

## Article

# Deep Learning-Based Diagnosis of Corneal Condition by Using Raw Optical Coherence Tomography Data

Maziar Mirsalehi <sup>1,\*</sup>, Michael Schwemm <sup>1</sup>, Elias Flockerzi <sup>2</sup>, Nóra Szentmáry <sup>1</sup>, Alaa Din Abdin <sup>2</sup>,  
Berthold Seitz <sup>2</sup> and Achim Langenbucher <sup>1</sup>

<sup>1</sup> Department of Experimental Ophthalmology, Saarland University, Kirrberger Street 100, 66424 Homburg, Saarland, Germany

<sup>2</sup> Department of Ophthalmology, Saarland University Medical Centre UKS, Kirrberger Street 100, 66424 Homburg, Saarland, Germany

\* Correspondence: maziar.mirsalehi@uni-saarland.de

## Abstract

**Background/Objectives:** Keratoconus (KC) is the most common corneal ectasia. This condition affects quality of vision, especially when it is progressive, and a timely and stage-related treatment is mandatory. Therefore, early diagnosis is crucial to preserve visual acuity. Medical data may be used either in their raw state or in a preprocessed form. Software modifications introduced through updates may potentially affect outcomes. Unlike preprocessed data, raw data preserve their original format across software versions and provide a more consistent basis for clinical analysis. The objective of this study was to distinguish between healthy and KC corneas from raw optical coherence tomography data by using a convolutional neural network. **Methods:** In total, 2737 eye examinations acquired with the Casia2 anterior-segment optical coherence tomography (Tomey, Nagoya, Japan) were decided by three experienced ophthalmologists to belong to one of three classes: ‘normal’, ‘ectasia’, or ‘other disease’. Each eye examination consisted of sixteen meridional slice images. The dataset included 744 examinations. DenseNet121, EfficientNet-B0, MobileNetV3-Large and ResNet18 were modified for use as convolutional neural networks for prediction. All reported metric values were rounded to four decimal places. **Results:** The overall accuracy for the modified DenseNet121, modified EfficientNet-B0, modified MobileNetV3-Large and modified ResNet18 is 91.27%, 91.27%, 92.86% and 89.68%, respectively. The macro-averaged sensitivity, macro-averaged specificity, macro-averaged Positive Predictive Value and macro-averaged F1 score for the modified DenseNet121, modified EfficientNet-B0, modified MobileNetV3-Large and modified ResNet18 are reported as 91.27%, 91.27%, 92.86% and 89.68%; 95.63%, 95.63%, 96.43% and 94.84%; 91.58% 91.65%, 92.91% and 90.24%; and 91.35%, 91.29%, 92.85% and 89.81%, respectively. **Conclusions:** The successful use of a convolutional neural network with raw optical coherence tomography data demonstrates the potential of raw data to be used instead of preprocessed data for diagnosing KC in ophthalmology.

**Keywords:** CNN; cornea; deep learning; ectasia; eye; GUI; keratoconus; OCT; raw data; vision



Academic Editor: Jae-Ho Han

Received: 25 October 2025

Revised: 1 December 2025

Accepted: 3 December 2025

Published: 8 December 2025

**Citation:** Mirsalehi, M.; Schwemm, M.; Flockerzi, E.; Szentmáry, N.; Abdin, A.D.; Seitz, B.; Langenbucher, A. Deep Learning-Based Diagnosis of Corneal Condition by Using Raw Optical Coherence Tomography Data.

*Diagnostics* **2025**, *15*, 3115.

<https://doi.org/10.3390/diagnostics15243115>

**Copyright:** © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license

(<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Sight is an important ability for human beings, and quality of life can be affected by eye problems. Human eyes contain tissues and structures such as the cornea, lens, retina

and optic nerve. The cornea covers the outer surface of the eyeball. Corneal conditions commonly include refractive errors such as myopia, hyperopia and astigmatism, as well as keratoconus (KC) and inflammation [1].

KC, as a quasi-inflammatory disease, is a bilateral condition that causes progressive thinning and steepening of the cornea and affects between 50 and 230 individuals per 100,000 [2,3]. Diagnosis of this disease typically occurs in the second or third decade of life. Eye rubbing is regarded as a major risk factor in the KC development [4]. Early KC can be detected using various diagnostic methods [5], including handheld keratoscopes (Placido disks) [6], slit-lamp biomicroscopy [7] and ultrasonic pachymetry [8], as well as corneal topography [9] and tomography techniques [10] such as Scheimpflug imaging [6] and anterior-segment Optical Coherence Tomography (OCT) [11]. OCT employs low-coherence interferometry with near-infrared light to generate high-resolution images of tissue morphology, including corneal layer-thickness maps [12].

Artificial intelligence (AI) has become increasingly valuable in ophthalmology, particularly for image analysis. Since the 1970s, its application in diagnosing eye diseases has grown substantially. Applications of artificial intelligence are being explored in the diagnosis and management of glaucoma, KC, cataracts and other anterior segment diseases [13]. There has been a growing volume of research assessing the implementation of AI-based methods for diagnosing anterior segment diseases, with particular emphasis on KC detection through anterior-segment OCT imaging [14]. Using AI to detect KC dates back to the earliest study by Maeda et al. [15] in 1994. Several deep learning approaches have been proposed in recent years for KC detection and the classification of the cornea through various corneal imaging techniques [16–21].

In this study, raw OCT data obtained from the Casia2 (Tomey, Nagoya, Japan) were used to diagnose the condition of the cornea. Data can generally be used in their original format or in a preprocessed form, which bears the risk of changes with software updates. In contrast to preprocessed data, raw data grabbed by the instrument typically remain unchanged with new software tools and updates, and potentially offer a more reliable foundation for analysis. The objective of this study is to use raw OCT data to diagnose corneal conditions and to develop a method for automatically distinguishing healthy corneas from KC and other conditions, using data labelled by three experienced ophthalmologists (PD Dr. Alaa Din Abdin, PD Dr. Elias Flockerzi and Univ.-Prof. Dr. Nóra Szentmáry).

## 2. Materials and Methods

### 2.1. Data

Patient data were collected at the Department of Ophthalmology, Saarland University Medical Centre, Homburg, Germany, using the cornea/anterior-segment optical coherence tomography device Casia2 (Tomey, Nagoya, Japan) [22], a swept-source OCT-based device that captures anterior segment images of the eye. The Casia2 generates raw data in 3dv format, with each file linked to a corneal map. Each 3dv file contains information for sixteen equi-angular meridional images. The images were stored in sixteen-bit unsigned integer format, with pixel values ranging from 0 to 65,535. Figure 1 shows information from a raw data file of the Casia2.

