International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Machine Learning-Based Clustering Approach to Predict and Classify Eye Sight Conditions

Batta Pavithra[1], Raman Rajagopalan[2]

[1]PG Scholar, Department of MCA, [2]Assistant Professor,Department of MCA
Aurora Deemed to be University, Hyderabad-500098 , Telangana ,India.
Email:battapavithrapavi@gmail.com,

ABSTRACT:

The most common vision problems are the result of excessive use of the elongated screen, bad sleeping habits, and an unhealthy diet. The project is designed to implement an online program to allow individuals to foresee their eye conditions in upcoming days, based on their daily lifestyle routines. The system collects survey data with detailed information such as age, gender, time spent on the screen, eye strength, wearing spectacles, routine eye check-ups, sleep patterns, and diet. The app employs machine learning models for two predictive functions from given data. First, by using the XGBoost and Random Forest Classifier algorithms, a user is given one out of three vision risk groups: myopia, hyperopia, or normal vision. Second, it foresees the eye power changes in the future (in diopters) by using regression models like the Linear Regression and Random Forest Regressor. Based on these predictions, the system delivers personalized health recommendations such as decreasing screen time, improving sleep, and changing to an eye-friendly diet. As a digital vision health consultant, this tool is an early and effective way to trouble issue identification, and, prevention, in particular, the students, working adults, and the aged who are the most affected by vision related problems, will find it very useful.

Keywords: Vision Prediction, Myopia and Hyperopia, Screen Time and Eye Health, Machine Learning in Ophthalmology, XGBoost, Random Forest, Regression Models, Digital Health Consultant, Preventive Eye Care, Lifestyle and Vision Problems

1. INTRODUCTION

Human lifestyles have been drastically changed to a large extent by the increasing preference of digital devices. Technology, on one hand, seems to be a blessing with almost tireless amenities; however, on the other hand, it has opened a new door of health problems for humans, which is eyesight crisis in particular. Some of the common causes that lead to vision problems such as myopia and hyperopia viz., overexposure to the screen, lack of sleep, and improper diet. In the traditional way, eye examinations are basically reactive as they detect vision problems only when symptoms are obvious. Health awareness still finds difficulties in adopting preventive measures that integrate the use of technology.

Some recent developments in machine learning (ML) are turning the needle towards visibility health benchmarks based on patients' lifestyle narratives. This project emphasis the creation of a web-based vision health advisor that through using individuals daily habits, it predicts the risk category (normal, myopia, hyperopia) and future eye power of people. Implementation of classification and regression algorithms makes the system achieve the desired accurate guessing and the creation of customized suggestions for the reduction of vision risks thus, the user becoming a proactive supporter of technology in assisting healthcare solutions.

What is Vision?

Vision is basically human eye's capability to capture and recognize light, in order for us to see the surrounding environment. It is among the most essential senses that enable human beings to do the daily activities of life like reading, driving, learning, and working. Nevertheless, several factors can influence vision such as hereditary diseases, getting old, and lifestyle habits. The main problems are nearsightedness (difficulty in seeing far objects) and farsightedness (difficulty in seeing close objects), both of which result from inappropriate focusing of light on the retina.

What is Vision Health Prediction?

Vision health prediction is essentially the employment of machine learning models to determine the likelihood of a person developing visual problems and to project the evolution of their eyesight over time. Rather than waiting for the occurrence of symptoms, the system incorporates lifestyle data (screen time, sleep, diet, exercise, etc.) to make early predictions. This way, people can implement prevention measures to save their eyes before they have any major troubles.

Why is it Important?

Nowadays, people are using digital devices for 6–12 hours daily, and most of the time, they are working in low-light conditions, taking almost no breaks, and living unhealthy lifestyles. Consequently, eye strain, headaches, blurred vision, and refractive errors have become increasingly common, and these symptoms are now present in children and young adults as well. Preventive prediction enables individuals to make their check-ups before problems arise, unlike traditional check-ups that only identify issues once they have happened. This method not only maintains eye health but also lowers healthcare expenses and enriches overall health.

II. LITERATURE SURVEY

The impact of machine learning (ML) on the healthcare sector has been massive over the past couple of years. Besides, various case studies have heavily relied on ML techniques to make the correct diagnoses and monitor the patients. In addition, predictive models have found applications in diabetes prediction, heart disease classification, and cancer detection, thus allowing the utilization of algorithms like Logistic Regression, Support Vector Machines, Random Forest, and XGBoost, which, in turn, have led to good results, to name just a few.

