

HEALTHCARE EFFICIENT DIABETES DISEASE PREDICTION USING SUPERVISED MACHINE LEARNING :A REVIEW

Saurabh Kumar Pandey
M.Tech Schollar
Saurabhkumarpandey70913@gmail.com
Mittal Institute of Technology, Bhopal

Prof. Preeti Mishra
Assistant Professor
pritypreet85@gmail.com
Mittal Institute of Technology, Bhopal

ABSTRACT :- The classification of objects into the right classes has long been one of the most crucial objectives of machine learning or deep learning. Due to the similarities between various things, their textures, colors, and other physical properties, object recognition still presents significant challenges despite the importance of categorizing specific groups of images. In computer vision, object detection has a wide range of uses, such as face and vehicle detection, video surveillance, and plant leaf detection. When conducting studies on flowers, using flowers as medicine, analyzing floral patents, etc., automatic classification of flowers is crucial. Traditionally, low-level characteristics like color, shape, texture, and geometry are used to classify flowers. The feature description has a significant impact on how accurately and robustly flowers are classified. In recent years, deep features have demonstrated good performance on high-resolution photos, but they are unable to extract precise global features from low-resolution images. Deep neural networks have been widely used in computer vision applications because they are effective at identifying picture patterns. An advanced deep-learning model that can precisely calculate the similarity of floral photos is required in the field of flower image analysis. In comparison to the given model mentioned in the exhibit results, the proposed model performs adequately. The proposed network can still be improved in terms of learning parameters, validation accuracy, loss, and training time. Fine-tuned deep learning models for similarity computation on flower datasets have a bright and broad future ahead of them, with lots of room for development and use. In this research work, the future potential of optimized deep learning models for the calculation of similarity on floral datasets is promising for growth and innovation. These models have the power to fundamentally alter the way we see, value, and engage with the world of flowers. The proposed model is supported by augmentation so the distance method is used against two augmentation views of the trained model being calculated.

Keywords: *Machine Learning, Supervised Learning, Diabetes Mellitus, Health System, Gradient Boosting, Python, Colab Platform, Recall, Accuracy*

1.INTRODUCTION

In several countries including United States, China, UK healthcare is a thriving domain that has industrialized as a leading sector related to economic growth and employment as well as functional overheads. From the study of the global spending on healthcare, it is anticipated that the general healthcare expenditure will reach double in the following 5 years [1, 2]. In India, public health spending is expected to increase to ₹486000 crores by 2022 from ₹267000 crores in the year 2018 (Ministry of Health & Welfare 2017). The overall financial outlay to health and wellbeing is expected to rise 45% from the financial year 2018 to 2022. Non-value added and unproductive activities including diagnostic and medication errors, improper usage of antibiotics, readmissions, and deception create a substantial amount of spending for care. Approximately, 5.2 million Indians demise annually due to clinical faults and the consequences of malicious activities [3].

A suitable Clinical Decision Support System (CDSS) can handle these problems and accelerate the transformation towards a value based clinical system to deliver passable care and high services [4]. Digital therapeutics has been

recognized to be a renowned intervention for treatment in disease control, and both medical professionals and patients are benefitting from CDSS. At present, the healthcare industry is adopting information and communication technologies for its organizational venture to deliver high quality care services in both commercial and technical aspects [5].

DM has developed as a long term deadly disease across the world and especially in developing nations. Several worldwide and nationwide epidemiological studies have observed the increasing number of diabetics across the globe. Recently, a study exposed that around 463 million individuals between the ages of 20 and 79 years are diabetic patients and it is projected that 552 million people are expected to get affected by the year 2030 [6]. Besides, DM is considered the 5th leading cause of disease oriented demises and is claimed to cause one death in every 6 seconds globally [7]. Unfortunately, 46.5% of DM patients have not been identified.

