



The Microorganism Detection System (SDM) for microbiological control of cosmetic products

Paweł Zawadzki^{1,2,A-E}, Piotr Adamczuk^{3,A-C,F}, Konrad Jamka^{3,A-C,F},
Paula Wróblewska-Łuczka^{4,2,A-D,F}, Hubert Bojar^{3,A-C,F}, Grzegorz Raszewski^{3,A,E-F}

¹ Institute of Theory of Electrical Engineering, Measurement and Information Systems, University of Technology, Warsaw, Poland

² NVT Limited Liability Company, Warsaw, Poland

³ Institute of Rural Health, Lublin, Poland

⁴ Department of Occupational Medicine, Medical University, Lublin, Poland

A – Research concept and design, B – Collection and/or assembly of data, C – Data analysis and interpretation, D – Writing the article, E – Critical revision of the article, F – Final approval of the article

Zawadzki P, Adamczuk P, Jamka K, Wróblewska-Łuczka P, Bojar H, Raszewski G. The Microorganism Detection System (SDM) for microbiological control of cosmetic products. *Ann Agric Environ Med.* 2021; 28(4): 705–708. doi: 10.26444/aaem/144668

Abstract

The Microorganism Detection System (SDM) is a new solution using artificial intelligence, unique on the international scale, to correctly identify and count microorganisms, with particular emphasis on specific listed microorganisms (Document of Standard PN-EN ISO 17516–2014:11) – *Candida albicans*, *Escherichia coli*, *Pseudomonas aeruginosa*, *Staphylococcus aureus*. SDM enables the use of algorithms for microscopic image interpretation in the microbiological assessment of the cosmetics in accordance with the standard, providing an answer to whether the tested product complies with the standard. Apart from the software part of SDM, an integral part of the system is an innovative methodology for preparing a cosmetic sample for testing. The experiments confirm the high sensitivity and specificity of the SDM method, its repeatability and, above all, the comparability of the results with the methods of European standards.

Key words

artificial intelligence, detection of microorganisms, image interpretation algorithm, microbiological safety

INTRODUCTION

General purpose of the SDM. The presented Microorganism Detection System (SDM) is an internationally unique use of the artificial intelligence (AI) to optimize the microbiological control process of cosmetics with the use of algorithms for microscopic image interpretation.

The general idea of SDM is to make correct identification and counting of microorganisms, with particular emphasis on specific listed microorganisms (Document of Standards PN-EN ISO 17516-2014:11) – *Candida albicans*, *Escherichia coli*, *Pseudomonas aeruginosa*, and *Staphylococcus aureus* [1]. This standard defines the acceptable levels of microbiological contamination in accordance with the category of the cosmetic. Since July 2013, Regulation (EC) No. 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products [2], has been in force, which imposes an obligation to test the microbiological purity of cosmetic produce to ensure the safety of consumers.

METHOD AND RESULTS

Methodology of preparing a cosmetic sample. An integral part of the presented SDM is the methodology of preparing a cosmetic sample for testing. The methodology covers the steps from taking a cosmetic sample to obtaining separated live microorganisms through staining and taking pictures, which

in the final stage are analyzed by the artificial intelligence of the SDM.

The research team of the Institute of Rural Health tested numerous variants of the methods [3–5] based mainly on the concentration methods: flotation, sedimentation, concentration gradient centrifugation and filtration (on filters with a pore size of 0.2 μm) for the separation of living microorganisms. The final, validated methodology for the preparation of cosmetic samples is a modification of the different above-mentioned methods [3, 4]. Live microorganism cells obtained in the previous stage were filtered and then stained in the next stages.

The researchers then looked for a method of comprehensive staining, so that bacteria, yeasts and moulds could be assessed simultaneously in one microscopic preparation. After carrying out numerous tests, the focus was on Gram staining and modifications [6–8]. Next, numerous photos of the microscope preparations stained with Gram staining modifications were sent to a Consortium member – NVT for the development of SDM software and artificial intelligence learning to recognize specific microbial species.