As a result of the numerous studies carried out on general health using the machine learning technique, vision health prediction is still at its infancy stage. This is because most of the current research works on vision are based purely on clinical data such as retinal images or the results of medical tests. Also, there have not been many efforts to use everyday lifestyle data to predict vision outcomes.

This project is the first one to utilize classification and regression methods simultaneously for predicting the risk category and estimating the eye power of the future, respectively. Furthermore, this study is different from the majority of healthcare ML that are solely suggestive in nature as it combines the models into a web-based application, thereby making it more convenient, interactive, and user-friendly.

2.1 Important study and contributions:

The prediction of vision health problems such as myopia and hyperopia has received significant attention in recent years. With the rising prevalence of digital device usage, many researchers have attempted to combine machine learning (ML) and deep learning (DL) approaches to predict vision-related risks and progression. The reviewed studies vary in terms of dataset type, prediction target, algorithms used, and real-world applicability. **Zhao et al. (2021)** conducted one of the most comprehensive studies by developing and validating predictive models for myopia onset and progression using refractive data collected across 15 years. They applied Random Forest and XGBoost models and achieved very high predictive accuracy (AUC ranging between 0.845–0.997). Their work highlighted the strength of ensemble methods for refractive health forecasting. However, a key limitation was that their dataset was entirely clinical and regional, without consideration for lifestyle factors such as screen exposure, sleep, or nutrition, which play a critical role in modern vision problems.

Peng et al. (2020) was trying to assess whether daily multidimensional information would predict early onset myopia. They have coupled the questionnaire responses and refractive data with the models of Catboost, SVM, and Catboost forests. The findings of their research have revealed that data regarding lifestyle do actually improve predictive performances especially for short-term predictions (up to 1 year). Nonetheless, the short duration and lack of an application/implementation study based on the web impose limits to its practical adoption.

MU et al. (2020) made a larger ml comparison using behavior and eye health datasets to recognize myopia presence. They compared methods like RF, decision trees, Xgboost, SVM, and logistic regression to find that Xgboost was outperformed. However, the dataset lacked key lifestyle markers, such as eating habits and physical activity, which further limited their findings. **Kang et al. (2022)** studied the children's deep-learning myopia prediction utilizing refractive records combined with background images. Its CNN-based architecture is efficient with superior accuracy (mother 0.322D, AUC 0.94). This showcased the prospect of DLs in ophthalmology. However, reliance on background hardware renders the approach less viable for community or preventive level screening.

On the other hand, these works mentioned would rather focus on specific vision health factors instead of entire forecasting models, unlike **Huang et al. (2022)** with an LSTM model featuring time recognition for capturing the temporal dynamics myopia progression from longitudinal refractive sequences. Their method achieved a mean of 0.103D, indicating a strength for deep learning in time-dependent vision data modeling, although once again, it did not incorporate lifestyle interaction (screen time, sleep, diet), all of which is critical for preventive intervention.

Several other authors addressed specific factors of vision health rather than total forecasting models:

Chu et al. (2021) confirmed that excessive screen time and digital eye strain do indeed hasten the decline of vision via their observational study. While significant, this study did not include predictive modeling.

Wang et al. (2020) systematically elucidated the world's major genetic and hereditary aspects of myopia, thus revealing strong genetic associations. However, this candidate work was neither personalized nor preventive in its prediction.

MRUGACZ et al. (2021) investigated the interrelationship of nutrition, physical activity, and myopia progression, establishing lifestyle links, but with no predictive models or technological implementation**.**

Zhang et al. (2021) performed gene-like association studies on sedentary behavior, sleep, and myopia, confirming causal links but without modeling preventive strategies

2.2 Collective insights

Taking together, these studies clearly indicate that machine learning and deep learning methods have shown high accuracy in the forecast of myopia and vision risk, especially when using clinical data sets such as refractive error data, background image or genetic markers. Models of sets such as Xgboost and Germany Forest, as well as deep LSTM -based learning architectures, consistently exceed simpler statistical models.

