International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0353

e ISSN: 2584-2854
Volume: 02
Issue: 07 July 2024
Page No: 2448-2456

An Investigative Study into Predictive Modeling for Early Detection of Eye Disorders: Challenges, Strategies and Mitigations

Ms. Vaishali Kale[1], Dr. Kavita Surawanshi[2]

[1,2]Research Scholar, D.Y Patil Institute of Management of Computer Applications and Management. Pune, Maharashtra, India.

Emails: *vaishalibarse1234@gmail.com[1], kavita1104@yahoo.com[2]*

Abstract

Eye disorders pose significant health challenges globally, affecting millions of individuals. Early detection is crucial for timely intervention and improved outcomes [1]. Our research investigates predictive modeling techniques to identify eye diseases at an early stage. We explore the impact of unique patient identifiers (Patient_ID), age (P_Age), and gender (P_Gender) on eye health. Understanding these factors is essential for personalized care and risk assessment. The study aims to bridge the gap between clinical practice and data-driven approaches, emphasizing the importance of accurate predictions for disease severity. To address our research objectives, we employ machine learning algorithms and statistical techniques. We leverage diagnostic information from both eyes using Left_Diagnostic_Keywords and Right_Diagnostic_Keywords. These keywords capture specific eye conditions, including glaucoma (G), cataracts (C), age-related macular degeneration (A), hypertensive retinopathy (H), macular edema (M), and optic neuritis (O). Our methodology integrates feature engineering, model selection, and cross-validation. We explore the trade-offs between sensitivity, specificity, and interpretability. Our findings reveal promising predictive capabilities for early detection. We examine machine learning algorithms to classify retinal images for glaucoma, diabetic retinopathy, and normal eyes, stressing the need for early detection. In our study, we compare the effectiveness of various machine learning methods. KNN proves to be the most effective for the given dataset. We focus on feature importance, analyzing attributes such as age, gender, diagnostic keywords, vision clarity, and specific conditions like glaucoma and cataracts to identify key predictors of eye diseases. By embracing data-driven insights,[2] we empower healthcare professionals to enhance patient outcomes and reduce the burden of eye diseases.

***Keywords**: Early detection; KNN; LR; Predictive modelling; RF*

1. Introduction

The role of this research is to reveal the profound implications of unique patient identifiers on eye checking. More specifically, we need to clarify the role of the patient's diagnostic data and the connection between P_Age, P_Gender, and the patient who determines the multinomial of ocular well-being. In practice, we measured the diagnostic data from each eye, which were presented with Left_Diagnostic_Keywords and Right_Diagnostic_Keywords, then our main figure is the predictions of the presence and grip of the disease. Specifically, this involves careful study of various common ophthalmic disorders: Glaucoma (G) is an ocular nerve multi-factorial disorder characterized by intraocular pressure [3]. Cataract (C) or lens is opacity, significantly reduces vision, and is natural in the elderly. Age-Related Macular Degeneration (A) is the central evolution of vision-loss pathology. Hypertensive Retinopathy (H), retinal changes due to opulent life pressure. Macular Edema (M) is the macula's swelling, expressed by blurring the centre vision, and the optic neuritis (O) obtained by young people [3]. We need to admit that early-stage cases are critical subjects at the moment. This is because once we detect them before they start causing a defect, possible involvement would change the path We want our healthcare provider to check their eyes occasionally, as this guarantees the moderation of good watch control of molecular changes from its dashboard [4]. Furthermore, we may need to paddle up every patient identifier concepts, clinical process information, and any eye professional knowledge into our AI technologies that are in progress now.

