Machine Learning for Glaucoma Identification: A Comparative Study of Algorithms

Robson Paulino da Silva[1], **Janier Arias García**[1], **Frederico Gualberto Ferreira Coelho**[1]

[1]Graduate Program in Electrical Engineering - Universidade Federal de Minas Gerais - Av. Antônio Carlos 6627, 31270-901, Belo Horizonte, MG, Brazil

robsonpaulino@ufmg.br, janier-arias@ufmg.br, fredgfc@ufmg.br

Abstract. *Glaucoma is one of the leading causes of irreversible blindness worldwide, and its early detection remains a significant challenge. Although machine learning models have shown promising performance in the classification of transcriptomic data, there is still a gap in understanding the robustness of these models under realistic perturbations and dataset heterogeneity. In this context, this work proposes a comprehensive evaluation of 13 machine learning algorithms for glaucoma identification using gene expression data obtained from public repositories. A unified experimental pipeline was adopted, including preprocessing steps, batch effect correction, and validation through repeated stratified hold-out, with evaluation based on metrics such as accuracy, precision, recall, F1-score, and AUC. Statistical comparisons were performed using Friedman and Nemenyi tests, and model interpretability was analyzed using SHAP. Additionally, robustness was investigated through the introduction of controlled transcriptomic perturbations, including Gaussian noise and gene-level mutations. The results indicate that the best performance was achieved with the combination of datasets GSE27276 and GSE9944, reaching an AUC of 0.99, with Logistic Regression and Stacking models standing out. In contrast, the inclusion of multiple additional datasets resulted in significant performance degradation, with AUC reduced to approximately 0.78, highlighting the impact of transcriptomic heterogeneity on generalization capability. Furthermore, the best-performing models demonstrated high robustness, maintaining stable AUC values even under perturbation scenarios. These findings indicate that regularized linear models can provide competitive performance and greater stability compared to more complex approaches, reinforcing their suitability for high-dimensional biomedical data.*

1. Introduction

Glaucoma is one of the leading causes of irreversible blindness worldwide, characterized by the progressive degeneration of the optic nerve and often diagnosed at late stages due to the absence of symptoms in early stages of the disease. In this context, gene expression data have been explored for the identification of molecular biomarkers, although they present significant challenges, such as high dimensionality and a limited number of samples.

Machine learning techniques have shown promising performance in this scenario. However, there are still few systematic evaluations regarding model stability under realistic perturbations and the impact of integrating multiple datasets, which are often character-

ized by high heterogeneity. Furthermore, the lack of standardized experimental protocols hinders comparisons across studies.

In this work, a systematic evaluation of 13 machine learning algorithms for glaucoma identification based on transcriptomic data is conducted. The study investigates model behavior under controlled perturbations, including Gaussian noise and gene-level mutations, and analyzes the impact of combining multiple datasets on generalization capability.

The main contributions include the analysis of model robustness under transcriptomic perturbations, the evaluation of the impact of heterogeneity resulting from the integration of multiple datasets, and the use of performance metrics, statistical tests, and SHAP-based interpretability techniques. The results show that regularized linear models achieve competitive performance and greater stability in high-dimensional biomedical scenarios.

2. Literature Review

The identification of transcriptomic biomarkers for glaucoma has been widely explored using bioinformatics and machine learning techniques. Previous studies have demonstrated significant advances in the identification of disease-associated genes and in the development of predictive models. However, important limitations remain regarding model robustness, interpretability, and generalization capability, particularly in the presence of the inherent variability of biological data.

2.1. Transcriptomic Biomarkers

Gene expression-based studies have identified patterns associated with glaucoma, including the dysregulation of inflammatory processes, oxidative stress, and neuronal apoptosis. Approaches such as differential expression analysis and co-expression networks have been widely employed to identify potential biomarkers [Qiu et al. 2018, Ullah et al. 2024, Suo et al. 2022, Xu et al. 2023, Xu et al. 2025, Feng and Xu 2019]. Despite these advances, most studies focus primarily on identifying relevant genes, with less emphasis on predictive evaluation and robust validation of the findings.