Worldwide around 422 million people suffers from diabetes. Diabetes mellitus can be understood to dissimilar types, first category being Type 1 diabetes and the other

one being Type 2 diabetes [5, 6]. For case of about all the diabetes cases about 5 to 10% of cases have only the Diabetes “Type1”. The Diabetes “Type 2” is about remaining 90% from all the aroused cases. The type 1 diabetes occurs to often kids or adults or during their adolescence days which is caused by the partial dis functioning of the Pancreas. It will not show any symptoms at the beginning stage because the Pancreas will be functioning partially [7]. This type 1 diabetes gets critical only when 80 90% of Insulin creating Pancreatic cells devastated. In insulin

dependent diabetes which is resultant of very known chronic Hyper glycemias and the body is unable to regulate its own sugar, level which will lead to the shooting of sugar levels in the blood. This type 2 diabetes happens mostly to grown up people and affect more heavy and obese adults. When conducting the quantitative research, the diagnosis of the diabetes mellitus is the difficult part because the terms like the A1C, WBC (white blood cell) count, fibrinogen and parameters such as the hematological indices are ineffective because of some shortcomings [8].

1.2 Diabetes Mellitus Occurs

Digestion is a process which requires food to be broken down into various forms of energy or sources of nutrients. Carbohydrates being taken by our body get converted into glucose by our body. Now, we have glucose present in the body which needs insulin to reach to its final destination i.e. the cells within the body so that which helps in building tissues and muscles. Pancreas which are present behind our stomach are responsible for making insulin. When the insulin is released into the blood, glucose enters into the blood by the body cells. If someone, has diabetes then this may be possible due to the various reasons listed below:

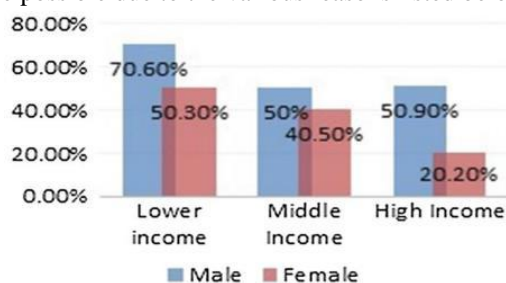


Figure 1.1: Comparison of death rate of diabetes among different classes of people

Pancreases are unable to form the enough amount of insulin required. When the pancreas forms the required amount of insulin, but the beta cells present in the blood does not retaliate to the production thus leaving it in the blood. Thus, Glucose remains unabsorbed thus allowing it to rise beyond levels [10]. Type 1 and 2 are the chronic diabetes problems since it can't be reversed, but gestational and pre diabetic condition can be reversed. Many health complications may arise due to the diabetes. It is also known as silent killer. In

the pandemic times also, people with the diabetes were the most affected ones [11].

1.3 ML Technique: Future for Diabetes Care

There has been a colossal impact in the healthcare sector with the proliferation of cutting-edge technologies including AI, DM, ML etc. These techniques have demonstrated their effectiveness in disease prediction from massive amounts of healthcare data [11]. They are used to perform prognostic modeling, that is, the exploitation of data and statistics to calculate imminent results from past experiences.

Efficient ML approaches have been employed to develop algorithms in order to support frameworks in diagnosing diabetes in early-stage or its subsequent complications. ML enables a constant and problem-free system for monitoring biomarkers and symptoms. Several ML approaches have been proposed, including unsupervised, supervised, and reinforcement learning approaches. This is obviously practical because ML techniques are driven by data. Since the huge volume of big disease-related data fed into the diagnosis system, ML can save substantial human effort. In the ML approach, models are trained on this data and deliver the more accurate output according to the input data. Some might analyze features of clinical scans, whereas others use blood test data collected from patients. Because there are several signs of the disease, the parameters differ accordingly. With several established techniques, investigators have investigated different algorithms and tweaked many hyper parameters to achieve outcomes that seem most appropriate for real time applications.