[[21921 21192 22633 ... 20013 22933 22006]	[[18940 21607 20113 ... 21528 21788 18504]	[[20552 24271 22668 ... 18336 21444 22130]	[[22159 22541 19216 ... 22992 20004 20141]
[22572 20962 21910 ... 21382 17792 20625]	[21955 21603 12128 ... 23074 22440 18976]	[21421 23992 22109 ... 16914 22879 22859]	[21349 20258 21148 ... 22912 17940 22246]
[21118 20652 20811 ... 21996 19096 20088]	[20873 21968 20391 ... 24197 21972 19191]	[21214 22238 22113 ... 18490 21163 20956]	[20618 20194 23429 ... 17843 18921 22802]
...	...	...	...
[18834 17166 19530 ... 21903 20096 19138]	[17301 20973 16561 ... 18738 21055 16377]	[20265 21758 20918 ... 20595 20274 19718]	[21091 21742 19963 ... 21366 22090 22139]
[17910 17389 18664 ... 21219 19509 18244]	[20323 22383 19523 ... 16434 19913 21335]	[15296 23392 20598 ... 18750 19295 20958]	[14121 19561 16450 ... 20313 20411 21525]
[19031 19773 20264 ... 16458 20707 21481]]	[19234 19713 17978 ... 18471 15376 21338]]	[14901 21571 20483 ... 18601 14649 21167]]	[20131 20120 20405 ... 12038 21453 21815]]
...	...	...	...
[[22726 20896 17598 ... 20444 22712 22851]	[[22105 20014 21047 ... 18607 22655 20969]	[[20872 20576 21691 ... 20104 23566 21136]	[[21657 24626 23315 ... 20970 20773 21881]
[20804 20521 20263 ... 21002 20707 21120]	[22039 20433 22933 ... 21347 17573 19108]	[20580 18873 21834 ... 21460 23953 17863]	[21822 23395 20727 ... 19330 22665 23545]
[20401 22026 22391 ... 21251 22427 21029]	[21749 19770 23928 ... 21014 18763 20393]	[19695 20971 22883 ... 20274 22984 17875]	[22429 20335 18138 ... 17926 21890 22844]
...	...	...	...
[19645 22655 17436 ... 17196 21861 18429]	[21161 16498 20163 ... 15846 19719 17549]	[18486 14397 18028 ... 21948 21726 18307]	[23754 24240 19932 ... 22461 15284 19877]
[22397 18713 18049 ... 19618 20239 20362]	[21738 17887 20999 ... 19382 20359 21287]	[18026 19685 22318 ... 19563 20649 17968]	[21128 22779 18695 ... 20147 21749 21264]
[22020 19972 19553 ... 19569 17315 21472]]	[16521 17306 18495 ... 15489 19909 19830]]	[11393 13115 19673 ... 19406 18823 21081]]	[24063 24490 21229 ... 20110 20676 23532]]
...	...	...	...
[[23104 20847 21830 ... 21884 19955 22610]	[[19987 21831 22445 ... 22301 16345 21888]	[[20409 19143 22157 ... 20446 24083 19433]	[[20835 22145 23403 ... 22135 22815 21211]
[22762 23300 24425 ... 15954 22232 22961]	[13453 19992 18449 ... 22547 21897 21540]	[20718 23197 20985 ... 19199 23393 19913]	[19886 19552 21835 ... 23463 20550 19176]
[19831 22981 23539 ... 19227 20033 21835]	[21711 21700 21240 ... 22086 23707 23850]	[20045 23857 16483 ... 19899 18558 17976]	[19610 20999 19567 ... 23730 19855 18558]
...	...	...	...
[18039 17801 22688 ... 20827 18991 15708]	[19730 19217 21133 ... 13105 20804 20573]	[19104 18427 19069 ... 17071 22094 19496]	[19866 20336 17638 ... 19987 20520 19278]
[21646 17696 21510 ... 16451 12535 19175]	[19726 17338 20509 ... 17629 21744 19995]	[13203 21024 19813 ... 20580 22384 20312]	[19659 21373 16881 ... 18079 19715 21871]
[17888 18480 17936 ... 20738 18971 20167]]	[22015 19586 19094 ... 8386 20970 21615]]	[14938 21156 17455 ... 18535 20741 16244]]	[21734 22588 19514 ... 15169 21147 21597]]
...	...	...	...
[[23556 17761 18973 ... 16580 22224 15423]	[[22284 19708 21404 ... 21412 22021 14419]	[[19101 22405 21029 ... 14991 18741 22507]	[[22533 24898 19669 ... 19753 21331 16022]
[19966 16883 21224 ... 21727 22596 19918]	[23323 18896 21300 ... 21371 20418 20662]	[16285 22049 18592 ... 18984 15091 20432]	[22805 23330 23149 ... 21594 23892 19124]
[19871 22345 21794 ... 23484 22943 21717]	[23210 22889 22490 ... 8695 20637 21205]	[19710 22771 18491 ... 20778 21687 22524]	[21562 18215 21635 ... 20764 24497 20840]
...	...	...	...
[19743 19695 22890 ... 21097 20690 18563]	[19207 20193 20706 ... 21675 17128 19704]	[16140 17221 18499 ... 22165 18819 17146]	[19367 19450 22377 ... 18098 19761 16552]
[17945 16701 19817 ... 16875 22464 18036]	[20184 18618 13772 ... 21631 21145 23383]	[21292 20355 17546 ... 22293 18520 19480]	[16859 20756 20356 ... 22290 20387 17875]
[20300 18451 19873 ... 20276 20607 21677]]	[18317 19153 17882 ... 22877 19537 21502]]	[21118 17865 19763 ... 17652 19213 22178]]	[18741 21531 18664 ... 19492 19040 20594]]

**Figure 1.** Information from a raw data file.

A Python script was developed to extract sixteen images from each 3dv file. Each image, originally sized at 800 pixels wide and 1464 pixels tall, was stored as a greyscale Portable Network Graphics (PNG) file. Figure 2 shows sixteen images extracted from the same raw data file as presented in Figure 1.

A total of 2737 eye examinations, each containing sixteen equi-angular meridional images with the aspect ratio adjusted to 1.629 (width divided by height) to better represent the realistic shape of the eye, were decided by three experienced ophthalmologists (PD Dr. Alaa Din Abdin, PD Dr. Elias Flockerzi and Univ.-Prof. Dr. Nóra Szentmáry) to belong to one of three classes: ‘normal’, which indicates a normal eye; ‘ectasia’, which indicates corneal ectasia; or ‘other disease’, which included eyes with penetrating keratoplasty, subepithelial or stromal scarring, corneal dystrophies, Salzmann’s nodules or pterygia. All three experienced ophthalmologists were in agreement in labelling 1325 eye examinations as ‘normal’, 212 eye examinations as ‘ectasia’ and 266 eye examinations as ‘other disease’. The split was performed at the eye examination level,

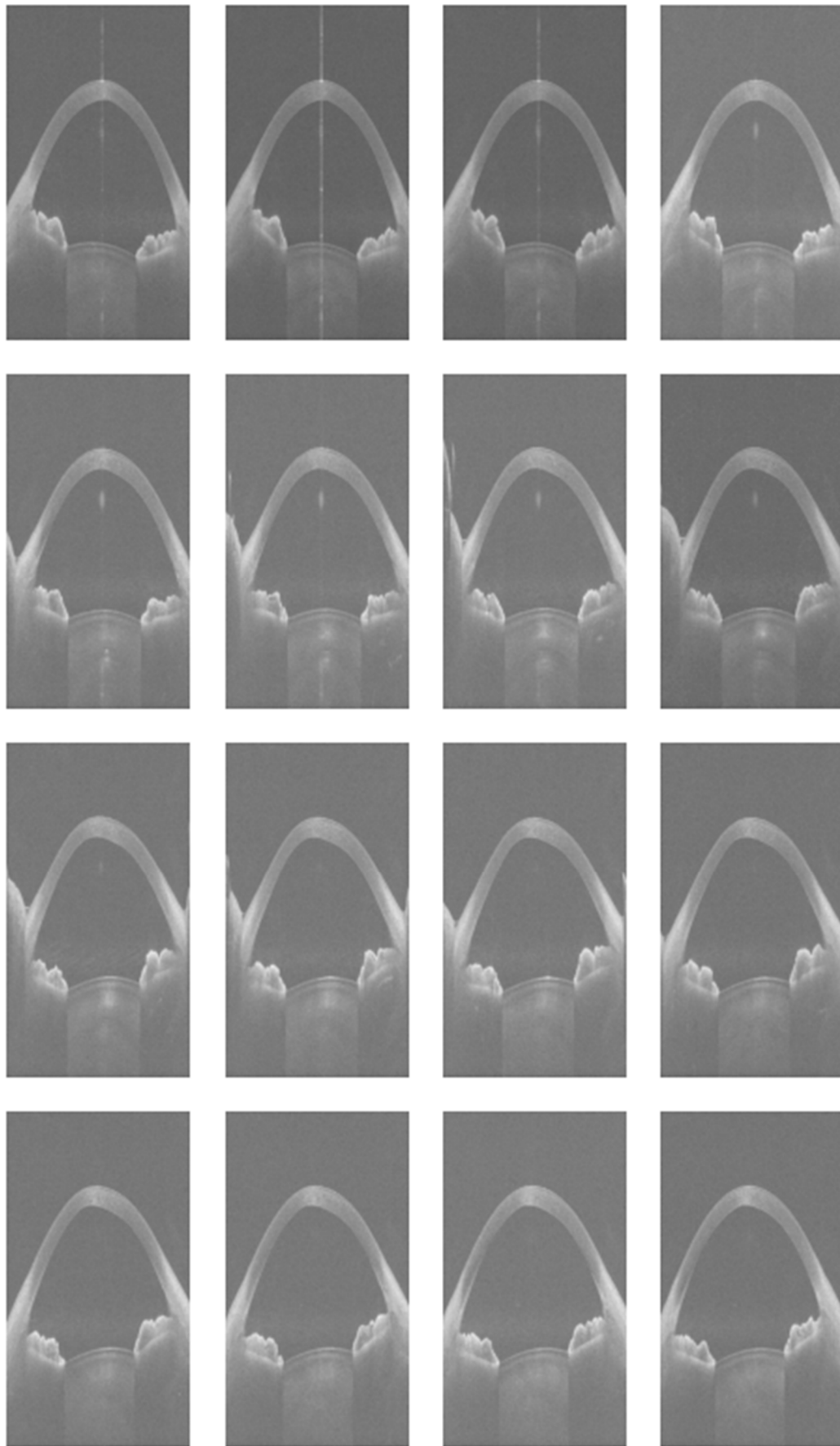
based on the sixteen equi-angular meridional images, rather than at the patient level. Since the labelled eye examinations were not balanced, it was decided to reduce the number of ‘normal’ eye examinations to make the dataset more balanced. All 212 eye examinations categorised as ‘ectasia’ and all 266 examinations classified as ‘other disease’ were used, while 266 ‘normal’ eye examinations were randomly selected to achieve a more balanced dataset. It was decided to create a balanced dataset for both the validation and test datasets. To allocate approximately 60% of the dataset for training, 20% for validation and 20% for testing, it was decided to use 42 of the 212 ‘ectasia’-labelled eye examinations (approximately 20%), for the test dataset and another 42 (approximately 20%) for the validation dataset and the remaining 128 for the training dataset. To create balanced validation and test datasets, it was decided to use 42 ‘normal’-labelled eye examinations and 42 ‘other disease’-labelled eye examinations for the validation dataset, another 42 ‘normal’-labelled and 42 ‘other disease’-labelled eye examinations for the test dataset, and the remaining 182 ‘normal’-labelled and 182 ‘other disease’-labelled eye examinations for training. After randomly shuffling the rows corresponding to the ‘normal’ class in a CSV file containing the image file names and labels of 212 eye examinations categorised as ‘ectasia’, 266 eye examinations categorised as ‘other disease’ and 266 eye examinations categorised as ‘normal’, the first 182 eye examinations were selected for the training dataset, the next 42 for the validation dataset and the following 42 for the test dataset. The ‘ectasia’ class was split in a similar manner. After randomly shuffling the rows corresponding to this class in the same CSV file, the first 128 eye examinations were selected for the training dataset, the next 42 for the validation dataset and the following 42 for the test dataset. The ‘other disease’ class was also split in the same way. After randomly shuffling the rows corresponding to this class in the same CSV file, the first 182 eye examinations were selected for the training dataset, the next 42 for the validation dataset and the following 42 for the test dataset.