2.2. Machine Learning

Machine learning techniques, including linear models and ensemble methods, have been successfully applied to transcriptomic data classification. These methods have shown strong predictive performance, especially in high-dimensional scenarios [Cortes and Vapnik 1995, Zhou 2012]. Nevertheless, the literature still lacks systematic evaluations that consider not only performance but also stability and sensitivity of models to data variations.

2.3. Robustness and Generalization

The variability of transcriptomic data, resulting from factors such as experimental noise, batch effects, and biological heterogeneity, represents a significant challenge for machine learning models. Small variations in the data may substantially impact predictions, highlighting the sensitivity of many algorithms in this context [Tezel 2011, Kaur et al. 2023, Garcia and Ramos 2022]. Despite this, most studies evaluate models under controlled conditions without explicitly considering the effects of realistic perturbations, limiting the understanding of their applicability in practical scenarios.

2.4. Interpretability

Interpretability is a central aspect of biomedical applications. Methods such as *SHapley Additive exPlanations* (SHAP) have been employed to identify the contribution of individual genes to model predictions [Lundberg and Lee 2017]. However, these approaches present limitations, including high computational cost and potential instability in the explanations, especially in high-dimensional data, which may compromise their reliability in clinical applications.

2.5. Statistical Comparison

Model comparison requires the use of appropriate statistical tests. The Friedman test, followed by the Nemenyi post-hoc test, is widely used to identify significant differences among classifiers, providing a more robust performance assessment in comparative studies [Friedman 1937, Nemenyi 1963, Demsar 2006].

2.6. Critical Discussion of the Literature

Table 1. Comparison of major studies on glaucoma detection using biomedical data and machine learning

Author	Approach	Data	Advantages	Limitations
Hemelings et al. (2021)	Deep Learning (CNN)	Fundus images	High predictive capability; captures complex patterns	Low interpretability; dependence on large volumes of data
Phene et al. (2019)	Deep Learning + specialists	Fundus images	Comparison with human specialists; clinical validation	Restricted focus on images; limited generalization to molecular data
Medeiros et al. (2019)	Deep Learning (transfer learning)	OCT + images	Multimodal integration; improved damage quantification	High computational cost; dependence on structured data
Qiu et al. (2018)	Bioinformatics (differential expression)	Microarray	Identification of glaucoma-associated genes	Lack of robust validation and predictive models
Dai et al. (2022)	Machine Learning (multiple models)	Gene expression	Use of multiple algorithms; biomarker identification	Limited evaluation of robustness and generalization
Xie et al. (2024)	ML + feature selection	RNA-Seq	Integration of multiple models; analysis of relevant genes	Lack of evaluation under perturbations and heterogeneity
Ullah et al. (2024)	Integrated bioinformatics	Gene expression	Focus on neuroinflammation; biological validation	Absence of robust predictive analysis
Wang et al. (2024)	Bioinformatics + ML	Gene expression	Identification of immune biomarkers	Limited analysis of model stability and variability

Table 1 highlights that different approaches present complementary advantages and limitations. Image-based methods tend to achieve high predictive performance; however, they often exhibit low interpretability and high dependence on large volumes of data. In contrast, gene expression-based approaches frequently prioritize biomarker identification, with less emphasis on the systematic evaluation of predictive capability.

It is also observed that most studies do not explicitly consider the impact of realistic perturbations or the heterogeneity among different datasets, which limits the assessment of model robustness and generalization. Therefore, a gap remains in the literature regarding the systematic evaluation of the robustness, stability, and generalization of machine learning models applied to transcriptomic data under scenarios that reflect real-world data variability conditions.

3. Materials and Methods

This work adopts a structured experimental pipeline for the evaluation of machine learning models applied to glaucoma identification using gene expression data. As illustrated in Figure 1, the pipeline begins with data acquisition, followed by preprocessing, integration, and quality control stages. During the modeling stage, batch effect correction is applied to mitigate technical variations among different datasets. Subsequently, machine learning models are trained and evaluated through stratified validation using appropriate performance metrics. Finally, robustness analyses are conducted through the introduction of controlled perturbations into the data. The details of each stage are presented in the following subsections.

