Grenze International Journal of Engineering and Technology, January Issue

A Study of Concept for Predicting Myopia using Machine Learning

Aarthy. S[1] and Dr. S. Mangayarkarasi[2]
[1]Research Scholar, Department of Advanced Computing and Analytics, School of Computing Sciences, VISTAS
Email: saarthiy@gmail.com
[2]Professor, Department of Advanced Computing and Analytics, School of Computing Sciences, VISTAS
Email: drmangaiprabu@gmail.com

***Abstract*— In summary, we created a thorough machine learning framework in this study to forecast the course of myopia and the likelihood of developing high myopia. Our machine learning algorithms will help identify individuals who are at high risk and direct the advancement of therapeutic strategies for this prevalent ailment.**
Our study provides useful information for customizing and optimizing the medical treatment of myopia in children by correctly estimating the probability of developing excessive myopia and the course of myopia.

***Keywords*— Machine learning, Myopia.**

I. Introduction

Myopia is one of the major causes of blurred vision globally. It is still difficult to predict with accuracy when myopia will advance and when a person is at high risk of developing myopia. The goal of this research is to create a predictive model for myopia development.
Artificial intelligence (AI) is poised to transform disease diagnosis and prognosis, perhaps resulting in an overall higher level of care, thanks to new machine learning (ML) techniques and their use in the medical area.
The exponential rise in myopic cases over the last decade has proven it possible to gather large datasets that may be used as a research tool to forecast whether myopia need monitoring and active management to avoid his risk of a specified time periods.
The development of myopia and the prediction of myopic progression in five separate cohorts of children have myopia. The assessments using machine learning (ML) can offer well-informed forecasts about the course of myopia and the likelihood of developing high myopia.

II. Literature Review

“Accurate prediction of myopia progression by machine learning” [1]
To forecast the course of myopia and the likelihood of developing high myopia, we designed a multivariate linear regression algorithm model.
An AI system is trained and validated using cohorts of myopic patients.
A wealth of information could be found in the GMS data applied for training, which primarily consisted of bilateral eye tests carried out on children and young people.

Grenze ID: 01.GIJET.12.1.558

The left and right eyes were regarded to offer distinct data.
Examination of the elements causing the growth of high myopia
We applied the nonparametric Kaplan–Meier method to estimate survival curves in order to evaluate the factors impacting the likelihood of progression to high myopia. For univariate analysis of categorical data, the log-rank test was employed to identify curve differences. P less than 0.05 was regarded as significant.

Table 1.
Cohorts of myopic patients used for training and validation of an AI system.

Cohort	Training	Internal validation	External validations			
	GMS	GMS	Validation study 1	Validation study 2	SCMS	BCES
Medical records (*n*)	273 307	29 445	203 462	100 213	4 148	1 955
Patients (*n*)	88 111	10 023	76 314	50 957	2 007	131
Male (*n*)	49 993	5 778	43 418	29 031	1 072	68
Female (*n*)	38 118	4 245	32 896	21 925	935	63
Mean interval between visits ± SD (years)	1.99 ± 1.32	1.93 ± 1.26	1.99 ± 1.33	1.99 ± 1.33	0.72 ± 0.21	2.47 ± 1.26
Mean age at first exam ± SD (years)	8.18 ± 3.54	8.14 ± 3.52	8.16 ± 3.54	8.19 ± 3.56	7.91 ± 3.19	7.71 ± 1.36
SE[a] at first exam ± SD (D)	−0.67 ± 3.45	−0.63 ± 3.44	−0.67 ± 3.45	−0.66 ± 3.45	−2.03 ± 1.81	−0.50 ± 1.37

Fig 1

The SE was calculated from the refractive error, defined by the following equation:

$$Spherical\ Equivalent = Spherical\ Diopter + \frac{1}{2}(Cylindrical\ Diopter) \quad (1)$$

The annual progression of refractive error was calculated for each visit using the equation:

$$Progression = \frac{Spherical\ Equivalent_{baseline} - Spherical\ Equivalent}{t_{baseline} - t} \quad (2)$$

“Machine-learning models to predict myopia in children and adolescents” [5]
Better feature sub-set construction for model training is achieved using a feature selection technique based on both univariate and multivariate correlation analysis. In order to fill in the gaps in the original data, a GBRT-based technique is offered.
The SVM model serves as the basis for the prediction model. The exact accuracy of predictions has been found by the application of data of transformation. The outcomes says that our approach could produce accuracy and performance that is realistic.

