# Assessment of a Vision-Based Technique for an Automatic Van Herick Measurement System

Tommaso Fedullo *Student Member, IEEE*, Davide Cassanelli *Student Member, IEEE*, Giovanni Gibertoni, Federico Tramarin *Member, IEEE*, Luciano Quaranta (MD, PhD), Ivano Riva, Lucia Tanga (MD), Francesco Oddone (MD, PhD), Luigi Rovati *Member, IEEE*

***Abstract*—The adoption of Artificial Intelligence methods within the instrumentation and measurements field is nowadays an attractive research area. On the one hand, making machines learn from data how to perform an activity, rather than hard code sequential instructions, is a convenient and effective solution in many modern research areas. On the other hand, AI allows for the compensation of inaccurate or not complete models of specific phenomena or systems. In this context, this paper investigates the possibility to exploit suitable Machine Learning techniques in a vision-based ophthalmic instrument to perform automatic Anterior Chamber Angle (ACA) measurements. In particular, two CNN–based networks have been identified to automatically classify acquired images and select the ones suitable for the Van–Herick procedure. Extensive clinical trials have been conducted by clinicians, from which a realistic and heterogeneous image dataset has been collected. The measurement accuracy of the proposed instrument is derived by extracting measures from the images of the aforementioned dataset, as well as the system performances have been assessed with respect to differences in patients' eye color. Currently, the ACA measurement procedure is performed manually by appropriately trained medical personnel. For this reason, Machine Learning and Vision–Based techniques may greatly improve both test objectiveness and diagnostic accessibility, by enabling an automatic measurement procedure.**

***Index Terms*—Artificial Intelligence, Machine Learning, CNN, Vision–Based Measurement, Van Herick, Computer Vision**

## I. Introduction

Glaucoma and other eye diseases are affecting more and more people in the last few years. Prevention is essential to avoid the progression of that disease, but in some cases, the screening exams are invasive or quite expensive, and it is not possible to periodically monitor the eye condition [1]. It has been demonstrated that people with a narrower Anterior Chamber Angle (ACA), i.e. the iridocorneal angle, are more vulnerable to the most aggressive form of glaucoma, the Primary Angle Closure Glaucoma (PACG) [2]. Nowadays, the gold technique used for ACA measurement is the gonioscopy, but it is invasive and requires high medical skills [3]. Among the various assessed techniques, one of the most interesting is the Van Herick (VH) maneuver, which exploits the correlation between the thickness of the cornea and the ACA [4]. The ratio between these two thicknesses represents the width of the ACA, hence allowing to detect the PACG. Van Herick approach requires a slit lamp, illuminating the limbus with a 60° angle between the light source and the eye optical axis [5]. Despite the simplicity of the approach, any PACG diagnosis derived from the estimation of the ACA with the van Herick technique must be performed by direct observation of an expert ophthalmologist with the help of a traditional slit lamp and a microscope. It appears then clear that, necessarily, this type of diagnosis is intrinsically subjective and results are strictly related to the ability and experience of the observing ophthalmologist [3].

In this paper, we discuss the adoption of suitable Machine Learning (ML) strategies to define a vision-based approach for the measurement of ACA. This novel Machine Vision (MV) method, which aims at overcoming the aforementioned subjective limitations, relies on a modified optical setup to perform the Van Herick measurement in a semi-automated way. In particular, the traditional slit light is obtained with an RGB LED-based digital light projector (DLP), whereas a compact CMOS camera acquires images of the human eye. The collected set of images is then classified by exploiting suitably designed MV techniques, thus identifying the specific images used for the automatic measurement of the ACA. That is, the presented test case perfectly fits the definition of Vision–Based Measurement (VBM) system, i.e. the application of MV to the instrumentation and measurement field [6]. The latter is actually a hot research topic. This article aim is to apply a solid and well–known ML algorithm to build a smart, accurate and intelligent measurement system, where the ML algorithm is evaluated through suitable metrological metrics.

This manuscript represents a technical extension of our preliminary study [7]. In particular, for the sake of a thorough and comprehensive assessment of the proposed measurement instrument, a new and realistic dataset has been acquired on a significant set of patients during clinical trials, carried out at IRCCS Fondazione G.B. Bietti Rome, Italy. Furthermore,

Author Tommaso Fedullo (tommaso.fedullo@unimore.it) is both with the Department of Management and Engineering of the University of Padova, Stradella S. Nicola 3 - 36100 Vicenza, Italy and the Department of Engineering "Enzo Ferrari" of the University of Modena and Reggio Emilia, Via Vivarelli 10, 41125, Modena, Italy. Authors Davide Cassanelli (davide.cassanelli@unimore.it), Giovanni Gibertoni (giovanni.gibertoni@unimore.it), Federico Tramarin (federico.tramarin@unimore.it) and Luigi Rovati (luigi.rovati@unimore.it) are with the Department of Engineering "Enzo Ferrari", University of Modena and Reggio Emilia, Via Vivarelli 10, 41125, Modena, Italy. Author Luciano Quaranta (lquaranta@centrooculisticoitaliano.it) is with the Centro Oculistico Italiano, Brescia, Italy. Author Ivano Riva (ivano.riva@virgilio.it) is with the Istituto Clinico Sant'Anna, Brescia, Italy. Authors Lucia Tanga (lucia.tanga@fondazionebietti.it) and Francesco Oddone (oddonef@gmail.com) are with the IRCCS Fondazione Bietti, Rome, Italy.

Manuscript received XXX, XXX. This work is part of the project n. 1278 "Development of a new non-contact screening method and instrument for the detection of narrow ocular anterior chamber angle" funded by Velux Stiftung foundation (ZÜRICH, SWITZERLAND). This work was funded in part by the UniMoRe – FAR Mission Oriented 2021 Grant.

the performance analysis has been extended to consider the effect of the eye colors and the impact in terms of measurement accuracy. Finally, the manuscript also takes into account several Machine Learning techniques, discussing a preliminary comparison of their performance in the considered application, and then specifically addressing the implementation of the instrument using the two most promising ones.

This paper is organized as follows. First, in Section II we present the current state of advancement and the technique already presented in the literature for objectively evaluating Van Herick grading on slit lamp based pictures. In Section III we sketch the experimental setup and the main goals of this research activity. Moreover, Section IV presents the specific ML techniques used and the evaluation techniques. Specifically, Section IV-A, presents a comparison between different ML techniques. CNN performances are presented in Section V. Preliminary results obtained from experiments in our laboratory are presented in Section VI-B while in Section VI-C we present results relative to a new dataset collected with our instrument by expert clinical personnel. The Paper concludes, after the presentation of the future research activities, in Section VII.

## II. State of Art

As introduced in the previous section, the Van Herick technique is a very promising approach for non-invasive ACA measurement [3]. Despite its advantages with respect to other techniques, some specific limitations of the Van Herick maneuver have narrowed its usage. In recent years, many studies have been carried out to improve the objectiveness and accuracy of the VH grading technique [8], [9]. Foster *et al.* [10] proposed a new grading scale system to improve the sensitivity of the evaluation. According to this work, the new scale reduces the possibility of wrongly diagnosing "open-angle" cases with a close angle condition. Moreover, Sihota *et.al* [11] introduced a modified grading system. However, in order to use these new grading systems, the opticians are required to have acquired a good level of training.

