INTERNATIONAL JOURNAL OF RESEARCH CULTURE SOCIETY ISSN: 2456-6683 Volume - 4, Issue - 6, June – 2020
Monthly, Peer-Reviewed, Refereed, Indexed Journal Scientific Journal Impact Factor: 5.245
Received on : 04/06/2020 Accepted on : 19/06/2020 Publication Date: 30/06/2020

Diabetes risk prediction model of connected organs using retinal images

[1]M. Vidhyasree, [2]Dr. R. Parameswari
[1]Research Scholar, [2]Associate Professor
[1, 2] Computer Science, Vels Institute of Science Technology& Advanced Studies, Chennai, India
Email – [1]mailtovidhyasree@gmail.com, [2]dr.r.parameswari16@gmail.com

Abstract: *Healthcare is the fastest-growing domain that aims to provide the appropriate treatment to delay the disease progression. Healthcare domain consists of the most important subdomain called disease prediction which is in need of an efficient technique to predict the disease in early stages. Data mining provides efficient techniques for early disease prediction. This work uses existing classification techniques to build the risk prediction model for early prediction of diabetes. This risk model predicts the possibility of the patient affected by diabetes after the age of 40. The dataset consists of set of retinal images of high myopia and Central Serous Retinopathy of PCOS, pancreatitis and lifestyle disorder. This work also explains the correlation of diseases like PCOS and pancreatic cancer with diabetes. Classification techniques like Adaboost, Gradient boosting, Decision tree, Navies Bayes, Logistic regression and neural network are compared to find an algorithm with high accuracy.*

Key words*: Metabolic Disease, Disease Correlation, Early Prediction, Adult-Onset Diabetes, PCOS, Pancreatitis.*

1. INTRODUCTION:

Data mining is defined as the technique used to find the anomalies, hidden patterns and correlations from the large databases [11]. Data mining consists of approaches like statistical and machine learning approaches used to build the predictive model efficiently. These approaches consist of classification and regression techniques can be applied in healthcare domain for disease prediction. Data mining plays the vital role to extract the hidden patterns and datasets from the database accordingly [12]. The datasets like scan images or recorded pathological values are used for early prediction of diabetes as shown in fig 4. Data mining consists of different learning techniques such as supervised, unsupervised, deep learning and transfer learning used for chronic disease prediction. This work uses supervised learning to predict the diabetes using retinal images. Supervised learning is defined as the technique used to map [13] the corresponding input and output labels for accurate disease prediction. The early prediction of the diabetes can be done by understanding the correlation of diseases like PCOS, pancreatitis and lifestyle disorders to diabetes. This work concentrates on early prediction of diabetes by retinal images of the PCOS, pancreatitis and lifestyle disorder patients. PCOS is the hormonal imbalance categorized into four types such as Insulin resistant PCOS; Pill induced PCOS, Inflammatory PCOS and Hidden PCOS [14]. Insulin resistant PCOS can be diagnosed in three stages such as mild, moderate and severe PCOS that lead to diabetes in later stages. The main reason for PCOS patient with diabetic eye complications will have high estrogen levels as shown in fig 1.

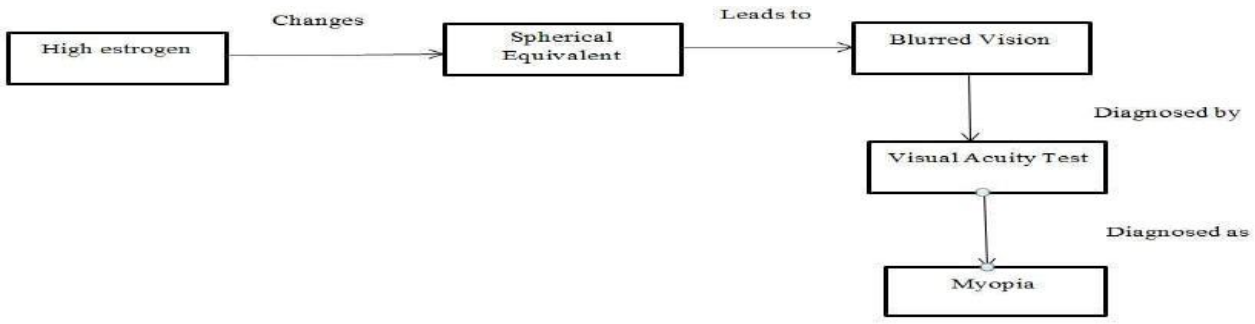

Fig 1.Eye Disease Correlated with PCOS

The pancreatic problem can be diagnosed at three stages such as acute pancreatitis, chronic pancreatitis and pancreatic cancer leads to the sudden loss of vision or dry eyes accordingly. Chronic pancreatitis combined with Sjogren's syndrome leads to dry eyes because the lacrimal glands will gets affected as shown in fig 2.

