Journal of King Saud University – Computer and Information Sciences xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

Using convolutional neural networks for corneal arcus detection towards familial hypercholesterolemia screening

Tomasz Kocejko [a,*], Jacek Ruminski [a], Magdalena Mazur-Milecka [a], Marzena Romanowska-Kocejko [b], Krzysztof Chlebus [b], Kang-Hyun Jo [c]

[a] Faculty of Electronics, Telecommunications and Informatic, Gdansk University of Technology, Department of Biomedical Engineering, Gdansk, Poland
[b] Gdansk Medical University, National Center for Familial Hypercholesterolemia, Gdansk, Poland
[c] Department of Electrical Engineering, University of Ulsan, Ulsan, South Korea

ARTICLE INFO

Article history:
Received 12 April 2021
Revised 6 August 2021
Accepted 1 September 2021
Available online xxxx

Keywords:
Neural networks
Image analysis
Decision support systems
Computer aided diagnosis
Corneal arcus detection
Familial hypercholesterolemia screening

ABSTRACT

Familial hypercholesterolemia (FH) is a highly undiagnosed disease. Among FH patients, the onset of premature coronary artery disease is 13 times higher than in the general population. Early diagnosis and treatment is essential to prevent cardiovascular diseases and their complications, and to prolong life. One of the clinical criteria of FH is the occurrence of a corneal arcus (CA) among patients, especially those under 45 years old. Therefore, by detecting a CA, it might be possible to reduce the number of undiagnosed FH cases. In this paper, we propose using convolutional neural networks (CNN) for automatic recognition of the presence of a corneal arcus. To achieve this goal, we created a dataset of images of irises containing different stages of CA as well as irises without a CA. The core of the dataset consists of images acquired from patients with a corneal arcus, enroled in the National Centre of Familial Hypercholesterolemia in Gdansk. To increase the number of images, the dataset was complemented with images downloaded from the Internet. This dataset created for training and testing the model consisted of nearly 4000 images. To detect a CA in photographic images, we tested neural network models based on the VGG16, ResNet and Inception architectures. Finally, the performance of the models was evaluated on a set of images acquired from volunteers with a custom mobile application. The accuracy of CA detection in a real life scenario was 88% and the F1 score was 86%

1. Introduction

Familial hypercholesterolemia (FH) is a common autosomal dominant genetic disorder of the lipid metabolism, characterised by a life-course elevated blood low-density lipoprotein cholesterol (LDL-C) concentration, leading to early atherosclerosis and premature coronary heart disease (Raal and Santos, 2012). Mortality due to cardiovascular diseases among FH patients between 20 and 39 years is 100 times higher and the onset of premature coronary artery disease is 13 times higher than in the general population (Hypercholesterolemia, 1999). The prevalence of FH in Poland has been estimated at one in 250 adults (Pajak et al., 2016). Chlebus et al. (2018) estimate that 102,000 patients in the Polish population are undiagnosed (nearly 98%). Through early diagnosis, intensive lifestyle modifications and pharmacotherapy, the incidence of cardiovascular events might be significantly decreased. Among the clinical features of FH, we distinguish a corneal arcus, tendon xanthomatosis, and xanthelasma. A corneal arcus (CA) is a lipid-rich and predominantly extracellular deposit that forms at the corneoscleral limbus. The prevalence of a corneal arcus in the FH population is approximately 50% (Ogura et al., 2016; Pajak et al., 2016). A corneal arcus reflects widespread tissue lipid deposition (Zech and Hoeg, 2008), therefore its presence (especially among people under 45 years old) is an indicator for examination for lipid abnormalities (Fernández et al., 2007).

Image processing can help to determine the presence of a CA, and can therefore play an important role in indicating further examination toward FH, among others. Most CA detection methods rely on an image threshold (using the Otsu algorithm) and histogram operations. Usually, these operators are applied on a

* Corresponding author at: Gdansk University of Technology.
E-mail address: tomasz.kocejko@pg.edu.pl (T. Kocejko).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2021.09.001
1319-1578/

Please cite this article as: T. Kocejko, J. Ruminski, M. Mazur-Milecka et al., Using convolutional neural networks for corneal arcus detection towards familial hypercholesterolemia screening, Journal of King Saud University – Computer and Information Sciences, https://doi.org/10.1016/j.jksuci.2021.09.001

normalised image of an iris. The authors (Ramlee et al., 2011; Songire and Joshi, 2016) state that the values of a histogram of an image with a CA present concentrate near the higher values, as opposed to images of a CA-free iris. A similar approach was proposed by Bhangdiya (2014). Kumar proposed using a mean value, standard deviation, entropy, skewness, kurtosis and a Support Vector Machine (SVM) for corneal arcus classification. A feature-based approach was also presented by Nasution and Kusuma (2009), who used a hybrid N-feature neural network (HNFNN) instead of an SVM for classification.

The rapid development in the area of machine learning (ML) and image oriented deep learning (DL) is allowing solutions to be found for more and more complex tasks. Developments in artificial intelligence (AI) make it attractive for medical applications allowing new areas to be explored where computer algorithms can improve medical processes (Patel et al., 2009; Laksanasopin et al., 2015; Castaneda et al., 2015). Convolutional neural networks (CNN) constitute a valuable method in visual recognition and are also widely applied in the medical field. They are widely appreciated in a variety of applications including breast tumor detection (Rouhi et al., 2015), lung image patches (Li et al., 2014; Ciompi et al., 2015), skin cancer classification (Esteva et al., 2017), radiology workflow triage (Titano et al., 2018), ECG analysis (Khamis et al., 2018) or improving the accuracy of classification of conjunctivitis (Li et al., 2019). Moreover, neural networks proved to be very useful in terms of organs segmentation. In their work (Zhao et al., 2019) presented combination of deep-learning and traditional methods for small organs segmentation despite the limited training data while (Wang et al., 2019) showed how deep-learning approach can resolved the problem of gland segmentation in histology images. Simonyan and Zisserman designed a CNN model (Simonyan and Zisserman, 2014) (VGG) of substantially deeper architecture with 16 or 19 weight layers. The fundamental part of the Inception model is the concept of the inception layer or inception block that essentially is a combination of convolutional layers of different kernel size with their outputs concatenated. The resulting single vector constitutes an input to another block (Szegedy et al., 2016). The ResNet model consists of so-called residual blocks. The idea behind the residual block is to pass the data to the consecutive layers or skip some layers and feed a layer after a few steps. The deep network structure of the ResNet model was designed to overcome the backpropagation problem.

Finding undiagnosed patients is a challenging task. We decided to exploit the fact that corneal arcus is a visible factor of familial hypercholesterolemia. Developing the algorithm and application that can recognize presence of corneal arcus in an photographic image of an eye can influence potential patient's decision to visit a specialist towards lipid screaming.

In this paper, we present a new model inspired by the VGG architecture and its application for corneal arcus detection. The motivation behind this study is to provide a reliable algorithm and a tool for investigating the presence of CA, and via that, screening against potential familial hypercholesterolemia. The algorithm is designed to work with images recorded with a phone camera. In addition to the algorithm and the mobile application, we contributed to the creation of a dataset of irises with a corneal arcus.

The rest of the paper is organised as follows: the methodology is presented in section II. Section III describes the implementation and testing. The obtained results are presented in section IV and discussed in section V. The whole paper is concluded in section VI.

