Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 196 (2022) 454–460

Procedia Computer Science

www.elsevier.com/locate/procedia

CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies 2021

Machine Learning automatic assessment for glaucoma and myopia based on Corvis ST data

Danilo Leite[a,b], Maria Campelos[c], Ana Fernandes[c], Pedro Batista[c,e], João Beirão[c,e], Pedro Menéres[c,e], António Cunha[c,f*]

[a] *Empresa Brasileira de Serviços Hospitalares- Ebserh, Paraíba, Brasil*
[b] *Universidade Federal da Paraíba, Paraíba, Brasil*
[c] *Universidade de Trás-os-Montes e Alto Douro, Vila Real 5000-80, Portugal*
[d] *Instituto de Ciências Biomédicas Abel Salazar, Porto 4050-313, Portugal*
[e] *Centro Hospitalar Universitário do Porto, Porto 4099-001, Portugal*
[f] *Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência,Porto 4200-465, Portugal*

Abstract

Glaucoma is a silent disease characterized by progressive degeneration of retinal ganglion cells and, when not detected or treated early, can lead to blindness. Computer systems have demonstrated their efficiency in the medical decision-making process and Artificial Intelligence (AI) techniques have helped advances in ophthalmology, allowing for faster and more effective detection of glaucoma. Machine learning is a very promising subfield of AI that supports research in understanding the development, progression and treatment of glaucoma, identifying new risk factors and assessing the importance of existing ones.
This study aims to test and analyze the results of different models of supervised machine learning in the detection and classification of ophthalmic diseases (Glaucoma, high myopia and low myopia) based on data from Corvis ST. The most important characteristics were selected based on a variance greater than 0.02. In terms of accuracy, the models that obtained the best results were Random Forrest 0.73, Stochastic Gradient Descent (SGD) 0.75, Gradient Boosting Classifier (GBC) 0.76 and K-Nearest Neighbors 0.71. The GBC model achieved the best results in accuracy, AUC, Recall and F1Score 76.00, 52.5, 78.00, 70.2 respectively.

Peer-review under responsibility of the scientific committee of the CENTERIS –International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies 2021

* Corresponding author. Tel.: +351931636373.
E-mail address: acunha@utad.pt

1877-0509
Peer-review under responsibility of the scientific committee of the CENTERIS –International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies 2021
10.1016/j.procs.2021.12.036

Keywords: Corvis ST; Machine Learning; Glaucoma; Cornea dynamics.

1. Introduction

Glaucoma is a disease classified as optic neuropathies, characterized by the progressive degeneration of retinal ganglion cells [1]. An abnormality typifies retinal glaucoma that originates in the optic nerve of an eye, commonly due to high pressure. The main risk factors are high intraocular pressure (IOP), age over 40 years, family history of glaucoma, associated eye pathologies such as retinal detachment, inflammation, and eye tumors, people who have had diabetes, myopia, and use of drug-based of corticosteroids. Also in this context, complications of pathological myopia are one of the main causes of visual impairment worldwide. Eyes with pathological myopia may develop ocular pathologies in the macula, peripheral retina, and optic nerve [2], [3].

In this sense, the characterization of the biomechanical attributes of the cornea has important implications for the management of eye diseases and the prediction of surgical responses [4]. In recent years, there has been an increase in the number of researches related to corneal biomechanics due to its prospective applications in the diagnosis, management, and treatment of various clinical conditions, including glaucoma, elective keratorefractive surgery, and different corneal diseases [4], [5].

In this scenario, studies reports that the integration of tomographic and biomechanical data using artificial intelligence techniques are having good results in the detection and classification of ophthalmic diseases. From this perspective, computer-assisted medical systems that use AI, specifically, MF methods, are being increasingly used to help professionals in the process of constructing the diagnosis, classification, as well as early detection of ophthalmic diseases [6].