Eye examinations for which the labels did not match were excluded from the analysis. The training dataset included 182 eye examinations for the ‘normal’ class, 128 eye examinations for the ‘ectasia’ class and 182 eye examinations for the ‘other disease’ class (approximately 66.13% of the total dataset). The validation dataset included 42 eye examinations for each class (approximately 16.93% of the total dataset), and the test dataset also included 42 eye examinations for each class (approximately 16.93% of the total dataset) to ensure a balanced distribution. Table 1 summarises the label distribution across the training, validation and test datasets.

**Table 1.** Label distribution across training, validation and test datasets.

Dataset	Normal	Ectasia	Other Disease	Total
Training	182	128	182	492
Validation	42	42	42	126
Test	42	42	42	126





**Figure 2.** Sixteen equi-angular meridional images extracted from a raw data file.

## 2.2. Training Architectures

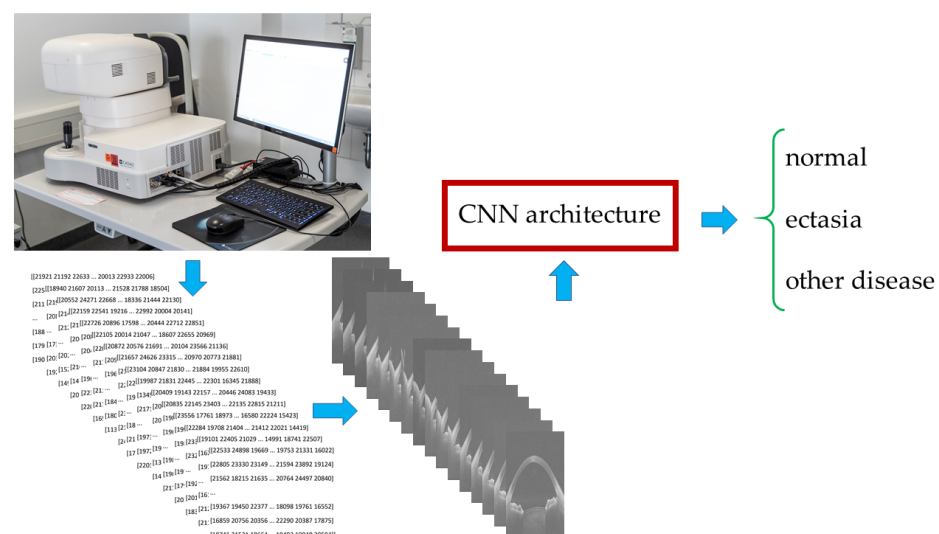
A Convolutional Neural Network (CNN) is a specialised form of artificial neural network structured for image data. It operates by means of convolutional layers that apply kernels, or filters, to detect features and form feature maps [23,24]. In this study, four CNN models from well-established deep learning architectures were selected. DenseNets offer advantages, such as easing the vanishing-gradient problem, strengthening feature propagation and reducing the number of parameters, and they were evaluated on the CIFAR-10, CIFAR-100, ImageNet and SVHN datasets [25]. EfficientNets use a compound scaling method, which can improve accuracy better than other single-dimension scaling methods, and they were evaluated on the Birdsnap, CIFAR-10, CIFAR-100, FGVC Aircraft, Flowers, Food-101, ImageNet, Oxford-IIIT Pets and Stanford Cars datasets [26]. MobileNets can be applied to mobile and embedded vision applications, and they were evaluated on ImageNet dataset [27]. ResNets have residual networks that are easier to optimise, and they were evaluated on the CIFAR-10 and ImageNet datasets [28]. In this study, DenseNet121, EfficientNet-B0, MobileNetV3-Large and ResNet18 were used as the CNN models. The images were resized to  $224 \times 224$  pixels to match the input dimensions required by the models. The resized images were normalised between  $-1$  and  $1$ . The models were trained from scratch using Python (version 3.13.5) and the PyTorch library (version 2.7.1) [29]. Training was performed on a system with an Intel(R) Xeon(R) Central Processing Unit (CPU) E5-1650 v2 @ 3.50 GHz 3.50 GHz processor and 32 GB of Random-Access Memory (RAM). The training process consisted of 200 full passes through the dataset. The Python scripts were designed to test the CNN models corresponding to the epoch that achieved the highest validation accuracy among the 200 epochs on the test dataset; if more than one epoch achieved the same highest validation accuracy, the model from the epoch with the highest training accuracy among them was used for evaluation on the test dataset. A batch size of 8 was applied consistently across the training, validation and test sets to ensure stable learning dynamics. To optimise the classification performance, the cross-entropy loss function [30] was selected as the loss metric. The models were optimised using the AdamW optimiser [31]. The initial learning rate was set to 0.01. To adjust the learning rate dynamically during training, a scheduler was employed to reduce the rate by a factor of 0.1 when no improvement occurred over 10 consecutive epochs. Additionally, a weight decay [32] value of 0.05 was applied to regularise the models and reduce the risk of overfitting [33] by constraining the magnitude of the learned parameters.

To accommodate the specific nature of the input data, each group of sixteen resized images was treated as a single volumetric instance and combined into a sixteen-channel input. This stack was then fed into the modified models. To enable the models to process this multi-channel input, the architectures were altered by modifying the first convolutional layer to accept sixteen input channels instead of the standard 3-channel Red–Green–Blue (RGB) format. Furthermore, the classification head of the networks, originally designed for a 1000-class output in ImageNet tasks, was adapted to output three predictions, corresponding to one of the three target classes: ‘normal’, ‘ectasia’ or ‘other disease’. Table 2 shows the differences between the original DenseNet121, EfficientNet-B0, MobileNetV3-Large and ResNet18 models and their modified versions used in this study.

Figure 3 shows the process, from obtaining data from the Casia2 to predicting one of the three classes.

**Table 2.** Comparison of the original CNN models with their modified versions. Abbreviation: CNN = Convolutional Neural Network.

CNN Model	Number of Input Channels	Number of Output Features
DenseNet121	3	1000
Modified DenseNet121	16	3
EfficientNet-B0	3	1000
Modified EfficientNet-B0	16	3
MobileNetV3-Large	3	1000
Modified MobileNetV3-Large	16	3
ResNet18	3	1000
Modified ResNet18	16	3



**Figure 3.** Data acquisition to three-class classification workflow. Abbreviation: CNN = Convolutional Neural Network.

### 2.3. Evaluation Metrics

For each class, the numbers of True Positives (TPs), False Positives (FPs), True Negatives (TNs) and False Negatives (FNs) were computed using a one-vs.-rest approach. These values were used to derive standard evaluation metrics adapted for multi-class classification, including accuracy, sensitivity, specificity, Positive Predictive Value (PPV) and F1 score. The metrics were calculated according to the definitions provided in [34], with macro-averaging strategies applied where appropriate to ensure fair assessment across all classes. All reported metric values were rounded to four decimal places.