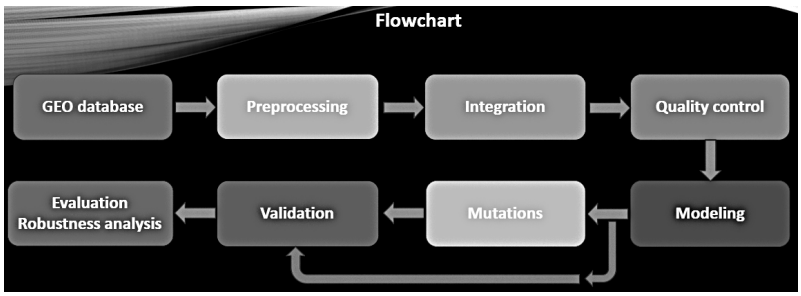


Figure 1. Methodological workflow of the study.

3.1. Datasets

Four public datasets from the Gene Expression Omnibus (GEO) were used: GSE27276, GSE9944, GSE2378, and GSE4316. These datasets contain transcriptomic profiles of ocular tissues and have been widely employed in bioinformatics studies focused on the identification of molecular signatures associated with glaucoma [Qiu et al. 2018, Suo et al. 2022, Ullah et al. 2024, Xu et al. 2025, Wu et al. 2023].

3.2. Preprocessing and Integration

The preprocessing stage consisted of standardization and integration of data obtained from different microarray platforms. Initially, *probe IDs* were mapped to gene symbols, since each platform uses specific probe codes to measure gene expression. The average expression value was used when multiple probes represented the same gene [Pulukool et al. 2021, Howell et al. 2012]. Subsequently, the datasets were integrated through the intersection of genes shared across the datasets, ensuring comparability among experiments.

Missing values were handled using median imputation, and a logarithmic transformation $\log_2(x + 1)$ was applied, together with variability-based filtering, to reduce

noise and dimensionality [Xu et al. 2023, Feng and Xu 2019]. All preprocessing steps were applied consistently across the analyzed datasets. Different dataset combinations, ranging from two to four datasets, were considered in order to evaluate the impact of data integration on model development:

1. GSE27276 + GSE9944;
2. GSE27276 + GSE9944 + GSE2378;
3. GSE27276 + GSE9944 + GSE4316;
4. GSE27276 + GSE9944 + GSE2378 + GSE4316.

3.3. Quality Control

Quality control was performed before and after preprocessing using boxplots, density curves, correlation matrices, principal component analysis (PCA), and hierarchical clustering (Ward's method), following established practices in transcriptomic studies [Peng et al. 2025, Zhang et al. 2025]. Additionally, a list of glaucoma-associated genes based on the literature was considered [Wang et al. 2024, Xu et al. 2025, Liu et al. 2023].

3.4. Machine Learning Models

Different machine learning algorithms were evaluated (Table 2), including linear, margin-based, probabilistic, and ensemble models, to compare performance, stability, and generalization capability. The use of multiple algorithms enables a more robust comparative evaluation and is a common practice in transcriptomic studies to increase the reliability of results [Dai et al. 2022, Xie and Yu 2024].

Table 2. Machine learning models used

MLP (Multilayer Perceptron)	LogReg (Logistic Regression)
Elastic Net	Ridge Classifier
SGDClassifier	Perceptron
LDA (Linear Discriminant Analysis)	SVM (Support Vector Machine)
Gradient Boosting	XGBoost
CatBoost	GaussianNB
Stacking	

3.5. Training and Validation

Model evaluation was conducted using a *repeated stratified hold-out* validation protocol implemented with the *StratifiedShuffleSplit* class from the *scikit-learn* library. Ten independent data splits were performed while preserving the original class proportion (glaucoma/control) in each partition. In each repetition, 70% of the samples were used for training and 30% for external evaluation.

3.6. Metrics and Batch Effect Correction

Performance was evaluated using accuracy, precision, recall, F1-score, and *Area Under the Curve* (AUC). The AUC metric was prioritized due to its robustness to class imbalance. Batch effect correction was performed using the ComBat method. To avoid data leakage, the parameters were estimated exclusively on the training set.

