifiers makes a mistake, which would not be possible in the case of a decision tree. An additional advantage of the implemented methods above the decision tree is the classification list sorted in the descending order of priority from the highest weight to the lowest. This means that in addition to the best classification it is possible to present alternatives, which is associated with additional metrics that can be used to assess the quality of the classifier. They are as follows:

- The position of the correct classification in the classification list. Assume Pos_j is the index of the position of W_j in the list of weights sorted in descending order. The species and genera of bacteria B_j for which Pos_j is equal 1 is the classification with the highest probability of being correct.

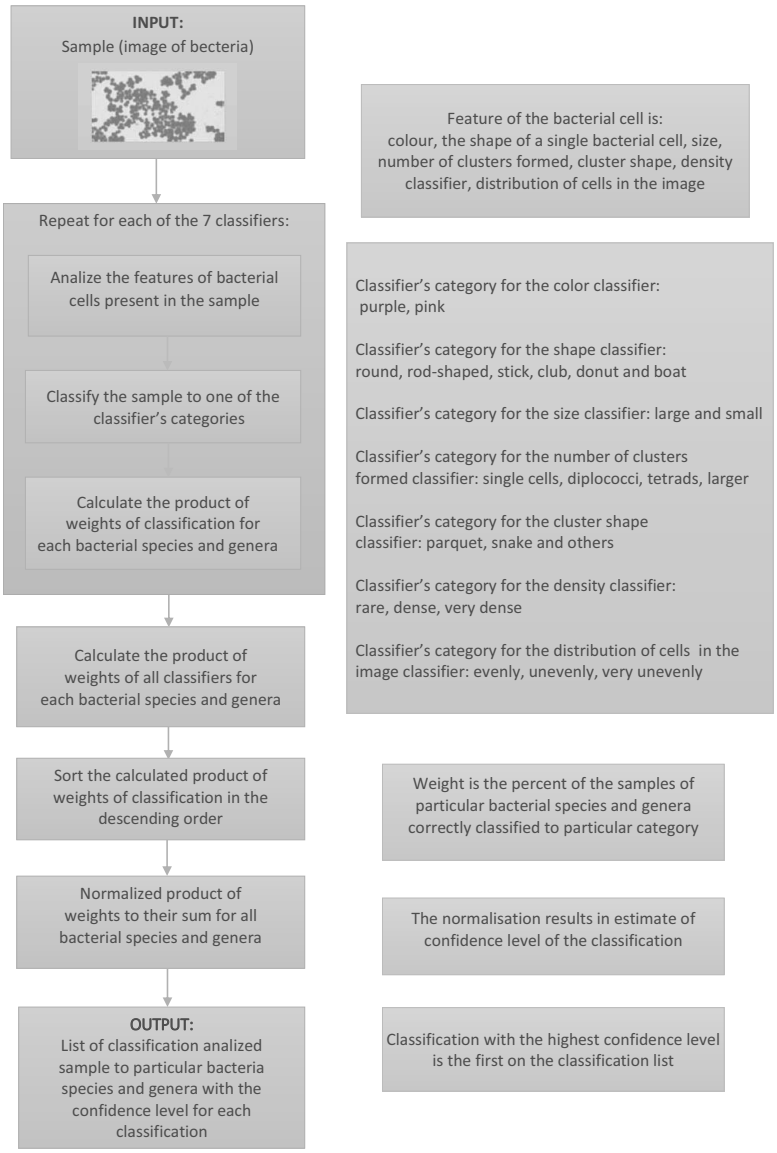

Fig. 2. Algorithm for calculating the product of weights of the classifiers.

In the case of an incorrect classification, the analysis of the B_j for which $\text{Pos}_j = 2$ or $\text{Pos}_j = 3$ may be valuable and may be used to improve the obtained results.

- The confidence of the classification. This value can be calculated in many ways. The most basic one could be

$$C = \frac{W_a}{\sum_j W_j}, \quad \text{Pos}_a = 1 \tag{4}$$

as the ratio of the best classification weight, W_a to the sum of the weights of all classifications, a is used to indicate such weight W_a for which Pos_a is equal 1. This value can be used for quality control – samples classified as those with a low confidence can be declared as unclassified, which shows that the described method does not correctly classify such samples.