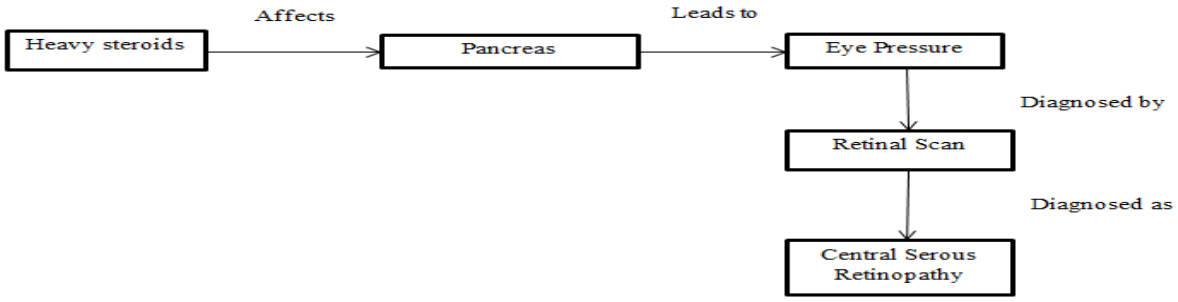

Fig2. Eye Disease Correlated with Pancreatitis

Lifestyle factors are defined as the habits followed in day-to-day life shows significant changes in blood glucose levels. High stress is the most dangerous lifestyle disorder reflected in eye as shown in fig 3.

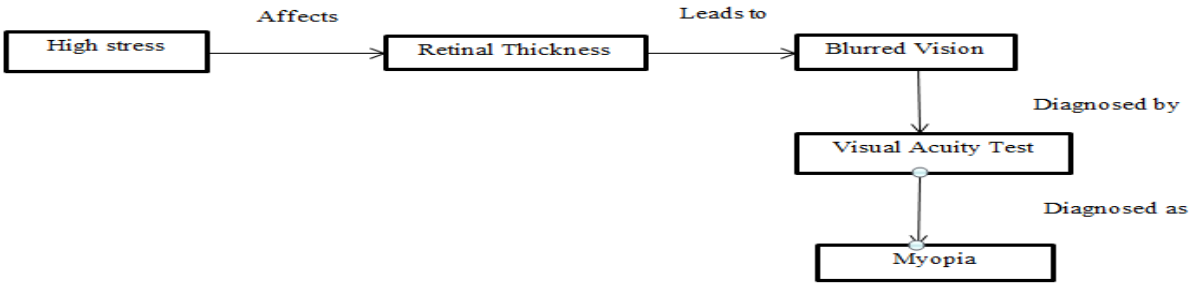

Fig3. Eye Disease Correlated with Lifestyle Factors

The main objective of this work is to predict the diabetes risk by the factors like PCOD, pancreatitis and stress. The retinal images also capture unhealthy lifestyle followed by the person diagnosed with PCOS and pancreatitis. Retina is the most sensitive part of the visual receptor organ that reflects the severity of disease like PCOD, pancreatitis and lifestyle disorders. Retinal images can also reflect the initial stage of PCOD and pancreatitis leads to diabetes. The set of retinal images are used to build diabetes risk prediction model to predict the disease at earlier stages.