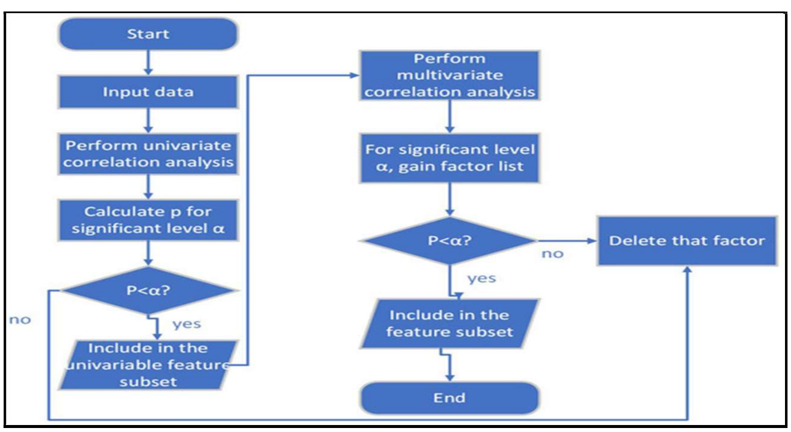

Figure 2. Flow of the Feature Selection Method

A. Assessment of Prediction Precision

We select a number of baseline approaches to compare with our approach in order to attest to its effectiveness. These foundational techniques include of BP neural network, Random Forest, KNN, Naive Bayes, and logistic regression[8].

This study presents a prediction model for teenage myopia based on behavior and measurement data from elementary school pupils.

Univariate correlation analysis helps in understanding the individual relationship between each feature and the target variable, while multivariate analysis can uncover more complex relationships and interactions between features. By using both, you can get a more comprehensive view of feature importance and redundancy.

Combining both univariate and multivariate correlation analysis for feature selection can definitely help in identifying a robust feature subset for model training.

A GBRT-based technique is offered to assist in completing the gaps in RA data. The outcomes demonstrate that our approach may yield predictions with a respectable degree of accuracy.

"Prediction of myopia in adolescents through machine learning methods".[6]

Questionnaires, ocular biological parameter measurements, and myopia screening were carried out. Five myopia prediction models were created using the random forest (RF), decision tree (DT), extreme gradient boosting trees (XGBoost), support vector machine (SVM), and logistic regression (LR) techniques.

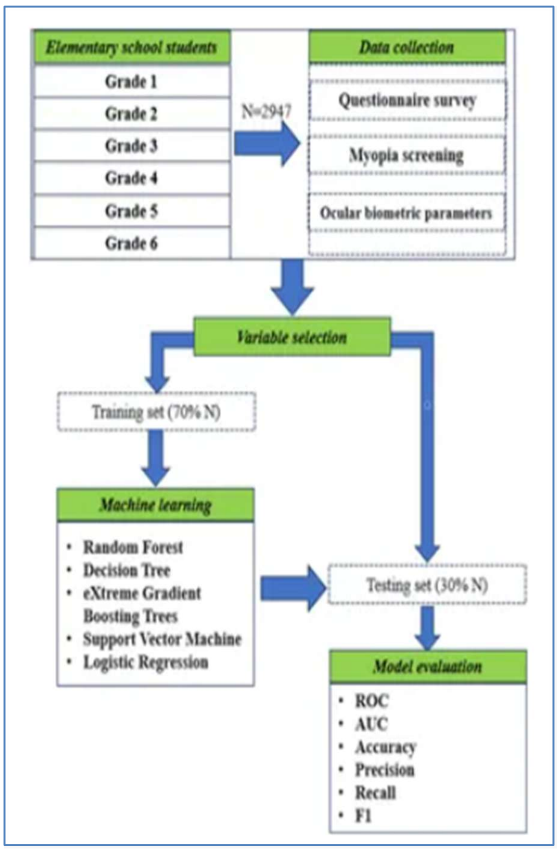

Figure 3. Machine Learning Modeling Process

TABLE I: FIVE DIFFERENT ALGORITHMS TRAINED AND EVALUATED

Algorithm	AUC (Test Set)	Accuracy	Precision	F1-Score
Support Vector Machine (SVM)	0.846	0.844	0.851	0.831
Logistic Regression (LR)	0.837	0.783	0.830	0.746
Random Forest (RF)	0.833	0.764	0.738	0.756
XGBoost	0.815	0.754	0.765	0.727
Decision Tree (DT)	0.791	0.757	0.785	0.723