Dataset: This dataset consists of 6,392 entries, and contains the following critical attributes related to eye diseases: each patient is uniquely identified by a Patient_ID, enabling efficient record-keeping. The patient's age is critical in assessing eye health, considering that some eye conditions are age-dependent. Additionally, the patient's P_Gender is included because some eye diseases manifest gender-specific severity and prevalence. Furthermore, the Left_Diagnostic_Keywords and Right_Diagnostic_Keywords may contain diagnostic information for a patient's left and right eye, respectively. The target variable is meant to refer to disease presence or severity- target. Finally, the following abbreviations present other critical attributes of eye health problems: N indicates Near Vision; D indicates Distance Vision; G represents all forms of Glaucoma; C indicates cataracts; A indicates Age-related Macular Degeneration (AMD); H indicates Hypertensive Retinopathy; M indicates Macular Edema; O denotes Optic Neuritis as the cause of blindness.[5]However, in this context, if the patient has been referred due to the presence of the abovementioned disease, absence is categorized as 'No'. Early diagnosis and proper eye care are vital aspects of regular eye check-ups.We annotated our dataset with meticulous expertise from active professionals to create a high-fidelity proxy that accurately represents the underlying conditions of eyes to minimize the rate of misclassification and bias in model performance. Thus, in this paper, we give the reader an opportunity to familiarize themselves with a dataset that has been artfully positioned to dissect the nature of eye conditions in correlation with patient age, gender, and diagnostic cues [6] with the intent to contribute to improved levels of precision in diagnostic algorithms in the domain of ophthalmology. The image overleaf, is represented by conditions in a large cohort of patients. dissect the nature of eye conditions in correlation with patient age, gender, and diagnostic cues [6] with the intent to contribute to improved levels of precision in diagnostic algorithms in the domain of ophthalmology. Figure 1 Shows the Data Explorations.

	Patient_ID	P_Age	P_Gender	Left_Diagnostic_Keywords	Right_Diagnostic_Keywords	target	N_Near_Vis	D_Dist_Vis	G_glaucoma	C_cataracts	A_age_m_de
0	0	69	Female	cataract	normal fundus	[1, 0, 0, 0, 0, 0, 0, 0]	0	0	0	1	0
1	1	57	Male	normal fundus	normal fundus	[1, 0, 0, 0, 0, 0, 0, 0]	1	0	0	0	0
2	2	42	Male	laser spot，moderate non proliferative retinopathy	moderate non proliferative retinopathy	[0, 1, 0, 0, 0, 0, 0, 0]	0	1	0	0	0
3	4	53	Male	macular epiretinal membrane	mild nonproliferative retinopathy	[0, 1, 0, 0, 0, 0, 0, 0]	0	1	0	0	0
4	5	50	Female	moderate non proliferative retinopathy	moderate non proliferative retinopathy	[0, 1, 0, 0, 0, 0, 0, 0]	0	1	0	0	0

Figure 1 Data Explorations - Getting Information from Data

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0353
e ISSN: 2584-2854
Volume: 02
Issue: 07 July 2024
Page No: 2448-2456

2. Data Analysis and Utilization

Demographic analysis allows for the integration of age and gender with selected ocular conditions, which can help to determine patterns in ocular health between such categories. Diagnostic keywords are entry points associated with common conditions of the eye and their manifestations as bilateral. [7] Additionally, condition vectors and labels facilitate straightforward classification tasks and enable the training of machine learning models to predict ocular diseases based on demographic and clinical input features[8].Here, the sample entries include: Patient 0, 69-year-old, female, 1 cataract, left eye, fundus normal, N. this corresponds to the presence of cataracts, as indicated, with the categorical label ['N'] being used to imply the non-identification of other diseases. Patient 2 is a male, 42 years old, 2 laser spots, 2 non-proliferative-M, both eyes, fundus distant-both directions, distant N. In this case, the patient displays normal distance vision but the classification label used to code other absent diagnoses, such as macular edema or optic neuritis, is ['D'], indicating disease Thus, the structured nature of the dataset provides an opportunity for a wide and diverse study of ocular pathologies for creating predictive models and increasing the level of detail in ophthalmologic epidemiology. The record of diagnostic tags in combination with binarized parameters and pre-built labels makes it possible to use the database as an educational sample toolkit for training diagnostic bots and understanding the relationship between various ocular pathologies and demographics. Furthermore, alongside medicine, certain robot-assisted surgeries have been conducted successfully. It makes a doctor's work more precise and efficient. Nowadays, AI-assisted medical screening and image-based diagnoses are emerging [9-10] Figure 2 below provides an in-depth description of the dataset used to explore and study ocular conditions in a large cohort of patients. Each record in this dataset, representing an individual patient, consists of various characteristics and diagnosis outcomes. This dataset is structured in a Data Frame format using the pandas library of Python, ranging from index 0 to 6391. As a result, each index of this dataset represents discrete patient records, as represented by row numbers. It contains 15 columns, consisting of both numeric and object data types. Among the columns of this dataset, Patient_ID is the unique identifier number for individual patients and therefore acts as the primary key of the dataset. P_Age, representing the age of the patient in years, is an essential column for age-based ocular condition studies. All the columns have 6,392 non-null entries, indicating our dataset is complete and there are no missing values. As a result, the dataset is complete and offers a robust basis for statistical analysis and model design.