Software part of SDM. This consists of 3 stages: detection of individual microorganisms (cells), classification of the detected microorganisms, and processing of the results into the final form in terms of compliance with the document of the standard ISO. SDM architecture schema is showed in Fig. 1.

Data description. The study used data consisting of 5 sets of images. Each set contains 5 images taken of 19 microorganisms (within 4 specific microorganisms). In other

Address for correspondence: Grzegorz Raszewski, Institute of Rural Health, Lublin, Poland
E-mail: raszewski.g@imw.lublin.pl

Received: 16.11.2021; accepted: 07.12.2021; first published: 16.12.2021

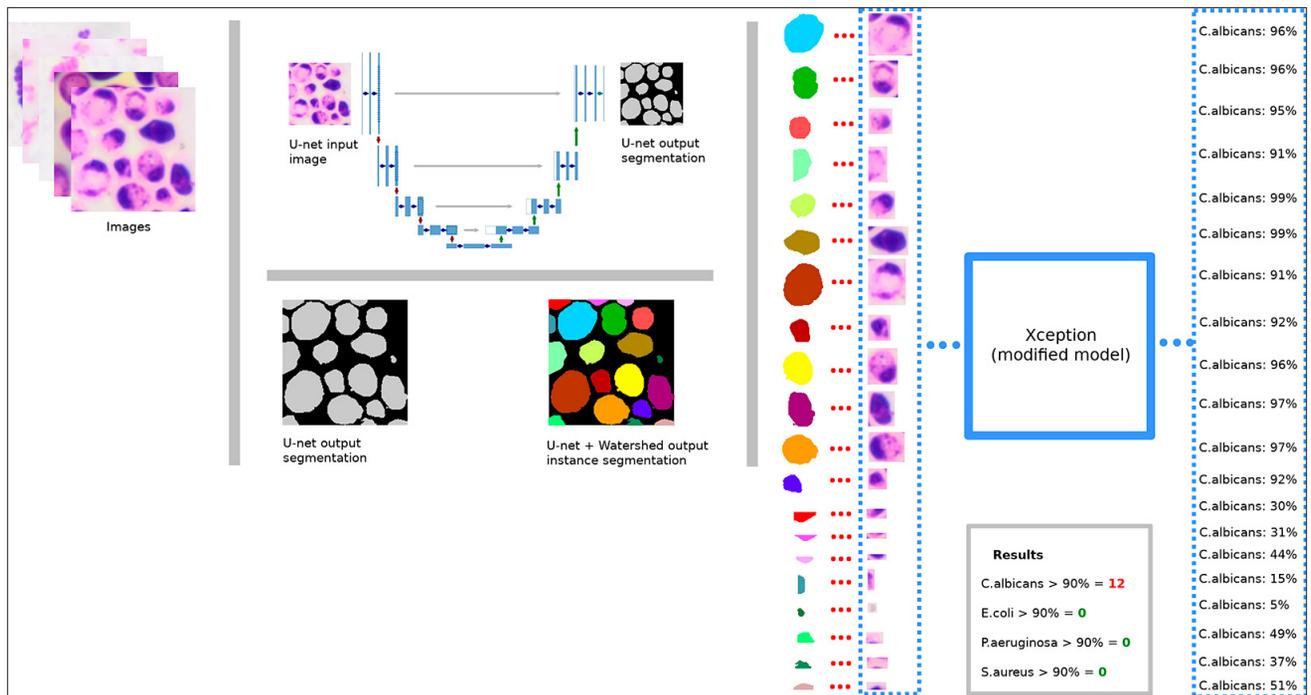


Figure 1. SDM architecture schema.

Left part: a set of selected input data (images); centre upper part: semantic segmentation (cells and background); centre bottom part: instance segmentation (each cell has its own color). Right part: classification of segmented cells (bounding boxes) and presentation of the final results (in bottom frame)

words, 25 images were made of each of the microorganisms; the sum of all images was $25 \times 19 = 475$. Each image was taken at the resolution: 3088×2076 pixels.

