Received: 21 September 2024 | Accepted: 16 December 2024 | Published online: 26 December 2024

DOI: 10.1111/opo.13435

OPO THE COLLEGE OF OPTOMETRISTS

ORIGINAL ARTICLE

Artificial intelligence virtual assistants in primary eye care practice

Leandro Stuermer[1,2] | Sabrina Braga[1,2] | Raul Martin[2,3] | James S. Wolffsohn[4]

[1]Department of Optometry, University of Contestado, Canoinhas, Brazil

[2]Optometry Research Group, School of Optometry, IOBA Eye Institute, University of Valladolid, Valladolid, Spain

[3]Departamento de Física Teórica, Atómica y Óptica, Universidad de Valladolid, Valladolid, Spain

[4]Optometry and Vision Sciences Research Group, Aston University, Birmingham, UK

Correspondence
Raul Martin, IOBA Eye Institute, University of Valladolid, Valladolid, Spain.
Email: raul@ioba.med.uva.es

Abstract

Purpose: To propose a novel artificial intelligence (AI)-based virtual assistant trained on tabular clinical data that can provide decision-making support in primary eye care practice and optometry education programmes.

Method: Anonymised clinical data from 1125 complete optometric examinations (2250 eyes; 63% women, 37% men) were used to train different machine learning algorithm models to predict eye examination classification (refractive, binocular vision dysfunction, ocular disorder or any combination of these three options). After modelling, adjustment, mining and preprocessing (one-hot encoding and SMOTE techniques), 75 input (preliminary data, history, oculomotor test and ocular examinations) and three output (refractive, binocular vision status and eye disease) features were defined. The data were split into training (80%) and test (20%) sets. Five machine learning algorithms were trained, and the best algorithms were subjected to fivefold cross-validation. Model performance was evaluated for accuracy, precision, sensitivity, F1 score and specificity.

Results: The random forest algorithm was the best for classifying eye examination results with a performance >95.2% (based on 35 input features from preliminary data and history), to propose a subclassification of ocular disorders with a performance >98.1% (based on 65 features from preliminary data, history and ocular examinations) and to differentiate binocular vision dysfunctions with a performance >99.7% (based on 30 features from preliminary data and oculomotor tests). These models were integrated into a responsive web application, available in three languages, allowing intuitive access to the AI models via conventional clinical terms.

Conclusions: An AI-based virtual assistant that performed well in predicting patient classification, eye disorders or binocular vision dysfunction has been developed with potential use in primary eye care practice and education programmes.

KEYWORDS
artificial intelligence, clinical decision support, machine learning, optometry, virtual assistant

INTRODUCTION

Artificial intelligence (AI) seeks to imitate human thought in a previously unprogrammed way through the manipulation of data through different machine learning (ML) algorithms, deep learning (DL) and a variety of neural networks (NNs).[1] In the fields of optometry and ophthalmology, there has been strong interest from researchers because AI can improve the diagnosis, therapeutic management and care of several eye conditions,[2] primarily related to posterior eye disorders (such as diabetic retinopathy, age-related macular degeneration and glaucoma but also to detect retinopathy of prematurity[3] or macular hole assessment[4]). Additionally, AI can be applied to anterior segment abnormalities[5] (such as keratoconus, infectious keratitis, cataracts, corneal topography assessment or surgical procedures such as corneal transplant and refractive surgery).[6] Furthermore, AI has applications in ocular refraction[7] (e.g., after cataract surgery)[8] and visual development in paediatric patients[3] (such as amblyopia, binocular vision disorders and strabismus).

However, AI has not been widely used in clinical practice because of its availability and heterogeneity in imaging techniques (typically fundus photography, optical coherence tomography (OCT), corneal topography and tomography and even slit-lamp photography).[9] This results in a gap in AI applications that utilise clinical findings in primary eye care practice.

This study aimed to create a novel AI-based virtual assistant trained on tabular clinical data to provide decision-making support in primary eye care practice and preclinical support to eye care professionals and eye care professional education programmes.

Key points

- Artificial intelligence is of significant interest in eye care practice, as it can improve diagnosis, therapeutic management and patient care in several ocular conditions.
- Artificial intelligence applications have not been widely applied in clinical practice, highlighting a gap in applications using clinical data to provide decision-making support in primary eye care practice.
- A novel artificial intelligence-based virtual assistant trained on tabular clinical data is proposed to support primary eye care practice decision-making and be used in education programmes.

MATERIALS AND METHODS

Clinical data

Anonymised clinical data corresponding to 1125 complete optometric examinations (2250 eyes; 708 [63%] women and 417 [37%] men; mean age 35 ± 20 years) conducted between 2012 and 2021 were retrieved from the University of Contestado Visual Health Clinic, Brazil. Only complete records with clear diagnoses and management strategies were included. This study followed the principles of the Declaration of Helsinki, and the study protocol was approved by the ethics committee of the University of Contestado, Brazil.

Data modelling

Excel and Access 365 (Microsoft.com) and Orange Data Mining 3.32.0 (orangedatamining.com) software were used for data adjustment, mining and modelling on an individual eye basis, checking for incorrect data, inconsistent entries, typing mistakes, etc., to ensure the integrity of the data for evaluating different AI algorithms.

Experienced optometrists classified all cases appropriately into three output features: refractive, ocular segment disorder or binocular vision dysfunction or any combination of these options. Suspected ocular disorders were also classified as related to the anterior eye, posterior eye or both. If the patient presented with any binocular vision dysfunction, it was further classified as accommodation dysfunction, fusional vergence dysfunction or both possibilities (Figure 1). In all cases, specific concurrences for each eye were separated.

Clinical data were classified into 75 input features distributed across the preliminary data and history of the patient (general characteristics, visual acuity, symptoms, personal and family history), ocular motility (including accommodation, cover test, fusional vergences, Hirschberg test and near point of convergence), ocular examinations of ophthalmoscopy (signs of the optic disc and macula), and biomicroscopic evaluation of the anterior eye (eyelids, conjunctiva and cornea), as summarised in Figure 1.

Three models were trained with machine learning algorithms: Model 1 to predict the most likely case classification (M1), Model 2 to predict pathology in different ocular segments (M2) and Model 3 to predict any type of binocular vision dysfunction. Table 1 summarises the input and output features used in the three models.

Preprocessing

The input features used to train each model were selected, excluding records with missing values or outliers. Categorical variables were handled using the one-hot encoding technique.[10] To mitigate the sensitivity of machine learning algorithms to numerical variations and sample imbalance, four versions of the base data set were generated: original unbalanced data (Ou), original balanced data (Ob), unbalanced normalised data (Nu) and balanced normalised data (Nb).[11] Figure 2 overviews the data processing and model validation steps.

Balancing was performed using the synthetic minority oversampling technique (SMOTE), which generates synthetic examples for minority output classes to balance the data distribution.[12] Normalisation, in turn, adjusts the numerical values to a common scale between 0 and 1. Testing these different data versions is essential, as machine learning algorithms can be sensitive to data distribution and scale. These variations allow for the identification of the most effective approach for each model.[11]

FIGURE 1 Names and formats of the features used to classify the eyes. ISNT, inferior, superior, nasal, temporal; NPC, near point of convergence.

