International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 29 Issue 4 – MAY 2022.

EYE DISEASE PREDICTION USING MACHINE LEARNING TECHNIQUES

Aruna Ramanan P,Gowtham Rajan K,Balaji S,Nilesh P [1], Dr.K.Dhanalakshmi [2]
UG scholar, Department of CSE, PSNA College of Engineering and Technology, Dindigul [1]
Professor,Department of CSE, PSNA College of Engineering and Technology, Dindigul[2]

.

Abstract:Medical health systems have been concentrating on artificial intelligence techniques for speedy diagnosis of eye disease. However, the recording of health data in a standard form still requires attention so that machine learning can be more accurate and reliable by considering multiple features. The aim of this study is to develop a general framework for recording diagnostic data in an international standard format to facilitate prediction of eye disease diagnosis such as Cataract, Glaucoma and retinal, diagnosis based on symptoms using decision tree algorithms. Efforts were made to ensure error-free data entry by developing a user-friendly interface. This data was formatted according to structured hierarchies designed by medical experts, whereas diagnosis was made as per the ICD-10 coding developed by the American Academy of Ophthalmology. Furthermore, the system is designed to evolve through self-learning by adding new classifications for both diagnosis and symptoms. The classification results from tree-based methods demonstrated that the proposed framework performs satisfactorily, given a sufficient amount of data. Owing to a structured data arrangement, the decision tree algorithms prediction rate is more than 90% as compared to more complex methods such as neural networks and the naïve Bayes algorithm.

Keywords:Diagnosis, ICD-10, Cataract, Glaucoma, Retinal, Naïve Baye, Opthamologys.

I. INTRODUCTION

Classifying data is a common task in Machine learning. Data mining in health care is an emerging field of high importance for providing prognosis of and a deeper understanding of medical data. Most data mining methods depend on a set of features that define the behaviour of the learning algorithm and directly or indirectly influence the complexity of resulting models. Eye disease is the most common disease, is caused by complications that occurs when blood vessels in the retina weakens or distracted. We have applied machine learning methods to predict the early detection of eye disease diabetic retinopathy and found that Decision Tree method to be 90% accurate. The performance was also measured by sensitivity and specificity. The dominant causes of visual impairment worldwide are Cataract, Glaucoma, and retinal diseases among patients. The alarming cases of these diseases call for an urgent intervention by early diagnosis.

The planned system is designed and developed to easily facilitate the detection of cataract, glaucoma and retinal diseases among patients. The Logistic Regression, Random Forest, Gradient Boosting and Support Vector Machine algorithms [1] are used for detection. The proposed system will help people to get the proper treatment of the aforementioned diseases at an early stage thus reducing the percentage of blindness being caused. The proposed system evaluates the effectiveness and safety of cataract surgery in eyes with agerelated degeneration along with glaucoma and retinal diseases detection. This paper shows the accuracy of algorithms and SVM classifiers based upon the glaucoma, retina, cataract and normal eye's fundus images.

Data mining is the process of discovering actionable information from large sets of data. Data mining uses mathematical analysis to derive patterns and trends that exist in data. Typically, these patterns cannot be discovered by traditional data exploration because the relationships are too

complex or because there is too much data. It is used to classify and predict eye disease like Cataract, Glaucoma, and retinal diseases. It is also used to implement decision tree based classifier for better performances.

II. LITERATURE REVIEW

An, G.; Omodaka, K.; Tsuda, S.; Shiga, Y.; Takada, N.; Kikawa, T.; Nakazawa, T.; Yokota, H.; Akiba, M. Comparison of machine-learning classification models for glaucoma management. J. Healthc. 2021. [CrossRef] [PubMed].

In this paper, to diagnose diabetic retinopathy, three models like Neural network (NN), Decision Tree (DT) and Support vector machine (SVM) [1] are described and their performances are compared. NN has an accuracy of 88.19% DT has an accuracy of 91.66% and SVM has an accuracy of86.19%.

Saito, K.; Nakano, R. Medical diagnostic expert system based on PDP model. In Proceedings of the IEEE International Conference on Neural Networks, San Diego, CA, USA, 24–27 July 2020 pp. 255–262.

