Training of an artificial intelligence algorithm for automatic detection of the Van Herick grade

Davide Cassanelli[a], Tommaso Fedullo[a,b], Giovanni Gibertoni[a], Giorgia Saporito[c], Manuela Ferrazza[c], Francesco Oddone[c], Ivano Riva[d], Luciano Quaranta[d], Federico Tramarin[a], and Luigi Rovati[a]

[a]Dip. di Ingegneria "Enzo Ferrari", Univ. di Modena e Reggio Emilia, Via Vivarelli 10 - 41125 Modena (Italy)
[b]Dept. of Management and Engineering, University of Padova, Vicenza, Italy,
[c]IRCCS Fondazione G.B. Bietti per lo Studio e la Ricerca in Oftalmologia, Italy,
[d]Istituto Clinico Sant'Anna, Italy,

ABSTRACT

Van Herick technique is a qualitative tool for assessing the anterior chamber angle and can be exploited as a simple screening alternative to gonioscopy. In our previous papers, we presented a novel instrument able to automatically perform the Van Herick manoeuvre. Therefore, to fully automate the screening method from the acquired images, it is still necessary to automatically determine the Van Herick grade. In this paper, we present a deep learning algorithm for automatically determining the Van Herick grade. In particular, the performances of three different Convolutional Neural Networks have been verified by acquiring the eye images of 80 patients. All the networks return the Van Herick grade classification with sufficient accuracy for a screening system and, after proper training, can offer a real-time response.

Keywords: Van Herick technique, Primary Angle Closure Glaucoma, Narrow Anterior Chamber Angle measurements, ophthalmic Instrumentation, Imaging measurement methods, Deep Learning, Artificial Intelligence, Convolutional Neural Networks

1. INTRODUCTION

Glaucoma is one of the most diffused ophthalmic diseases worldwide and, specifically, Primary Angle Closure Glaucoma (PACG) is the most aggressive type of Glaucoma; it may cause bilateral blindness in a very short time making essential the prevention and screening.[1] In particular, people with a Narrow Anterior Chamber Angle (NACA), i.e. narrow iridocorneal angle, are more exposed to the development of PACG.[2] Van Herick's technique can be used as a screening method to detect NACA and, consequentially, predict potential PACG development. According to this non-contact, qualitative approach, the ACA width is estimated comparing the Peripheral Anterior Chamber Depth (PACD) and the thickness of the cornea (CT).[3,4] The width is then classified into 4 grades, i.e 1 for narrow-angle and 4 for wide-angle. The technique usually requires a slit lamp, enlightening the limbus with a 60° angle, and trained personnel to perform the manoeuvre. The estimation comes from the visual or digital image comparison between PACD and CT and so some uncertainty due to the operator and conditions of the manoeuvre (illumination, position, etc.) can be introduced.[5] Digital eye images can be analyzed according to two complementary possible approaches: i) image recognition algorithms extracting the PACD and CT measurements, and ii) Artificial Intelligence (AI) algorithms for classification of the images according to the Van Herick grade (VHG).

In this paper, we consider AI approaches. In recent years, many studies have been brought on considering the possible applications of machine learning on ophthalmic images.[6,7] Considering the NACA detection, Fu[8] proposed a deep learning algorithm applied to images captured with optical coherent tomography whereas Theeraworn *et. al*[9,10] proposed a Support Vector Machine Method applied on images acquired using the classic slight lamp approach. These approaches are not suitable for PACG screening operations mainly due to the cost of the

---

e-mail corresponding author: davide.cassanelli@unimore.it

instrumentation and for the need for qualified and highly trained personnel. In this context, the exploitation of AI methods allows the reduction of complexity, and therefore the cost of the instrumentation. Moreover, the use of such intelligent and automatic techniques, together with a suitable User Interface, allows execution of the test even by less qualified personnel. In our previous paper, we presented a compact instrument capable of automatically performing the manoeuvre[11] and returning the most significant images[12] of the eye. From these images an expert can perform the visual comparison and hence assigns a Van Herick grade. To completely automatize the technique, it is necessary to determine, in a fully automated way, the two key parameters, i.e. PACD and CT, but this operation is not trivial and requires heavy image processing. The high computational cost may result in a rather difficult implementation within a low-cost portable embedded system.