AI algorithms training and validation

Each model was trained with five different machine learning algorithms for multiclass classification tasks: (Random Forest [RF], Decision Tree [DT], Neural Network [NN], Support Vector Machine [SVM] and Logistic Regression [LR]) using Python (3.9.7) (Python Software Foundation, python.org) and the Pandas (1.5.3) (Pandas Development Team, pandas.pydata.org/), NumPy (1.26.4) (NumPy Developers, numpy.org/) and Scikit-Learn (1.3.0) (Scikit-learn Developers, scikit-learn.org/) libraries. The test was performed with each database version, using 80% of the data for training and 20% for testing, which were randomly allocated. The accuracy performance of each algorithm was compared.

TABLE 1 Summary of input and output features used to develop the models.

	Model 1	Model 2	Model 3
Features (IN/OUT)	35/1	65/1	30/1
Aim	**Case classification?**	**Where could a possible pathological ocular problem be?**	**Vergence or accommodation dysfunction?**
Input (*group predictor variables*)	Preliminary examinations History	Preliminary examinations History Ocular tests	Preliminary examinations Oculomotor tests
Outcome (*target variable*)	Clinical management	Ocular segment disorder	Binocular vision dysfunction
Output classes (*possible values*)	None Oculomotor Ocular pathology Refractive Oculomotor and pathology Refractive and oculomotor Refractive and pathology Refractive, oculomotor and pathology	None Anterior pole Posterior pole Anterior and posterior pole	None Accommodative dysfunction Vergence dysfunction Accommodative and vergence dysfunction

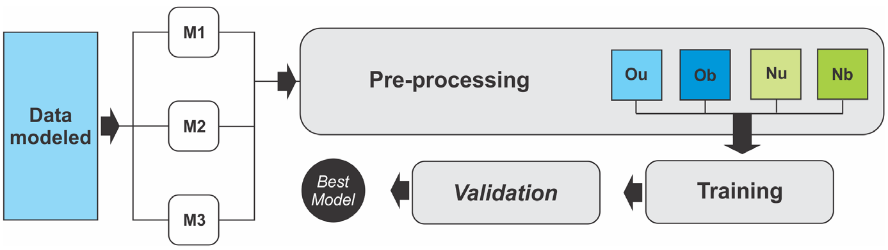

FIGURE 2 Diagram of the data processing and model validation steps. M1, model to predict the most likely case classification; M2, Model to predict pathology in different ocular segments; M3, Model to predict any binocular vision dysfunction type; Nb, balanced normalised; Nu, unbalanced normalised; Ob, original balanced; Ou, original unbalanced.

The two best algorithms for each model were subsequently subjected to fivefold cross-validation, where the training data set was randomly partitioned into five subsamples of equal size. In each iteration, one subsample was used as a validation set to test the model, while the other four subsamples were used for training. This process was repeated five times, ensuring each subsample was used once as validation data. Cross-validation is essential for assessing the generalisability of the model. It follows the recommendation of using a significant fraction of the training data to drive a well-performing model.[7] An additional purity-based feature importance analysis[13] was included in the best-performing algorithm to examine the most impactful predictors on the model.

Statistical analysis

The number of features necessary to train the machine learning algorithm is determined by the data set size. Because a fivefold cross-validation technique was applied to train the algorithms, with a sample size of 2250 eyes, the effective training size must be 1800, and it is accepted that the maximum number of parameters for optimal model performance is limited to one-tenth of the effective size.[7] In all the trained models, the number of variables was significantly lower, indicating that the sample size was sufficient for testing the ML algorithms.

The performance of the models was evaluated using five different indicators: accuracy (Ac), defined as the proportion of correct predictions in relation to the total predictions; precision (Pr) and sensitivity (Se), defined as the weighted average of the proportions of true positives in relation to the total positives (true and false, respectively) for each class; the F1 score (Fs), defined as the harmonic mean between precision and sensitivity for each class, weighted by the proportion of each class and specificity (Sp), defined as the simple average of the proportions of true negatives relative to the total true negatives for each class. In addition, for each metric, a paired *t*-test was conducted between the results of each iteration of the cross-validation of the two algorithms tested, considering a p-value < 0.05 as statistically significant.[14]

Implementation of the assistant

The implementation of the virtual assistant was carried out, encompassing the creation of the operational interface to facilitate interaction with the AI models, allowing users to consult the models in a straightforward manner without technical concerns. The process was divided into three stages: *design and prototyping*, which focused on defining the interfaces with an emphasis on usability and accessibility for vision science professionals; *project planning*, in which the technical procedures and logical developmental flow were defined with the ASP.NET (VB) (Microsoft, learn.microsoft.com/dotnet/visual-basic) for the server and JavaScript (jQuery) (OpenJS Foundation, jquery.com/) for the client and *development*, which included coding functionalities and integrating AI models with the virtual assistant, adjusting data formats until the predictions were presented.

RESULTS

Model (M1) classified patients based on 35 input features related to the *preliminary data* and *history* groups; Model (M2) proposed a subclassification of ocular disorders by analysing 65 features from the *preliminary data*, *history* and *ocular examinations* groups and Model (M3) differentiated binocular vision dysfunctions by evaluating 30 features from the *preliminary data* and *oculomotor tests* groups. Table 2 summarises the developed models and the preprocessed data used in each model for training.

The accuracy of the three developed models showed that the SVM and LR algorithms performed worst among the models. In contrast, the best accuracy of all the models was achieved with the RF algorithm, with accuracies of 94.6% (M1), 97.1% (M2) and 99.4% (M3), followed by the DT algorithm in M1 (90.0%) and the NN algorithm in M2 (95.2%) and M3 (97.3%). The best performance was observed for the balanced data, while the differences between normalised and non-normalised values were not remarkable. Table 3 summarises the accuracy achieved by each algorithm in each model.

A detailed analysis of the RF and NN algorithms revealed that the RF algorithm yielded significantly better results across the three developed models, with averages of 95%, 98% and 99% for M1, M2 and M3, respectively. In comparison, the decision tree (DT) algorithm achieved an average accuracy of 90% for M1, while the NN algorithm reached 97% for M2 and 98% for M3 (Table 4).

The importance of individual features across the three random forest models is displayed in Figure 3, where patient age consistently appears to be the most significant predictor (Figure 3b—right). The analysis also reveals that 15 predictors account for 95.4% of the relevance in Model 1, whereas Models 2 and 3 account for 72.4% and 90.6%, respectively. In these latter models, 34 (M2) and 18 (M3) features together contribute >95% of the model's importance (Figure 3a—left).

A detailed analysis of the proposed models is provided in Figure 4. Model 1 demonstrated good performance in case classification even in combinations of refractive disease, ocular disease or binocular vision dysfunction over 80% with the RF algorithm (Figure 4a). Additionally, subclassification of eye disease location by Model 2 was excellent, with agreement >98% (Figure 4b). Finally, subclassification of binocular vision dysfunction by Model 3 also showed excellent performance (Figure 4c).