The most important element of the method dedicated to the classification of the selected species and genera of bacteria is proposed in this paper. Their correctness and sensitivity has a decisive impact on the final classification of the analyzed samples. In the classification using the product of the weights of classifiers, the basic factor determining the correctness of the classification method is the number of correctly classified species and genera of bacteria. Because in this method a classification list is sorted in descending order according to the product weights of classifiers, the correct classification is the one with the biggest product of the weight of the

classifiers. The confusion matrix for the product of weights of classifiers id presented in Table 1. The correct classification by this method for all analyzed species and genera of bacteria is 90.45%, the sensitivity is 100%. Some of the analyzed samples (*Bifidobacterium* spp., *Escherichia coli*, *Fusobacterium* spp., *Proteus* spp.) were recognized as the appropriate genera and species of bacteria with 100% correctness. In the other samples, due to the same bacterial cell color, a very similar shape, a similar cell size or a similar shape of the cluster of cells, or a similarity of other analyzed features, the classification to the correct species and genera is at a lower level.

The implemented classifiers have 100% sensitivity, with the exception of the classifier of bacterial cell shapes, whose sensitivity is 98.81%. High sensitivity facilitates the classification process, especially if it is conducted by the method described, as this does not lead to a premature rejection of the sample due to the lack of classification. Correct classification by means of other classifiers may allow one to obtain the correct final classification result.

In some cases, the described classifiers do not allow the correct classification of certain species and genera of bacteria. In particular, density classifiers and the uniformity of the distribution of bacterial cells in the image will yield the lowest correctness of classification. The results obtained in this manner are significantly better than the results obtained using decision trees. The biggest advantage of this classification method is the fact that it provides a list of possible alternatives, which allows the assessment of both the confidence level of one's decision and possible other classifications in the case of suspected incorrect classification of the sample for a particular species and genera of bacteria.

4.3. Position of correct classification in the classification list. The metric, which helps in the assessment of the quality of the classification, is to determine the position of a correct classification in the list of classifications with the largest product of the weights of classifiers. In the case of most incorrect results, the correct answer may be in the second or third position of this list. The position of the correct classification result is presented in Table 2.

The results presented in Table 2 show that in 98.56% of cases the correct classification is among the first four items on the classification list, and in 97.37% among the first two. The implemented classifiers are not able to distinguish between two very similar species and genera of bacteria. However, the proposed method allows, with high probability, the rejection of most incorrect classifications, narrowing down the choice to several answers that are likely to contain the correct answer. Analyzing the species and genera of bacteria whose classification was most problematic (the smallest number of samples was correctly classified), it can be seen that the biggest mistakes are made for those bacteria that differ only by one classified feature. In a majority of cases, the classifier that makes a mistake has a low level of confidence.

5. Confidence level: The difference between correct classification and other results in the classification list

The quality of the obtained classification can also be analyzed taking into account the difference between the products of the weights of classifiers for the classification located in the first position and the classifications appearing in the subsequent positions in the classification list. This creates a statistic of the confidence level of the selection made.

Table 3 presents the average confidence level of the classification if the correct classification was at the first position or at one of the following positions in the classification list. As a result, an average confidence level for each species and genera of bacteria is obtained. It was calculated for each position in the classification list based on the number of correctly classified samples of the specific bacterium at the specified position in the classification list. The average confidence level is calculated as follows:

1. Take the sample of the particular species and genus of bacteria.
2. Create the classification list.
3. Check the position on the list including the correct classification of bacteria.
4. Save the correct classification of bacteria in the classification list. Save the position of the correct classification in the classification list (position 1, 2, 3, 4 or the next position in the list) and save the confidence level of that classification.
5. Repeat Steps 2, 3 i 4 for all samples belonging to the particular species and genera of bacteria.

The quality of the obtained classification can also be calculated using following confidence value:

$$C_x = W_a - W_b, \quad \text{Pos}_a = X \wedge \text{Pos}_b = X - 1, \quad (5)$$

where C_1 would be the confidence of the classification with the highest probability of being correct. X stands for the order of the metric C_x. The average confidence of the classification of all samples for given species and genera of bacteria can be used to determine the quality of the classification of this species and genera of bacteria. Table 3 presents values C_1, C_2, C_3, C_4 and the average of values $C_5, \ldots, C_{m-1}$ for individual species and genera of bacteria.