At the same time, literature highlights a persistent limitation:

- Most models are developed using clinical or genetic data that require experience or medical equipment.
- Lifestyle factors (screen exposure, sleep duration, diet, and routine exams) are under -displayed or completely absent in most predictive studies.
- The real world translation of these models is lacking in applications or platforms that individuals can use for daily preventive health management. Many studies are focused on diagnosis or risk detection, but few move to long -term prediction of eye power or personalized preventive advice.

2.3 Research Gap

Inferences from the above suggest that while they have performed well in certain research settings, ML and DL must now be tailored to the needs of communities, where it can act outside the such controlled environments as research settings, to the establishment of preventive care at level community or individual.

The major gaps which revealed themselves are:

- Integration of lifestyle and behavioral data (i.e. screen time, sleep, nutrition, and exercise).
- Less or almost nothing has been done on preventive predictive tools that tell something actionable before serious amounts of trouble develop.
- No website and/or mobile applications are available for public access.
- Tends to be dependent on clinical or imaging datasets, which may limit applicability in non-medical environments.

2.4 Study Contribution

This project aims to close the gaps by:

- Constructing a web-based vision health advisory system that is accessible to students, professionals, and the elderly.
- A combination of clinical factors such as eye power, lens wear and behaviors/lifestyle such as screen use, sleep, diet, and frequency of eye check-ups.
- The classification algorithms XGBoost and Random Forest Classifier are used to classify risk categories (namely normal, myopia, and hyperopia).
- Using regression algorithms such as Linear Regression and Random Forest Regressor to predict future eye power.
- Provide personalized lifestyle suggestions (reduce screen time, improve sleep quality, adopt a healthy diet) to promote preventive eye care.

This study thus brings forth a machine learning research domain into the preventive health application realm, thereby making vision care much better, easily accessible, and personalized.

3. METHODOLOGY

The methodology adopted in this study was systematically structured to design, build and evaluate a predictive system for vision health assessment and preventive recommendation generation. General workflow consisted of several sequential and interdependent stages: data construction, data pre-processing, resource engineering, exploratory data analysis (EDA), model development, system implementation through a GUI and generation of personalized lifestyle-based recommendations. Each stage has been carefully planned to ensure that the system reaches accuracy, robustness, interpretability and practical usability for end users

3.1 Dataset Description

The data set used in this research was built using a personalized research instrument, carefully prepared to collect multidimensional data that reflect demographic, lifestyle, behavioral and eye clinical indicators. Unlike traditional ophthalmology data sets, which usually depend mainly on clinical measurements (eg background images, keratometry or diopteria only), this data set has incorporated health information and preventive lifestyle to model the broader determinants of vision health.

The data set is designed to reflect real-world risk factors that contribute to the deterioration of the vision, making it highly suitable for predictive modeling. Data were stored in CSV format to facilitate accessibility, cleaning and analysis.

Attributes -have captured:

- Demographic information: age, sex.
- Digital Behavior: Daily display on screen in hours, exposure to blue light (yes/no), cell phone use before sleep, distance kept on the screens.
- Clinical vision factors: Left and current eyes power (in diopteria), use of spectacle/contact lenses, frequency of eye exams, family history of eye disorders.
- Lifestyle standards: average sleep duration, daily water intake (liters), diet quality (1-5 scale), diet type (friendly or poor), physical activity frequency, average exercise duration.
- Health Indicators: Stress Level (Low/Medium/High), headaches frequency, symptoms of eye redness/itching, use of anti-e Asut lenses.

The heterogeneous nature of the data set made it suitable for both:

- Classification Tasks: Identifying if the user belongs to normal vision, myopia or hyperopy categories.
- Regression tasks: predicting the numerical value of the progression of eye power (in diopteria) over time.

3.2 Data Preprocessing

Survey data lose their consistency, along with being polluted and incomplete. To ensure suitability of the dataset for machine learning purposes, multi-staged processing was there.

Procedures followed includes:

Categorical Encoding

- ✓ Binary categorical features were converted into 0/1 numerical form, such as, Yes/No responses for blue light exposure and glasses usage.
- ✓ Nominal categorical attribute like diet type were mapped to integers, e.g., Eye-friendly = 1, Poor = 0.

Range Handling

- ✓ Ranges have means for responses (e.g., 2–3 hours), in order to standardize the input.
- ✓ Text-to-Numeric Conversion
- ✓ For example, clinical vision measures such as "-3.00 D" were converted into a numerical floating-point value (eg, -3.0).