3.7. SHAP-Based Interpretability Analysis

Feature importance was estimated throughout repeated stratified hold-out validation and computed after training each model in every partition. For tree-based and linear models, the SHAP methods *TreeExplainer* and *LinearExplainer* were employed, respectively [Lundberg and Lee 2017]. For models with computational or methodological limitations, the *Permutation Importance* technique was applied.

3.8. Statistical Analysis

The Friedman test was used to detect overall differences among classifiers. When significant, the Nemenyi post-hoc test was applied for pairwise comparisons [Friedman 1937, Nemenyi 1963, Demsar 2006].

3.9. Robustness Analysis

Robustness was evaluated through the injection of stochastic Gaussian noise [Zhao and Wang 2020, Lee and Kim 2020] and the generation of gene-specific mutations [Tezel 2011, Kaur et al. 2023, Garcia and Ramos 2022]. These experiments simulate biological variability and measurement noise, enabling the assessment of model stability.

4. Results

This section presents the results obtained from the evaluation of machine learning models applied to gene expression data. Initially, the overall performance of the models is presented, followed by analyses across different dataset combinations. Subsequently, model robustness is evaluated under perturbation scenarios. Finally, interpretability aspects and statistical comparisons among the algorithms are discussed.

4.1. Overall Model Performance

Table 3 presents the model performance for the GSE (27276 + 9944) combination. Table 4 summarizes the results for the different dataset combinations. The GSEs (27276 + 9944) combination achieved the best performance (AUC = 0.99), with Logistic Regression and Stacking standing out, whereas the inclusion of additional datasets resulted in reduced performance. Tables 5 and 6 present the results under Gaussian noise and targeted mutation scenarios, respectively.

Table 3. Model performance in the scenario without mutations.

Metric	MLP	LogReg	GB	SVM	ENet	LDA	Ridge	SGD	Perc.	GNB	XGB	CatB	Stack
Accuracy	0.52	**0.94**	0.50	0.50	0.58	0.46	0.58	0.88	0.53	0.44	0.51	0.50	**0.94**
F1-score	0.43	**0.95**	0.36	0.36	0.55	0.07	0.67	0.89	0.44	0.00	0.43	0.36	**0.95**
AUC	0.73	**0.99**	0.51	0.47	0.84	0.50	0.92	0.94	0.89	0.50	0.50	0.48	**0.99**

4.2. Perturbation Scenario Analysis

To evaluate model robustness, three scenarios were considered: (i) original data, (ii) addition of Gaussian noise, and (iii) gene-directed mutation. Figure 2 shows the ROC curves for Logistic Regression across the three scenarios. Performance remained close to ideal even under perturbations.

Table 4. Simulation results (summary) – without mutation

Combination	Best Model	Accuracy	F1-score	AUC
27276 + 9944	LogReg / Stacking	**0.94**	**0.95**	**0.99**
27276 + 9944 + 2378	Stacking	0.86	0.87	0.91
27276 + 9944 + 4316	LogReg / Stacking	0.72	0.72	0.78
27276 + 9944 + 2378 + 4316	Stacking	0.72	0.72	0.79

Table 5. Simulation results (summary) – Gaussian noise mutation

Combination	Best Model	Accuracy	F1-score	AUC
27276 + 9944	LogReg / Stacking	**0.94**	**0.95**	**0.99**
27276 + 9944 + 2378	LogReg	0.86	0.86	0.90
27276 + 9944 + 4316	Stacking	0.75	0.74	0.77
27276 + 9944 + 2378 + 4316	Stacking	0.73	0.75	0.79