2. Method

Usually, deep learning performs best when trained on a large dataset of images. However, it is quite difficult to build a large dataset of medical images, especially when considering a rare medical condition. In our research, we focused on the detection of the presence of a corneal arcus in photographic images. In general, there are two types of corneal arcus: arcus senilis and arcus juvenilis. Arcus senilis appears as a white/whitish ring or arc around the cornea of the eye. The condition is usually seen in older adults but can affect people of all ages, even appearing at birth. Sometimes it appears in people under 40 and is called arcus juvenilis. In these cases, the rings may be a result of high cholesterol levels in the bloodstream (Desnick et al., 2001; Brewer et al., 2005; Winder et al., 1998).

2.1. Dataset

Collecting the dataset started with images acquired at the National Centre of Familial Hypercholesterolemia (University Clinical Centre Gdansk, Poland). The images of an iris were acquired from 50 patients (34 female, 16 male, age: 27–58). Each patient was asked to expose the affected area of their iris two times (2 images were acquired per patient). Exemplary views are presented in Fig. 1. All of the images were taken by two experts (specialists in cardiology with at least 3 years work experience in the familial hypercholesterolemia clinic). To increase the number of iris images, it was decided to extend the dataset with images obtained from various Internet resources. We queried the Google Image browser for "corneal arcus", "corneal juvenilis" and "corneal senilis". A retrieved image was included into the dataset after the content and the quality of the image were visually inspected and approved by the experts. To avoid false classification based on extreme condition of the eye (evident presence of corneal arcus or extreme degradation of the iris) such images as well as the different instances of the same image were excluded from the dataset. As a result, the dataset labeled as "CA class" contained 195 images. A corresponding dataset of images labeled as "NO CA class", presenting a healthy iris, was generated in similar way.

(a) (b)

Fig. 1. Example images of an iris taken from two different angles. (a) The CA partially exposed. (b) The CA visible around the whole iris.

The entire dataset contained 390 images equally divided into the two classes.

2.2. Data preparation

Because of the different angles of image acquisition, we decided to extract only the iris area. The main assumption was to get the classification based on the medical conditions and not on the presence of additional features such as skin colour, eyelashes etc. To detect the iris, previously designed algorithms for pupil detection were utilised (Kocejko et al., 2009; Kocejko and Wtorek, 2013). The pupil detection algorithm needed some adjustments such as the adaptive threshold and the size of the structuring element used for morphological image operations. In general, the iris extraction algorithm was composed of two parts. The first part was pupil region detection. The V-channel (the YUV color model) was extracted from the original input image. Next, a Gaussian blur was applied along with erosion and closing operations. The pupil region intensity was enhanced by squaring up the image. Next, the adaptive threshold was applied. The assumption was that the pupil region boundaries will be represented by the largest inbound contour which was approximated with a bounding box (square). The pupil center and radius were represented by bounding box parameters. The overview of the pupil detection algorithm is presented in Algorithm 1.

Algorithm 1: pupil detection algorithm

```
Get V-component
Gaussian filtering (kernel size: 9)
Erosion (3x3 rectangular structuring element)
Morphological Opening (5x5 eliptical structuring element)
Intensity enhancement (squaring up the image)
Adaptive Threshold
procedure Modified Longest Segment Detection
    Points ← x,y coordinates of beginning and end of longest vertical segment
    l_vert ← length_vert of longest vertical segment
    Points ← x,y coordinates of beginning and end of longest horizontal segment
    l_hor ← length_hor of longest horizontal segment
    for each row in image do
        l ←Get Segment length
        if l > 0.5 · l_hor and l < 0.9 · l_hor then
            Points ← x,y coordinates of beginning and end of segment
        for each col in image do
            l ←Get Segment length
            if l > 0.5 · l_ver and l < 0.9 · l_ver then
                Points ← x,y coordinates of beginning and end of segment
Fit ellipse to Points
Fit bounding box around ellipse
```

The second part of the iris extraction algorithm relied on the detected pupil size and position. The iris candidates were detected by the circle Hough transform applied on the V-channel image. To limit the number of detected circles, the accepted size of the detected circles was related to the pupil size estimated in the previous step (parameters of the bounding box). The constrains were that the detected circle diameter should be in the range from one to five lengths of the bounding box. Next, the iris candidate was selected based on the minimum distance from the pupil center to the center of the detected circle. The overview of the iris detection is presented in Algorithm 2. Finally the imaged was cropped to the iris bounding box and resized to 512 by 512 pixels. The whole dataset was reviewed for correct iris estimation. An incorrectly detected iris was manually validated. A similar procedure was applied on images representing the second, NO CA class. Additionally, an image mask was applied to separate the iris from the background. It is worth mentioning that the corneal arcus appears on the boundary of the iris. Therefore the mask was also applied over the inner part of the iris. The radius of applied masks were 250 and 170 pixels. The general overview of the algorithm is presented in Fig. 2. We utilised black and white background masks. As a result, two datasets were obtained with different background masks. It was important to verify if different distributions of iris colours in the dataset was an important factor in any image classification problem. To overcome the problem of the differing quality levels of the acquired images, all images were normalised according to formula:

$$X_{norm} = \frac{X - min(X)}{max(X) - min(X)} \quad (1)$$

where: X_{norm} is normalized value, X is original value, $max(X)$ and $min(X)$ are maximum and minimum value of image.

Algorithm 2: Iris detection algorithm

```
Get V-component
Gaussian filtering (kernel size: 9)
Inpaint pupil region
Otsu Threshold
circles ← Hough circle detection
for circel in circles do
    center ← circle center
    dist ← distance between center and pupilcenter
    if dist < mindist then
        mindist ← dist
        radius_iris ← circle radius
        center_iris ← center
Fit bounding box around circle
```

A data augmentation procedure was used to increase the size and diversity of the dataset available for the training phase. Each image was rotated nine times in the range from 1 to 359 degrees. This was done to simulate the presence of a CA in different spatial locations of an eye. As a result, each dataset was increased in size to 3900 images (390 unrotated images and 9x390 rotated ones).

Fig. 3 presents example of images from each dataset. The manual procedure of dividing the data into training, test and validation set was adopted to avoid including an image acquired from the same person in the training and validation set, or in the training and test set. It has to be underlined that images from the Internet and the clinic were stored in separate folders. Images acquired from the interent were divided into the training, validation and test sets. Images acquired from the clinic were stored in separate subfolders for each subject (for each subject enrolled into studies separate subfolder was created to store obtained images). The subfolders were divided into training, test and validation set. Finally, images from all training, validation and test folders and their subfolders (internet and clinic) were moved to the corresponding training, validation and test folder.

The described algorithms of pupil and iris detection were implemented using Python3.6 and the OpenCV library. The software for manual iris region correction was written in QT C++.

