ISSN 2278-3091
Volume 10, No.2, March - April 2021
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse571022021.pdf
https://doi.org/10.30534/ijatcse/2021/571022021

A detailed survey on Prognostication of diabetes diagnosis on the basis of machine learning techniques and the detection approaches to diabetic retinopathy using Artificial Intelligence

S.Prasanth[1], M. Roshni Thanka[2], E. Bijolin Edwin[3]
[1]Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India. E-mail ids: sprasanth19@karunya.edu.in
[2]Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India. E-mail ids: roshni@karunya.edu
[3]Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India. E-mail ids: bijolin@karunya.edu

ABSTRACT

The aim of the study is to compare, assess the optimum tools as well as the techniques and advanced features focused on prediction of diabetes diagnosis based on machine learning tactics and diabetic retinopathy using Artificial Intelligence. The literature on data science, Artificial Intelligence (AI) contains important knowledge and understanding of AI entities such as Data science, machine learning, deep learning, Medical image processing, feature extraction, classification techniques, etc. Diabetes diagnosis is a phenomenon that impacts individuals around the globe. Now, with diabetes impacting people from children to the elderly, the out-dated approaches to diabetes diagnosis should be replaced with new, time-saving technologies. There's several studies carried out by researchers to recognise and predict diabetes. Here plenty of classifiers in machine learning can be used, such as KNN, Random Tree, etc.They can save time and get more precise outcome when using these techniques to predict diabetes. Diabetic retinopathy (DR) is a typical disorder of diabetic disease that induces vision-impacting lesions in the retina. It also can turn to visual impairment if it is not addressed early. DR therapy only helps vision. Deep learning has in recent times being one of the most widely used approaches that has accomplished higher outcomes in so many fields, especially in the analysing and identification of medical image classification. In medical image processing, convolutional neural networks (CNN) using transfer learning are commonly used as a deep learning approach and they are incredibly beneficial.

Key words: Diabetic Retinopathy (DR), Artificial Intelligence , Machine Learning strategies, Deep learning , Transfer learning, Medical Image Processing ,Feature extraction, Classifiers.

1. INTRODUCTION

Diabetes diagnosis is a group of chronic illnesses wherein glucose levels or sugar levels stay abnormally high for extended spans of time Frequent urination, elevated appetite, and increased hunger are signs of high blood sugar. Diabetes can cause multiple problems if left unchecked. Diabetic ketoacidosis, hyperosmolar hyper-glycaemic condition, or death may be acute complications. Cardiovascular disease, stroke, progressive kidney disease, foot ulcers, and eye injury are serious long-term risks. Diabetes mellitus, also known as diabetes, is another metabolic disorder that causes elevated blood sugar. The enzyme insulin transports sugar from the blood into the cells for absorption or use for eating. The body either doesn't have sufficient insulin for diabetes or doesn't use the insulin it produces effectively. But also untreated excessive blood sugar from diabetes can be affected in the liver, eye, lungs, kidneys, and other organs.

One of the most severe and chronic conditions that cause blood sugar to increase is considered to be diabetes. Several risks occur as diabetes remains unchecked and unexplained. The exhausting identification process consists of a patient visiting and consulting a consultant at a medical centre. Yet this key issue is answered by the rise in approaches to machine learning.

Classification approaches are widely used in the medical field to classify data subject to such constraints according to a single classifier in multiple classes. Diabetes is a disorder that inhibits the bodies natural capacity to regulate insulin receptors, which in turn induces unhealthy carbohydrate metabolism and increases blood glucose levels. High blood pressure is typically caused by diabetes. Intensifying hunger, intensifying appetite, and frequent urination are all the signs induced by increased blood sugar. Because diabetes is left uncontrolled, it causes a

slew of issues. Metabolic acidosis and – anti metabolic derangement syndrome are two important signs. Diabetes is recognized as a global public health issue in which calculating sugar consumption is challenging. However, diabetes is not only caused by various variables such as height, weight, genetic factor, and insulin, but the main reason is also the concentration of sugar among all variables. Early warning is the safest solution to staying away from the issues. Several researchers use numerous of machine learning approach for classification algorithms to run disease diagnosis trials such as MLP, SVM, Naive Bayes, Decision Tree, Decision Table, etc. Machine-learning algorithms have been shown to be better at diagnosing various illnesses. Data mining and machine learning algorithms are gaining attention as a result of their ability to handle a large number of data sources from many formats, blending data and inserting historical data into the study.

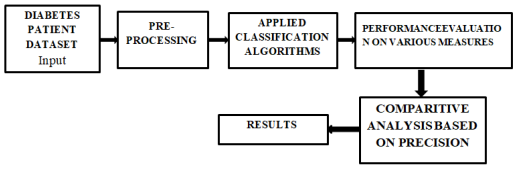

Figure 1.1: Overview of Diabetes prediction process

Diabetic Retinopathy (DR) is a chronic diabetes-associated eye condition. Diabetic retinopathy is the leading cause of vision loss in active working age individuals; Furthermore, it is claimed that over 93 million people could have been infected. If DR is identified in time, progression to vision deficiency may be delayed or controlled, although this can be problematic since the condition sometimes displays few signs before it is too late to have successful care.DR diagnosis is currently a time-consuming and manual procedure that involves an ophthalmologist or qualified clinician to review and analyse images of the retina in the image fundus, to distinguish DR by the appearance of lesions associated with the disease-induced vascular anomalies and in below fig 1.3 the overview of DR which helps to understand better.

Medical specialists are currently conducting an assessment of the seriousness and intensity of retinopathy associated with a person with diabetes depending on the fundus or retinal images of the patient's eyes.If the number of diabetic patients is growing exponentially, the number of retinal images generated by screening systems will also rise, resulting in a substantial union - intensive workload on medical practitioners and the expense of healthcare services.Based on the methodology of deep learning and medical image processing, high sensitivity and high precision can be achieved in the identification of referable diabetic retinopathy, identified as moderate or worse diabetic retinopathy.

A current study in India revealed the estimated incidence of type-2 diabetes and DR in the rural demographic of South India, finding symptoms of patients with type 2 diabetes in roughly 1 in 10 people over the age of forty.