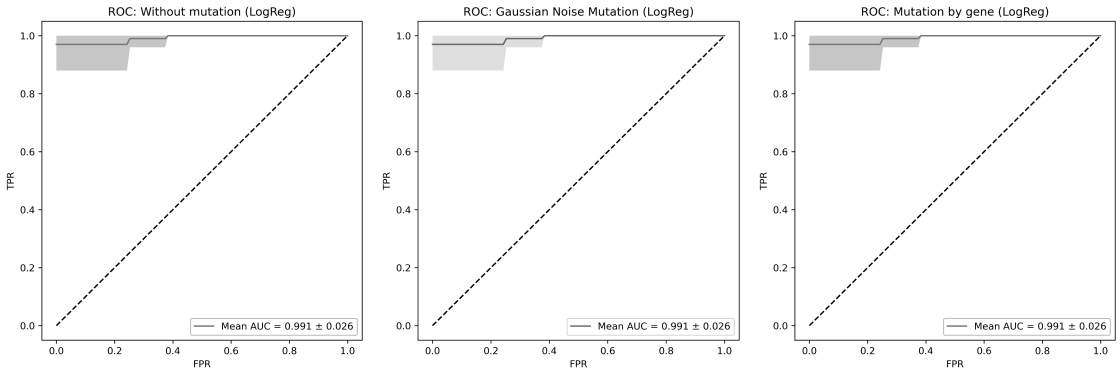


Figure 2. ROC curves for the three analyzed scenarios.

Table 6. Simulation results (summary) – Mutation by gene

Combination	Best Model	Accuracy	F1-score	AUC
27276 + 9944	LogReg / Stacking	**0.94**	**0.95**	**0.99**
27276 + 9944 + 2378	LogReg / Stacking	0.94	0.95	0.99
27276 + 9944 + 4316	LogReg / Stacking	0.73	0.73	0.78
27276 + 9944 + 2378 + 4316	Stacking	0.71	0.72	0.78

4.3. Model Interpretability

Interpretability analysis was conducted using SHAP, considering one of the best-performing models (Logistic Regression). Figure 3 shows gene importance according to SHAP. A reduced set of genes concentrates the highest contribution, maintaining consistency across scenarios.

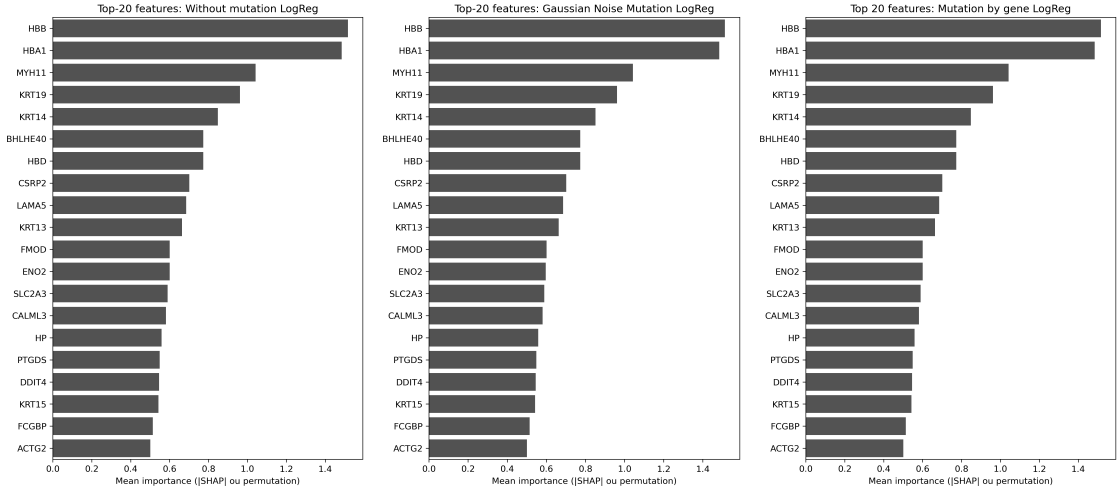


Figure 3. Gene importance (SHAP) for the three analyzed scenarios.

4.4. Statistical Comparison of Models

Table 7 presents the results of the Friedman test. Statistically significant differences were observed across all scenarios ($p < 0.001$), with high agreement (Kendall's W). Figure 4 shows the critical difference diagrams, highlighting groups of statistically equivalent models and significant differences among distinct groups. The *Logistic Regression* and *Stacking* models achieved the lowest average ranks, with no statistically significant difference between them.

Table 7. Friedman and Kendall results for AUC GSEs (27276 + 9944).