Table 1. Confusion matrix for the product of weights of the classifiers.

No.	Species and genera of bacteria	*Acinetobacter baumannii*	*Actinomyces israelii*	*Bacteroides fragilis*	*Bifidobacterium* spp.	*Candida albicans*	*Escherichia coli*	*Fusobacterium* spp.	*Lactobacillus crispatus*	*Lactobacillus delbrueckii*	*Lactobacillus johnsonii*	*Lactobacillus plantarum*	*Listeria monocytogenes*	*Micrococcus* spp.	*Porphyromonas gingivalis*	*Proteus* spp.	*Pseudomonas aeruginosa*	*Staphylococcus aureus*	*Staphylococcus saprophyticus*	*Streptococcus agalactiae*	*Veillonella* spp.	no classification
1	*Acinetobacter baumannii*	85%		5%	4%								5%								10%	
2	*Actinomyces israelii*		96%																			
3	*Bacteroides fragilis*			96%																	4%	
4	*Bifidobacterium* spp.				100%																	
5	*Candida albicans*		5%			90%							5%									
6	*Escherichia coli*						100%															
7	*Fusobacterium* spp.							100%														
8	*Lactobacillus crispatus*				5%				80%				10%									
9	*Lactobacillus delbrueckii*									75%	10%	15%	5%									
10	*Lactobacillus johnsonii*				5%					25%	70%											
11	*Lactobacillus plantarum*									10%		90%										
12	*Listeria monocytogenes*				5%								95%									
13	*Micrococcus* spp.													95%				5%				
14	*Porphyromonas gingivalis*								4%						96%							
15	*Proteus* spp.															100%						
16	*Pseudomonas aeruginosa*															5%	95%					
17	*Staphylococcus aureus*								5%									85%				
18	*Staphylococcus saprophyticus*									5%			5%					20%	75%			
19	*Streptococcus agalactiae*			5%															10%	5%	90%		
20	*Veillonella* spp.	5%																				90%

Table 2. Percentage of correct classification results in the first and the next position in the classification list.

No.	Species and genera of bacteria	Position on the classification list				
		1	2	3	4	higher
1	*Acinetobacter baumannii*	85.00%		5.00%	5.00%	5.00%
2	*Actinomyces israelii*	95.65%				4.35%
3	*Bacteroides fragilis*	95.65%	4.35%			
4	*Bifidobacterium* spp.	100.00%				
5	*Candida albicans*	90.00%	10.00%			
6	*Escherichia coli*	100.00%				
7	*Fusobacterium* spp.	100.00%				
8	*Lactobacillus crispatus*	80.00%	20.00%			
9	*Lactobacillus delbrueckii*	75.00%	25.00%			
10	*Lactobacillus johnsonii*	70.00%	25.00%			5.00%
11	*Lactobacillus plantarum*	90.00%	5.00%	5.00%		
12	*Listeria monocytogenes*	95.45%				4.55%
13	*Micrococcus* spp.	95.00%				5.00%
14	*Porphyromonas gingivalis*	95.65%		4.35%		
15	*Proteus* spp.	100.00%				
16	*Pseudomonas aeruginosa*	95.00%	5.00%			
17	*Staphylococcus aureus*	85.00%	15.00%			
18	*Staphylococcus saprophyticus*	75.00%	20.00%			5.00%
19	*Streptococcus agalactiae*	90.00%	5.00%	5.00%		
20	*Veillonella* spp.	90.91%	9.09%			
Summarised results		90.45%	6.92%	0.95%	0.24%	1.44%

The results in Table 3 show that the average confidence level of classification for all analyzed data samples is the highest if the correct classification is in the first position on the list (the last column contains individual samples and, therefore, it can be omitted). Based on these results, one can introduce a mechanism to determine in which cases this method is not able to correctly classify bacterial samples. One can do this by marking results having too low confidence levels as unclassified. As a result, at the expense of the reduction in the sensitivity of the method, its higher correctness may be achieved.

Figure 3 presents this relationship in the form of a graph of correctness and sensitivity in relation to the level of confidence (it was calculated on the basis of the ratio of the best classification to all classifications), at which the sample data are labeled as unclassified. The Y-axis represents correctness and sensitivity, whereas the X-axis shows the percentage of the confidence level of the sample being defined as unclassified.