To overcome the intrinsic subjectivity of this approach, an automatic measurement of the ACA could be performed. Theeraworn *et al.* [12], [13] present a Support Vector Machine algorithm able to extract the measurement of the ACA from slit-lamp images. This approach is very promising, but it still requires the presence of an operator that correctly sets the slit lamp and performs the maneuver. In fact, the algorithm just extracts the information from the images. Other automatic procedures have been proposed over the years. However, such works foresee the usage of other techniques such as gonioscopy [14] or OCT [15], even exploiting Machine Learning techniques. In the first case, the technique is invasive while in the second the procedure requires an OCT machine, which is expensive, thus being not suitable for continuous screening. Moreover, Shimizu *et al.* have recently presented a portable device that can be equipped on a smartphone, whose aim is to replicate slit-lamp microscope system behavior in a cheaper and more compact design [16]. However, the analysis of the collected image is still subjective and relies on expert medical opticians' evaluation.

Differently, our work is focused on the development of an instrument that automatically and objectively performs the VH procedure in a cheap, compact, and portable system. The proposed automatic Vision–Based Measurement System, exploits Machine Learning techniques for image classification, whose behavior is analyzed from a metrological perspective. Machine Learning techniques, in literature, have been widely applied to VBM systems [17], especially CNN–based ones. Indeed, CNNs reviled their suitability for VBM systems in many different applications [18]. For example, the authors [19] proposed a CNN-based system for railway network inspection. Differently, [20] presented a method for head pose estimation in vehicles, demonstrating high versatility in the application of CNNs. Furthermore, an interesting review paper [21] summarizes the most widespread Deep Learning (DL) techniques for Image Classification, with regard to ophthalmic applications. They identified several CNN network models that proved to be promising such as, among others, AlexNet [22], GoogleNet, [23] and ResNet [24]. These networks showed the best results in the ImageNet Large Scale Visual Recognition Challenge [25] and are all based on Convolutional Layers.

## III. Experimental Setup and Research Goals

The Van Herick procedure used to measure the ACA must be accomplished with specific alignment constraints [3]. It has been shown by Leung et al. [26] that both illumination and observation angles affect the ACA openness assessment. As a consequence, particular attention was paid during the realization of the optical setup for such an experimental evaluation. A schematic diagram of the optical setup used to perform the Van Herick measurement is shown in Fig. 1.

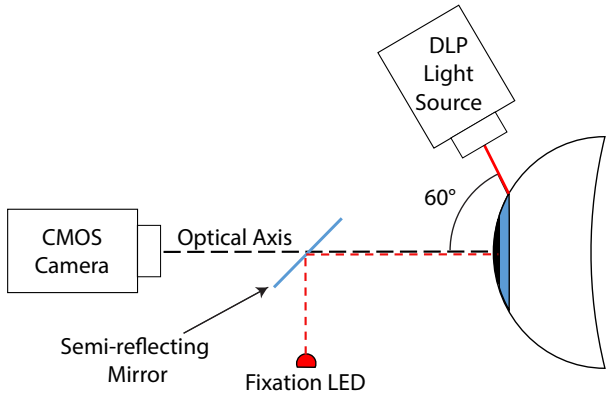



Figure 1. Schematic diagram of the optical setup.

The optical setup consists of two main devices. Firstly, a digital CMOS camera Basler Dart (daA1600-60uc S-Mount, Basler© AG, Ahrensburg, Germany) is positioned in front of the analyzed eye, aligned with its optical axis. A 16 mm focal length lens (Evetar Lens M12B1618IRM12 F1.8) is also used in the optical setup. Secondly, a LED DLP (DLP2010EVM-LC, Texas Instruments©, Dallas, Texas, U.S.) was used as the illumination unit, instead of the traditional slit light. The DLP relies on modern micro-mirror technology to project structured light onto a specific target. The illumination unit was placed at a 60-degree rotation angle with respect to the eye and camera optical axis as shown in Fig. 1. Moreover, the DLP

has been configured to emit a uniform red slit light that scans the whole surface of the eye under examination. Thus, during the measurement procedure, the emitted slit light scans the eye from the external corner of the sclera towards the nose. This scan takes place during the frames acquisition with the digital CMOS camera. A fixation target, i.e. a small light pointer, was placed on the eye-camera optical axis, through a 45° semi-reflecting mirror, to help the patient to look straight ahead. The patient's head is then placed on a chin rest to guarantee the steadiness and alignment to the optical setup. A single measurement procedure has been designed to perform two entire scans of the eye in 4 seconds, allowing for a total of 120 raw pictures to be collected at 30 fps. The entire system is connected to a Windows® based embedded computer (Lattepanda Alpha 864s) with an Intel® Core™ m3-8100Y processor unit. The computer runs a Python-based graphics user interface (GUI) that is in charge of peripheral control (Camera and DLP) as well as real-time image processing. Actually, the set of images previously acquired by the camera are then processed by the Machine Learning algorithm, aiming at the identification of the central images.

As previously mentioned, the Van Herick technique foresees the comparison of the depth of the peripheral anterior chamber with the thickness of the cornea. This measure can be derived when a narrow slit of light shines within the limbus, i.e. the edge between the cornea and the sclera. As a consequence, within the entire set of images acquired during the scan, only a few of these (referred to as *central* images) can be used to measure the ACA. Both previous and subsequent acquisitions, where the light is placed respectively on the left and on the right of the limbus, must be discarded. Indeed, since the patient eye position may change between different measurements, an a–priori images selection can not be performed. Consequently, a wider area must be scanned.

The outcome of a single measurement, that is, the dataset coming from the acquisition system consists of two scans of the entire eye, each one composed of 60 images with a $1600x1200$ px resolution. As an example, a set of images acquired during a scan is represented in Fig. 2.

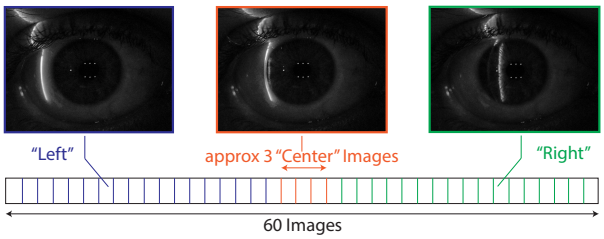



Figure 2. Example of a set of images acquired during a scan.

The approach we followed in this work is based on a Convolutional Neural Network (CNN), solving a three-class image classification problem. The CNN takes as inputs the images acquired by the setup and performs the classification activity by tagging the images as *left*, *central* or *right* respectively labeled with 0, 1, 2.

In the following, after a comparative analysis among different ML approaches, we identified the two CNNs that revealed most suitable to solve the classification problem. Both of them have been then specifically analyzed, through an objective approach presented in Sec. VI. As a last remark, as outlined in Section II, to the best of our knowledge there are no other contributions in the literature addressing an automatic procedure to perform the Van Herick maneuver, whereas other works focus on the of ML techniques to process images acquired with the standard manual Van Herick technique. As a result, it is not possible to identify a literature baseline, with which to compare our approach.

For this reason, in this work, the behavior of the classification algorithm is evaluated by means of two different approaches:

1) We firstly used classical Machine Learning metrics, like accuracy, precision and recall to evaluate the network behaviour and choose the best model. These tasks have been addressed in Sections IV and V.
2) Secondly, we have tested the developed network by means of a metrological evaluation strategy, described and presented in Section VI. It is worth observing that the test data used to carry out this last evaluation have not been used for the network training, neither for the accuracy, precision and recall evaluations.