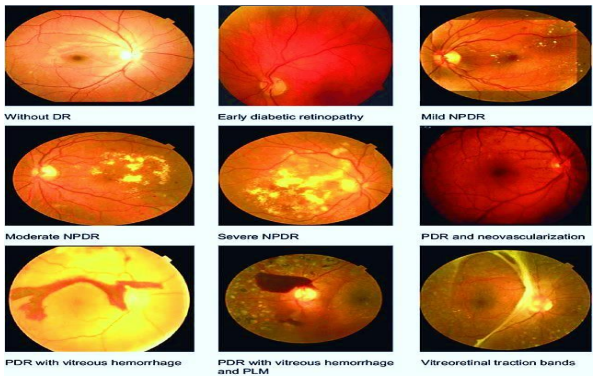

Figure 1.2: Diabetic Retinopathy Stages

Clinical DR severity criteria are divided into five categories. Many patients with DM may not have clinically significant DR early on, but genetic and metabolic variations in the retina part of the eye, such as reduced retinal blood flow, increased leukocyte adhesion, inflammation of the base cells and lack of retinal pericytes, are documented. Mild non-proliferative is the first scientifically evident form of DR. The emergence of microaneurysms characterises diabetic retinopathy. Where further DR lesions arise, including venous calibre modifications and intra-retinal micro-vascular anomalies, Diabetic retinopathy can progress to moderate non-proliferative DR, which Increases in NPDR extremes and retinal blood flow are increasingly affected by the severity and duration of these lesions. As a result, the non-perfused retina regions transmit signals that induce the formation of new blood vessels, taking the direction to proliferative diabetic retinopathy. The new one Blood vesselis irregular, friable, and can quickly bleed, often causing serious loss of vision.Diabetic macular edema happenswhile the retina swells due to fluid escaping from the blood vessels within the macula, which can occur when the retina is swollen during every DR point.

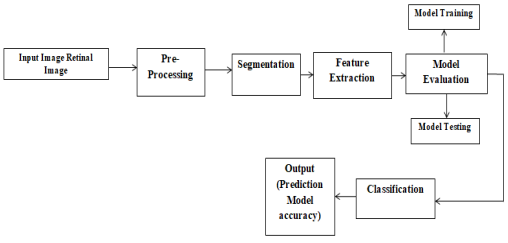

Figure 1.3: Overview of Detection of Diabetic Retinopathy (DR) process

1.1 Pre-Processing in Diabetes diagnosis

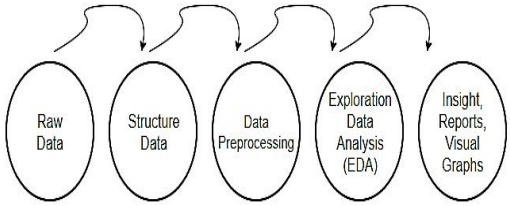

Figure 1.4: Pre-Processing process in diabetes diagnosis

Every repository is a collection of objects with related documents. Sample data, events, observations, and records are all appropriate names. Interestingly, every of them has been characterised by a set of features. In data science terminology, these are referred to as characteristics or features are until creating a model with these attributes, data Preprocessing is needed. By data Preprocessing, we:

- Improve the accuracy of our database. We remove any values that are inaccurate or missing as a result of medical errors or defects.
- Consistency should be improved. The precision is harmed when there are data inconsistencies or duplicate entries.
- Make the data as achieve as possible. If necessary, we could even replace in the missing characteristics.
- The collected data should be clean. We find things simpler to use and perceive this way.

1.1.2 Pre-Processing in Diabetic Retinopathy:

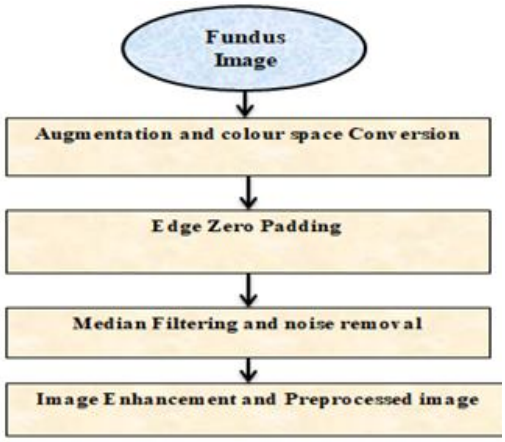

Figure 1.5: Pre-Processing process in diabetic retinopathy

The light that reaches the retina is not standardized, resulting in contrast differences and brightness in particular areas at every given time, the illumination will be far more or less, and some part will be missing. In poor contrast or low visibility, defects are not visible. Furthermore, images can vary in terms of hue and accuracy. As a consequence and here image Preprocessing is necessary.

In Pre - processing stage, one of the first steps is to resize the files. Size is used to translate the images to grey scale before feeding them into the classification framework. Instead, turn oneself into a L shape. It was a monochrome representation that's used to display micro aneurysms and blood vessels in fundus pictures. Also, for further editing, flatten the single-dimensional images.

1.2 Approaches and Methods for classifications(classifiers) used in this project

Machine learning methods are used in multiple prognostication classifiers. The machine learning process of training a mapping function an input to even an output relying on instance input-output pairs is known as supervised learning. It uses named training data and a series of training experiences to conclude a feature.
Define the Features to their respective characteristics using classification algorithms. In, we use some typical classifiers like:

- Multi-layer Perceptron
- K-Nearest Neighbour
- Gaussian Mixture Model
- Hidden Markov Model
- Support Vector Machine
- singular value Decomposition
- Artificial Neural Network
- Random Forest
- Decision Trees
- Boosting
- Naïve Bayes
- Classification And Regression Trees

In data science, machine learning is the classification of a supervised theory of learning that separates a set of data into categories. Voice, detection, facial identification, and other classification problems are the most critical. It may be a problem with binary sorting or a problem with different groups. There are several common machine learning algorithms for classification and recognition in machine learning.
The structure of the visual cortex of organisms where neurons are organized in certain manner that they lead to stimuli and cluster centres that tile the field of vision determines the template of contact among its neurons in a cnn.The cnn employs a complex structure of textured strands in which the visuals are ideally appropriate for deep learning classifiers. CNN provides Robustness and Delicate results for each feature present in the images.
We use transfer learning approaches of cnn based models to classify the model, and some of them are:

- LeNet- 1990
- AlexNet- 2012
- ZFNet- 2013
- GoogLeNet- 2014
- VCGNet- 2014
- Resent- 2015
- Inception v3- 2016
- EyeNet- 2015
- MobileNet- 2017
- EfficientNet- 2019

1.3 Objective

- The motivation of this project is to create a simplest yet complete representative of prediction of diabetes diagnosis detection and identification of diabetic retinopathy,
- This study uses a distinctive classifier to determine whether or not a patient will develop diabetes. As previously mentioned, the dataset includes the labels for the model's dummy variable Result. The goal is to predict the risk of diabetes based on a subset of variables such as blood pressure, insulin levels, glucose, skin thickness, and BMI.
- The objectives of this work is to develop a modern technologies that can combine the impacts of multiple machine learning tactics to provide a more precise earlier start diabetes prognostication for a service user. SVM, Logistic regression, ANN, MLP, Random Forest, Decision Tree, Boosting, KNN, and other machine learning techniques were used in this project plan to prognosticate diabetes. This study also aims to recommend an essential method for detecting diabetes disease early on.
- The detection of diabetic retinopathy is done using deep learning and medical image processing tactics.