Finally, a responsive web application (for desktops and smartphones) integrating the developed AI models was created. The application is freely accessible and available in three languages (English, Spanish and Brazilian Portuguese), allowing users to consult the AI models via conventional clinical terminology, with a feature that converts these terms to the formats interpreted by the models, making the use simple and intuitive (Figure 5).

DISCUSSION

AI virtual assistants are software-based systems designed to simulate human conversations and interactions, offering real-time support and information widely used in different services, including healthcare (scheduling appointments, screening symptoms, managing chronic conditions, choosing treatments based on clinical evidence, etc.).[15] Their use has improved patients' medication adherence,[16] promoted healthy lifestyles[17] and provided continuous, personalised patient support.[18] Additionally, virtual assistants help healthcare providers by automating routine tasks (reducing administrative workload, etc.) to help health professionals focus on personal interactions with patients.[18]

To the best of the author's knowledge, no previous AI-based virtual assistants have been developed for primary eye care practice. Earlier use of AI in optometry and ophthalmology has focused primarily on diagnosing posterior eye diseases,[2,4] with other minor applications in anterior eye imaging-based diagnosis[6] and ocular refraction.[7,8] Previous machine and deep learning algorithms have generally achieved similar performance to that described in this study (>95% with the RF algorithm) under different eye conditions.[19] For example, convolutional neural networks (CNNs) can detect referrable diabetic retinopathy with a sensitivity between 91.7% and 97.2%,[20,21] and have accuracies ranging from 91% to 96% in detecting age-related macular degeneration[22,23] or from 90% to 99.5% in keratoconus identification.[24–26] Other machine learning algorithms have demonstrated 97.6% and 95.4% accuracy for meibomian gland assessment[27] and tear meniscus thickness measurements,[28] respectively. These algorithms have demonstrated great utility in glaucoma disease,[19] for example, classifying colour fundus photography as early glaucoma, with accuracies ranging from 91.4% to 98.2%,[29–31] discriminating glaucomatous optic nerves from healthy eyes using OCTs with accuracies ranging from 90.2% to 96.6%,[32,33] and detecting the progression of glaucomatous nerve damage with a sensitivity of 85%.[34,35] In paediatric patients, machine learning algorithms have also exhibited high sensitivity and specificity[19] for detecting retinopathy of prematurity

TABLE 2 Descriptive summary of the different preprocessed (balanced sets) data sets used to train the AI models (M1, M2 and M3).

				Case classification			
Data set	***n***	**Age (years)**	**Sex (women/men)**	**Refractive (%)**	**Binocular vision (%)**	**Ocular/systemic disease (%)**	**Input feature**
All samples	2250	35 ± 20	63% 37%	87.6	22.4	32.4	75
Case classification (M1)							
M1							
Unbalanced							35
Total	2214	35 ± 19	63% 37%	87.7	22.4	32.4	
Trained	1771	35 ± 19	63% 37%	87.4	22.6	32.3	
Test	443	36 ± 20	65% 35%	88.9	21.7	32.7	
Balanced							
Total	8176	28 ± 19	60% 40%	50.0	50.0	50.0	
Trained	6540	28 ± 19	60% 40%	49.7	50.1	49.9	
Test	1636	29 ± 19	60% 40%	51.1	49.6	50.2	
Ocular segment disorder classification (M2)							
M2							
Unbalanced							
Total	2206	35 ± 19	63% 37%	–	–	26.0	65
Trained	1746	35 ± 19	63% 37%	–	–	25.9	
Test	442	36 ± 19	64% 36%	–	–	26.2	
Balanced							
Total	6532	40 ± 20	52% 48%	–	–	75.0	
Trained	5225	40 ± 20	52% 48%	–	–	74.4	
Test	1307	40 ± 20	52% 48%	–	–	74.4	
Binocular vision dysfunction classification (M3)							
M3							
Unbalanced							
Total	1982	35 ± 19	63% 37%	–	21.0	–	30
Trained	1585	35 ± 19	63% 37%	–	21.4	–	
Test	397	36 ± 19	65% 35%	–	19.4	–	
Balanced							
Total	6264	28 ± 18	64% 36%	–	75.0	–	
Trained	5011	28 ± 18	63% 37%	–	75.1	–	
Test	1253	27 ± 18	65% 35%	–	74.5	–	

Note: M1: Model to predict the most likely case classification; M2: Model to predict pathology in different ocular segments; M3: Model to predict any binocular vision dysfunction type.

TABLE 3 Summary of the accuracy of testing different algorithms for each developed model.

Data set version	Unbalanced		SMOTE balanced		Overall average (%)
	Original	Normalised	Original	Normalised	
M1 *n (training/test)*	*2214 (1771/443)*		*8176 (6540/1636)*		–
Random forest **(RF)**	83.1%	82.8%	**94.6%**	94.6%	*88.8*
Decision tree **(DT)**	71.3%	71.3%	89.8%	**90.0%**	*80.6*
Neural network (NN)	59.1%	78.6%	77.3%	*84.6%*	*74.9*
Support vector machines (SVM)	49.4%	54.6%	32.4%	*64.3%*	*50.2*
Logistic regression (LR)	47.9%	47.4%	39.9%	*39.2%*	*43.6*
M2 *n (training/test)*	*2206 (1764/442)*		*6532 (5225/1307)*		–
Random forest **(RF)**	89.4%	89.4%	**97.1%**	97.1%	*93.2*
Decision tree (DT)	81.0%	81.0%	92.0%	92.0%	*86.5*
Neural network **(NN)**	90.0%	91.2%	94.7%	**95.2%**	*92.8*
Support vector machines (SVM)	73.8%	75.6%	51.2%	88.8%	*72.3*
Logistic regression (LR)	73.3%	73.5%	71.1%	70.4%	*72.0*
M3 *n (training/test)*	*1982 (1585/397)*		*6264 (5011/1253)*		–
Random forest **(RF)**	95.5%	95.5%	**99.4%**	99.4%	*97.5*
Decision tree (DT)	90.9%	90.7%	95.1%	95.1%	*93.0*
Neural network **(NN)**	91.4%	93.5%	96.6%	**97.3%**	*94.8*
Support vector machines (SVM)	80.6%	80.9%	52.7%	83.8%	*74.5*
Logistic regression (LR)	80.1%	80.9%	53.2%	52.4%	*66.7*

Note: The best results are highlighted with a grey background and bold text, and the worst results are underlined. M1: Model to predict the most likely case classification; M2: Model to predict pathology in different ocular segments; M3: Model to predict any binocular vision dysfunction type.
Abbreviation: SMOTE, synthetic minority oversampling technique.