2.3. Model architecture

In this study, the most common CNN architectures were evaluated in terms of the corneal arcus presence classification problem including Inception v3, ResNet-50 and VGG-16. An overview of the

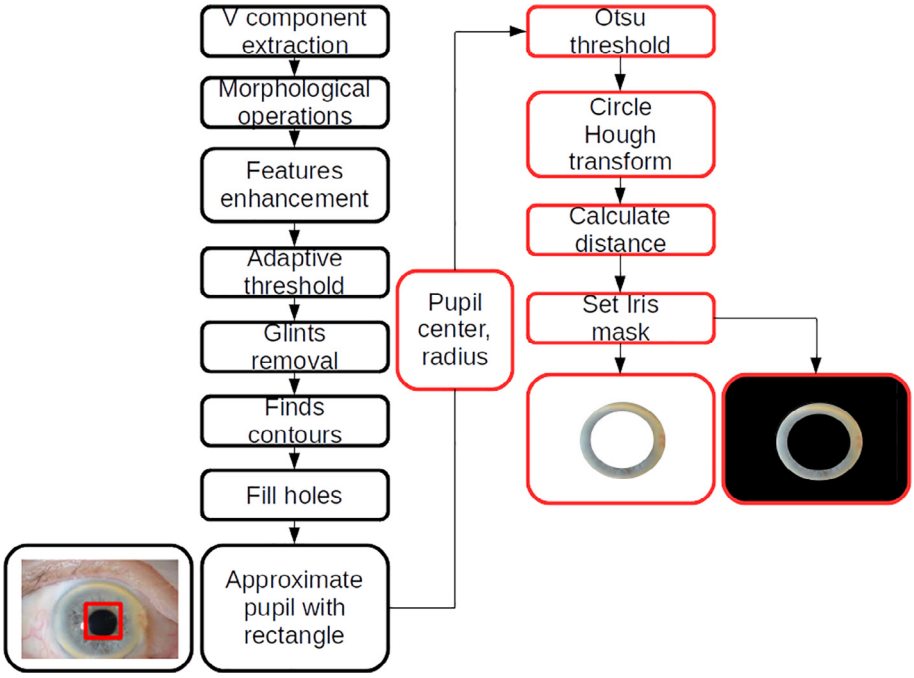

Fig. 2. Simplified diagram of iris detection and isolation.

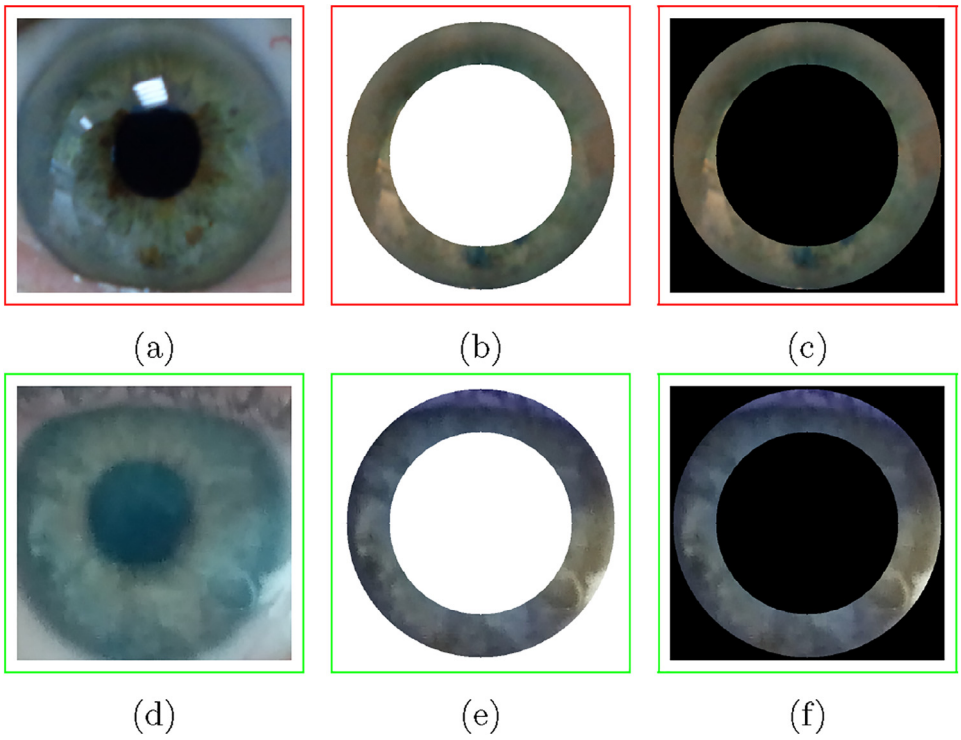

Fig. 3. Examples of images included in a dataset, (a) and (d) are original images, (b) and (e) are normalized images with white mask applied, (c) and (f) are normalized images with black mask applied.

designs of these models is presented in Fig. 4. The choice of these models was justified by the models' performance in the ImageNet competition. The final layers of the models were experimentally chosen to obtain best classification results. Additionally, a batch normalization layer was added before the dense layer. According to Simon et al. (2016), batch normalisation appears to be crucial for the successful training and convergence of the model. In general, networks with batch normalisation train faster and converge much more quickly. Such a network should produce more reliable and consistent results.

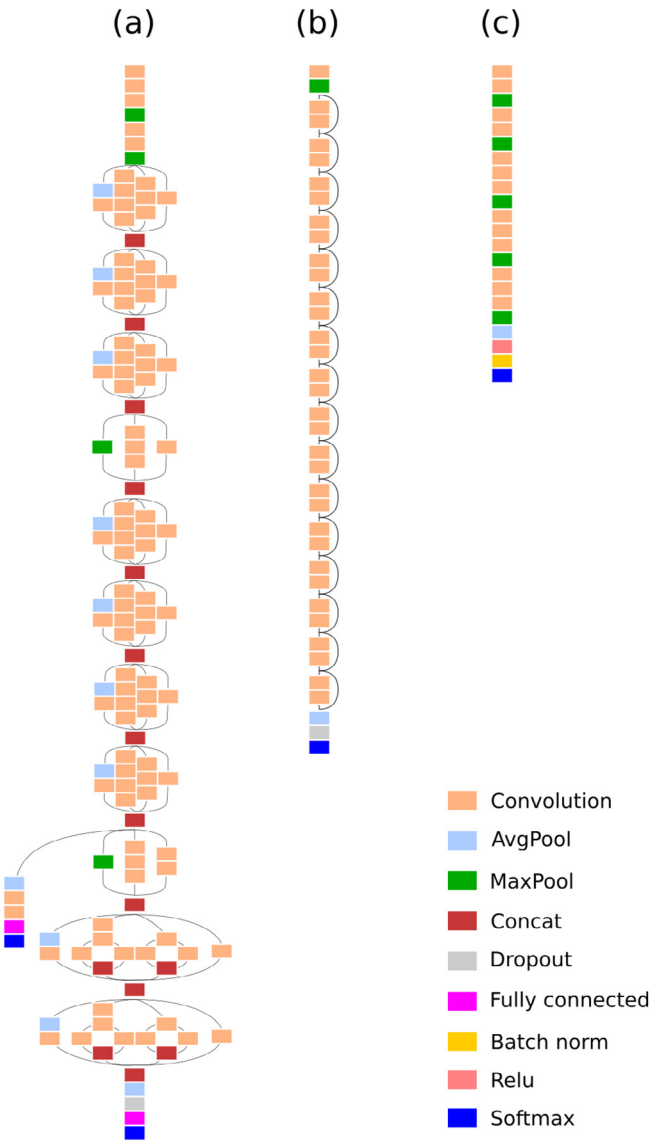

Fig. 4. Overview of utilized architectures: (a) Inception v3, (b) ResNet 50, (c) VGG-16.