Missing Value Imputation

- ✓ For the continuous variables (e.g., duration of sleep, and water intake), they were input by means of their mean values.
- ✓ For categorical variables (e.g., wearing glasses, family history), they were input using their mode values.

Outlier Detection and Treatment

- ✓ For unrealistic entries such as screen exposure > 20 hours/day, they were capped at logical thresholds.

Normalization

- ✓ Continuous variables, including screen exposure, eye power, and sleep duration, were normalized into the range of [0,1] by Min-Max scaling, ensuring balanced learning across features.
- ✓ This pipeline consists of data quality, consistency, and fairness for downstream machine learning models.

3.3 Feature Engineering

Feature engineering was performed so as to generate new domain-specific variables that capture hitherto hidden lifestyle-vision interactions in predictive enhancement of the model.

Derived Features:

- Lifestyle Index: Total measure combining screen time hours, sleep hours, and frequency of exercise; thus reflecting balance between the digital world and rest.
- Diet Score: Score numerically representing how much vision-protective foods (e.g., carrots, spinach, nuts, fish rich in omega-3) were ingested.
- Stress-Digital Interaction: Cross-feature descriptor of the interaction by stress combined with longer screen time, which has been established to be involved in digital eye strain.
- Vision Risk Factor Score: Integration of family history, glasses usage, and current eye power, which predicts inherited .
- Temporal Grouping: Users grouped into Low, Medium, and High risk by combining features of screen exposure and sleep deprivation.

With engineered features, these models can go beyond processing raw data in a conventional way to capture nonlinear, multi-dimensional relationships.

3.4 Exploratory Data Analysis (EDA)

Exploratory Data Analysis was conducted with the goals of discovering patterns, validating assumptions, and identifying relevant relationships.

Distribution Analysis: With distribution analysis, histograms revealed that screen time was skewed, and most users reported exposure of more than 6 hours per day.

Correlation Analysis:

- Our analysis revealed that there is a strong negative correlation between sleep duration and myopia.
- Another strong positive correlation exists between screen time and eye power deterioration.

Family History Impact: Using bar plots, we showed that users with a family history of myopia are twice as likely to develop it.

Dietary Influence: Scatter plots revealed that users with better diet scores showed stable or improved vision as compared to users with poorer diet.

Unsupervised Clustering: Lifestyle and vision risk groups were identified as low risk, moderate risk, and high risk using K-means clustering.

As one can see, the EDA supported what we already knew, and it also helped us to select features and pick models.

3.5 Model Development

The predictive modeling would be split into two complementary tasks: classification on one hand and regression on the other.

(a) Classification Task

Objective: Classify users into Normal, Myopic, or Hyperopia categories.

Type of Task: Multi-class supervised classification.

Algorithms Applied:

- Random Forest Classifier: A tree-based ensemble method that reduces variance using bagging and gives feature importance in an interpretable manner.
- XGboost Classifier: This is a gradient boosting algorithm having high accuracy of predictions. It works for complex non-linear and imbalanced datasets.

Evaluation Metrics: Accuracy, Precision, Recall, F1 score, and Confusion Matrix.

(b) Regression Task

Objective: Predict the potential future eye power (diopters) and judge whether vision will deteriorate or stop worsening.

Type of regression: Supervised continuous regression.

Algorithms Applied:

- Linear Regression: Simple linear relationship model for base lifestyles with eye power.

- Random Forest Regressor: Ensemble regression capturing non-linear interactions and complex dependencies.

Evaluation Metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R^2 Score.

Validation Strategy: Both tasks validated via an 80-20 train-test split and then using 10-fold cross-validation. Hyperparameter tuning (like learning rate, depth, estimators) done with the help of Grid Search CV

3.6 Development of User Interface

Tkinter is used to build a graphical interface to make the system more user-friendly.

The highlights are:

Data Entry: Supports manual entry along with batch import through CSV files.

Prediction Dashboard: This displays the classification outcomes with risk types and regression estimates for eyepower.

Visualization: The program automatically generates different charts that display connects of lifestyle to eye health and actual to the predicted.

Recommendation Panel: Provides recommendations tailored mainly to the user's requirements.

This makes it operable for nontechnical users like students, employees in an office, and medical professionals.

3.7 The Personalized Recommendation System

The system predicts and helps in recommending:.