2. LITERATURE SURVEY

2.1Related Works

There has been a significant amount of study has been done in the area of Machine learning based prediction on diabetes diagnosis detection.

In [1] they have a proposed a Prediction of Diabetes using Classification Algorithms and The focus of this study is to build a model that can more accurately assess the threats of diabetes in patients. As a result, this thesis hires three machine learning classification algorithms and techniques also a tentative diagnosis of diabetes. The results of all three algorithms are calculated using various metrics. Accuracy is measured in terms of correctly categorized and incorrectly classified cases. The findings show that Naive Bayes outperforms other algorithms, with an accuracy of 76.30 percentage growth.

In [2] they have recommended theMachine Learning Based Unified Framework for Diabetes Prediction. We suggested a system for the estimation, tracking and execution of diabetes in real time. Our mission is to create an optimised and effective framework for machine learning (ML) that can identify and forecast the state of diabetes effectively. In this research, the five most important classification strategies for machine learning were considered for predicting diabetes. Fortunately, in order to evaluate the feasibility of these classification processes, more research is needed. As compared to the other four classifiers, the analysis showed that Nave Bayes had the highest performance, with an F1 measure of 0.74.

In [3] they have proposed an Analysis and Prediction of Diabetes based on Machine Learning.The key objective is to detect new trends and then to analyse these patterns and provide users with relevant and usable data. Diabetes leads to heart failure, kidney disease, blindness and nerve damage. A main issue is the effective mining of diabetes data. The techniques and techniques of data mining will be discovered to identify the required methods and techniques for effective diabetes dataset classification and extraction of useful patterns.To construct an accurate model, the dataset was studied and analysed. Diabetes disease detection and diagnosis. In this research, we plan to use the resampling technique of bootstrapping to improve precision and then add Naïve Bayes classifier, Decision Trees algorithm and Knn classifiers and compare their efficiency.

In [4] they have proposed a Machine learning framework for diabetes disease diagnosis and detection. The main goal of this study is to develop a machine learning-based approach for forecasting diabetic patients. Primarily focused on the p value and odds ratio, logistic regression is used to classify risk factors for disease prediction. .To determine diabetic patients, we used four distinct classifiers. These protocols have also been followed and replicated in 20 trails by three groups of partition protocols (K2, K5, and K10). The precision of these classifiers is used to test their reliability.

In [5] they have explored implementing Machine Learning Techniques to predict diabetes. In this study, various machine learning strategies were added to the input, and detection was performed using various classifiers, with Logistic Regression ensuring the finest of 96 % accuracy. Pipeline implementation offered the Adaboost classifier with 98.8 % accuracy as the best model. There was also a variation in performance between the machine learning model and two separate samples. In comparison to current data sets, it is apparent that the methodology improves the reliability as well as accuracy of diabetes prediction with this dataset.

In [6] A Comparative Study of Machine Learning Methods to Predict Diabetic Mellitus was discussed. For the estimation of diabetic mellitus in adult population data, we use four widely used machine learning algorithms: Support Vector Machine, Naive Bayes, K-Nearest Neighbour, and C4.5 Decision Tree. In comparison to other machine learning models, the C4.5 decision tree obtained better accuracy according to our results.

In [7] they suggested the Detection of diabetes Disease Prediction Based ML on Public health care Big Data analytics. This work aims to build a binary classifier using the WEKA tool to predict future diabetes diagnosis using various classifiers, and the Base model is CART classifier.Centered on promising actual outcomes, the study intends to recommend the best model for diabetes disease diagnosis. Taking a look at the test findings of each classifier in the dataset. It was discovered that the SVM performed well in disease diagnosis, with the greatest precision the highest, which is 79.13. SVM has a significantly better accuracy rate in anticipating diabetes than other classification algorithm.

In [8] they discussed by using auto tuning multi-layer perceptron to build an accomplished method for diabetes prognostication. The thesis builds an unique outlier recognition approach by combining an AutoMLP mechanism instead of an Enhanced Class Outlier Detection exceptional case recognition system that uses a distance-based methodology. The Automation Multi-Layer Perceptron proposed approach is auto-tunable, which means it instantaneously optimizes parameters during in the training process of the test, which would otherwise require human interference. Outlier detection is performed by our system during the pre-processing of data. The machine achieved an accuracy of 88.7%, which was higher than any previous findings.

In [9] they've suggested incorporating machine learning approaches to evaluate diabetes. In the world of data science, ML is a modern area of research that explores how computers learn from their experiences. The aim of this research is to develop technologies which could incorporate the results of different machine learning methods to provide a more early risk diabetes predictor for an individual. This article intends to anticipate diabetic using various supervised machine learning tactics SVM, logistic regression, and ANN.This initiative seems to have the aim of recommending an appropriate strategy. We selected SVM to anticipate diabetes as it is an efficient method for comparative classifier. The justification for this is that SVM is well known for its discriminatory identification power, specifically when a substantial majority of classification methods are concerned in the features, and the component of the function in our case is 7.

In [10] they have proposed a Prediction of Type 2 Diabetes using Machine Learning Classification Methods.This article investigated the likelihood of diabetic in individuals based on their diet and personal health records. Using multiple machine learning algorithms, the probability of type 2 diabetes was estimated because these algorithms are highly reliable, which is really important in the health care profession. When the model is educated with reasonable specificity, people will determine the risk of diabetes on their own. The precision of our dataset's Random Forest is 94.10%, which is the best among the others. For the PIMA dataset, Random Forest often has the best precision. Learning algorithms applied to six different machines of classifiers

In [11] they suggested a Diabetes Analysis Using Various Machine Learning Methodologies. The aim of this research is to develop a tool for determining a patient's diabetes perceived risk which is more reliable. Pattern is applied using classification techniques such as Decision Trees, ANNs, and SVMs.The models for Decision Tree have a performance of 85 % 77 % for NB, and 77.3 % for SVM. The results point to a high level of accuracy in the techniques.

In [12] they suggest implementing machine learning approaches to classify diabetes medical data. The objective of this work is to discover insights by observing the trends in the repository and using predictive analytics to classify these diseases. In addition, the solution to the neural network is often used to identify current diabetic patient data to predict the disease of the patient based on qualified data that can contribute to the detection of various levels of diabetes affected people. It is often compared to the extraction system of collaboration rule for model evaluation to ensure that the classification is accurate.