### 2.4. Gradient-Weighted Class Activation Mapping

Gradient-Weighted Class Activation Mapping (Grad-CAM) [35] was used to visualise the regions of the CNN architectures that are most influential in distinguishing between the three classes: ‘normal’, ‘ectasia’ and ‘other disease’ on the test dataset. The Grad-CAMs for each predicted class were averaged to produce a representative Grad-CAM pattern for that class. The Python scripts were designed to use the model corresponding to the epoch that achieved the highest validation accuracy among the 200 epochs to generate Grad-CAM on the test dataset; if more than one epoch achieved the same highest validation accuracy, the model from the epoch with the highest training accuracy among them was used for

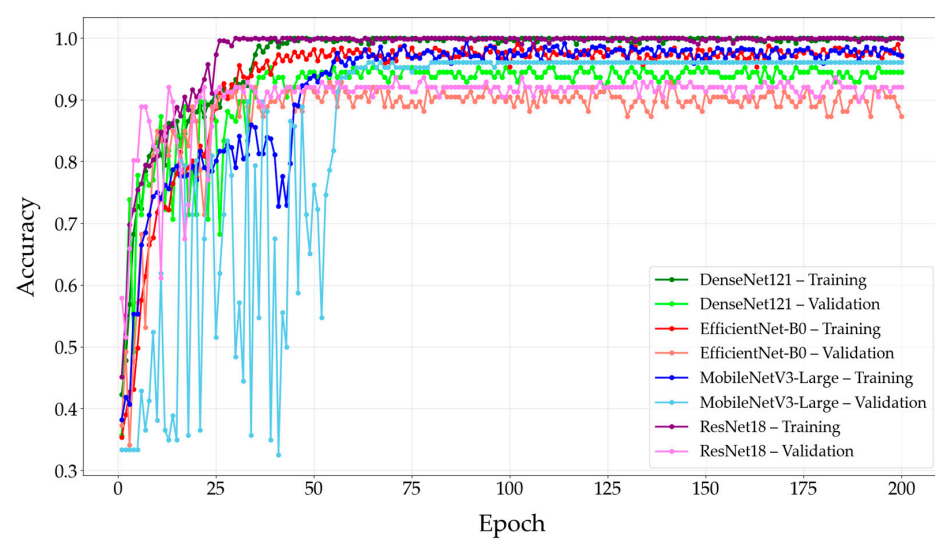
generating Grad-CAM on the test dataset. For each modified CNN model, Grad-CAM was applied to the last convolutional layer.

### 2.5. Graphical User Interface

A Graphical User Interface (GUI) can help users visualize the prediction results. The GUI was created by using Tkinter (version 8.6), OpenCV (cv2, version 4.12.0), Matplotlib (version 3.10.3), NumPy (version 2.2.6) and Pathlib (Python version 3.13.5). The selected CNN architecture, based on its performance on the test dataset, was implemented by using PyTorch (version 2.7.1). The GUI was developed on the system running Python (version 3.13.5).

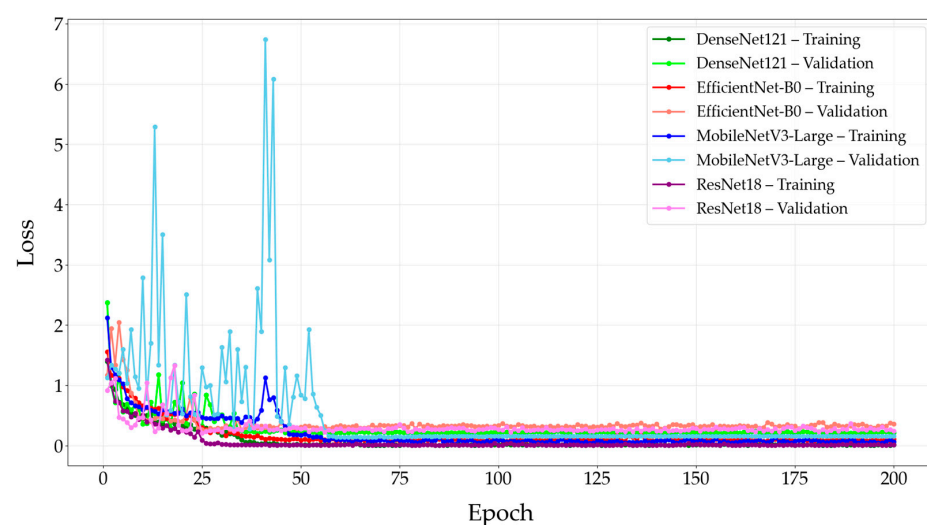
## 3. Results

The training and validation accuracies of the modified CNN models over 200 epochs are shown in Figure 4.



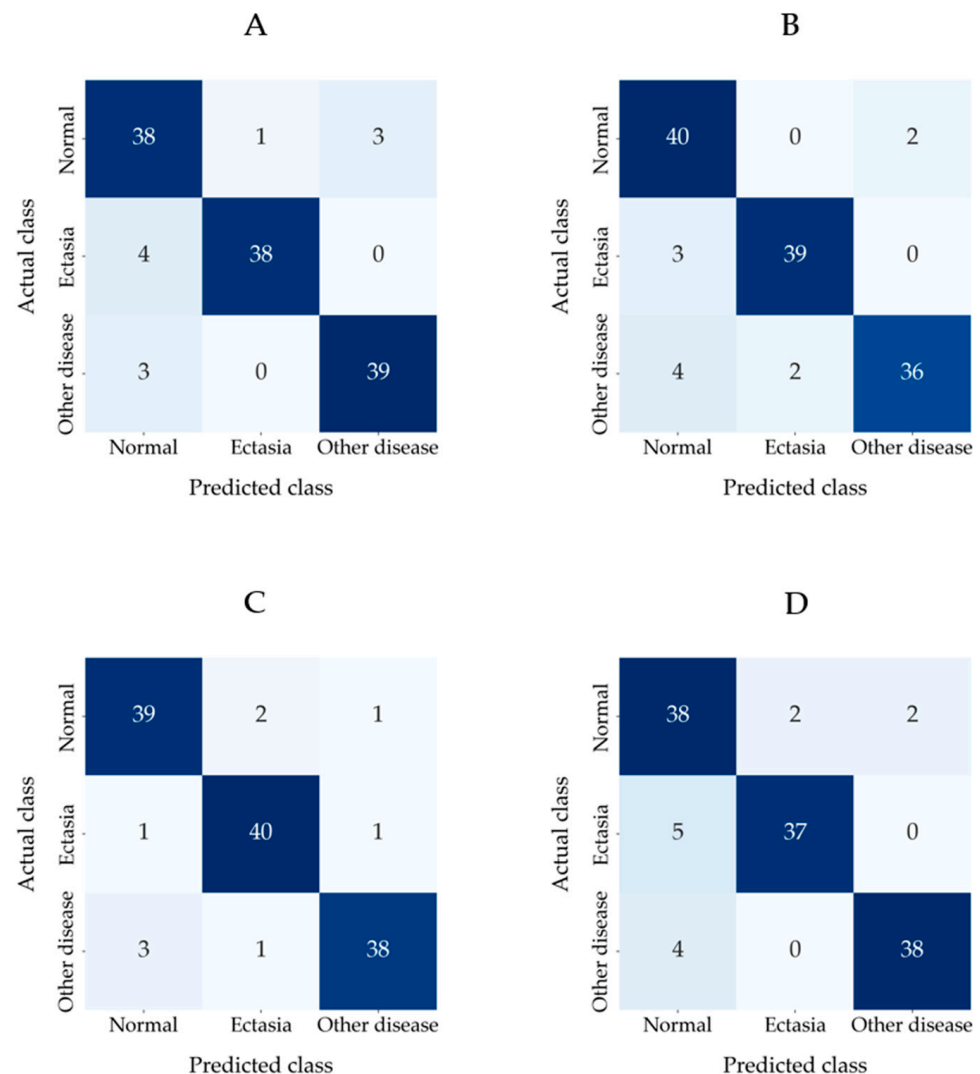
**Figure 4.** Training and validation accuracies of the modified CNN models over epochs. Abbreviation: CNN = Convolutional Neural Network.

The training and validation losses of the modified CNN models over 200 epochs are shown in Figure 5.



**Figure 5.** Training and validation losses of the modified CNN models over epochs. Abbreviation: CNN = Convolutional Neural Network.

The distribution of predictions among the classes for the modified CNN models is shown in the confusion matrix in Figure 6.