Cell detection. For the cell detection step (the first step of the described process), an algorithm based on the U-net deep neural network [9] and the Watershed method [10] was developed. As a result of this algorithm, the contours of individual cells were obtained, which were then transformed into bounding boxes. In fact, cells marked in this way (by bounding boxes) are cut out of the image and saved as an independent image (detected cells of microorganisms). However, in order for the whole process to work in its final form, it must first be trained.

Data for training cell detection method. On N random-selected images (from 2–10 depending on the number of visible cells) of each microorganism, manual cell segmentation was performed with QuPath [11]. The results of manual segmentation were then saved in the resulting form (JSON), and the original images cut into small parts, each with a size of 304×304 pixels; the size inspired by YOLO v3 [12] where the images are 4 times bigger: 608×608 pixels. For each cropped image, a mask file was created with the selection of individual cells (the area of a single cell consists of pixels with the same value).

Training and testing. The data prepared in the manner described above was split into independent training and testing sets with 1,816 and 455 pairs of images (photo + mask), respectively.

The cell detection algorithm was trained for 200 epochs and the input data pre-normalized. The deep neural network (DNN) during training was assessed with different metrics, but due to the specificity of the task (segmentation) and

the importance of the assessment, it was decided to use the Intersection over Union (IOU) metric, which takes values in the range $<0, 1>$ (the closer to 1 the better).

The entire training was repeated 10 times, and the average results of IOU were equal to 0.7713 for the training set and 0.7134 for the testing set.

Although according to the statistics of the learning results, the segmentation results could be much more accurate, it was decided that this was not necessary because the segmentation results were to be transformed into a bounding boxes determined by the coordinates of the extreme points of the segmented cells.

Cell classification. For the microorganisms cell classification step (the second step of described process), a deep convolutional neural network (CNN) based on Xception net [13] was developed. As a result of the operation of this network, individual images of microorganisms cells were assigned to one of the classes of analyzed microorganisms. The input size of the image in the modified Xception model was limited to the smallest possible size of 71×71 pixels. In fact, this size is still much larger than the size of the detected (segmented) microorganisms as the average size for the largest cells (*Candida dubliniensis*) was 91×93 pixels, and for the smallest cells (*Streptococcus agalactiae*) – 16×16 pixels.

The Xception network mostly consists of blocks of layers: convolution (looking for features in the image), batch normalization (preventing model overfitting), activation (adding non-linearity to the calculations). The input size of the DNN had a particular effect on the final result of classification, as the successive layers of the network were becoming smaller and smaller, with the result that some information may be lost, and thus the classification may not be exact. With the selected input size of DNN (71×71), the last block of layers was only 3×3 pixels, which made it

very difficult for the network to correctly classify cells. To improve the results, the last layers were modified by reducing the number of considered network parameters (2 blocks of layers: dropout and batch normalization).

Data to train DNN for cell classification. Images described in paragraph II were used again, but this time as data for the classifying network. For each of the 19 microorganisms, 20 ($19 \times 20 = 380$) images were selected and the cell detection process described in paragraph II was performed on each one of them. Results of detection were saved as single images (one image per cell) in 19 directories. Additionally, an extra (twentieth) class was created for outliers microorganisms, which with their size (too small or too large) were significantly different as incorrectly marked microorganisms in the image. In this way, 20 classes were created for the image classification CNN. Among these 20 classes there were 4 classes of specific microorganisms (*Candida albicans*, *Escherichia coli*, *Pseudomonas aeruginosa*, *Staphylococcus aureus*).

The sum of cells of some microorganisms were much bigger than others, thus a common upper sum limit was adopted, and in this way 2,000 images of detected cells were randomly selected for each of the 20 classes. The only exception was the *Candida tropicalis* class, for which only 968 cells were collected as single images.