Many already existing methods are not so effective in diagnosis or prediction. Enough data collected by various healthcare organizations for the decision making for diabetes mellitus. Several new models are used in early diabetes prediction nowadays such as the classifiers of machine learning systems, In order to get the biomedical statistics various data mining methods used for prediction of the disease [12]. The process of analyzing and collecting various databases and extracting important chunk of data which is used for the effective analysis is known as the Data mining technique. In the diabetic field, predictive analytics method is being used for diagnosis, prediction, self management, and prevention being done. In the recent trend, the research analysis helps in the diabetes prediction that plays a vigorous role in the case of High mortality and high morbidity, disease compilation and prevention [11].

2. LITERATURE REVIEW

Isfaffuzaman Tasin et al. [1], in diagnosis or detection of temperature variations in the human skin are obtained from good resolution thermal images captured by the highly sensitive thermal imaging system. Historically, the first thermal image of human was acquired with a British prototype system; the “Pyroscan” were made in London for

research in rheumatology, chronic fatigue, and pain management between 1959 and 1961. The thermal imaging system has been utilized for many clinical applications in hospitals since the late 20th century. The thermal imaging system has been used for mass fever detection localization of swelling in dentistry biometric authentication of facial signature and palm vein. Recently, it has also been utilized to record the presence and position of the fetus and other physiological factors associated with pregnancy. This technique is effectively utilized for screening of temperature variations in minimally symptomatic patients, during the outbreak of the coronavirus disease –19.

Olisah et al. [2], is characterized in the form of intensity differences on a thermal image. The difference in the thermal pattern of pixel intensities results in texture variations in an image. These texture variations can be extracted in the form of statistical features and used for classification. As the temperature distribution of DM patients is different from healthy subjects, asymmetry amongst the features may be analyzed to gather significant information about the pathology. The human examination of plantar thermal images is susceptible to faults owing to carelessness, color blindness, fatigue, and repetitive tasks. The utility of the textural and higher order statistical features in finding the temperature distribution changes in plantar thermal images has been determined. These features have been utilized to differentiate diabetic feet from healthy ones.

Deberneh et al. [3], in this work, many feature extraction approaches are explored to design and develop efficient machine learning frameworks intended for detecting diabetic feet in plantar thermal images. In extension to the broadly utilized asymmetry method using textural and temperature features, the subsequent feature extraction approaches are also analyzed, for plantar thermal images. When the performance of several extracted features could not be used seemingly, machine learning methods have been utilized for training and automated classification. Many CNN models are employed to automatically segment, classify diabetic foot ulcers from color (RGB) images. The most significant features are also extracted from different layers of the pre-trained CNN models using transfer learning. The process of analysing and collecting various databases and extracting important chunk of data which is used for the effective analysis is known as the Data mining technique. In the diabetic field, predictive analytics method is being used for diagnosis, prediction, self-management, and prevention being done. In the recent trend, the research analysis helps in the diabetes prediction that plays a vigorous role in the case of High mortality and high morbidity, disease compilation and prevention.

Nikos Fazakis et al. [4], the dataset contained nine variables, eight of which contained patient information.

The ninth variable was the class that predicted whether patients would develop diabetes or not. Outliers and missing values make up the dataset. We have removed the outliers from the dataset using our proposed method. The mean filter method was used to impute missing values that were present in the dataset, preserving the dataset's consistency. Weka was used in each and every one of the experiments. It is used for "Data Mining" and classification purpose. Classification algorithms may be applied to a dataset. The other applications of weka are visualizing, regression, purpose of data mining methods. Full form of Weka is Waikato Environment for knowledge Analysis.

