SUMMARY OF THE DOCTORAL DISSERTATION

MSc. Agnieszka Suchwałko

Application of statistical analysis for the bacteria identification on the basis of bacterial colonies diffraction patterns

Supervisor: Professor Halina Podbielska
Secondary supervisor: dr. Igor Buzalewicz

Purpose of the dissertation:

Development of rapid and accurate method for bacteria identification based on statistical analysis of numerical features extracted from the Fresnel diffraction patterns of the bacterial colonies grown on solid media and diffraction patterns of bacterial colonies database, suitable for implementation in microbiological practice.

Theses of the dissertation:

1. Taking into account the natural distribution of the patterns into rings and the diameter of diffraction patterns has a significant impact on the possibility of bacteria identification.
2. Calculation of the relevant numerical features of the bacterial diffraction patterns based on their pixel values that takes into account the distinctive texture and morphology of the registered Fresnel diffraction patterns of bacterial colonies ensures optimal use of the information carried by the patterns. Optimal use of the available information allows for the construction of the best fitting predictive models thanks to the use of only those numerical features that best describe various bacterial groups.
3. Selection of appropriate numerical features of those extracted from the Fresnel diffraction patterns of bacterial colonies allows building of optimal classification models.
4. Application of the pixel values normalization on registered Fresnel diffraction patterns of bacterial colonies ensures their accurate comparison between each other, which leads to reduction of the bacteria identification error.
5. The method has a high accuracy, that allow the identification at the level of bacterial serovars.
6. Introduction of minor modifications to the optical system enables provision of the repeatability of measurements, regardless of the person performing the registration with just a slight deterioration of the results obtained.
7. Application of appropriate algorithms for classification, identification and verification of the results allows obtaining bacterial identification results with the least error.

Conducted research and the obtained results:

Results of performed research work were presented in 6 publications (chapter in the English book, two works from Philadelphia List, 3 proceedings publications), patent application.

First of the research, the results of which were published in [1] were intended to show that the data extracted from the Fresnel diffraction spectra of bacterial colonies recorded using a dedicated optical system enables the identification of bacteria with high accuracy. Division of the patterns into cylindrical areas from which numerical features were extracted was proposed. The

partitioning is analogous to the natural distribution of the patterns into the annular areas that differ between the groups of bacteria. Mean and standard deviation of pixel values were used as a measure of the brightness and roughness in all areas of diffraction patterns as quantitative features describing the patterns. Exploratory data analysis has been carried out in order to illustrate the potential of data on the differentiation of bacterial groups. Then the best numerical features were chosen with use of ANOVA and the classification model LDA (Linear Discriminant Analysis) was build. **Verification of the results obtained using cross-validation was the final task and showed identification error at the level of 5.97%.**

Another of conducted works has been started on extending the database of bacterial colonies diffraction patterns [2]. Analysis of the new, much larger set of data was necessary to confirm the possibility of bacterial identification based on only a simple numerical features extracted from the bacterial colonies diffraction patterns. Achieving a 8.11% identification error for the new, approximately twice as big database allowed to assume that the initial data preparation combined with application of a more advanced statistical data analysis apparatus and extraction combined with selection of numerical features better differentiating bacterial group may significantly improve the results.

Next experiments described in the paper [3], in addition to the previously used techniques of processing and analysis of the data were supplemented with normalization of the patterns, which aimed to harmonize the gray-scale digital images of bacterial colonies diffraction patterns, test the effect of classifiers: QDA (Quadratic Discriminant Analysis) and SVM (Support Vector Machine) and the calculation of sensitivity and specificity for multiple classes. Application of normalization and QDA classifier with the partitioning of the diffraction patterns into 10 cylindrical areas has **reduced the identification error to a value of 3.14% with a sensitivity of 100% and a specificity of 96.35%**.

The next step of improving the bacteria identification method, presented at the conference SPIE BIOS Munich [4] was to test morphological and textural features of the diffraction patterns, based on statistical moments and compare the impact of the results of ANOVA with Fisher divergence (SNR–Signal to Noise Ratio). In addition, it was necessary to introduce a two-step selection of the best separating bacteria features: first, best separating bacterial groups sets of features without division into regions (e.g. mean, standard deviation), then best separating bacterial groups features from the previously selected sets (e.g. mean in the areas 1 and 3 starting from center of the pattern). **The most effective was a QDA classifier, giving an error of 0.857% in combination with the ranking features selected by SNR for the 18 best discriminatory features. Obtained for the best result sensitivity and specificity of multiclass are close to 100%.**

A summary of results obtained so far and a comparison with competitive method for identifying bacteria proposed by the American group of researchers (also using an optical system and data analysis) is contained in the chapter of the book [5]. Research conducted by the US group differ from ours not only in means of collecting and analyzing the data, but also in a way of presenting the results and the size of data sets analyzed. Our data are larger and more diverse. Conducted in both groups study also assume different conditions of cultivation, so unambiguous comparison was not possible. However, an attempt to statement of the results published by the two groups for the largest data collections, which for Americans means a set of 4 strains, and for us a set of 7 groups of bacteria, **higher sensitivity and more than 1% lower value of identification error speak in favor of our method.**