However, there is a lack of techniques to detect and classify problems more efficiently and quickly, allowing a more agile screening for glaucoma based on biomechanical response. However, with the advent of the one-dimensional (1D) air blow-based technique to measure corneal surface response in 2005, advances in clinical imaging technology have produced increasingly sophisticated approaches to characterize the biomechanical properties of the cornea [7], [8].

The Corvis ST (Oculus Corvis ST, Scheimpflug Technology; Wetzlar, Germany) is a new non-contact tonometry system integrated with Scheimpflug's ultra-fast camera enabling a new approach to biomechanical assessment by capturing multiple dynamic corneal flattening images in front of a breath of air [9], [10]. Thus, the factors described highlight the opportunity of inserting artificial intelligence (AI) techniques, specifically Machine Learning (ML), in the corneal subspecialty to assist in screening and diagnosis in ophthalmology.

That could result in predicting responses to medical or surgical interventions; detection of susceptibility to early corneal ectasia; interrelationship with intraocular pressure measurements, with the possibility of developing an eyeball biomechanics marker that may be predictive of glaucoma susceptibility [6], [11]. Due to its irreversible nature, several studies have proposed methods for diagnosing glaucoma to assist in the specialist's decision-making process. Some of these studies use ML techniques to detect and classify the disease early.

Upon that, this study aims to test and analyze the results of different ML models to determine the most important characteristics and detect and classify the presence of ophthalmic disease (Glaucoma, high myopia, and low myopia). The data used in these studies were collected in a Portuguese public hospital; therefore, the data are private.

This study may help further research to build tools for the assessment and monitoring of glaucoma and contribute to the production of knowledge and training of new professionals, whether academics or professionals.

2. Methodology

The pipeline of the worked developed is shown in Figure 1. It illustrates the three different phases performed during this study, such as data collection and data processing, data transformation, model building and performance evaluation.

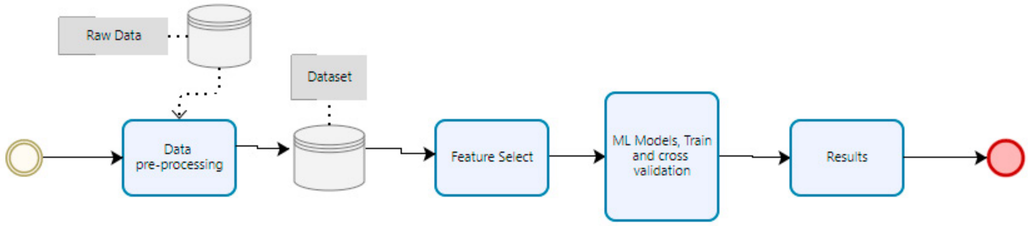

Figure 1. Metodology pipeline.

The steps that describe the construction of the proposed model are listed below.

2.1. Data preparation

The data set used for the analysis consisted of 331 samples, classified as Normal (53), Low Myopia (69 samples), High Myopia (106 samples) and With glaucoma (103 samples), from a Portuguese public hospital. The models were trained and tested to determine the most important characteristics to be used, as well as to detect and classify the presence of eye disease.

At this stage, data normalization was performed to leave the frequency distribution of the data as a normal distribution, or at least to make it symmetrical and minimize data bias, and the transformation of categorical class data into attributes in a Numerical representation was done through one-hot coding. Finally, a specific Dataset was created for training and evaluating ML models.

2.2. Feature selection

Feature selection is the process of reducing the number of input variables when developing a predictive model, it refers to the process of identifying the most important variables or attributes that are essential in a model for more accurate prediction. It is advisable to reduce the number of input variables to reduce the computational cost of modeling and, in some cases, to improve model performance. One of the main goals is to remove uninformative or redundant predictors from the model[12]. Therefore, in this study, we chose to use variance to select the most important measures from the data set to train the models and, consequently, detect and classify patients with absence or presence of voice deviation. Therefore, we chose the variance because, according to [12], it reduces the computational cost, is statistically robust against overfitting and minimizes data bias.