Therefore, we analyzed the potential of machine learning approaches by comparing the performance of three types of Convolutional Neural Networks (CNN). The analysis has been performed using the MATLAB Deep Network Designer tool. In the following, Section 2 describes the considered networks and the methods used to compare their performance. The results are reported in Section 3, where the accuracy and confusion matrix of the considered networks are presented. Finally, discussion and conclusions are drawn in Section 4.

2. METHODS

Eighty adult subjects, 30 males and 50 females, with ages ranging from 25 to 70 years were recruited from the "IRCCS Fondazione G.B. Bietti per lo Studio e la Ricerca in Oftalmologia", and informed consent was obtained from each subject. The collected dataset has the characteristics described in in Table 1.

Charateristic		N° of Patients
Sex	Male	30
	Female	50
Age	<30	4
	30-50	48
	50-70	28
Eye Colour	Black	9
	Blue	15
	Brown	47
	Green	9

Table 1. Enrolled subjects data.

Note that the subjects had also different eye pigmentation, thus allowing to train a network being capable to generalize its functionality to a variety of eye colors. Standard Van Herick's technique was firstly applied to each patient's right eye by highly trained personnel and the corresponding VHG was assigned. This value was considered as the reference for the analysis of the CNNs performance. Immediately afterward, the subject was analyzed using our automatic instrument, and the images of the right eye generated were used for the training and verification of the CNNs. Since the instrument generated 2 to 4 images for each measurement, the total number of images considered for this study was 277; the distribution of these images in the four VHGs is shown in Table 2.

VHG	Eye Images
1	55
2	62
3	54
4	106

Table 2. Number of images collected for each VHG.

It is worth noting that, by remaining faithful to the real distribution of the VHGs in the population, we made no prior selection of the subjects, thus resulting in an higher number of collected images associated with VHG = 4. As an example, in Figure 1 two typical images acquired by the instrument at the two extreme conditions, i.e. VHG = 1 and VHG = 4, are shown.
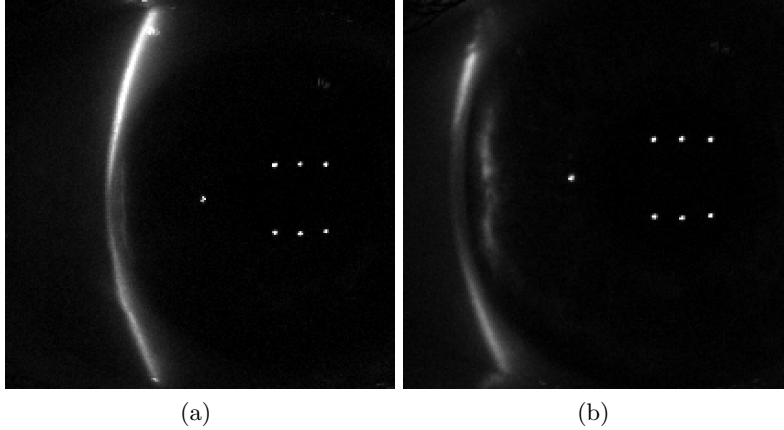
Figure 1. Typical images acquired with the instrument. Figure (a) shows a typical VHG = 1 eye, in which the two line projections are quite closed, while Figure (b) shows a VHG = 4 with a wide distance between the two line projections.

Exploiting the Deep Network Designer tool of MATLAB, three different types of CNN were tested. Specifically, the tested networks were: i) AlexNet, already used for finding the most significant images in our previous work,[12] ii) GoogLeNet, and iii) ResNet. All the CNNs were just pre-trained by the MATLAB tool. The validation of the networks was performed using the 40% images of the original dataset, while the remaining 60% images were augmented and used for the networks training. Because the collected images were imbalanced towards the VHG = 4, data augmentation was performed just on the training images, thus obtaining a balanced final training dataset. Data augmentation was performed using a Python algorithm that randomly shifted, rotated, zoomed, and changed the brightness of the original images. This operation was not performed using MATLAB, since the Deep Network Designer tool was not able to control the number of images of the augmented set. It is important to highlight that all the augmented images were used, with no post-processing and no manual selection, in order to obtain a more robust training of the networks. The final augmented dataset, used for both training and validation of the CNNs, is summarized in Table 3.