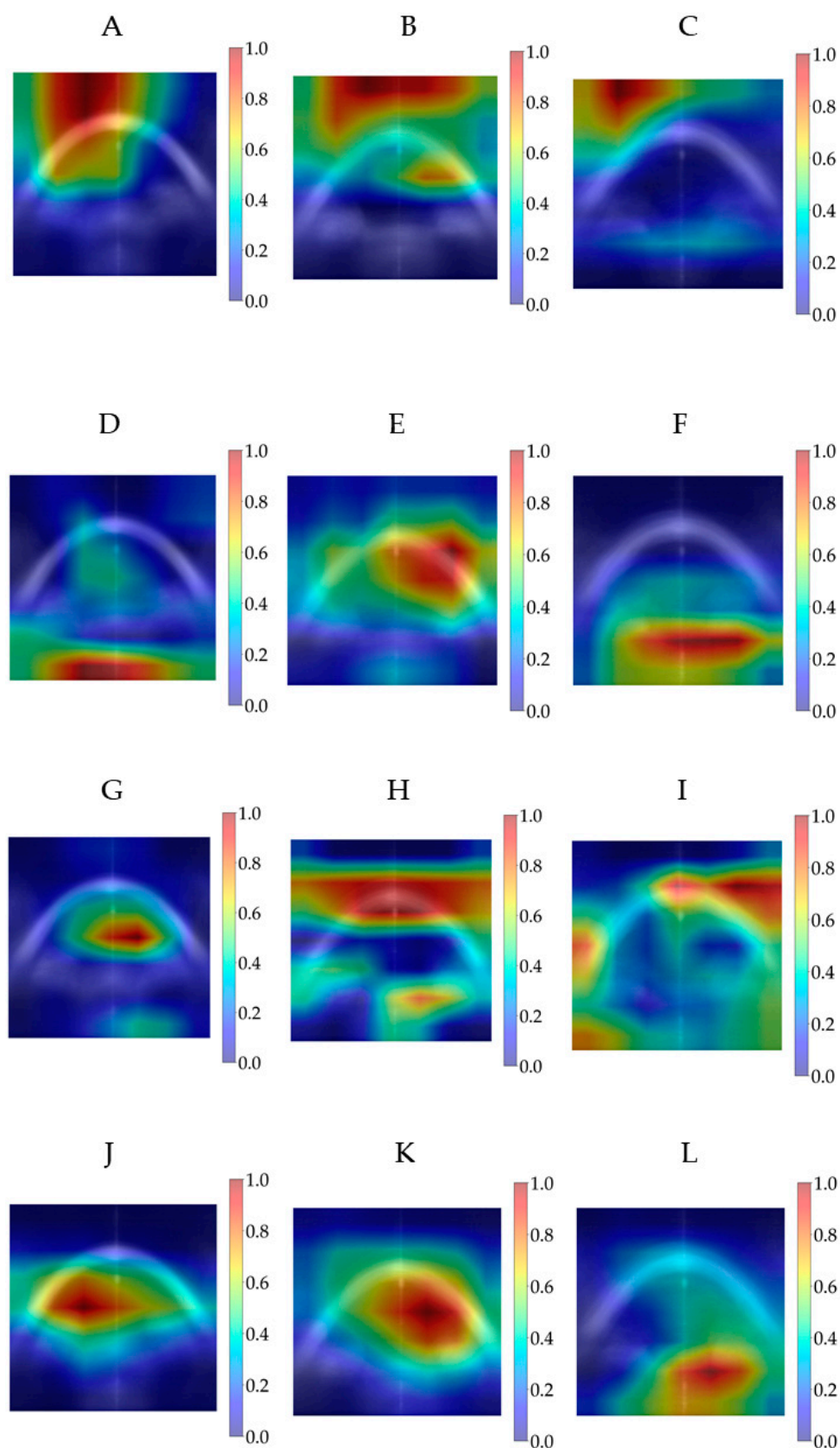
**Figure 6.** Confusion matrices of the modified CNN models: (A) modified DenseNet121, (B) modified EfficientNet-B0, (C) modified MobileNetV3-Large and (D) modified ResNet18.

Figure 7 shows the averaged Grad-CAM for each predicted class of the modified CNN models in the test dataset. The CNNs' decision importance was colour-coded from blue, for little or no influence, to red, for strong influence on the prediction. Each Grad-CAM for each prediction was overlaid on the average of all sixteen resized extracted images. The averaged Grad-CAM for each class was overlaid on the average of all those averaged images, where the lighter areas represent the averaged background images.

Table 3 shows the performance metrics of the modified DenseNet121, modified EfficientNet-B0, modified MobileNetV3-Large and modified ResNet18 evaluated on the test dataset by using seed number one.

Figure 8 shows the GUI for corneal condition prediction. The GUI has three buttons: 'File', 'Images' and 'Prediction'. The user can select a raw data file in .3dv format by clicking the 'File' button. By clicking the 'Images' button, the user can view the extracted images, which are automatically scaled, as shown in Figure 9. The user can obtain a prediction of the corneal condition and the probabilities of each class, which are rounded to two decimal places, by clicking the 'Prediction' button. Figure 10 shows the prediction output for a raw data file.





**Figure 7.** Averaged Grad-CAM for each predicted class in the test dataset for modified DenseNet121, (A) 'normal', (B) 'ectasia' and (C) 'other disease'; for modified EfficientNet-B0, (D) 'normal', (E) 'ectasia' and (F) 'other disease'; for modified MobileNetV3-Large, (G) 'normal', (H) 'ectasia' and (I) 'other disease'; and for modified ResNet18, (J) 'normal', (K) 'ectasia' and (L) 'other disease'. Abbreviation: Grad-CAM = Gradient-Weighted Class Activation Mapping.

**Table 3.** Performance metrics of the modified CNN models evaluated on the test dataset. Abbreviation: PPV = Positive Predictive Value.

Model	Group	Sensitivity	Specificity	PPV	F1 Score	Overall Accuracy
Modified DenseNet121	Normal	0.9048	0.9167	0.8444	0.8736	0.9127
	Ectasia	0.9048	0.9881	0.9744	0.9383	
	Other disease	0.9286	0.9643	0.9286	0.9286	
	Macro-average	0.9127	0.9563	0.9158	0.9135	
Modified EfficientNet-B0	Normal	0.9524	0.9167	0.8511	0.8989	0.9127
	Ectasia	0.9286	0.9762	0.9512	0.9398	
	Other disease	0.8571	0.9762	0.9474	0.9	
	Macro-average	0.9127	0.9563	0.9165	0.9129	
Modified MobileNetV3-Large	Normal	0.9286	0.9524	0.9070	0.9176	0.9286
	Ectasia	0.9524	0.9643	0.9302	0.9412	
	Other disease	0.9048	0.9762	0.95	0.9268	
	Macro-average	0.9286	0.9643	0.9291	0.9285	
Modified ResNet18	Normal	0.9048	0.8929	0.8085	0.8539	0.8968
	Ectasia	0.8809	0.9762	0.9487	0.9136	
	Other disease	0.9048	0.9762	0.95	0.9268	
	Macro-average	0.8968	0.9484	0.9024	0.8981	

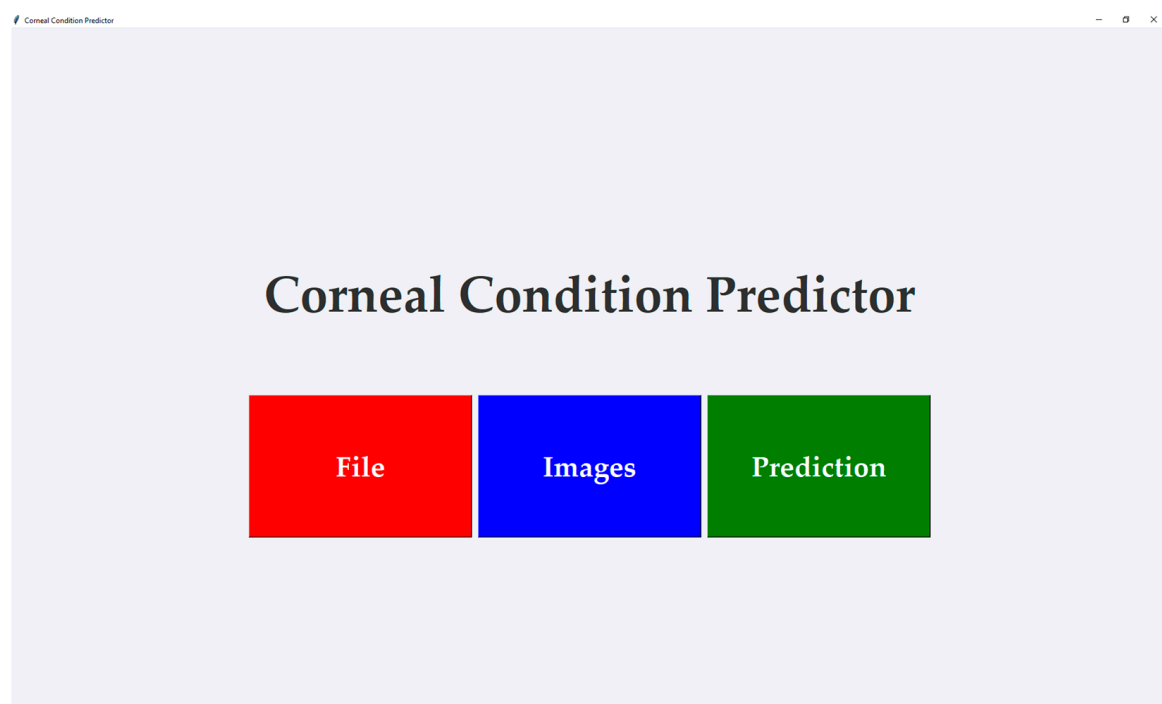
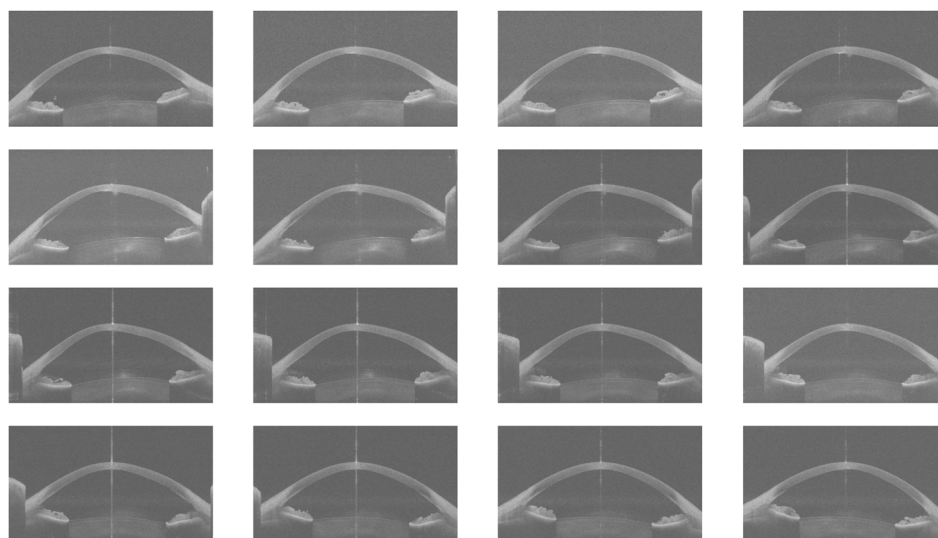
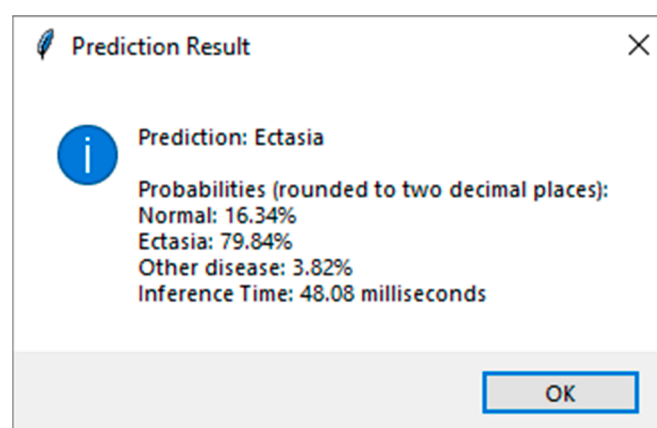
**Figure 8.** GUI for corneal condition prediction. Abbreviation: GUI = Graphical User Interface.