2.4. Training the models

A series of experiments were performed to identify the best models and to establish which dataset should be used to train our models. Each model was trained for over 100 epochs. The following methods to prevent the overfitting problem were used: data split (training, validation and test), cross-validation (five folds), data augmentation, dropout/batch normalization and early stopping. Using callback functions allowed us to save the weights when the model was considered the best. Before the training session, each dataset was randomly divided into a training set (60%), a validation set (20%) and a test set (20%). This operation was repeated five times. It was assumed that the images from a patient that belongs to the test set can not be present in a training and validation set. Therefore the augmentation procedure was applied after data split.

The other training parameters were: the *Adam* optimiser with an initial learning rate of 0.0001, and the *categorical_crossentropy* loss function. The accuracy of the CNN-based classification was compared to the accuracy of a human-based classification performed by our experts (two experts with over 3 years of work experience each, in a FH clinic). Additionally to analyse to role of the mask applied for a original image a subset of 50 randomly selected images with mask from the test set were again evaluated by experts and by the model. All experiments were performed on an NVIDIA DGX Station platform using Keras (v. 2.2) with a TensorFlow (v. 1.8.0) back-end. The software to support the experts during the evaluation process was written in QT C++ (random image selection, image presentation and GUI widgets to provide the expert's feedback).

3. Implementation and testing on data acquired with mobile phone

In many situations general practitioners, physicians and other care givers does not have access to professional equipment. With a dedicated application they can easily acquire image with a mobile phone and process an image over on-line classification service. This allows not only for current disease detection but in a future also the long term monitoring of changes and differential analysis. To check this, we designed a simple mobile application for taking images of the eye, extracting the iris and applying the mask. The pre-processed images were then transferred to the machine learning server for classification. Using this application, 24 volunteers (age 33 to 67 yo.) were tested (12 with a CA and 12 without). The images of the iris acquired in this part of the study were not used during the network training and the evaluating phases. The application was used for image acquisition in a noisy environment, but in the future it could be extended to detect and observe CA changes. The mock-up of the application GUI is presented in Fig. 5.

4. Results

The metrics that were used in this study for the analysis of the results are: Recall, Specificity, Precision, F1 score, and Accuracy.

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$Specificity = \frac{TN}{TN + FP} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$F1score = \frac{2TP}{2TP + FP + FN} \tag{5}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

where: TP - true positive, TN - true negative, FP - false positive, FN - false negative.

First, the selected models were applied on pre-processed images from original dataset (pre-augmentation procedure). Table 1 shows the results obtained for images with the white mask applied while Table 2 presents the results obtained for images with the black mask applied.

The analysis of the results obtained for the particular network trained and tested on images with the white mask applied are presented in Table 3 while Table 4 shows the results obtained for the networks trained and tested on images with the black mask.

In pursuit of optimal model performance the models were trained with different learning rate values (0.0001, 0.00001 and 0.000001). The values of accuracy function for the trained models are presented in Figs. 6–8. Although we trained each model five

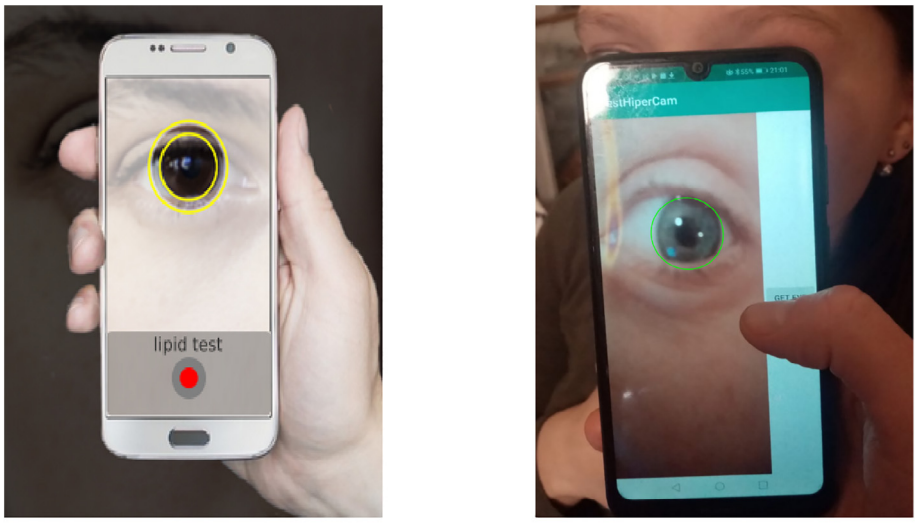

Fig. 5. Application demonstrating how to test for the presence of a CA (a) Mock-up (b) Demo application "TestHiperCam".

Table 1
Analysis of the results obtained across the utilized network architectures for the original dataset (no augmentation) with the white mask applied.

Model	lr.	Acc.	Recall	Spec.	Prec.	F1
ResNet	10^{-4}	0.88	0.94	0.80	0.86	0.90
Inception	10^{-4}	0.77	0.84	0.68	0.77	0.81
VGG	10^{-4}	0.72	0.77	0.64	0.77	0.77

Table 2
Analysis of the results obtained across the utilized network architectures for the original dataset (no augmentation) with the black mask applied.

Model	lr.	Acc.	Recall	Spec.	Prec.	F1
ResNet	10^{-4}	0.77	0.82	0.7	0.8	0.81
Inception	10^{-4}	0.77	0.82	0.70	0.80	0.81
VGG	10^{-4}	0.75	0.82	0.67	0.77	0.79

times, for further analysis, we selected the ones with the highest accuracy.

Further, we juxtaposed the best obtained accuracy of all five training sessions for all tested models. The values of Accuracy, Recall, Specificity, Precision and F1 score obtained for the different configurations of the models are presented in Tables 3 and 4. One of the challenges related to the classification approach based on a neural network is to avoid false classification results based on inappropriate image features. In this case, we tried to design a model that will classify images based on the presence of a specific medical condition (corneal arcus). To verify if the trained model is not sensitive only on image quality, we calculated the signal to noise ratio (SNR) for each image in the dataset. The results are presented in Fig. 9, 10 and Table 7. For comparison, we also calculated the average SNR and standard deviation (std. dev.).

Finally, the models that generated the best results were also tested on a set of iris images acquired from 24 volunteers and never used during the network training and evaluating phases.

Table 3
Analysis of the results obtained across the utilized network architectures for the augmented dataset with the white mask applied.

Model	lr.	Acc.	Recall	Spec.	Prec.	F1
ResNet	10^{-4}	0.92	0.97	0.85	0.9	0.93
ResNet	10^{-5}	0.93	1.0	0.85	0.89	0.94
ResNet	10^{-6}	0.92	0.98	0.85	0.89	0.93
Inception	10^{-4}	0.85	0.95	0.75	0.81	0.87
Inception	10^{-5}	0.83	0.94	0.73	0.78	0.85
Inception	10^{-6}	0.81	0.86	0.74	0.82	0.84
VGG	10^{-4}	0.81	0.88	0.73	0.8	0.84
VGG	10^{-5}	0.81	0.9	0.7	0.77	0.83
VGG	10^{-6}	0.81	0.87	0.74	0.83	0.85

Table 4
Analysis of the results obtained across the utilized network architectures for the augmented dataset with the black mask applied.

