

Keratoconus diagnosis based on dynamic corneal imaging using 3D convolutional neural networks

Jakub Gęca^{1*}, Dariusz Głuchowski², Arkadiusz Podkowiński³,
Tomasz Chorągiewicz⁴, Dominika Wróbel-Dudzińska⁵, Robert Karpiński^{6,7},
Arkadiusz Syta⁸, Katarzyna E Jonak⁹, Agnieszka Wolińska⁷, Robert Rejdak⁴

¹ Faculty of Electrical Engineering and Computer Science, Department of Electrical Drives and Machines, Lublin University of Technology, Nadbystrzycka 38A, 20-618, Lublin, Poland

² Faculty of Mathematics and Information Technology, Department of Applied Informatics, Lublin University of Technology, Nadbystrzycka 38, 20-618 Lublin, Poland

³ Da Vinci NeuroClinic, Tomasza Zana 11A, Lublin, Poland

⁴ Chair and Department of General and Pediatric Ophthalmology, Medical University of Lublin, Chmielna 1, 20-079 Lublin, Poland

⁵ Department of Diagnostics and Microsurgery of Glaucoma, Medical University of Lublin, 20-079 Lublin, Poland

⁶ Faculty of Mechanical Engineering, Department of Machine Design and Mechatronics, Lublin University of Technology, Nadbystrzycka 36, 20-618 Lublin, Poland

⁷ Institute of Medical Sciences, The John Paul II Catholic University of Lublin, Konstantynów 1F, 20-708 Lublin, Poland

⁸ Faculty of Mathematics and Information Technology, Department of Technical Computer Science, Lublin University of Technology, Nadbystrzycka 38, 20-618 Lublin, Poland

⁹ Doctoral School, The John Paul II Catholic University of Lublin, Al. Raławickie 14, 20-950 Lublin, Poland

* Corresponding author's e-mail: j.geca@pollub.pl

ABSTRACT

Keratoconus is a progressive disease that requires precise and rapid diagnosis, as well as the initiation of treatment, to prevent serious and permanent visual impairment. This article presents a comparison of 3D convolutional neural network models for the diagnosis of keratoconus based on dynamic corneal imaging results obtained with the CORVIS device. The article describes the data preprocessing and compares models of varying complexity in terms of accuracy, inference time, number of parameters, and GPU memory usage. To ensure adequate generalization capability during algorithm training, 5-fold stratified cross-validation was used, and the average metrics from all splits were compared. The best model achieved an average keratoconus detection accuracy exceeding 88%, confirming that deep neural networks can be a promising tool to support physicians in diagnosing corneal diseases such as keratoconus. Future work includes plans to gather a larger patient database and apply more advanced preprocessing methods for the video data.

Keywords: keratoconus, cornea, deep learning, convolutional neural network, medical imaging.

INTRODUCTION

The cornea is a key component of the eye's refractive system, responsible for 70% of the total refractive power. Visual quality depends on its shape and transparency, and therefore on the stability and structural integrity of the cornea, which are

determined by its biomechanical properties [1, 2]. These properties dictate how the cornea responds to stress and deformation, and this process is influenced by its viscoelastic characteristics. The cornea can be regarded as a complex biomechanical composite. Its stroma, which accounts for more than 90% of its thickness, consists of 300 to 500

lamellar layers composed of unbranched collagen fibers extending across the entire cornea from limbus to limbus. Each lamella has a highly regular organization: the fibers are parallel and evenly spaced [3, 4]. The biomechanical properties of the cornea are important in the measurement of intraocular pressure (IOP) and are relevant in clinical conditions such as keratoconus, corneal refractive surgery, and glaucoma [5].

Keratoconus is a progressive, non-inflammatory disorder characterized by corneal thinning, conical protrusion, and biomechanical alterations [6]. This results in irregular astigmatism, visual deterioration, and reduced visual acuity. The etiology of the disease is not fully understood, but it is suggested that genetic, biomechanical, and environmental factors may contribute to its development [7]. Early diagnosis is crucial for halting disease progression and initiating treatment [8]. Traditional diagnostic methods focus on analyzing the corneal shape and thickness using topography and tomography (Scheimpflug, OCT) [9]. Although these techniques are effective in detecting advanced keratoconus, their accuracy is limited in identifying early-stage changes [8, 10, 11]. For this reason, increasing attention is being given to methods that assess the biomechanical properties of the cornea.

A modern diagnostic tool is the Ocular Response Analyzer (ORA), which measures the cornea's response to dynamic deformation through the corneal hysteresis (CH) and corneal resistance factor (CRF) [12,13]. These values are reduced in patients with keratoconus, indicating structural weakening of the tissue [14].

Another commonly used method is the Corvis ST (Corneal Visualization Scheimpflug Technology), which employs an air puff and Scheimpflug imaging to record dynamic corneal deformation [15, 16]. The device generates a range of mechanical indices, such as stiffness and deformation parameters, enabling earlier detection of degenerative changes. Corvis ST (Oculus Optikgeräte GmbH, Wetzlar, Germany) is a non-contact, semi-automated device that analyzes corneal biomechanics in vivo. It uses an air pulse to induce corneal deformation and captures its progression in real time with a high-speed Scheimpflug camera, acquiring 4300 horizontal frames per second [17–19]. Corneal deformation can be observed in slow motion on the device's screen.

Scheimpflug imaging enables the assessment of central corneal thickness (CCT). During the first

of two corneal applanations, intraocular pressure (IOP) is calculated [20]. The most important biomechanical parameters include, among others: applanation times and lengths (A-time1/A-time2, A-length1/A-length2), corneal velocities (Vin/Vout), time and radius of highest concavity, deformation amplitude (DA), peak distance (PD), and the radius of curvature at the point of highest concavity (R).

The use of Scheimpflug imaging combined with artificial intelligence methods and multimodal data (both imaging and numerical) enables effective classification of keratoconus cases. The extraction of corneal geometric features based on deformation curve profiles allowed for 99% diagnostic accuracy using a neural network [21]. Expanding the feature set with topographic data from the Pentacam device enabled a Random Forest model to achieve 89% classification accuracy [22]. Similar combinations of features yielded an AUC of 93% [23]. Analysis of data from three devices (Sirius, Pentacam, Corvis) enabled identification of the most accurate indicators (Sirius – AUC = 91%, Corvis – AUC = 82%) [24]. Transfer learning has been applied to classify images from the Corvis device - comparison of the ResNet152, VGG16, and InceptionV3 models showed the highest diagnostic accuracy for ResNet152 (AUC = 99%) [25]. The latest review [26] of literature on AI applications in corneal disease diagnostics highlights their high effectiveness, as well as the need for external validation and greater population heterogeneity.

The implementation potential of artificial intelligence models depends not only on their diagnostic performance for a given task but also on their hardware requirements and their ability to meet real-time constraints, if such are critical in a specific context. The aim of this study was to evaluate the capabilities of three-dimensional convolutional neural network (3D-CNN) architectures for the diagnosis of keratoconus, as well as to assess their inference time and memory consumption. This approach makes it possible to evaluate algorithmic performance while simultaneously addressing their feasibility for practical deployment in a clinical setting, which is an aspect often overlooked in medical data analysis. The conducted literature review indicates that previous studies have focused primarily on the diagnostic accuracy of artificial intelligence algorithms for keratoconus based on biomechanical parameters derived from dynamic corneal imaging. In contrast, our approach relies on the

analysis of complete video sequences captured by the high-speed camera integrated in the Corvis ST system. This eliminates the need for explicit computation of biomechanical parameters and requires only the direct use of the recorded video. Furthermore, our method makes it possible to define implementation-related requirements that are critical for the translation of developed models into medical practice.

The novelty of this study lies in the comparative analysis of the performance and hardware requirements of 3D convolutional neural networks for keratoconus detection based solely on dynamic corneal imaging using video sequences acquired with the CORVIS ST device. Our findings offers a novel perspective on adapting the 3D CNN architecture to the specific requirements of keratoconus diagnosis and available computational resources, while achieving high classification accuracy. The integration of the proposed models into clinical workflows has the potential to serve as a valuable decision-support tool for ophthalmologists in the diagnosis of keratoconus. This contribution, to the best of our knowledge, has not yet been reported in the existing literature.

