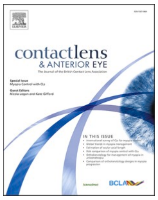
Contents lists available at ScienceDirect

Contact Lens and Anterior Eye

journal homepage: www.elsevier.com/locate/clae



Machine learning techniques in the diagnosis of meibomian glands related alterations from clinical indicators

Elena Fernández-Jiménez [a], Elena Diz-Arias [b,*], Jose A. Gomez-Pedrero [b], Assumpta Peral [a]

[a] Department of Optometry and Vision, Faculty of Optics and Optometry, University Complutense of Madrid, Spain
[b] Optics Department, Faculty of Optics and Optometry, University Complutense of Madrid, Spain

ARTICLE INFO

Keywords:
Artificial intelligence
Contact lens
Diagnosis
Machine learning
Meibomian glands

ABSTRACT

Purpose: There is no "Gold Standard" test that allows the diagnosis and classification of alterations and pathologies related to Meibomian glands (MG). A global evaluation of objective and subjective tests is necessary to determine the final diagnosis.
In recent years, Artificial Intelligence (AI) and Machine Learning (ML) techniques have experienced great progress in the field of health sciences, as promising techniques for predicting pathologies from data and images. The main objective of this study has been to train ML classifiers for the classification of three groups of participants with and without MG alterations. The secondary objective was to study the precision, specificity and sensitivity of the ML classifiers.
Methods: A retrospective comparative study was carried out on a total of 135 participants (control, contact lens wearers and MG pathology). Symptomatology and clinical tests were performed to evaluate the ocular surface and adnexa. The numerical data obtained from these tests were used to train ML classifiers and the top 5 were subsequently verified.
Results: Accuracies greater than 76 % were obtained for the training group and greater than 79 % for the verification group, for five classifiers previously described in Matlab. Subspace KNN was the classifier with the highest accuracies, specificities and sensitivities, these being moderate-high (greater than 79 %).
Conclusions: ML algorithms can be useful for classifying groups of participants with various meibomian gland disorders using clinical data. A large number of participants is needed for reliable diagnostic accuracy.

1. Introduction

There are different objective and subjective tests that allow the diagnosis and classification of alterations and pathologies related to Meibomian glands (MG). However, there is no "Gold Standard" test, nor a direct correlation between clinical tests and the symptoms commonly described by patients. Therefore, it is necessary to carry out a general evaluation and perform a variety of tests by an expert in the field which determines the diagnosis of alterations or pathologies related to MG [1,2].

The alterations in the morphology of the MG with more prevalence are those related to Meibomian gland dysfunction (MGD). Also, the daily prolonged use of contact lenses (CL) can result in changes in the MG secretion, particularly in the composition and quality of the tear film layer. In addition, it has been reported [3,4] that contact lens usage also has an impact on the morphology and function of MG.

Nowadays, the use of Artificial Intelligence, through Deep Learning and Machine Learning techniques represents an advance in the world of biomedical sciences by providing data-driven, objective prediction tools. These techniques help professionals from different areas to develop objective prediction data-based methods, whether numerical or images, for the diagnosis and monitoring of diseases [5]. Historically, before the introduction of generative artificial intelligence techniques, the use of artificial intelligence techniques in clinical research was primarily focused on classification and regression tasks. Classification is particularly useful in clinical fields such as computer aided diagnosis, analysis of medical images, monitoring of clinical essays, etc. Regression enables modelling complex biomedical systems from experimental data. Machine Learning algorithms are particularly suited for analysing datasets with moderate amounts of data presenting medium to low complexity [6,7]. In clinical applications, machine learning algorithms have been employed in assisting autistic patients' triage [8],

---

* Corresponding author at: Street Arcos del Jalón 118 San Blas – Canillejas, 28037 Madrid, Spain.
E-mail address: elenadiz@ucm.es (E. Diz-Arias).

https://doi.org/10.1016/j.clae.2025.102479
Received 12 December 2024; Received in revised form 13 July 2025; Accepted 15 July 2025

enhancement of biomedical imaging analysis and interpretation [9,10], analysing bio-signals through Brain and Body Machine Interfaces [11], and may other applications [12]. In ophthalmology, deep learning and machine learning algorithms range from image analysis of cornea diseases [13] to diagnosis of retinal conditions from OCT images [14].

Focusing on the application of deep learning and machine leaning techniques in the context of Dry Eye Disease and tear alterations, recent studies have developed artificial intelligence models to detect diseases related to MGD [15–21]. These studies use artificial Intelligence to analyse meibography images, automatically segmenting glands, and detecting morphological anomalies related to MG dysfunction such as gland length, width, and atrophy. Other studies employ machine learning algorithms to predict outcomes of new therapies for Dry Eye Disease [22] and to automatically grade MG loss from meibography images [23]. Clearly, artificial intelligence applications in the diagnosis and treatment of Dry Eye Disease and similar pathologies represent a promising research field.

Most previous work has focused on machine learning and deep learning techniques for analysing MG images. Recently, Graham et al. [24] introduced a novel approach that combines meibography image analysis with clinical symptom data to predict MG dysfunction, aqueous deficiency, and blepharitis. The Graham et al. [24] study highlights the importance of gland morphology in predicting signs and symptoms of dry eye disease and suggests that combining clinical data with gland morphology analysis could improve automatic diagnostic accuracy, as is standard in manual diagnosis.