## IV. ML–based classification Technique

In the following, the Neural Network design is specifically addressed. At first, we choose the best ML techniques for the considered application, by a comparative approach. An investigation of the adopted CNNs is then proposed.

### A. Typical metrics and ML Algorithm Comparison

Typically, a first evaluation of the network performance includes the analysis of the accuracy. This is simply calculated as the percentage of correctly labeled data ($y_t$) over the total observations ($y_t + y_f$):

$$accuracy(\%) = \frac{y_t}{y_t + y_f} * 100. \tag{1}$$

Accuracy, however, becomes a not representative metric if the validation dataset is unbalanced. Among all the *central* predictions obtained by the network (Total Positive, TotP), a group of them (True Positive, TP) is well predicted while others are not (False Positive, FP). Moreover, it is possible that the network wrongly labels some *central* images as *right* or *left*, namely False Negative (FN). Given these definitions, precision and recall are defined as:

$$Recall(\%) = \frac{TP}{TP + FN} * 100. \tag{2}$$

$$Precision(\%) = \frac{TP}{TP + FP} * 100, \tag{3}$$

Recall indicates the network ability to mark all the real *Central* images as *Central*. On the other hand, precision represents the network capability not to label *left* and *right* images as *central*. If the number of False Positives (i.e. left and right images marked as central) decreases, precision increases.

In general, it would be important to obtain high values of both recall and precision, but for the application considered

in this work, the precision is more important. Indeed, if some *left* or *right* images are wrongly interpreted as central, the measurement outcome taken from those images becomes intrinsically less accurate. Considering the aforementioned metrics, different ML techniques have been tested and evaluated to identify the most suitable one for the application at hand. As underlined in Section II, image classification problems can be solved by exploiting CNN-based architectures. To this aim, several networks structures have been trained, since the level of complexity of the given problem does not preclude the selection of any specific model, nor imposes an *a priori* choice of one model with respect to another. Therefore, we tested and compared the classification performance of the most commonly used CNN structures, as well as of a linear *Space Vector Machine* classifier (SVM) and of a *K-means* unsupervised classification algorithm.

In Table I the results of the performance comparison among the aforementioned techniques are reported. Namely, the performance of the different machine learning approaches over a *test data-set*, composed by 8677 non-augmented images, have been assessed in terms of validation accuracy (VAL-ACC), *center* class prediction recall (CTR-REC) and precision (CTR-PR), single frame prediction time (SFPD) and the respective standard deviation over the whole test data set. It is worth observing that, the *test data-set* used for the comparison has not been used before during the training phase. Moreover, the total number of parameters of each CNN, regarding convolutional layers only, have been reported to give an idea of overall networks complexity.

Table I
CLASSIFICATION METHOD COMPARISON

| Method | VAL-ACC | *CTR-PR* | *CTR-REC* | SFPD (ms) | $\sigma$ SFPD (ms) | Parameters |
|---|---|---|---|---|---|---|
| **AlexNet** | 98.82 % | 86.04 % | 86.04 % | 29.98 | 3.43 | 6088768 |
| **VGG16** | 99.02 % | 88.44 % | 87.18 % | 33.42 | 3.53 | 14713536 |
| **ResNet50** | 98.29 % | 79.56 % | 82.05 % | 39.94 | 6.28 | 23581440 |
| **SVM** | 92.90 % | 26.92 % | 7.98 % | 0.39 | 0.21 | - |
| **K-means** | 59.31 % | 3.73 % | 12.82 % | 0.71 | 0.25 | - |

Among the different ML techniques, CNNs show a superior performance in terms of accuracy if compared to both the linear SVM classifier and the unsupervised *k*-means approaches. It is worth noting that the SVM accuracy is quite high while prediction of *center* images is not equally acceptable. Indeed, the linear classifier can correctly predict most of the *left* or *right* images, while confusing the central images. The same can be said for the k-means ML algorithm, but with even worst performances: In this case, the unsupervised method failed to recognize a common pattern among the presented training images, hence resulting in a significantly lower performance in the validation phase.

From Table I, AlexNet and VGG16 [27] are the networks providing the best performance in terms of precision and recall, and in general with respect to all the paramenters. As a general consideration, the medical operator needs to review the images classified as 'central' just after the procedure, possibly repeating the measurement if the acquisition was not successful. Besides the accuracy of the classification, also the overall time taken by the CNN to label all the acquired images is an important performance index. From Table I Alexnet revealed a lower complexity in terms of total parameters (last column of Table I) and showed to be slightly faster than VGG16, that is however able to provide better values of precision and recall. For these reasons, both will be considered in the following analysis, and as potential candidates for the implementation within the proposed VBM.

### B. Convolutional Neural Network Design

As described in the previous section, both AlexNet [22] and VGG-16 [27] structures can be finally chosen for this application. Different Alexnet and VGG-16 settings have been tested during an extensive experimental campaign, where a trial and error methodology has been adopted. In particular, several typical parameters, such as precision and recall, have been used to optimize the training activity. Afterwards, the best training configuration for our application has been chosen and it is presented in this section. Moreover, as the adopted method is a supervised ML technique, we took particular care of the labelling process. The collected images have been carefully divided into the three classes, by exploiting the guidelines of expert ophthalmologists. *Relu* has been adopted, both for AlexNet and VGG-16, as it is a commonly used activation function [28], that is in charge to manage the Input-Output behavior of neurons. Moreover, in this situation, it achieves excellent performances and it demonstrated to learn faster. AlexNet structure, as typical CNNs, is formed by both Convolutional and Pooling layers, where the latter are placed between each convolutional layer and the subsequent one. Specific image patterns are identified by means of a Kernel slicing on the entire layer. Afterwards, a common practice consists in the usage of a pooling layer, that downsamples the input patterns, aiming to increase the robustness of the network to slight variations of the detected features. The extraction of such meaningful features is done taking the maximum from each kernel acquisition, namely *max pooling*. Finally, the last fully connected layers and the *softmax activation* compute the probability of each image to belong to each of the 3 classes. A slight modification has been made in respect to the typical AlexNet structure: we used a bilinear interpolation algorithm to obtain resized $400x300$ px input images. It is worth noting that downsized images are used only for classification purposes, while the ACA measurement will be performed with native resolution ones. Indeed, there are no network performance improvements when using the typical $227x227$ px AlexNet input size. Likely, an higher quality of the image is, for our purposes, more important than a fine tuning of layers and kernel sizes. Fig. 3 shows the network structure we used.

As it is possible to see from Fig. 3, the dropout technique, which consists on the deactivation of some randomly chosen neurons, is used to reduce inter–dependency between fully connected layers, thus preventing over–fitting. The Dropout Rate has been set to 50%.

Analogously, Fig. 4 represents the VGG-16 architecture, which is basically formed by the same components already discussed for the AlexNet one. Actually, the difference resides in the number and dimensions of the various layers.