Model	lr.	Acc.	Recall	Spec.	Prec.	F1
ResNet	10^{-4}	0.9	0.97	0.81	0.86	0.91
ResNet	10^{-5}	0.9	1.0	0.8	0.85	0.91
ResNet	10^{-6}	0.89	0.97	0.8	0.85	0.91
Inception	10^{-4}	0.84	0.9	0.76	0.83	0.87
Inception	10^{-5}	0.84	0.95	0.73	0.79	0.86
Inception	10^{-6}	0.81	0.89	0.72	0.79	0.84
VGG	10^{-4}	0.84	0.87	0.78	0.86	0.87
VGG	10^{-5}	0.85	0.95	0.74	0.79	0.86
VGG	10^{-6}	0.8	0.89	0.71	0.78	0.83

Table 5
Analysis of the results obtained across the utilized network architectures for a dataset simulating a "real life" scenario with the white mask applied.

Model	lr.	Acc.	Recall	Spec.	Prec.	F1
ResNet	10^{-5}	0.88	1.0	0.8	0.75	0.86
Inception	10^{-4}	0.88	1.0	0.8	0.75	0.86
VGG	10^{-4}	0.83	0.79	0.9	0.92	0.85

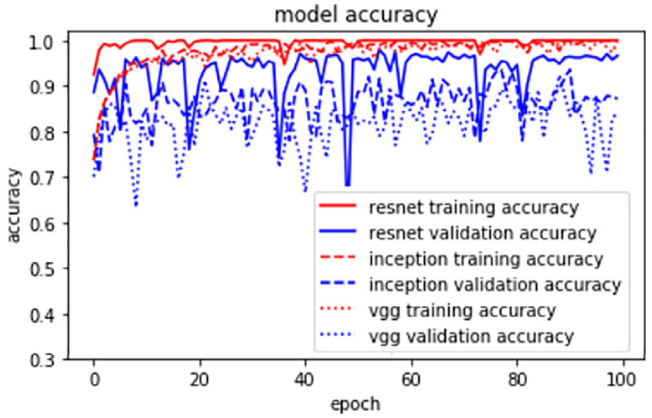

Fig. 6. Training and validation accuracy for compared models trained with learning rate lr = 0.0001.

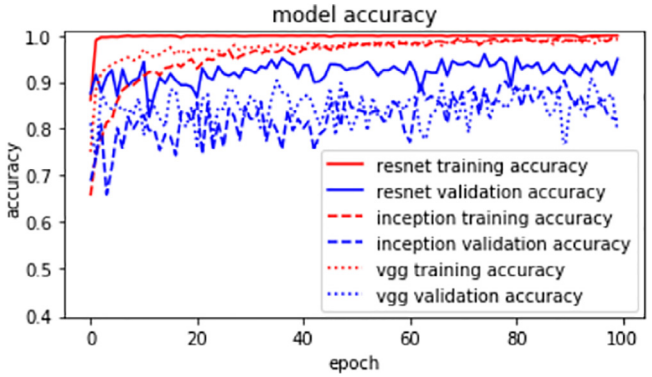

Fig. 7. Training and validation accuracy for compared models trained with learning rate lr = 0.00001.

These images were pre-processed and duplicated to create two datasets, one for the white mask and one for the black mask application. The analysis of the obtained results is presented in Fig. 5 for the images with the white mask and 6 for the images with the black mask.

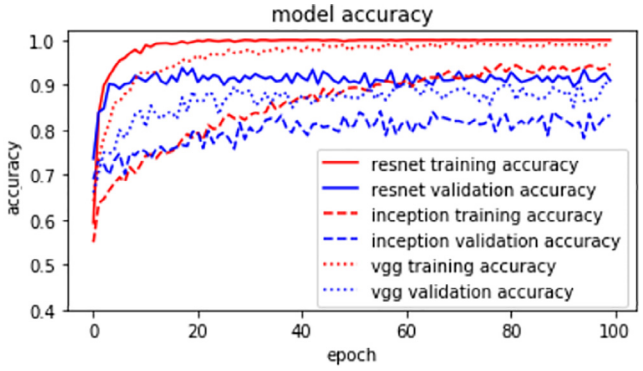

Fig. 8. Training and validation accuracy for compared models trained with learning rate lr = 0.000001.

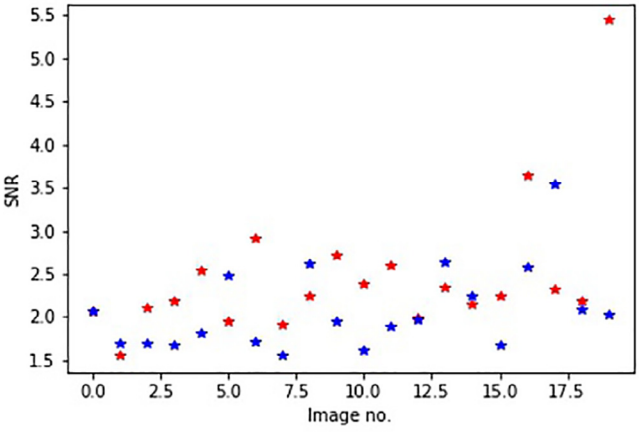

Fig. 9. Graphic presentation of SNR values calculated for sample images from a training set divided into two classes: "CA class" (blue) and "NO CA class" (red).

To identify if the CNN model is classifying based on the correct features (the presence of a corneal arcus), we checked Class Activation Maps for all of the images acquired using the mobile application. This technique allows the scores associated with an output class to be visualised - the features the trained model is focusing

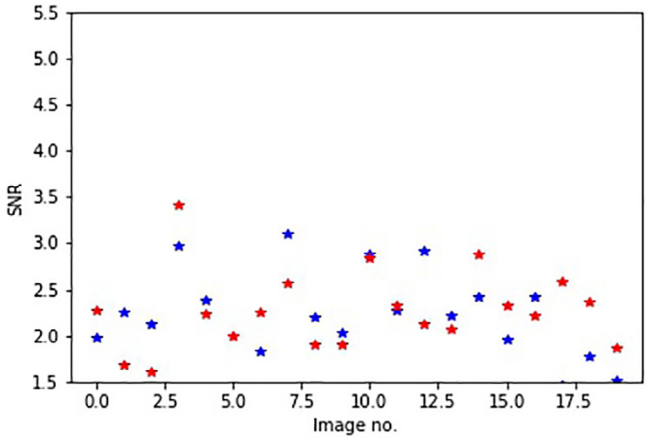

Fig. 10. Graphic presentation of SNR values calculated for sample images from a test set divided into two classes: "CA class" (blue) and "NO CA class" (red).

on (Kwaśniewska et al., 2017; Selvaraju et al., 2016). Fig. 11 shows what part of the images has activated the neurons of a particular network.

In most cases, the corneal arcus is easy to recognize by an expert. However, we wanted to compare the results of the classification performed by experts and by the trained CNN. Likewise, the images included in the dataset were also geometrically normalised. The image mask was implemented to focus on the iris region. Fig. 12 shows the comparison of the accuracy between the better expert and the chosen model.

