# Cloud Infrastructure for Skin Cancer Scalable Detection System

Pavels Osipovs*
Faculty of Computer Science and Information Technology
Riga Technical University
Riga, Latvia
pavels.osipovs@gmail.com

Dmitrijs Bliznuks
Faculty of Computer Science and Information Technology
Riga Technical University
Riga, Latvia
dmitrijs.bliznuks@rtu.lv

Ilona Kuzmina
Biophotonics Laboratory, Institute of Atomic Physics and Spectroscopy
University of Latvia
Riga, Latvia
ilona.kuzmina28@gmail.com

***Abstract*— Skin cancer diagnostics is one of the medical areas where early diagnostic allows achieving patients' high survival rate. Typically, skin cancer diagnostic is performed by dermatologist, since the amount of such specialists is limited, mortality rate is high [1]. By creating the low cost and easy to use diagnostic device, it is possible to bring skin cancer diagnostic to primary care physicians and allow to check much more persons and diagnose skin cancer on the early stages. There are several existing devices, that provide skin cancer diagnostics [2]. Most of them process the skin images locally and have limited diagnostic capabilities; some of them send images to dermatologists for manual analysis to achieve higher diagnostic quality. Therefore, there is a lack of diagnostic quality or response time.**

**To be able to use the latest diagnostic algorithms and still have fast acting automated diagnostic system, we propose using distributed cloud-based system. In that system, diagnostic device is used only for image acquisition under special multispectral illumination (405nm, 535nm, 660nm and 950nm). Obtained skin imaged are sent further to cloud system for analysis and diagnostic results visualization. By means of proposed approach, images could be processed by using the same Matlab [3] algorithms [4] that skin cancer research team is using. That will eliminate the need of adopting each algorithm to a specific architecture of diagnostic device. Moreover, the proposed system keeps relation between multiple skin analysis from each patient and could be used to track skin lesions changes in time. Proposed cloud system has architecture that allows fast scaling according to real-time requirements. Proposed system uses central load balancing server, that accepts diagnostic requests and sends image processing request to less loaded Matlab processing station. In case of high load, balancing server can launch an additional processing station. Therefore, it brings main cloud system advantages – efficient resource usage and fast adopting to current needs by increasing processing power.**

**The system is being tested in ongoing European project by the biophotonic research team and medical personal. The results of clinical testing will follow after completing first stage of clinical tests.**

***Keywords — cloud processing, skin cancer, automated diagnostics***

## I. Introduction

The use of a portable non-contact diagnostic device (SD) for early diagnosis of malignant lesions on the skin helps even a nonprofessional in skin analysis (for example, a family doctor), to give a qualitative conclusion on a suspicious skin area. If there are suspicions, the patient is more likely to visit a dermatologist. And given the high accuracy of the diagnostic algorithm used [5], this device can also help the dermatologist draw attention to clearly suspicious formations and do not waste time on long analysis of less dangerous skin defects. In addition, even an expert may not pay attention to signs of malignancy due to his own fatigue, or other distractions, whereas automated analysis has a constant level of accuracy.

In the case of using such devices, it is possible to use several different implementations of the system architecture. What approach to use, the creators of the system choose depending on the specifics of the task. Each of them has both advantages and disadvantages. The authors offer the following classification of implementation options:

- Smart diagnostic device (SD-S);
- Simple device + offline human expert (SD-OE);
- Simple device + remote human expert (SD-RE);
- Simple device + automatic analyze system (SD-AA).

### *A. Smart diagnostics device (SD-S)*

In the case of using a smart device, all the logic of scanning, analyzing the resulting images and displaying the result, everything is done directly on the device [6]. All algorithms and programs are stored in the SD-S memory and, in general, cannot be changed without updating it. The doctor, as a device user, has limited options for configuring customizable options available in the menu of the software used. Such a device can work both in fully automatic mode and in conjunction with connected PCs, for printing and printing more detailed information. In addition, it can be connected to the Internet, but basically only to send data to remote services, whereas all analysis steps are performed using only the resources of the device itself. Since device requires processing the images locally, hardware should have high processing power. Therefore, the price of the device should be relatively high. Moreover, in case of small, portable device, it could not effectively process Matlab and similar scripts (that are typical in biophotonics), therefore scripts should be redesigned and adopted to the specific hardware architecture. That requires additional time and might negatively affect the quality of the diagnostic.