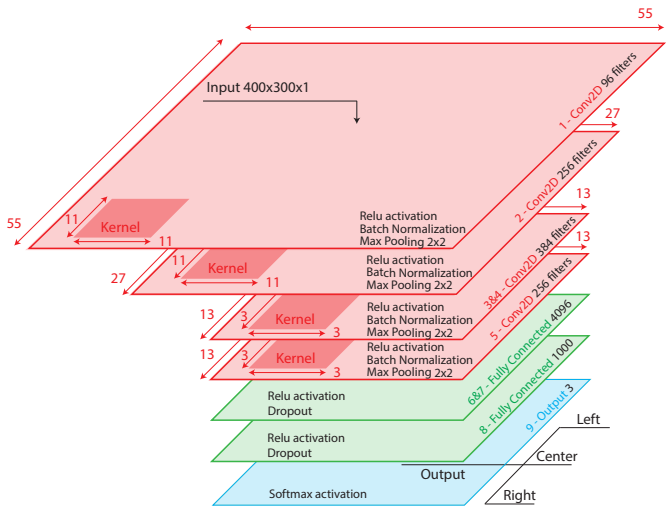



Figure 3. AlexNet [22] structure.

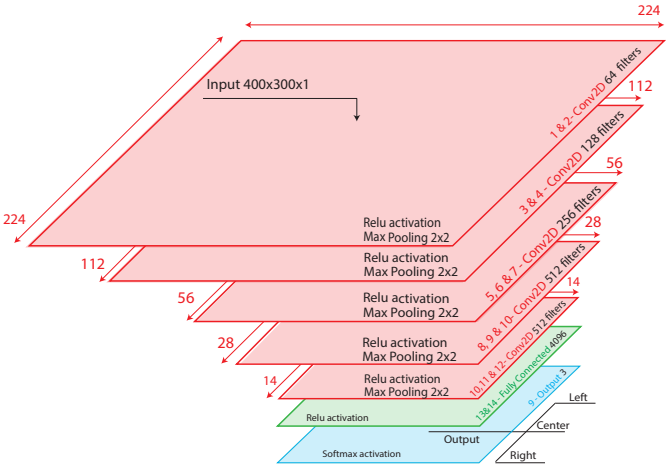



Figure 4. VGG - 16 [27] structure.

Data preparation is a fundamental task to be done before the training stage. Hundreds of acquisitions were made and the data have been manually split into the three classes to train the network. Data are normalized between 0 and 1, and then 70% of the total have been used to train the network, while the other portion for validation purposes. One important feature of the available dataset is the low number of *central* images (i.e. those ones to be identified with the network), usually three or four out of 120 images acquired during the measurement procedure. To over the hump, *data augmentation techniques* have been adopted. Data augmentation usually foresees the generation of modified training examples, applying different and pre–tuned image transformations, e.g. horizontal and vertical shifts, rotations and brightness modification, only to name a few. This technique is typically used to generalize the behavior of the network, trying to generate new and meaningful examples targeted for the application. Instead, we used data augmentation also to generate more *central* images, this way giving the possibility to train the network with comparable amounts of *left*, *right* and *central* images. The data augmentation technique has been applied to the already labelled images, to ensure that the augmented images are correctly labelled. In Table II are listed our data augmentation parameters.

Table II
DATA AUGMENTATION SETUP PARAMETERS.

| | |
|---|---|
| Horizontal Shift Range (% of the width) | (−1%, 1%) |
| Vertical Shift Range (% of the height) | (−1%, 1%) |
| Rotation Range (deg) | (−7, 7) |
| Brightness Multiplication Factor Range | (0.5 – 1.75) |
| Zoom Range (% of the picture size) | (90% – 110%) |

Indeed, little movements to the right and to the left, rotations and zooms may occur. Furthermore, different brightness levels could take place in different surrounding environments. It is worth observing that we generated more central images than the other two classes, aiming to increase the precision of the classification of the *central* images. After the data preparation stage, the prepared data set had the properties listed in Table III

Table III
TRAIN DATA SET PROPERTIES.

| | Original | Original Training | Augmented | Total Training |
|---|---|---|---|---|
| Left | 8537 | 5976 | 5711 | 11637 |
| Central | 593 | 416 | 14994 | 15410 |
| Right | 7877 | 5513 | 5471 | 10984 |

In Table III, as said before, the original images have been splitted in validation and training data, being respectively the 30% and 70% of the total. Afterwards, only the training images have been augmented.

## V. NEURAL NETWORK EVALUATION

The CNN *AlexNet* and *VGG-16* described in Section IV-B have been implemented in Python, within the *Keras* (Tensorflow Version 2.1.0) framework and trained according to the set–up described in Table IV. Clearly, as the hyper-parameters have been chosen after a trial-and-error stage, the final settings can be slightly different for each CNN.

Table IV
CNN TRAINING SETTINGS.

| | **AlexNet** | **VGG-16** |
|---|---|---|
| Optimizer | SGD | Adam |
| Learning Rate (LR) | $10^{-2}$ | $2 \cdot 10^{-4}$ |
| Epochs | 75 | 150 |
| Batch Number (BN) | 128 | 332 |
| Loss Function | SCC | SCC |
| Validation Data (% of original images) | 30 | 30 |

At each training stage, the error gradient was calculated through the Stochastic Gradient Descent (SGD) algorithm for Alexnet, and by exploiting the Adam optimizer [29] for the VGG-16 approach. By controlling the error gradient the optimizer updated the weights so that the error decreased step by step. The next weights choice was done by evaluating the *error cost*, through a specific loss function. Given the training settings of Table IV, considering the n–th observation was labeled as a specific class $c$ with probability $p_{n,c}$ and N was the total number of observations, the Sparse Categorical Crossentropy (SCC) loss function was expressed as:

$$Loss(p_{n,c}) = -\frac{1}{N}\sum_{n=1}^{N}[log(p_{n,c})]. \quad (4)$$

In this context, the Learning Rate (LR), which represents how quickly the network learns, was surely a key parameter, that has been chosen after several experimental tests. Weights were updated at each step proportionally to the LR and to the calculated error. Both training and validation data have been divided into BN chunks that were individually loaded in memory to increase computational efficiency.

It is worth noting that in our test case the train dataset was used several times to feed the network, 75 and 150 Epochs for AlexNet and VGG-16, respectively. The training has been stopped when the loss and accuracy did not improve for several consecutive epochs, to avoid overfitting. Results are encouraging since it is possible to achieve high accuracy values, already presented in Table I. Despite this, a more accurate analysis is needed, and a deeper evaluation is now conducted by means of different metrics. Indeed, the objective of the network is to identify images belonging to the *center* class, that has very low validation examples. For this reason, high accuracy values can be obtained also if the *central* class is not well predicted. Moreover, there is an impact of the threshold value on the precision. The threshold is defined as the probability level above which an observation is labeled as *center*. The resulting recall and precision curves obtained are shown in Fig. 5 and Fig. 6 for AlexNet and VGG-16, respectively.