(93%),[36,37] congenital cataracts (94.4%–98.9%)[38] or amblyopia (81.6%).[39] Previous clinical decision support systems used in clinical practice by primary eye care physicians,[40] eye care practitioners[41,42] and students[43] have not been informed by AI. It is critical to differentiate this approach from large-scale generative language models (LLMs), such as ChatGPT, which are trained on vast volumes of general data aimed at generating informative text across a wide range of contexts, including vision health,[44] but are not optimised to provide decision support in specific clinical contexts. In contrast, predictive models, as proposed in this research, are specifically developed and trained using different clinical data (tabular, imaging or other data) to support accurate and safe decisions assisting healthcare decision-making.

In this study, different machine learning regression algorithms were trained to predict eye health using clinical eye examination data classified as healthy, refractive, ocular disease or binocular vision dysfunction, following technical recommendations to develop deep learning algorithms to be used in eye care practice (preprocessing, grading, training, validating and testing data sets and performance of the metrics chosen).[45] The RF algorithm provided the best model, with an accuracy >95%, which shows high potential for use in primary eye care practice and preclinical training of healthcare students.

Early detection of eye problems from refractive or pathological causes significantly impacts well-being and quality of life.[46] It is estimated that more than 1 billion people worldwide suffer from moderate or severe vision impairment or blindness from preventable or potentially correctable causes, such as refractive errors, presbyopia or cataracts.[46] Additionally, >2 billion people globally will be older than 60 years in 2050 (most in low- and middle-income countries).[47] Given the clear relationship between age and ocular diseases[48] such as diabetic retinopathy, glaucoma, age macular degeneration or cataracts, providing adequate eye care and ocular disease screening in these ageing patients will create challenges for healthcare systems, as well as increasing expenditures. The use of deep learning algorithms has been proposed for screening highly prevalent eye diseases to address these personnel and expertise shortages.[47]

Therefore, it is necessary to implement effective strategies for screening and clinical decision-making to expand and/or improve access to visual and eye health, especially in areas without access to eye care practitioners (optometrists or ophthalmologists), thereby ensuring that patients receive necessary care in a timely and accurate manner.[49] In this way, advanced technologies applied to eye care represent a valuable opportunity to overcome traditional challenges, promising tools to support clinical decision-making that assist optometrists, ophthalmologists and other health professionals, allowing a personalised and efficient approach to visual care, reducing waiting times and improving access to specialists.[50]

TABLE 4 Summary of cross-validation results for the three designed models.

M1	Random forest **(RF)***[Ob]*	Decision tree **(DT)***[Nb]*	*p*-Value
Ac	95.72% ± 0.70%	90.14% ± 0.52%	<0.001
Pr	95.73% ± 0.70%	90.09% ± 0.60%	<0.001
Se	95.72% ± 0.70%	90.14% ± 0.52%	<0.001
Fs	95.70% ± 0.70%	90.09% ± 0.56%	<0.001
Sp	95.73% ± 0.70%	90.09% ± 0.60%	<0.001
Av	**95.72% ± 0.70%**	90.11% ± 0.56%	<0.001
M2	Random forest **(RF)***[Ob]*	Neural network **(NN)***[Nb]*	*p*-Value
Ac	98.13% ± 2.05%	97.08% ± 1.93%	<0.01
Pr	98.31% ± 1.72%	97.20% ± 1.71%	<0.01
Se	98.13% ± 2.05%	97.08% ± 1.93%	<0.01
Fs	98.15% ± 2.02%	97.08% ± 1.92%	<0.01
Sp	98.31% ± 1.72%	97.20% ± 1.70%	<0.01
Av	**98.21% ± 1.91%**	97.13% ± 1.84%	<0.01
M3	Random forest **(RF)***[Ob]*	Neural network **(NN)***[Nb]*	*p*-Value
Ac	99.73% ± 0.34%	97.54% ± 1.10%	<0.01
Pr	99.74% ± 0.33%	97.56% ± 1.07%	<0.01
Se	99.73% ± 0.34%	97.54% ± 1.10%	<0.01
Fs	99.73% ± 0.34%	97.54% ± 1.10%	<0.01
Sp	99.74% ± 0.33%	97.56% ± 1.07%	<0.01
Av	**99.73% ± 0.34%**	97.55% ± 1.09%	<0.01

Note: M1: Model to predict the most likely case classification; M2: Model to predict pathology in different ocular segments; M3: Model to predict any binocular vision dysfunction type.

Abbreviations: Ac, accuracy; Av, average; Fs, F1 score; *Nb*, normalised dataset with balancing; *Ob*, original dataset with balancing; Pr, precision; Se, sensitivity; Sp, specificity.

The first developed model (Model 1) classified patients to identify any eye problem (refractive, ocular disease or binocular vision dysfunction) based on small amounts of information from a preliminary examination (including patients' ocular and familiar history, anamnesis and visual acuity) facilitating early detection. In this model, the most impactful features were age, visual acuity at far and near distances and sex, followed by symptoms of ocular pain, itching, headache, far and near complaints, burning and photophobia, which showed similar relevance (Figure 3 top). Web-based, open-access AI assistants have the potential for widespread use by different healthcare practitioners for eye health screening purposes (Table 1 and Figure 4a). Model 2 allows subclassification of the suspected eye disease (anterior, posterior or both poles) that will be highly useful in referral processes to reduce chair time and facilitate patient access to a specialist (Figure 4b). In this model, the most impactful features were age, far distance complaints and sex, followed by clinical signs in the lens and optic disc (papillary excavation and shape and the inferior–superior–nasal–temporal (ISNT) pattern) with similar relevance, as well as visual acuity at far and near distances. Other relevant features were headache, Bruck test results, hypertension, optic disc borders, conjunctival hyperaemia and near complaints (Figure 3 middle). Additionally, if binocular vision dysfunction is suspected, Model 3 allows a subclassification of accommodative or vergence-related causes that could greatly benefit patient management (Figure 4c). In this model, the most impactful features were age and oculomotor tests of the near point of convergence (with a real object, light and red filter) and cover test (at near and with prism), followed by far and near visual acuity, sex, angle Kappa, far cover test, headache, burning and far and near complaints (Figure 3 bottom). This analysis suggests that the developed models have potential for further optimisation, particularly by focusing on the most relevant features. In each model, a significant proportion of features (43% in M1 [15/35], 52% in M2 [34/65] and 60% in M3 [18/30]) account for over 95% of the total model importance. This may indicate potential for further improvement by prioritising these key predictors.