Figure 1

- □ ×



**Figure 9.** Images extracted and scaled from the selected raw data file.



**Figure 10.** Prediction result of the selected raw data file.

#### 4. Discussion

In this study, four deep learning models were modified to diagnose corneal conditions using the raw data from the Casia2. The raw data exported from the Casia2 consist of sixteen meridional images. Accordingly, sixteen channels were introduced to the modified models to receive the sixteen images.

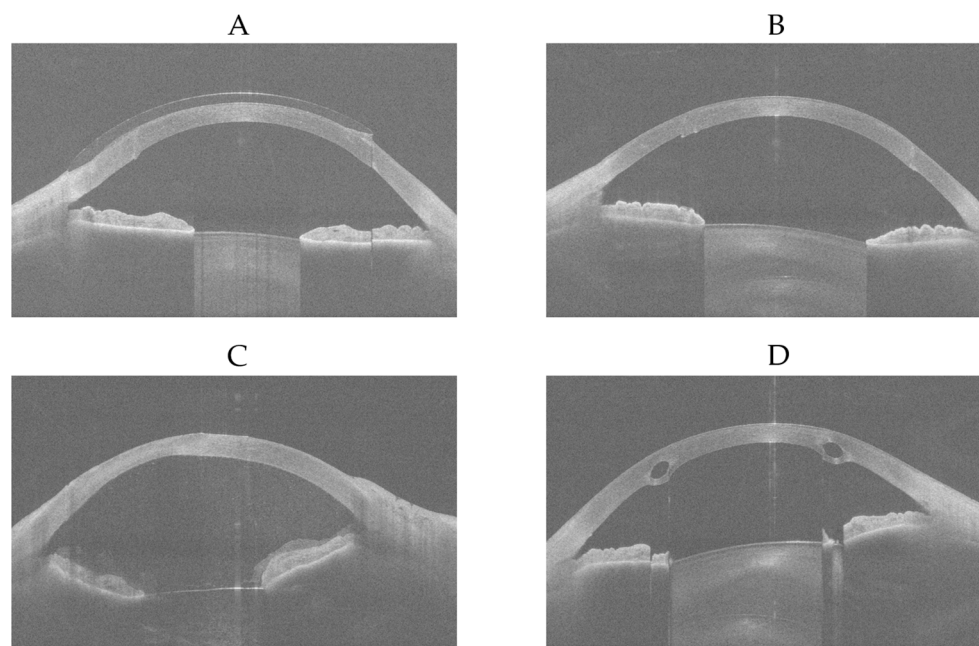
Among four modified CNN models, the modified MobileNetV3-Large achieved the highest overall accuracy of 92.86%. For the modified MobileNetV3-Large, among the 200 epochs, epoch 89 achieved the highest accuracy on the validation dataset. The model from epoch 89 was used for evaluation on the test dataset and for generating Grad-CAM. According to the results, the modified MobileNetV3-Large correctly predicted 117 out of 126 examinations. The 'ectasia' class had the highest number of correct predictions, with 40 out of 42, followed by 'normal', with 39 out of 42, and 'other disease', with 38 out of 42. Since the modified MobileNetV3-Large had higher overall accuracy than the other modified CNN models, it was run four additional times with seed numbers from two to five

to validate the robustness and reliability of the results. Table 4 shows the mean  $\pm$  standard deviation for each performance metric across five runs of the modified MobileNetV3-Large by using seed numbers from one to five.

**Table 4.** Performance metrics (mean  $\pm$  standard deviation) of the modified MobileNetV3-Large evaluated on the test dataset across five runs with different seed numbers. Abbreviation: PPV = Positive Predictive Value.

Model	Group	Sensitivity	Specificity	PPV	F1 Score	Overall Accuracy
Modified MobileNetV3-Large	Normal	0.9095 $\pm$ 0.0391	0.9048 $\pm$ 0.0895	0.8424 $\pm$ 0.1125	0.8703 $\pm$ 0.0631	0.8937 $\pm$ 0.0519
	Ectasia	0.9191 $\pm$ 0.1019	0.9643 $\pm$ 0.0266	0.9332 $\pm$ 0.0475	0.9214 $\pm$ 0.0418	
	Other disease	0.8524 $\pm$ 0.0865	0.9714 $\pm$ 0.0065	0.9366 $\pm$ 0.0165	0.8910 $\pm$ 0.0528	
	Macro-average	0.8937 $\pm$ 0.0519	0.9468 $\pm$ 0.0260	0.9041 $\pm$ 0.0357	0.8942 $\pm$ 0.0502	

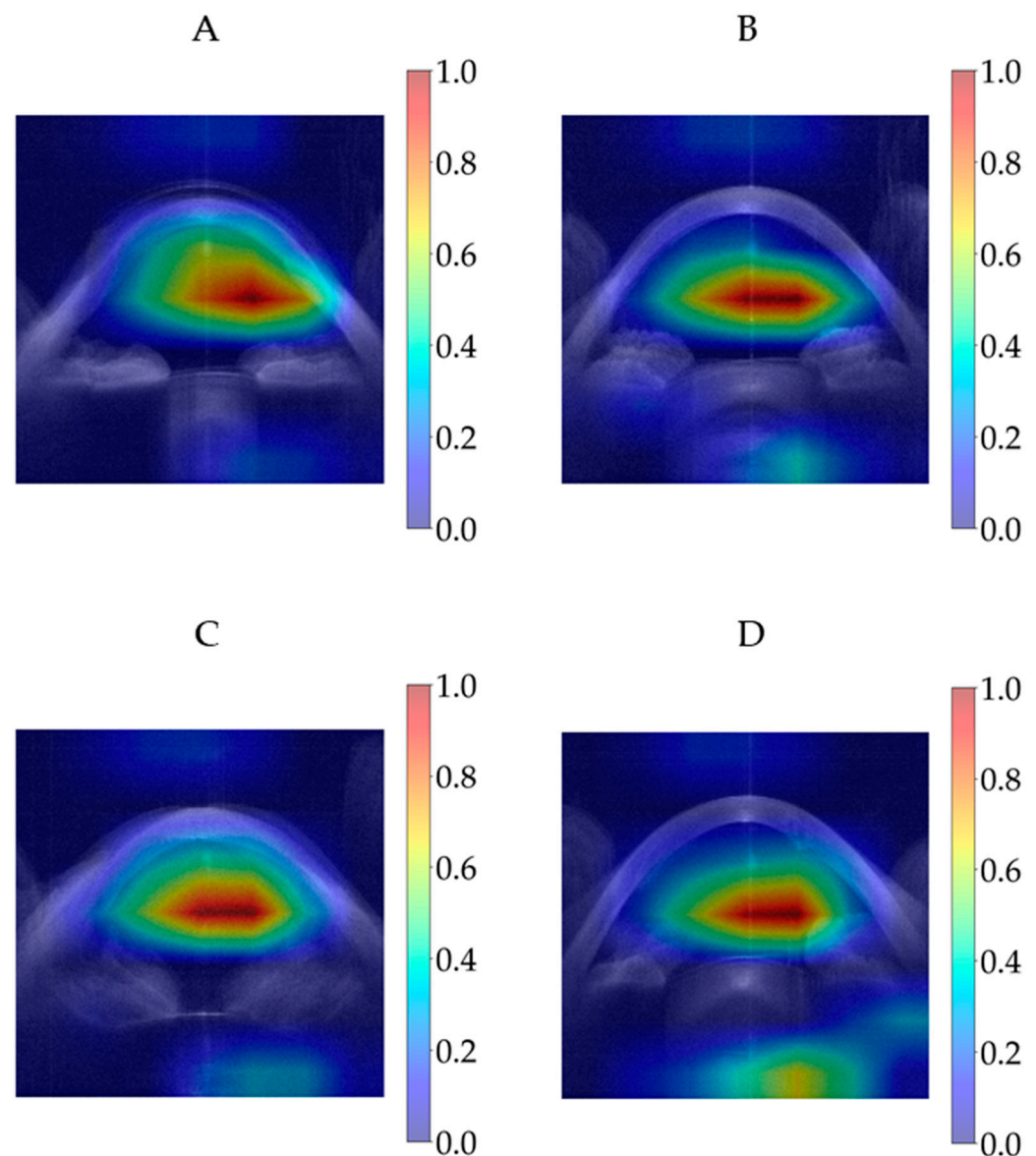
There were four eye examinations that were predicted as ‘normal’ by the modified MobileNetV3-Large but were diagnosed as ‘ectasia’ or ‘other disease’ by all three experienced ophthalmologists. For each of these four eye examinations, the first of sixteen equi-angular meridional images, with the aspect ratio adjusted to 1.629 (width divided by height) to better represent the realistic shape of the eye, is shown in Figure 11.