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 6392 entries, 0 to 6391
Data columns (total 15 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Patient_ID                6392 non-null   int64
 1   P_Age                     6392 non-null   int64
 2   P_Gender                  6392 non-null   object
 3   Left_Diagnostic_Keywords  6392 non-null   object
 4   Right_Diagnostic_Keywords 6392 non-null   object
 5   target                    6392 non-null   object
 6   N_Near_Vis                6392 non-null   int64
 7   D_Dist_Vis                6392 non-null   int64
 8   G_glaucoma                6392 non-null   int64
 9   C_cataracts               6392 non-null   int64
 10  A_age_m_de                6392 non-null   int64
 11  H_hy_retinopathy          6392 non-null   int64
 12  M_mac_edema               6392 non-null   int64
 13  O_opt_neuritis            6392 non-null   int64
 14  labels                    6392 non-null   object
dtypes: int64(10), object(5)
memory usage: 799.0+ KB
```

Figure 2 Description of The Dataset Used

In the present study, we conducted a robust statistical analysis using a dataset comprising over 6000 patients on numerous ocular health indicators and conditions. This kind of dataset is vital in realizing the extent and distribution of ocular conditions among various age group and demographic characteristics.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0353
e ISSN: 2584-2854
Volume: 02
Issue: 07 July 2024
Page No: 2448-2456

Table 1 Descriptive Statistics of Patient Ocular Conditions

Variable	Count	Mean	Std. Dev.	Min	25th Percentile	Median	75th Percentile	Max
Patient_ID	6392	2271.151	1417.559	0	920.750	2419.5	3294.000	4784
P_Age	6392	57.858	11.728	1	51.000	59.000	66.000	91
N_Near_Vis	6392	0.329	0.470	0	0.000	0.000	1.000	1
D_Dist_Vis	6392	0.332	0.471	0	0.000	0.000	1.000	1
G_glaucoma	6392	0.062	0.241	0	0.000	0.000	0.000	1
C_cataracts	6392	0.063	0.243	0	0.000	0.000	0.000	1
A_age_m_de	6392	0.050	0.218	0	0.000	0.000	0.000	1
H_hy_retinopathy	6392	0.032	0.175	0	0.000	0.000	0.000	1
M_mac_edema	6392	0.048	0.214	0	0.000	0.000	0.000	1
O_opt_neuritis	6392	0.248	0.432	0	0.000	0.000	0.000	1

Table 1 contains items concerning patient demographics. With a mean age of about 57.9 years and a standard deviation of 11.73 years, there is a moderate distribution of ages in the study population. The minimum and maximum ages are 1 and 91 years, respectively, showing that the data recorded was well broad in age group and stage of life. With regard to the ocular conditions, the following are observed: i. Near and Distant Vision: About 32.9% and 33.2% of the patients have normal near and distant vision, respectively. This indicates that a considerable proportion of the population can experience unclear vision in multiple lenses. ii. Glaucoma and Cataracts: Both Glaucoma and Cataracts affect about 6.2% and 6.3% of the patients, respectively. These two conditions are essential and threaten potential visual loss hence requiring optimal diagnosis and management. iii. Age-related Macular Degeneration: Noteworthy, 5.0% of the patients are affected by AMD, a severe leading condition of vision loss among older adult. iv. Hypertensive Retinopathy and Macular Edema: These conditions are relatively rare at 3.2% and 4.8% of the cohort. v. Optic Neuritis: Surprisingly, 24.8% of the patients exhibit optic neuritis, suggesting it is a prevalent inflammatory condition of the optic nerve in the dataset. Consequently, the above dataset serves as a valuable source of insights for analysing the prevalence and distribution of the most common ocular conditions among a wide range of patients. [11] Most importantly, the data reveals a critical requirement to conduct regular screenings of eye health, as the majority of patients are diagnosed specifically with glaucoma, cataracts, and age-related macular degeneration, which may not cause early symptoms; yet, they severely affect the quality of life. Finally, the discussed representation allows conducting additional research focused on identifying the risk factor for the development of each condition and developing effective measures to decrease the burden.

Table 2 Attributes with Correlation Coefficient

Correlation Coefficient	1
Mean Absolute Error (MAE)	0.4585
Root Mean Squared Error (RMSE)	1.5597
Relative Absolute Error (RAE)	0.1289
Root Relative Squared Error (RRSE)	0.14

Most of the model research predictions were accurate. Model performance is measured by Mean Absolute Error and Root Mean Squared Error. RMSE is a standard prediction deviation. In our case, the RMSE is 1.56, which is 10% more than the MAE of 0.46 It implies that the assumption is linear in this

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0353
e ISSN: 2584-2854
Volume: 02
Issue: 07 July 2024
Page No: 2448-2456

case and that it was found. The model shows that the Type of the disease has a negative correlation with macular edema M, and a positive one with glaucoma G, and cataracts C [12]. Second, predict and plot the estimated value from the test result. Take note of the difference between the actual and expected wise. Then our model is first class since the actual and predicted values are almost similar. We have our model. As shown, the Root Mean Squared Error is 1.5597, which is 10% more than the MAE of 0.4585. Our model made accurate predictions. This study used multivariate linear regression via the Python Scikit- Learn machine learning library. K-Nearest Neighbors algorithm is a non-parametric supervised learning method and a classification and regression task. [13] The fundamental premise of the algorithm is that similar data points tend to cluster together. KNN is a versatile algorithm that can handle both classification and regression tasks. For classification, the algorithm assigns a data point a class label based on a majority vote from the data points in the vicinity. For regression, the algorithm predicts the continuous value of a data point by equaling the averages of the k-nearest points Algorithm Workflow: [90-thesis] Given a new, unlabeled data point

- Calculate its distance from each training example using the Euclidean distance.
- Identify the K-Nearest Neighbors (closest data points).
- Determine the majority class (classification) or average value (regression) among these neighbors.