### *B. Simple device with offline human expert (SD-OE)*

This is the simplest of the approaches considered. The scanning device only takes a picture and displays all the images on the screen or sends them to the connected PC. The doctor, as the user himself is responsible for the analysis and

evaluation of the results. Device is used only for image acquisition and transmission. The cost of device itself is relatively small and there are no other limitations. However, there is a dependence of the quality of the result on the qualification of the doctor performing the examination and interpreting the results of the scan.

### *C. Simple device with remote human expert (SD-RE)*

This approach is similar to SD-OE, the difference is that the device [9] sends the received images to a remote expert who analyzes them. The analysis can be performed immediately, in real time or after receiving the images. Since a remote expert can be much more qualified than a device operator, this increases the effectiveness of such an analysis. However, the imaging device has still got issues with quality of illumination and skin pictures. As well as there is a dependence on the capabilities of the remote expert, the availability of time for analysis, and other human or technical factors may affect the quality of the analysis, for example, the well-being of the assessor or the connection with the Internet. Such an approach is popular in existing diagnostic applications [7, 8], where image quality and especially illumination might affect diagnosis precision.

### *D. Simple device with remote automatic analysing system (SD-AA)*

This approach allows to obtain a stable, predictable analysis quality when using an inexpensive diagnostic device. Device is kept simple and cheap, since it is used just to obtain and transmit images to the remote servers and does not contain high processing power hardware. Even though device is simple, it still does not require skilled personal for operation, since diagnostics is done by remote automated algorithms. That allows bringing diagnostic results almost instantly, without waiting for image analysis by skilled medical person. In the case of an increase in the number of scanning devices, there is no need to hire more staff, you just need to increase the number of computing nodes. The quality of the result is limited only by the accuracy of the algorithms used. Automated diagnostic algorithms quality is equal to skilled dermatologists, therefore such a simple device could replace dermatologists at least on the first stage of skin lesions diagnostics.

Along mentioned advantages, SD-AA approach allows to bring newest improvements in diagnostic algorithms without delays if compared to Smart diagnostic device (SD-S) approach. SD-AA approach uses the same diagnostic scripts, that are used by biophotonic scientists. Typically, after testing diagnostic algorithm in laboratory, the script should be reprogrammed and adopted to the specific device architecture. That will bring noticeable delay. Moreover, biophotonic scientists' group might not have programmers at all, that could perform algorithm adoption. In SD-AA approach script is transferred to the remote servers without any changes and makes it instantly available for the diagnostic devices. Matlab environment is typical for biophotonic scientists and it will be used further for realization of proposed approach.

### *E. Comparative evaluation of approaches*

To compare the effectiveness of the approaches, each of them was evaluated according to the criteria described below. The main criteria for assessing the quality of the resulting system are its following characteristics:

- Diagnostic quality (DQ);
- Cost of the device (DP);
- Cost of one analysis (SAP);
- Time of obtaining the results of the analysis (processing speed) (AT);
- Ability to make changes (adjustments) in the analysis process (AU);
- The cost of increasing system capacity (CISC).

The value for each criterion is estimated on a three-point scale, where 1 – is bad, 2 – is normal, and 3 – is good. Values were calculated as average from three dermatology experts.

Diagnostic quality (DQ) evaluation for each system:

- SD-S. Diagnostic quality depends on device built in algorithms, that typically are close to dermatologist skills. The result is slightly less than maximum – 2,5.
- SD-OE. Allows achieving highest diagnostic quality, but only in case of skilled medical personal, that is used to analyze the images. The results are maximal – 3, since images are judged by expert.
- SD-RES. Results will achieve maximum - 3, but only if high quality illumination and camera are used.
- SD-AA. Diagnostic quality depends on device built in algorithms, that typically are close to dermatologist skills. The result is slightly less than maximum – 2,5.

Device price (DP) evaluation for each system:

- SD-S. Device price is highest, since high processing power is required. Criteria is lowest – 1.
- SD-OE, SD-RES and SD-AA. Device price is low, since device does not contain high processing power hardware. Criteria is highest – 3.

Single analysis price (SAP):

- SD-S. Price is lowest, since device owning expenses are low and result is obtained without spending skilled medical labour. Criteria is highest – 3.
- SD-OE. Price is highest, since requires high skilled medical labour to analyze the results. Criteria is lowest – 1.
- SD-RES. Price is less that SD-OE approach, since skilled medics can better utilize their time. Right after judging the images, they can return to their job. Criteria is average – 2.
- SD-AA. Price is higher than in SD-S approach, since remote servers require monthly payments. Criteria is above average – 2,5.