In [13] they suggested to use data science of ML tactics to recognize diabetes in data sets. They expect to apply the bootstrapping-like approach in this analysis to upgrade the accuracy and then apply it to different classification strategies and learn about their application. After Bootstrapping (Accuracy Rates %) comparing all classifiers the SVM and Ada boost gives highest precision of 94.44%

In [14] they proposed Analyzing Machine Learning Techniques and Achieving the Greatest Performance for Diabetes Prognostication in Female.We worked to identify the best fitting algorithm for this function in order to successfully predict and diagnose diabetes. To achieve the highest precision, the key purpose is to compare the various algorithms. Few classifiers were compared for finding the best outcome. With the aid of K-Fold and Cross Validation, the final outcome gave us an accuracy of 81.1%.

In [15] they proposed a Research on Diabetes Prediction Method Based on Machine Learning.We use supervised ML algorithms such as SVM, Classifier Naive Bayes classifier and LightGBM in this manuscript to train based on the real data of 520 diabetic and possible diabetic patients between the ages of 16 and 90. The efficiency of the support vector machine is the greatest, by comparative study of classification and recognition accuracy.

In [16] they suggested incorporating Deep Learning to diagnose severity of blindness, which is diabetic retinopathy. This work proposes an automatic classification system in which, using AI systems such as Convolutional network, transfer learning based model like VGG-16 and VGG-19, it analyses input images using different lighting and DR intensity is determined by the number of viewing angles and the seriousness class.This system achieves 80 % of accuracy, 82 % resolution, 82 % of precision, and 0.904 AUC score for the classifying of image data into 5classes ranging from 0 to 4, where 0 is no DR and 4 is proliferative Diabetic Retinopathy stage.

In [17] they suggested a DR Detection using AI technologies which is deep learning model of the Fundus eye images of the patient in this study are being used as a input constraints Furthermore, a training set (DenseNet Architecture, which is a transfer learning based model) can isolate the task of fundus images acquired and send the result after the activation function. The architecture gave the DR detection a precision of 0.9611 and QWK score of 0.8981. And at last in the evaluation stage, we examine the two model of CNN. In which are VGG16 and DenseNet121 architecture.

In [18] they have proposed a DR Detection using Deep CNN.Using deep learning, this study intends to automatically detect the disorder in its multiple phases. This work proposes the creation and implementation of deep convolutional neural networks enhanced by GPU to instantly detect and determine optimal retinal images into 5 phases. Dependent on seriousness of the disorder.On a QWK metric, the single accuracy on the model of the CNN described in this work is 0.386 and the assembly of trio such related models culminated in a score of 0.3996.

In [19] A Deep Learning Ensemble Solution for DR Identification was suggested by them. This study used a publicly available repository of retinal fundus images to build an aggregate of deep CNN transfer learning of classification algorithms to represent the abundant features and improve recognition for specific DR stages. The findings of the experiments showed that, with the exception of current models, the developed method recognizes all phases of DR and outperforms them on the very same data source.

In [20] they introduce a Diabetic Retinopathy Detection Deep Learning Approach. We propose an AI-based methodology for stage detection of DR in this analysis using single imagery of the person fundus. We also recommend the inter transfer learning system which utilizes similar databases with different labeling. On the APTOS 2019 Blindness Identification Test, the device listed will be used as an efficient clinical diagnostic tool for DR, with a precision and efficiency of 0.99.

In [21] DR Detection Using Artificial Intelligence and Conceptual Models is something they've suggested.In this study, we incorporate the use of various texture characteristics, primarily Local Ternary Pattern and Local Energy-based Form Histogram for DR. We illustrate that LBP extracted features outperform them. For the classification of an extracted histogram, SVM are used.A histogram binning method is proposed for representation of characteristics. The experimental findings indicate that LESH is the highest performing technique using Support Vector Machine classifier with a Radial Basis Function kernel with an accuracy of 90.04%. Similarly, the ROC curve analysis reveals that with 93.01%, LESH with SVM-RBF provides the best AUC results.

In [22] they suggested an AI-based Earlier DR Detection and Evaluation Using Retinal Image Data. To accomplish early diagnosis, they introduced a simple deep cnn which detects all microaneurysms, the very first symptoms of DR, and appropriately assigning labeling to features extracted graded into five groups.We checked our model on the largest publicly available open source provided DR repository and achieved a QWK score of 0.851 and an AUC score of 0.844, indicating intensity classification results. We were able to detect with an intensity of 98 % and an accuracy of over 94 %, demonstrating the results of the developed procedure.

In [23] Envisioning Artificial Intelligence Systems for the Identification of Referable Eye Disease and Glaucoma is a suggestion they made. In CNN, to ferociously visualize two validated deep learning strategies for the diagnosis of related diabetic retinopathy. Abnormalities commonly seen during cases of pertaining DR (exudate, haemorrhage, or vessel abnormality) have been recognised as a significant pathological area in 96 out of 100 accurate DR cases. In 39 of 46 deceptive DR cases, the heat map showed visual representation of – anti fundus areas even without retinal venous system.

In [24] they suggested an Image Analysis and Cnn -based Model of Diabetic Retinopathy diseases. This study determines how retinal fundus images may be used to detect retinopathy using image recognition and AI technologies. The fundus images edge detection strategy was using a practical procedure that included HSV, V transformations, and a histogram equalization

technique. Following the image processing, the classification was generated using CNN. The experiments demonstrated that the proposed method for diagnosing DR using image processing algorithms is highly effective and productive.

In [25] they've suggested incorporating simple computational methods to detect diabetic retinopathy. In this study, the unbalanced dataset reduces the precision such that before classification, oversampling is completed. Svm, Random Forests, Boosting algorithms, and Gaussian Naive Bayes were used to identify the DR. Each classifier has its own collection of benefits and degrees of accuracy.SVM has good precision relative to other classifiers.

In [26] for recognizing diabetic retinopathy, they recommend a hybrid deep learning platform. In this study, we propose a hybrid deep learning-based approach for diagnosing DR in fundus images. We use a cnn with a linear svm classifier to train the model on the standard evaluation using an open source data set.Whenever it came to identifying DR, our suggested system has a good tolerance and precision, according to the results of our experiments.