STUDY PARTICIPANTS

The study included two groups of patients presenting at the outpatient clinic: a pilot group diagnosed with keratoconus (57 eyes) and a control group (47 eyes). All examinations were conducted at the Department of Ophthalmology, Medical University of Lublin, between March and August 2024. Each patient provided written informed consent to participate in the study. The study was conducted in accordance with Good Clinical Practice (GCP) guidelines and the Declaration of Helsinki, and its implementation was approved by the Local Ethics Committee of the Medical University of Lublin (approval number: KE-0254/98/03/2023). The preliminary examinations included: objective and subjective refraction, slit-lamp examination, intraocular pressure measurement, and corneal tomography. Patients were enrolled in the study based on the following inclusion criteria:

- Inclusion criteria for healthy eyes:
 - normal corneal topography and tomography results ($K_{\max} < 47$ D, inferior–superior difference < 1.5 D, skewed radial axis index $< 22^\circ$)

- normal elevation maps of the anterior and posterior corneal surfaces
- uniform corneal thickness distribution ($CCT > 480 \mu\text{m}$)
- no corneal scarring
- no clinical signs of keratoconus
- no family history of keratoconus
- Inclusion criteria for keratoconic eyes:
 - abnormal corneal tomography and topography results ($K_{\max} > 47$ D, inferior–superior difference > 1.5 D, skewed radial axis index $> 22^\circ$),
 - abnormal elevation maps of the anterior and posterior corneal surfaces,
 - central or inferior corneal protrusion,
 - thin cornea,
 - no corneal scarring;
- Exclusion criteria included: other forms of corneal ectasia (e.g., pellucid marginal degeneration, keratoglobus), endothelial diseases, prior ocular surgeries, and eye infections.
- Rationale for inclusion and exclusion criteria.

The inclusion criteria for healthy eyes were based on tomographic and topographic thresholds ($K_{\max} < 47$ D, inferior–superior difference < 1.5 D, skewed radial axis index $< 22^\circ$, central corneal thickness $> 480 \mu\text{m}$) that are well established in the literature to rule out sub-clinical keratoconus and ensure a truly normal corneal profile. Additional requirements, such as normal elevation maps and absence of clinical signs of keratoconus or family history, were applied to minimize the risk of enrolling eyes with forme fruste disease. The inclusion criteria for keratoconic eyes were based on characteristic tomographic features ($K_{\max} > 47$ D, inferior–superior difference > 1.5 D, skewed radial axis index $> 22^\circ$) and morphological changes, such as central or inferior protrusion and corneal thinning, which are diagnostic hallmarks of keratoconus. The absence of corneal scarring was required to avoid irregularities that could interfere with optical quality and biomechanical measurements. Exclusion criteria (other corneal ectasias such as pellucid marginal degeneration or keratoglobus, endothelial diseases, prior ocular surgeries, or active eye infections) were introduced to eliminate potential confounders that could independently affect corneal biomechanics, thereby ensuring the validity of the comparative analysis between groups.

All available biomechanical parameters obtained using the Corvis ST device were recorded for analysis. Table 1 contains the most important information about the device used. Measurements were performed by a qualified operator in accordance with the manufacturer's instructions.

DATA PREPROCESSING

This article presents a study on keratoconus detection based on videos from the CORVIS ST device using 3D convolutional neural network models of varying complexity. The input data consisted of 139 videos, of which 87 represented keratoconus eyes and 52 were healthy eyes. Each video comprised 139 frames, each with a resolution of 200×576 pixels. The duration of the videos as well as the frame dimensions were consistent across all recordings.

Proper data preparation is a crucial aspect of medical imaging and the use of artificial intelligence methods for diagnosing various diseases. Medical images may contain different types of noise or artifacts that degrade the quality of algorithm training and, in some cases, even prevent pattern recognition within the data. The absence of preprocessing in medical imaging can hinder the generation of consistent and interpretable results [27, 28]. Digital image processing plays a vital role in medical technologies such as radiography, ultrasonography, computed tomography, and computer-based analysis and interpretation of microscopic cellular images. As noted by the authors in [29], the goal of image processing is to

enhance relevant image features to facilitate interpretation. Furthermore, image processing can positively influence training speed and reduce the hardware requirements needed for analysis. This approach can enable the deployment of diagnostic models directly on measuring devices and edge devices [30].

Each video obtained from the CORVIS ST imaging device contains essential information about the examination, such as patient data, date of acquisition, and the examined eye. Additionally, each video frame includes a watermark with the logo of Oculus – the device manufacturer. Therefore, before proceeding with the study, it was necessary to remove this text to prevent patient-identifiable information from influencing the diagnostic features. Another important issue was the removal of the manufacturer's logo, which, due to its sharp edges and high brightness, could reduce the sensitivity of vision models to the edges of the cornea and their deformation over time.

The video processing was performed using the OpenCV library, because it is one of the most commonly used media processing library, supports wide range of video formats and is optimized for speed and performance, resulting in the initial processing of all videos taking approximately 84 seconds. Initially, edge detection was carried out using the Canny algorithm [31], followed by dilation to enhance the fine edges of the text overlays. In the next step, the `cv2.findContours` function was used to detect the contours of the text. For each contour, a bounding rectangle was calculated and a mask was created based on matching the rectangle size to predefined minimum and maximum character dimensions and aspect ratio. Using this mask, the text was removed through inpainting, i.e., filling in the masked regions based on an estimation of the background appearance. The results of this procedure are presented in Figure 1 for several frames from a demo video.

In the next stage of preprocessing, the video frames were resized to 100×100 pixels. This step was necessary to adjust the image size to the memory capacity of the graphics card used for computations. Processing full-size frames from videos consisting of 139 images would not have been feasible on an RTX 4070 Super GPU, as confirmed by preliminary tests, during which out of memory (OOM) errors occurred.

While the initial preparation of individual frames aimed to facilitate the algorithm's identification of spatial features - such as texture, contours,

Table 1. Corvis ST specification and key features

General information	
Manufacturer	Oculus
Model	Corvis ST
Tonometer	
Measurement range	6–60 mmHg
Measurement distance	11 mm
Inner fixation light	Red LED
Scheimpflug camera	
Frame rate	4330 images per sec
Measurement range	8.5 mm horizontal coverage
Pachymeter measurement range	200–1200 μm
Measuring points	576 per image (80640 per examination)
Source of light	Blue LED (470 nm UV free)