The ‘k’ in KNN represents the number of neighbors considered. If k = 1, the label of the nearest neighbor is assigned. For k > 1, the majority vote among the k neighbors determines the label. Distance measures (e.g., Euclidean distance) play a crucial role in supervised learning. They reveal patterns and similarities in input data. These metrics aid in understanding relationships between results. In summary, KNN’s accuracy make it a foundational algorithm in machine learning, although its efficiency decreases with larger datasets [X] Acito, F. (2023). k Nearest Neighbors. In: Predictive Analytics with KNIME. Springer, Cham. https://doi.org/10.1007/978-3-031-45630-5_10

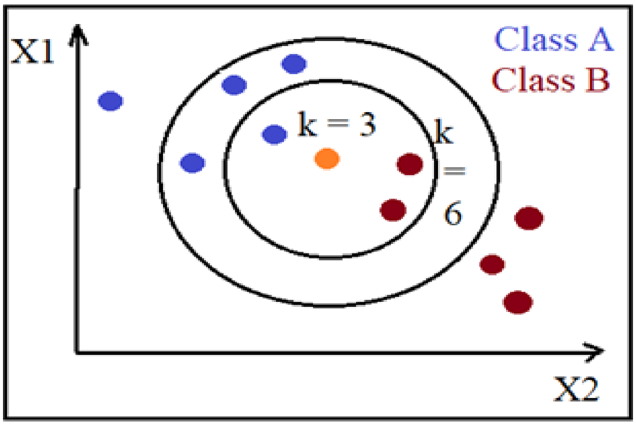

Figure 3 KNN Classifier [82-th]

In the above Figure 3, blue and maroon points are Class A and Class B in training data. The saffron star is testing data to be classified. When k = 3, it predicts the output as Class B, and when k = 6, it predicts the output as Class A.

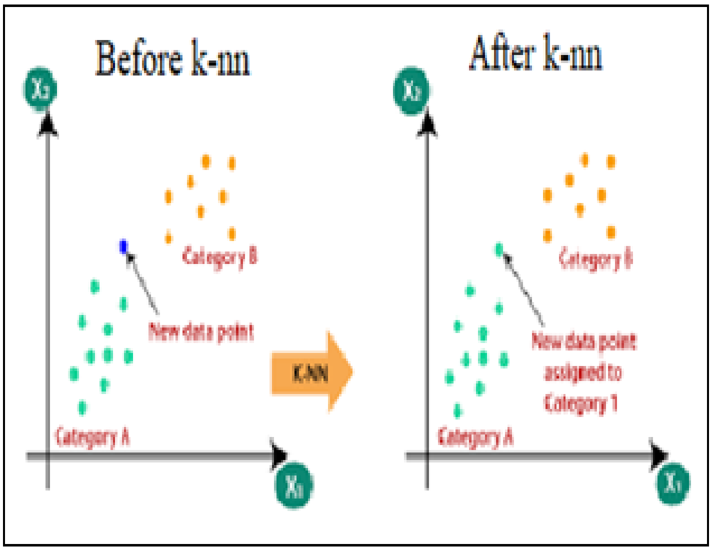

Figure 4 K-Nearest Neighbor

For KNN classification Figure 4, a majority vote is taken over all of its k-nearest data points. For KNN regression, the output is the mean of k nearest data points. As a good practice, we recommend choosing odd numbers to ask [82-thesis]. The output is determined by a majority vote of the k-nearest points in classification and their average in regression. [14] This model resembles a lazy learner since it only calculates at runtime and tests. During testing, the

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0353
e ISSN: 2584-2854
Volume: 02
Issue: 07 July 2024
Page No: 2448-2456

$z = \theta_0 + \theta_0 x_1 + \theta_0 x_2$
$h\theta = g(z)$
$g(z) = 1 / 1 + e\text{-}z \quad (1)$

The h(θ) value here corresponds to P(y=1|x), i.e., probability of output to be binary 1, given input 'x'. Whenever 'z' is closest k-neighbors decide. The key hyper parameters are 'k' and the distance, which is Euclidean by default. The proper choice of k is crucial because a relatively small value consumes many resources to handle large datasets while too high a value implies too much model simplicity Logistic Regression: Sigmoid function is the frequently used logistic function. The 'z' value is same as that of the linear regression output in Equation (1).

$$z = \theta_0 + \theta_0 x_1 + \theta_0 x_2$$
$$h\theta = g(z)$$
$$g(z) = 1 / 1 + e\text{-}z \quad (1)$$

The h(θ) value here corresponds to P(y=1|x), i.e., probability of output to be binary 1, given input 'x'. Whenever 'z' is positive, h(θ) will be greater than 0.5 and output will be binary 1

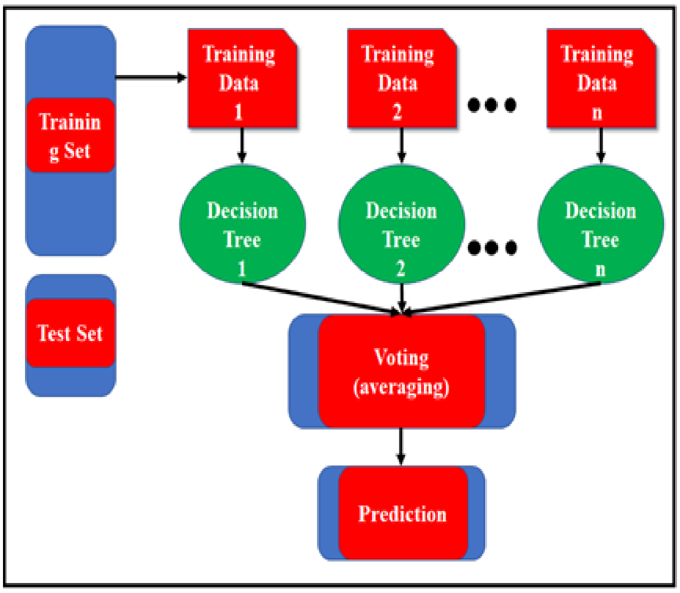

Figure 5 Working of RF Algorithm

By following above Figure 5 and considering our research problem, following are the steps in RF algorithm

- Initially 1000 random events are taken from the dataset with 2 centers, 1000 features.
- For n-estimators, consider good values might be a log scale from 10 to 1,000. Thus, decision trees are constructed individually for each sample.

3. Significance for Research

The availability of detailed diagnostic keywords for both eyes, combined with demographic information and binary indicators of various ocular conditions, provides a rich dataset for conducting multifactorial analysis, like- Proposed is a novel model based on the concept that employs the instrument 'Eli5' to enhance the analysis of eye diseases. For debugging Machine with 7 number of folds and 1 to half the number of input features as maximum features, define grid search. Each decision tree will generate an output. Summarize final output with means i.e. based on averaging for classification respectively Learning classifiers and describing predictions, the Eli5 Python package is utilized. In this instance, the binary classification model produced by the Eli-5 utility was 81% accurate. The results of the Binary Classifier, Confusion Matrix, and Correlation Matrix are found to be highly correlated and positive, with a probability score exceeding 5, as shown in Table 3.