5. Discussion

One of the reasons that familial hypercholesterolemia is underdiagnosed among the population is that it relies on a blood cholesterol level test, which in Europe is not standard among patients under 40 years old (Mach et al., 2020). The presence of a corneal arcus does not unequivocally indicate that the patient has familial hypercholesterolemia, but it is considered to be one of the indicators of this disease (Ogura et al., 2016; Pajak et al., 2016; Fernández et al., 2007). The hypothesis of conducted research was quite similar that the distinctive part of FH patients also have developed a corneal arcus. Earlier studies focused on the detection of a corneal arcus have been mainly based on hand-crafted features (e.g. colour features) (Kumar and Gunasundari, 2016; SV and Gunasundari, 2018). There were also examples to implement such algorithms in the form of a mobile application. Some of these papers clearly indicated that the future of corneal arcus detection lies with machine learning (Alhasawi et al., 2018). In this paper, we applied VGG16, ResNet and Incetion v3-based models to the problem of corneal arcus classification. These architectures proved to be successful on image classification tasks. Their models, pre-trained on the ImageNet dataset, can be easily used for fine-tuning and transfer learning. It was very helpful at the early stage of this project when we just started to collect the samples to our dataset. It must be underlined that the process of building such a dataset is very time consuming. Despite that fact we were able to fine-tuned these networks and start investigate which architecture (sequential or the one with "building blocks") can produce better results. Moreover, these architectures are suitable for conversion to the models that can be deployed in mobile application. To train the classifier, we built a dataset that is, before data augmentation, more than two times larger then the dataset utilised by (Kumar and Gunasundari, 2016; SV and Gunasundari, 2018; Nasution and Kusuma, 2009). We also developed a deep learning algorithm for CA detection as well as a basic mobile application. The small number of patients diagnosed with hypercholesterolemia limits the possibility of enrolling these patients into a study on corneal arcus detection. The dataset was augmented by rotating all of the images, which simulated different special locations of corneal arcus. Moreover, we created our dataset based on images acquired in a similar way to those taken with a hand-held camera or smartphone (close proximity to the eye).Therefore the processed images are not a high resolution images that can be acquired from high class medical devices. The intention is, that these images are similar to the images every general practitioner (care giver or patient) can take with his personal device. The models presented in this paper have proven to be very successful in medical applications solving classification problems. However, these models are usually applied on a dataset acquired from professional medical equipment whose solely purpose is to capture precise medical images such as retinal images (Yu et al., 2017). In this paper, we utilized popular models on images captured by regular phone cameras in very unpredictable conditions. Photographic images taken by a smartphone or hand-held camera are versatile in terms of resolution, zoom and lightning. All of these factors can influence the classification process, making it quite challenging. Even though there are databases containing iris images (Proença et al., 2009), it is difficult to find a dataset that corresponds to our dataset. In this study, we focused on the practical application and the possibility of using neural networks in a real-life application for CA screening. Reducing distortion in the images included in our dataset was desired. Therefore we decided to apply masks that cover the irrelevant parts of the image. We evaluated the performance of the utilized networks on images covered with white and black masks. The results are presented in Figs. 6–8 as well as in Tables 3 and 4. We tried to minimize validation loss by changing the initial learning rate. After all of this, we came to the conclusion that in this particular case, the increase of the learning rate does not produce better results in terms of validation loss. We decided to assess the networks by obtained accuracy and F1 score and therefore in the final test performed on the images acquired from 24 volunteers, a limited number of evaluated network models and configurations was used. The results from Tables 3–5 suggest that using the ResNet-based models produces the best results. We decided to evaluate

Table 7
Average SNR and std. dev. values calculated for images for the training and test datasets.

	Training set		Test set	
	CA	no CA	CA	no CA
Average SNR	2.37	2.19	2.43	2.20
Std Dev.	0.68	0.50	0.57	0.47

Table 6
Analysis of the results obtained across the utilized network architectures for a dataset simulating a "real life" scenario with the black mask applied.

Model	lr.	Acc.	Recall	Spec.	Prec.	F1
ResNet	10^{-4}	0.83	0.79	0.9	0.92	0.85
Inception	10^{-4}	0.75	0.75	0.75	0.75	0.75
VGG	10^{-4}	0.79	0.82	0.77	0.75	0.78

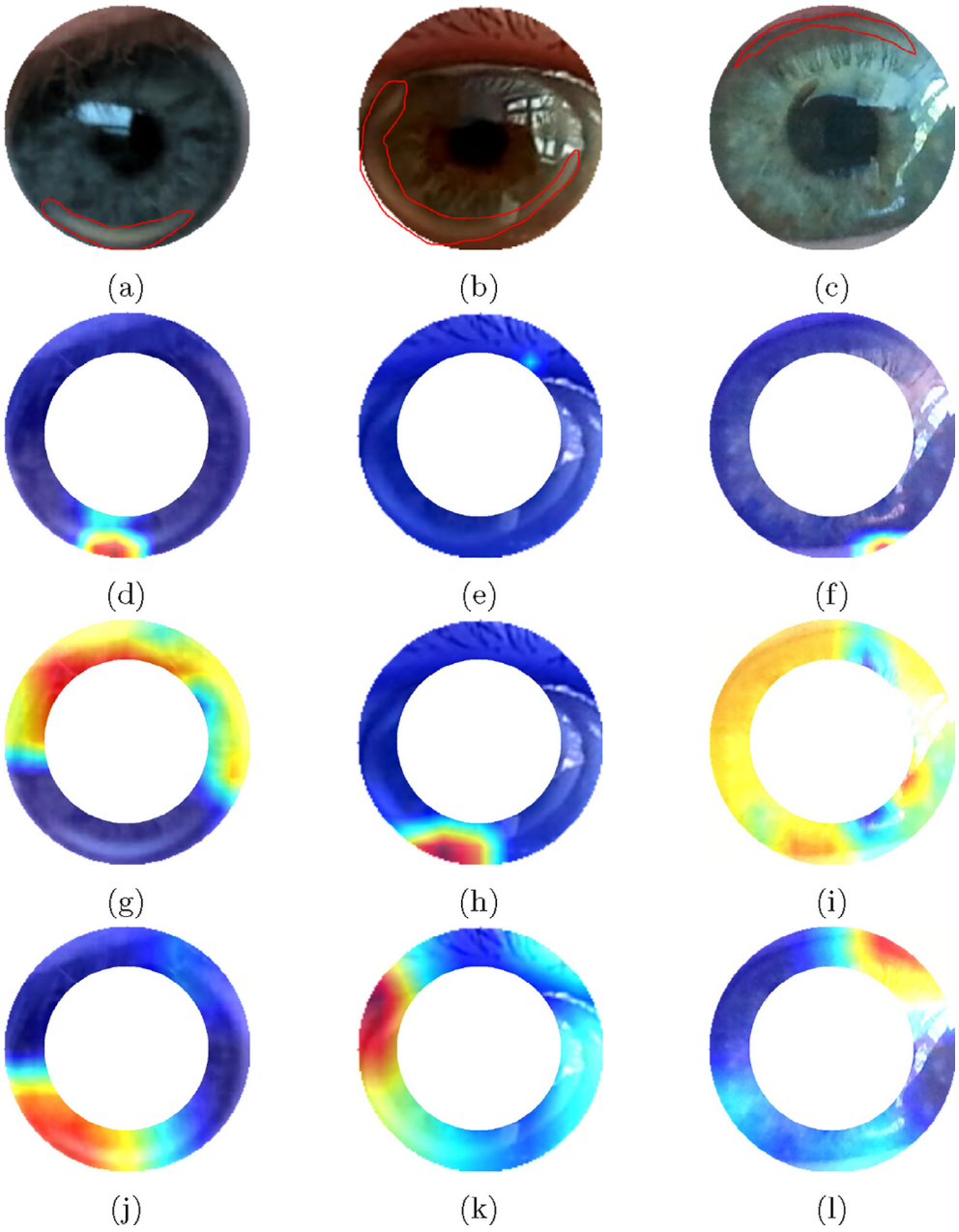

Fig. 11. Image of an eye with a corneal arcus detected. The ground truth images, with the corneal arcus marked by an expert are presented in (a,b,c) while the corresponding Class Activation Maps for the trained models are presented in (d,e,f) for VGG (white mask); (g,h,i) ResNet (white mask); and (j,k,l) Inception (white mask).