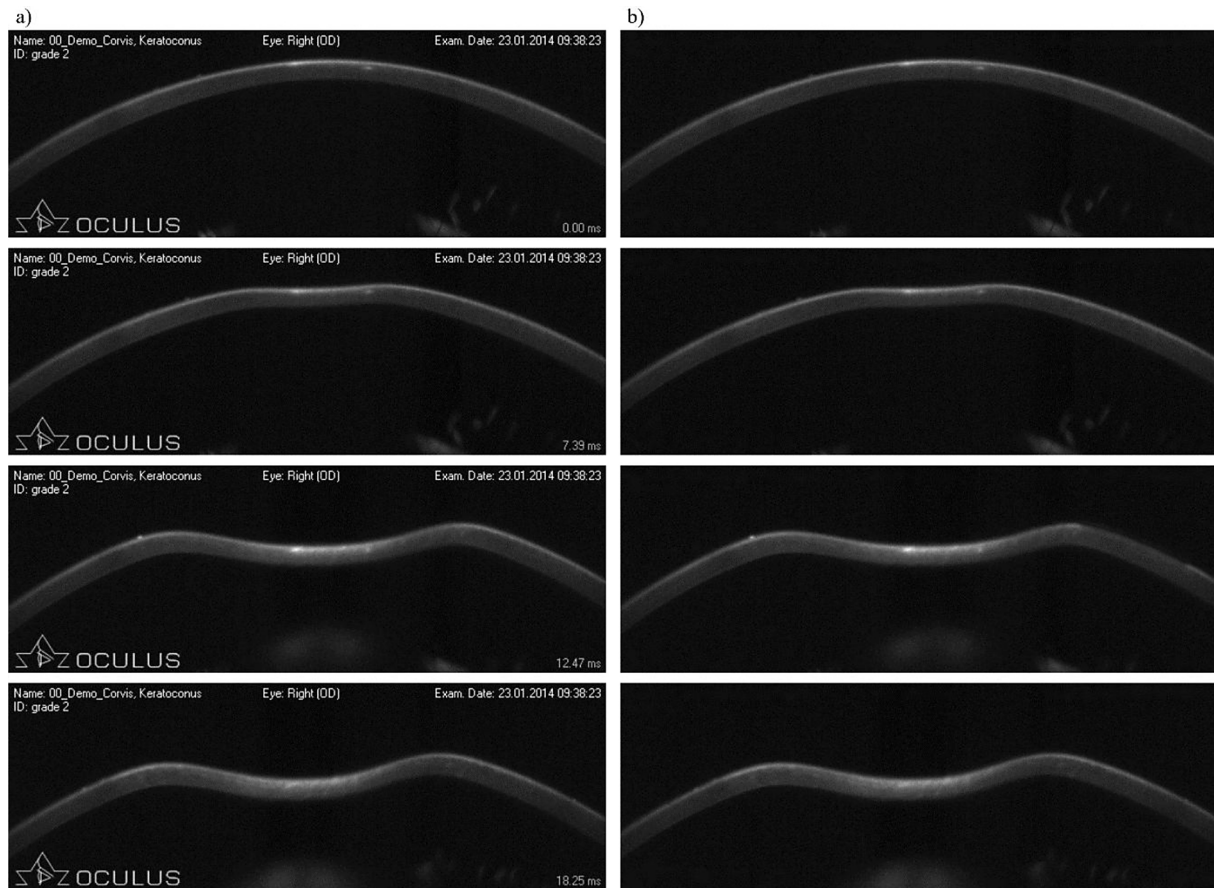


Figure 1. Examples of video frames before (a) and after (b) preprocessing, based on demo videos from the CORVIS ST standard database (excerpts from demo videos were shown to avoid revealing sensitive patient data)

shape, and the arrangement of elements – for the classification of entire videos, temporal features also play a significant role. These include object movements, changes in brightness and structure, as well as inter-frame dependencies. Analysis of the video material, composed of 139 frames, indicates that only a subset contains relevant information about corneal dynamics (i.e. its motion). The initial part of the video captures a static state – the patient is waiting for the test to begin, and the cornea remains motionless, aside from slight movements related, for example, to breathing. In the subsequent part of the video, the cornea's response to the air impulse is observed – from the initial state, through the point of maximum deformation, to the return to baseline position. In addition to the video, the CORVIS device also records three key corneal frames corresponding to the following states: initial, applanation, and maximum concavity, which are presented in the device report (Figure 2). Considering that these three states and the transitions between them are most crucial in the diagnostic process, a decision was made to

trim the videos to include only the time range in which they occur. Each video was truncated to 55 frames, covering the corneal deformation from the initial state to the point of maximum deflection and containing the most relevant information. An additional advantage of this approach was the reduction of spatially near-identical frames, which decreases the risk of algorithmic bias toward repeated features. Furthermore, this step helped to reduce GPU memory requirements and accelerate the training process.

Due to the limited size of the dataset and the need to test multiple models to develop an effective solution for keratoconus diagnosis, a 5-fold stratified cross-validation was applied to compare the predictive performance of the models and to fine-tune their parameters. This method involves dividing the dataset into five equal parts and training the algorithm on four of them while testing on the remaining one. In each iteration of the process, a different subset is used for testing, ensuring that each subset maintains a class distribution similar to that of the original dataset, that is, the

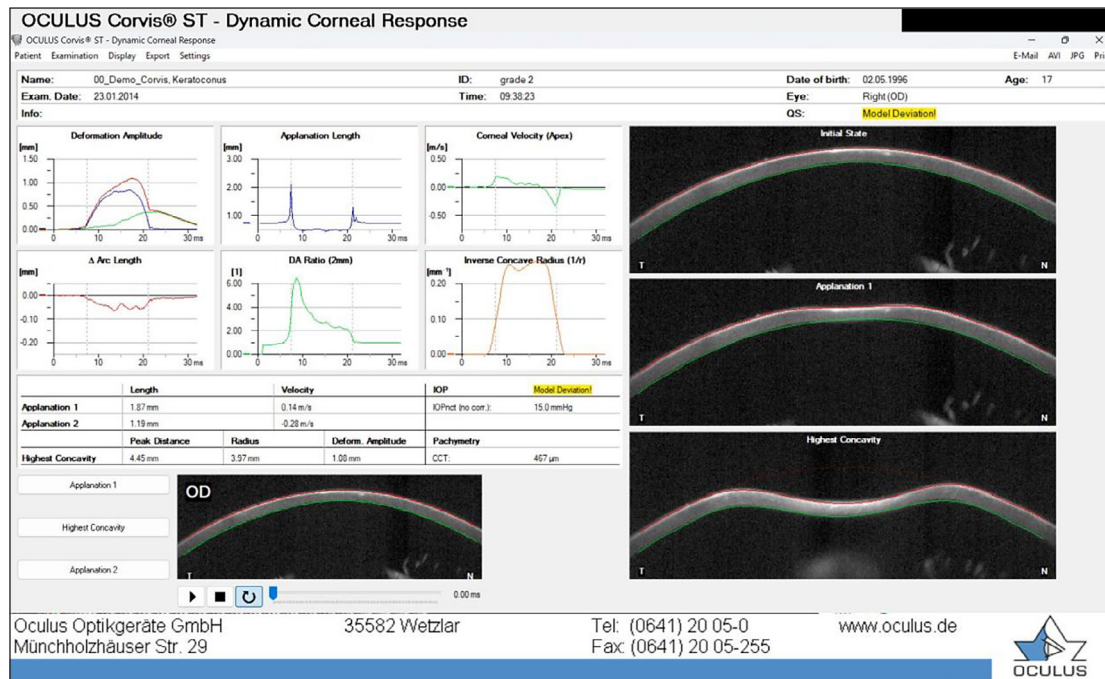


Figure 2. Sample report from the CORVIS device showing the three corneal states, based on demo videos from the CORVIS ST standard database (excerpts from demo videos were shown to avoid revealing sensitive patient data)

number of keratoconus and healthy cases is approximately balanced. This approach helps avoid bias associated with the selection of the test set when determining the model architecture and parameters. According to the data reported in [32], keratoconus affects between 0.2 and 4.790 individuals per 100,000. Consequently, it is difficult to collect a sufficiently large dataset of affected patients. The limited size of the dataset reduces the stability of model predictions, particularly when ambiguous cases are present. For this reason, both parameter selection and classification evaluation were performed using average accuracy values across five validation folds. Although this approach does not provide a definitive estimate of model performance on previously unseen data, it offers a practical compromise that enables meaningful comparison between models.

3D CONVOLUTIONAL NEURAL NETWORKS

The structure of the 3D convolutional neural network model is very similar to standard models used in image analysis, with the key difference being the inclusion of an additional temporal dimension representing changes in the image over

time. In 2D CNNs, two-dimensional convolutions are applied to extract features from local neighborhoods on the feature maps of the previous layer. Weight units are then added, and the result is passed through a sigmoidal activation function. The value at a given position (x, y) in a layer of the network can be expressed as:

$$Y(x, y) = f\left(b + \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{c=0}^{C-1} W(m, n, c) X(x+m, y+n, c)\right) \quad (1)$$

where: f is the sigmoidal activation function (e.g., ReLU), b is the bias unit for the feature map, $X(x+m, y+n, c)$ is the value at position $(x+m, y+n)$ in channel c of the input image, $W(m, n, c)$ is the value at position (m, n) of the filter for channel c , $M \times N$ is the size of the filter, and C is the number of input channels (e.g., 3 for RGB images).

Typically, to reduce sensitivity to distortions in the input data, convolutional layers are followed by pooling layers, which aggregate local neighborhoods on the feature maps. An example of the convolution mechanism for two-dimensional networks is shown in Figure 3.

In video analysis, it is desirable to capture not only spatial features but also temporal features, that is, changes occurring in images over time. At

the input of a 3D convolutional neural network model, the entire video or a selected segment is typically provided, and the convolutional filter takes the shape of a cube. This structure enables the feature maps in the convolutional layer to be connected to multiple adjacent video frames, allowing the network to capture motion information (see Figure 4). Formally, the value at position (x, y, z) can be expressed as:

$$Y(x, y, z) = f\left(b + \sum_{d=0}^{D-1} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{c=0}^{C-1} W(d, m, n, c) X(x+d, y+m, z+n, c)\right) \quad (2)$$

where: D is the size of the filter along the new temporal dimension.

In the diagnosis of keratoconus based on dynamic corneal imaging, both spatial features, related to the cornea's shape, structure, and thickness, and temporal features, i.e., corneal biomechanics – are highly important. Three-dimensional convolutional neural network (3D CNN) models have gained popularity in the scientific community precisely due to their ability to learn spatiotemporal features [33]. For this reason, such models are frequently used for human action recognition [34], and in medicine they can support the detection of discomfort in hospitalized infants [35]. Applications of 3D CNNs in the diagnosis of forme fruste keratoconus have also been reported in the literature, in [36] the authors achieved an accuracy of 87%.