Scenario	χ^2 (df=12)	p-value	W (Kendall)
Without mutation	106.43	< 0.001	0.887
Gaussian noise	108.58	< 0.001	0.905
Directed mutation	105.28	< 0.001	0.877

5. Discussion

The results are interpreted in light of the characteristics of transcriptomic data and the existing literature, with emphasis on performance, the impact of integrating different datasets, robustness, interpretability, and implications for biomedical applications.

5.1. Model Performance

The results (Table 4) indicate that linear models can achieve performance comparable to that of more complex approaches when applied to gene expression data. This suggests

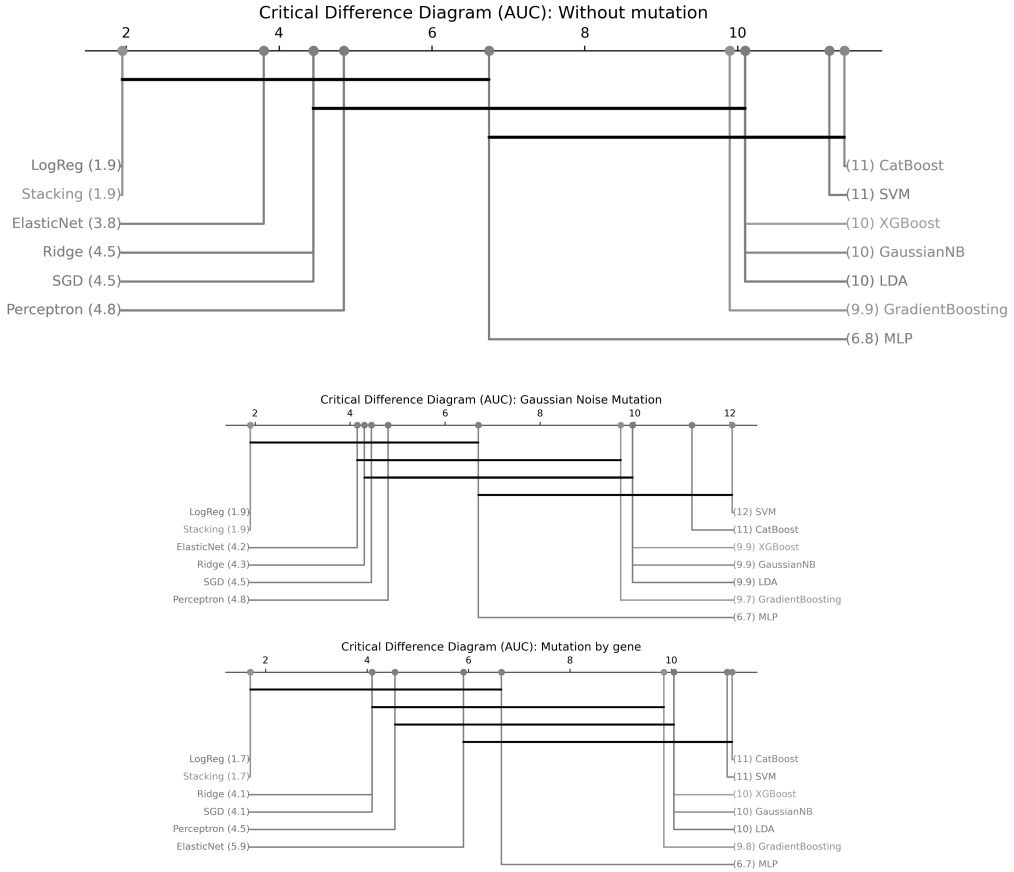


Figure 4. Critical difference diagrams (AUC) for the three analyzed scenarios.

that, after preprocessing, part of the discriminative patterns can be adequately captured by simpler decision boundaries, which is consistent with the high dimensionality and possible linear separability of relevant gene subsets. On the other hand, the performance reduction observed with the integration of multiple datasets indicates that increasing the number of samples does not necessarily imply improved predictive capability. This effect suggests the presence of transcriptomic heterogeneity, possibly associated with experimental and biological variations, which hinders model generalization.