VHG	Training Images	Validation Images
1	4026	22
2	3996	25
3	3968	22
4	3969	43

Table 3. Dataset used for the networks training and validation.

After training, the performances of the CNNs were compared in terms of the capability of correctly classifying an image according to its VHG class. The images classified by a network resulted in true-positive (TP), i.e. correctly classified, false-positive (FP), i.e. incorrectly classified in a class, and false-negative (FN), i.e. incorrectly not classified in a class. The performances of the networks were evaluated according to three metrics. Firstly, Accuracy, defined as in 1 for the entire dataset describes the capability of a network to correctly classify an image. Then, the Precision for a specific class C, defined as in 2, indicates the network capability to not label different class images to class C. Finally, Recall for a specific class C, defined as in 3 represents the network capability to not confuse class C images with other class images.

$$Accuracy(\%) = \frac{Correct}{Total} * 100, \tag{1}$$

$$Precision_C(\%) = \frac{TP_C}{TP_C + FP_C} * 100, \tag{2}$$

$$Recall_C(\%) = \frac{TP_C}{TP_C + FN_C} * 100, \tag{3}$$

where subscript C identifies the specific VHG class.

3. RESULTS

The results of the training of the GoogLeNet, ResNet and AlexNet are shown, respectively, in Figures 2.(a), 2.(b) and 2.(c). In these images, the plot of the training accuracy, in blue, and the validation accuracy, in orange, are shown with respect to the number of software iterations performed to train the networks. As shown in the figure, the number of iterations is not equal for the different networks, since the training was performed until the validation accuracy stabilizes. From these plots it is possible to observe that the ResNet network reaches more faster a high value of accuracy, both for training and validation. Since the training and validation accuracy grow together it is possible to exclude the possibility of overfitting. Moreover, AlexNet was the network that, depending on the i-th iterations, had the higher variability in the accuracy; probably this result is due to the fact that this network had the simplest configuration among the tested network.
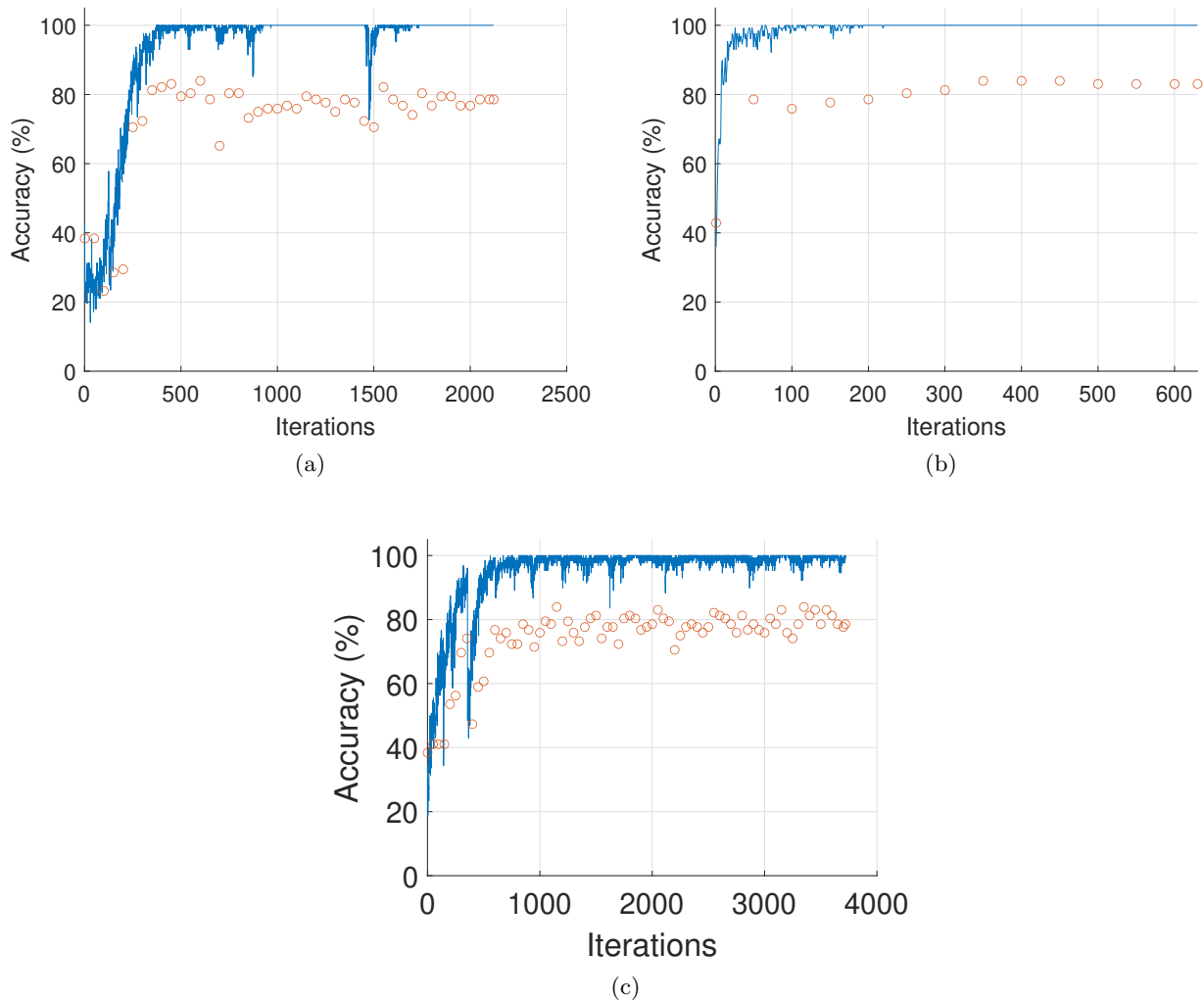