Analysis time (AT):

- SD-S. Time delay from analysis till results is low, since device uses automatic analysis system and receives results instantly. Criteria is highest – 3.
- SD-OE. Time delay from analysis till results is low, since medical personal can instantly give diagnostic result. Criteria is highest – 3.

- SD-RES. Time delay from analysis till results is high, since remote expert might perform analysis only when free time is available. Criteria is lowest – 1.
- SD-AA. Time delay from analysis till results is low, since device uses automatic analysis system and receives results instantly. Criteria is highest – 3.

Analysis process adjustments (AU):

- SD-S. Making any changes in the device is expensive and time-consuming process. Typically, devices are rarely updated or not updated at all. Criteria is lowest – 1.
- SD-OE. Since skilled medical is in charge of the whole process, changes are easy to implement. Criteria is highest – 3.
- SD-RES. Skilled medical is in charge of the diagnostic process, but changing the device might be complicated. Criteria is average – 2.
- SD-AA. Diagnostic can be easily changed, but changing the device might be complicated. Criteria is average – 2.

The cost of increasing system capacity (CISC):

- SD-S. High costs, since it will need to pay full system price for each additional device and device is expensive. Criteria is lowest – 1.
- SD-OE. High costs, since it will require additional high skilled personal at each additional device. Criteria is lowest – 1.
- SD-RES. Low costs, since remote personnel will spend additional only for analyzing images. Maximal system capacity is limited to time, that skilled personnel can spend on image analysis. Criteria is below highest – 2.5.
- SD-AA. Lowest cost, since diagnostic process is automated and maximal capacity is almost unlimited. Criteria is highest – 3.

Each application might have its own priorities concerning diagnostic quality, price and other criteria. As for average application all criteria will have same weight. Total results are calculated as sum of all values (see Table I).

TABLE I. COMPARATIVE ASSESSMENTS OF THE CHARACTERISTICS OF APPROACHES (HIGHER VALUE IS BETTER)

| **Approach Criteria** | *DQ* | *DP* | *SAP* | *AT* | *AU* | *CISC* | ***Result*** |
|---|---|---|---|---|---|---|---|
| *SD-S* | 2,5 | 1 | 3 | 3 | 1 | 1 | **11.5** |
| *SD-OE* | 3 | 3 | 1 | 3 | 3 | 1 | **14.0** |
| *SD-RES* | 3 | 3 | 2 | 1 | 2 | 2.5 | **13.5** |
| *SD-AA* | 2,5 | 3 | 2.5 | 3 | 2 | 3 | **16.0** |

The SD-AA approach (16 points) is the most effective in terms of the aggregate values of the evaluation criteria. It is used to build the system described in this article.

## II. DIAGNOSTIC DEVICE USED

As seen before, device with remote analysis is optimal. To further increase its effectiveness, device uses modular structure, where most of the system modules have standard interfaces and are easily interchangeable. That does not require high electronics and programming skills to make changes in hardware, allowing even faster adoption of new diagnostic systems by biophotonic scientists. Camera and communication modules (WiFi and 4G cellular modem) are using USB interface, central system for process control is Raspberry. Modules are popular and it will be easy to find replacement. Of course, illumination module is custom made, since there are specific requirements that could not be found in market available devices. Nevertheless, illumination module has modular structure as well and allows making changes in illumination parameters relatively easy, without affecting other modules. On the Figure 1 device overall

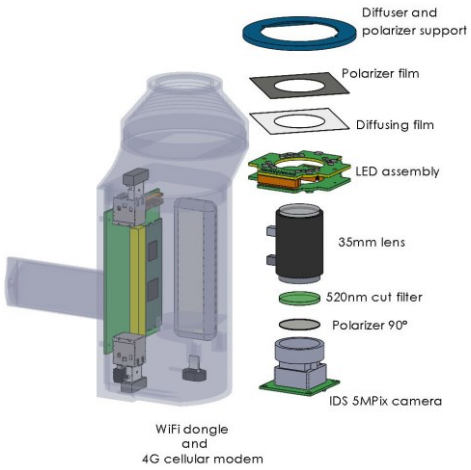


Fig. 1 General structure of the created system

design and each module is depicted.

The main modules are:

1) Optics and illumination. Device uses custom built board with five groups of LEDs — 405nm, 535nm, 660nm, 950nm and white. Each group has four identical LEDs distributed in circle.