Training and testing of classification. The CNN training process was performed in accordance with the generally applicable principles in deep learning [14–16]. The data set was divided in the proportion of 7:3 into a training set and the validation set.

In order to improve the learning process and prevent overfitting of the model, a data augmentation mechanism (zoom, rotation, horizontal and vertical flipping), selected experimentally by trial and error, was implemented.

Learning the CNN took 200 epochs, with a batch size equal to 32, learning rate equal 0.0001, Adam as the optimization method, and categorical cross-entropy as the loss function. The categorical accuracy metric was used to assess the effectiveness of the learning process.

After the training process, the trained weights of the CNN model was tested on a data set independent of the training set, and the result summarized in a confusion matrix (Fig. 2), and the Receiver Operating Characteristic (ROC) [17] curve (Fig. 3) for multi-class variant was plotted (the closer to the upper left corner, the more accurate the results of the classification; the closer to the $y = x$ curve, the more randomly the classifier behaves).

It is worth noting that the proposed computational methodology, as well as the exact methodology of sample preparation and staining, together with the software of the MDS, form mutually dependent parts and are provided only to cosmetic companies cooperating with the Consortium.

The trained CNN, for the test data, obtained an average accuracy of 80%, but a more detailed analysis of the confusion matrix showed that for some microorganisms the obtained results were much higher (*Candida* cells: over 95% accuracy) or lower (*Proteus vulgaris* cells: 59% accuracy).

From the point of view of the SDM, the most important point was to recognize specific microorganisms: *Candida albicans* – 97% accuracy, *Escherichia coli* – 76% accuracy, *Pseudomonas aeruginosa* – 70% accuracy, and *Staphylococcus aureus* – 85% accuracy.

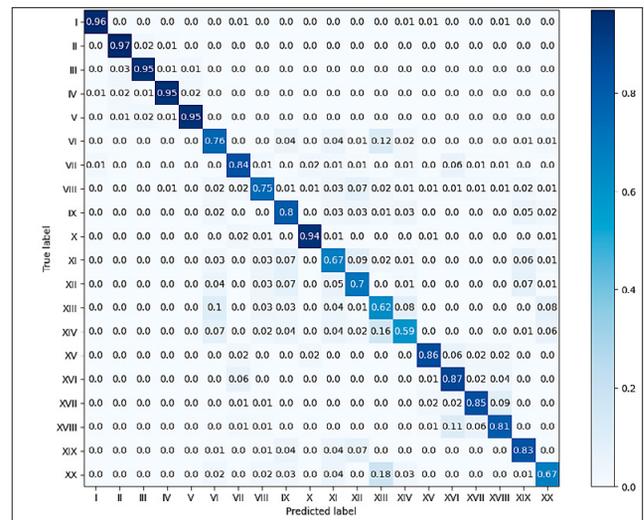


Figure 2. Confusion matrix of classification microorganisms. I – *B.subtilis*; II – *C.albicans*; III – *C.dubliniensis*; IV – *C.glabrata*; V – *C.tropicalis*; VI – *E.coli*; VII – *E.faecalis*; VIII – outliers cells; IX – *K.pneumoniae*; X – *L.monocytogenes*; XI – *M.morganii*; XII – *P.aeruginosa*; XIII – *P.mirabilis*; XIV – *P.vulgaris*; XV – *R.equi*; XVI – *S.galactiae*; XVII – *S.aureus*; XVIII – *S.epidermidis*; XIX – *S.marcescens*; XX – *S.typhimurium*.