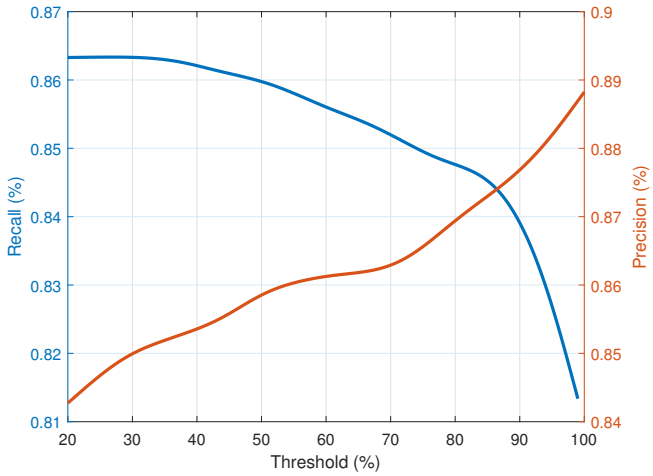



Figure 5. Precision and Recall: AlexNet

An evident effect is that, as the threshold increases, recall decreases with increasing precision. Indeed, a threshold increase reflects on a lower number of total positive, i.e. $TotP = TP + FP$. The growth of the precision is a clear indication that the total positive decrease goes together with a decrease of false positives that mostly become true negatives. The recall decrease is slighter than the precision increase. This allows choosing a high value of the threshold to increase the precision of the whole network. It is worth observing that all the reasoning is made on the global number of test sequences, but high values of the threshold may involve in a small number

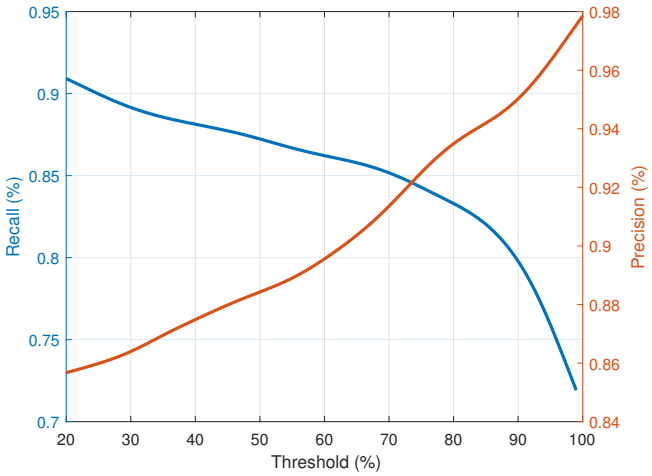



Figure 6. Precision and Recall: VGG-16

of total positive for a single data–set. For this reason, in this work the chosen threshold is 89%, allowing to achieve a precision much greater than the 80% for both the CNNs.

The trained network can be finally used to predict *central* images on the newly acquired data set. The classification algorithm can be run in the described embedded system, with additional few seconds of execution that does not increase the acquisition time, as it is performed after the scans are completed. This is surely an important aspect to underline, as it is demonstrated that patient's ability to look steady toward the fixation target, is time-limited.

## VI. Test results

The performance of the networks and the instrument were tested by exploiting the evaluation technique presented in Section VI-A. The images were collected in grey scale by extracting the red channel from the original images, to emphasize the light line. In distinction, the reference image was collected in RGB to simplify the iris recognition. In the following, results obtained by two different datasets are presented: i) the preliminary results with the old dataset, already presented in our previous work [7], and ii) the clinical results obtained with the new dataset, collected by expert ophthalmologists at IRCCS Fondazione G.B.Bietti. For both datasets written informed consent was obtained from all patients via an agreement document.

The original dataset was acquired by the authors at OptoLab, Department of Engineering "Enzo Ferrari", UNIMORE, Modena, Italy. Six different subjects have been tested five times, for a total of 60 eye scans (i.e. 30 measurements). The number of *center* images chosen by the network ranged from 3 to 5 per each measurement. The total amount of analyzed images was 125. The six subjects on which the measurements were carried out were chosen with different colors of the iris to exclude this variable as a likely possible source of error in the validation phase.

The new acquired dataset, instead, has been collected on a total of 36 different subjects by expert ophthalmologists

in a clinical environment. As before, the number of *center* images chosen by the network ranged from 3 to 5 per each measurement. The total amount of analyzed central images was 140. In this work, the preliminary results have been compared with the new ones and an analysis of the influence of the iris pigmentation on the networks performances is presented. It is also worth observing that, in the following, all the histograms reporting the resulting error between the light line and limbus position have been derived by using the AlexNet network. Those related to the use of the VGG network have not been presented, as they are not informative and the comparison between the two networks has been conducted only by analyzing the mean and standard deviation values of the considered error.

### A. Metrological Evaluation Technique

The goodness of the CNNs performance is evaluated, from a metrological point of view, through the study of the difference in pixels between the position of the limbus and the position of the light line on the eye surface: the lower this difference the higher the CNN accuracy. This procedure has been divided into 4 steps: i) recognition of the iris from a reference image to have a reference for the intensity profile extraction; ii) extraction of the limbus position from a reference image captured at the first instant of the measurement procedure; iii) evaluation of the line position for every image labeled as *central*; iv) computation of the difference between these two positions. It is worth observing that all the procedure is conducted in a semi-automatic way, thus revealing to be not directly applicable to identify the center images. Indeed, the aim is to build a totally automatic VBM system.

The extraction of the limbus position is performed by considering the limbus intensity profile, that can be fitted as an error function such as the one in Eq.(5), where A, B, C and D are fitting parameters that change for each reference image. Then, the maximum of the fit function derivative, Eq.(6), is computed to represent the limbus position, $X_{limbus}$.

$$f(x) = A + B\frac{2}{\sqrt{\pi}}\int_0^{D(x-C)} e^{-t^2}dt \tag{5}$$

$$f'(x) = D \cdot B\frac{2}{\sqrt{\pi}}e^{-D^2(x-C)^2} \tag{6}$$

Conversely, since the light line intensity profile can not be easily fitted by a standard distribution, the $X_{line}$ position of the light line is estimated as in [30] [31]:

$$X_{line} = \frac{\sum_i I_i x_i}{\sum_i I_i}, \tag{7}$$

where $x_i$ is the single–pixel x-coordinates, while $I_i$ is the single pixel intensity. The error between the limbus position $X_{limbus}$ and the line position $X_{line}$ is finally computed as:

$$E = X_{limbus} - X_{line}. \tag{8}$$

The value of $E$ can be positive or negative (i.e. left or right, respectively) based on the light line position with respect to the limbus on the images selected by the neural network.

The error E measurement is evaluated through a statistical analysis with a histogram of the error occurrences, observing its center of gravity and standard deviation.

### B. Preliminary Results

In the following, the results obtained with the dataset collected in the OptoLab Laboratory are presented. A typical reference image, from which the limbus position $X_{Limbus}$ is extracted, is depicted in Fig. 7(a). In particular, a reference image is collected for each measurement. Furthermore, in Fig. 7(b) the limbus intensity profile, extracted in correspondence to the eye center y-coordinate, is plotted along with its fitting as an error function (orange solid line) and its derivative (light green solid line). The maximum of the derivative represents the limbus position. In Fig. 8 a typical *center* image used to extract light line position is shown. After retrieving the position of the Limbus $X_{Limbus}$ and of the light line $X_{Line}$, exploiting Eq.8 the displacement $E$ is computed.

In Fig. 9, the histogram of the error $E$ occurrences, for the AlexNet network, is shown. As it is possible to see, the histogram has a Gaussian-like behavior with parameters: $\mu$ = 5.99 px and $\sigma$ = 5.92 px. By exploiting the same technique with the VGG-16 approach, it is possible to derive $\mu$ = 6.27 px and $\sigma$ = 6.3 px. Actually, VGG-16 results in a higher error and standard deviation. Despite this, as the values are greatly lower than the light line thickness, both networks can not perform better, and both errors are totally negligible.