Naveen Kishore G et al. [5], One of the most fatal and long-lasting diseases that causes an increase in glucose is diabetes. If diabetes is left undiagnosed and untreated, many problems arise. However, the advancement of AI approaches addresses this primary problem. Estimates of exactness are made based on correctly and incorrectly ordered cases. With a precision of 76.30 percent, the obtained results demonstrate that Naive Bayes prevails over relatively different calculations. R.O.C. bends are used in a legitimate and methodical way to verify these results. For the diagnosis of the diabetes these indices are necessary for research papers these indices are used. In-order to increase the A1C several treatment methods are used such as the ingestion of liquor, narcotics and salicylates. A1C can be increased by the process of vitamin-c ingestion but when analysed by the method of chromatography the count of A1C remains the same the increased output of A1C count can be obtained only through the electrophoresis analysis.

Chatrati et al. [6], there are a couple of simulated intelligence methodologies that are used to carry out judicious examination over colossal data in various fields. Perceptive assessment in clinical benefits is a troublesome endeavor regardless can assist experts with making gigantic data taught ideal decisions about open minded's prosperity. Perceptive examination in clinical consideration, man-made intelligence estimations are used in this assessment work. Six distinct AI calculations are applied to a dataset of the patient's clinical record for the purpose of analysis. The applied calculations' execution and precision are examined and analyzed. The best calculation for predicting diabetes is determined by correlating the various AI methods used in this study. One of the leading causes of disability and death is type II diabetes. These are incomplete and noisy data. A comprehensive estimation of missing data is presented in this paper. Correlation-based estimation models and k-NN were two of the five approaches to imputation of missing values that were compared.

Hasan et al. [7], have proposes the affirmation of Indian correspondence through marking movements using a mind blowing man-made awareness gadget, convolutional brain associations (CNN). The catch technique used in this work

is selfie mode continuous gesture-based communication video, allowing a conference-weakened individual to freely use the S.L.R. versatile application. CNN preparation is carried out using three distinct example estimates, each containing distinct subject and survey point arrangements. The prepared CNN is tested using the remaining two examples. With our selfie-based communication data, distinct CNN models were planned and tested for improved acknowledgment precision. In contrast to other classifier models written about the same dataset, our acknowledgment rate was 92.88%.

Cervantes and others [8], SVMs and how they can be utilized in various logical fields certainly stand out lately. SVMs are one of the most effective and long-lasting classification and regression algorithms for a wide range of applications. Design acknowledgment, which is a famous and dynamic area of exploration, has benefited incredibly from the SVM's commitments. Moreover, SVM has been integrated into other state of the art draws near, for example, develop calculations, to further develop arrangement capacities and advance boundaries. SVM algorithms are now utilized extensively in numerous engineering and scientific fields. SVMs are momentarily presented, various applications are talked about, and difficulties and patterns are summed up in this paper. The limitations of SVMs will also be determined. In addition to the SVMs' future, we will talk about additional applications. SVMs' applications, especially in certain fields, will likewise be analyzed.

Pranto et al. [9], previously unseen portion of PIMA. The study's goal is to demonstrate how well a number of classifiers tested on patients from another country and trained on a diabetes dataset from one nation perform. In this review, we saw choice trees, K-closest neighbors, irregular backwoods, and gullible bayes classifiers. When the levels of sugar remain prolonged in blood over a larger period metabolic disorder occurs which is acknowledged as diabetes. It is a cluster or collection of diseases that may get affected by the high amount of glucose present in the body. The most vital source of energy required by our body is glucose because it helps our body in building muscles, tissues, and cells of, the body.

Mohan et al. [10], it is feasible to control such an illness and save survives early forecast. Utilizing AI techniques, this concentrate principally examines an assortment of elements to accomplish its goal. Demonstrative clinical datasets accumulated from diabetic patients, AI methods give productive outcomes to separating information. In case of the diagnosis of diabetes a single parameter or indices will not be very effective which may be an error factor in arriving at a decision. Various parameters are combined to get the early prediction or diagnosis of the diabetes. Learning from such data can help predict diabetic patients.