A limitation of the Graham et al. [24] study, however, is the complexity of the analysis, which relies on a deep learning model trained to extract gland morphology from images. Designing and training such networks require vast amounts of data, large patient numbers, and substantial expertise in programming. In contrast, machine learning algorithms generally require less data and can be implemented using accessible software such as MATLAB or BigML [BIGML], which offers user-friendly applications that don't require extensive programming knowledge. Given that most clinical data on dry eye are obtained from standardized tests routinely carried out by practitioners and since Graham et al. [24] identify gland loss—a feature easily assessed in clinical exams—as a significant morphological parameter. For this reason, the proposed approach to automatically classify patients based on standard clinical data is believed to be feasible and beneficial.

The method presented in this paper offers several significant benefits for clinical practice and research. First, it provides an objective and data-driven classification of patients, which can significantly reduce inter-observer variability in MGD diagnosis, particularly in early stages before more pronounced symptoms appear. Second, it serves as a diagnostic aid for clinicians, enhancing the diagnostic process by identifying complex patterns within high-dimensional clinical data that might not be immediately apparent. Third, it can provide a tool for less experienced practitioners, such as early graduates or general practitioners in remote or underserved areas, to systematically assess patients with dry eye symptoms or discomfort from contact lens use, serving as an educational and preliminary diagnostic aid. Fourth, by streamlining the diagnostic workflow, it contributes to more efficient patient management and can potentially guide personalized diagnostics and treatments. Finally, this work lays foundational groundwork for future clinical decision support systems; the combination and customization of different AI technologies, including this machine learning model, can form the basis for powerful comprehensive tools, especially given the emergence of generative artificial intelligence algorithms like Large Language Models. The usefulness of the present work lies in demonstrating the feasibility of Machine Learning as a practical tool to assist clinicians in the classification of patients based on diverse test results, by, for example, indicating which machine learning models could be more promising for classifying clinical data. While training a machine learning model can be a complex task, this is increasingly mitigated by the availability of user-friendly platforms that allow for the design, interaction, and usage of machine learning models without extensive programming knowledge [BIGML].

Therefore, the main objective of the present study has been testing the feasibility of using some of the machine learning classifiers described in the literature to classify a set of subjects in three groups, subjects free of MG pathology, subjects presenting MG pathology and CL wearers from the numerical clinical data obtained through the performance of the tests used for the assessment of MG [5]. The second objective has been to study the accuracy, specificity, and sensitivity of machine learning classifiers in three different groups of participants.

2. Materials and methods

A retrospective comparative study was carried out from the data collected in previous clinical trials. The results were taken from 135 subjects including both eyes in the analysis so a total of 270 eyes were used to train and verify the machine learning classifiers (130 control, 68 with MGD and 72CL users). The database was obtained from previous studies approved by the Ethics Committee of the Hospital Clínico San Carlos, Madrid, Spain C.P. EO4/DPI2016-75272-R- C.I. 16/550-E and C. I. 21/176-E which adhered to the procedures described by the Declaration of Helsinki.

The participants were classified into three groups: a control group made up of healthy participants, another group with pathology made up of participants with alterations in MG, specifically MGD, and a third group made up of CL wearers. The inclusion criteria for the control group were being older than 18 years of age, not being CL users, having a score on the OSDI test <15 points, and not having a confirmed diagnosis of MG-related disorders [25]. For the pathology group, the inclusion criteria were being over than 18 years of age, having a score on the OSDI test ≥15 points, and having a diagnosis of MGD confirmed by a specialist [25]. The diagnosis was made by the ophthalmologist from the Clinic of Optometry of the Faculty of Optics and Optometry from UCM Dr. Cristina Niño Rueda. The ophthalmology diagnosis was based on the analysis of ocular symptomatology, observation of the eyelid structures and adnexa (irregularity and eyelid vascularization, eyelashes, tarsal conjunctiva), meibography images and meiboscores.

The ophthalmologist verified that all participants from the group with pathology had symptoms and signs compatible with dry eye and had a degree of glandular loss ≥25 % which is equivalent to grade 2 on the scale described by Pult et al. [26]. The inclusion criteria for the CL wearers group were: being over 18 years of age, being a user of monthly soft CL for a period of more than two years, and not having a confirmed diagnosis of MG disorders. The criterion for duration of CL wear was based on the articles by Alghamdi et al. [27] and Arita et al. [4] which showed that changes in MG morphology and function are not reversible after 2 years of wear. The following participants were excluded from the study: those with presence of ocular inflammation, disease that compromised ocular health, participants with a history of ocular trauma or surgery (including refractive surgery), and those with ocular or systemic medication, in addition to those who did not meet the inclusion criteria for each group. The sequence of the clinical tests performed on each participant is described in Table 1. In Table 1 K5M stands for Keratograph 5M (Oculus GmbH, Germany) and ETDRS for Early Diabetic Retinopathy Study, which was the chart used for assessing the patients' visual acuity.

2.1. Machine learning algorithms

Machine Learning is defined as a set of programmed algorithms that receive and analyse input data to predict output values. Its main feature is that the performance and optimization of the algorithms increase with experience that is, a greater number of input data implies better performance for predicting output values within an acceptable range. Therefore, a machine learning algorithm can be contrived as an

Table 1
Description of the clinical tests carried out on the participants.