2. LITERATURE REVIEW:

Diabetes is the metabolic disorder can be classified into four types such as prediabetes, type 1, type 2 and gastro intestinal diabetes. This can be diagnosed by taking pathological samples and scan images of particular organs. Arif khan implements graph theory to predict type 2 diabetes with the efficiency of 87% [1]. Zhongxian Xu implements weighted feature selection, XG boost ensemble and random forest classifier to build risk prediction model using Pima diabetes database with the accuracy of 88% [2].Mohammed Azeem and Kamal nazir implements machine learning algorithm for diabetic prediction with accuracy of 82% [3]. Saeed Farsi and Sahar Kannian implements machine learning algorithm to predict diabetic complications with accuracy of 83% [4]. Deeraj Shetty uses data mining techniques to predict diabetes with prediction accuracy of 84% [5]. Dyah Ceni Adelina uses iris images to predict the diabetes by monitoring pancreatic cells with prediction accuracy of 80% [6].

Khaled implements machine learning techniques to predict diabetic macular edema with publically available datasets and has got prediction accuracy of 85% [7]. Mustafa Ali Shahin implements deep learning techniques to predict insomnia due to heavy usage of internet with prediction accuracy of 84% [8]. C.Y. Chan implements transfer learning techniques to predict diabetic macular edema using DME patient dataset with 94% accuracy [9]. Kishore Kumar Reddy builds predictive model using clustering techniques with the accuracy of 86% [10]. This section discusses about the early prediction of diabetes in terms of prediction accuracy of the techniques used. The algorithm's efficiency does not only depends on prediction accuracy but also with the other attributes like processing time, attributes taken for model creation, number of images or dataset used. These techniques can be improved by improving processing time and the attributes taken for model training.

According to the previous study, the pathological samples or scan images are taken for diabetic patients to build predictive models using classification or clustering techniques. According to the current status, prediction model build by using pathological values or scan images of eye, heart, kidney to predict diabetic complications. Early prediction of diabetes can be done by considering retinal images of PCOS, Pancreatitis and lifestyle disorders.

3. DIABETES RISK PREDICTION MODEL USING RETINAL IMAGES:

This section discusses the techniques used to build a diabetic risk prediction model with retinal images of PCOS, pancreatitis and lifestyle disorders. This methodology consists of three phases such as input, pre-processing, classification and output phase discussed briefly in fig 4.

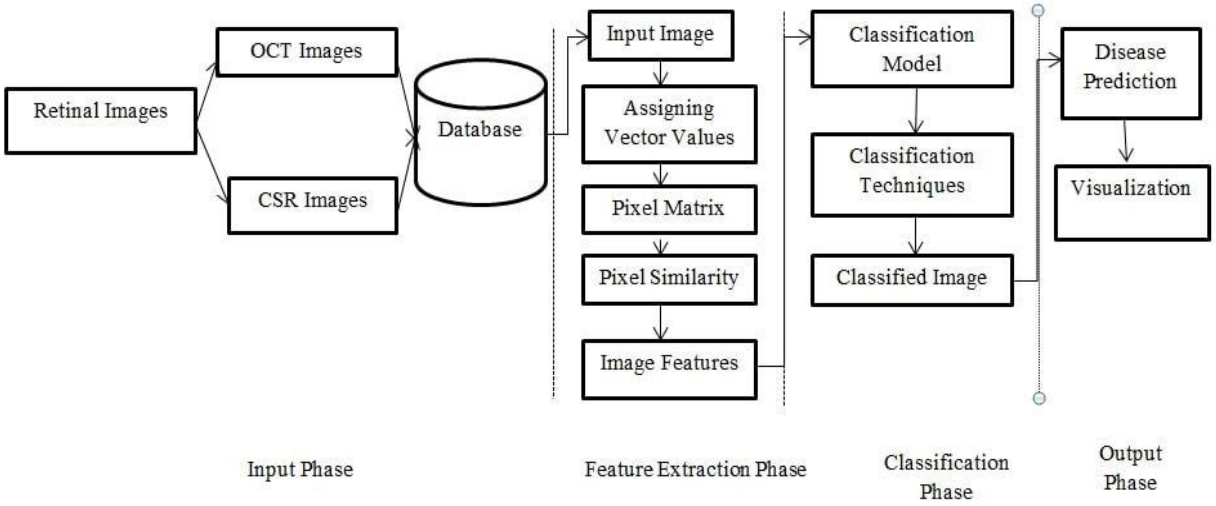

Fig 4. System Architecture

a) Input Phase

Input phase consists of retinal images collected in different sources is stored in database for further processing. The input image taken from the database will undergo pre-processing stage for feature extraction and image enhancement.