Use of moment invariant features of pixels as part of the feature set for training data is employed. the 20 D feature vector comprises of the maximum Gabor transform responses over different angles, Hu moment invariants, Hessian multiscale filter response, Local Binary Pattern (LBP) of an image and Gray-Level Co-Occurrence Matrix (GLCM) features [2]. Before training the NN, a Principal Component which is used in the NN for training. A manually selected training point obtained from the training set of the DRIVE dataset, covering all possible manifestations were used for training the ANN-based binary classifier. The method was evaluated on unknown test samples of DRIVE and STARE databases and returned accuracies of 0.945 and 0.9486 respectively, outperforming other existing supervised learning methods.

Huang, M.L.; Chen, H.Y. Development and comparison of automated classifiers for glaucoma diagnosis using Stratus optical coherence tomography. Investig. Ophthalmol. Vis. Sci. 2018, 46, 4121–4129. [CrossRef] [PubMed].

In this paper, to diagnose diabetic retinopathy, three models like Probabilistic Neural network (PNN), Bayesian Classification and Support vector machine (SVM) are described and their performances are compared. PNN has an accuracy of 87.69% Bayes Classifier [3] has an accuracy of 90.76% and SVM has an accuracy of 95.38%.

Naser, S.S.A.; Ola, A.Z.A. An Expert System for Diagnosing Eye Diseases using CLIPS. J. Theor. Appl. Inf. Technol. 2018, 4, 923–927.

This paper presents a new supervised method for segmentation of blood vessels in retinal photographs. This method uses an ensemble system of bagged and boosted decision trees and utilizes a feature vector based on the orientation analysis of gradient vector field, morphological transformation, line strength measures and Gabor filter [4] responses. The algorithm a suitable tool for automated retinal image analysis.

Farooq, U.; Sattar, N.Y. Improved automatic localization of optic disc in Retinal Fundus using image enhancement techniques and SVM. In Proceedings of the IEEE International Conference on Control Systems, Computing and Engineering, Penang, Malaysia, 27–29 November 2018; pp. 532–537.

This paper presents a supervised method for blood vessel detection in digital retinal image. The use of digital images for eye disease diagnosis could be used for early detection of Diabetic Retinopathy (DR) [5].This method uses the DRIVE database, which has different image conditions.

III. SYSTEM ARCHITECTURE

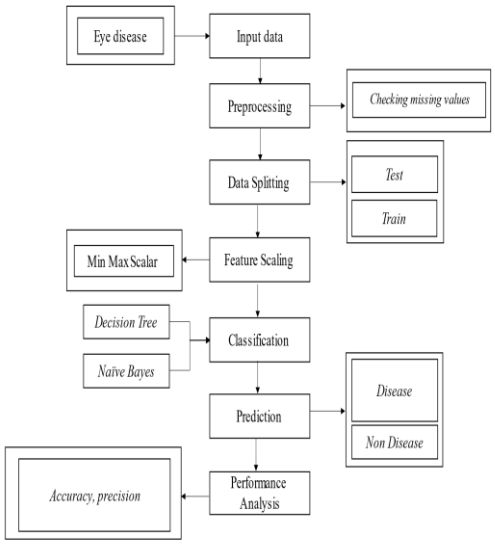


Fig.1.System Architecture

IV. METHODOLOGY

DECISION TREE CLASSIFIER

Several widely used techniques include decision tree classifier multilayer perceptron and Naïve Bayes classifier.Output of a C4.5 decision tree classifier is structural data in the form a binary tree. A C4.5 tree is modelled as follows. A training set is a set of base tuples to determine classes related to these tuples. A tuple X is represented by an adjective vector X = (x1, x2, …, xn). Assume that a tuple belongs to a predefined class that is determined by an adjective called as class label. The training set is randomly selected from the base; this step is called the learning step. This technique is very efficient and extensively uses classification. The structure of the tree can be implemented with the following factors:

1. A node of the tree represents a test on an adjective;
2. A branch exiting from a node represents possible outputs of a test;
3. A leaf represents a class label.

A decision tree includes a rule set by which objective functions can be predicted. The algorithm used for this model uses greedy techniques.