Visit 0
Information on the characteristics of the study and signing of the informed consent
Inclusion/exclusion criteria
Visit 1
Ophthalmological evaluation
Symptomatology test (OSDI and VAS) [25,28]
Clinical measures
- Visual Acuity (ETDRS 4 m) [29]
- Tear meniscus height (K5M) [30]
- Tear Break Up Time (TBUT) [31]
- Slit lamp examination of the ocular surface and adnexa [32]
o Corneal staining
o Conjunctival staining
o Eyelid irregularity
o Eyelid vascularity
o MG exploration [32]
▪ Meibum quality
▪ Meibum expression
- Meibography of both eyelids with K5M [32]

algorithm designed to carry out a task whose performance increases with the acquired experience [5,33].

Machine Learning Algorithms can be categorized into classification algorithms and regression algorithms. Classification algorithms aim to create models capable of taking input data, identifying patterns, and making predictions based on these characteristics. In contrast, regression models predict continuous numerical output while classification models predict discrete categorical output. The most widely used learning processes include supervised, unsupervised, and reinforcement learning [34–36].

2.2. Classification of numerical data using machine learning techniques

The Matlab® (Mathworks Inc, United States) program was used to classify the data collected from previous studies. Specifically, the Classification Learner application was used for this task [37]. This app trains data classification models and allows users to perform a supervised machine learning sequence using multiple classifiers, analyse the data, train models, evaluate results or select features. With the Classification Learner app, the user can simultaneously train several machine learning models, grouped by families of algorithms such as classification trees. This allows for the comparison of the performance of various machine learning algorithms using the same data set during both the training and validation stages. After training and validation, the app sorts the machine learning algorithms based on the accuracy, enabling the user to select the best-suited model for their data.

A total of 24 machine learning classifiers already programmed in Matlab® were trained using this app. To do so, dataset was split into two subsets: training and test. The training set was used to train and validate the models; to do so, Matlab® Classification Learner app internally partitioned the training set into a proper training set and a validation set. Therefore, this initial training set will henceforth be referred to as the train/validation set. Once trained, the five classifiers that yielded the highest accuracy in the training/validation set were tested with the test set.

2.3. Data used for training and testing of machine learning classifiers

Forty-four numerical clinical variables were used for the training and testing of the machine learning classifiers. The variables were those obtained from the symptomatology and ocular surface tests, and from grading of meibographic images. The description of the 44 variables used for the training and testing of the machine learning classifiers is shown in Table 2.

3. Data analysis

To analyse the accuracies of the classifiers, the Matlab® program and the Statistics and Machine Learning Toolbox were used [38]. Given the high number of variables compared to the number of cases, a Principal Component Analysis (PCA) was carried out to try to detect correlations between variables and thus reduce the dimensionality (number of variables) of the problem [39]. This analysis gave a negative result, meaning that no linear combination of variables was found that could substitute several variables with a single linear combination of variables. Thus, all the variables were used in the training and testing of the machine learning classifiers. The accuracies of the 24 machine learning classifiers were calculated, and the five with best accuracies were selected. These top five were later verified against the test data set. Due to the multiclass nature of the data, the confusion matrices obtained for the five best classifiers were analysed through the micro-averaging analysis, from which the precision, sensitivity and specificity were calculated, to evaluate the effectiveness of the classifiers [40].

4. Results

A total of 282 eyes were recruited for the study, of which 270 were finally included in the study. Fig. 1 shows the flow chart of the total number of participants recruited, as well as those excluded and those who made up each study group.

As can be seen in Fig. 1, 12 participants out of a total of 282 were excluded because six of them had missing data/variables, two were contact lens users and were diagnosed with blepharitis, and four were under pharmacological treatments not compatible with the study.

Of the total of 270 eyes included, 216 (80 %) were used to train and verify the 24 machine learning classifiers and the remaining 54 (20 %) for testing the 5 best classifiers, that is, those that obtained the highest precision for classifying each participant correctly in each group. Therefore, the whole dataset of 270 observations was effectively divided into the three customary sets of machine learning: a training set (N = 184 observations), a validation set (N = 32 observations), and a test set (N = 54 observations). Fig. 2 shows the distribution of the participants in the different study groups for the training/validation and test sets.

Table 2
Description of the 44 clinical numerical variables used for training and testing of the machine learning classifiers.

44 Variables used for the analysis of the results	
Demographic variables	Sex △ Age
Symptomatology test	OSDI [25] VAS: pain, dryness, irritation, burning, itching, photophobia, foreign body sensation [28]
Clinical measures	
- Visual acuity (ETDRS a 4 m) [29]	Palpebral irregularity ◊
- Tear meniscus height (K5M) [30]	Palpebral vascularization ◊
- TBUT [31]	Visibility of the MG orifices ◊
- Slit lamp examination of the ocular surface and adnexa [32]	Plugging of the MG orifices ◊ Corneal staining: central, superior, inferior, nasal and temporal.
- MG exploration [32]	
- Meibography with K5M [32] ◊	Conjunctival staining: superior, inferior, superior temporal, inferior temporal, superior nasal, inferior nasal. Conjunctival redness Limbal redness Eyelid redness ◊ Eyelid roughness ◊
	Meibum quality ◊ Meibum expression ◊

△ Categorical variables were converted to numerical variables before processing the data, assigning a number to each category. ◊ Data were taken on each eyelid separately.