the networks using the Class Activation Matrix approach. The results presented in Fig. 11 show that the results produced by the Inception-based model are most consistent with the ground truth images, but only when the white mask is applied. On the other hand, the ResNet-based model although it showed high accuracy and F1 score, seems to make its decision based on the wrong features. However when it comes to images covered with the black mask, the ResNet architecture produces better results. In general, the best results can be obtained when the white mask is applied, as the results obtained on images acquired with the mobile application are slightly better for images covered with the white mask.

Due to the specific character of the problem and the dataset it is difficult to compare our findings to other similar research. In they work (Kumar and Gunasundari, 2016; SV and Gunasundari, 2018 and Nasution and Kusuma, 2009) reports nearly 0.96 and 0.93 accuracy of corneal arcus detection respectively. It has to be underline though that theses research focused on detecting arcus senilis and even cataract (features that are very distinctive eye condition easily noticed even by untrained personnel). In case of our studies we focused on detecting arcus juvenilis that is less distinctive and therefore more difficult to detect. We have made every effort to build the dataset used in this study based mostly on the images with arcus juvenilis. The feature based approach may provide better results for small datasets but its performance can decrease when introduced to new slightly different data (like early stage of corneal arcus). The advantage of using neural network over feature based approach is that the network can be re-trained to new data even if they will slightly differ from initial core of the data (different stages of corneal arcus). The performance of the designed network was compared to the performance of experts. Although the data presented in Fig. 12 suggest that the CNN model classifies with a higher level of accuracy than experts (0.98 compared to

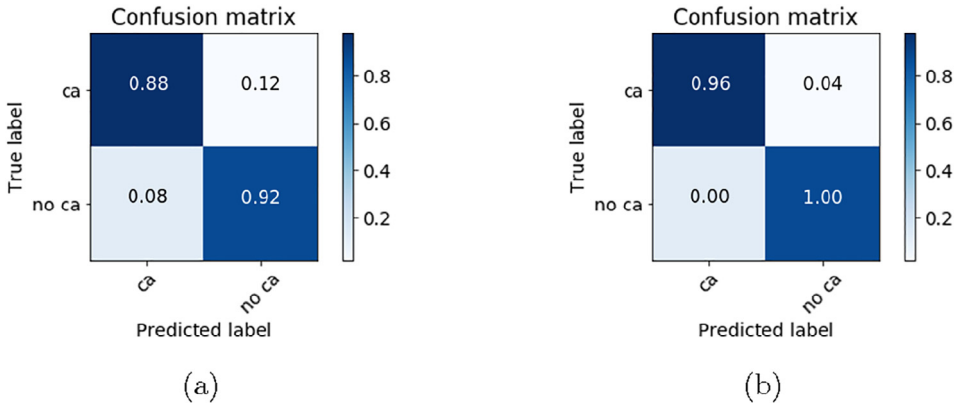

Fig. 12. Confusion matrix for a classification performed by, (a) human expert (b) chosen model.

0.90), it should be underlined that the experts were asked to make a decision based on the pre-processed data (the same images that the CNN model was trained on). The accuracy of corneal arcus detection was higher, nearly 100% accurate, when the specialist was presented with the whole image of the eye (not within the mask). A future iteration of the model might benefit from processing the entire eye (including the iris and sclera).

6. Conclusions

In our work, we showed another application where a CNN-based model can be successfully utilised to accurately detect the presence of an illness. Although using a CNN for object (or feature) detection is not original, the application of a CNN for corneal arcus classification in images has not been previously reported. The practical aim of this project was to design a mobile application that will help detect the presence of a CA. Like it was mentioned before, in many situations physicians and other care givers does not have access to professional equipment. Dedicated application will provide service for instant assessment of a CA presence but more importantly it will allow the long term monitoring of its changes and possibility of differential analysis. Designing the application that can help in distance diagnosis of potential disease has a grate importance in all situations were access to health care and specialists is limited. Moreover, we have evaluated the most popular networks in terms of the features that activated the neurons in each tested model. It is worth mentioning that selected machine learning models can be deployed on mobile device or as a web service. We also showed that, generally it is better to use a white mask when covering irrelevant part of the image, at least for corneal arcus classification. Although the dataset was larger than in previous studies, it still needs to be enlarged for future deep learning applications. However, the obtained high accuracy values and high correlation with the decisions of experts are satisfying, and constitute a good foundation for further studies. Currently, we are working on enlarging the dataset with new CA cases. The enlarged dataset will allow further studies focused on corneal arcus localisation, and classification of the different CA stages. Creating an algorithm that classifies the medical condition with high accuracy provides great potential for screening toward a disease such as familial hypercholesterolemia, and to hopefully reduce the number of undiagnosed patients.

Declaration of Competing Interest

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

Acknowledgment

This work was partially supported by Statutory Funds of Electronics, Telecommunications and Informatics Faculty, Gdansk University of Technology.

References

Alhasawi, Y., Mullachery, B., Chatterjee, S., 2018. Design of a mobile-app for non-invasively detecting high blood cholesterol using eye images. In: Proceedings of the 51st Hawaii International Conference on System Sciences.

Bhangdiya, V., 2014. Cholesterol presence detection using iris recognition. Int. J. Technol. Sci 1 (1), 22–25.

Brewer, H., Santamarina-Fojo, S.M., Shamburek, R., 2005. Genetic dyslipoproteinemias. In: Philadelphia, P.A. (Ed.), Atherothrombosis and coronary artery disease. Lippincott Williams & Wilkins, pp. 55–83.

Castaneda, C., Nalley, K., Mannion, C., Bhattacharyya, P., Blake, P., Pecora, A., Goy, A., Suh, K.S., 2015. Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine. J. Clinical Bioinform. 5 (1), 1–16.

Chlebus, K., Cybulska, B., Gruchała, M., Smaga, A., Wróbel, K., Wojtyniak, B., Pajkowski, M., Jankowski, P., Zdrojewski, T., 2018. Prevalence, diagnosis, and treatment of familial hypercholestero laemia in outpatient practice in poland. Kardiologia Polska=. Polish Heart J. 76 (6).

Ciompi, F., de Hoop, B., van Riel, S.J., Chung, K., Scholten, E.T., Oudkerk, M., de Jong, P. A., Prokop, M., van Ginneken, B., 2015. Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2d views and a convolutional neural network out-of-the-box. Medical Image Anal. 26 (1), 195–202.