6. Comparison of classification by the product of weights of classifiers and by decision trees

To verify the obtained results, the proposed method was compared with a decision tree. Both the decision tree and the method described in this study are based on the recognition of physical features of bacterial cells visible in the images. In both methods, the same images and the same set of seven implemented classifiers were used to classify the samples and recognize the analyzed species and genera of bacteria; all the images were taken from DIBaS DB database resources (Plichta, 2019).

The proposed decision tree tended to use highly correct classifiers, such as the bacterial cell color, which has the highest correctness of all. This decision tree can be further optimized to provide the best possible results using the boosted decision trees method. The accuracy (correctness) of this method amounted to 83.77%, and although changes in the decision tree remained at 95.94%, it had no impact on its sensitivity. Classification by means of the method based on the product of the weights of the classifiers brought better results. For all the analyzed

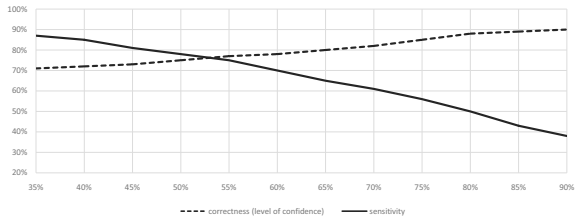

Fig. 3. Chart of correctness and sensitivity with regard to the confidence level.

Table 3. Average confidence level for all classified species and genera of bacteria when the correct classification is in the first or next positions in the classification list.

No.	Species ad genera of bacteria	Position on the classification list Average confidence level				
		1	2	3	4	next position
1	*Acinetobacter baumannii*	70.99%		38.95%	42.88%	98.89%
2	*Actinomyces israelii*	85.98%				87.86%
3	*Bacteroides fragilis*	53.95%	75.07%			
4	*Bifidobacterium* spp.	97.12%				
5	*Candida albicans*	82.45%	55.24%			
6	*Escherichia coli*	97.88%				
7	*Fusobacterium* spp.	99.57%				
8	*Lactobacillus crispatus*	84.52%	67.39%			
9	*Lactobacillus delbrueckii*	71.03%	78.35%			
10	*Lactobacillus johnsonii*	75.08%	66.43%			74.96%
11	*Lactobacillus plantarum*	76.81%	69.15%	68.85%		
12	*Listeria monocytogenes*	98.88%				58.34%
13	*Micrococcus* spp.	93.97%				89.11%
14	*Porphyromonas gingivalis*	95.90%		63.75%		
15	*Proteus* spp.	91.81%				
16	*Pseudomonas aeruginosa*	96.99%	92.41%			
17	*Staphylococcus aureus*	87.16%	77.08%			
18	*Staphylococcus saprophyticus*	94.34%	80.82%			91.76%
19	*Streptococcus agalactiae*	91.76%	66.95%	47.52%		
20	*Veillonella* spp.	94.26%	54.97%			
Summarised result		87.56%	71.45%	54.77%	42.88%	83.49%

species and genera of bacteria, correct classification by this method amounted to 90.45% and its sensitivity to 100%. Moreover, the classification list shows that in 98.56% of cases the correct classification is among the first four items on the classification list whereas in 97.37% it is among the first two items. The comparison of these two methods with the use of the same classifiers and the same test data confirms the correctness of the method proposed in the paper.

7. Summary

The proposed algorithms dedicated to the extraction of physical features of cells of different species and genera of bacteria enabled their proper recognition and implementation of classifiers based on such extracted features. The correct classification by this method for all analyzed species and genera of bacteria is 90.45%. The results of the experiment presented in the paper show that in 98.56% of cases the correct classification is among the first four items on the classification list, and in 97.37% among the first two.

An innovative element in the proposed method was the use of classifiers that simultaneously analyze the following seven physical characteristics of bacterial cells: color shape, size of a single cell, number of clusters formed, cluster shape, density and cell distribution in the image. Until now, the color and the shape or the size of bacterial cells have been taken into account in the classification. The shape of the cluster, the density or the distribution of cells in the picture were not taken into account as the features affecting the correct identification of the analyzed samples.

It is also worth noticing that the previous classification methods concerned one, several or a dozen species and genera of bacteria, most often of the same type or showing similar morphological features. The number of twenty different species and genera of bacteria analyzed in the conducted tests is therefore an innovative element.