2) Central processing module. Based on custom built circuit board with quick change connector for Raspberry Pi Compute module. It will allow quick upgrading to the future Raspberry modules without changing the custom board.

3) Wireless connectivity modules. Devices uses WiFi and 4G cellular modems with USB interface. Similarly, that will allow quick and easy future upgrades.

## III. OVERALL STRUCTURE OF THE CREATED SYSTEM

The SD-AA approach is based on the use of the cloud-based system architecture. The implementation consists of one basic module (*BM*) - which is the main program code for checkyourskin.eu, and a module installed on each activated computation node (*CN*). BM provides a set of system administration panels, as well as a set of programming interfaces (APIs) for interacting with compute nodes and diagnostic devices. Each *CN* contains only tools for interaction with Matlab and an API for interaction with the *BM*.

Figure 2 shows the general structure of the created system. The central node - checkyouskin.eu is a web server that contains the code of the administration system, and also acts as a router and load balancer, redistributing the incoming images between available *CN*s for analysis.

Each *CN* performs only the required calculations and nothing more. Additional *CN* can be dynamically added or deactivated as the number of images arriving for analysis increases or decreases. At the moment, all calculations on the

*CN* are performed only using *Matlab*. In this case, one *CN* can provide several parallel computational *Matlab* processes.

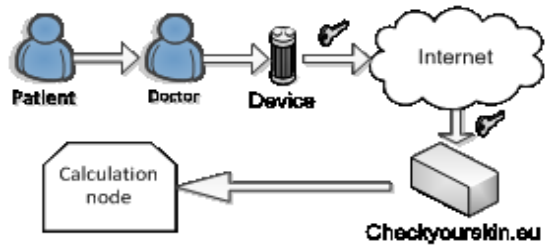


Fig. 2 General structure of the created system

## IV. Common Architecture of the Main System Module

The common structure of the main system module is shown in Figure 3. The main PHP programming language is used by the *Laravel*. *Nginx* web-server is used to provide communication of the system with the Internet. Using *Laravel* allowed us to focus on the main application logic implementation, since the framework already provides most of the low-level functionality.

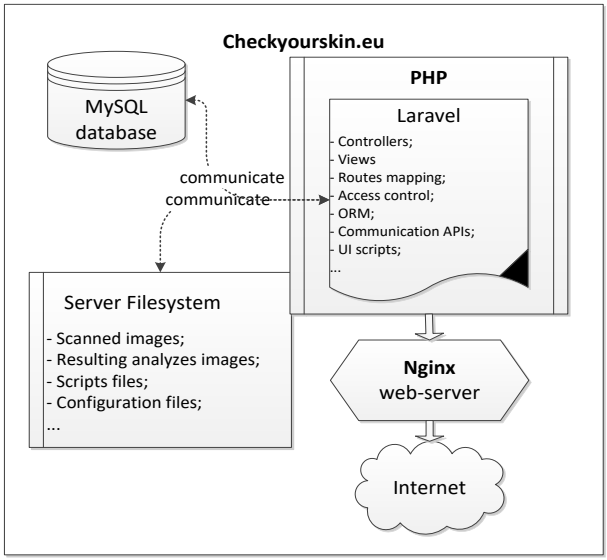


Fig. 3 Structure of the created system Main Module

## V. System Efficiency

The overall efficiency of a distributed system depends on a large number of parameters. Some of them can be controlled and some not. The main criteria for the effectiveness of this system is the time between sending scanned images and obtaining the results of their analysis by the doctor. In part, it can be reduced by using more *CN*, more efficient load balancing policies between *CN*, optimizing the *M*-scripts used by the implementation of analysis algorithms in the *Matalb* environment. However, other characteristics cannot be changed directly when creating the system, for example, the network bandwidth between *SD* and *checkyourskin.eu*, the data transfer rate between internal nodes of the system, and the time spent by *Matlab* itself.

### A. System effectivity per cost

The main factor influencing the time of obtaining the results of the analysis is the speed of image processing by the *CN*. And its reduction will reduce the time the doctor receives scan results. The following basic ways to do this are possible:

- Increase the number of *CN*;
- Increase the calculation power of all or some *CN*;
- Increase the effectiveness of the *CN* selection policy for processing the incoming analysis.

For greater efficiency, all these methods can be combined. Since maintenance of each *CN* requires money, then there is the problem of the optimal ratio of the number and capacity of the available *CN*s. Depending of the type of computations performed, in order to ensure a better processing time, in some cases it is more efficient to have a large number of low-power *CN*s, in others, a small number of powerful nodes will be more effective.