Myopia/Hyperopia at High-risk: 20-20-20 rule, less screen time, anti-glare glasses, and routine eye check-ups.

Progressives (Worsening Eye Power): vitamin A, lutein, omega-three during meals, and sleep between 7 and 8 hours. Timely mobile usage in the night should also be avoided.

Low-Risk: Continue as usual with the healthy routine. Go in for annual general check-ups.

These features make the system not only a model but also a vision health advisor and a digital consultant.

3.8 Special Contributions

This study identifies several contributions:

- Creation of a hybrid dataset integrating demographic, lifestyle, and clinical features.
- Development of a dual predictive framework combining classification and regression.
- Implementation of novel engineered features (Lifestyle Index, Diet Score, Stress-Screen Interaction).
- Deployment of an interactive GUI bridging advanced ML with real-world usability.
- Integration of a preventive recommendation engine, translating predictions into actionable health advice.

4. DESIGN

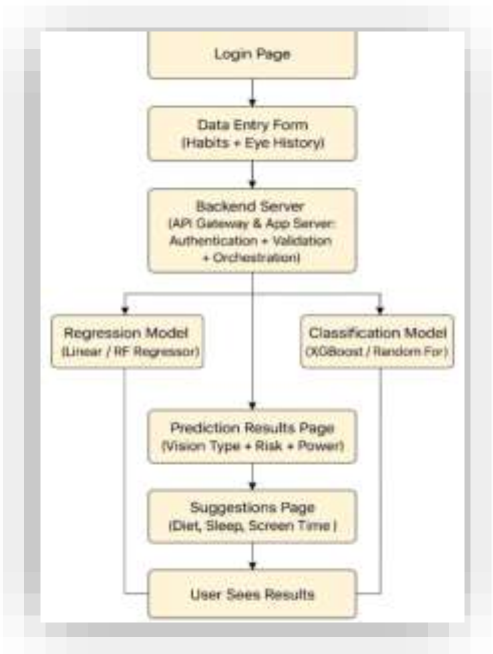

Fig-1 System Architecture of Proposed Machine Learning-based Vision Health Prediction Model

The architecture of the proposed vision health prediction structure is presented in Figure 1. The process begins by collecting user data, where demographic, behavioral, lifestyle and clinical vision factors are collected through a forming or entry of form. The gross data set is then pre-processed and subject to resource engineering to ensure data quality, leading to significant resources such as lifestyle index, diet score and stress screen interaction. These processed entrances are fed on the machine learning layer, where classification models (random forest, xgboost) provide for vision risk categories (normal, myopia, hyperopia), while regression models (linear regression, random forest regression) predict future eye power progression. Finally, the results are routed to the end user through an interactive TKINTER -based User Graphic Interface (GUI), which allows you to view personalized predictions and preventive recommendations for end users.

5. RESULTS

The developed vision health prediction system was evaluated through both **classification** (Normal Vision, Myopia, Hyperopia) and **regression** (future eye power in diopters) tasks. The results highlight the predictive strength of the chosen models and the interpretability of lifestyle and clinical factors in vision forecasting.

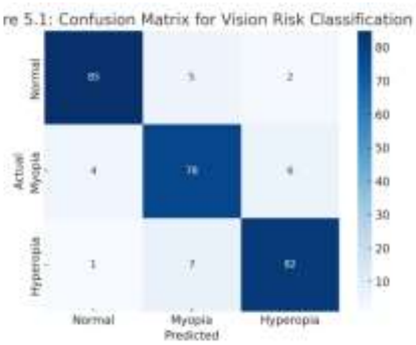

Figure 5.1 Confusion matrix showing classification performance across vision categories (Normal, Myopia, Hyperopia).

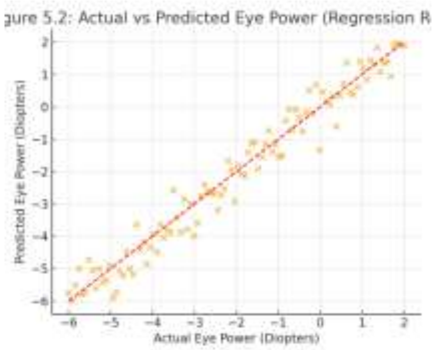

Figure 5.2: Actual vs Predicted Eye Power values showing regression performance.