In [27] they have recommended a deep neural network to predict diabetic retinopathy.The current study uses neural networks processing based on the principal component analysis. To define the derived characteristics of diabetic retinopathy, the network model uses the Grey Wolf Optimization algorithm.On the basis of success metrics, precision, memory, sensitivity and specificity, the developed framework technique has been sorely tested. The model was applied to machine learning models such as the regular SVM, Naive Bayes Algorithm, Decision Tree, and XGBoost classifier. The findings demonstrate that improved efficiency is given by the proposed model .Evaluated to the machine learning already stated.

In [28] they recommended incorporating efficient deep learning to diagnose diabetic retinopathy through phase classification in eye fundus videos. This analysis developed an intelligent DR stage recognition came up with a new multi-layer model of active deep learning.To develop the ADL architecture, we used the Cnn model to instantaneously select knowledge comparable to network of individual's features. The avg SE for the ADL-CNN model was 92.20 %, 95.10 % SP and 98 % of accuracy. DR-related lesion recognition is also improved by the updated ADL-CNN model.

In [29] they explored using cnn on a small number of features to detect diabetic retinopathy automatically.In this study, we recommend a CNN colour fundus photography transfer learning network framework which executes comparatively on the whole, also In recognizing diabetic retinopathy categories, a relatively shorter set of data of skewed categories of 3050 training samples and 419 validation images was used. The transfer learning-based Cnn model allows for initial detection of Diabetes-1 and Diabetes-2 on colour fundus images.

In [30] A Study on Diabetic Predictions Using Machine Learning has been suggested by them. It aims to use machine learning models such as SVM and Naive Bayes. Using such system to determine diabetes will help save other time and provide more accurate outcomes.

In [31] they have proposed diabetes to Heart Disease: A Survey. In addition to glucose level, heart rhythm, BMI, era, circulatory pressure, we understand heart disease by the diabetes side effects. The suggested system is to classify the heart condition for which we implement the prediction system based on diabetes and separate the heart disease using SVM measurement.

In [32] they have recommended a Survey on recent developments in automatic detection of diabetic retinopathy. Here we use various tactics on deep learning, machine learning and medical image processing. DR is diagnosed with a CAD protocol. It is used to analyse medical images, Medical diagnosis Image Recognition.

In [33] they recommended a machine learning method for predicting and diagnosing possible diabetes threat We attempted to focus forward in this review on the onset of diabetes, and is one of the global 's highest degenerative illnesses, according to the Health Organization. We have attempted to show numerous approaches including certain Classification Algorithms like GB, LR, and NB, which can be used to diagnose diabetes disease with 86 %accuracy for Gradient Boosting, % accuracy for Logistic Regression, and % accuracy for Naive Bayes.

In [34] they suggested a machine learning methodology for diabetes disease prognostication and diagnosis. The combination of a classifier based on LR and RF works well. For the prediction of diabetic patients, this mixture would be really useful. The ML-based system's overall precision is 90.62%. The fusion of LR-based feature selection and Random Forest based classification algorithm provides 94.25 % accuracy and 0.95 precision for the K10 protocol.

2.2 Title, classifiers used, Performance Metrics, and Merits of an existing system are:

Ref No	Classification (Classifier) and Algorithm used	Performance Metrics	Merits	Demerits
[1]	Decision Tree classifier , SVM and NB	The findings reveal that Naive Bayes outperforms other classifiers, with an accuracy of 76.30 %. These results are verified using Receiver Operating Characteristic curves in a reasonable and validated manner.	Constant Improvement. ML algorithms are able to learn from the knowledge that we have.	The model's performance can be impacted by algorithm selection.
[2]	Five classifiers are used	The findings of the study show that Naïve Bayes obtained the best efficiency, achieving the F1 measure of 0.74, compared with the other classifiers.	Time-consuming	There's a risk that ML model just might make a significant error.
[3]	Naive Bayes Decision Tree, KNN	The WEKA software has been used to diagnosis diabetes as a mining tool. High precision with an accuracy score of 90.36 % and Stump's judgement offered less precision than some by offering 83.72 %.	Handling multi-dimensional and multi-variety data	For train and test the data, we collect a large volume of data. Such procedure can occasionally result in data uncertainty.
[4]	NB , Decision Tree, Ada Boost, RF classifier	The cumulative precision of the ML-based model is 90.62%. The combined effect of LR-based function choice and RF-based classifier offers 94.25 % of accuracy and 0.95 AUC for the K10 protocol.	Easily identifies trends and patterns	ML can consider taking time and effort to achieve outputs
[5]	Decision Tree Gaussian NB , SVM ,Random Forest ,Extra Trees AdaBoost ,MLP ,LR , Gradient Boost Classifier , KNN ,Bagging	Logistic regression offers 96 % precision. The pipeline application gave the Ada Boost classifier 98.8 % as the best model.	Helps in Identifying Diseases and Diagnosis	In order to assess the usefulness of ml strategies it is therefore essential to analyse the data.
[6]	SVM, Naïve Bayes, KNN, C4.5	According to the investigation, and discussion the C4.5 decision tree outperforms those certain diabetes input classifier model by 73.5 %.	Helps in Identifying Diseases and Diagnosis	Over Fitting of the Training Samples
[7]	Naïve Bayes, SVM, RF, simple CART	On the highest, which is 0.7913, is according to the Classification Accuracy of SVM. SVM's overall success in predicting diabetes is higher than that of NB, Random Forest classifier and Simple CART.	Improving quality of life	Interpretation of the findings
[8]	Auto MLP	Auto MLP Outlier Identification has an accuracy of 88.7%, a Mean Score Recall of 88.5, and a Weighted Mean Precision of 85.8%.	Improving efficiency and quality for care.	It necessitates a considerable variety of data.
[9]	SVM, Logistic regression, ANN	SVM is an effective approach for binary supervised learning, so we chose SVM to predict diabetes.	Fast Processing and Real-Time Predictions	When the sample data contains more distortion, such as conflicting target classes, SVM classifier does not accomplish efficiently.