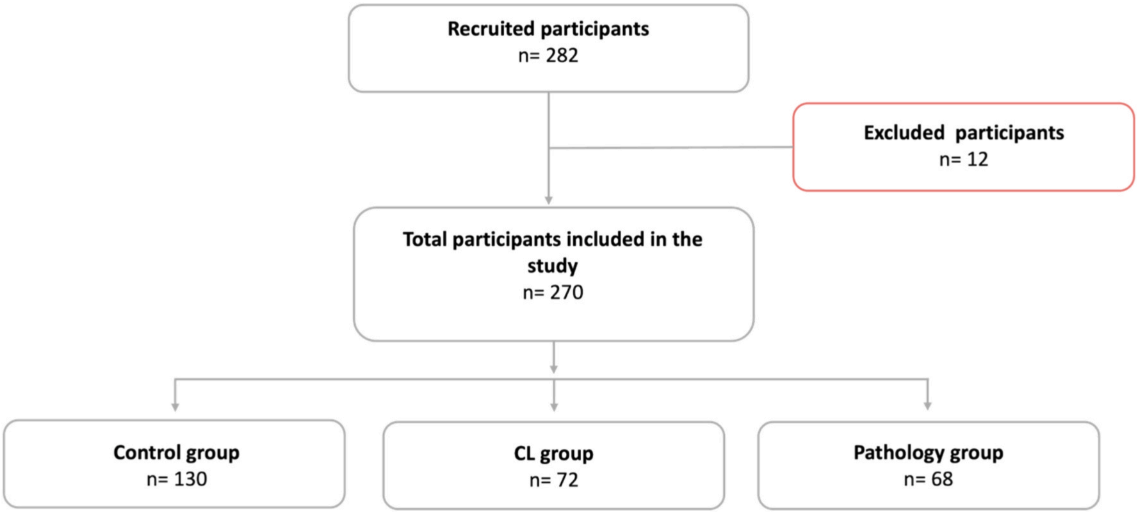


Fig. 1. Flow chart showing the total number of participants recruited for the study and those who were finally included in each group.

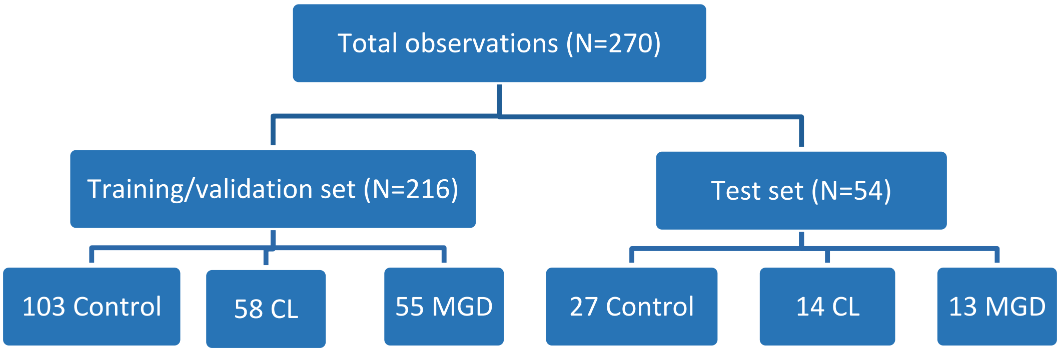


Fig. 2. Distribution of the participants in the different study groups, for the training/validation and test sets.

4.1. Training of machine learning classifiers

From the first analysis of the classifiers carried out with the data of the 216 eyes in the training/validation group, accuracies between 76 % and 81 % were obtained for the 5 most accurate classifiers. The names and accuracies obtained for these 5 machine learning classifiers are listed in Table 3.

The classifier for which the highest accuracy was obtained to classify each participant correctly for the training set, was Subspace KNN, with an accuracy of 81.4 %. Fig. 3 shows the confusion matrix obtained for this classifier, where the true successes are shown on the main diagonal, observing a good-moderate result. The total errors in the classification were 40, compared to the 176 in which it was correct classification. As shown in Fig. 3, the most common mistakes were confusion between controls and CL wearers, totalling 24 occurrences. Additionally, there were 4 confusion errors in the classifications between CL and MGD, and 12 between controls and MGD.

Table 3
Full names of the 5 classifiers for which the best accuracies were obtained with the training/validation group (n = 216). Results of the accuracies obtained for each classifier in %.

Full Name	Accuracy (%)
Subspace KNN	81.4
Bagged Trees	80.1
Boosted Trees	79.6
RUSBoosted Trees	77.3
Fine KNN	76.9

Accuracy, sensitivity and specificity are common clinical metrics to evaluate the performance of a diagnostic test. However, for multi-class classifiers the computation of these metrics is not straightforward. In this work, accuracy, sensitivity and specificity metrics have been

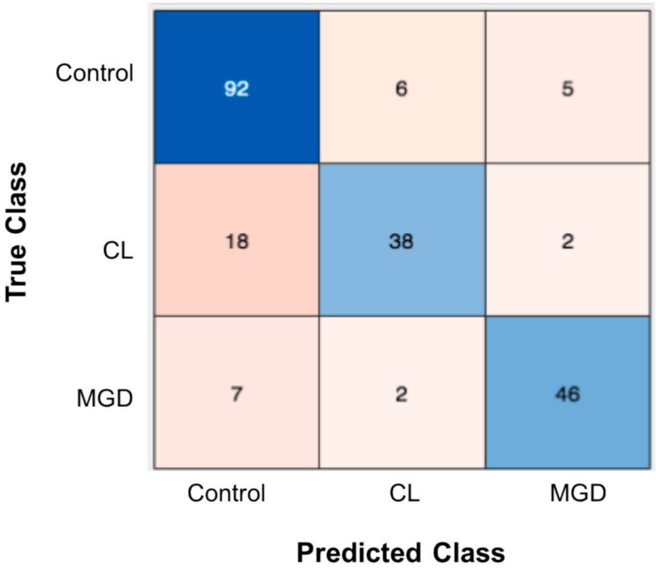


Fig. 3. Confusion matrix obtained for Subspace KNN classifier in the training/validation set.