Table 3 Classification Report (CR)

	Precision	Recall	F1-Score	Support
0	0.92	0.90	0.86	413
1	0.96	0.55	0.61	105/(508)

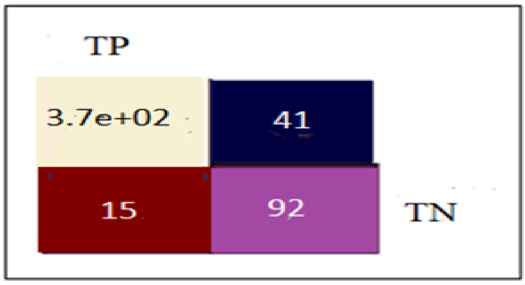

Figure 6 Confusion Matrix for Binary Classifier

After model creation with Eli5, the Classification Report (CR) shows good precision (P), recall (R), and F1-score for the 0th class: 0.92, 0.90, and 0.86, respectively. The Confusion Matrix (CM), as depicted in Figure 6, indicates 89% accuracy (462

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0353
e ISSN: 2584-2854
Volume: 02
Issue: 07 July 2024
Page No: 2448-2456

correct out of 518) with a 14% False Positive Rate (15/107). The matrix is constructed by testing the validation dataset, making predictions, and comparing them to expected outcomes. This tool helps balance precision and recall, especially useful when data is imbalanced, giving a 76% average PRF for both classes. A study on eye diseases utilized machine learning to identify the most effective models: K-Nearest Neighbor (KNN), Logistic Regression (LR), and Random Forest (RF). Evaluation metrics included the classification report confusion matrix, and AUROC. Results showed KNN outperforming LR and RF, suggesting its potential to bridge knowledge gaps in this field. This comprehensive analysis demonstrates KNN's superiority through popular classification metrics, highlighting its effectiveness in disease diagnosis. The performance of three classification models—K-Nearest Neighbor, Logistic Regression, and Random Forest—was assessed using the 'Feature Importance' metric, which aids in understanding the eye disease landscape and may inspire model enhancements. Glaucoma, age-related macular degeneration, and hypertensive retinopathy were identified as the most significant features, crucial for early disease detection. Additionally, the dataset facilitates the exploration of correlations between ocular diseases and demographic factors, enhancing predictive modeling and subgroup analysis. [15]

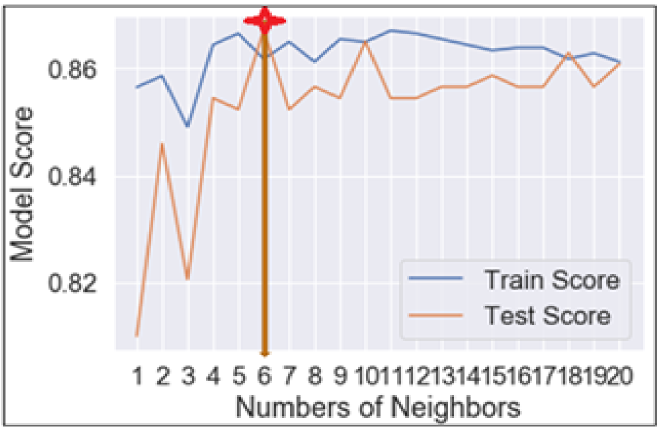

Figure 7 **KNN Model Score**

To determine the best 'k' value for KNN, a range of values from 1 to 21, mostly odd numbers, is considered to identify the optimal hyper parameters The process involves training the model with different 'k' values and assessing accuracy. In Figure 7, it can be observed that accuracy levels off at around 0.85 after k=6, indicating that k=6 is a suitable choice. Important hyper parameters to consider are 'n_neighbors', 'metric': 'manhattan', and 'weights': 'uniform'. This setup results in consistent accuracy levels, as shown in the classification report based on this configuration. Therefore, for k-neighbors = 6, the classification report from this research study is as follows- For k-neighbors = 6 – Classification Report is shown in Table 4

Table 4 **Classification Report (CR)**

	Precision	Recall	F1-Score	Support