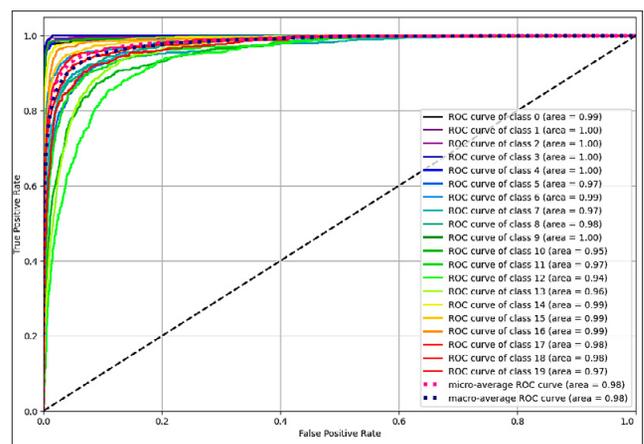


Figure 3. ROC curve of all classes. The smallest value of area under the curve (AUC) is for 12 class (indexed from 0) -- *P.mirabilis*

Acknowledgements

The project was co-financed by the European Union from the European Regional Development Fund under the Intelligent Development programme. The project implemented under the competition 1 / 4.1.4 / 2018 / POIR: 'Microorganism detection system – SDM', National Centre for Research and Development.

REFERENCES

1. Norma PN-EN ISO 17516:2014-11: Kosmetyki–Mikrobiologia–Limity mikrobiologiczne.
2. Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products. <https://eur-lex.europa.eu/eli/reg/2009/1223/oj> (access: 2021.11.07).
3. Fukushima H, Katsube K, Hata Y, Kishi R, Fujiwara S. Rapid separation and concentration of food-borne pathogens in food samples prior to quantification by viable-cell counting and real-time PCR. *Appl Environ Microbiol.* 2007; 73(1): 92–100. doi: 10.1128/AEM.01772-06
4. Pertoff H. Fractionation of cells and subcellular particles with Percoll. *J Biochem Biophys Methods.* 2000; 44: 1–30. [https://doi.org/10.1016/S0165-022X\(00\)00066-X](https://doi.org/10.1016/S0165-022X(00)00066-X)

5. Garbacz M, Malec A, Duda-Saternus S, Suchorab Z, Guz Ł, Łagód G. Methods for Early Detection of Microbiological Infestation of Buildings Based on Gas Sensor Technologies. *Chemosensors*. 2020; 8: 7. doi: 10.3390/chemosensors8010007
6. Bartholomew JW, Mittwer T. The Gram stain. *Bacteriol Rev*. 1952; 16(1): 1–29. doi: 10.1128/br.16.1.1-29.1952
7. Beveridge TJ. Use of the gram stain in microbiology. *Biotech Histochem*. 2001; 76(3): 111–8.
8. Smith AC, Hussey MA. Gram Stain Protocols. *Am Soc Microbiol*. 2016.
9. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation, *CoRR*: <http://lmb.informatik.uni-freiburg.de/> 2015 (access: 2021.11.07).
10. Beucher S. The Watershed Transformation Applied to Image Segmentation, *Scanning Microscopy*, 1992, <https://www.researchgate.net/publication/303201436> (access: 2021.11.07).
11. Bankhead P, Loughrey MB, Fernández JA, et al. QuPath. Open source software for digital pathology image analysis. *Sci Rep*. 2017; 7: 16878. doi.org/10.1038/s41598-017-17204-5
12. Redmon J, Farhadi A. YOLOv3: An Incremental Improvement, *CoRR*, 2018, <https://arxiv.org/pdf/1804.02767.pdf> (access: 2021.11.07).
13. Chollet F. Xception: Deep Learning with Depthwise Separable Convolution, *CoRR*, 2017 <https://arxiv.org/pdf/1610.02357.pdf> (access: 2021.11.07).
14. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; 521: 436–444.
15. Alzubaidi L, Zhang J, Amjad I, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data*. 2021; 8(1): 53. doi: 10.1186/s40537-021-00444-8
16. Shrestha YR, Krishna V, von Krogh G. Augmenting Organizational Decision Making with Deep Learning Algorithms: Principles, Promises, and Challenges. *J Bus Res*. 2021; 123: 588–603. <https://doi.org/10.1016/j.jbusres.2020.09.068>
17. Fawcett T. An introduction to ROC analysis. Elsevier; 2005.