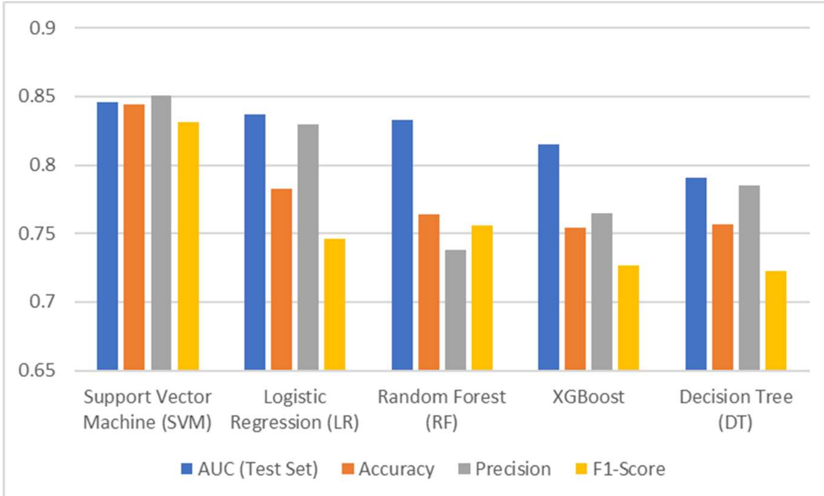

Figure 4. Algorithm Vs Evaluated values

"The association between screen time exposure and myopia" [13]

Screen time exposure was categorized as either categorical or continuous, and device types included computers, televisions, smartphones, laptops, tablets, and related digital media platforms. When sufficient data were available, we conducted a quantitative synthesis to derive pooled effect estimates. The risk of bias for each included study was independently assessed, and the certainty of evidence across studies was evaluated using established methodological frameworks.

The main advantage of this study is that it provides the latest and most comprehensive evidence on the association between screen time exposure and myopia in children and adolescents. Second, we pooled effect estimates from categorical and continuous exposure to screen time separately. Previous meta-analyses in related fields did not consider to separate processing of categorical data and continuous data

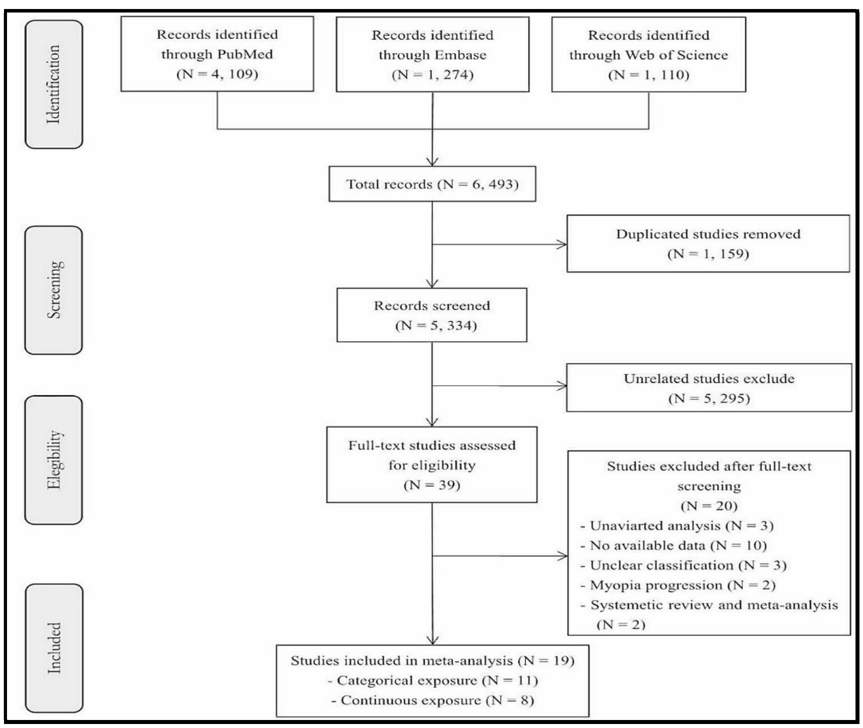

Figure 5. Literature Search for screen time exposure and myopia.