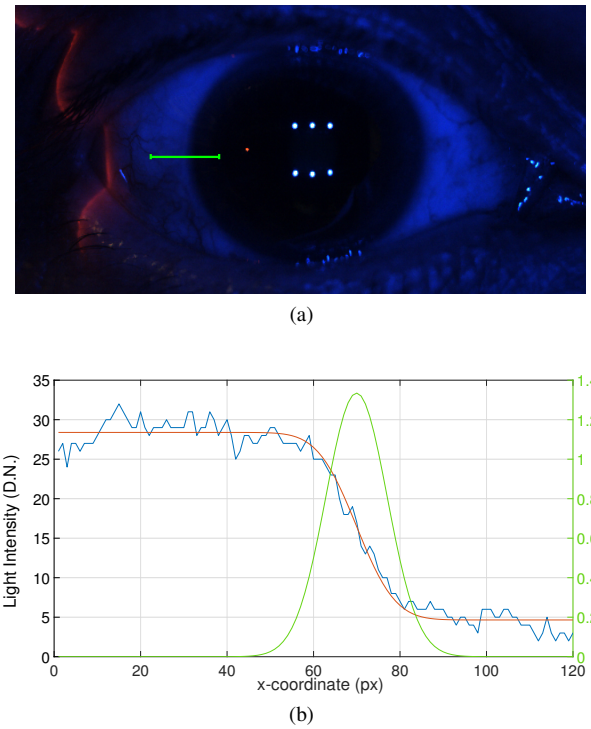



Figure 7. Typical reference image used to estimate the limbus position (a); the corresponding intensity profile along the y-coordinate of the eye center (•) (green line in (a)), the fitting error function (•) and, its derivative (•) (b).

### C. Clinical Results

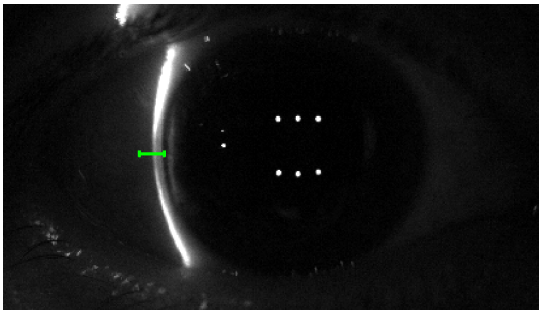


Figure 8. Typical *center* image used to estimate the light line position; The intensity profile of the line is extracted along the y-coordinate of the eye center (green line).

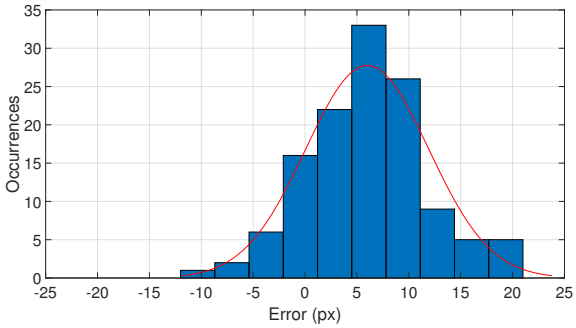



Figure 9. Histogram of occurrences of the error between the limbus and line position in pixel considering the preliminary dataset with the AlexNet network. The (•) line represents the Gaussian fitting of the histogram.

To evaluate the influence of different iris pigmentation on network capability to extract *center* images, a new dataset has been collected on a total of 36 volunteers by specialized ophthalmologists. It is worth noting that, for each volunteer, a different number of measurements have been performed, as to monitor the progression of the glaucoma several measures are needed. The dataset includes eyes of four different colors: brown, blue, black, and green. The exploitation of this novel dataset, acquired by professionals, is surely of fundamental importance to analyze the accuracy of the whole system. The number of measurements comprised in this new dataset for each color is presented in Tab. V. It is worth observing that, for each measurement, the number of *central* images can be different, so that the analysis was conducted by using 35 images for each iris pigmentation. More precisely, the balanced dataset has been obtained by randomly removing a few images for some eye colors.

Table V
NEW DATASET PROPERTIES.

| Color | Number of Measurements |
|---|---|
| Black | 14 |
| Brown | 13 |
| Green | 11 |
| Blue | 12 |

As before, the error between the limbus and the line position, in pixel, is evaluated. The error distribution histogram of the complete dataset, as well as the Gaussian line-fit, is depicted in Fig. 10, considering the AlexNet network. The mean experimental error is $\mu = 4.12$ with standard deviation $\sigma = 5.67$, as also summarized in Tab. VI, where a comparison with the VGG-16 network has been conducted.

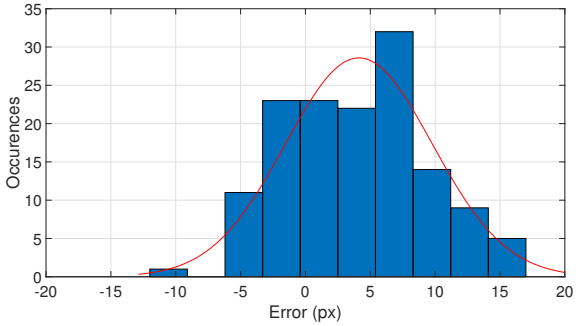



Figure 10. Histogram of occurrences of the error between the limbus and line position in pixel for the new dataset, exploiting the AlexNet network. The (•) line represents the Gaussian fitting of the histogram.

By comparing the results obtained from the two datasets, respectively of Fig. 9 and 10 it is possible to notice that both revealed similar standard deviations (slightly lower for the new dataset). This result is totally reasonable as the two datasets present similar number of statistical samples (slightly higher for the new dataset). Moreover, it is possible to underline that in both situations the mean error is similar, but always positive. Indeed, we decided, by suitably labeling the training data, to instruct the network to choose central images which present the line slightly inside the iris, rather than slightly outside. Actually, when the line is somewhat inside the iris, it is still possible to see the refraction of the light, and derive the Van Herick grade. To evaluate the iris pigmentation influence we extract the error distribution histogram for each color, as shown in Fig. 11, that refers to the AlexNet network. The same reasoning can be made referring to the results obtained with the VGG-16 network, presented in Tab. VI. Moreover, the VGG-16, as before, still results in a higher error and standard deviation. As said before, this comparison still depends on the specific hardware used (in particular the light line thickness) which makes substantially this error totally negligible.

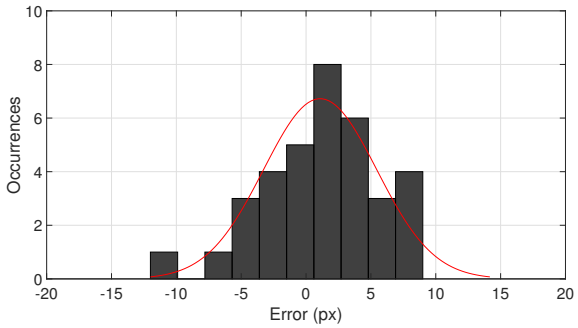
According to the Table VI, the best results have been obtained for black and brown eyes. Differently, green and blue eyes show a slightly higher mean error and standard deviations. This trend is reasonable since we used red-colored light line.