[10]	LR, KNN, SVM,NB, Decision Tree and RF classifiers were used	Hence the precision of used dataset's Random Forest is 94.10 % which is the best among the others. For the PIMA dataset, Random Forest often has the best precision. Learning algorithms applied to six different classifiers	Machine Learning Improves Over Time	Utilizing the KNN classifier The consistency of the data is determined by its efficiency.
[11]	Decision tree classifier, ANN, NBand SVM model.	The models with the good success are Decision Tree of 85% accuracy, Naïve Bayes obtains of 77%, and SVM had given 77.3%precision.	Contingency ratings that are reliable and appropriate allow reliable as well as consistent resource distribution, resulting in high levels of efficiency results.	We can resolve the difficulties of exhibiting the challenge to the network by using ANN.
[12]	Rule-based strategies Neural network	Rule-based method gives 88.6% of accuracy and Neural network gives 88.5% precision	No Human Intervention Needed (Automation)	It's complex and time-consuming to enumerate any of the rules.
[13]	SVM, Decision Trees , Ada boost , Linear regression	After Bootstrapping (Accuracy Rates %) comparing all classifiers the SVM and Ada boost gives highest precision of 94.44%	machine learning enables computers to access hidden insights	Linear regression classifier only considers the dependent variable's mean value.
[14]	DT, Logistic Regression classifier, NB, SVM and KNN algorithm	With the aid of KFold and Cross Validation, the final outcome gave us an accuracy of 81.1%.	Wide Applications are used here	The concept of linearity here across variables of the study is a significant drawback of LR classifier.
[15]	SVM, Naïve Bayes, Light GBM	SVM has the highest accuracy rate, with an accuracy rate of 96.54%. This illustrates that the most effective diabetes classification algorithm SVM is a forecast.	With a strong margin of differentiation, SVM fits pretty well. In high-dimensional spaces, it is efficient.	The presumption of autonomous predictive properties is one of NB classifier key flaws.
[16]	Cnn model, Transfer leaning model of VGG-16 and VGG-19.	For the classification of DR into 5 ranges scale from 0 to 4, where 0 is normal DR and 4 is proliferative DR, This test has an AUC of 0.904 and produces 80% accuracy, 72 % accuracy, and 82 % precision.	Efficient in the processing of greater outcomes. Sickness Diagnosis using medical image processing	Negative friction caused where the current system's output or consistency suffers as a result including its image classification of transfer learning.
[17]	DenseNet, CNN VGG16 architecture and DenseNet121	After the activation function, the outcome. The design gave the DR detection a precision of 0.96111 and QWK score of 0. 8981. Also we will compare two CNN model at the end.	Best Results with Unstructured Data. Through medical image processing, practitioners see a good view of what is happening on throughout the patient's body.	Images in diverse situations are graded.
[18]	CNN, Quadratic weighted kappa metric, openCV	On a QWK score of the metric, the single accuracy of the model of the CNN described in this work is 0.386 and the assembly of 3 such related models culminated in a score of 0.3996.	Robustness is taught instantly against normal variations in the results. Doctors will reliably estimate the risk that you will contract a disease by medical imaging.	Deep learning necessitates the use of pricey GPUs and a large number of connections.
[19]	5 deep CNN transfer learning models are Resnet50, Inceptionv3, Xception, Dense121, and Dense169	On the given repository the experimental findings showed that, including current models, the suggested model senses all stages of DR and outperforms state-of-the-art approaches.	Almost the exact cnn methodology is used for a variety of applications and database schema. Rapid and affordable cost of processing using medical image processing	Over fitting problem can arise in transfer learning
[20]	CNN	The system described could be used as an early diagnosis	The deep learning architecture is scalable for	The location and direction of an object are

		screening method for diabetic retinopathy with a specificity and sensitivity precision with 0.99 and QWK score of 0.925466	potential adaptation to new issues. To take and process digital images, no processing or fixing chemicals are required.	not encoded by CNN.
[21]	SVM Textures features (LPB, LESH, LTP)	The experimental findings indicate that LESH is the highest performing technique using SVM-RBF kernel with an precision of 0.904.	Deep learning in diagnostic systems has made remarkable strides. Digital image processing made digital image can be noise free.	Deep networks demand amazingly huge data sources to reach great efficiency.
[22]	CNN	They evaluated the model and received a QWK score of 0.851 and an AUC score of 0.844, indicating that it performs well in intensity scoring. They were possible to diagnose with a sensitivity of 98 % and an accuracy of over 94 %.	Receiving adequate tools for streamlining and simplifying detailed data analysis	Inability to be directionally irreducible with respect to the information of input data
[23]	CNN	Hence the visualization of heat map model displays the detection of diabetic retinopathy	Understand the correlation between outcomes and performance. Visual content happens to be significantly easier for the neural network to perceive than visual data.	Deep Learning is overly susceptible to picture shifts that might be apparent to a reasonable eye.
[24]	HSV transform algorithm and histogram equalization tactics, CNN model	Efficiency was 97 %sensitivity was 96.67 %, accuracy was 93.33 %, accuracy was 97.78 %, and performance was 93.33 % in this data analysis.	Images are acquired more easily and at a reduced rate. Increase the accuracy of diagnosis	Lack of insight of the input data's precise reflects rule
[25]	SVM, RF, Gradient boost, AdaBoost, Gaussian NB classifiers	SVM has a high precision of 88.71 %, 83.34 % accuracy of random forest methodology, 83.34 % precision of gradient algorithm, The AdaBoost and Gaussian NB approaches, on the other hand, attained 54.3 %and 37.09 % accuracy respectively.	Image correction and retouching have been simpler in image processing. Best Results with Unstructured Data	AI technology, unlike human beings, can indeed be enhanced by learning.
[26]	Linear SVM CNN	Experimental findings indicate high sensitivity and accuracy obtained by their suggested model while detecting DR.	It is possible to smooth out images. Images can be categorised instantaneously based on the content they have.	The most significant issue in AI is data collection and processing
[27]	Grey Wolf Optimization (GWO), PCA, DNN, SVM, Naive Bayes Classifier, Decision Tree and XGBoost.	The experimental results indicate that the DNN-PCAGWO algorithm proposed outperformed the algorithms described earlier.	It is possible to increase or decrease visual features of images.	Prior to PCA, data normalization is required.
[28]	ADL- Active deep learning ADL-CNN	The Active deep learning CNN model obtained an avg SE of 92.20 %, 95.10 % , F-measure and 98% of Precision	It's used to analyse diagnostic images.	Memorization is often needed.