In this article, a comparative study of 21 3D-CNN models of varying complexity and architecture was undertaken in the context of keratoconus detection. The model-building approach was inspired by the VGG16 architecture [37], which has repeatedly demonstrated its effectiveness in extracting features from medical images. For instance, in the study presented in [38], the authors achieved an accuracy of 88% in brain tumor classification using this model. For every model, data were first normalized using a BatchNormalization layer. Subsequent models were constructed by incrementally adding blocks composed of two 3D convolutional layers (kernel size = $3 \times 3 \times 3$, stride = $1 \times 1 \times 1$, ReLU activation, and 'same' padding) followed by one MaxPooling3D layer (pool size = $2 \times 2 \times 2$). The number of Dense layers at the end of the network equaled the number of blocks; when more than one Dense layer was present, a Dropout layer with a rate of 0.3 was inserted between them. In addition, L2 weight regularization with a coefficient of 0.0001 was applied to every layer.

For each network depth, different layer widths were also tested, with a minimum width of 16 filters. Two width profiles were analyzed: constant and a pyramidal profile (increasing width). The inverted-pyramid topology (decreasing width) was not tested because a large number of filters in the initial layers, where the feature-map size had not yet been reduced by pooling or stride, caused Out-Of-Memory (OOM) errors during preliminary trials. A summary of all built and evaluated models is presented in Table 2. It is worth noting

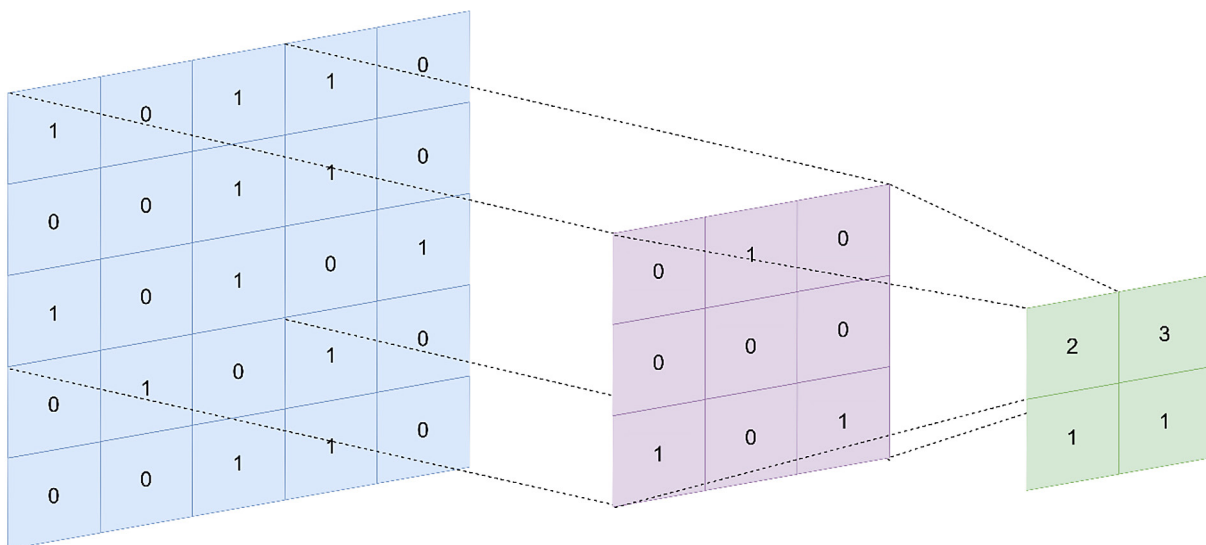


Figure 3. 2D convolution diagram

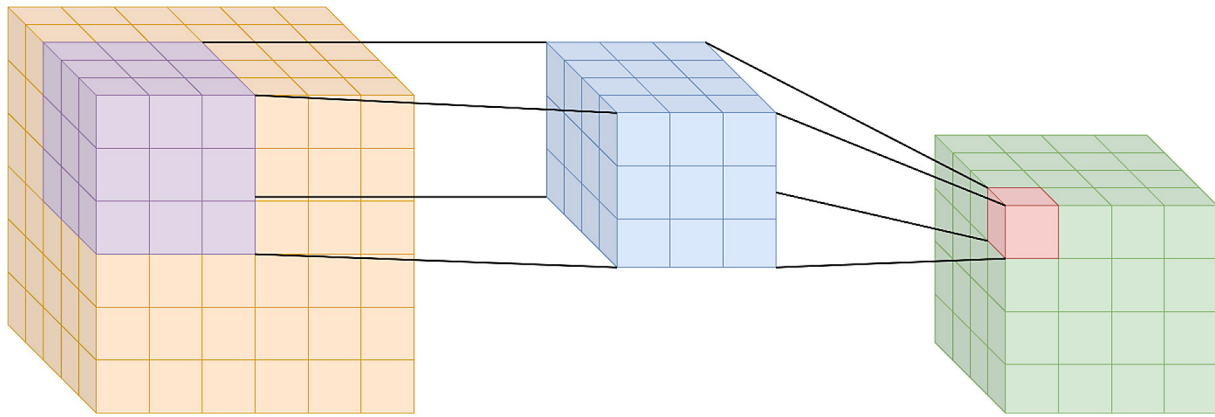


Figure 4. 3D convolution diagram

that the models were expanded by sequentially adding blocks and increasing the number of filters until OOM errors occurred.

The final classification layer consisted of a fully connected layer with a single neuron and a sigmoid activation function. Each developed 3D CNN model was trained for 100 epochs using the Binary Crossentropy loss function and the Adam optimizer with a learning rate set to 0.0001. During training, a Model Checkpoint callback was introduced to enable restoration of the model that achieved the best validation accuracy in terms of loss. A summary of the modeling approach is presented in Figure 5.

RESULTS AND DISCUSSION

For each of the models presented in the previous section, an evaluation was conducted based on four criteria: accuracy, inference time, number of parameters, and GPU memory usage. The accuracy was calculated according to formula:

$$\frac{1}{i} \sum_{i=0}^k \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \quad (3)$$

where: i is the fold index, TP – number of true positives (correctly detected keratoconus cases), TN – number of true negatives (correctly classified healthy cases), FP – number of false positives (videos incorrectly classified as keratoconus), FN – number of false negatives (videos incorrectly classified as healthy).

Since cross-validation was applied, the accuracy and inference time were calculated as the

mean across all folds, while the number of parameters remained constant for each fold, and GPU memory usage was reported as the maximum observed value. These evaluation criteria allowed for a comparison of the models not only in terms of keratoconus detection performance but also in terms of computational complexity and processing speed. Figure 6 presents a plot of model accuracy as a function of model complexity for both constant-width and pyramidal network architectures. It is evident that for both network types, the best results are similar, around 89% accuracy, and are achieved with networks composed of four convolutional blocks. Moreover, it is worth noting that for a given number of blocks, increasing the width improves the performance of constant-width networks, whereas it decreases accuracy in pyramidal architectures. An exception to this is observed in constant-width models with four convolutional blocks, where increasing the width beyond 64 filters causes a slight drop in accuracy.