(a)

(b)

(c)

Figure 2. Accuracy plot of GoogLeNet (a), ResNet (b) and AlexNet (c) respect to the training iterations. Training Accuracy (-) and Validation Accuracy (○)

The performances of the networks were evaluated in terms of Recall and Precision exploiting the Validation dataset. In Figures 3.(a), 3.(b) and 3.(c) the confusion matrix for each trained network is shown. The columns represent the reference VHG (Target) while the rows the predicted VHG (Output). The main diagonal represents the images that are correctly classified, while the other cells represent the images that are not correctly classified and the incorrect VHG estimated for them. The percentage in each cell is calculated with respect to the total

validation image number. The last column represents the Precision, in green; the last row represents the Recall, in green. The bottom right cell shows the overall accuracy for the trained network.

(a)

	VH$_1$	VH$_2$	VH$_3$	VH$_4$	
VH$_1$	19.6%	1.8%	0.9%	0.9%	84.6%
VH$_2$	0.0%	17.9%	0.0%	8.9%	66.7%
VH$_3$	0.0%	0.0%	13.4%	0.9%	93.8%
VH$_4$	0.0%	2.7%	5.4%	27.7%	77.5%
	100%	80.0%	68.2%	72.1%	**78.6%**

(b)

	VH$_1$	VH$_2$	VH$_3$	VH$_4$	
VH$_1$	17.9%	0.0%	0.0%	0.0%	100%
VH$_2$	1.8%	19.6%	0.9%	7.1%	66.7%
VH$_3$	0.0%	0.0%	16.1%	2.7%	85.7%
VH$_4$	0.0%	2.7%	2.7%	28.6%	84.2%
	90.9%	88.0%	81.8%	74.4%	**82.1%**

(c)

