Diagnosis of Keratoconus Using Machine Learning

[1]Dr. Bharati Bidikar, [2]Karanam Sagar Kiran

[1]Adjunct Professor, [2]Student
Department of CS&SE, AU College of Engineering
Department of Information Technology and Computer Applications
Andhra University, Visakhapatnam, India

Abstract: *This paper focuses on using machine learning to detect and classify keratoconus, an eye disease, using corneal tomography data. The data is first cleaned by removing missing values, duplicates, and unnecessary columns. Then the data is explored using graphs and statistics to understand patterns and outliers. To reduce the number of features while keeping important information, Principal Component Analysis (PCA) is applied. It is also identified which features are most important for the prediction Several machine learning models such as Random Forest, Logistic Regression, SVM, Decision Tree, and KNN are trained and tested both with and without PCA. Their performance is compared using accuracy, precision, recall, and F1-score. The results show that machine learning models, combined with good data preparation, can accurately detect different stages of keratoconus and could help in medical diagnosis. For corneal tomography data Random Forest algorithm gave the best result*

Index Terms: Keratoconus, Corneal Tomography, Machine Learning, Principal Component Analysis (PCA), Feature Selection, Random Forest, Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbors (KNN), Decision Tree, Data Preprocessing, Medical Diagnosis.

Introduction: Keratoconus is a progressive eye disease that affects the shape and structure of the cornea, leading to visual impairment. Early and accurate diagnosis is essential to manage the disease effectively and prevent further vision loss. Traditional diagnostic methods rely heavily on clinical examination and corneal topography, which can sometimes miss subtle changes in the early stages of the condition.

With the advancement of medical imaging technologies, corneal tomography has become a valuable tool in detecting structural abnormalities of the cornea. However, the large volume and complexity of data generated from these scans make manual analysis time-consuming and prone to error.

Machine learning offers a powerful approach to analyze medical imaging data efficiently. By applying statistical and computational techniques, machine learning models can uncover hidden patterns in the data that may not be apparent through conventional methods. In particular, dimensionality reduction methods such as Principal Component Analysis (PCA) help in selecting the most relevant features, improving both the accuracy and speed of the classification process.

We explore a machine learning-based framework for detecting Keratoconus using features extracted from corneal tomography images. The process involves data preprocessing, feature selection using PCA, and classification using various algorithms such as Random Forest, Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbors (KNN), and Decision Tree. The goal is to develop a reliable and efficient diagnostic model that can assist ophthalmologists in early and accurate detection of Keratoconus. To achieve this goal, machine learning model that can detect early signs of keratoconus is created. Different machine learning methods are tested on collected data.

II. LITERATURE REVIEW

Hallett et al. introduced a deep learning model that used both unsupervised and semi-supervised learning techniques to detect keratoconus in its early stages. The goal was to give clinicians enough time to choose suitable treatment strategies. Their model achieved an accuracy of 80.3% when tested on 124 eyes affected by keratoconus. However, the limited dataset may impact the reliability and applicability of the results to larger populations.

In early-stage keratoconus, researchers used a logistic regression model to make their identification. However, the study depended on a single diagnostic measure, auto keratometer readings, which limited the depth of the analysis. Similarly, other researchers applied a classification method based on corneal shape data gathered from Optical Coherence Tomography (OCT) instruments. They reported an accuracy of 92% on a sample of 244 eyes. However, the study did not specify which stages of keratoconus were included in the data, especially whether early cases were analyzed. Moreover, the relatively small sample size may limit the generalizability of their conclusions. Artificial intelligence has also been used successfully to guide the placement of intracorneal ring segments (ICRS) as part of keratoconus treatment. This further shows the flexibility and usefulness of machine learning in various aspects of managing keratoconus, beyond just diagnosis.

III. METHODOLOGY

(A) **Machine Learning Overview:** There are different machine learning approaches to detect keratoconus. Supervised learning is applied to labeled data to train models for accurate predictions. Unsupervised learning helps find patterns in data without labels. Semi-supervised learning combines a small amount of labeled data with lots of unlabeled data to improve performance. Using all three methods allows the system to work well with both clinical records and image data. This ensures a more complete and accurate analysis.

(B) **Machine Learning Process :** This is a clear step-by-step machine learning pipeline. It starts by defining the goal of detecting keratoconus early. Data is collected from corneal imaging tools like OCT, Pentacam, and CASIA. This data is cleaned and prepared through preprocessing. Then, Exploratory Data Analysis (EDA) is used to find patterns. Predictive models are built, tested with techniques like cross-validation, and improved for better accuracy. The final model predicts keratoconus in new, unseen cases.