Table 4

Accuracy, sensitivity and specificity according to the values obtained in Fig. 3 for the Subspace KNN classifier on the training/validation set. Average indicates the average between classes or macro-average.

	Accuracy (%)	Sensitivity (%)	Specificity (%)
Control	83.33	89.32	77.88
CL	87.04	65.52	94.94
MGL	92.59	83.64	95.65
Average	87.65	79.49	89.49

computed for each class using the confusion matrix data following the micro-averaging procedure [40]. For each class, $C_k(k = 1, 2, 3)$, the following metrics are computed: true positive, $tp_k = m_{kk}$, true negative, $tn_k = \sum_{i\neq k, j\neq k} m_{ij}$, false positive, $fp_k = \sum_{j\neq k} m_{kj}$, and false negative, $fn_k = \sum_{i\neq k} m_{ik}$, where $m_{i,j}(i, j = 1, 2, 3)$ are the elements of the confusion matrix. Then, for a given class, the accuracy, sensitivity and specificity are defined as: $Accuracy_k = \frac{tp_k + tn_k}{N}$, $Sensitivity_k = \frac{tp_k}{tp_k + fn_k}$ and $Specificity_k = \frac{tp_k}{tp_k + fp_k}$, N being the total number of elements of the confusion matrix. Finally, the accuracy, sensitivity and specificity indicators are averaged across all classes to get the final values. Table 4 shows the accuracy, sensitivity and specificity computed for the classifier Subspace KNN obtained from the confusion matrix defined in Fig. 3, following the procedure described above. The average sensitivity of KNN classifier was 79.5 % and the specificity was 89.5 %.

4.2. Testing machine learning classifiers

Once the 5 best classifiers were chosen, the testing of the classifiers was carried out with data from the 54 participants designated for this task. An increase in the accuracy of the classifiers previously selected in the training/validation set was obtained. The test set global accuracies obtained for all the classifiers are shown in Table 5.

The classifier for which the highest accuracy was obtained to classify each participant correctly for the test set was Subspace KNN, with an accuracy of 92.6 %. Fig. 4 shows the confusion matrix obtained for this classifier, where the true successes are shown on the main diagonal, observing a good result. The total errors in the classification were 4, compared to the 50 observations of the test set that were correctly classified. As shown in Fig. 4, the most common mistakes were between controls with CL and MGD, two errors with each group. However, these results should be interpreted with caution. While the observed higher accuracy on the independent test set compared to the validation set, coupled with the relatively small dataset size (N = 270 total) and high dimensionality (44 variables), suggests an increased potential for overfitting, the consistent performance of diverse machine learning algorithms on the test set indicates the models' ability to capture robust patterns within the data. This robust performance, despite the dataset limitations, supports the potential for these models to provide valuable diagnostic assistance in a clinical setting.

Table 5

Full names of the 5 classifiers for which the best accuracies were obtained with the test group (n = 54). Results of the accuracies obtained for each classifier in %.

Full Name	Accuracy (%)
Subspace KNN	92.6
Bagged Trees	81.5
Boosted Trees	82.3
RUSBoosted Trees	79.1
Fine KNN	83.3

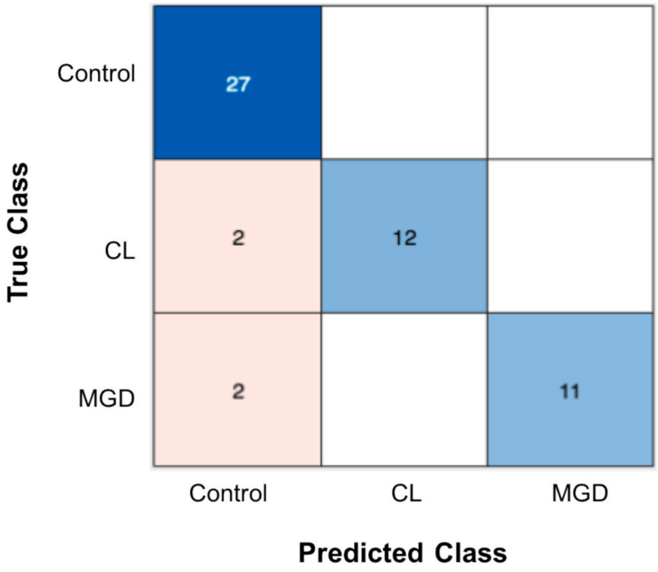


Fig. 4. Confusion matrix obtained for Subspace KNN classifier in the test set.

Table 6

Accuracy, sensitivity and specificity according to the values obtained in Fig. 4 for the Subspace KNN classifier on the test set. Average indicates the average between classes or macro-average.

	Accuracy (%)	Sensitivity (%)	Specificity (%)
Control	92.6	100	85.2
CL	96.3	85.7	100
MGD	96.3	84.6	100
Average	95.1	90.1	95.1

Table 5 shows the accuracy, sensitivity and specificity obtained for the classifier Subspace KNN in which study group for the test set. The average sensitivity was 90.1 % and the specificity was 95.1 %. The accuracy, sensitivity and specificity were higher than the ones obtained for the training/validation set (Table 6).