Additionally, these AI-based assistants have considerable potential to educate healthcare practitioners, particularly in primary eye care. If they are introduced in preclinical teaching, they can help students in clinical assessment and analysis training prior to honing their skills on actual patients. This should improve the experience and safety of university clinic patients, as they can be assessed by better-trained students.[51,52]

However, using AI algorithms is not free of technical and clinical considerations, and this research has limitations. First, limited-quantity, single-centre clinical data with a specific profile (school-university clinic) were used to develop and evaluate AI algorithms. This issue is relevant

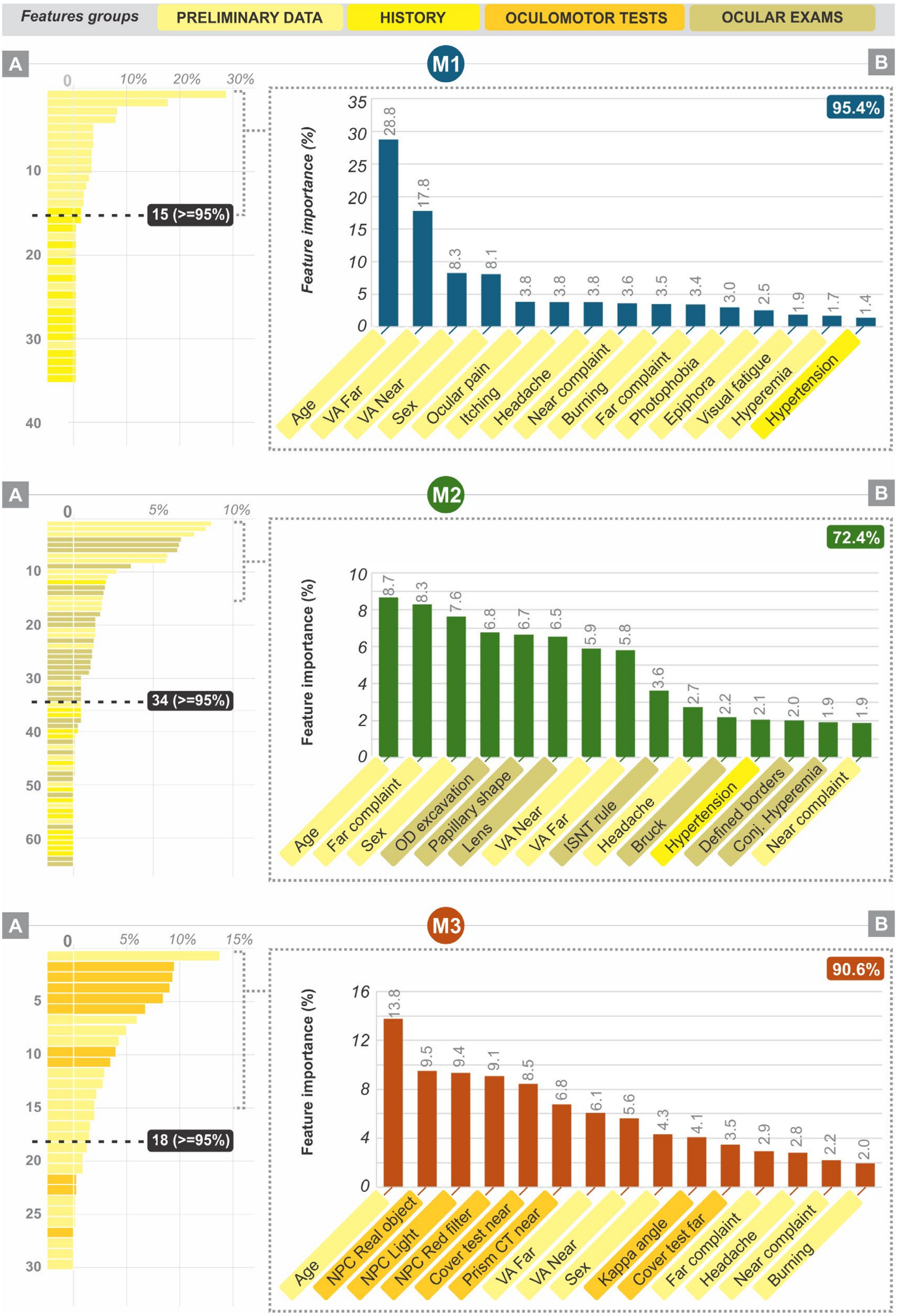

FIGURE 3 Summary of the most impactful features in each model. The top panel presents the Model (M1) used to predict the most likely case classification, the middle Model (M2) used to predict pathology in different ocular segments and the bottom Model (M3) used to predict any binocular vision dysfunction type. Left (a): An ordered overview of feature importance is presented in each model (dashed line marking the point where at least 95% importance is reached), and right (b): A breakdown of the top 15 features most relevant in each model is presented. Conj, conjunctiva; CT, cover test; ISNT, inferior, superior, nasal, temporal; NPC, near point of convergence; OD, optic disc; VA, visual acuity.