The last stage of the research undertaken in the doctorate is summarized in paper from the Philadelphia List [6]. The paper concerns optimization of the entire method: the way of sample preparation, the optical system and data analysis. Method required clarification of procedures due to an absolute requirement of ensuring repeatability of experiments, regardless of conditions, including people carrying them. Further modifications of the method consisted of testing the ability to identify

bacteria that were recorded for 7 different sample distance (Petri dish with grown bacterial colonies) from the camera lens and the choice of optimal distance. Moreover, various incubation times were tested required to grow bacterial colonies with the most optimal diameter for further measurements. New numerical feature (that differentiates well bacteria groups) describing the radius of the diffraction pattern of a bacterial colony has also been added. Stratified sampling method was used to increase the representativeness of the samples for cross-validation. Two experimental tracks were also compared: the original and optimized. **The modified method was more effective than its previous version reaching the error of identification at 1.34%. Multiclass sensitivity and specificity obtained respectively 97.59% and 99.03%.** SVM classifier turned out to be the optimal one.

Conclusions:

In the course of the research method for bacteria identification based on analysis of bacterial colonies Fresnel diffraction patterns was subjected to numerous modifications that have significantly influenced both its accuracy and suitability for use in practice. From a purely research method it was transformed into a technology that meets the requirements of technique used commercially. Naturally spectrum of possible modifications and improvements has not been fully exhausted, but achieved by applied changes results are entirely satisfactory.

The method uses the unique properties of scattered light on bacterial colony in the form of Fresnel diffraction pattern. The use of properly selected statistical data analysis apparatus allowed the optimal usage of the information recorded in the form of bacterial colonies diffraction patterns. The method does not require any dedicated reagents, only the standard materials present in every microbiological laboratory, for which there is only the a requirement of good transparency. These requirements ensure low operating costs and the lack of expensive staff training. The operating time is limited to the time necessary for bacterial colony cultivation in diameter allowing the registration of its diffraction pattern. Performing the analysis takes just a few minutes, which is competitive with most methods with the exception of tests (e.g. immunoassay). By means of the proposed method bacteria whose diffraction patterns had previously been stored in the database can be identified. New (ie. mutated) species or strains of bacteria can be added to the database and immediately be identified without interfering with the method itself. The method is able to identify also a mixture of bacteria, since with appropriate dilution, each colony arises from a single bacterial cell. If the colony is not in contact with other colonies on a plate, it is a subject to the identification independently of the other colonies present in the sample. The effectiveness of the method for identification of bacterial serovars has also been demonstrated.

Great application potential of the method has been recognized by the private sector. R&D work of the method were conducted at Wroclaw University of Technology in cooperation with a Bioavlee company. Currently the company is engaged in implementation of a commercial device based on the concepts developed by the Bio-Optics Group from the Department of Biomedical Engineering.

The works that form the basis of applying for doctoral degree:

1. **Suchwalko A**, Buzalewicz I, Podbielska H (2012) *Computer-based classification of bacteria species by analysis of their colonies Fresnel diffraction patterns.* In: Miller BL, Fauchet PM, editors. Proceedings of SPIE. Vol. 11. p. 82120R – 82120R – 13. http://www.opticsinfobase.org/abstract.cfm?URI=BIOMED-2012-BSu5A.5

2. Podbielska H, Buzalewicz I, **Suchwalko A**, Wieliczko A (2012) *Bacteria Classification by Means of the Statistical Analysis of Fresnel Diffraction Patterns of Bacteria Colonies.* Biomedical Optics and 3-D

Imaging. Washington, D.C.: OSA. p. BSu5A.5. http://www.opticsinfobase.org/abstract.cfm?URI=BIOMED-2012-BSu5A.5

3. **Suchwalko A**, Buzalewicz I, Wieliczko A, Podbielska H (2013) *Bacteria species identification by the statistical analysis of bacterial colonies Fresnel patterns.* Opt Express 21: 11322–11337, **IF: 3.525, Punkt MNiSW (2013): 45** http://www.opticsinfobase.org/abstract.cfm?URI=oe-21-9-11322

4. **Suchwalko A**, Buzalewicz I, Podbielska H (2013) *Identification of bacteria species by using morphological and textural properties of bacterial colonies diffraction patterns.* In: Remondino F, Shortis MR, Beyerer J, Puente León F, editors. Proceedings of SPIE. pp. 87911M – 1–87911M – 7. http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1691637

5. **Suchwalko A**, Buzalewicz I, Podbielska H (2013) *Statistical identification of bacteria species*. In: Méndez-Vilas A, editor. Microbial pathogens and strategies for combating them: science, technology and education. Badajoz, Spain: Formatex Research Center. pp. 711–721. http://www.formatex.info/microbiology4/vol1/711-721.pdf

6. **Suchwałko A**, Buzalewicz I, Podbielska H (2014) *Bacteria identification in an optical system with optimized diffraction pattern registration condition supported by enhanced statistical analysis*. Opt Express 22: 26312–26327. **IF: 3.488, Punkt MNiSW (2014): 45** http://www.ncbi.nlm.nih.gov/pubmed/25401664.

7. **A. Suchwałko**, H. Podbielska, I. Buzalewicz, „Sposób identyfikacji gatunku bakterii" , nr zgłoszenia P 400116 z dn. 24.07.2012