Table VI
MEAN AND STANDARD DEVIATION OF THE ERROR E BETWEEN THE LIMBUS AND THE LIGHT LINE POSITION FOR DIFFERENT IRIS PIGMENTATION.

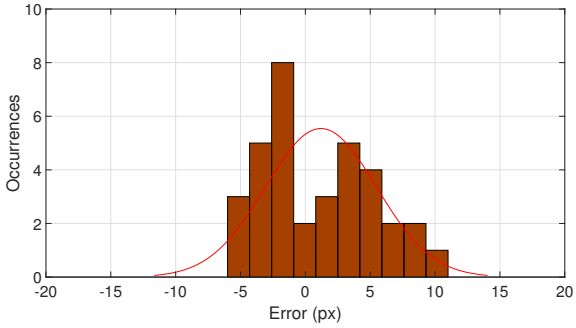
| Color | AlexNet | | VGG-16 | |
|---|---|---|---|---|
| | $\mu$ (px) | $\sigma$ (px) | $\mu$ (px) | $\sigma$ (px) |
| Black | 1.09 | 4.36 | 2, 6 | 3, 2 |
| Brown | 1.20 | 4.28 | 3, 7 | 5, 9 |
| Green | 6.69 | 5.19 | 6, 2 | 5, 6 |
| Blue | 7.51 | 5.48 | 7, 2 | 6, 2 |
| Total | 4.12 | 5.67 | 5, 37 | 6, 1 |

For this reason, the system acquires the red RGB channel from the images, to better highlight the line. This operation leads to different light line intensities, depending on the specific eye pigmentation. In Fig. 12 it is possible to observe

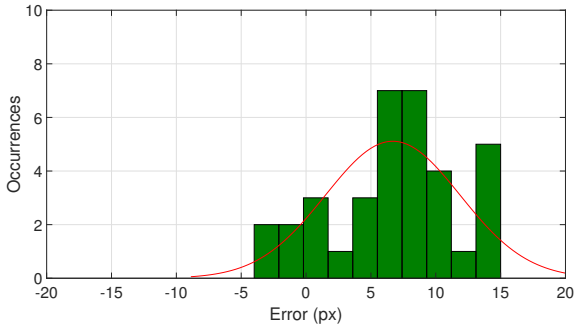
the mean maximum line intensity per eye pigmentation.



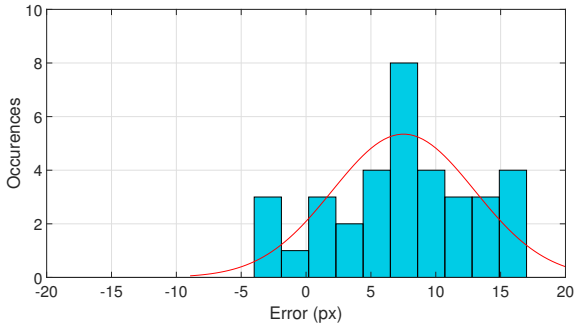

(a)




(b)




(c)




(d)

Figure 11. Histograms of occurrences of the error $E$ between the limbus and the light line position for different eye pigmentation, were obtained with the AlexNet network. (a) black eyes, (b) brown eyes, (c) green eyes and (d) blue eyes. In every figure the • line represents the Gaussian fitting for each histogram.

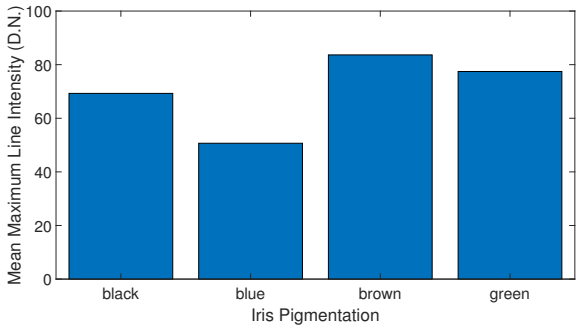



Figure 12. Histogram of the mean intensity of the light line for each different eye color.

The mean maximum intensity is retrieved by extracting ten contiguous horizontal line intensity profiles from ten different center images per color. The profiles have been extracted in correspondence with the eye y-center. As it is possible to see in Fig. 12 the brown eyes show the highest light line intensity, while the blue eyes present the lowest maximum intensity. The rationale of the aforementioned behavior is that brown eyes have a higher red component with respect to green and blue eyes. For this reason, brown eyes images are clearer, resulting in a lower mean error. Black iris represents an exception as it shows a lower mean maximum intensity while performing better in terms of mean error. In this case, the results are reasonable because black is a neutral color, thus its behavior does not depend on the specific RGB channel acquired. At the same time, black is the iris color that shows a better contrast with respect to the white color of the sclera, which probably leads to a better light line recognition by the neural network.

## VII. Conclusions and Future Work

The use of Machine Learning techniques to extract measurement information from ophthalmic images is a very promising technique that can overcome some critical aspects of the manual Van Herick approach. After a research phase where different ML techniques have been tested, CNN networks proved to perform better in terms of precision and accuracy for our specific test case. Both Alexnet and VGG-16 structures proved to be suitable for the considered application. Indeed, AlexNet revealed to be less resource-demanding, providing slightly lower prediction times. Besides, VGG-16 is a little more accurate, and considering that due to the specific employed hardware and the fact that the analysis is made after the image collection phase, the proposed Neural Networks provide very good results in the *center* images choice while maintaining optimal computational complexity.

In this work, a new dataset has been collected by an expert ophthalmologist. This extensive measurement activity is of fundamental importance to, on the one hand, evaluate the neural network with realistic statistical samples. On the other hand, these measures have been taken in different environmental conditions with the respect to the training images. Moreover, the effect of different iris pigmentation has been observed. The presented results show low mean and standard deviation values of the error between the light line and the limbus positions. Moreover, also the maximum error observed

during the experiment, i.e. $E \approx 18$ px, is visually comparable to the line width. Moreover, the trained CNN has proven to work as expected in different environmental conditions, in a realistic clinic field, and with eyes of different colors. More precisely, the detection of the line on darker irises, i.e. black and brown, reported a noticeable lower average $\mu$ error than eyes with lighter colors. For this reason, as a future improvement, the system will be automatically able to change the light line color, depending on the specific patient's iris pigmentation. After the *center* images are identified, a Vision–Based algorithm is currently under development to measure the displacement between the slit light that hits the cornea and the refracted light inside the anterior chamber. A refined tuning of the Vision–Based system will be carried out, based on a comprehensive inter–subject experimental campaign. Moreover, the ongoing activity is aimed at a careful analysis of the measurement uncertainty.

## REFERENCES


[1] Y.-C. Tham, X. Li, T. Y. Wong, H. A. Quigley, T. Aung, and C.-Y. Cheng, "Global prevalence of glaucoma and projections of glaucoma burden through 2040: A systematic review and meta-analysis," *Ophthalmology*, vol. 121, no. 11, pp. 2081–2090, 2014.

[2] A. Javed, M. Loutfi, S. Kaye, and M. Batterbury, "Interobserver reliability when using the Van Herick method to measure anterior chamber depth," *Oman Journal of Ophthalmology*, vol. 10, no. 1, pp. 9–12, Apr. 2017.