5. Discussion

This study analyses the use of artificial intelligence techniques, specifically machine learning classifiers, for the classification of three groups of participants, control, CL users and participants with MGD. These classifiers were trained and tested with numerical data obtained from symptomatology and clinical tests carried out on each participant.

It has been observed that it is possible to train machine learning classifiers to classify the three groups of participants using clinical data. Accuracies greater than 76 % have been obtained for five of the machine learning classifiers described in Matlab® for the training/validation data set. Specifically, the best classifier was the Subspace KNN, for which a classification accuracy of 81.4 % and a sensitivity of 79.5 % were obtained, which implies a high capacity of the classifier to detect the disease in actual sick subjects. The specificity obtained for this classifier was 89.5 %, which shows the ability of the classifier to consider healthy cases those that really are. Subsequently, in the test set, higher precision of 92.6 %, sensitivity of 90.1 % and specificity of 95.1 % have been obtained for this classifier, Subspace KNN. Therefore, it is observed that the classifier has increased its ability to discriminate between the three groups of participants. This somewhat paradoxical result of an increment of the algorithm precision between training and test sets may be due to different causes: algorithm overfitting, different internal structures between training/validation and test sets, reduced ratio between dataset cardinality and number of variables, etc. To address this point conclusively, further studies involving an expansion of both training/validation and test datasets are required [41]. The Subspace KNN or, more precisely, random Subspace KNN is a classifier well suited to problems with a high number of features [42], such is the case of this study where there are 44 numerical variables (features).

Regarding classification errors, there is a tendency for errors to occur in the CL user group, which could be due to the interaction of CL with the MG due to the morphological alterations that occur in the glands caused by daily CL reported in some studies [3]. These alterations make variables such as those used for the classification of the three groups (symptom test, corneal and conjunctival stains, and conjunctival or eyelid redness) very similar in the CL and MGD groups. [3,4,27,43,44]. In contrast, other studies found no relationship between CL use and morphological alterations in MG [45,46]. Another point to consider is the symptoms and discomfort suffered by CL users. These symptoms may be like those experienced by patients with MGD [47].

In the field of vision, the use of machine learning techniques for the diagnosis of ocular pathologies has not yet been widely studied. A study has been found that may be comparable to the one carried out in this article. Kim *et al.* used machine learning techniques to aid in the diagnosis of glaucoma. They analysed the usefulness of the classifiers to support the diagnosis of glaucoma. [48] Kim et al. observed that the family of classifiers called *ensemble*, to which the Subspace KNN belongs, was the one with the best precision, which is consistent with what has been obtained in the present study [48].

Once the use of these classifiers has been analysed and verified, it would be interesting to study, using artificial intelligence techniques, which of the clinical parameters used to train and verify the classifiers has the greatest clinical relevance for the classification of each participant in the study group, and be able to perform a more exhaustive diagnosis without having to perform the complete battery of diagnostic tests.

Among the limitations of the present study, the relatively small size of the data set must be mentioned first. Although machine learning model training does not always require such a huge training dataset as deep learning models do, the size of the combined training/validation and test dataset ($N = 270$ observations in total) employed in this work is indeed reduced, particularly given the high number of independent variables studied (44 variables). This fact directly results in several issues, such as significant challenges to obtain truly unbiased estimates of model performance, an increased overfitting risk, and potential limitations in model generalization to new, unseen patient populations.

While the methodology represented in this study incorporated standard data splitting into training, validation, and test sets to mitigate these issues as much as possible, the absolute data scarcity inherently impacts the robustness of the study conclusions. Unfortunately, addressing this is a common challenge for many machine learning applications, particularly in specialized clinical fields where large, well-curated datasets are difficult to acquire. However, techniques such as incremental learning can be implemented to continuously improve model performance. With these techniques, a pretrained model could learn from more data gradually as it is added. Indeed, some machine learning algorithms, such as decision trees, are particularly adept at this kind of incremental training. This opens an interesting future line of research in the application of incremental training to machine learning classifiers of MG dysfunction

6. Conclusions

This study demonstrates the feasibility of using machine learning classifiers to differentiate between individuals presenting Meibomian gland dysfunction (MGD), contact lens wearers, and healthy controls based solely on clinical data and using an easy-to-use app native from MATLAB language. The results obtained underscore the potential of machine learning techniques as valuable tools in healthcare, offering clinicians a tool for enhancing diagnostic accuracy, especially in the diagnosis of MGD.

While not intended to replace clinical judgment, these models can serve as reliable support, assisting specialists in diagnosing, monitoring, and managing patients with dry eye symptoms and MGD. The use of machine learning in this context brings distinct benefits, including accessibility and ease of use, and could prove especially valuable in resource-limited settings or for general practitioners working without access to advanced imaging technologies and computational tools.

The implications of this work are potentially significant, suggesting that machine learning can streamline routine clinical workflows in eye care, support data-driven decision-making, and improve early intervention strategies. It also further demonstrates that integration of clinical and imaging data is a strategy that can yield greater benefits in the automated diagnosis of MGD and related pathologies. Future studies could extend this research by incorporating larger, more diverse datasets, ultimately contributing to broader, more robust applications of machine learning in clinical practice.

Funding sources

This work was supported by the European fund for Regional Development; Santander-UCM grant for pre-doctoral research students; Spanish Government́s Agencia Estatal de Investigación [Grant number DPI2016-75272-R].