**Figure 11.** First of sixteen equi-angular meridional images, with the aspect ratio adjusted to 1.629 (width divided by height). (A) ‘other disease’-labelled but predicted as ‘normal’. (B) ‘other disease’-labelled but predicted as ‘normal’. (C) ‘other disease’-labelled but predicted as ‘normal’. (D) ‘ectasia’-labelled but predicted as ‘normal’.

It was determined that Figure 11A shows an eye examined while wearing a contact lens after penetrating keratoplasty; as the contact lens regularised the corneal surface, it was predicted as ‘normal’ by the modified MobileNetV3-Large. Figure 11B shows an eye that has undergone penetrating keratoplasty, which likely resulted in regularisation of the corneal surface, and therefore it was also predicted as ‘normal’ by the modified MobileNetV3-Large. Since the ‘other disease’ class contained eyes with penetrating keratoplasty, subepithelial or stromal scarring, corneal dystrophies, Salzmann’s nodules, or pterygia, the wide variety of

conditions within the 266 eye examinations may have limited the modified MobileNetV3-Large's ability to learn the differences among these diseases effectively. However, with a larger number of keratoplasty samples, it is possible that the model would recognise this condition better. Figure 11C was diagnosed as showing stromal corneal scarring and Descemet's folds indicative of endothelial dysfunction, along with vitreous prolapse into the anterior chamber. No explanation could be provided as to why this case was predicted as 'normal' by the modified MobileNetV3-Large. Figure 11D shows intrastromal corneal ring segments implanted surgically due to ectasia; the resulting regularisation of the corneal surface likely led to it being predicted as 'normal' by the modified MobileNetV3-Large. Figure 12 shows the Grad-CAMs of these four eye examinations generated by the modified MobileNetV3-Large.



**Figure 12.** Grad-CAMs of eye examinations generated by the modified MobileNetV3-Large. (A) 'other disease'-labelled but predicted as 'normal'. (B) 'other disease'-labelled but predicted as 'normal'. (C) 'other disease'-labelled but predicted as 'normal'. (D) 'ectasia'-labelled but predicted as 'normal'. Abbreviation: Grad-CAM = Gradient-Weighted Class Activation Mapping.

Quanchareonsap et al. [16] tested three AI models based on EfficientNet-B7 to differentiate between normal cornea, subclinical keratoconus and keratoconus using tomographic



maps from the Pentacam and corneal biomechanics from the Corvis ST. AI model 1, which used refractive maps from the Pentacam, achieved a macro-average accuracy of 93.6%, a macro-average sensitivity of 86%, a macro-average specificity of 95.7% and a macro-average PPV of 81.1%. For AI model 2, the dynamic corneal response and the Vinciguerra screening report from the Corvis ST were added to AI model 1. AI model 2 achieved a macro-average accuracy of 95.7%, a macro-average sensitivity of 73.7%, a macro-average specificity of 96.1% and a macro-average PPV of 95.3%. For AI model 3, the corneal biomechanical index was incorporated into AI model 2. AI model 3 achieved a macro-average accuracy of 95.7%, a macro-average sensitivity of 73.7%, a macro-average specificity of 96.1% and a macro-average PPV of 95.3%. Zhang et al. [17] used the CorNet model for the diagnosis of keratoconus using Corvis ST raw data. The dataset consisted of 1786 Corvis ST raw data samples, with 70% allocated to the training set and 30% to the validation set. The CorNet model achieved an accuracy of 92.13%, a sensitivity of 92.49%, a specificity of 91.54%, a PPV of 94.77% and a F1 score of 93.62%. Abdelmotaal et al. [18] developed a DenseNet121-based CNN model to distinguish between normal eyes and eyes with keratoconus using 734 Corvis ST videos from 734 eyes. The model achieved an accuracy of 89% on the test set, which comprised 30% of a dataset of 502 subjects, with the remaining 70% used for training and an accuracy of 88% on a separate dataset of 232 subjects. Fassbind et al. [19] employed preprocessed OCT data from a Casia2 device to diagnose corneal conditions, including healthy, keratoconus, post-laser, keratoglobus, pellucid marginal corneal degeneration, other and not appreciable, using the CorNeXt CNN model, which builds on the ConvNeXt architecture [36]. The model achieved a weighted-average accuracy of 93.52%, a weighted-average sensitivity of 84.30%, a weighted-average specificity of 99% and a weighted-average F1 score of 88.17%. For keratoconus detection specifically, it achieved an accuracy of 92.56%, a sensitivity of 84.07%, a specificity of 100% and an F1 score of 91.34%.

In comparison with [19], the macro-average F1 score of the modified MobileNetV3-Large achieved in this study (92.85%) exceeded the weighted-average F1 score of 88.17%, derived from the value reported in [19]. Moreover, the F1 score of the modified MobileNetV3-Large in this study for the ‘ectasia’ class (94.12%) is higher than the F1 score of 91.34% for the Keratoconus class, derived from the value reported in [19]. These findings indicate that using raw OCT data can outperform approaches based on preprocessed data in diagnosing corneal conditions.

Since the modified MobileNetV3-Large had higher overall accuracy than the other modified CNN models, it was selected for application in the GUI. The modified MobileNetV3-Large with a seed number of four achieved higher overall accuracy (94.44%) than the model with seed number one reported in Table 3 (92.86%); therefore, this model was used in the GUI. The GUI provides a classification of the corneal condition and the probability for each class. The user can select a raw data file in .3dv format and has the option to view the sixteen automatically scaled extracted images from the selected file, with the ability to zoom in on each image. This feature allows the user to compare the diagnosis provided by the GUI with what they observe in the images. The GUI helps ophthalmologists obtain a prediction from an eye examination taken with the Casia2 anterior-segment OCT and allows them to determine the corneal condition regardless of changes to the Casia2 software version.

However, this study has four limitations that must be considered. Firstly, only four CNN models were modified and tested, and it is possible that other models could achieve better results. Secondly, although 2737 eye examinations were decided to belong to one of the three classes, 1325 of these were decided to be ‘normal’ by all three experienced ophthalmologists. This represents approximately 73.49% (rounded to two decimal places) of the 1803 eye examinations for which all the three experienced ophthalmologists decided

on the same labels. To create a balanced test dataset, 42 eye examinations were selected for each class, and this corresponded to approximately 16.93% (rounded to two decimal places) of the total dataset. If the ‘ectasia’ and ‘other disease’ classes had more samples, it would have been possible to train the architecture and test a larger number of samples from these classes. Thirdly, mismatched cases between three experienced ophthalmologists were disregarded, which may have affected the generalisability of the findings to all patient populations. Fourthly, the dataset was monocentric. These limitations indicate areas for further research and refinement, which may affect overall accuracy.

## 5. Conclusions

In this study, raw OCT data were used to diagnose corneal conditions, including ‘normal’, ‘ectasia’ and ‘other disease’. Using raw data has advantages over preprocessed data, such as remaining unchanged with new software tools and updates and providing a more reliable foundation for analysis. The successful application of four CNN architectures with raw OCT data validates the use of raw OCT data for the diagnosis of corneal conditions in ophthalmology. Moreover, the GUI helps ophthalmologists obtain a prediction from an eye examination performed with the Casia2 anterior-segment OCT.

**Author Contributions:** Conceptualization, A.L. and M.M.; data labelling, A.D.A., E.F. and N.S.; methodology, M.M.; software, M.M. and M.S.; validation, A.L.; formal analysis, M.M.; investigation, M.M.; resources, A.L.; data curation, M.M.; writing—original draft preparation, M.M.; writing—review and editing, A.D.A., B.S., A.L., E.F. and N.S.; visualization, M.M.; supervision, A.L.; project administration, A.L.; funding acquisition, A.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** M.M. was supported in part by the Dr. Rolf M. Schwiete Stiftung under project no. 2020-024.