[29]	Transfer learning based CNN	On the validation phase, the four phases of DR got a Cohen Kappa value of 0.8836, and on the test data, 0.9809. Our model obtained instruction in four different classes.	For quicker file delivery across the network, files may be compressed and decompressed.	Transfer learning only succeeds when both replicas' original and goal concerns are sufficiently close.
[30]	SVM and Naïve Bayes classifiers were used.	It could conserve resources and time by doing so to achieve more reliable results by using these algorithms to anticipate diabetes.	This data analysis methodology is used by many hospitals to determine admissions patterns.	It requires more time to process.
[31]	Machine Learning Classifiers, SVM	The suggested system is to classify the heart condition for which we implement the prediction system based on diabetes and separate the heart disease using SVM measurement.	Unlimited Resources Incorporating Information. With timely analysis and evaluation, machine learning can effectively ingest infinite volumes of data.	Automation is a con
[32]	ML model, Deep learning algorithm, Medical Image processing	CAD system is used for diagnosing of DR	It is used to analyse medical images Medical diagnosis Image Recognition.	When it comes to bringing existing models into practice, ML remains a challenge.
[33]	Gradient Boosting, Logistic Regression and Naive Bayes	The obtained accuracy for the model are86 % for Gradient Boosting, 79 % for Logistic Regression and 77 % for NB the diagnosis of diabetes.	No human activity was required (automation) Multi-dimensional and multi-variety information handling.	It generally requires innovation, exploration, and perseverance.
[34]	Naïve Bayes algorithm , Decision tree classifier, RF and Adaboost	The ML-based system's overall precision is 90.62 % . For the K10 protocol, The integration of LR based selecting features and an RF based classification model yields a reliability of 0.95 and an accuracy of 94.25 %.	Continuous development, continuous improvement in precision and efficacy	Hence the possibility of riskon mass underemployment though they are trying to replace humankind in certain areas.

Table 2.1 Title, classifiers used, Performance Metrics, Merits of an existing system

2. 3 Data Sources used are:

The Datasets are used from public and open-source platforms like kaggle, etc.

S.no	Tends to require Database (Dataset)
1.	Pima Indian diabetes Database
2.	Diabetic Retinopathy Detection from kaggle
3.	Aptos2019-blindness-detection
4.	Diabetic Retinopathy Detection Competition Dataset Resized/Cropped

Table 2.2 Data Sources used

2.4 Review on Findings (Summary):

The focus of the survey is to compare and assess the optimum tools and techniques and advanced features focused on Prediction of diabetes diagnosis detection based on machine learning tactics and algorithm and detection of diabetic retinopathy using AI. The overview of detection of diabetes diagnosis and diabetic retinopathy is shown in below fig 2.1 and 2.2.

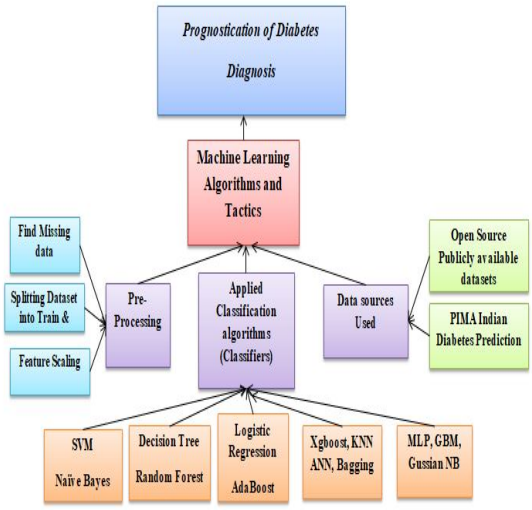

Figure 2.1: Overview of Prognostication of diabetes diagnosis

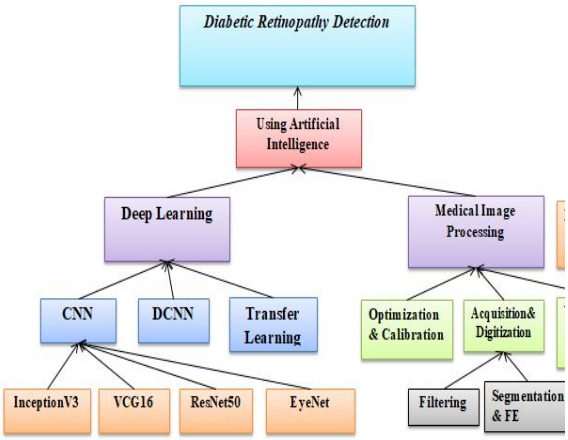

Figure 2.2: Overview of Diabetic Retinopathy Detection

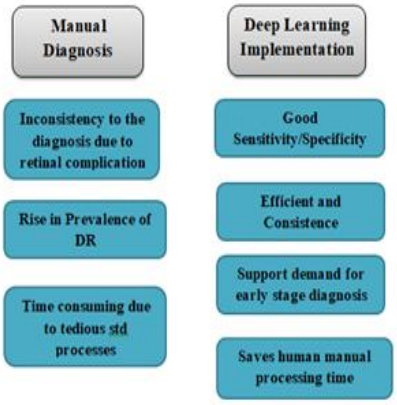

Figure 2.3: Comparison of Manual diagnosis and AI technologies

3. APPLICATIONS USED

Identifying Diseases and Diagnosis	Smart Health Records	Predict chronic disease
Personalized Treatment & Behavioural Modification	Clinical Trial Research	Medical Diagnosis

Table 3.1: Applications

4. CONCLUSION

The Study on Prognostication of diabetes diagnosis and diabetic retinopathy detection is still evolving, due to the difficulty of disease diagnosis and detection synopsis and modelling. Different scholars and researchers are working hard to find an advanced method, using multiple techniques. The results from this study will support the upcoming researchers which will give an idea of the various techniques of diabetes diagnosis and diabetic retinopathy These has been accomplished by various pre-processing, selection of features, Transfer learning , classification as classifiers, machine learning and deep learning techniques and exceptional precision in detection of diabetes diagnosis and diabetic retinopathy.

REFERENCES:

1. Sisodia, D. and Sisodia, D.S., 2018. **Prediction of diabetes using classification algorithms**. Procedia computer science, 132,pp.1578-1585.
2. Mahmud, S.H., Hossin, M.A., Ahmed, M.R., Noori, S.R.H. and Sarkar, M.N.I., 2018, August**. Machine Learning Based Unified Framework for Diabetes Prediction.** In Proceedings of the 2018 International Conference on Big Data Engineering and Technology (pp. 46-50).
3. Saru, S. and Subashree, S., Analysis and **Prediction of Diabetes Using Machine Learning** (April 2, 2019). International Journal of Emerging Technology and Innovative Engineering, Volume 5, Issue 4, April 2019, Available at SSRN: https://ssrn.com/abstract=3368308
4. Maniruzzaman, M., Rahman, M.J., Ahammed, B. and Abedin, M.M., 2020. **Classification and prediction of diabetes disease using machine learning paradigm.** Health Information Science and Systems, 8(1), p.7.
5. Mujumdar, A. and Vaidehi, V., 2019. **Diabetes prediction using machine learning algorithms.** Procedia Computer Science, 165, pp.292-299.
6. M. F. Faruque, Asaduzzaman and I. H. Sarker, "**Performance Analysis of Machine Learning Techniques to Predict Diabetes Mellitus**," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox'sBazar, Bangladesh, 2019, pp. 1-4, doi: 10.1109/ECACE.2019.8679365.
7. A. Mir and S. N. Dhage, "**Diabetes Disease Prediction Using Machine Learning on Big Data of Healthcare**," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-6, doi: 10.1109/ICCUBEA.2018.8697439.
8. M. Jahangir, H. Afzal, M. Ahmed, K. Khurshid and R. Nawaz, "**An expert system for diabetes prediction using auto tuned multi-layer perceptron**," 2017 Intelligent Systems Conference (IntelliSys), London, 2017, pp. 722-728, doi: 10.1109/IntelliSys.2017.8324209.