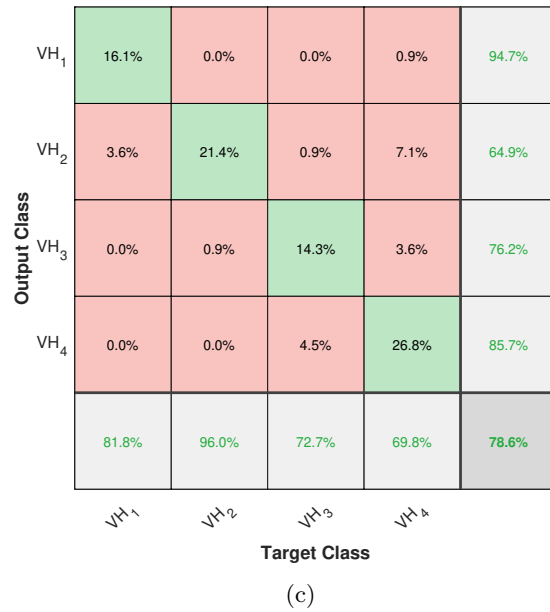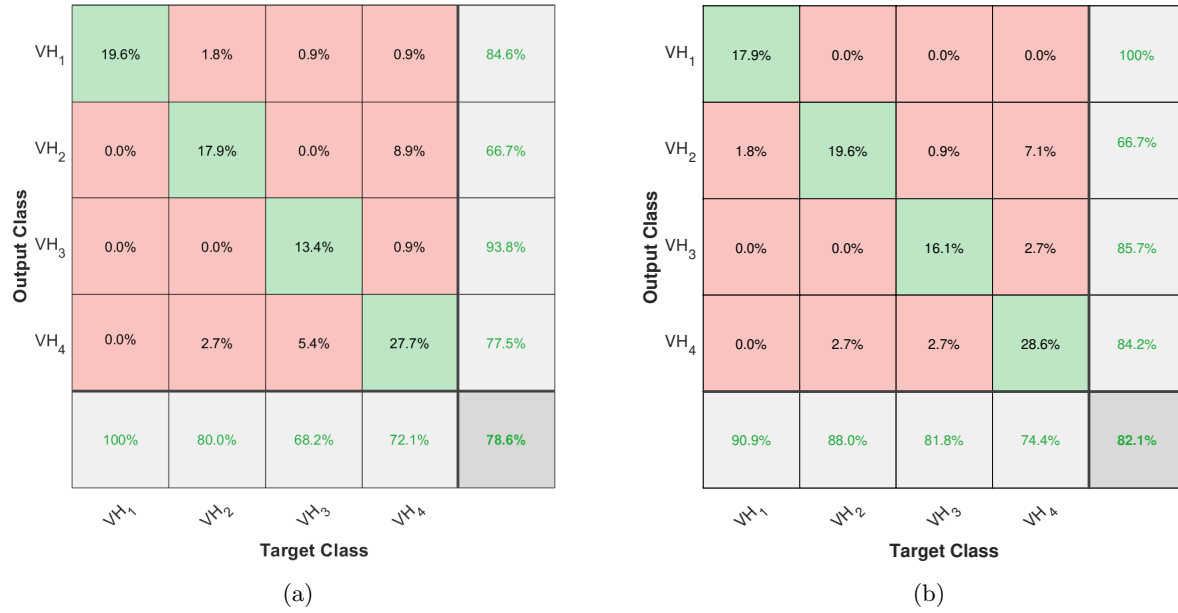|	VH$_1$	VH$_2$	VH$_3$	VH$_4$	
VH$_1$	16.1%	0.0%	0.0%	0.9%	94.7%
VH$_2$	3.6%	21.4%	0.9%	7.1%	64.9%
VH$_3$	0.0%	0.9%	14.3%	3.6%	76.2%
VH$_4$	0.0%	0.0%	4.5%	26.8%	85.7%
	81.8%	96.0%	72.7%	69.8%	**78.6%**

Figure 3. Confusion matrix for the GoogLeNet (a), ResNet (b) and AlexNet (c). The main diagonal, green cells, represents the correctly classify images. The light grey row shows the Recall data, in green. The light grey column shows the Precision data, in green. In the bottom right cell there is the overall Accuracy data.

4. DISCUSSION AND CONCLUSION

In this preliminary study, we have shown that CNNs can be used to estimate the Van Herick grade. The performances of all the tested networks are suitable for real-time estimation of Van Herick grade in portable screening instruments, without requiring a huge computational power.

All the networks have similar characteristics. The overall accuracy is in the order of 80% for each network, showing a good accuracy even if the training dataset is still limited. Analyzing the obtained results, even if the networks work in a very similar way, it is possible to appreciate that ResNet has better behaviour, but at the same time is the most complex network, with the highest computational cost in training phase, among the CNNs that have been tested.