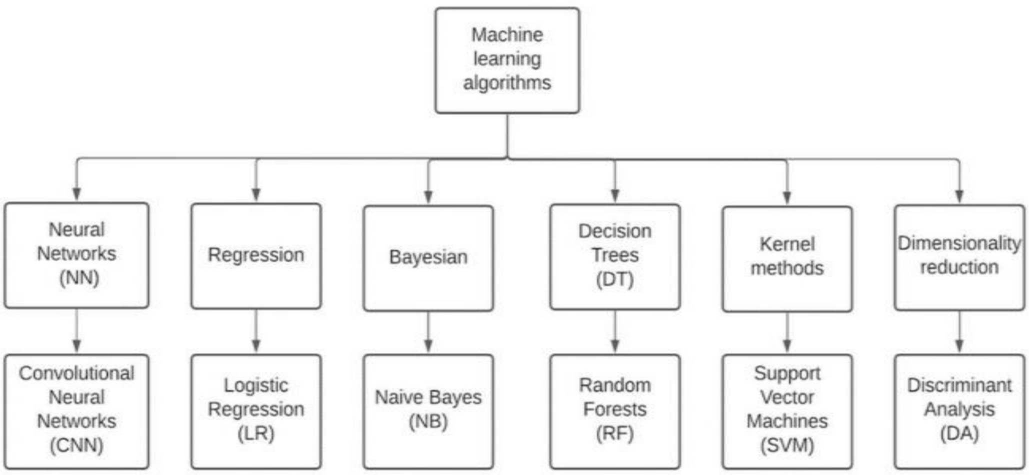

Fig1 : Overview of Machine Learning Algorithms Used

(C) **Feature Selection :** With 423 features in the dataset, selecting the most important ones is essential. The study uses univariate selection to find features that strongly relate to the target outcome. This helps reduce model complexity, speeds up training, and improves accuracy. Important features like ESI.Posterior and K.Max..8mm..1 are chosen for their strong ability to detect subclinical keratoconus. These selected features play a key role in making accurate prediction. When dealing with high-dimensional data, it becomes essential to identify features that significantly influence the outcome variable while eliminating those that add noise or redundancy. Feature selection helps improve model performance, reduces training time, and minimizes the risk of overfitting.

(E) Existing Methodology: The existing method by Yousefi et al. used unsupervised learning to classify keratoconus severity. They first applied PCA to reduce 420 features to 8 key components. Then, manifold learning further reduced these to 2 main variables. Finally, density-based clustering grouped the eyes into four severity levels. The method was validated using the Ectasia Status Index (ESI), showing its effectiveness in classifying keratoconus stages**.**

(F)Proposed Methodology : The new method improves past studies by adding doctor-diagnosed labels for better accuracy. It also uses more data from tools like the Pentacam and checks new keratoconus signs like the BA index. From 423 features, only the most important ones are chosen using univariate selection. This makes the model faster, more accurate, and easier to use. The goal is to make it work better in real hospitals and clinics.

IV RESULTS AND DISCUSSION

The research evaluated the effectiveness of various machine learning algorithms in diagnosing keratoconus using corneal imaging data. The models were assessed based on key performance metrics accuracy, precision, and recall.Among the tested models, the **Random Forest classifier** emerged as the most accurate, achieving an accuracy of **95.8%**, a recall of **91.7%**, and a precision of **91.4%**. This indicates the model's strong ability to correctly identify keratoconus cases while minimizing false positives.

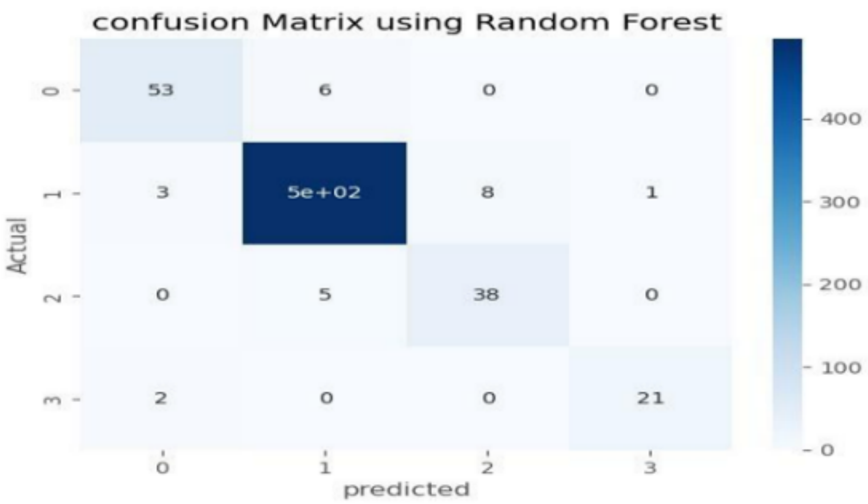

Fig 2: Confusion matrix using Random Forrest

The **Optimized k-Nearest Neighbors (kNN)** algorithm also performed well, with a **95.2%** accuracy and high precision, although its recall was slightly lower than that of Random Forest. Similarly, **Support Vector**

Machines (SVMs)—both linear and kernel variants—achieved competitive results accuracy 94.6%with good recall and precision balances, demonstrating their suitability for high-dimensional biomedical datasets.