The effectiveness of *CN* can be estimated as the maximum number of simultaneous computational processes (*PW*) that it can provide. Basically, the PW value depends on the CPU power, the amount of operational memory (RAM) and the specificity of the problem being solved. Different its configurations can provide different amounts of PW in different tasks (Maximal Process Count - *MPC*).

Since there is an accurate data on the maximum number of processing flows that can be performed on a single node, the following approach is used to calculate the optimum quantity, and power, of the nodes of the system.

Let assume, that:

$MPC_x$ - one of the possible server configurations;

$PW(x)$ - the maximum number of computational processes on the node for the configuration *x*;

The variants of cost and performance of various configurations are given at Table II.

TABLE II. Cost and Performance of Configurations

| *MPC* | *MPC_1* | *MPC_2* | *MPC_3* | *MPC_4* | *MPC_5* |
|---|---|---|---|---|---|
| *Price* | € 100 | € 150 | € 210 | € 273 | € 327 |
| *PW* | 5 | 12 | 19 | 29 | 55 |

When effectivity of configuration *x* is calculated like $Eff(MPC_x) = \frac{PW_x}{Price_x}$. To do this, a cost table is created for a different number of nodes, for all possible configurations (Table III).

TABLE III. Cost of Configurations Setups

| *MPCs count* | *MPC_1* | *MPC_2* | *MPC_3* | *MPC_4* | *MPC_5* |
|---|---|---|---|---|---|
| *1* | € 100.00 | € 150.00 | € 210.00 | € 273.00 | € 327.60 |
| *2* | € 200.00 | € 300.00 | € 420.00 | € 546.00 | € 655.20 |
| *3* | € 300.00 | € 450.00 | € 630.00 | € 819.00 | € 982.80 |
| *4* | € 400.00 | € 600.00 | € 840.00 | € 1,092.00 | € 1,310.40 |
| *5* | € 500.00 | € 750.00 | € 1,050.00 | € 1,365.00 | € 1,638.00 |

From Table III can be seen that with a budget of 1000 €, it is possible to use 3x*CN* configurations of type *MPC_5*, 4x*CN* configurations of type *MPC_4* and 5x*CN* configurations of type *MPC_3*.

Further, for a different number of nodes and configurations, a matrix of efficiency values is calculated for each possible configurations (Table IV).

TABLE IV. VALUES OF THE EFFICIENCY OF VARIOUS CONFIGURATIONS WHERE 1.0 – FULL UTILISATION

| *MPCs count* | *MPC_1* | *MPC_2* | *MPC_3* | *MPC_4* | *MPC_5* |
|---|---|---|---|---|---|
| *1* | 0.05 | 0.08 | 0.09 | 0.11 | 0.17 |
| *2* | 0.10 | 0.16 | 0.18 | 0.21 | 0.34 |
| *3* | 0.15 | 0.24 | 0.27 | 0.32 | 0.50 |
| *4* | 0.20 | 0.32 | 0.36 | 0.42 | 0.67 |
| *5* | 0.25 | 0.40 | 0.45 | 0.53 | 0.84 |

For the selected three configurations (*MPC_3, MPC_4, MPC_5*), the efficiency values are given in Table V. As can be seen, in conditions of the initially specified target performance parameters, the use of three *CN*s configurations of type *MPC_5* will be most effective.

TABLE V. THE EFFECTIVENESS OF SELECTED CONFIGURATIONS

| *Configuration* | *Effectivity* |
|---|---|
| 5 x MPC_3 | 0.45 |
| 4 x MPC_4 | 0.42 |
| 3 x MPC_5 | 0.5 |

### B. *System extendibility for new types of devices*

The system is based on a flexible architecture that allows not only to use devices and algorithms implemented in the context of contactless skin analysis, but also other devices that perform other tasks, but also require a cloud part for the organization and execution of resource-intensive calculations.

For each type of supported scanning device, a pool of computing nodes is created (Figure 5). Since the requirements for processing power for different types of *CN* can be very different, this approach allows you to flexibly organize the processing power for each of the supported types of *SD*.