Pseudo code for Input Phase

```
//Input Image
Import required libraries
For each image
        Do
                Define the image path
                Load images in array format
                Return image files
        End
End
```

The above pseudo code lists the steps to be followed while reading the image from image file. The images in the file are arranged in array format to read multiple images at a time.

b) Pre-processing Phase

Pre-processing phase plays a significant role in image-based disease prediction because unclear information of data is removed from the input image. The enhanced image features extracted manually (fig 5) by feature extraction algorithm. These features will be given to the next phase.

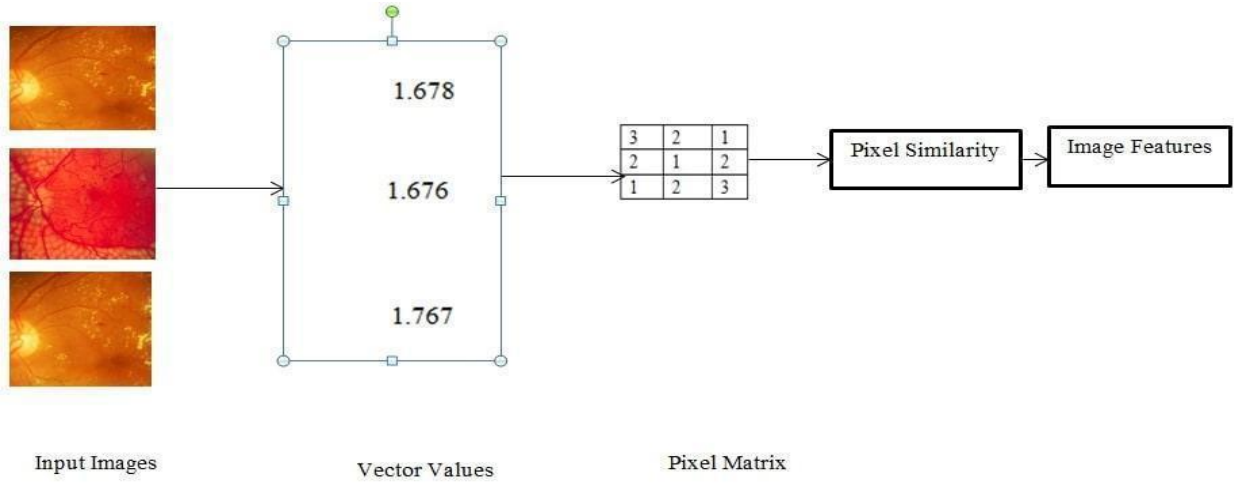

Fig5. Pre-processing Phase

Pseudo code for Pre-processing Phase

```
For each input image
Do
//Image resizing
        Read original image
        Change height and width of each image
        Display Image
//Image denoising
For each image
Do
        for(i= 1..len)
                Arrange the image pixel in array
                blur () //Blur the noise
                displayimage () //Display image
//feature extraction
For each image
Do
        For (i = 1...n)
                Extract image pixels

                Determine pixel similarity

                Exact image features
        End
        End
        End
End
```

The above pseudo code consists of three major steps such as image resizing, Image de-noising and feature extraction. The main aim of resizing and de-noising is to enhance the image and the features extracted are given to the next phase.

c) Classification Phase

The features extracted from the image are given to the classification phase to classify the images under given label. The image features extracted in fig 5 is used to build classification model and this model is trained to get accurate results as shown in fig 6.

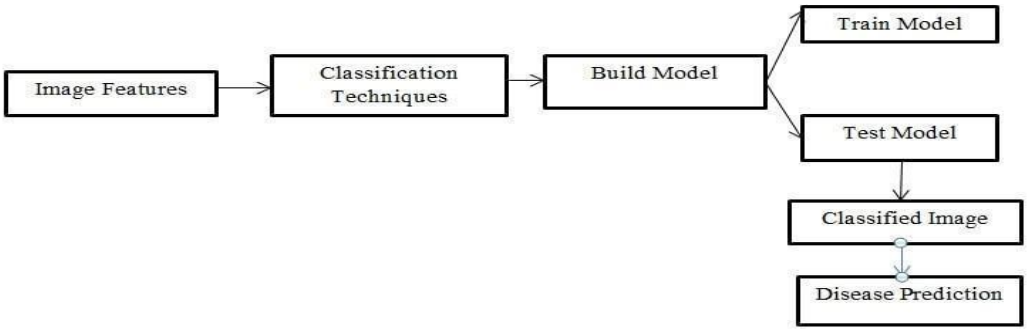

Fig 6. Classification Phase

Pseudo code for Classification Phase

```
For each image features
Do
        for(i =1...n)
        buildmodel ()
        trainmodel ()
        testmodel ()
End
```