5.2. Behavior under Data Perturbations

The observed stability under perturbations (Figure 2) suggests that the models are capable of capturing robust patterns that are minimally sensitive to data variations. This behavior is particularly relevant in biomedical applications, where experimental noise and biological variability are inherent. The consistency in model ranking across different scenarios (Figure 4) reinforces this interpretation, indicating that the relative performance among algorithms remains stable even under adverse conditions. Taken together, these results point to the reliability of the evaluated approaches in contexts closer to real-world practice.

5.3. Model Interpretability

The concentration of predictive importance in a reduced set of genes (Figure 3) suggests that class distinction may be associated with specific molecular mechanisms. The

consistency of these genes across different scenarios reinforces the hypothesis that the models capture relevant structural signals rather than dataset-specific patterns. This result is aligned with the literature, which highlights the existence of transcriptomic biomarkers associated with glaucoma, and reinforces the potential of interpretable approaches to support the understanding of disease mechanisms.

5.4. Main Contributions

This study advances the literature by demonstrating that the evaluation of models on transcriptomic data should go beyond predictive performance, incorporating robustness and stability analyses. The results show that simpler models can provide competitive and more stable solutions, whereas the integration of multiple datasets introduces relevant challenges related to heterogeneity. Furthermore, the incorporation of interpretability analyses enables not only the evaluation of model performance, but also understanding the biological patterns underlying the predictions, expanding the potential for application in biomedical contexts.

5.5. Comparison with the Literature

The results corroborate previous studies highlighting the role of molecular alterations in glaucoma progression [Tezel 2011], as well as the potential of machine learning techniques for biomolecular data analysis. However, unlike most studies in the literature, which evaluate models under controlled scenarios, this work explicitly incorporates data perturbations, enabling a more comprehensive robustness analysis. This approach contributes to a more realistic evaluation of model performance, an aspect that remains underexplored in previous studies.

5.6. Limitations

Despite the obtained results, some limitations should be considered. The limited number of samples may impact model generalization, while the integration of different datasets introduces challenges related to heterogeneity and batch effects. In addition, the absence of external validation on independent datasets limits the assessment of applicability in real-world scenarios.

6. Conclusion

This work evaluated the performance of different machine learning algorithms for glaucoma classification using gene expression data. The findings demonstrated that linear approaches, such as Logistic Regression, can achieve performance equivalent to or even superior to more complex methods, with AUC values close to 0.99. High predictive stability was also observed under different types of data perturbations, with performance maintained even after the introduction of noise and structured modifications. Furthermore, the SHAP-based analysis revealed that a reduced set of genes concentrates most of the predictive relevance, consistently across the evaluated scenarios, reinforcing the potential of these approaches for biomarker identification.

As future work, validation using independent datasets is recommended, as well as the incorporation of additional sources of information, such as clinical data and ophthalmologic images. Additionally, the exploration of more advanced interpretability methods

may contribute to a deeper understanding of the biological mechanisms associated with glaucoma. Finally, the application of these approaches in clinical contexts represents an important step toward the use of computational tools as diagnostic support systems.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001

References

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3):273–297.

Dai, M., Hu, Z., Kang, Z., and Zheng, Z. (2022). Based on multiple machine learning to identify the eno2 as diagnosis biomarkers of glaucoma. *BMC ophthalmology*, 22(1):155.

Demsar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30.

Feng, J. and Xu, J. (2019). Identification of pathogenic genes and transcription factors in glaucoma. *Molecular medicine reports*, 20(1):216–224.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32(200):675–701.

Garcia, D. and Ramos, J. (2022). Neuroinflammatory pathways and their role in glaucoma. *Frontiers in Immunology*, 13:889302.

Howell, G. R., Soto, I., Zhu, X., Ryan, M., Macalinao, D. G., Sousa, G. L., Caddle, L. B., MacNicoll, K. H., Barbay, J. M., Porciatti, V., et al. (2012). Radiation treatment inhibits monocyte entry into the optic nerve head and prevents neuronal damage in a mouse model of glaucoma. *The Journal of clinical investigation*, 122(4):1246–1261.