The implemented classifiers are assigned a confidence level of classification for each of the classified species and genera of bacteria, which is later used, among others, to assess the quality of the classification. In addition, the analysis of the classification confidence level allows us not only to indicate the most probable classification of the tested sample to the genera and species, but also the prospective possible responses, which is very important in microbiological diagnostics, especially in ambiguous cases, when the correct classification is difficult, for example, due to the poor quality of the image taken. The original approach is also the classification of samples using the product of the weights of the above-mentioned classifiers. Thanks to the automation of the process of proper recognition, the proposed method shortens the time necessary for

identification and classification and hence for the correct recognition of the species and genera of bacteria in the image, which supports and significantly improves the microbiological diagnosis process. In addition, the participation of a specialist in the diagnosis of bacteria was limited to the proper preparation of bacterial cell cultures and taking an image of the sample visible under the microscope — the identification and classification of the bacterial cells under study for individual species and genera will no longer be his or her task.

The use of the method does not require the use of specialized computer equipment, which widens the scope of applications regardless of the laboratory IT infrastructure. It can be applied not only in microbiological diagnostics, but also in other laboratories, for example, veterinary or epidemiological ones, where species and genera of bacteria or other microorganisms are analyzed and classified.

The described method of identifying selected species and genera of bacteria can be used separately, but in the case of more complex images containing many different cells of bacteria or other microorganisms mixed together in one image, it can be only the first stage of classification. In this approach, it is possible to reject or narrow down the possibility of classifying the different types of cells visible in the image so that in the next step, other known methods, for example, neural networks, can be used. This method can also be used to verify the correctness of the classification obtained by using other dedicated tools, for example machine learning or neural networks. Classifiers directly refer to the physical characteristics of bacterial cells. Therefore, they can be used to extract only the characteristics of these samples that have not yet been classified using the proposed method. Thanks to this, it is possible to use the method to recognize new species and genera of bacteria added to the DIBaS DB database and analysis of images of many other microorganisms that have characteristics similar to those of bacterial cells.

References

Abdullah, A., Jing, T., Sie, C., Yusuf, N., Zakaria, A., Omar, M., Shakaff, A.M., Adom, A.H., Kamarudin, L., Juan, Y. and et al. (2014). Rapid identification method of aerobic bacteria in diabetic foot ulcers using electronic nose, *Advanced Science Letters* **20**(1): 37–41.

Alvarez-Ordonez, A., Mouwen, D., Lopez, M. and Prieto, M. (2011). Fourier transform infrared spectroscopy as a tool to characterize molecular composition and stress response in foodborne pathogenic bacteria, *Journal of Microbiological Methods* **84**(3): 369–378.

Arabestani, M.R., Fazzeli, H. and Esfahani, B.N. (2014). Identification of the most common pathogenic bacteria in patients with suspected sepsis by multiplex PCR, *Journal of Infection in Developing Countries* **8**(4): 461–468.

Ates, H. and Gerek, O.N. (2009). An image-processing based automated bacteria colony counter, *Proceedings: International Symposium on Computer and Information Sciences ISCIS, Guzelyurt, Cyprus*, pp. 18–23.

Blackburn, N., Hagström, Å., Wikner, J., Cuadros-Hansson, R. and Bjórnsen, P.K. (1998). Rapid determination of bacterial abundance, biovolume, morphology, and growth by neural network-based image analysis, *Applied and Environmental Microbiology* **64**: 3246–3255.

Bruyne, D.K., Slabbinck, B., Waegeman, W., Vauterin, P., De Baets, B. and Vandamme, P. (2011). Bacterial species identification from MALDI-TOF mass spectra through data analysis and machine learning, *Systematic and Applied Microbiology* **34**(1): 20–29.

Bulanda, M. and Brzychczy-Włoch, M. (Eds) (2015). *Microbiology and Parasitology: Lecture Notes for 2nd Year Students of the Faculty of Medicine of Jagiellonian University Collegium Medicum*, Cracow Scientific Publishers Tekst, (in Polish).

Cimpoi, M., Maji, S., Kokkinos, I. and Vedaldi, A. (2016). Deep filter banks for texture recognition, description, and segmentation, *International Journal of Computer Vision* **118**: 65–94.