With the dataset created, we apply the variation to select the most relevant features to be used by the models in ranking. In the first round of tests, the complete set was used, while in the second test, a variance greater than 0.02, the latter with better results[11]. In the first round of tests, there were 41 resources (Table 1), the results showed the low efficiency of the models when all resources were used. Also in this first round of testing, we tested stratified cross-validation 4, 6, 8, and 10 times. In the second round of tests, the classification obtained was better efficiency with 8 k-fold, with an accuracy greater than 73.0 for some models. Therefore, we use the variance of 0.02 to select the best features, to train and adjust the models. The selected features are highlighted in Table 1.

2.3. Machine Learning Models train and cross-validation

We select ten of the most used models in supervised machine learning for data classification namely: Random Forest (RF), Naive Bayes (NB), Support Vector Machines (SVM), Multi-layer Perceptron Classifier (MLPC), Decision Tree (DT), Gradient Boosting Classifier (GBC), K-Nearest Neighbor (KNN), Stochastic Gradient Descent (SGDC), AdaBoostClassifier (ABC) and Regression Logistic (RL) [13], [14], [15].

Table 1 - Dataset features

Age	IOP [mmHg]	Pachy [µm]	Def. Amp. Max [mm]
A1 Time [ms]	A1 Velocity [m/s]	A2 Time [ms]	SSI
A2 Velocity [m/s]	HC Time [ms]	Peak Dist. [mm]	Radius [mm]
A1 Deformation Amp. [mm]	HC Deformation Amp. [mm]	A2 Deformation Amp. [mm]	A1 Deflection Length [mm]
HC Deflection Length [mm]	A2 Deflection Length [mm]	A1 Deflection Amp. [mm]'	HC Deflection Amp. [mm]
A2 Deflection Amp. [mm]	Deflection Amp. Max [mm]	Deflection Amp. Max [ms]	Whole Eye Movement Max [mm]
Whole Eye Movement Max [ms]	PachySlope [µm]	DA Ratio Max (1mm)	ARTh
bIOP	Integrated Radius [mm^-1]	SP A1	A1 Deflection Area [mm²]
HC Deflection Area [mm²]	A2 Deflection Area [mm²]	A1 dArc Length [mm]	HC dArc Length [mm]
A2 dArc Length [mm]	dArcLengthMax [mm]	Max InverseRadius [mm^-1]	DA Ratio Max (2mm)
CBI			

The hyperparameters for each ML model were selected to obtain the best performance and best classification of the data. The ML models initially used their default settings so that, as each model was adjusted to the data in the training process, the hyperparameters were adjusted. The hyperparameters of the selected models were:

- RF - {'bootstrap': True, 'criterion': 'entropy', 'max_depth': 3, 'max_features': 'auto', 'min_samples_leaf': 1, 'n_estimators': 16, 'random_state': 123}
- NB – {'var_smoothing': 2.310129700083158e-05}
- SVM – {'C': 10, 'degree': 2, 'gamma': 'scale', 'kernel': 'linear'};
- MLPC – {'activation': 'tanh', 'alpha': 0.05, 'hidden_layer_sizes': (50, 50), 'learning_rate': 'constant', 'solver': 'adam'};
- DT – {'max_depth': 7, 'max_features': None, 'max_leaf_nodes': 10, 'min_samples_leaf': 1, 'min_weight_fraction_leaf': 0.1, 'splitter': 'best'};
-
- GBC - {'learning_rate': 0.1, 'max_depth': 3, 'n_estimators': 14}
- KNN - {'algorithm': 'brute', 'leaf_size': 1, 'n_neighbors': 10, 'weights': 'uniform'}
- SGDC - {'alpha': 0.01, 'loss': 'modified_huber', 'penalty': 'l1'}
- RL - {'C': 3.727593720314938, 'penalty': 'l2', 'random_state': 0}
- ABC - {'algorithm': 'SAMME', 'learning_rate': 0.05, 'n_estimators': 500}