On grown-up populace information, we utilize four notable AI calculations and other AI strategies regarding precision. When the glucose is left unabsorbed in the body and the cells are unable to utilize it, then diabetes happens due to the existence of extra levels of blood sugar. When our 14 pancreas does not absorb the excessive amount of sugar in the blood, then diabetes mellitus occurs.

Azeem Muhammad Sarwar et al. [11], prescient examination over a lot of information are a typical use of an assortment of AI strategies in various fields. Although it is a challenging task, predictive analytics in healthcare, this study's different AI strategies are contrasted with figure out which calculation is generally appropriate for diabetes expectation. The purpose of this paper is to assist practitioners and doctors in making early diabetes predictions using machine learning techniques. Firstly, if the Insulin is not made by the body at an enough amount, secondly, the insulin is made by the body but the body will not respond to the insulin properly only in the above two condition hyper-glycemia occurs. Insulin is an important need of blood so that the glucose can be converted as energy by the body. If the utilization of glucose is not being done to produce energy then the Hyper-glycemia can occur. When the above condition exists in a human body which will results in a serious of health implications like diabetic Keto-acidosis, stroke related paralysis disease, cardiology related heart diseases like cardiovascular disease, nonketotic hyperosmolar etc.

G. A. Rao et al. [12] it is thought to be a challenging problem in computer vision to extract. This paper proposes convolutional neural networks (CNN), a powerful artificial intelligence method, for recognizing Indian sign language gestures. In a video, each sign involved 60 casings or pictures. The prepared CNN is tested using the remaining two examples. With our selfie communication through signing information, different CNN structures were planned and tried for further developed acknowledgment precision. Problems in risk management, portfolio construction, and securities pricing are all examples of financial prediction problems that involve large data sets and intricate data interactions that are presently troublesome or difficult to determine in a full financial model. DL methods, when applied to these problems, may yield outcomes that are more useful than those produced by standard finance techniques. Particularly, deep learning is able to identify and take advantage of data interactions that, at least for the time being, are invisible to any existing financial economic theory.

Quan Zou et al. [13], hyperglycemia is a characteristic feature of diabetes mellitus, a chronic condition. It might result in numerous problems. One in ten adults in the future will have diabetes, as predicted by rising morbidity in recent years, when the number of diabetics worldwide

reaches 642 million in 2040. This alarming figure absolutely requires significant attention. With the quick advancement of AI, AI has been applied to numerous parts of clinical wellbeing. We conducted independent test experiments on a few of the methods that performed the best in order to confirm their universal applicability. As a training set, we selected the data of 68994 diabetic patients and healthy individuals at random. We extracted the data five times at random due to the imbalance in the data. Additionally, the outcome is the sum of these five experiments. PCA and mRMR were utilized in this study to reduce the dimensionality.

3. CONCLUSION AND FUTURE WORK

3.1 Conclusion

One significant health challenge worldwide is Diabetes detection early and proactively. The present study aims to establish a suitable prediction model that depends on a Machine Learning scheme for predicting Diabetes. Diabetes is a metabolic condition brought on by elevated levels of glucose in the blood. Patients with diabetes don't have enough insulin in their bodies to control their sugar levels. Additionally, diabetes contributes to a number of other dangerous diseases. Due to the fact that diabetes is the root cause of numerous diseases in the human body, it is essential to detect this serious condition as soon as possible.

There have been a number of approaches taken in the past to make it easier for doctors to diagnose diabetes. However, there is a classification issue with it: a selection of an excessive number of samples, but it does not improve classification accuracy. The speed of the algorithm is better in a few situations, but the accuracy of the data classification is lower. A selected classification algorithm is used to test the diabetic data set. We applied a gradient-boosting machine learning algorithm to the Diabetes dataset's attributes in order to get the best results.