[3] I. Riva, E. Micheletti, F. Oddone, C. Bruttini, S. Montescani, G. De Angelis, L. Rovati, R. N. Weinreb, and L. Quaranta, "Anterior chamber angle assessment techniques: A review," *Journal of Clinical Medicine*, vol. 9, no. 12, p. 3814, Nov 2020.

[4] J. Gispets, G. Cardona, M. Verdú, and N. Tomàs, "Sources of variability of the van Herick technique for anterior angle estimation," *Clinical & Experimental Optometry*, vol. 97, no. 2, pp. 147–151, Mar. 2014.

[5] P. L. Dabasia, D. F. Edgar, I. E. Murdoch, and J. G. Lawrenson, "Noncontact Screening Methods for the Detection of Narrow Anterior Chamber Angles," *Investigative Ophthalmology & Visual Science*, vol. 56, no. 6, pp. 3929–3935, Jun. 2015.

[6] E. S. dos Santos, W. B. Xavier, R. N. Rodrigues, S. S. da C. Botelho, and A. V. Werhli, "Vision based measurement applied to industrial instrumentation," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 788 – 793, 2017, 20th IFAC World Congress.

[7] T. Fedullo, D. Cassanelli, G. Gibertoni, F. Tramarin, L. Quaranta, G. de Angelis, and L. Rovati, "A machine learning approach for a vision-based van-herick measurement system," in *2021 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 2021, pp. 1–6.

[8] D. Ferreiro, S. García-Montero, J. Garcia-Queiruga, E. P. Dorrio, and E. Yebra-Pimentel, "Assessment of Van Herick technique by using ImageJ software," in *Fourth International Conference on Applications of Optics and Photonics*. SPIE, Oct 2019, vol. 11207, pp. 310–315.

[9] J. Gispets, G. Cardona, M. Verdú, and N. Tomàs, "Sources of variability of the van Herick technique for anterior angle estimation," *Clinical and Experimental Optometry*, vol. 97, no. 2, pp. 147–151, Mar 2014.

[10] P. J. Foster, J. G. Devereux, P. H. Alsbirk, P. S. Lee, D. Uranchimeg, D. Machin, G. J. Johnson, and J. Baasanhu, "Detection of gonioscopically occludable angles and primary angle closure glaucoma by estimation of limbal chamber depth in Asians: modified grading scheme," *Br J Ophthalmol.*, vol. 84 (2), no. 2, pp. 186–192, Feb. 2000.

[11] R. Sihota, N. Kamble, A. K. Sharma, A. Bhari, A. Gupta, N. Midha, H. Selvan, T. Dada, V. Gupta, and R. M. Pandey, "'Van Herick Plus': a modified grading scheme for the assessment of peripheral anterior chamber depth and angle," *British Journal of Ophthalmology*, vol. 103, no. 7, pp. 960–965, 2019, publisher: BMJ Publishing Group Ltd _eprint: https://bjo.bmj.com/content/103/7/960.full.pdf. [Online]. Available: https://bjo.bmj.com/content/103/7/960

[12] C. Theeraworn, W. Kongprawechnon, T. Kondo, P. Bunnun, A. Nishihara, and A. Manassakorn, "Automatic screening of narrow anterior chamber angle and angle-closure glaucoma based on slit-lamp image analysis by using support vector machine," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013, pp. 5887–5890.

[13] ——, "Automatic screening algorithm for narrow anterior chamber angle and angle-closure glaucoma based on slit-lamp image analysis," vol. 47, no. 6, p. 940 – 952, 2013.

[14] F. Teixeira, D. C. Sousa, I. Leal, A. Barata, C. M. Neves, and L. A. Pinto, "Automated gonioscopy photography for iridocorneal angle grading," *European Journal of Ophthalmology*, vol. 30, no. 1, pp. 112–118, 2020, pMID: 30360660. [Online]. Available: https://doi.org/10.1177/1120672118806436

[15] H. Fu, M. Baskaran, Y. Xu, S. Lin, D. W. K. Wong, J. Liu, T. A. Tun, M. Mahesh, S. A. Perera, and T. Aung, "A Deep Learning System for Automated Angle-Closure Detection in Anterior Segment Optical Coherence Tomography Images," *American Journal of Ophthalmology*, vol. 203, pp. 37–45, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S000293941930087X

[16] E. Shimizu, H. Yazu, N. Aketa, R. Yokoiwa, S. Sato, J. Yajima, T. Katayama, R. Sato, M. Tanji, Y. Sato, Y. Ogawa, and K. Tsubota, "A Study Validating the Estimation of Anterior Chamber Depth and Iridocorneal Angle with Portable and Non-Portable Slit-Lamp Microscopy," *Sensors*, vol. 21, no. 4, p. 1436, Feb 2021.

[17] M. Khanafer and S. Shirmohammadi, "Applied ai in instrumentation and measurement: The deep learning revolution," *IEEE Instrumentation Measurement Magazine*, vol. 23, no. 6, pp. 10–17, Oct 2020.

[18] G. Lee and H. Fujita, *Deep learning in medical image analysis: challenges and applications*. Springer, 2020, vol. 1213.

[19] J. Zhong, Z. Liu, Z. Han, Y. Han, and W. Zhang, "A cnn-based defect inspection method for catenary split pins in high-speed railway," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 8, pp. 2849–2860, Aug 2019.

[20] M. Venturelli, G. Borghi, R. Vezzani, and R. Cucchiara, "Deep Head Pose Estimation from Depth Data for In-Car Automotive Applications," in *Understanding Human Activities Through 3D Sensors*, ser. Lecture Notes in Computer Science, H. Wannous, P. Pala, M. Daoudi, and F. Flórez-Revuelta, Eds. Cham: Springer International Publishing, 2018, pp. 74–85.

[21] Y. Tong, W. Lu, Y. Yu, and Y. Shen, "Application of machine learning in ophthalmic imaging modalities," *Eye and Vision*, vol. 7, 2020. [Online]. Available: https://doi.org/10.1186/s40662-020-00183-6

[22] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," *Neural Information Processing Systems*, vol. 25, 01 2012.

[23] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," 2014.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.

[25] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[26] M. Leung, S. S. O. Kang, J. Turuwhenua, and R. Jacobs, "Effects of illumination and observation angle on the van Herick procedure," *Clinical and Experimental Optometry*, vol. 95, no. 1, pp. 72–77, Jan 2012.

[27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015. [Online]. Available: http://arxiv.org/abs/1409.1556

[28] Y. Wang, Y. Li, Y. Song, and X. Rong, "The influence of the activation function in a convolution neural network model of facial expression recognition," *Applied Sciences*, vol. 10, no. 5, p. 1897, 2020.

[29] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.

[30] B. H. Pui, B. Hayes-Gill, M. Clark, M. Somekh, C. See, J.-F. Piéri, S. Morgan, and A. Ng, "The Design and Characterisation of an Optical VLSI Processor for Real Time Centroid Detection," *Analog Integrated Circuits and Signal Processing*, vol. 32, pp. 67–75, Jul. 2002.

[31] K. K. Chavali, P. M. Furth, and P. R. Surkanti, "A low-voltage, adaptive CMOS centroid image sensor with improved bandwidth," in *2014 IEEE 57th International Midwest Symposium on Circuits and Systems (MWSCAS)*, Aug. 2014, pp. 937–940, iSSN: 1558-3899.