Recall results show that the probability to correctly classify a VHG = 1 and VHG = 2 is higher than 85%. Differently, the Recall value for VHG = 4 is lower. This recall reduction could be due to the intrinsic difference of the positions of the light line projections among the images at different grades. In fact, considering the VHG = 4, a small misalignment of the light line respect to the the limbus position produces a large variation of the position of the projected line on the iris; thus VHG = 4 can be easily classified as a VHG = 3 or worstly as VHG = 2. This phenomenon result also in a low value of Precision for VHG = 2, for all the tested networks, and VHG= 3, mostly for AlexNet. In fact, networks confuse VHG = 4 with the other grades increasing the number of images incorrectly classified as other grades. It is possible to observe that each network incorrectly classifies at least the 7% of VHG = 4 images as VHG = 2. Differently, for the other grades, this critical issues tends to be reduced, causing less confusion in the classification.

Much more work is still needed to develop a specific CNN for our automatic instrument. However, we can confirm that the results obtained to date are encouraging and that hopefully, they will improve by having a larger set of training images. In addition, it will be important to improve the detection of the correct images for measurement in order to feed the VHG convolutional network with more accurate images and reduce classification confusion. Simultaneously, we are also developing a deterministic approach capable of providing a measure of the distance between the projected light lines. In this case, the effort is to reduce the computational load in order to make the algorithm implementable in a compact and portable system.

REFERENCES

[1] Tham, Y.-C., Li, X., Wong, T. Y., Quigley, H. A., Aung, T., and Cheng, C.-Y., "Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis," *Ophthalmology* **121(11)**, 2081–2090 (Nov 2014).

[2] Riva, I., Micheletti, E., Oddone, F., Bruttini, C., Montescani, S., De Angelis, G., Rovati, L., Weinreb, R. N., and Quaranta, L., "Anterior chamber angle assessment techniques: A review," *J Clin Med* **9**, 3814 (Nov 2020).

[3] Van Herick, W., Shaffer, R. N., and Schwartz, A., "Estimation of width of angle of anterior chamber. incidence and significance of the narrow angle. incidence and significance of the narrow angle," *Am J Ophthalmol.* **68(4)**, 626–629 (Oct 1969).

[4] Foster, P. J., Devereux, J. G., Alsbirk, P. H., Lee, P. S., Uranchimeg, D., Machin, D., Johnson, G. J., and Baasanhu, J., "Detection of gonioscopically occludable angles and primary angle closure glaucoma by estimation of limbal chamber depth in Asians: modified grading scheme," *Br J Ophthalmol.* **84 (2)**, 186–192 (Feb. 2000).

[5] Leung, M., Kang, S. S. O., Turuwhenua, J., and Jacobs, R., "Effects of illumination and observation angle on the Van Herick procedure," *Clin Exp Optom.* **95(1)**, 72–77 (Jan 2012).

[6] Tong, Y., Lu, W., Yu, Y., and Shen, Y., "Application of machine learning in ophthalmic imaging modalities," *Eye and Vision* **7** (2020).

[7] Khanafer, M. and Shirmohammadi, S., "Applied ai in instrumentation and measurement: The deep learning revolution," *IEEE Instrumentation Measurement Magazine* **23**, 10–17 (Oct 2020).

[8] Fu, H., Baskaran, M., Xu, Y., Lin, S., Wong, D. W. K., Liu, J., Tun, T. A., Mahesh, M., Perera, S. A., and Aung, T., "A Deep Learning System for Automated Angle-Closure Detection in Anterior Segment Optical Coherence Tomography Images," *American Journal of Ophthalmology* **203**, 37–45 (2019).

[9] Theeraworn, C., Kongprawechnon, W., Kondo, T., Bunnun, P., Nishihara, A., and Manassakorn, A., "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)*], 5887–5890 (2013).

[10] Theeraworn, C., Kongprawechnon, W., Kondo, T., Bunnun, P., Nishihara, A., and Manassakorn, A., "Automatic screening algorithm for narrow anterior chamber angle and angle-closure glaucoma based on slit-lamp image analysis," *Kasetsart Journal - Natural Science* **47**, 940–952 (01 2013).

[11] Cassanelli, D., Gibertoni, G., Cattini, S., Quaranta, L., Riva, I., Bruttini, C., Angelis, G. D., and Rovati, L., "A new screening system for the estimation of ocular anterior chamber angle width," in [*Ophthalmic Technologies XXXI*], Hammer, D. X., Joos, K. M., and Palanker, D. V., eds., **11623**, 101 – 105, International Society for Optics and Photonics, SPIE (2021).

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