NAÏVE BAYES CLASSIFIER

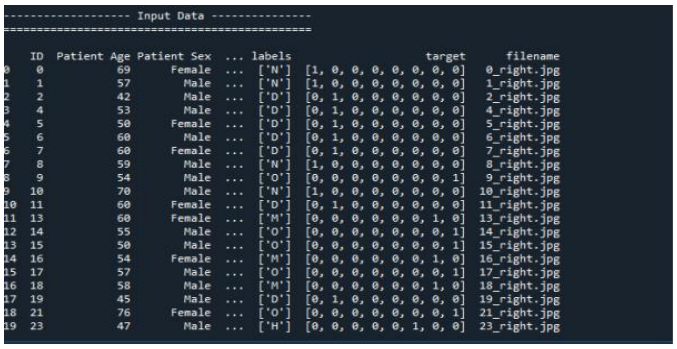
In machine learning, naive Bayes classifiers are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong (naive) independence assumptions between the features. Naive Bayes has been studied extensively since the 1950s. It was introduced under a different name into the text retrieval community in the early 1960s and remains a popular (baseline) method for text categorization, the problem of judging documents as belonging to one category or the other (such as spam or legitimate, sports or politics, etc.) with word frequencies as the features. With appropriate pre-processing, it is competitive in this domain with more advanced methods including support vector machines.t also finds application in automatic medical diagnosis.

Naive Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem. Maximum-likelihood training can be done by evaluating a closed-form expression, which takes linear time, rather than by expensive iterative approximation as used for many other types of classifiers.

V. EXPERIMENTAL RESULTS AND DISCUSSION

Data Selection:

The input data was collected from dataset repository. The data selection is the process of selecting the data for predicting the human activity. In this project, we have to use the eye disease dataset. The dataset which contains the information about the eye disease symptoms. In python, we have to read the dataset by using the pandas packages.Our dataset, is in the form of '.csv' file extension.

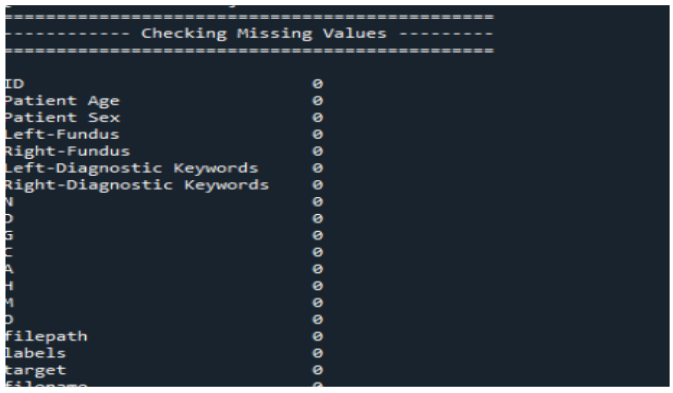


```
------------------ Input Data ---------------
=============================================

    ID  Patient Age Patient Sex  ... labels                  target     filename
0    0           69      Female  ... ['N']  [1, 0, 0, 0, 0, 0, 0, 0]  0_right.jpg
1    1           57        Male  ... ['N']  [1, 0, 0, 0, 0, 0, 0, 0]  1_right.jpg
2    2           42        Male  ... ['D']  [0, 1, 0, 0, 0, 0, 0, 0]  2_right.jpg
3    4           53        Male  ... ['D']  [0, 1, 0, 0, 0, 0, 0, 0]  4_right.jpg
4    5           50      Female  ... ['D']  [0, 1, 0, 0, 0, 0, 0, 0]  5_right.jpg
5    6           60        Male  ... ['D']  [0, 1, 0, 0, 0, 0, 0, 0]  6_right.jpg
6    7           60      Female  ... ['D']  [0, 1, 0, 0, 0, 0, 0, 0]  7_right.jpg
7    8           59        Male  ... ['N']  [1, 0, 0, 0, 0, 0, 0, 0]  8_right.jpg
8    9           54        Male  ... ['O']  [0, 0, 0, 0, 0, 0, 0, 1]  9_right.jpg
9   10           70        Male  ... ['N']  [1, 0, 0, 0, 0, 0, 0, 0]  10_right.jpg
10  11           60      Female  ... ['D']  [0, 1, 0, 0, 0, 0, 0, 0]  11_right.jpg
11  13           60      Female  ... ['M']  [0, 0, 0, 0, 0, 0, 1, 0]  13_right.jpg
12  14           55        Male  ... ['O']  [0, 0, 0, 0, 0, 0, 0, 1]  14_right.jpg
13  15           50        Male  ... ['O']  [0, 0, 0, 0, 0, 0, 0, 1]  15_right.jpg
14  16           54      Female  ... ['M']  [0, 0, 0, 0, 0, 0, 1, 0]  16_right.jpg
15  17           57        Male  ... ['O']  [0, 0, 0, 0, 0, 0, 0, 1]  17_right.jpg
16  18           58        Male  ... ['M']  [0, 0, 0, 0, 0, 0, 1, 0]  18_right.jpg
17  19           45        Male  ... ['D']  [0, 1, 0, 0, 0, 0, 0, 0]  19_right.jpg
18  21           76      Female  ... ['O']  [0, 0, 0, 0, 0, 0, 0, 1]  21_right.jpg
19  23           47        Male  ... ['H']  [0, 0, 0, 0, 0, 1, 0, 0]  23_right.jpg
```