Figure 7 presents a comparison of the number of parameters in relation to model complexity. First, it should be noted that the number of parameters in both architectures is not linearly dependent on either the number of convolutional blocks or their width. Interestingly, the number of parameters decreases as the number of convolutional blocks increases. This occurs because each block contains a MaxPooling layer, which reduces the data shape, and consequently lowers the number of parameters in the first Dense layer (since the number of parameters in that layer depends on the number and shape of its inputs). It is also evident that increasing the width of the layers results in a higher number of parameters; however, this growth is significantly more rapid in pyramidal models. This is due to the

Table 2. Compared models

ConvNet Configuration																							
A1	A2	B1	B2	B3	C1	C2	D1	D2	D3	D4	E1	E2	E3	F1	F2	F3	F4	G1	G2	G3			
Video input 55 x (100 x 100 RGB frames)																							
Conv3D																							
16	32	16	32	64	1	32	16	32	64	128	16	32	64	16	32	64	128	16	32	64			
Conv3D																							
16	32	16	32	64	16	32	16	32	64	128	16	32	64	16	32	64	128	16	32	64			
Maxpool																							
	Conv3D																						
	16	32	64	32	64	16	32	64	128	32	64	128	16	32	64	128	32	64	128				
	Conv3D																						
	16	32	64	32	64	16	32	64	128	32	64	128	16	32	64	128	32	64	128				
	Maxpool																						
							Conv3D																
							16	32	64	128	64	128	256	16	32	64	128	64	128	256			
							Conv3D																
							16	32	64	128	64	128	256	16	32	64	128	64	128	256			
							Maxpool																
																	Conv3D						
																	16	32	64	128	128	256	512
																	Conv3D						
																	16	32	64	128	128	256	512
											Maxpool												
Dense																							
16	32	16	32	64	64	128	16	32	64	128	128	256	512	16	32	64	128	256	512	1024			
	Dropout (0.3)																						
	Dense																						
	16	32	64	64	128	16	32	64	128	128	256	512	16	32	64	128	256	512	1024				
							Dropout (0.3)																
							Dense																
							16	32	64	128	128	256	512	16	32	64	128	256	512	1024			
																	Dropout (0.3)						
																	Dense						
											16	32	64	128	256	512	1024						
Softmax																							

fact that the X-axis of the plot represents the number of filters in the first convolutional layer. In pyramidal architectures, the number of filters is doubled in each subsequent convolutional block and also doubled again when transitioning to the Dense layers. This means that in a model with three blocks, if the first block contains 32 filters, the corresponding Dense layer would consist of 256 neurons. As a result, pyramidal models have significantly more parameters than constant-width models.

Another criterion related to model complexity was GPU memory usage, as shown in

Figure 8. Although the shape of the plots is similar to those depicting the number of parameters, it is noticeable that for constant-width models, memory usage increases much more rapidly than the number of parameters. This is primarily due to the fact that memory consumption depends not only on the number of parameters but also on the data shape, and even on the specific implementation of the model within the framework. For both constant-width and pyramidal models, the maximum memory usage was comparable and did not exceed 7.5 GB.

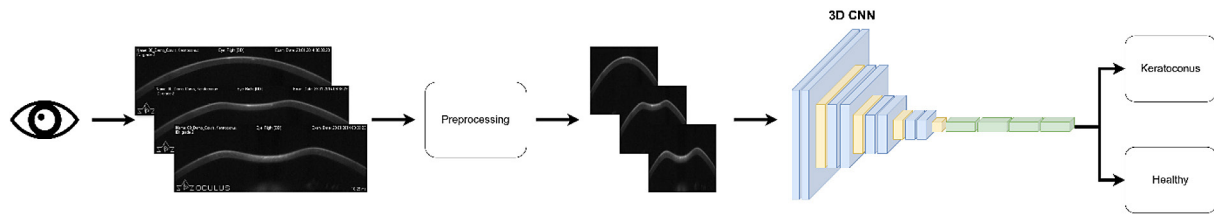


Figure 5. Diagram showing the individual stages of developing a model for diagnosing keratoconus

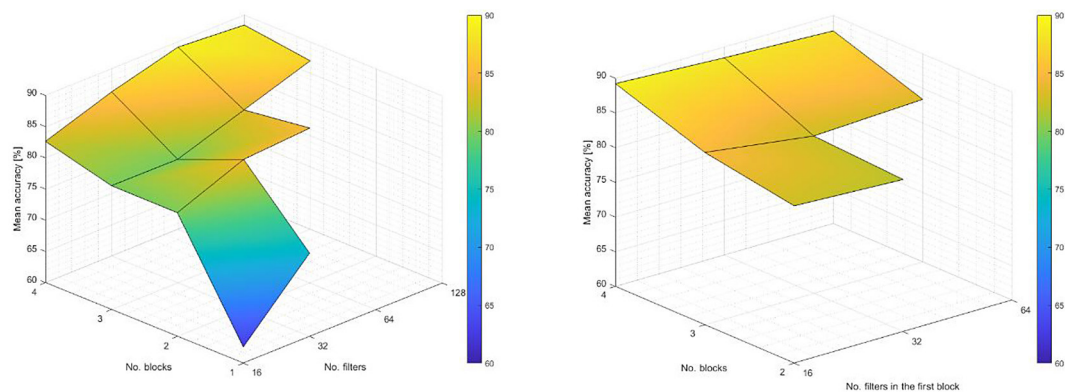


Figure 6. Relationship between mean accuracy, the number of filters and the number of blocks for constant-width layer models (left) and pyramidal models (right)

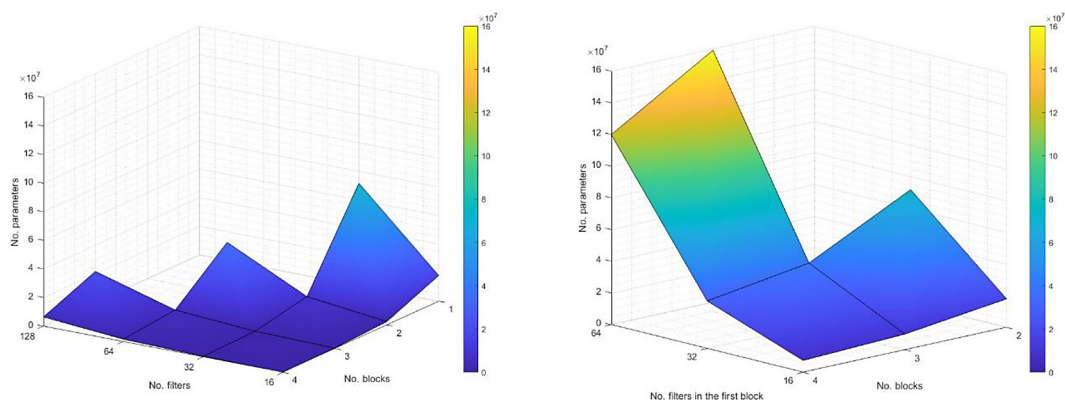


Figure 7. Relationship between number of parameters, the number of filters and the number of blocks for constant-width layer models (left) and pyramidal models (right)

When implementing decision-support systems, an important aspect is the prediction time for a single sample. In the case of keratoconus diagnosis, this refers to the time required to analyze one examination video from the CORVIS ST device. Although keratoconus is not an acute condition that requires immediate action to save a patient's life or health, the diagnosis time should still be short enough to allow comparison with the analysis conducted by a physician. Figure 9 presents the inference time as a function of model complexity for both constant-width and pyramidal architectures. Most importantly, it should be

noted that the prediction time for a single video did not exceed 0.2 seconds, which is well within acceptable limits for building a medical decision-support system. Naturally, inference time depends on model complexity. It is evident that as the number of convolutional blocks increases, inference time also increases, and the relationship appears approximately linear - though more data points would be needed to confirm this precisely. However, it can be assumed that the relationship becomes nonlinear for increasing layer width.

Figure 10 presents a comparison of the accuracy of constant-width and pyramidal models.

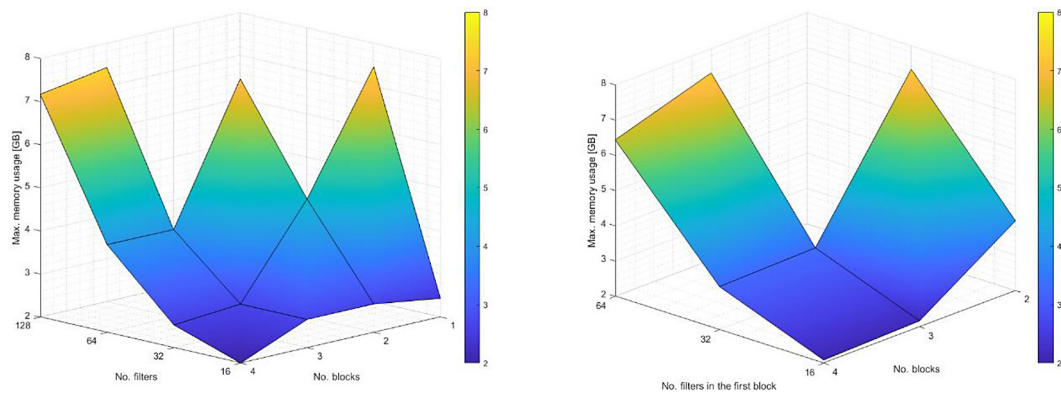


Figure 8. Relationship between maximum GPU memory usage, the number of filters and the number of blocks for constant-width models (left) and pyramidal models (right)