**Institutional Review Board Statement:** This study is a retrospective data analysis and has been exempted by the local institutional review board (Ärztchamber des Saarlandes, registration number 157/21) on 31 May 2021.

**Informed Consent Statement:** All data processed in this study were already anonymised at the source before being transferred to us for processing. This precludes any back-tracing of the identity, and therefore, informed consent of the patients was not necessary.

**Data Availability Statement:** The Python scripts and some data files are openly available from Zenodo at <https://doi.org/10.5281/zenodo.17711223>.

**Acknowledgments:** During the preparation of this manuscript, the corresponding author used ChatGPT (version 5-mini) for the purpose of text correction in British English. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
CNN	Convolutional Neural Network
CPU	Central Processing Unit
CSV	Comma-Separated Values
FP	False Positive
FN	False Negative
Grad-CAM	Gradient-Weighted Class Activation Mapping

GUI	Graphical User Interface
KC	Keratoconus
OCT	Optical Coherence Tomography
PNG	Portable Network Graphics
PPV	Positive Predictive Value
RAM	Random-Access Memory
RGB	Red–Green–Blue
TP	True Positive
TN	True Negative

## References

- Pang, G.; Wang, C.; Wang, X.; Li, X.; Meng, Q. A review of human cornea finite element modeling: Geometry modeling, constitutive modeling, and outlooks. *Front. Bioeng. Biotechnol.* **2024**, *12*, 1455027. [\[CrossRef\]](#) [\[PubMed\]](#)
- Santodomingo-Rubido, J.; Carracedo, G.; Suzaki, A.; Villa-Collar, C.; Vincent, S.J.; Wolffsohn, J.S. Keratoconus: An updated review. *Contact Lens Anterior Eye* **2022**, *45*, 101559. [\[CrossRef\]](#)
- Espandar, L.; Meyer, J. Keratoconus: Overview and update on treatment. *Middle East Afr. J. Ophthalmol.* **2010**, *17*, 15–20. [\[CrossRef\]](#)
- Singh, R.B.; Koh, S.; Sharma, N.; Woreta, F.A.; Hafezi, F.; Dua, H.S.; Jhanji, V. Keratoconus. *Nat. Rev. Dis. Primers* **2024**, *10*, 81. [\[CrossRef\]](#)
- Rabinowitz, Y.S. Keratoconus. *Surv. Ophthalmol.* **1998**, *42*, 297–319. [\[CrossRef\]](#)
- Martin, R. Cornea and anterior eye assessment with placido-disc keratometry, slit scanning evaluation topography and scheimpflug imaging tomography. *Indian J. Ophthalmol.* **2018**, *66*, 360–366. [\[CrossRef\]](#) [\[PubMed\]](#)
- Martin, R. Cornea and anterior eye assessment with slit lamp biomicroscopy, specular microscopy, confocal microscopy, and ultrasound biomicroscopy. *Indian J. Ophthalmol.* **2018**, *66*, 195–201. [\[CrossRef\]](#)
- Mohan, S.; Aggarwal, A.; Dada, T.; Vanathi, M.; Panda, A. Pachymetry: A review. *DOS Times* **2007**, *12*, 19–28.
- Seitz, B.; Behrens, A.; Langenbucher, A. Corneal topography. *Curr. Opin. Ophthalmol.* **1997**, *8*, 8–24. [\[CrossRef\]](#)
- Fan, R.; Chan, T.C.; Prakash, G.; Jhanji, V. Applications of corneal topography and tomography: A review. *Clin. Exp. Ophthalmol.* **2018**, *46*, 133–146. [\[CrossRef\]](#)
- Schmitt, J.M. Optical coherence tomography (oct): A review. *IEEE J. Sel. Top. Quantum Electron.* **2002**, *5*, 1205–1215. [\[CrossRef\]](#)
- Zhang, X.; Munir, S.Z.; Sami Karim, S.A.; Munir, W.M. A review of imaging modalities for detecting early keratoconus. *Eye* **2021**, *35*, 173–187. [\[CrossRef\]](#) [\[PubMed\]](#)
- Anton, N.; Doroftei, B.; Curteanu, S.; Catalin, L.; Ilie, O.D.; Tărcoveanu, F.; Bogdănici, C.M. Comprehensive review on the use of artificial intelligence in ophthalmology and future research directions. *Diagnostics* **2022**, *13*, 100. [\[CrossRef\]](#)
- Goodman, D.; Zhu, A.Y. Utility of artificial intelligence in the diagnosis and management of keratoconus: A systematic review. *Front. Ophthalmol.* **2024**, *4*, 1380701. [\[CrossRef\]](#) [\[PubMed\]](#)
- Maeda, N.; Klyce, S.D.; Smolek, M.K.; Thompson, H.W. Automated keratoconus screening with corneal topography analysis. *Investig. Ophthalmol. Vis. Sci.* **1994**, *35*, 2749–2757.
- Quanchareonsap, W.; Kasetuwan, N.; Reinprayoon, U.; Piyacomn, Y.; Wungcharoen, T.; Jermjutitham, M. Deep Learning Algorithm for Keratoconus Detection from Tomographic Maps and Corneal Biomechanics: A Diagnostic Study. *J. Curr. Ophthalmol.* **2024**, *36*, 46–53. [\[CrossRef\]](#)
- Zhang, P.; Yang, L.; Mao, Y.; Zhang, X.; Cheng, J.; Miao, Y.; Bao, F.; Chen, S.; Zheng, Q.; Wang, J. CorNet: Autonomous feature learning in raw Corvis ST data for keratoconus diagnosis via residual CNN approach. *Comput. Biol. Med.* **2024**, *172*, 108286. [\[CrossRef\]](#)
- Abdelmotaal, H.; Hazarbassanov, R.M.; Salouti, R.; Nowroozzadeh, M.H.; Taneri, S.; Al-Timemy, A.H.; Lavric, A.; Yousefi, S. Keratoconus Detection-based on Dynamic Corneal Deformation Videos Using Deep Learning. *Ophthalmol. Sci.* **2023**, *4*, 100380. [\[CrossRef\]](#)
- Fassbind, B.; Langenbucher, A.; Streich, A. Automated cornea diagnosis using deep convolutional neural networks based on cornea topography maps. *Sci. Rep.* **2023**, *13*, 6566. [\[CrossRef\]](#)
- Feng, R.; Xu, Z.; Zheng, X.; Hu, H.; Jin, X.; Chen, D.Z.; Yao, K.; Wu, J. Kernet: A novel deep learning approach for keratoconus and sub-clinical keratoconus detection based on raw data of the pentacam hr system. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 3898–3910. [\[CrossRef\]](#) [\[PubMed\]](#)
- Al-Timemy, A.H.; Mosa, Z.M.; Alyasseri, Z.; Lavric, A.; Lui, M.M.; Hazarbassanov, R.M.; Yousefi, S. A Hybrid Deep Learning Construct for Detecting Keratoconus From Corneal Maps. *Transl. Vis. Sci. Technol.* **2021**, *10*, 16. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hara, N.; Benedikt, K.; Owaki, H. CASIA2: Anterior Segment 3D Swept-Source OCT. In *Intraocular Lens Calculations*; Aramberri, J., Hoffer, K.J., Olsen, T., Savini, G., Shammas, H.J., Eds.; Essentials in Ophthalmology; Springer: Cham, Switzerland, 2024.

23. Li, Z.; Liu, F.; Yang, W.; Peng, S.; Zhou, J. A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 6999–7019. [\[CrossRef\]](#)
24. Taye, M.M. Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions. *Computation* **2023**, *11*, 52. [\[CrossRef\]](#)
25. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
26. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
27. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861. [\[CrossRef\]](#)
28. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
29. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. Pytorch: An imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 1–12.
30. Wang, Q.; Ma, Y.; Zhao, K.; Tian, Y. A comprehensive survey of loss functions in machine learning. *Ann. Data Sci.* **2022**, *9*, 187–212. [\[CrossRef\]](#)
31. Zhou, P.; Xie, X.; Lin, Z.; Yan, S. Towards understanding convergence and generalization of adamw. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 6486–6493. [\[CrossRef\]](#)
32. Krogh, A.; Hertz, J. A simple weight decay can improve generalization. *Adv. Neural Inf. Process. Syst.* **1991**, *4*, 950–957.
33. Ying, X. An overview of overfitting and its solutions. *J. Phys. Conf. Ser.* **2019**, *1168*, 022022. [\[CrossRef\]](#)
34. Hicks, S.A.; Strümke, I.; Thambawita, V.; Hammou, M.; Riegler, M.A.; Halvorsen, P.; Parasa, S. On evaluation metrics for medical applications of artificial intelligence. *Sci. Rep.* **2022**, *12*, 5979. [\[CrossRef\]](#)
35. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 618–626.
36. Liu, Z.; Mao, H.; Wu, C.Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A convnet for the 2020s. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 11976–11986.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.