9. Joshi, T.N. and Chawan, P.P.M., 2018. **Diabetes Prediction Using Machine Learning Techniques**. Ijera, 8(1), pp.9-13.
10. Tigga, N.P. and Garg, S., 2020. **Prediction of Type 2 Diabetes using Machine Learning Classification Methods**. Procedia Computer Science, 167, pp.706-716.
11. P. Sonar and K. JayaMalini, "**Diabetes Prediction Using Different Machine Learning Approaches**," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 367-371, doi: 10.1109/ICCMC.2019.8819841
12. Singh, P.P., Prasad, S., Das, B., Poddar, U. and Choudhury, D.R., 2018. **Classification of diabetic patient data using machine learning techniques.** In Ambient Communications and Computer Systems (pp. 427-436). Springer, Singapore.
13. Aada, A. and Tiwari, S., 2019. **Predicting diabetes in medical datasets using machine learning techniques**. Int. J. Sci. Eng. Res, 5(2).
14. Agarwal and A. Saxena, "**Analysis of Machine Learning Algorithms and Obtaining Highest Accuracy for Prediction of Diabetes in Women**," 2019 6th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2019, pp. 686-690.
15. Xue, J., Min, F. and Ma, F., 2020, November. **Research on Diabetes Prediction Method Based on Machine Learning.** In Journal of Physics: Conference Series (Vol. 1684, No. 1, p. 012062). IOP Publishing.
16. Nguyen, Q.H., Muthuraman, R., Singh, L., Sen, G., Tran, A.C., Nguyen, B.P. and Chua, M., 2020, January. **Diabetic retinopathy detection using deep learning.** In Proceedings of the 4th International Conference on Machine Learning and Soft Computing (pp. 103-107).
17. Mishra, S., Hanchate, S. and Saquib, Z., 2020, October. **Diabetic Retinopathy Detection using Deep Learning**. In 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE) (pp. 515-520). IEEE.
18. Doshi, D., Shenoy, A., Sidhpura, D. and Gharpure, P., 2016, December. **Diabetic retinopathy detection using deep convolutional neural networks**. In 2016 International Conference on Computing, Analytics and Security Trends (CAST) (pp. 261-266). IEEE.
19. "**A Deep learning Ensemble approach for Diabetic Retinopathy Detection**," in IEEE Access, vol. 7, pp. 150530-150539, 2019, doi: 10.1109/ACCESS.2019.2947484.
20. Tymchenko, B., Marchenko, P. and Spodarets, D., 2020. **Deep learning approach to diabetic retinopathy detection**. arXiv preprint arXiv:2003.02261.
21. M. Chetoui, M. A. Akhloufi and M. Kardouchi, "**Diabetic Retinopathy Detection Using Machine Learning and Texture Features**," 2018 IEEE Canadian Conference on Electrical & Computer Engineering (CCECE), Quebec City, QC, 2018, pp. 1-4, doi: 10.1109/CCECE.2018.8447809.
22. Islam, S.M.S., Hasan, M.M. and Abdullah, S., 2018. **Deep learning based early detection and grading of diabetic retinopathy using retinal fundus images**. arXiv preprint arXiv:1812.10595.
23. Keel, S., Wu, J., Lee, P.Y., Scheetz, J. and He, M., 2019. **Visualizing deep learning models for the detection of referable diabetic retinopathy and glaucoma**. JAMA ophthalmology, 137(3), pp.288-292.
24. Deperlıoğlu, Ö. and Köse, U., 2018, October. **Diagnosis of diabetic retinopathy by using image processing and convolutional neural network**. In 2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) (pp. 1-5). IEEE.
25. J. D. Labhade, L. K. Chouthmol and S. Deshmukh, "**Diabetic retinopathy detection using soft computing techniques**," 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), Pune, 2016, pp. 175-178, doi: 10.1109/ICACDOT.2016.7877573.
26. Seth, S. and Agarwal, B., 2018. **A hybrid deep learning model for detecting diabetic retinopathy**. Journal of Statistics and Management Systems, 21(4), pp.569-574.
27. Gadekallu, T.R., Khare, N., Bhattacharya, S. et al. **Deep neural networks to predict diabetic retinopathy**. J Ambient Intell Human Comput (2020).https://doi.org/10.1007/s12652-020-01963-7
28. Qureshi, I., Ma, J. & Abbas, Q. **Diabetic retinopathy detection and stage classification in eye fundus images using active deep learning**. Multimed Tools Appl (2021).

https://doi.org/10.1007/s11042-020-10238-4

29. Samanta, A., Saha, A., Satapathy, S.C., Fernandes, S.L. and Zhang, Y.D., 2020. **Automated detection of diabetic retinopathy using convolutional neural networks on a small dataset. Pattern Recognition Letters**, 135, pp.293-298.
30. Amulya, K.J., Divya, S., Deepali, H.V. and Ravikumar, V., 2021. A Survey on **Diabetes Prediction Using Machine Learning**. In ICCCE 2020 (pp. 1049-1057). Springer, Singapore.
31. R. A. Canessane, R. Dhanalakshmi, B. Muthukumar, C. V. Sailaja, B. Jyothi and S. Dhamodaran, "**Diabetes To Heart Disease: A Survey**," 2019 Fifth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), Chennai, India, 2019, pp. 69-74, doi: 10.1109/ICONSTEM.2019.8918854.
32. Bilal, A., Sun, G. and Mazhar, S., 2021. **Survey on recent developments in automatic detection of diabetic retinopathy**. Journal Françaisd'Ophtalmologie.
33. Birjais, R., Mourya, A.K., Chauhan, R. and Kaur, H., 2019. **Prediction and diagnosis of future diabetes risk: a machine learning approach**. SN Applied Sciences, 1(9), pp.1-8.
34. Maniruzzaman, M., Rahman, M.J., Ahammed, B. et al. **Classification and prediction of diabetes disease using machine learning paradigm**. Health InfSciSyst 8, 7 (2020). https://doi.org/10.1007/s13755-019-0095-z