To assess classifier models, we used the following metrics based on the number of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), which were:

- Accuracy, which measures the test's ability to correctly identify when there is and when there is no deviation. It is defined as the relationship between the number of correctly classified cases and all cases exposed to the classifier:

$$acc = \frac{VP + VN}{VP + VN + FP + FN} \tag{1}$$

- Recall, which captures the ability of the classifier to find all the positive samples;

$$r = \frac{TP}{TP+FN} \tag{2}$$

- Precision, which is the ability of the classifier not to label a negative sample positive;

$$p = \frac{TP}{TP+FP} \tag{3}$$

- F1-Score which is the harmonic mean of precision and recall computes values in the range [0,1];

$$f1 = 2 * \frac{(precision*recall)}{precision+recall} \quad (4)$$

- AUC, which demonstrates the model's ability to distinguish categories;

The test is considered positive (diseased) or negative (non-diseased), and the deviation present or absent. The test is correct when it is positive in the presence of deviation (True Positive-VP) or negative in the absence of deviation (True Negative-TN). Furthermore, the test is wrong when it is positive in the absence of the deviation (False Positive-PF) or negative when the deviation is present (False Negative-FN).

The dataset was divided into k parts, with the k-1 part for training and the rest for testing during the first iteration. The entire process was iterated k-times, where each bend is arbitrarily selected to avoid bias by selecting specific samples. Cross-validation of 4, 6, 8, and 10 times was used, with 8 having the best result. Some studies report better results of the RF, KNN, SVM, NB [1], [8], [16] models among others, but few have tested the efficiency of the GBC and SGD models.

3. Results and discussions

The ten classifer's performance was compared. We selected the RF, SGD, GBC, and KNN models that meet the exclusion criteria mean accuracy, F1 score, recall, and precision below 0.70. Their confusion matrix and results per class are shown in the Tables 2.

Table 2 - Models matrix confusion and per class metrics for the RF, SGD, GBC, and KNN models.

			True class							
		Class	Normal	Low myopia	High Myopia	Glaucoma	**Accuracy**	**Recall**	**Precision**	**F1**
GBC	Predicted classes	Normal	6	0	2	0	0.92	0.75	0.75	0.75
		Low myopia	0	12	0	0	**1.00**	**1.00**	**1.00**	**1.00**
		High Myopia	1	0	13	4	0.78	0.72	0.65	0.68
		Glaucoma	1	0	5	11	0.81	0.65	0.73	0.69
RF	Predicted classes	Normal	4	0	1	3	0.87	0.50	0.57	0.53
		Low myopia	0	12	0	0	**1.00**	**1.00**	**1.00**	**1.00**
		High Myopia	0	0	14	4	0.84	0.78	0.74	0.76
		Glaucoma	3	0	4	10	0.75	0.59	0.59	0.59
SGDC	Predicted classes	Normal	3	0	0	6	0.89	0.33	**1.00**	0.50
		Low myopia	0	12	0	0	**1.00**	**1.00**	**1.00**	**1.00**
		High Myopia	0	0	13	4	0.85	0.76	0.76	0.76
		Glaucoma	0	0	4	13	0.75	0.76	0.57	0.65
KNN	Predicted classes	Normal	7	0	0	1	0.91	0.88	0.64	0.74
		Low myopia	0	12	0	0	**0.98**	**1.00**	**0.92**	**0.94**
		High Myopia	1	1	12	4	0.78	0.67	0.67	0.67
		Glaucoma	3	0	6	8	0.75	0.47	0.62	0.53

As we can see in Table 2, for the 8 selected features, the GBC model had a better accuracy performance for the Normal classes with 0.92 and glaucoma with 0.81, in the F1-score performance measure, the GBC had 0.75 for normal, and for the glaucoma class it achieved 0.81 and 0.73 for accuracy and precision. However, the RF model obtained the best results for the High Myopia class in recall and F1, obtaining 0.78 and 0.76. On the other hand, the SGDC model achieved the best results in the classification of High Myopia in accuracy, precision, and F1 with 0.85; 0.76, and 0.76 respectively. Finally, the KNN classification model obtained the best result in the Normal classification, with a recall of 0.88. It is noteworthy that the Low myopia class obtained the best classification among the ML.