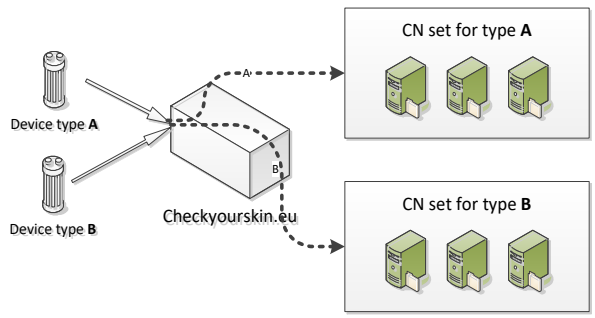


Fig. 5 Distribution of compute nodes depending on the type of scanning devices

### C. *System extendibility for new types of analyse algorithms*

Same as with the addition of new *SD* types, new data processing algorithms for already supported scanning devices can easily be added to the system (Figure 6). All processing is performed on internal computing nodes, and all the algorithms used can be centralized in the event that you need to add a new one or update an existing algorithm.

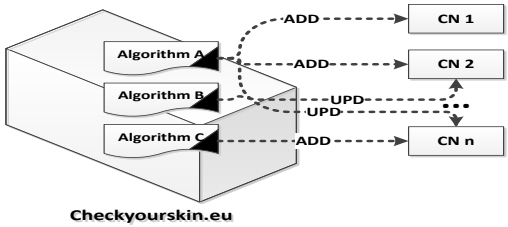


Fig. 6 Centralized algorithms management

## VI. CONCLUSION

The article describes the architecture of the distributed cloud-based system, allowing to organize data processing from a large number of non-invasive skin cancer diagnostics devices. Approaches to the evaluation of the effectiveness of the system are given and the schemes of organization of distributed computing are shown.

The created system allows not only to organize data processing of existing scanning devices, but also allows the addition of new types of devices, and new algorithms for data processing and analysis. Cloud approach, which combines a large number of distributed resources, allowed us to create an expensive system consisting of a large number of easily accessible devices. This in turn makes it possible to make early diagnostics of skin diseases in large quantities and in places that previously did not have access to such technologies. An additional advantage is the centralized management of the algorithms used for analysis, which allowed us to make updates quickly, as well as add new methods for analysing and processing the data.

## REFERENCES

[1] Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh JWW, Comber H, Forman D, Bray F. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. Eur J Cancer. doi: 10.1016/j.ejca.2012.12.027, 2012

[2] D. Bliznuks, D. Jakovels, I. Saknite and J. Spigulis, "Mobile platform for online processing of multimodal skin optical images: Using online Matlab server for processing remission, fluorescence and laser speckle images, obtained by using novel handheld device," 2015 International Conference on BioPhotonics (BioPhotonics), Florence, 2015, pp. 1-4. doi: 10.1109/BioPhotonics.2015.7304024

[3] *MATLAB* and Statistics Toolbox Release 2014a The MathWorks, Inc., Natick, Massachusetts, United States.

[4] Bļizņuks, D., Kuzmina, I., Boločko, K., Lihachev, A. Image Quality Enhancement for Skin Cancer Optical Diagnostics. Proceedings of SPIE, 2017, Vol.10592: Biophotonics—Riga 2017, pp.1-8. ISSN 0277-786X. e-ISSN 1996-756X. Available from: doi:10.1117/12.2297579.

[5] Lihacova, I., Bolocko, K., and Lihachev, A., "Semi-automated non-invasive diagnostics method for melanoma differentiation from nevi and pigmented basal cell carcinomas", Proc. SPIE 10592, Biophotonics—Riga 2017, 1059206 (2017).

[6] SkImager: a concept device for in-vivo skin assessment by multimodal imaging, Janis Spigulis, Uldis Rubins, Edgars Kviesis-Kipge, Oskars Rubenis, Proceedings of the Estonian Academy of Sciences, 2014, 63, 301–308

[7] DermDOC portable dermatoscope description, http://www.dermamedicalsystems.com/index.php?menu_id=118

[8] DermLite handheld dermatoscope, http://dermlite.com/

[9] Joanna Jaworek-Korjakowska* and Ryszard Tadeusiewicz, "Design of a teledermatology system to support, the consultation of dermoscopic cases using, mobile technologies and cloud platform", Bio-Algorithms and Med-Systems. 2015-02-28 | DOI: https://doi.org/10.1515/bams-2015-0004

This work has been supported by European Regional Development Fund project 'Portable Device for Non-Contact Early Diagnostics of Skin Cancer' under grant agreement # 1.1.1.1/16/A/197. This work has been supported by European Regional Development Fund project 'Portable Device for Non-Contact Early Diagnostics of Skin Cancer' under grant agreement # 1.1.1.1/16/A/197.