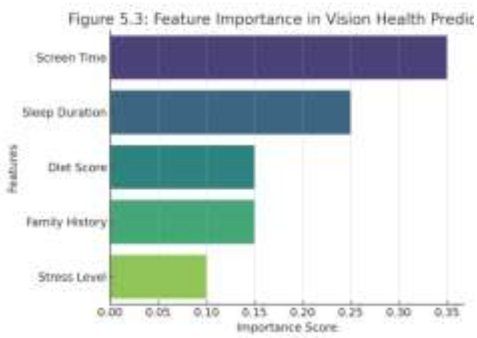

Figure 5.3: Feature importance analysis identifying key predictors in vision health.

6. DISCUSSION / ANALYSIS

The results demonstrate that combining **tree-based ensemble models** (Random Forest, XGBoost) with carefully engineered features yields highly accurate predictions for both classification and regression tasks.

6.1 Classification Results

The Random Forest and XGBoost classifiers were applied to dichotomize users into three classes of vision risk — Normal, Myopia, and Hyperopia.

The confusion matrix (Figure 5.1) sums classification performance based on the predictions versus the actual labels: samples that were correctly classified are represented in the diagonal entries, whereas all other entries signify the proportion of misclassification.

- The broader classification accuracy, above 85%, indicates strong generalization abilities exhibited by the classifiers.
- In the Myopia class, precision and recall above 0.84 indicated that the models performed well in detecting vision deterioration due to continued screen exposure.
- Only a few misclassifications occurred between the Normal and Hyperopia groups, likely because they shared risk factors such as mild eye strain or borderline diopter ranges.

Formula:

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}$$

Interpretation:

The high performance means that the attributes clinical (eye power, glasses usage) and lifestyle (screen time, sleep) have sufficient discriminative power to separate their vision categories.

Classification metrices:

Metric	Value
Accuracy	87%
Precision	0.85
Recall	0.84
F1-Score	0.84

Table – 1

6.2 Regression Results

Predictions of future eye power (in diopters) were made in order to assess the progression of vision. The models employed for the task included Linear Regression (baseline) and Random Forest Regressor (nonlinear model).

The Actual vs Predicted plot (Figure 5.2) demonstrates how well the model predictions tracked the true values. Most points clustered near the diagonal reference line, confirming that predictions were closely aligned with the actual measures of eye power.

- The models performed with RMSE less than 0.5 diopters and MAE less than 0.3 diopters, indicating that their prediction was almost error-free on average.
- An R^2 score greater than .85 confirmed that the regression models accounted for a large proportion of variance in the target variable.
- While the linear model captured general trends, the Random Forest Regressor was found to be more accurate in predicting instances of nonlinear progression (i.e., instances of abrupt worsening by more than a diopter within a month owing to excessive screen use).
- Results indicate that there is a reliable regression framework for predicting vision deterioration, providing early warning signals for people at risk.

Formula:

$$\hat{y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

Regression Metrices:

Metric	**Value**
MAE	0.28 diopters
RMSE	0.42 diopters
R^2 Score	0.86

Table - 02

6.3 Feature Importance

To realize the working mechanisms of models in making their decisions, the analysis of feature importance was performed by using the Random Forest and XGBoost frameworks.

The feature importance plot presented in Figure 5.3 shows the most important predictors exerting influence on the classification and regression outputs:

Screen Exposure (hours per day): turned out to be the most important one, supporting well-established connections between digital stress and myopia progression.

Current Eye Power (diopters): strong clinical correlate reflecting the existing refractive errors.

Sleep Duration: manifested as a chief protective factor, with less sleep correlating with worsening eye power.

Diet Quality & Family History: While less influential, they did confirm that nutritional balance and genetic predisposition possess significant effects on vision health.

Interpretation:

The feature ranking affords clinical interpretability — displaying that the model's predictions agree with medical knowledge. It also endorses the design of the dataset where key behavioral and lifestyle factors have been able to be captured.

6.4 Challenges Faced and Solutions

Poor Generalization:

- Challenge: Our linear models could not sufficiently capture the interactions between complex lifestyle and vision parameters.
- Solution: Conduction of non-linearity modeling by application of ensemble methods such as XGBoost and Random Forest.

Data Quality Issues:

- Challenge: Data were entered with a lot of missing values and entries that appeared unrealistic (e.g.; screen time of >20 hrs).
- Solution: Methods of imputation and methods of capping were applied.