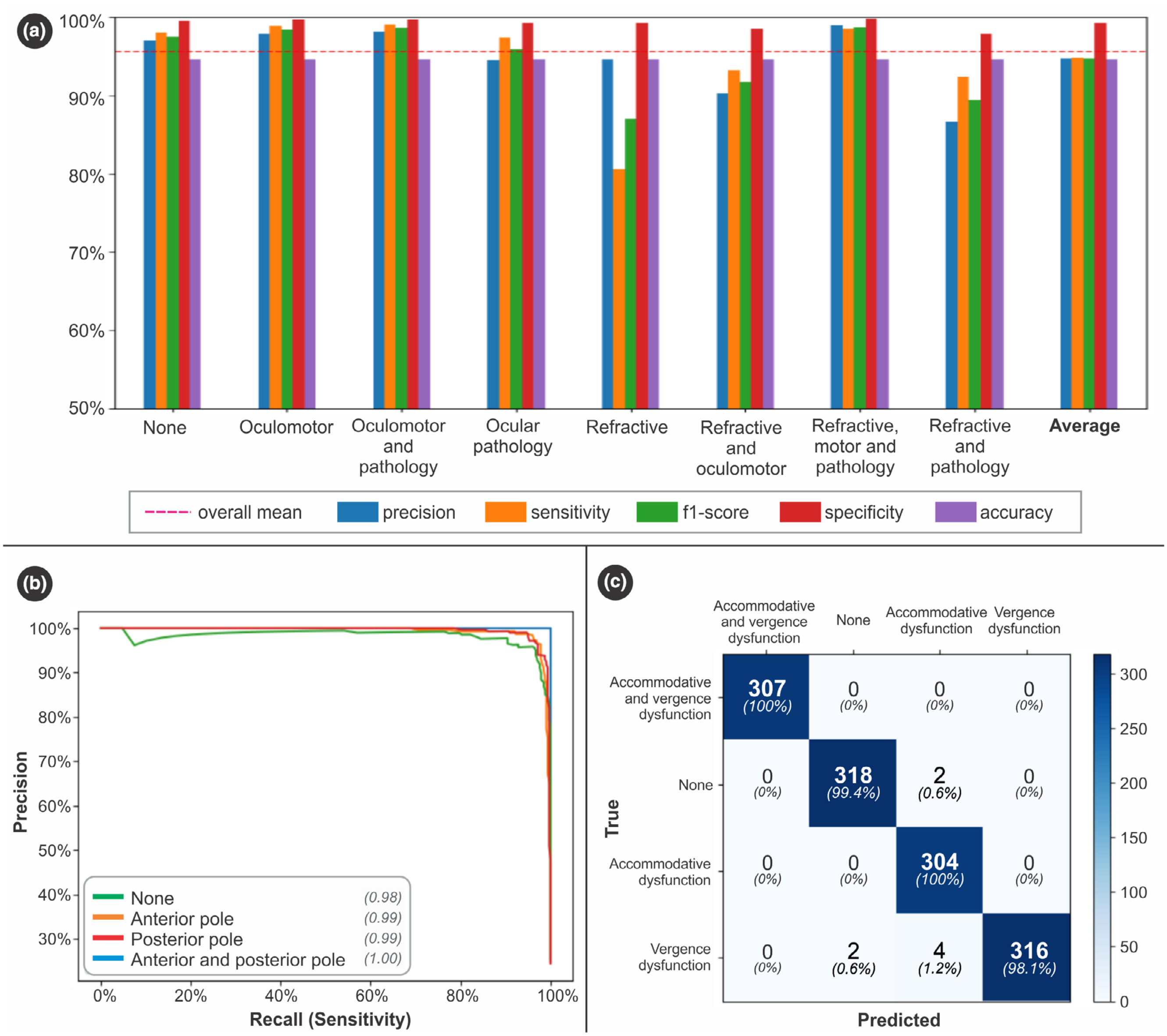

FIGURE 4 Summary of the random forest (RF) algorithm performance metrics for subclassification. (a) Results for Model 1 subject subclassification. (b) Precision–recall curve of Model 2 for segment subclassification of eye diseases. (c) Confusion matrix of Model 3 for subclassification of binocular vision dysfunction.

because the generalisability of the developed model could require further evaluation. Therefore, additional research with larger samples or patients, taking into account multicentre collaboration and multiethnic populations, will be needed to assess the actual real-world clinical applicability of the developed assistant. Such research is needed to determine the accuracy, reliability, robustness and stability of the algorithm's performance, generalisability (ruling out overfitting) and algorithm clinical utility in a real-time environment, by evaluating temporal validation sets with new longitudinal and external geographic data sets, as well as usability and cost-effectiveness validation.[53] Eyecare clinical data collection is not standardised, and differences between practitioners/centres could affect data quality. Standardising eye examination data collection, modelling and preprocessing for use in AI algorithms could facilitate data sharing, increasing network collaboration that must follow regulations and state privacy rules, guaranteeing data anonymisation and ensuring patients' privacy (obtaining patients' consent before data sharing). Additionally, collecting large amounts of multicentre origin data requires technical solutions to guarantee data storage and management. Finally, some ethical issues have been described[54–56] regarding the use of deep learning applications in healthcare practice (ranging from racial bias in algorithm development to liability for misdiagnosis); thus, the use of any AI-based assistant must not be perceived as competent because it does not substitute for an eye care practitioner, and the specific interest of the patient must remain the priority.

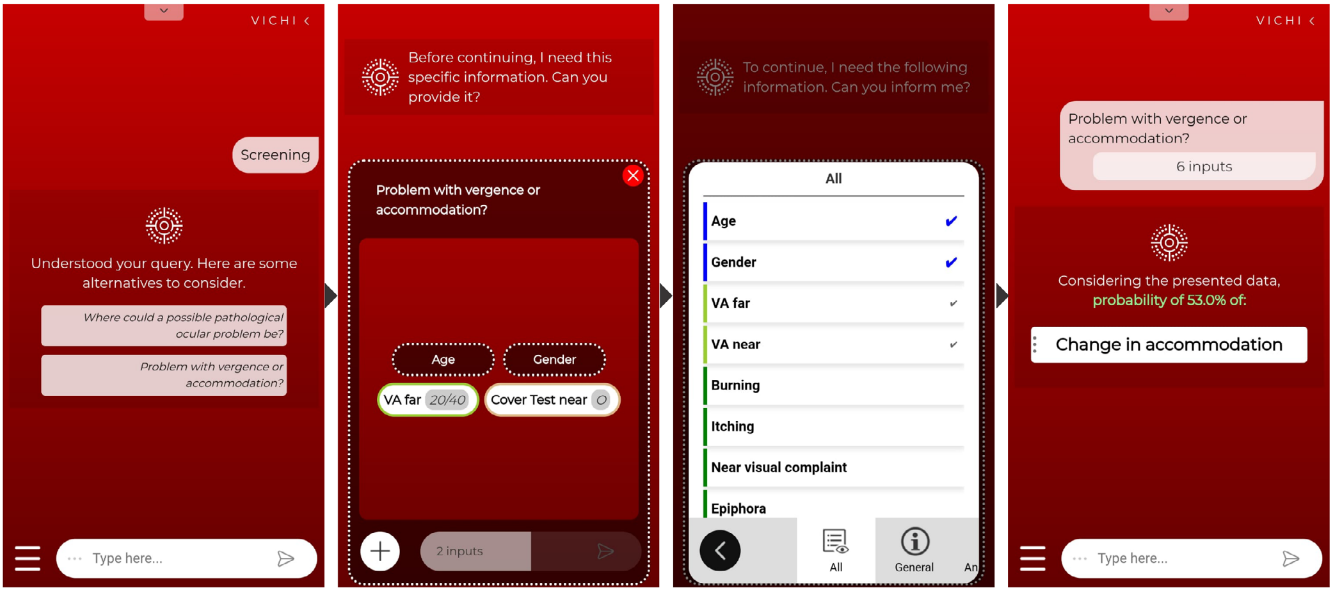

FIGURE 5 Interface of the web-based application designed to use the developed artificial intelligence (AI) algorithms. Vision Care Helper Intelligence (VICHI) is open access and available at https://www.visioncare.digital/vichi. The figure shows a sequence of screens, starting with the search for available solutions, going through the input data entry and modulation (conversion of nomenclature to model data format) and ending with the prediction result of the consulted model.

Finally, future integration of new models could expand AI-based assistants to other eye areas, such as dry eye and myopia control. In addition, it is possible to expand the database and test other modelling techniques, including deep learning libraries and images as data input. This would allow for integration with health devices and the development of other applied technological innovations.

CONCLUSION

A virtual assistant based on three different AI models (the RF algorithm) has been developed with potential use in primary eye care practice to predict patient classification, eye disorders or binocular vision dysfunction with an accuracy >95%. This virtual assistance can aid in screening for ocular problems, support primary eye care practitioners with their patient load and aid in preclinical eye care education programmes. There are still challenges to overcome before widespread dissemination of AI-based virtual assistants in primary eye care, such as standardisation of clinical data collection, development of clinical findings repositories that ensure data integration and patient confidentiality and further validation research with different patient samples. However, the developed AI-based assistants show that the technology is accessible and can be implemented with conventional clinical data without needing large image sets.

AUTHOR CONTRIBUTIONS

Leandro Stuermer: Conceptualization (equal); data curation (equal); formal analysis (equal); funding acquisition (equal); investigation (equal); methodology (equal); resources (equal); software (lead); validation (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal). **Sabrina Braga:** Conceptualization (equal); data curation (equal); formal analysis (equal); funding acquisition (equal); investigation (equal); methodology (equal); resources (equal); writing – original draft (equal); writing – review and editing (equal). **Raul Martin:** Conceptualization (equal); data curation (equal); formal analysis (equal); funding acquisition (equal); investigation (equal); methodology (equal); project administration (equal); resources (equal); supervision (lead); validation (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal). **James S. Wolffsohn:** Conceptualization (equal); investigation (equal); methodology (equal); resources (equal); writing – review and editing (equal).