Desnick, R., Ioannou, Y., Eng, C., Scriver, C., Beaudet, A., Sly, W., 2001. The metabolic and molecular bases of inherited disease. In: Beaude, A.L., Sly, W.S., Valle, D., Scriver, C.R. (Eds.). McGraw-Hill, New York, pp. 3733–3774.

Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S., 2017. Dermatologist-level classification of skin cancer with deep neural networks. Nature 542 (7639), 115–118.

Fernández, A., Sorokin, A., Thompson, P.D., 2007. Corneal arcus as coronary artery disease risk factor. Atherosclerosis 193 (2), 235–240.

Hypercholesterolemia, F., 1999. report of a second who consultation. World Health Organization, Geneva.

Khamis, H., Chen, J., Redmond, J.S., Lovell, N.H., 2018. Detection of atrial fibrillation from rr intervals and pqrst morphology using a neural network ensemble. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp. 5998–6001.

Kocejko, T., Bujnowski, A., Wtorek, J., 2009. Complex human computer interface for las patient. In: 2009 2nd Conference on Human System Interactions. IEEE, pp. 272–275.

Kocejko, T., Wtorek, J., 2013. Gaze tracking in multi-display environment. In: 2013 6th International Conference on Human System Interactions (HSI). IEEE, pp. 626–631.

Kumar, S.M., Gunasundari, R., 2016. Diagnosis of corneal arcus using statistical feature extraction and support vector machine. In: Artificial Intelligence and Evolutionary Computations in Engineering Systems. Springer, pp. 481–492.

Kwaśniewska, A., Rumiński, J., Rad, P., 2017. Deep features class activation map for thermal face detection and tracking. In: 2017 10Th international conference on human system interactions (HSI). IEEE, pp. 41–47.

Laksanasopin, T., Guo, T.W., Nayak, S., Sridhara, A.A., Xie, S., Olowookere, O.O., Cadinu, P., Meng, F., Chee, N.H., Kim, J., et al. (2015). A smartphone dongle for diagnosis of infectious diseases at the point of care. Science translational medicine, 7(273):273re1–273re1.

Li, Q., Cai, W., Wang, X., Zhou, Y., Feng, D.D., Chen, M., 2014. Medical image classification with convolutional neural network. In: 2014 13th international

conference on control automation robotics & vision (ICARCV). IEEE, pp. 844–848.
Li, W., Liu, X., Liu, J., Chen, P., Wan, S., Cui, X., 2019. On improving the accuracy with auto-encoder on conjunctivitis. Appl. Soft Computing 81, 105489.
Mach, F., Baigent, C., Catapano, A.L., Koskinas, K.C., Casula, M., Badimon, L., Chapman, M.J., De Backer, G.G., Delgado, V., Ference, B.A., et al., 2020. 2019 esc/eas guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk: the task force for the management of dyslipidaemias of the european society of cardiology (esc) and european atherosclerosis society (eas). Eur. Heart J. 41 (1), 111–188.
Nasution, A., Kusuma, W., 2009. Recognition of incremental changes in corneal ring (arcus-senilis) using the hybrid n-feature neural network (hnfnn). In: Proc. 4th Asian and Pacific-Rim Symposium on Biophotonic (APBP). Citeseer.
Ogura, M., Hori, M., Harada-Shiba, M., 2016. Association between cholesterol efflux capacity and atherosclerotic cardiovascular disease in patients with familial hypercholesterolemia. Arteriosclerosis, Thrombosis, and Vascular Biology 36 (1), 181–188.
Pajak, A., Szafraniec, K., Polak, M., Drygas, W., Piotrowski, W., Zdrojewski, T., Jankowski, P., 2016. Prevalence of familial hypercholesterolemia: a meta-analysis of six large, observational, population-based studies in poland. Arch. Med. Sci.: AMS 12 (4), 687.
Patel, V.L., Shortliffe, E.H., Stefanelli, M., Szolovits, P., Berthold, M.R., Bellazzi, R., Abu-Hanna, A., 2009. The coming of age of artificial intelligence in medicine. Artif. Intell. Med. 46 (1), 5–17.
Proença, H., Filipe, S., Santos, R., Oliveira, J., Alexandre, L.A., 2009. The ubiris. v2: A database of visible wavelength iris images captured on-the-move and at-a-distance. IEEE Trans. Pattern Anal. Mach. Intell. 32 (8), 1529–1535.
Raal, F.J., Santos, R.D., 2012. Homozygous familial hypercholesterolemia: current perspectives on diagnosis and treatment. Atherosclerosis 223 (2), 262–268.
Ramlee, R., Aziz, K., Ranjit, S., Esro, M., 2011. Automated detecting arcus senilis, symptom for cholesterol presence using iris recognition algorithm. J. Telecommun., Electr. Computer Eng. (JTEC) 3 (2), 29–39.
Rouhi, R., Jafari, M., Kasaei, S., Keshavarzian, P., 2015. Benign and malignant breast tumors classification based on region growing and cnn segmentation. Expert Syst. Appl. 42 (3), 990–1002.
Selvaraju, R.R., Das, A., Vedantam, R., Cogswell, M., Parikh, D., and Batra, D. (2016). Grad-cam: Why did you say that? arXiv preprint arXiv:1611.07450.
Simon, M., Rodner, E., and Denzler, J. (2016). Imagenet pre-trained models with batch normalization. arXiv preprint arXiv:1612.01452.
Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
Songire, S.G., Joshi, M.S., 2016. Automated detection of cholesterol presence using iris recognition algorithm. Int. J. Computer Appl. 133 (6), 41–45.
SV, M.K. and Gunasundari, R. (2018). Computer-aided diagnosis of anterior segment eye abnormalities using visible wavelength image analysis based machine learning. Journal of medical systems, 42(7):1–12.
Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818–2826.
Titano, J.J., Badgeley, M., Schefflein, J., Pain, M., Su, A., Cai, M., Swinburne, N., Zech, J., Kim, J., Bederson, J., et al., 2018. Automated deep-neural-network surveillance of cranial images for acute neurologic events. Nature Med. 24 (9), 1337–1341.
Wang, L., Zhen, H., Fang, X., Wan, S., Ding, W., Guo, Y., 2019. A unified two-parallel-branch deep neural network for joint gland contour and segmentation learning. Future Generation Computer Systems 100, 316–324.
Winder, A.F., Jolleys, J.C., Day, L.B., Butowski, P.F., 1998. Corneal arcus, case finding and definition of individual clinical risk in heterozygous familial hypercholesterolaemia. Clinical Genetics 54 (6), 497–502.
Yu, F., Sun, J., Li, A., Cheng, J., Wan, C., Liu, J., 2017. Image quality classification for dr screening using deep learning. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp. 664–667.
Zech, L.A., Hoeg, J.M., 2008. Correlating corneal arcus with atherosclerosis in familial hypercholesterolemia. Lipids Health Disease 7 (1), 1–9.
Zhao, Y., Li, H., Wan, S., Sekuboyina, A., Hu, X., Tetteh, G., Piraud, M., Menze, B., 2019. Knowledge-aided convolutional neural network for small organ segmentation. IEEE J. Biomed. Health Inform. 23 (4), 1363–1373.