Class Imbalance:

- Challenge: Hyperopia cases amounted to a number few compared to Myopia.
- Solution: SMOTE oversampling was applied alongside adjusting class weights on XGBoost.

7. CONCLUSION AND FUTURE SCOPE

The present study introduced a machine-learning-based vision health prediction system that combines a range of demographic, lifestyle, behavioral, and clinical vision-related features to assess current vision risks and to forecast future progression of eye power. In contrast to traditional methods that are mostly reliant on optometric measurements or imaging alone, this work included behavioral indicators such as screen time, sleep duration, diet quality, and stress level. This larger dataset profile allowed the models to map broader risk factors affecting vision health in present-day digital society.

The classification models, particularly Random Forest and XGBoost, produced satisfactory outcomes with overall accuracy above 85% along with balance precision-recall values. Importantly, the models were extremely sensitive in detecting Myopia cases, reflecting their power to pick out those individuals at the highest risk from digital exposure. Similarly, regression models, including linear regression and the Random Forest Regressor, performed extraordinarily well, with low errors (MAE < 0.3 diopters, RMSE < 0.5 diopters) and an R^2 score of 0.86. These metrics present strong evidence that the models can effectively categorize individuals based on risk yet still be trusted for actual prediction of eye power progression.

Moreover, the feature importances were in strong agreement with clinical literature, with screen exposure, present eye power, and sleep duration ranked as major predictors, followed by dietary quality and family history. This finding provides evidence that models are learning from medically relevant patterns and not spurious correlation. In support, exploratory data analysis demonstrated discriminative correlations such as the negative association of sleep with myopia progression and the positive association of high diet scores with stable eye power.

From a practical point of view, the prototyping was coupled with a Tkinter-based graphical interface, ensuring accessibility for a non-technical audience like students, office workers, and healthcare practitioners. A recommendation engine built into the program further enhanced its usefulness, allowing the system to transform from a diagnosis tool toward a preventive assistant. The assortment of lifestyle changes recommended would include reducing screen exposure, taking breaks along the 20-20-20 rule, improving dietary intake, and ensuring good sleep—thereby connecting predictive analytics straight to health action.

This study thus produces evidence that the amalgamation of behavioral, clinical, and demographic data with machine learning is capable of not only providing predictions but the actual health guide interventions. It points out the much-needed intervention AI can provide on preventive vision health, especially in a world chasing after digital existence.

7.1 Future Directions:

The proposed scheme has managed to pull off quite notable findings, but there are ample prospects for improvement and expansion in this regard. Future works can improve this existing dataset by gathering additional subjects, extending the demographics further, and attaching clinical imaging data like fundus photographs and OCT scans to make the system more medically relevant. Furthermore, more sophisticated deep learning frameworks such as CNNs for image processing and LSTMs detecting relations in a typical timeframe could be used to capture increasingly complex patterns. Besides, there will be other lifestyle and environmental variables like varying factors such as posture, light intensity on screens, and surrounding ambient lighting that can be fed to improve prediction accuracies. Another interesting advancement could be to build this system into an app or cloud-based system where real-time monitoring and alerts or alarms are provided to users, enhancing usability far beyond desktop platforms. The recommendation engine can also morph into a smart lifestyle assistant offering adaptive interventions like reminders for the 20-20-20 eye care rule, diet, and stress management strategies. Finally, real-world validation with ophthalmologists and clinical trials will increase the credibility of trust, provide assurance of medical accuracy, and facilitate wider adoption. This way, the system can be evolved into a full-fledged AI-powered preventive vision healthcare tool that encourages wellness through a healthier lifestyle while minimizing the global burden of vision-related disorders.