The **Decision Tree** and **Logistic Regression** models, while slightly less accurate (93.3% and 94.6% respectively), still showed reliable classification capabilities. Each model's performance confirms that machine learning can serve as a valuable tool for early keratoconus detection using structured corneal feature data.

Table 1: Performance metrics comparison different algorithms

Algorithm	Accuracy	Recall	Precision
Optimized kNN	0.952	0.852	0.922
Decision Tree	0.933	0.861	0.847
Random Forest	0.958	0.917	0.914
Linear SVM	0.946	0.913	0.941
Kernel SVM	0.946	0.837	0.913
Logistic Regression	0.946	0.859	0.923

The table1 compares different machine learning algorithms based on their accuracy, recall, and precision. Random Forest shows the highest accuracy at 0.958 and also performs well in recall and precision. Optimized kNN has an accuracy of 0.952, with good precision (0.922) but lower recall (0.852). Linear SVM and Logistic Regression both show the same accuracy (0.946) but Linear SVM has the highest precision (0.941). Kernel SVM has the lowest recall (0.837) among all models despite decent accuracy. Decision Tree has the lowest accuracy (0.933) and also the lowest precision (0.847).

CONCLUSION

The project demonstrates that machine learning techniques, especially supervised and unsupervised algorithms, can significantly aid in the early detection of subclinical keratoconus. By using advanced data-driven approaches like feature selection, dimensionality reduction (PCA), and models such as SVM, Random Forest, and Neural Networks, the system can accurately distinguish between normal and keratoconic eyes. The use of a large dataset with over 400 corneal parameters and application of univariate selection has enhanced model precision while reducing noise and overfitting. The study also highlights the potential of unsupervised clustering to identify disease stages without relying on clinical labels. Although current results are promising, further validation with external datasets and clinical diagnostic labels is recommended to improve generalizability and real-world applicability. Overall, this methodology offers a strong foundation for building intelligent diagnostic tools that support early intervention and better management of keratoconus.

FUTURE WORK

Follow up studies can explore the efficiency of deep learning models using corneal images. The other aspect that can be examined in the future is KCN progression detection, which is critical in clinical practice to adjust treatment plan based on progression of the disease.

REFERENCES

[1] N. Hallett, K. Yi, J. Dick, C. Hodge, G. Sutton, Y. G. Wang, and J. You, "Deep learning based unsupervised and semi-supervised classification for keratoconus," Jan. 2020, arXiv: 2001, vol. 105, pp. 11653. [Online] [Available: http://arxiv.org/abs/2001.11653]

[2] T. Kojima, T. Nishida, T. Nakamura, A. Tamaoki, A. Hasegawa, Y. Takagi, H. Sato, andK. Ichikawa, ‘‘Keratoconus screening using values derived from auto-keratometer measurements: A multicenter study,’’ Amer. J. Ophthalmol., Jul 2020, vol. 215, pp. 127–134, [doi: 10.1016/j.ajo.2020.02.017]

[3] M. D. Twa, S. Parthasarathy, C. Roberts, A. M. Mahmoud, T. W. Raasch, and M. A. Bullimore, ‘‘Automated decision tree classification of corneal shape,’’ Optometry Vis. Sci., Off. Publication Amer. Acad. Optometry, Dec 2005, vol. 82, no. 12, pp. 1038– 1046.

[4] C. Fariselli, A. Vega-Estrada, F. Arnalich-Montiel, and J. L. Alio, ‘‘Artificial neural network to guide intracorneal ring segments implantation for keratoconus treatment: A pilot study,’’ Eye Vis., Dec 2020, vol. 7, no. 1, pp. 1–12.

[5] S. R. Lin, J. G. Ladas, G. G. Bahadur, S. Al-Hashim, and R. Pineda, ‘‘A review of machine learning techniques for keratoconus detection and refractive surgery screening,’’ Seminars Ophthalmol., 2019, vol. 34, no. 4, pp. 317–326, [doi: 10.1080/08820538.2019.1620812]

[6] G. W. Armstrong and A. C. Lorch, ‘‘A(eye): A review of current applications of artificial intelligence and machine learning in ophthalmology,’’ Int. Ophthalmol. Clinics, 2020, vol. 60, no. 1, pp. 57–71.