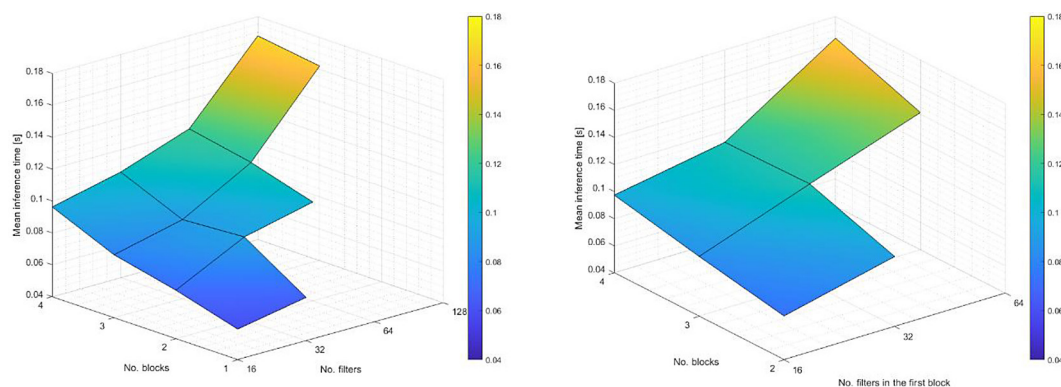


Figure 9. Relationship between inference time, the number of filters and the number of blocks for constant-width models (left) and pyramidal models (right)

Interestingly, the median accuracy is exactly the same for both architectures; however, the constant-width models exhibit a much greater variability in results. Additionally, the pyramidal models did not achieve 100% accuracy in any of the cross-validation folds, in contrast to the constant-width models.

A comparison of inference times for the evaluated 3D convolutional network architectures is shown in Figure 11. It can be observed that, once again, the median value for both architectures is very similar, although not identical in this case. While constant-width models exhibit less variability, the plot contains more outliers compared to pyramidal structures. Naturally, inference time is also influenced by the load on the CPU, GPU and RAM caused by other running processes, which could explain the presence of these outlier values. However, it is important to note that even the outlier inference times did not exceed 0.3 seconds, which remains acceptable from the perspective of keratoconus diagnosis support systems.

Considering the results presented above, it is evident that the highest accuracy among constant-width models was achieved by the model with 4 blocks, each containing 64 filters. In contrast, the best-performing pyramidal architecture was the model with 4 blocks and 16 filters in the first block. A full comparison of the results for these two best models is provided in Table 3. It is worth noting that the constant-width model achieved nearly the same accuracy as the best pyramidal model (a difference of only 0.05%) while using significantly fewer parameters, which translates into a shorter training time. However, the advantage of the pyramidal model lies in its lower memory consumption and shorter inference time.

In medical diagnostics, it is important not only to ensure adequate diagnostic accuracy. Another crucial aspect is minimizing the number of false alarms, as well as maximizing prediction coverage, i.e. the proportion of correctly identified patients. Therefore, to properly evaluate the model, it is necessary to analyze classification

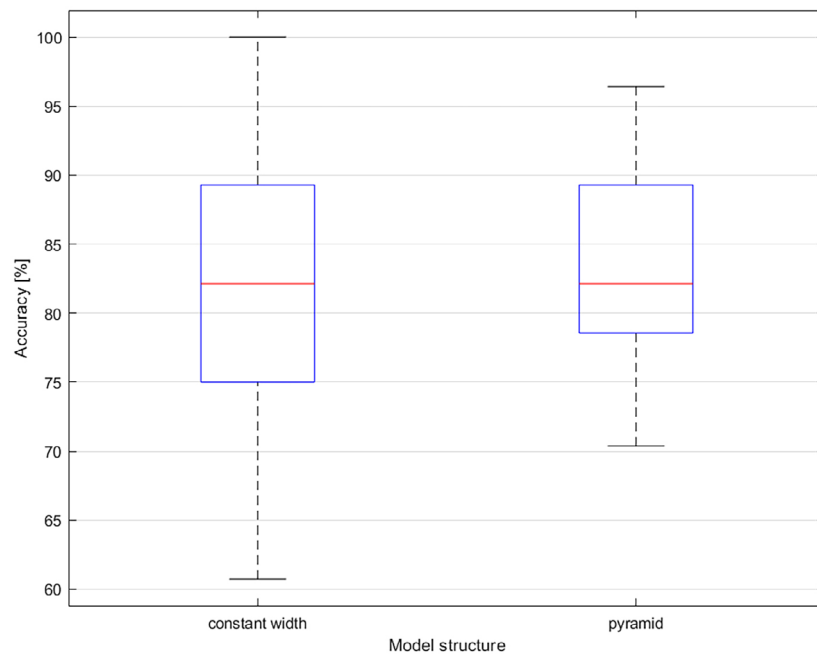


Figure 10. Accuracy comparison for constant width and pyramid model architectures

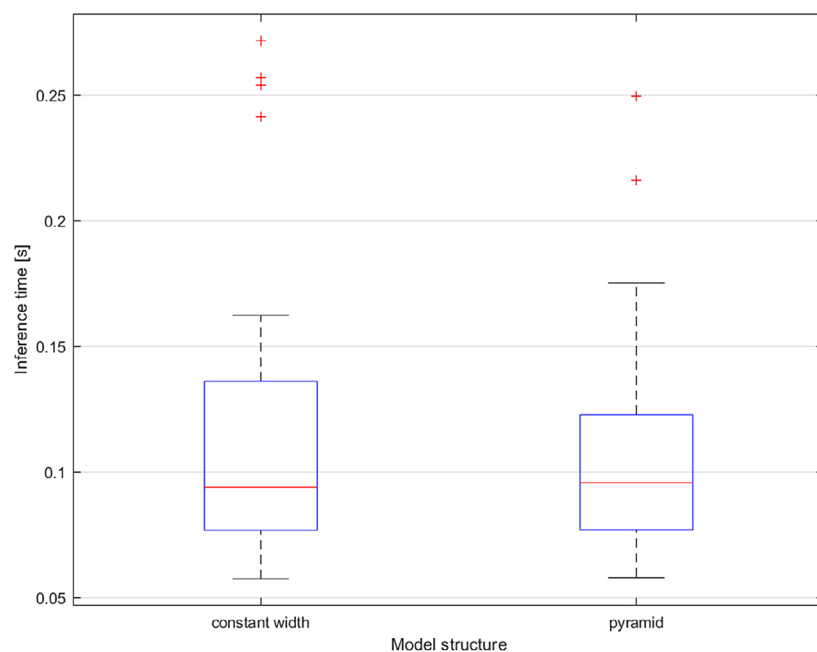


Figure 11. Inference time comparison for constant width and pyramid model architectures

Table 3. Comparison of the best-performing constant-width and pyramidal models

Parameter	Mean accuracy [%]	Mean inference time [s]	Number of parameters [mln]	Memory usage [GB]
Constant width	89.18	0.1214	1.595	4.037
Pyramid	89.23	0.0974	7.500	2.087

metrics such as precision, recall, and F1 score. The values of these metrics for the best-performing model (pyramidal) are presented in Figure 12. It is evident that the accuracy varies considerably

across different cross-validation splits, with a discrepancy of approximately 10%. The other metrics also exhibit substantial variation, which may be attributed to the limited size of the dataset. The

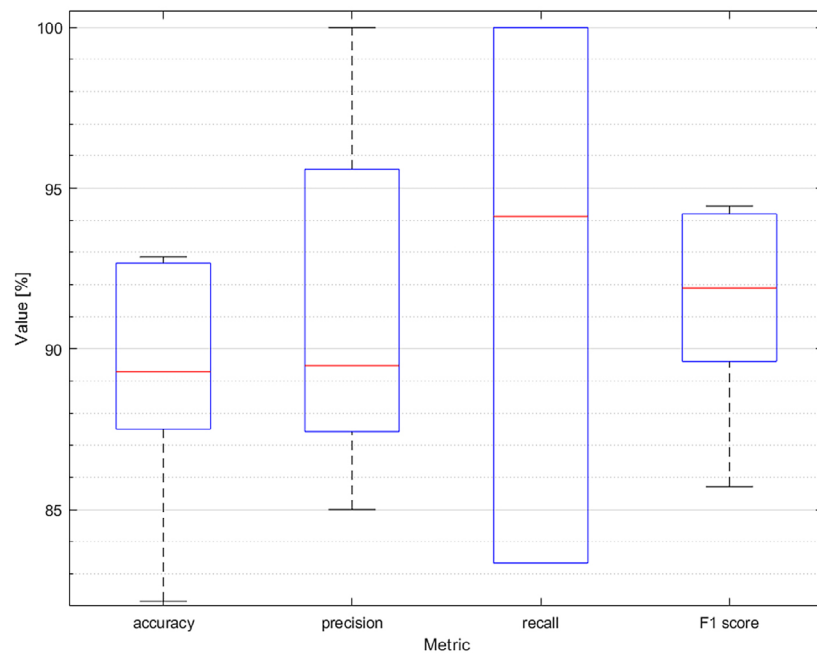


Figure 12. Classification metrics for the best-performing pyramidal model with 4 blocks and 16 filters in the first block