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, machine learning techniques, with Matlab software, have been used for the analysis of the benefits that artificial intelligence provide in the diagnosis of multifactorial alterations such as alterations related to the Meibomian glands. Generative AI (Microsoft Copilot and Google Gemini) has been employed only for improving grammar correctness of the text in the revised version of the manuscript. After using this tool, the authors reviewed and edited the content as necessary and take full responsibility for the content of the publication

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank all the study participants and staff at the study site.

References

[1] Nichols KK. The international workshop on meibomian gland dysfunction: introduction. Invest Ophthalmol Vis Sci 2011;52:1917–21. https://doi.org/10.1167/iovs.10-6997.

[2] Sullivan BD, Crews LA, Messmer EM, Foulks GN, Nichols KK, Baenninger P, et al. Correlations between commonly used objective signs and symptoms for the diagnosis of dry eye disease: clinical implications. Acta Ophthalmol 2014;92: 161–6. https://doi.org/10.1111/aos.12012.

[3] Osae EA, Jones L, Nichols JJ. The impact of contact lenses on meibomian gland morphology. Ocul Surf 2022;24:148–55. https://doi.org/10.1016/j.jtos.2022.04.001.

[4] Arita R, Itoh K, Inoue K, Kuchiba A, Yamaguchi T, Amano S. Contact lens wear is associated with decrease of meibomian glands. Ophthalmology 2025;2009(116): 379–84. https://doi.org/10.1016/j.ophtha.2008.10.012 [BIGML] BigML. BigML website. Available at: https://bigml.com/. Accessed 31 May.

[5] Murphy KP. Machine Learning: A Probabilistic Perspective. Illustrated. Massachusetts: The MIT press; 2012.

[6] Hasan Z, Jishkariani M. Machine learning and data mining methods for cyber security: a survey. Mesopotamian J CyberSecurity 2022;2022:47–56. https://doi.org/10.58496/MJCS/2022/006.

[7] Qamar R, Zardari BA. Artificial neural networks: an overview. Mesopotamian J Comp Sci 2023;2023:124–33. https://doi.org/10.58496/MJCSC/2023/015.

[8] Shayea GG, Zabil MHM, Albahri AS, Joudar SS, Hamid RA, Albahri OS, et al. Fuzzy evaluation and benchmarking framework for robust machine learning model in real-time autism triage applications. Int J Comput Intell Syst ;17:151. https://doi.org/10.1007/s44196-024-00543-3.

[9] Greenspan H, Van Ginneken B, Summers RM. Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. IEEE Trans Med Imaging 2016;35:1153–9.

[10] De Bruijne M. Machine learning approaches in medical image analysis: from detection to diagnosis. Med Image Anal 2016;33:94–7.

[11] Mahmud M, Vassanelli S. Processing and analysis of multichannel extracellular neuronal signals: state-of-the-art and challenges. Front Neurosci 2016;10:248.

[12] Zemouri R, Zerhouni N, Racoceanu D. Deep learning in the biomedical applications: recent and future status. Appl Sci 2019;9:1526. https://doi.org/10.3390/app9081526.

[13] Tey KY, Cheong EZK, Ang M. Potential applications of artificial intelligence in image analysis in cornea diseases: a review. Eye and vis 2024;11:10. https://doi.org/10.1186/s40662-024-00376-3.

[14] Karn PK, Abdulla WH. On machine learning in clinical interpretation of retinal diseases using OCT images. Bioengineering 2023;10:407. https://doi.org/10.3390/bioengineering10040407.

[15] Rajan S, Ponnan S. An efficient enhanced stacked auto encoder assisted optimized deep neural network for forecasting Dry Eye Disease. Sci Rep 2024;14:24945. https://doi.org/10.1038/s41598-024-75518-7.

[16] 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.

[17] Koh YW, Celik T, Lee HK, Petznick A, Tong L. Detection of meibomian glands and classification of meibography images. J Biomed Opt 2012;17:086008. https://doi.org/10.1117/1.JBO.17.8.086008.

[18] Setu MAK, Horstmann J, Schmidt S, Stern ME, Steven P. Deep learning-based automatic meibomian gland segmentation and morphology assessment in infrared meibography. Sci Rep 2021;11:7649. https://doi.org/10.1038/s41598-021-87314-8.

[19] Prabhu SM, Chakiat A, Shashank S, Vunnava KP, Shetty R. Deep learning segmentation and quantification of Meibomian glands. Biomed Signal Process Control 2020;57:101776. https://doi.org/10.1016/j.bspc.2019.101776.

[20] Wang J, Li S, Yeh TN, Chakraborty R, Graham AD, Stella XY, et al. Quantifying meibomian gland morphology using artificial intelligence. Optom Vis Sci 2021;98:1094–103. https://doi.org/10.1097/OPX.0000000000001767.

[21] Llorens-Quintana C, Rico-del-Viejo L, Syga P, Madrid-Costa D, Iskander DR. A novel automated approach for infrared-based assessment of meibomian gland morphology. Transl Vis Sci Technol 2019;8:17. https://doi.org/10.1167/tvst.8.4.17.

[22] Iradier MT, del Buey MÁ, Peris-Martínez C, Cedano P, Piñero DP. Characterization and prediction of the clinical outcome of intense pulsed light-based treatment in dry eye associated to meibomian gland dysfunction. J Clin Med 2021;10:3573. https://doi.org/10.3390/jcm10163573.

[23] Yu Y, Zhou Y, Tian M, Zhou Y, Tan Y, Wu L, et al. Automatic identification of meibomian gland dysfunction with meibography images using deep learning. Int Ophthalmol 2022;42:3275–84. https://doi.org/10.1007/s10792-022-02262-0.