The analysis of the GBC, RF, SGDC, and KNN classifiers was performed on two types of resource datasets. The first is the complete set of 41 features consisting of structural and non-structural features. Second, the set is reduced to 8 resources using a variance greater than 0.02 (Table 1). Furthermore, the analysis was performed with a stratified k-fold division and cross-validation method for both sets of features.

To optimize the performance of the considered classifier model, different metrics were adjusted to obtain better classification accuracy and the appropriate ones are selected. Table 3 illustrates the comprehensive explanation of the results obtained from different resource selection/sorting experiments and records the results of different performance metrics. In the first test, all resources were used and in the second, with a variance greater than 0.02.

Table 3 – Full features dataset vs Selected features dataset result analysis

	Accuracy		**AUC**		**Recall**		**Precision**		**F1-Score**	
	FDS	SDS	FDS	SDS	FDS	SDS	FDS	SDS	FDS	SDS
GBC	0.71	**0.76**	50,1	**52,5**	71	**78,0**	71	**78,3**	70,2	78,0
RF	0.70	**0.73**	45	**52**	68	**71,8**	68	**72,5**	65	72,0
SGDC	0.71	**0.75**	48	**52,25**	68	**71,3**	75	**83,3**	56,8	72,3
KNN	0.70	**0.71**	50	**52,5**	70	**75,5**	67	**71,3**	62,4	72,0

Legend: FDS – Full features dataset; SDS – Selected features dataset.

In terms of accuracy, the RF, SGDC, GBC, and KNN models obtained 0.73, 0.75, 0.76, and 0.71 respectively for 8 resources. On the other hand, the accuracy, Recall, Precision, and F1-Score values for 41 characteristics were lower.

The GBC models had the best results in accuracy, AUC, Recall, and F1Score 76.00, 52.5, 78.00, 70.2 respectively. On the other hand, the SGB model had better accuracy, with 83.25.

Thus, it is observed that the set of reduced resources proposed surpasses the complete set of 41 resources for all classifiers. In addition, the GBC classifier outperformed the other classifiers with greater precision and accuracy for both feature sets.

4. Conclusion

This work presented a classification based on machine learning using full-resources and reduced-resources datasets to detect and classify eye diseases (glaucoma, high myopia and low myopia) based on data from the XPTO machine. In existing studies, the use of resources, classifiers and datasets varies from study to study. Furthermore, few studies use the GBC and SGDC classifications, similarly few studies use the variance in resource selection. Although the databases are different, these studies gave us support for our research. For future work, we intend to increase our database, test other resource selection methods, as well as explore new classification methods. Finally, this study can help in further research in the construction of tools for the assessment and monitoring of glaucoma, as well as contributing to the production of knowledge and training of new professionals, whether academics or professionals.

Acknowledgments

This work is financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, within project UIDB/50014/2020.

References

[1] N. Thakur and M. Juneja, “Classification of glaucoma using hybrid features with machine learning approaches,” *Biomed. Signal Process. Control*, vol. 62, no. August, p. 102137, 2020, doi: 10.1016/j.bspc.2020.102137.

[2] Y. C. Kim *et al.*, “Machine learning prediction of pathologic myopia using tomographic elevation of the posterior sclera,” *Sci. Rep.*, vol. 11, no. 1, pp. 1–13, 2021, doi: 10.1038/s41598-021-85699-0.