Fig.2.Data Selection

Data Preprocessing:

Data pre-processing is the process of removing the unwanted data from the dataset. Pre-processing data transformation operations are used to transform the dataset into a structure suitable for machine learning. Missing data removal: In this process, the null values such as missing values and Nan values are replaced by 0. Encoding Categorical data: That categorical data is defined as variables with a finite set of label values.



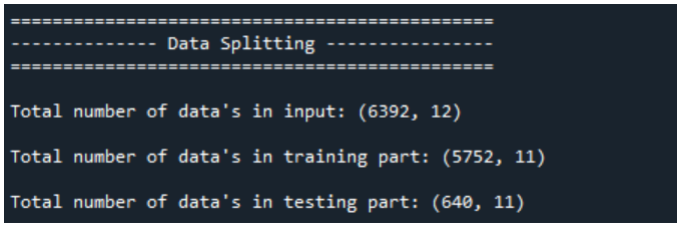
```
=============================================
------------ Checking Missing Values ---------
=============================================

ID                          0
Patient Age                 0
Patient Sex                 0
Left-Fundus                 0
Right-Fundus                0
Left-Diagnostic Keywords    0
Right-Diagnostic Keywords   0
N                           0
D                           0
G                           0
C                           0
A                           0
H                           0
M                           0
O                           0
Filepath                    0
labels                      0
target                      0
filename                    0
```

Fig.3.Data Pre-Processing

Data Splitting:

During the machine learning process, data are needed so that learning can take place. In addition to the data required for training, test data are needed to evaluate the performance of the algorithm in order to see how well it works. In our process, we considered 70% of the dataset to be the training data and the remaining 30% to be the testing data.

Data splitting is the act of partitioning available data into two portions, usually for crossvalidator purposes. One Portion of the data is used to develop a predictive model and the other to evaluate the model's performance.



```
=============================================
-------------- Data Splitting ----------------
=============================================

Total number of data's in input: (6392, 12)

Total number of data's in training part: (5752, 11)

Total number of data's in testing part: (640, 11)
```

Fig.4.Data Splitting

Feature Scaling:

Feature scaling is a method used to normalize the range of independent variables or features of data. In data processing, it is also known as data normalization and is generally performed during the data preprocessing step. Feature Scaling is a technique to standardize the independent features present in the data in a fixed range. It is performed during the data pre-processing to handle highly varying magnitudes or values or units. If feature scaling is not done, then a machine learning algorithm tends to weigh greater values, higher and consider smaller values as the lower values, regardless of the unit of the values. Calling of the data makes it easy for a model to learn and understand the problem.

```
=============================================
-------------- Standard Scalar ---------------
=============================================

C:\ProgramData\Anaconda3\lib\site-packages\sklearn\preprocessing\data.py:645:
DataConversionWarning: Data with input dtype int32, int64 were all converted to float64
by StandardScaler.
  return self.partial_fit(X, y)
C:\ProgramData\Anaconda3\lib\site-packages\sklearn\base.py:464: DataConversionWarning:
Data with input dtype int32, int64 were all converted to float64 by StandardScaler.
  return self.fit(X, **fit_params).transform(X)
C:/Eye Disease Prediction Using ML Techniques/Source_code/main.py:101:
DataConversionWarning: Data with input dtype int32, int64 were all converted to float64
by StandardScaler.
  X_test = sc.transform(X_test)
[[ 1.19900814 -2.01990008 -1.07251125 ... -0.18432166 -0.22450335
  -0.57253224]
 [ 0.08435369 -0.49476681 -1.07251125 ... -0.18432166 -0.22450335
  -0.57253224]
 [ 0.28920147  0.69144796 -1.07251125 ... -0.18432166 -0.22450335
  -0.57253224]
 ...
 [ 1.70688934 -1.34206307  0.93239116 ... -0.18432166 -0.22450335
  -0.57253224]
 [-0.46944167  0.35252945 -1.07251125 ... -0.18432166  4.45427656
  -0.57253224]
 [-1.09599331  0.01361095  0.93239116 ...  5.42529837 -0.22450335
   1.74662653]]
```