d) Output Phase

The classified image extracted from image is given to the output phase is displayed as output with the corresponding image label. The risk prediction model created with classification techniques are visualized by the techniques like scatter plot, confusion matrix etc. The pseudo code of output phase is shown

Pseudo code for Output Phase

For each classified image
Do
//Plot the image
for(i = 1..n)
ggplot()
//Calculate performance matrix
MSE = SQRT (Actual value – Predicted value)^2
End

4. RESULT & DISCUSSION:

Classification techniques like Adaboost, Gradient boosting, Logistic Regression, Navies Bayes, SVM, Decision Trees, Random Forest and neural networks are used to predict the diabetes earlier. The classification techniques are compared to find the algorithm with high prediction accuracy. Mean square error is the important attribute considered for calculating prediction accuracy shown in fig 7.

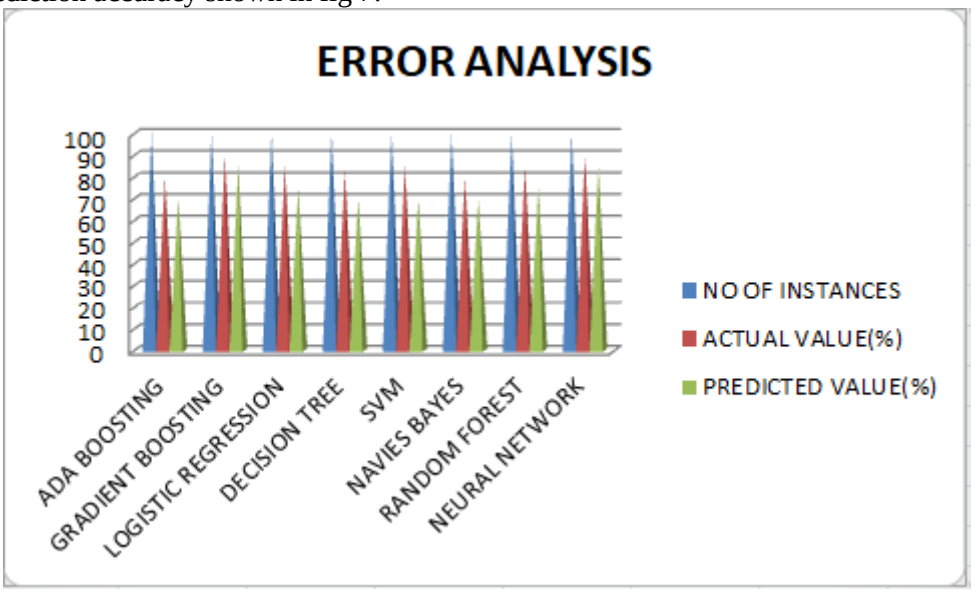

Fig 7. Error Analysis

Mean square value calculated to find the difference between actual and predicted values to find the efficiency of the algorithm and its corresponding visualization is shown in fig 7.