[3] B. T. Lopes *et al.*, “Enhanced Tomographic Assessment to Detect Corneal Ectasia Based on Artificial Intelligence,” *Am. J. Ophthalmol.*, vol. 195, pp. 223–232, 2018, doi: 10.1016/j.ajo.2018.08.005.

[4] L. P. G. Esporcatte *et al.*, “Biomechanical diagnostics of the cornea,” *Eye Vis.*, vol. 7, no. 1, pp. 1–12, Dec. 2020, doi: 10.1186/s40662-020-0174-x.

[5] T. K. Yoo *et al.*, “Adopting machine learning to automatically identify candidate patients for corneal refractive surgery,” *npj Digit. Med.*, vol. 2, no. 1, Dec. 2019, doi: 10.1038/s41746-019-0135-8.

[6] A. Consejo, J. Solarski, K. Karnowski, J. J. Rozema, M. Wojtkowski, and D. R. Iskander, “Keratoconus detection based on a single scheimpflu image,” *Transl. Vis. Sci. Technol.*, vol. 9, no. 7, pp. 1–12, 2020, doi: 10.1167/tvst.9.7.36.

[7] J. Chong and W. J. Dupps, “Corneal biomechanics: Measurement and structural correlations,” *Exp. Eye Res.*, vol. 205, no. November 2020, p. 108508, 2021, doi: 10.1016/j.exer.2021.108508.

[8] A. Bustamante-Arias, A. Cheddad, J. C. Jimenez-Perez, and A. Rodriguez-Garcia, “Digital Image Processing and Development of Machine Learning Models for the Discrimination of Corneal Pathology: An Experimental Model,” *Photonics*, vol. 8, no. 4, p. 118, 2021, doi: 10.3390/photonics8040118.

[9] M. Q. Salomão *et al.*, “The role of corneal biomechanics for the evaluation of ectasia patients,” *Int. J. Environ. Res. Public Health*, vol. 17, no. 6, p. 2113, Mar. 2020, doi: 10.3390/ijerph17062113.

[10] J. Chong and W. J. Dupps, “Corneal biomechanics: Measurement and structural correlations,” *Exp. Eye Res.*, vol. 205, no. November 2020, p. 108508, 2021, doi: 10.1016/j.exer.2021.108508.

[11] A. Septiarini, D. M. Khairina, A. H. Kridalaksana, and H. Hamdani, “Automatic glaucoma detection method applying a statistical approach to fundus images,” *Healthc. Inform. Res.*, vol. 24, no. 1, pp. 53–60, Jan. 2018, doi: 10.4258/hir.2018.24.1.53.

[12] D. Singh and B. Singh, “Investigating the impact of data normalization on classification performance,” *Appl. Soft Comput.*, vol. 97, no. xxxx, p. 105524, 2020, doi: 10.1016/j.asoc.2019.105524.

[13] T. M. Mitchell, “Machine Learning,” p. 432, 1997.

[14] S. Oh, Y. Park, K. J. Cho, and S. J. Kim, “Explainable Machine Learning Model for Glaucoma Diagnosis and Its Interpretation,” *Diagnostics*, vol. 11, no. 3, p. 510, 2021, doi: 10.3390/diagnostics11030510.

[15] Leite, Danilo Rangel Arruda, R. M. De Moraes, and L. W. Lopes, “Método de Aprendizagem de Máquina para Classificação da intensidade do desvio vocal utilizando Random Forest Machine Learning Method for Classifying Vocal Deviation Intensity Using Random Forest Método de aprendizaje automático para clasificar la intensid,” pp. 196–201, 2020.

[16] H. Salazar, V. Misra, and S. S. Swaminathan, “Artificial intelligence and complex statistical modeling in glaucoma diagnosis and management,” *Curr. Opin. Ophthalmol.*, vol. 32, no. 2, pp. 105–117, 2021, doi: 10.1097/ICU.0000000000000741.