IV. RESULTS

This study involved 2,947 primary school pupils, with 47.2% diagnosed with myopia (short-sightedness) [2]. The research employed machine learning to predict myopia using various features related to demographics, behavior, and ocular measurements. The models demonstrated strong predictive power, with AUC scores above 0.75—indicating good discrimination between myopic and non-myopic students.

The study evaluated five predictive models:

- Support Vector Machine (SVM): AUC = 0.846
- Logistic Regression (LR): AUC = 0.837
- Random Forest (RF): AUC = 0.833
- XGBoost: AUC = 0.815
- Decision Tree (DT): AUC = 0.791
  - Axial length appears central to both ML and statistical modeling, consistent with known biology of myopia progression.
  - Early-life and parental factors (e.g., feeding pattern, maternal myopia) also play a strong predictive role, suggesting an interaction of genetics and early environmental exposures.
  - Sex-based differences and reading posture highlight modifiable behavioral factors that schools and parents might address.
  - The inclusion of corneal radius and anterior chamber depth as protective factors suggests that biometric profiling could inform individual risk assessments.

This study shows that machine learning models [9], especially SVM and LR, can effectively predict myopia in primary school pupils, with AUCs >0.8. The most influential risk factor is axial length, followed by biological and behavioral contributors. These findings support early biometric screening and lifestyle interventions as potential strategies to curb the onset or progression of myopia.

Machine learning-based myopia prediction models showed good predictive performance and correctly identified myopia risk variables. As a result, they might help high-risk individuals execute myopia prevention and control strategies.

V. METHODS

If building a model for myopia prediction, here are common ML approaches used in related ophthalmic contexts [10]:

- Random Forest (RF) – Strong in handling complex feature sets like biometric and lifestyle data.
- XGBoost (Extreme Gradient Boosting) – High-performance gradient boosting often used in medical outcome prediction.
- Support Vector Machines (SVM) – Effective for small, high- dimensional datasets such as eye scan features.
- Convolutional Neural Networks (CNNs) – If retinal images (e.g., OCT or fundus) are available, CNNs are ideal for spatial feature recognition.

Analysis of prediction models

A number of algorithms that can forecast the start and course of myopia, especially in children and teenagers, have been developed as a result of recent developments in machine learning Random Forest and XGBoost Models: Excellent Prediction Accuracy Childhood myopia [7] can be reliably predicted by Random Forest and Extreme Gradient Boosting (XGBoost), RandomForest machine learning models.

For instance, the XGBoost model obtained an accurate value 0.970 and an area under the curve (AUC) of 0.983 when non-cycloplegic spherical equivalent (SE), uncorrected distance visual acuity (UCDVA), axial length (AL), and age were used [3].

A. Time-Aware Deep Learning Models

Long-Term Prediction A novel time-aware deep learning model [4]has been developed to predict myopia in adolescents by incorporating vision records and various quantitative environmental data. This approach enhances prediction accuracy and allows for early identification of individuals at risk. ScienceDirect.

B. Fundus Image-Based Prediction

Deep Learning Applications Deep learning systems have been employed to predict using fundus images for high myopia [4] . These models, which include image-based, clinical, and combined approaches, have shown results in forecasting high myopia development during teenage years.

C. PubMed Interpretable Machine Learning Models:

Understanding Risk Factors Interpretable machine learning models have been utilized to predict childhood myopia by assessing various diagnostic indicators. These models help in understanding the influence of different factors on myopia development, aiding in the implementation of prevention and control measures. These advancements in machine learning and deep learning offer promising tools for early detection and intervention in myopia, potentially mitigating its progression and associated complications.

VI. CONCLUSION

Despite the lack of direct studies on ML-based myopia prediction, algorithms such as Random Forest and XGBoost proven effective in other ophthalmic conditions are promising candidates.

A successful model would likely require multimodal inputs, including biometric, behavioral, and environmental data, with the potential to enhance early diagnosis and targeted interventions.

Machine training-based myopia prediction models showed good predictive performance and correctly predicted myopia risk variables; as a result, they might help high-risk individuals adopt myopia prevention and control strategies.

Our study demonstrates accurate prediction of myopia progression and risk of high myopia providing valuable insights for tailoring strategies to personalize and optimize the clinical management of myopia in children.

In summary, in this study we developed a comprehensive ML framework for prediction of myopia progression and the risk of high myopia. Our ML models will aid in identifying high-risk patients and help guide developments in advancing therapeutic interventions for this common condition.