The diabetes dataset of 1145 Pima Indians: The test uses 330 diabetic and 815 non-diabetic participants. An ensemble of gradient boosting was used in the proposed algorithm to achieve an accuracy of 91.23%. As can be seen, the majority vote-based model employs NB, DT, and SVM classifiers, and its accuracy for the diabetes disease dataset is 73.42%, 80.76%, and 82.51%, respectively. Subsequently, the Inclination helping calculation gives the best exactness to Diabetes finding than the past calculation.

REFERENCES

[1] Isfuzzaman Tasin, Tansin Ullah Nabil, Sanjida Islam, Riasat Khan, "Diabetes prediction using machine learning and explainable AI techniques", *Healthcare Technology Letters*, pp. 01-10, Wiley 2022.

[2] Olisah, C.C., Smith, L., Smith, M., "Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective", *Comput. Methods Programs Biomed.*, Vol. 20, pp. 1–12, 2022.

[3] Deberneh, H.M., Kim, I., "Prediction of type 2 diabetes based on machine learning algorithm", *Int. J. Environ. Res. Public Health*, Vol. 18, pp. 1–14, 2021.

[4] Nikos Fazakis, Otilia Kocsis, Elias Dritsas, Sotiris Alexiou, Nikos Fakotakis, and Konstantinos Moustakas, "Machine Learning Tools for Long-Term Type 2 Diabetes Risk Prediction", *IEEE Access* 2021.

[5] Naveen Kishore G, V.Rajesh, A.Vamsi Akki Reddy, K.Sumedh, T.Rajesh Sai Reddy, "Prediction of Diabetes using Machine Learning Classification Algorithms", *International Journal of Scientific & Technology Research*, Vol. 9, No. 01, 2020.

[6] Chatrati, S.P., Hossain, G., Goyal, A., "Smart home health monitoring system for predicting type 2 diabetes and hypertension", *J. King Saud Univ. Comput. Inf. Sci.*, Vol. 34, No. 3, pp. 862–870, 2020.

[7] Hasan, M.K., Alam, M.A., Das, D., Hossain, E., Hasan, M., "Diabetes prediction using ensembling of different machine learning classifiers", *IEEE Access*, Vol. 8, pp. 76516–76531, 2020.

[8] Cervantes, J., García-Lamont, F., Rodríguez, L., Lopez-Chau, A., "A comprehensive survey on support vector machine classification: Applications, challenges and trends", *Neurocomputing*, Vol. 408, pp. 189–215, 2020.

[9] Pranto, B., "Evaluating machine learning methods for predicting diabetes among female patients in Bangladesh", *Information* Vol. 11, pp. 1–20, 2020.

[10] Mohan, N., Jain, V., "Performance analysis of support vector machine in diabetes prediction", In: *International Conference on Electronics, Communication and Aerospace Technology*, pp. 1–3, 2020.

[11] Muhammad Azeem Sarwar, 2Nasir Kamal, 3Wajeeha Hamid, 4Munam Ali Shah, "Prediction of Diabetes Using Machine Learning Algorithms in Healthcare", 24th International Conference on Automation & Computing, Newcastle University, Newcastle upon Tyne, UK, 6-7 September 2019.

[12] Rao G.A., Syamala K., Kishore P.V.V., Sastry A.S.C.S., "Deep convolutional neural networks for sign language recognition", *International Journal of Engineering and Technology(UAE)*, Vol: 7, Issue 5, pp: 62-70, 2018.

[13] Quan Zou, Kaiyang Qu, Yamei Luo, Dehui Yin, Ying Ju and Hua Tang, "Predicting Diabetes Mellitus With Machine Learning Techniques", Springer, 2018.

[14] L. Zhou, S. Pan, J. Wang, and A. V. Vasilakos, "Machine learning on big data: Opportunities and challenges," *Neuro computing*, vol. 237, pp. 350–361, May 2017.

[15] J. B. Heaton, N. G. Polson, and J. H. Witte, "Deep learning for finance: deep portfolios," *Appl. Stoch. Model. Bus. Ind.*, vol. 33, no. 1, pp. 3–12, Jan. 2017.