ACKNOWLEDGEMENTS

Preliminary results of this AI simulator use were presented at the OPTOM2024 meeting in Madrid, Spain, April, 2024.

FUNDING INFORMATION

This research has not received any funding.

CONFLICT OF INTEREST STATEMENT

None of the authors has a financial or proprietary interest in any material or method mentioned. Additionally, the authors declare that this research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

ORCID

Leandro Sturmer https://orcid.org/0000-0002-6433-2333
Sabrina Braga https://orcid.org/0009-0005-7285-3509
Raul Martin https://orcid.org/0000-0001-7156-619X
James S. Wolffsohn https://orcid.org/0000-0003-4673-8927

REFERENCES

1. Stuermer L, Martin R. Characterization of technologies in digital health applied in vision care. *Aust J Optom*. 2022;15:S70–S81.
2. de la Torre-Díez I, Martínez-Pérez B, López-Coronado M, Díaz JR, López MM. Decision support systems and applications in ophthalmology: literature and commercial review focused on mobile apps. *J Med Syst*. 2015;39:174. https://doi.org/10.1007/s10916-014-0174-2
3. Oke I, VanderVeen D. Machine learning applications in pediatric ophthalmology. *Semin Ophthalmol*. 2021;36:210–7.
4. Rizzo S, Savastano A, Lenkowicz J, Savastano MC, Boldrini L, Bacherini D, et al. Artificial intelligence and OCT angiography in full thickness macular hole. New developments for personalized medicine. *Diagnostics*. 2021;11:2319. https://doi.org/10.3390/diagnostics11122319
5. Mahesh Kumar SV, Gunasundari R. Computer-aided diagnosis of anterior segment eye abnormalities using visible wavelength image analysis based machine learning. *J Med Syst*. 2018;42:128. https://doi.org/10.1007/s10916-018-0980-z
6. Ting DSJ, Foo VH, Yang LWY, Sia JT, Ang M, Lin H, et al. Artificial intelligence for anterior segment diseases: emerging applications in ophthalmology. *Br J Ophthalmol*. 2021;105:158–68.
7. Espinosa J, Pérez J, Villanueva A. Prediction of subjective refraction from anterior corneal surface, eye lengths, and age using machine learning algorithms. *Transl Vis Sci Technol*. 2022;11:8. https://doi.org/10.1167/tvst.11.4.8
8. Yamauchi T, Tabuchi H, Takase K, Masumoto H. Use of a machine learning method in predicting refraction after cataract surgery. *J Clin Med*. 2021;10:1103. https://doi.org/10.3390/jcm10051103
9. Wawer Matos PA, Reimer RP, Rokohl AC, Caldeira L, Heindl LM, Große HN. Artificial intelligence in ophthalmology—status quo and future perspectives. *Semin Ophthalmol*. 2023;38:226–37.
10. Potdar K, Pardawala TS, Pai CD. A comparative study of categorical variable encoding techniques for neural network classifiers. *Int J Comput Appl*. 2017;175:7–9.
11. Chahid I, Elmiad AK, Badaoui M. Data preprocessing for machine learning applications in healthcare: a review. Proceedings—SITA 2023: 2023 14th International Conference on Intelligent Systems: Theories and Applications. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers Inc.; 2023. https://doi.org/10.1109/SITA60746.2023.10373591.
12. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res*. 2002;16:321–57.
13. Géron A. Hands-on machine learning with Scikit-learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems. Newton, Massachusetts: O'Reilly Media; 2019.
14. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. *Inf Process Manag*. 2009;45:427–37.
15. Curtis RG, Bartel B, Ferguson T, Blake HT, Northcott C, Virgara R, et al. Improving user experience of virtual health assistants: scoping review. *J Med Internet Res*. 2021;23:e31737. https://doi.org/10.2196/31737
16. Roca S, Lozano ML, García J, Alesanco Á. Validation of a virtual assistant for improving medication adherence in patients with comorbid type 2 diabetes mellitus and depressive disorder. *Int J Environ Res Public Health*. 2021;18:12056. https://doi.org/10.3390/ijerph182212056
17. Aggarwal A, Tam CC, Wu D, Li X, Qiao S. Artificial intelligence-based Chatbots for promoting health behavioral changes: systematic review. *J Med Internet Res*. 2023;25:e40789. https://doi.org/10.2196/40789
18. Maher C, Singh B, Wylde A, Chastin S. Virtual health assistants: a grand challenge in health communications and behavior change. *Front Digit Health*. 2024;6:1418695. https://doi.org/10.3389/fdgth.2024.1418695
19. Srivastava O, Tennant M, Grewal P, Rubin U, Seamone M. Artificial intelligence and machine learning in ophthalmology: a review. *Indian J Ophthalmol*. 2023;71:11–7.
20. Heydon P, Egan C, Bolter L, Chambers R, Anderson J, Aldington S, et al. Prospective evaluation of an artificial intelligence-enabled algorithm for automated diabetic retinopathy screening of 30 000 patients. *Br J Ophthalmol*. 2020;105:723–8.
21. Wang X, Ji Z, Ma X, Zhang Z, Yi Z, Zheng H, et al. Automated grading of diabetic retinopathy with ultra-widefield fluorescein angiography and deep learning. *J Diabetes Res*. 2021;2021:250. https://doi.org/10.1155/2021/2611250
22. Burlina PM, Joshi N, Pekala M, Pacheco KD, Freund DE, Bressler NM. Automated grading of age-related macular degeneration from color fundus images using deep convolutional neural networks. *JAMA Ophthalmol*. 2017;135:1170. https://doi.org/10.1001/jamaophthalmol.2017.3782
23. Vaghefi E, Hill S, Kersten HM, Squirrell D. Multimodal retinal image analysis via deep learning for the diagnosis of intermediate dry age-related macular degeneration: a feasibility study. *J Ophthalmol*. 2020;2020:7493419. https://doi.org/10.1155/2020/7493419
24. Lavric A, Valentin P. KeratoDetect: keratoconus detection algorithm using convolutional neural networks. *Comput Intell Neurosci*. 2019;2019:8162567. https://doi.org/10.1155/2019/8162567
25. Kamiya K, Ayatsuka Y, Kato Y, Fujimura F, Takahashi M, Shoji N, et al. Keratoconus detection using deep learning of colour-coded maps with anterior segment optical coherence tomography: a diagnostic accuracy study. *BMJ Open*. 2019;9:e031313. https://doi.org/10.1136/bmjopen-2019-031313
26. Shi C, Wang M, Zhu T, Zhang Y, Ye Y, Jiang J, et al. Machine learning helps improve diagnostic ability of subclinical keratoconus using Scheimpflug and OCT imaging modalities. *Eye Vis*. 2020;7:48. https://doi.org/10.1186/S40662-020-00213-3
27. Wang J, Yeh TN, Chakraborty R, Yu SX, Lin MC. A deep learning approach for meibomian gland atrophy evaluation in meibography images. *Transl Vis Sci Technol*. 2019;8:37. https://doi.org/10.1167/TVST.8.6.37
28. Stegmann H, Werkmeister RM, Pfister M, Garhöfer G, Schmetterer L, dos Santos VA. Deep learning segmentation for optical coherence tomography measurements of the lower tear meniscus. *Biomed Opt Express*. 2020;11:1539–54.
29. Girard F, Hurtut T, Kavalec C, Cheriet F. Atlas-based score for automatic glaucoma risk stratification. *Comput Med Imaging Graph*. 2021;87:101797. https://doi.org/10.1016/j.compmedimag.2020.101797
30. Orlando JI, Fu H, Barbossa Breda J, Van Keer K, Bathula DR, Diaz-Pinto A, et al. REFUGE challenge: a unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. *Med Image Anal*. 2020;59:570. https://doi.org/10.1016/j.media.2019.101570
31. Thompson AC, Jammal AA, Medeiros FA. A deep learning algorithm to quantify neuroretinal rim loss from optic disc photographs. *Am J Ophthalmol*. 2019;201:9–18.
32. Ran AR, Cheung CY, Wang X, Chen H, Luo LY, Chan PP, et al. Detection of glaucomatous optic neuropathy with spectral-domain optical coherence tomography: a retrospective training and validation deep-learning analysis. *Lancet Digit Health*. 2019;1:e172–e182.
33. García G, Colomer A, Naranjo V. Glaucoma detection from raw SD-OCT volumes: a novel approach focused on spatial dependencies. *Comput Methods Programs Biomed*. 2021;200:5855. https://doi.org/10.1016/j.cmpb.2020.105855
34. Medeiros FA, Jammal AA, Mariottoni EB. Detection of progressive glaucomatous optic nerve damage on fundus photographs with deep learning. *Ophthalmology*. 2020;128:383–92.