Cortes, C. and Vapnik, V. (1995). Support-vector networks, *Machine Learning* **20**: 273–297.

Green, G., Chan, A. and Lin, M. (2014). Robust identification of bacteria based on repeated odor measurements from individual bacteria colonies, *Sensors and Actuators B: Chemical* **190**: 16–24.

Hasman, H., Saputra, D., Sicheritz-Ponten, T., Lund, O., Svendsen, C., Frimodt-Móller, N. and Aarestrup, F. (2013). Rapid whole genome sequencing for the detection and characterization of microorganisms directly from clinical samples, *Journal of Clinical Microbiology* **52**(1): 139–146.

Hiremath, P. and Bannigidad, P. (2009). Automated gram-staining characterization of digital bacterial cell images, *Procceedings: International Conference on Signal and Image Processing ICSIP, Amsterdam, The Netherlands*, pp. 209–211.

Holmberg, M., Gustafsson, F., Hörnsten, G.E., Winquist, F., Nilsson, L.E., Ljung, L. and Lundström, I. (1998). Bacteria classification based on feature extraction from sensor data, *Biotechnology Techniques* **12**(4): 319–324.

Kim, H., Doh, I.-J., Bhunia, A., King, G. and Bae, E. (2015). Scalar diffraction modeling of multispectral forward scatter patterns from bacterial colonies, *Optics Express* **23**(7): 8545–8554.

Krizhevsky, A., Sutskever, I. and Hinton, G. (2012). Imagenet classification with deep convolutional neural networks, *in* P. Bartlett (Ed.), *Advances in Neural Information Processing Systems,* NIPS, San Diego, CA, pp. 1097–1105.

Kusic, D., Kampe, B., Rösch, P. and Popp, J. (2014). Identification of water pathogens by Raman microspectroscopy, *Water Research* **48**: 179–189.

Liu, J., Dazzo, F., Glagoleva, O., Yu, B. and Jain, A. (2001). CMEIAS: A computer-aided system for the image analysis of bacterial morphotypes in microbial communities, *Microbial Ecology* **41**: 173–194.

Murray, P., Rosenthal, K. and Pfaller, M. (2015). *Medical Microbiology*, Elsevier, Amsterdam.

Perner, P. (2001). Classification of hep-2 cells using fluorescent image analysis and data mining, *International Symposium on Medical Data Analysis: Medical Data Analysis, Madrid, Spain*, pp. 219–224.

Plichta, A. (2019). Methods of classification of the genera and species of bacteria using decision tree, *Journal of Telecommunications & Information Technology* **4**: 74–82.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition, `https://arxiv.org/abs/1409.1556`.

Sommer, C. and Gerlich, D. (2013). Machine learning in cell biology—Teaching computers to recognize phenotypes, *Journal of Cell Science* **126**: 18–23.

Suchwałko, A., Buzalewicz, I. and Podbielska, H. (2014). Bacteria identification in an optical system with optimized diffraction pattern registration condition supported by enhanced statistical analysis, *Optics Express* **22**(21): 26312–26327.

Suchwałko, A., Buzalewicz, I., Wieliczko, A. and Podbielska, H. (2013). Bacteria species identification by the statistical analysis of bacterial colonies fresnel patterns, *Optics Express* **21**(9): 11322–11337.

Tadeusiewicz, R. and Wajs, W. (1999). *Health Informatics*, AGH University of Science and Technology Press, Cracow, (in Polish).

Trattner, S., Greenspan, H., Tepper, G. and Abboud, S. (2004). Automatic identification of bacterial types using statistical imaging methods, *IEEE Transactions on Medical Imaging* **23**: 807–820.

Zieliński, B., Plichta, A., Misztal, K., Spurek, P., Brzychczy-Włoch, M. and Ochońska, D. (2017). Deep learning approach to bacterial colony classification, *PloS One* **12**(9): e0184554.

Anna Plichta graduated in computer science from the Cracow University of Technology in 2010. In 2019 she obtained her PhD in computer science at the Wroclaw University of Science and Technology. Currently, she is an assistant professor at the Cracow University of Technology. The main topics of her research are pattern recognition, databases, artificial intelligent systems and e-learning technologies.

Received: 5 December 2019
Revised: 3 March 2020
Re-revised: 15 May 2020
Accepted: 29 May 2020