0	0.93	0.97	0.95	223
1	0.96	0.92	0.94	245
Avg/ Total	0.95	0.95	0.95	468

with k = 6, Here, we got 95% of Precision, recall and f1-score. For predictions, we'll select 'k' as 6, using Euclidean distance between existing and new points. This final model will make predictions.

4. Results and Discussion

This study paper provides a thorough examination of many conventional and innovative methods used to tackle the problem of eye disorders. The study encompasses Apply inferential analysis to investigate and assess links Machine learning models, such as regression leveraging Root Mean Square Error (RMSE)Conducting exploratory data analytics (EDA) to ascertain the most suitable number of classifiers.

5. Potential Application

The dataset plays a vital role in comprehending the prevalence of specific eye conditions in the population and can aid in evaluating risk factors, treatment results, and the efficiency of diagnostic procedures. It also forms the basis for epidemiological research that could guide public health strategies and resource distribution for eye health services. In summary, this dataset is essential for our continuous research endeavours to enhance

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0353
e ISSN: 2584-2854
Volume: 02
Issue: 07 July 2024
Page No: 2448-2456

our knowledge of eye health and disease patterns in various populations. It will enable a thorough analysis to help in creating focused interventions to lessen the impact of eye diseases on a global scale. [16]

Conclusion

Our research contributes to advancing the field of eye disease prediction through a focus on precise dataset labelling, robust machine learning models, and insights driven by features. We thoroughly analyse a variety of machine learning algorithms to categorize retinal fundus images linked to glaucoma, diabetic retinopathy, and healthy eyes. [17] These eye conditions have a significant impact on patients' quality of life, underscoring the importance of early detection and effective treatment. We compare three classification models: K-Nearest Neighbour (KNN), Logistic Regression (LR), and Random Forest (RF). Through extensive testing, we find that KNN outperforms both LR and RF [18] Our evaluation is based on the importance of features, highlighting the significance of specific features in predicting eye diseases. Future efforts should explore the avenues mentioned above to improve disease prediction and clinical decision-making support.

References

[1]. R. S. M. Lakshmi Patibandla, B. Tarakeswara Rao, M. Ramakrishna Murty, "Revolutionizing Diabetic Retinopathy Diagnostics and Therapy Through Artificial Intelligence: A Smart Vision Initiative", Source Title: Transformative Approaches to Patient Literacy and Healthcare Innovation, IGI Global Publishing Tomorrow's Research Today

[2]. Ji, P. X., Ramalingam, V., Balas, M., Pickel, L., & Mathew, D. J. (2024). Artificial Intelligence in Glaucoma: A New Landscape of Diagnosis and Management. Journal of Clinical & Translational Ophthalmology, 2(2), 47-63.

[3]. Schlosser, T., Beuth, F., Meyer, T., Kumar, A. S., Stolze, G., Furashova, O., ... & Kowerko, D. (2024). Visual acuity prediction on real-life patient data using a machine learning based multistage system. Scientific Reports, 14(1), 5532.