35. Ajitha S, Akkara JD, Judy MV. Identification of glaucoma from fundus images using deep learning techniques. *Indian J Ophthalmol*. 2021;69:2702–9.
36. Brown JM, Campbell JP, Beers A, Chang K, Ostmo S, Chan RVP, et al. Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks. *JAMA Ophthalmol*. 2018;136:803–10.
37. Tong Y, Lu W, Deng Q, Chen C, Shen Y. Automated identification of retinopathy of prematurity by image-based deep learning. *Eye Vis*. 2020;7:40. https://doi.org/10.1186/S40662-020-00206-2
38. Long E, Chen J, Wu X, Liu Z, Wang L, Jiang J, et al. Artificial intelligence manages congenital cataract with individualized prediction and telehealth computing. *NPJ Digit Med*. 2020;3:112. https://doi.org/10.1038/S41746-020-00319-X
39. Chun J, Kim Y, Shin KY, Han SH, Oh SY, Chung TY, et al. Deep learning-based prediction of refractive error using photorefraction images captured by a smartphone: model development and validation study. *JMIR Med Inform*. 2020;8:e16225. https://doi.org/10.2196/16225
40. López MM, López MM, de la Torre DI, Jimeno JCP, López-Coronado M. mHealth app for iOS to help in diagnostic decision in ophthalmology to primary care physicians. *J Med Syst*. 2017;41:81. https://doi.org/10.1007/s10916-017-0731-6
41. Ebrahimi F, Ayatollahi H, Aghaei H. A clinical decision support system for diagnosing and determining severity of dry eye disease. *Eye*. 2023;37:1619–24.
42. Guo X, Swenor BK, Smith K, Boland MV, Goldstein JE. Developing an ophthalmology clinical decision support system to identify patients for low vision rehabilitation. *Transl Vis Sci Technol*. 2021;10:24. https://doi.org/10.1167/tvst.10.3.24
43. López MM, López MM, de la Torre DI, Jimeno JCP, López-Coronado M. A mobile decision support system for red eye diseases diagnosis: experience with medical students. *J Med Syst*. 2016;40:151. https://doi.org/10.1007/s10916-016-0508-3
44. Biswas S, Davies LN, Sheppard AL, Logan NS, Wolffsohn JS. Utility of artificial intelligence-based large language models in ophthalmic care. *Ophthalmic Physiol Opt*. 2024;44:641–71.
45. Ting DSW, Peng L, Varadarajan AV, Keane PA, Burlina PM, Chiang MF, et al. Deep learning in ophthalmology: the technical and clinical considerations. *Prog Retin Eye Res*. 2019;72:100759. https://doi.org/10.1016/j.preteyeres.2019.04.003
46. Assi L, Chamseddine F, Ibrahim P, Sabbagh H, Rosman L, Congdon N, et al. A global assessment of eye health and quality of life: a systematic review of systematic reviews. *JAMA Ophthalmol*. 2021;139:526–41.
47. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol*. 2019;103:167–75.
48. Wang MTM, Muntz A, Lim J, Kim JS, Lacerda L, Arora A, et al. Ageing and the natural history of dry eye disease: a prospective registry-based cross-sectional study. *Ocul Surf*. 2020;18:736–41.
49. Burton MJ, Ramke J, Marques AP, Bourne RRA, Congdon N, Jones I, et al. The Lancet Global Health Commission on Global Eye Health: vision beyond 2020. *Lancet Glob Health*. 2021;9:e489–e551.
50. Li JPO, Liu H, Ting DSJ, Jeon S, Chan RVP, Kim JE, et al. Digital technology, tele-medicine and artificial intelligence in ophthalmology: a global perspective. *Prog Retin Eye Res*. 2021;82:100900. https://doi.org/10.1016/j.preteyeres.2020.100900
51. Han ER, Yeo S, Kim MJ, Lee YH, Park KH, Roh H. Medical education trends for future physicians in the era of advanced technology and artificial intelligence: an integrative review. *BMC Med Educ*. 2019;19:460. https://doi.org/10.1186/s12909-019-1891-5
52. Nagi F, Salih R, Alzubaidi M, Shah H, Alam T, Shah Z, et al. Applications of artificial intelligence (AI) in medical education: a scoping review. *Stud Health Technol Inform*. 2023;305:648–51.
53. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. *Future Healthc J*. 2021;8:e188–e194.
54. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: addressing ethical challenges. *PLoS Med*. 2018;15:e1002689. https://doi.org/10.1371/journal.pmed.1002689
55. Char DS, Shah NH, Magnus D. Implementing machine learning in health care—addressing ethical challenges. *N Engl J Med*. 2018;378:981–3.
56. Keskinbora K, Güven F. Artificial intelligence and ophthalmology. *Turk J Ophthalmol*. 2020;50:37–43.

How to cite this article: Sturmer L, Braga S, Martin R, Wolffsohn JS. Artificial intelligence virtual assistants in primary eye care practice. *Ophthalmic Physiol Opt*. 2025;45:437–449. https://doi.org/10.1111/opo.13435