A prediction model of myopia in adolescents based on both measurement and behaviour data of primary school students is presented. A feature selection method based on both univariate correlation analysis and multivariate correlation analysis is used to better construct a feature sub-set for model training. A method based on GBRT is provided to help fill in missing items in RA data. The prediction model is built based on SVM model. Data transformation has been used to improve the prediction accuracy. The results show that our method could provide reasonable prediction accuracy.

Also, according to our results, we found out that some factors have positive correlation with myopia, like DAI, JG, YG, AL, K1, K2, WHIM, which we call them protective factors. While some other ones, like JTR, YTR, PULSE, COLA, REDM have negative effect on the myopia, which we call them dangerous factors.

Since our goal is to build a model which could predict the myopia in adolescents, and deduce the relation between myopia and different factors, so that we could formula policy to help preventing myopia. Future work might include more thoroughly analysis of the result and discussing of possible improvements in the methodology.

Machine learning models—especially SVM and Random Forest—demonstrated strong predictive capability for identifying children at risk of myopia.

REFERENCES

[1] J. Mu, H. Zhong, and M. Jiang, “Machine-learning models to predict myopia in children and adolescents,” Frontiers in Medicine, vol. 11, 2024.

[2] Y. Wang, R. Du, S. Xie, C. Chen, H. Lu, and J. Xiong, “Machine learning models for predicting long-term visual acuity in highly myopic eyes,” JAMA Ophthalmology, vol. 141, no. 5, pp. 432–441, 2023.

[3] S.-M. Li, M.-Y. Ren, J. Gan, S.-G. Zhang, M.-T. Kang, H. Li, D. A. Atchison, J. Rozema, A. Grzybowski, and N. [Wang, “Machine learning to determine risk factors for myopia progression in primary school children: The Anyang Childhood Eye Study,” Eye and Vision, vol. 9, no. 4, pp. 1–10, 2022

[4] H. Zuo, B. Huang, J. He, L. Fang, and M. Huang, “Machine learning approaches in high myopia: systematic review and meta-analysis,” Journal of Medical Internet Research, vol. 27, 2025.

[5] J. Li, S. Zeng, Z. Li, J. Xu, Z. Sun, and J. Zhao, “Accurate prediction of myopic progression and high myopia by machine learning,” Precision Clinical Medicine, vol. 7, no. 2, pp. 105–116, 2024.

[6] H. Lin, E. Long, X. Ding, H. Diao, Z. Chen, and R. Liu, “Prediction of myopia development among Chinese school-aged children using refraction data from electronic medical records: a retrospective, multicentre machine learning study,” PLoS One, vol. 13, no. 2, e0195343, 2018.

[7] H.-J. Tong, Z.-M. Huang, Y.-L. Li, Y.-M. Chen, and B. Tian, “Machine learning to analyze the factors influencing myopia in students of different school periods,” Frontiers in Public Health, vol. 11, 2023.

[8] T. Tang, Z. Yu, Q. Xu, Z. Peng, Y. Fan, K. Wang, and Q. Ren, “A machine learning-based algorithm used to estimate the physiological elongation of ocular axial length in myopic children,” Eye and Vision, vol. 7, no. 1, 2020.

[9] X. Yang, G. Chen, Y. Qian, Y. Wang, and Y. Zhai, "Prediction of myopia in adolescents through machine learning methods," International Journal of Environmental Research and Public Health, vol. 17, no. 13, 2020.

[10] R. Zhang, L. Dong, Q. Yang, W. Zhou, H. Wu, Y. Li, and H. Li, "Screening for novel risk factors related to high myopia using machine learning," BMC Ophthalmology, vol. 22, no. 1, pp. 1–10, 2022.

[11] K. A. Rose, I. G. Morgan, J. Ip, A. Kifley, S. Huynh, W. Smith, and P. Mitchell, "Outdoor activity reduces the prevalence of myopia in children," Ophthalmology, vol. 115, no. 8, pp. 1279–1285, Aug. 2008.

[12] T. Tang, Z. Yu, Q. Xu, Z. Peng, Y. Fan, K. Wang, Q. Ren, J. Qu, and M. Zhao, "A machine learning-based algorithm used to estimate the physiological elongation of ocular axial length in myopic children," Eye and Vision, vol. 7, no. 50, Oct. 2020, doi: 10.1186/s40662-020-00214-2.

[13] The association between screen time exposure and myopia in children and adolescents: a meta-analysis," Research, vol. 24, art. no. 1625, Jun. 2024. [Online].