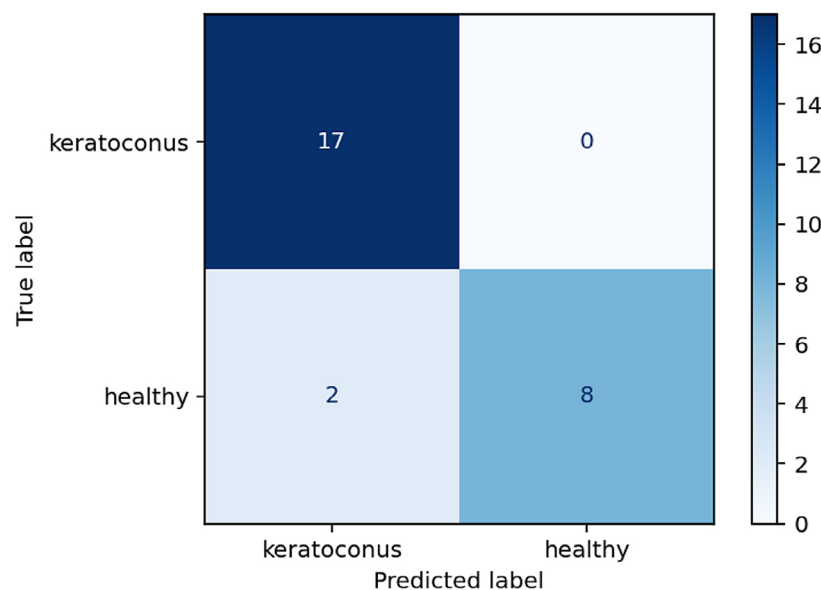


Figure 13. Confusion matrix for the best-performing pyramidal model with 4 blocks and 16 filters in the first block and the final cross-validation split

results further show that the model achieves a considerably higher median recall – 94%, compared to precision – 89%, while the F1 score did not exceed 95% in any of the splits

Figure 13 presents the confusion matrix for the final cross-validation split of the pyramidal model with 4 blocks and 16 filters in the first block. It can be observed that the model mistakenly classified two healthy patients as having

keratoconus, i.e. two false positives were recorded. However, it is important to note that all keratoconus cases were correctly identified. Such model behavior is beneficial from the perspective of physician-assisting systems, as borderline cases can be referred for further diagnostic testing, for example using the Pentacam device. Moreover, it is particularly important that all affected individuals were correctly detected.

Previous studies have highlighted the potential of artificial intelligence methods in keratoconus diagnosis, although most of them rely on pre-extracted features rather than on raw data. The extraction of corneal biomechanical features based on deformation curve profiles allowed neural networks to achieve up to 99% diagnostic accuracy [21]. While this result is remarkable, it does not address the computational requirements and time constraints. Similarly, the inclusion of topographic data obtained from Pentacam device enabled a Random Forest model to achieve 89% classification accuracy [22]. This result is comparable to the accuracy of our best-performing 3D CNN, but does not take into account the cornea dynamics and stiffness, which makes the approach less generalizable to heterogeneous populations. In other study, similar combination of features yielded an AUC of 93% [23], but not considering the inference time and memory usage. A broader multimodal analysis that combined data from three devices (Sirius, Pentacam and Corvis) identified Sirius-derived parameters as the most effective indicators (AUC = 91%), followed by Corvis parameters (AUC = 82%) [24]. Although this highlights the benefit of multimodal fusion, it also introduces challenges related to the simultaneous use of multiple diagnostic devices, which limits practical applicability in routine ophthalmic practice. In a different research, authors explored transfer learning approaches applied to images from the Corvis device [25]. A comparison of ResNet152, VGG16 and InceptionV3 showed that ResNet152 achieved the highest diagnostic accuracy (AUC = 99%). While this illustrates the strength of deep feature representations, the dependence on pretrained 2D CNN models does not fully capture the temporal dynamics of cornea deformation.

In contrast to the above approaches, our study employs complete video sequences directly obtained from the Corvis ST device and provides a systematic comparison of 21 distinct 3D CNN architectures. This design not only eliminates the need of dedicated feature extraction model but also enables the network to learn spatiotemporal features directly from raw cornea deformation videos. The evaluation of accuracy, inference time, number of parameters and GPU memory usage in keratoconus diagnosis adds a practical perspective that is absent in prior studies, where the main focus was achieving the best diagnostic effectiveness. Our results demonstrate that pyramidal architectures achieved superior accuracy and sensitivity while consuming less memory

than constant-width networks, which emphasizes that it is possible and beneficial to optimize models in this context. Nevertheless, further external validation on broader and more heterogeneous dataset is needed to confirm the robustness of the proposed approach. Although the achieved accuracy of over 89% and sensitivity of 85.9% are competitive compared to the results of [22] and exceed those of [36].

CONCLUSIONS

This study presents a comparison of the effectiveness of keratoconus detection using 3D convolutional neural network models of varying complexity and architecture, based on dynamic corneal imaging from the CORVIS ST device. The videos collected during the examinations were properly preprocessed and labeled. A 5-fold stratified cross-validation was employed to identify the model with the highest generalization capability.

The study indicates that both constant-width and pyramidal models can achieve comparable classification accuracy; however, the latter exhibit lower memory usage and shorter inference times. Additionally, pyramidal models demonstrate less variability in performance across different cross-validation splits. The best-performing model achieved an average keratoconus detection accuracy of 89%, while maintaining high sensitivity.

The obtained results confirm the potential of the model as an effective and reliable tool for early keratoconus detection or for use in screening programs. The high performance of the model demonstrates its ability to effectively complement existing diagnostic procedures, potentially reducing the number of misdiagnoses and supporting clinical decision-making.

Nevertheless, the model's tendency toward false positives represents an area for further improvement and should be addressed prior to its implementation in a medical environment. In addition, the generalizability of the model must be evaluated on larger and more diverse datasets. A further limitation of the proposed 3D neural network models lies in their relatively high GPU memory requirements, since diagnostic workstations are typically not equipped with dedicated computational units. A potential solution may involve the use of cloud-based infrastructures; however, this approach necessitates careful consideration of data security and the protection of

sensitive patient information contained in clinical examinations. The implementation of a clinical decision-support system will also require enhancing the interpretability of diagnostic results. This can be partially addressed by adopting explainable AI methods or providing the probability of class membership rather than a label, allowing physicians to pay particular attention to patients for whom the model is uncertain.

Acknowledgements

This research was supported by Polish Ministry of Education and Science, grant no. MEiN/2023/DPI/2194, project title: Lublin Digital Union – Use of Digital Solutions and Artificial Intelligence in Medicine – Research Project.