Decision Tree:

```
============================================
-------------- Decision Tree ---------------
============================================

1. Accuracy for decision tree:  84.78789986091793

--------Classification Report-------
              precision    recall  f1-score   support

           0       0.88      0.96      0.92       240
           1       0.77      0.92      0.84       266
           2       0.81      0.96      0.88      1438
           3       0.85      0.92      0.88       257
           4       0.95      0.78      0.86       118
           5       0.85      0.97      0.90       208
           6       0.93      0.81      0.87      2578
           7       0.65      0.61      0.63       647

   micro avg       0.85      0.85      0.85      5752
   macro avg       0.84      0.87      0.85      5752
weighted avg       0.85      0.85      0.85      5752
```

Fig.6.Decision Tree

Naïve Bayes

```
============================================
=

-------------- Naives Bayes ---------------

============================================
=

1. Accuracy for naïve bayes:  82.35396383866481

--------Classification Report-------

          precision    recall  f1-score   support

       0       0.84      0.98      0.90       240

       1       0.74      1.00      0.85       266

       2       0.79      0.91      0.85      1438

       3       0.76      0.96      0.85       257

       4       0.62      1.00      0.77       118

       5       0.80      1.00      0.89       208

       6       1.00      0.74      0.85      2578

       7       0.58      0.72      0.64       647


   micro avg       0.82      0.82      0.82      5752

   macro avg       0.77      0.91      0.82      5752

weighted avg       0.86      0.82      0.83      5752
```

Final Result:

```
Enter the value : 77
The predicted result is..... 77
====================================
Eye Disease ---- Cataract (C)
====================================
```

VI. CONCLUSION

Eye disease prediction is implemented by usingdecision tree and naïve bayes.The performance of this proposed approach is evaluated using some measures like, accuracy and precision. Then, we have to predict or to classify the eye diseases based on symptoms.Among the two classifiers decision tree technique is performing well with accuracy of 84.78789.

REFERENCES

1. An, G.; Omodaka, K.; Tsuda, S.; Shiga, Y.; Takada, N.; Kikawa, T.; Nakazawa, T.; Yokota, H.; Akiba, M. Comparison of machine-learning classification models for glaucoma management. J. Healthc. 2021. [CrossRef] [PubMed]

2. Saito, K.; Nakano, R. Medical diagnostic expert system based on PDP model. In Proceedings of the IEEE International Conference on Neural

Networks, San Diego, CA, USA, 24–27 July 2020 pp. 255–262.

3. Huang, M.L.; Chen, H.Y. Development and comparison of automated classifiers for glaucoma diagnosis using Stratus optical coherence tomography. Investig. Ophthalmol. Vis. Sci. 2018, 46, 4121–4129. [CrossRef] [PubMed]

4. Naser, S.S.A.; Ola, A.Z.A. An Expert System for Diagnosing Eye Diseases using CLIPS. J. Theor. Appl. Inf. Technol. 2018, 4, 923–927.

5. Farooq, U.; Sattar, N.Y. Improved automatic localization of optic disc in Retinal Fundus using image enhancement techniques and SVM. In Proceedings of the IEEE International Conference on Control Systems, Computing and Engineering, Penang, Malaysia, 27–29 November 2018; pp. 532–537.