Kaur, M. et al. (2023). Emerging molecular biomarkers and pathways in glaucoma: A comprehensive review. *Molecular Vision*, 29:552–570.

Lee, J. and Kim, S. (2020). Augmenting transcriptomic data for deep learning applications. *PLOS Computational Biology*, 16(12):e1008446.

Liu, M., Hu, F., Lei, C., Fu, M., Li, X., and Yu, L. (2023). Candidate genes identification and rna-seq based pathway analysis associated with primary angle-closure glaucoma with cataract. *BMC ophthalmology*, 23(1):192.

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems (NeurIPS)*, 30:4765–4774.

Nemenyi, P. (1963). *Distribution-free multiple comparisons*. PhD thesis, Princeton University.

Peng, C., Hu, L., Su, W., and Hu, X. (2025). Identification of diagnostic biomarkers and dissecting immune microenvironment with crosstalk genes in the poag and covid-19 nexus. *Scientific Reports*, 15(1):25244.

Pulukool, S. K., Bhagavatham, S. K. S., Kannan, V., Sukumar, P., Dandamudi, R. B., Ghaisas, S., Kunchala, H., Saieesh, D., Naik, A. A., Pargaonkar, A., et al. (2021). Elevated dimethylarginine, atp, cytokines, metabolic remodeling involving tryptophan metabolism and potential microglial inflammation characterize primary open angle glaucoma. *Scientific Reports*, 11(1):9766.

Qiu, H., Zhu, B., and Ni, S. (2018). Identification of genes associated with primary open-angle glaucoma by bioinformatics approach. *International Ophthalmology*, 38(1):19–28.

Suo, L., Dai, W., Qin, X., Li, G., Zhang, D., Cheng, T., Yao, T., and Zhang, C. (2022). Screening of primary open-angle glaucoma diagnostic markers based on immune-related genes and immune infiltration. *BMC Genomic Data*, 23(1):67.

Tezel, G. (2011). The molecular regulation of optic nerve degeneration in glaucoma. *Progress in Retinal and Eye Research*, 30(6):369–380.

Ullah, Z., Tao, Y., and Huang, J. (2024). Integrated bioinformatics-based identification and validation of neuroinflammation-related hub genes in primary open-angle glaucoma. *International Journal of Molecular Sciences*, 25(15):8193.

Wang, D., Pu, Y., Tan, S., Wang, X., Zeng, L., Lei, J., Gao, X., and Li, H. (2024). Identification of immune-related biomarkers for glaucoma using gene expression profiling. *Frontiers in Genetics*, 15:1366453.

Wu, J., Lin, C., Yang, C., Pan, L., Liu, H., Zhu, S., Wei, S., Jia, X., Zhang, Q., Yu, Z., et al. (2023). Identification and validation of key biomarkers and potential therapeutic targets for primary open-angle glaucoma. *Science China Life Sciences*, 66(12):2837–2850.

Xie, Y. and Yu, K. (2024). Identifying hub genes for glaucoma based on bulk rna sequencing data and multi-machine learning models. *Current Medicinal Chemistry*, 31(42):7059–7071.

Xu, W., Sun, Y., Zhao, S., Zhao, J., and Zhang, J. (2023). Identification and validation of autophagy-related genes in primary open-angle glaucoma. *BMC Medical Genomics*, 16(1):287.

Xu, W., Zhang, F., and Huang, J. (2025). Identification and validation of oxidative stress signature genes in the trabecular meshwork of glaucoma. *Scientific Reports*, 15(1):31260.

Zhang, X., Chen, J., Zhou, X., Zhou, D., Liao, L., Zhao, Y., Wu, P., Nie, F., Liao, Z., Cai, Z., et al. (2025). Exploring diagnostic m6a regulators in primary open-angle glaucoma: insight from gene signature and possible mechanisms by which key genes function. *BMC Medical Genomics*, 18(1):57.

Zhao, Y. and Wang, Y. (2020). Data augmentation for rna-seq using generative neural networks. *Bioinformatics*, 36(14):4280–4287.

Zhou, Z.-H. (2012). *Ensemble Methods: Foundations and Algorithms*. Chapman and Hall/CRC.