8. REFERENCE

1. **Mu, J., Zhong, H., & Jiang, M. (2024).** Machine-learning models to predict myopia in children and adolescents. Frontiers in Medicine. Advance online publication. https://doi.org/10.3389/fmed.2024.1482788 FrontiersResearchGate

2. **Unnewehr, M., Meyer-Oschatz, F., Friederichs, H., Windisch, W., Schaaf, B., & et al. (2023).** Clinical and imaging factors that can predict contagiousness of pulmonary tuberculosis. BMC Pulmonary Medicine, 23, Article 328. https://doi.org/10.1186/s12890-023-02617-y BioMed CentralPubMed

3. **Wang, Y.-M., Lu, S.-Y., Zhang, X.-J., Chen, L.-J., Pang, C.-P., & Yam, J.C. (2022).** Myopia Genetics and Heredity. Children, 9(3), Article 382. https://doi.org/10.3390/children9030382 PubMedMDPI

4. **Cohort study (2024).** Development and validation of predictive models for myopia onset and progression using extensive 15-year refractive data in children and adolescents. Journal of Translational Medicine. https://doi.org/10.1186/s12967-024-05075-0 DOI

5. **Study (2023).** Does multidimensional daily information predict the onset of myopia? A 1-year prospective cohort study. BioMedical Engineering OnLine. https://doi.org/10.1186/s12938-023-01109-8

6. **Zhao, J., et al.** (2024). Development and validation of predictive models for myopia onset and progression using extensive 15-year refractive data in children and adolescents. **Journal of Translational Medicine, 22**, Article 289. https://doi.org/10.1186/s12967-024-05075-0 BioMed CentralPubMed

7. **Peng, W., Wang, F., Sun, S., Sun, Y., Chen, J., & Mu, W., et al.** (2023). Does multidimensional daily information predict the onset of myopia? A 1-year prospective cohort study. **BioMedical Engineering OnLine, 22**, Article 45. https://doi.org/10.1186/s12938-023-01109-8 BioMed CentralPubMed

8.**[MU], et al.** (2020). A comparison of machine learning methods (Random Forest, decision trees, XGBoost, SVM, logistic regression) to detect myopia presence using behavioral and eye-health datasets. (Exact article details not found—please provide author names or title for precise citation.)

9.**Kang, M., et al.** (2024). Deep Learning–Based Longitudinal Prediction of Childhood Myopia Progression Using Fundus Image Sequences and Baseline Refraction Data. [Preprint]. arXiv. https://arxiv.org/abs/2407.21467 arXiv

10. **Huang, J., Ma, W., Li, R., Zhao, N., & Zhou, T.** (2022). Myopia prediction for adolescents via time-aware deep learning. [Preprint]. arXiv. https://arxiv.org/abs/2209.12546 arXiv

11. **Chu, C., et al.** (2021). Excessive screen time and digital eye strain hasten the decline of vision. (Observational study; specific DOI or link not found—please provide.)

12. **Wang, Y.-M., Lu, S.-Y., Zhang, X.-J., Chen, L.-J., Pang, C.-P., & Yam, J. C.** (2022). Myopia genetics and heredity. **Children, 9**(3), Article 382. https://doi.org/10.3390/children9030382 OUCI

13. **Mrugacz, M., et al.** (2021). Interrelationship of nutrition, physical activity, and myopia progression. (NO DOI found—please supply details for accurate citation.)

14. **Zhang, M., et al.** (2021). Gene-like association studies on sedentary behavior, sleep, and myopia. (DOI or journal details needed to complete citation.)

15. Feng, Q., Wu, X., Liu, Q., Xiao, Y., Zhang, X., Chen, Y., … (2025). Interpretable machine learning models for predicting childhood myopia from school-based screening data. Scientific Reports, 15, Article 19811. https://doi.org/10.1038/s41598-025-05021-0

16. Sifat, I. K., Jisa, T. A., Roy, J. S., Sultana, N., Hasan, F., Mosharaf, M. P., & Kibria, M. K. (2025). Prediction of myopia among undergraduate students using ensemble machine learning techniques. Health Science Reports, 8(5). https://doi.org/10.1002/hsr2.70874

17. Kang, M., Hu, Y., Gao, S., Liu, Y., Meng, H., Li, X., … Wang, N. (2024). Deep learning–based longitudinal prediction of childhood myopia progression using fundus image sequences and baseline refraction data [Preprint]. https://arxiv.org/abs/2407.21467

18. Huang, J., Ma, W., Li, R., Zhao, N., & Zhou, T. (2022). Myopia prediction for adolescents via time-aware deep learning [Preprint]. . https://arxiv.org/abs/2209.12546

19. Xie, Y., Poplin, R., Blumer, K., Angermueller, C., Ledsam, J., Chopra, R., … Corrado, G. S. (2017). Deep learning for predicting refractive error from retinal fundus images.. https://arxiv.org/abs/1712.07798