[24] Graham AD, Kothapalli T, Wang J, Ding J, Tse V, Asbell P et al. A machine learning approach to predicting dry eye-related signs, symptoms and diagnoses from meibography Images. https://doi.org/10.2139/ssrn.4724519.

[25] Schiffman RM, Christianson MD, Jacobsen G, Hirsch JD, Reis BL. Reliability and validity of the ocular surface disease index. JAMA Ophthalmol 2000;118:615–21. https://doi.org/10.1001/archopht.118.5.615.

[26] Pult H, Riede-Pult B. Comparison of subjective grading and objective assessment in meibography. Cont Lens Anterior Eye 2013;36:22–7. https://doi.org/10.1016/j.clae.2012.10.074.

[27] Alghamdi WM, Markoulli M, Holden BA, Papas EB. Impact of duration of contact lens wear on the structure and function of the meibomian glands. Ophthalmic Physiol Opt 2016;36:120–31. https://doi.org/10.1111/opo.12278.

[28] Price DD, McGrath PA, Rafii A, Buckingham B. The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. Pain 1983;17:45–56. https://doi.org/10.1016/0304-3959(83)90126-4.

[29] Ferris FL, Bailey I. Standardizing the measurement of visual acuity for clinical research studies. Ophthalmology 1996;103:181–2. https://doi.org/10.1016/s0161-6420(96)30742-2.

[30] Tomlinson A, Bron AJ, Korb DR, Amano S, Paugh JR, Pearce EI, et al. The international workshop on meibomian gland dysfunction: report of the diagnosis subcommittee. Invest Ophthalmol Vis Sci 2011;52:2006–49. https://doi.org/10.1167/iovs.10-6997f.

[31] Geerling G, Baudouin C, Aragona P, Rolando M, Boboridis KG, Benítez-Del-Castillo JM, et al. Emerging strategies for the diagnosis and treatment of meibomian gland dysfunction: proceedings of the OCEAN group meeting. Ocul Surf 2017;15:179–92. https://doi.org/10.1016/j.jtos.2017.01.006.

[32] Foulks GN, Bron AJ. Meibomian gland dysfunction: a clinical scheme for description, diagnosis, classification, and grading. Ocul Surf 2003;1:107–26. https://doi.org/10.1016/S1542-0124(12)70139-8.

[33] Mitchell TM. Machine learning. New York: EEUU; 1997.

[34] Mahesh B. Machine learning algorithms-a review. Int J Sci Res 2020;9:381–6.

[35] Ray S. A quick review of machine learning algorithms. International conference on machine learning, big data, cloud and parallel computing. 2019;35–9. https://doi.org/10.1109/COMITCon.2019.8862451.

[36] Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. Introduction to machine learning, neural networks, and deep learning. Transl Vis Sci Technol 2020;9:14. https://doi.org/10.1167/tvst.9.2.14.

[37] MathWorks®. Estados Unidos de América: MathWorks: 1984 [actualizado 15 Mar 2022; citado 7 Febrero 2024]. Disponible en: https://www.mathworks.com/help/stats/classificationlearner-app.html.

[38] Paluszek M, Thomas S. Practical matlab deep learning. A project-based approach, Michael Paluszek and Stephanie Thomas. Apress, 2020.

[39] Pang R, Lansdell BJ, Fairhall AI. Dimensionality reduction in neuroscience. Curr Biol 2016;26:r656–60.

[40] Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. Inf Process Manag 2009;45:427–37. https://doi.org/10.1016/j.ipm.2009.03.002.

[41] Ying X. An overview of overfitting and its solutions. En J Phys: Conf Series IOP 2019;1168:022022. https://doi.org/10.1088/1742-6596/1168/2/022022.

[42] Ferri FJ, Inesta JM, Amin A, Pudil .Advances in Pattern Recognition: Joint IAPR International Workshops SSPR 2000 and SPR 2000. Alicante, Spain, Springer, 2003.

[43] Iqbal A, Thomas R, Mahadevan R. Impact of modulus of elasticity of silicone hydrogel contact lenses on meibomian glands morphology and function. Clin Exp Optom 2021;104:760–6. https://doi.org/10.1080/08164622.2021.1887703.

[44] Villani E, Ceresara G, Beretta S, Magnani F, Viola F, Ratiglia R. In vivo confocal microscopy of meibomian glands in contact lens wearers. Invest Ophthalmol Vis Sci 2011;52:5215–9. https://doi.org/10.1080/08164622.2021.1887703.

[45] Hom MM, Martinson JR, Knapp IL, Paugh JR. Prevalence of meibomian gland dysfunction. Optom Vis Sci 1990;67:710–2.

[46] Arita R, Itoh K, Maeda S, Maeda K, Tomidokoro A, Amano S. Association of contact lens-related allergic conjunctivitis with changes in the morphology of meibomian glands. Jpn J Ophthalmol 2012;56:14–9. https://doi.org/10.1007/s10384-011-0103-6.

[47] Arita R, Fukuoka S, Morishige N. Meibomian gland dysfunction and contact lens discomfort. Eye Contact Lens 2017;43:17–22. https://doi.org/10.1097/ICL.0000000000000351.

[48] Kim SJ, Cho KJ, Oh S. Development of machine learning models for diagnosis of glaucoma. PLoS One 2017;12:e0177726. https://doi.org/10.1371/journal.pone.0177726.