$$MSE = \sqrt{(Actual\ value - Predicted\ Value)^2}$$

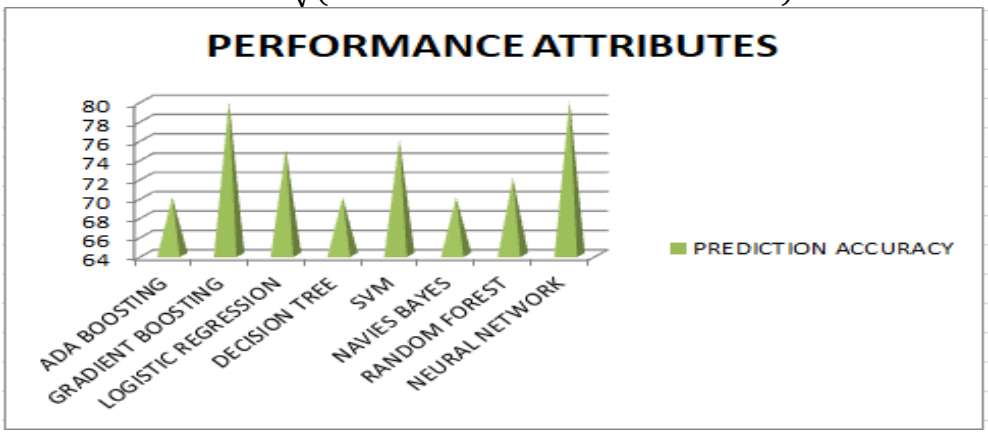

Fig8. Comparative Analysis

According to comparative analysis shown in fig 8 shows neural network has prediction accuracy of 80%. Prediction accuracy is calculated based on the values of number of hit and miss values.

5. CONCLUSION:

This work discusses about the prediction of diabetes by considering the risk factors like PCOS, Pancreatitis and lifestyle disorder with retinal images of high myopia and central serous retinopathy. Diabetes risk prediction model is built with retinal images of eye disease connected with PCOS, pancreatitis and lifestyle disorders leads to diabetes. Diabetes risk prediction is done with existing classification techniques like decision trees, adaboost, logistic regression, neural network etc. The prediction accuracy of existing algorithms are compared to find the algorithm with best accuracy. This work can be extended by using neural network to build diabetic risk prediction model.

REFERENCES:

1. Arif Khan, Shahadat Uddin and Uma Srinivasan, (2019) "Chronic Disease Prediction using Administrative Data and Graph Theory The Case of Type 2 Diabetes", Journal of Expert Systems and Applications, Vol 136, pp - 230 -241.
2. Zhongxian Xu and Zhiliang Wang, (2019)"A Risk Prediction Model Type 2 Diabetes Based on Weighted Feature Selection of Random Forest and XG boost Ensemble Classifier" IEEE International Conference on Communication and Electronic System pp 41- 46.
3. Mohamed Azeem, Nazir Kamal, (2018) "Prediction of Diabetes using Machine Learning Algorithm in healthcare" IEEE International Conference on Automation and Computing, pp 1 – 6.
4. Saeed Farzi , Sahar Kannian, (2017) "Predicting Serious Diabetic Complications using Hidden Pattern Detection" 4th International Conference on Knowledge Based Engineering and innovation, pp 63-68.
5. Deeraj Shetty , (2017) "Diabetes Disease Prediction using Data Mining", IEEE Conference on Innovations in Information and Embedded and Communication Systems, pp 338 – 343, 2017
6. Dyah Ceni Adelina & Riyatno Sigit, (2017) "Identification of Diabetes in Pancreatic Organs using Iridology" IEEE International Electronics Symposium on Knowledge Creation and Intelligent Computing pp 114-119, 2017.
7. Khaled Alsaih & Guillaume Lemaitre, (2017)"Machine Learning Technique for Diabetic Macular Edema Classification on SD- OCT Images", Journal on Biomedical Engineering, Volume 16 No 68, pp 1-12.
8. Mostafa Ali Shahin & Beena Ahamed, (2017)"Deep Learning and Insomnia: Assisting Clinicians with Diagnosis" IEEE Journal of Biomedical and Health Informatics vol 21(6) pp 1533-1542.
9. C.Y. Chan & Awais Mohamed, (2017)"Transfer Learning for Diabetic Macular Edema Detection on Optical Coherence Images" IEEE International Conference on Signal and Image Processing Applications, pp 493- 496.
10. Kishore Kumar Reddy , (2015) "Detection of Pancreatic Cancer using Clustering and Wavelet Transform Techniques" International Conference on Computational Intelligence and Communication Network, pp 332 336. 2015.
11. https://www.sas.com › en_in › insights › analytics › data-mining
12. https://www.usfhealthonline.com/resources/key-concepts/data-mining-in-healthcare/
13. https://en.wikipedia.org › wiki › Supervised_learning.
14. https://www.indiraivf.com › types-of-pcos.