[4]. Rajan, S., Sathiyanarayanan, M., Prashant, S., Prashant, S. B., & Nataraj, P. L. (2018, January). Prevention of avoidable blindness and improving eye healthcare system in india. In 2018 10th International Conference on Communication Systems & Networks (COMSNETS) (pp. 665-670). IEEE.

[5]. Karadeniz Ugurlu, S., Kocakaya Altundal, A. E., & Altin Ekin, M. (2017). Comparison of vision-related quality of life in primary open-angle glaucoma and dry-type age-related macular degeneration. Eye, 31(3), 395-405.

[6]. Giger, M. L. (2018). Machine learning in medical imaging. Journal of the American College of Radiology, 15(3), 512-520.

[7]. Ting, D. S., Peng, L., Varadarajan, A. V., Keane, P. A., Burlina, P. M., Chiang, M. F., ... & Wong, T. Y. (2019). Deep learning in ophthalmology: the technical and clinical considerations. Progress in retinal and eye research, 72, 100759.

[8]. Abbas, Q., Qureshi, I., Yan, J., & Shaheed, K. (2022). Machine learning methods for diagnosis of eye-related diseases: a systematic review study based on ophthalmic imaging modalities. Archives of Computational Methods in Engineering, 29(6), 3861-3918.

[9]. Du, X. L., Li, W. B., & Hu, B. J. (2018). Application of artificial intelligence in ophthalmology. International journal of ophthalmology, 11(9), 1555.

[10]. Salem, H., Negm, K. R., Shams, M. Y., & Elzeki, O. M. (2021). Recognition of ocular disease based optimized VGG-net models. In Medical Informatics and Bioimaging Using Artificial Intelligence: Challenges, Issues, Innovations and Recent Developments (pp. 93-111). Cham: Springer International Publishing. in Computational Intelligence, vol 1005. Springer

[11]. Parekh, P. H., Boente, C. S., Boente, R. D., Meeker, J. W., & Carlos, W. G. (2019). American Ophthalmology in critical care.

Annals of the Thoracic Society, 16(8), 957-966.

[12]. Kumar, Y., & Gupta, S. (2023). Deep transfer learning approaches to predict glaucoma, cataract, choroidal neovascularization, diabetic macular edema, drusen and healthy eyes: an experimental review. Archives of Computational Methods in Engineering, 30(1), 521-541.

[13]. Hackeling, G. (2017). Mastering Machine Learning with scikit-learn. Packt Publishing Ltd.

[14]. Wong, D. W. K., Zhang, Z., Srivastava, R., Liu, H., Chen, X., Duan, L., ... & Liu, J. (2014). A survey on computer aided diagnosis for ocular diseases.

[15]. Zhao, Y., Li, X., Li, S., Dong, M., Yu, H., Zhang, M., ... & Gao, Z. (2022). Using machine learning techniques to develop risk prediction models for the risk of incident diabetic retinopathy among patients with type 2 diabetes mellitus: a cohort study. Frontiers in endocrinology, 13, 876559.

[16]. Ting ShuWei [Ting, S., Cheung ChuiMing [Cheung, C., & Wong TienYin, W. T. (2016). Diabetic retinopathy: global prevalence, major risk factors, screening practices and public health challenges: a review.

[17]. Chalakkal, R. J., Abdulla, W. H., & Hong, S. C. (2020). Fundus retinal image analyses for screening and diagnosing diabetic retinopathy, macular edema, and glaucoma disorders. In Diabetes and fundus OCT (pp. 59-111). Elsevier.

[18]. Hassan, C. A. U., Khan, M. S., & Shah, M. A. (2018, September). Comparison of machine learning algorithms in data classification. In 2018 24th International Conference on Automation and Computing (ICAC) (pp. 1-6). IEEE.