REFERENCES

- Kotecha A. What biomechanical properties of the cornea are relevant for the clinician? *Surv Ophthalmol.* 2007;52 Suppl 2:S109–114. <https://doi.org/10.1016/j.survophthal.2007.08.004>
- Teraï N, Raiskup F, Haustein M, Pillunat LE, Spoerl E. Identification of biomechanical properties of the cornea: the ocular response analyzer. *Curr Eye Res.* 2012;37(7):553–62. <https://doi.org/10.3109/02713683.2012.669007>
- Maurice DM. CHAPTER 6 - The Cornea and Sclera. *Vegetative Physiology and Biochemistry.* Academic Press; 1962;289–368. <https://doi.org/10.1016/B978-1-4832-3090-0.50013-7>
- Meek KM. The Cornea and Sclera. *Collagen: Structure and Mechanics.* Boston, MA: Springer US; 2008;359–96. https://doi.org/10.1007/978-0-387-73906-9_13
- Vellara HR, Patel DV. Biomechanical properties of the keratoconic cornea: a review. *Clin Exp Optom.* 2015;98(1):31–8. <https://doi.org/10.1111/cxo.12211>
- Romero-Jiménez M, Santodomingo-Rubido J, Wolffsohn JS. Keratoconus: a review. *Cont Lens Anterior Eye.* 2010;33(4):157–66. <https://doi.org/10.1016/j.clae.2010.04.006>
- Vazirani J, Basu S. Keratoconus: current perspectives. *Clin Ophthalmol.* 2013;7:2019–30. <https://doi.org/10.2147/OPHTH.S50119>
- Shi Y. Strategies for improving the early diagnosis of keratoconus. *Clin Optom (Auckl).* 2016;8:13–21. <https://doi.org/10.2147/OPTO.S63486>
- Belin MW, Jang HS, Borgstrom M. Keratoconus: Diagnosis and Staging. *Cornea.* 2022;41(1):1–11. <https://doi.org/10.1097/ICO.0000000000002781>
- Fernández Pérez J, Valero Marcos A, Martínez Peña FJ. Early diagnosis of keratoconus: what difference is it making? *Br J Ophthalmol.* 2014;98(11):1465–6. <https://doi.org/10.1136/bjophthalmol-2014-305120>
- Zhang X, Munir SZ, Sami Karim SA, Munir WM. A review of imaging modalities for detecting early keratoconus. *Eye (Lond).* 2021;35(1):173–87. <https://doi.org/10.1038/s41433-020-1039-1>
- Schweitzer C, Roberts CJ, Mahmoud AM, Colin J, Maurice-Tison S, Kerautret J. Screening of forme fruste keratoconus with the ocular response analyzer. *Invest Ophthalmol Vis Sci.* 2010;51(5):2403–10. <https://doi.org/10.1167/iovs.09-3689>
- Fontes BM, Ambrósio R, Velarde GC, Nosé W. Ocular response analyzer measurements in keratoconus with normal central corneal thickness compared with matched normal control eyes. *J Refract Surg.* 2011;27(3):209–15. <https://doi.org/10.3928/1081597X-20100415-02>
- Touboul D, Bénard A, Mahmoud AM, Gallois A, Colin J, Roberts CJ. Early biomechanical keratoconus pattern measured with an ocular response analyzer: curve analysis. *J Cataract Refract Surg.* 2011;37(12):2144–50. <https://doi.org/10.1016/j.jcrs.2011.06.029>
- Tian L, Huang YF, Wang LQ, Bai H, Wang Q, Jiang JJ, i in. Corneal biomechanical assessment using corneal visualization scheimpflug technology in keratoconic and normal eyes. *J Ophthalmol.* 2014;2014:147516. <https://doi.org/10.1155/2014/147516>
- Qin X, Tian L, Zhang H, Chen X, Li L. Evaluation of corneal elastic modulus based on Corneal Visualization Scheimpflug Technology. *BioMedical Engineering OnLine.* 2019;18(1):42. <https://doi.org/10.1186/s12938-019-0662-1>
- Piñero DP, Alcón N. In vivo characterization of corneal biomechanics. *J Cataract Refract Surg.* 2014;40(6):870–87. <https://doi.org/10.1016/j.jcrs.2014.03.021>
- Huseynova T, Waring GO, Roberts C, Krueger RR, Tomita M. Corneal biomechanics as a function of intraocular pressure and pachymetry by dynamic infrared signal and Scheimpflug imaging analysis in normal eyes. *Am J Ophthalmol.* 2014;157(4):885–93. <https://doi.org/10.1016/j.ajo.2013.12.024>
- Hong J, Xu J, Wei A, Deng SX, Cui X, Yu X, i in. A new tonometer—the Corvis ST tonometer: clinical comparison with noncontact and Goldmann applanation tonometers. *Invest Ophthalmol Vis Sci.* 2013;54(1):659–65. <https://doi.org/10.1167/iovs.12-10984>
- Valbon BF, Ambrósio R, Fontes BM, Luz A, Roberts CJ, Alves MR. Ocular biomechanical metrics by CorVis ST in healthy Brazilian patients. *J Refract Surg.* 2014;30(7):468–73. <https://doi.org/10.3928/1081597X-20140521-01>

21. Tan Z, Chen X, Li K, Liu Y, Cao H, Li J, i in. Artificial intelligence-based diagnostic model for detecting keratoconus using videos of corneal force deformation. *Transl Vis Sci Technol.* 2022;11(9):32. <https://doi.org/10.1167/tvst.11.9.32>
22. Castro-Luna G, Jiménez-Rodríguez D, Castaño-Fernández AB, Pérez-Rueda A. Diagnosis of subclinical keratoconus based on machine learning techniques. *J Clin Med.* 2021;10(18):4281. <https://doi.org/10.3390/jcm10184281>
23. Liu Y, Zhang Y, Chen Y. Application of a scheimpflug-based biomechanical analyser and tomography in the early detection of subclinical keratoconus in chinese patients. *BMC Ophthalmology.* 2021;21(1):339. <https://doi.org/10.1186/s12886-021-02102-2>
24. Heidari Z, Hashemi H, Mohammadpour M, Amanzadeh K, Fotouhi A. Evaluation of corneal topographic, tomographic and biomechanical indices for detecting clinical and subclinical keratoconus: a comprehensive three-device study. *Int J Ophthalmol.* 2021;14(2):228–39. <https://doi.org/10.18240/ijo.2021.02.08>
25. Kuo BI, Chang WY, Liao TS, Liu FY, Liu HY, Chu HS, i in. Keratoconus screening based on deep learning approach of corneal topography. *Transl Vis Sci Technol.* 2020;9(2):53. <https://doi.org/10.1167/tvst.9.2.53>
26. Nguyen T, Ong J, Masalkhi M, Waisberg E, Zaman N, Sarker P, i in. Artificial intelligence in corneal diseases: A narrative review. *Cont Lens Anterior Eye.* 2024;47(6):102284. <https://doi.org/10.1016/j.clae.2024.102284>
27. Hu Z, Zheng H, Gui J. A Novel Interactive Image Processing Approach for DICOM Medical Image Data. W: 2009 2nd International Conference on Biomedical Engineering and Informatics. 2009. s. 1–4. <https://doi.org/10.1109/BMEI.2009.5304840>
28. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, i in. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* 2017;77(21):e104–7. <https://doi.org/10.1158/0008-5472.CAN-17-0339>
29. Chiuchisan I. Implementation of medical image processing algorithm on reconfigurable hardware. W: 2013 E-Health and Bioengineering Conference (EHB). 2013;1–4. <https://doi.org/10.1109/EHB.2013.6707298>
30. Memon K, Yahya N, Yusoff MZ, Remli R, Mustapha AWMM, Hashim H, i in. Edge computing for AI-Based brain MRI applications: A critical evaluation of real-time classification and segmentation. *Sensors.* 2024;24(21):7091. <https://doi.org/10.3390/s24217091>
31. Canny J. A Computational approach to edge detection. *IEEE transactions on pattern analysis and machine intelligence.* 1986;PAMI-8(6):679–98. <https://doi.org/10.1109/TPAMI.1986.4767851>
32. Bui AD, Truong A, Pasricha ND, Indaram M. Keratoconus diagnosis and treatment: recent advances and future directions. *Clin Ophthalmol.* 2023;17:2705–18.
33. Ji S, Xu W, Yang M, Yu K. 3D convolutional neural networks for human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 2013;35(1):221–31. <https://doi.org/10.1109/TPAMI.2012.59>
34. Al-Akam R, Paulus D, Gharabaghi D. Human action recognition based on 3D convolution neural networks from RGBD videos. *Computer Science Research Notes.* 2018;2803(May):18–26. <https://doi.org/10.24132/CSRN.2018.2803.3>
35. Sun Y, Hu J, Wang W, He M, de With PHN. Camera-based discomfort detection using multi-channel attention 3D-CNN for hospitalized infants. *Quant Imaging Med Surg.* 2021;11(7):3059–69. <https://doi.org/10.21037/qims-20-1302>
36. Rong H, Liu G, Wang Y, Hu J, Sun Z, Gao N, i in. Using 3D convolutional neural network and corvis st corneal dynamic video for detecting forme fruste keratoconus. *J Refract Surg.* 2025;41(4):e356–64. <https://doi.org/10.3928/1081597X-20250226-01>
37. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv.org.* 2014. <https://doi.org/10.48550/arXiv.1409.1556>
38. Sterniczuk B, Charytanowicz M. An ensemble transfer learning model for brain tumors classification using convolutional neural networks. *Adv Sci Technol Res J.* 2024;18(8):204–16. <https://doi.org/10.12913/22998624/193627>