Abhinav Sharma, 2021, 9:6
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

Survey Paper of Prediction of Diabetes using Machine Learning Algorithm for FPGA Application

M. Tech. Scholar Abhinav Sharma, Prof. Satyarth Tiwari, Prof. Suresh. S. Gawande
Department of Electronics and Communication,
Bhabha Engineering Research Institute,
Bhopal,MP,India

Abstract- In today's world, one of the major threats to human health is Diabetes Mellitus (DM). Diabetes is a metabolic disease where a person suffers from increased sugar level, is because of either the pancreas does not produce sufficient insulin for the body or the cells do not respond to the insulin. The persistent diabetes leads to malfunction, injury and failure of different organs such as kidneys, eyes, nerves, blood vessels and heart. In this research, the main objective is to classify the data as diabetic or non-diabetic and improve the classification accuracy. It presents an automatic prediction system for diabetes mellitus through machine learning techniques by taking into account of several limitations of traditional classifiers and provide a great relationship between patient's symptoms with diabetes diseases and the blood sugar rate. Machine learning provides a reliable and excellent support for prediction of a DM with correct case of training and testing. Diagnosis of diabetes mellitus desires great support of machine learning classifiers to detect diabetes disease in early stage, since it cannot be cured which brings great complication to our health system.

Keywords: Diabetes Mellitus (DM), Machine learning, Early Stage, FPGA Application.

I. INTRODUCTION

Diabetes is a extreme growing health problem in all over the world which causes death, especially in industrial and developing countries. This chronic disease may over to long term complications and death. It can cause high risk of kidney failure, nervous system damage, blindness and heart diseases. In this disease the body does not growth or rightly use the insulin, the hormone that release the cells of the body, grant the glucose to get in and ammunition them [1].

In the absence of insulin the cells become starving of glucose energy against the being of abundant glucose in blood stream. Complications of diabetes are linked to blood vessel diseases and generally classify into bitty vessel disease such as those affect the eyes, called as diabetic retinopathy.

It occurs in the patients who have diabetes for minimum five years with the drainage of protein in small blood vessels at the back of eyes and the blood in retina. Disease in blood vessel also causes the formation of small aneurysms and new but brittle blood vessel leaded to retinal scarring and retinal detachments thus impairing vision [2].

Kidney damage from diabetes is called diabetic nephropathy. Initially disease blood vessel in the kidney causes the leakage protein in the urine. Later on kidneys lose their capacity to scrub and channel blood. The collection of dangerous waste item in blood prompts the requirement for dialysis.

Nerve harm from diabetes is called diabetic neuropathy and it is additionally caused by ailment of little veins. The blood stream to the nerves is restricted leaving the nerves without blood stream and they get harmed or kick the bucket subsequently. The symptoms of nerve damage are numbness burning, aching of feet and lower extremities [3].

There are two major types of diabetes type I and type II. Type I is basically diagnosed in children

which is usually known as Juvenile diabetes and type II is most common form of diabetes. A patient with type II diabetes do not require insulin cure to remain alive, although up to 20% are treated with insulin to control blood glucose level [4].

Diabetes mellitus is a degenerative infection described by either absence of insulin or a protection from insulin, a hormone which is pivotal for digestion of glucose. In a sound individual, the pancreas produces insulin to help process sugar in the blood and keep up blood glucose (sugar) levels inside their ordinary range.

Diabetics can't deliver insulin or are impervious to insulin, and subsequently can't expel glucose from the circulatory system. Regardless of whether there is insufficient insulin or insulin protection, glucose levels in the blood increment and cause extreme medical issues. There are two noteworthy sorts of diabetes. Sort 1, or insulin-subordinate adolescent diabetics, is hereditary in cause and is described by the body's powerlessness to create insulin, and the subsequent development of glucose in the blood.

It normally happens amid pubescence or immaturity however can happen amid adulthood. Side effects incorporate outrageous craving and thirst, visit or over the top pee, and weight reduction. Definitive impacts of diabetes incorporate coronary illness, kidney malady, malignancy, hypertension, gangrene, diseases, visual deficiency, strokes, and demise.

II. RELATED WORK

Naveen Kishore et al. [1], diabetes is considered as one of the deadliest and persistent illnesses which causes an increment in glucose. Numerous entanglements happen if diabetes stays untreated and unidentified. The monotonous distinguishing measure brings about visiting of a patient to a symptomatic focus and counselling specialist. However, the ascent in AI approaches tackles this basic issue.

The thought process of this investigation is to plan a model which can visualize the probability of diabetes in patients with greatest exactness. Thusly three AI characterization calculations specifically Decision Tree, SVM and Naive Bayes are utilized in this analysis to identify diabetes at a beginning phase. Analyses are performed on Pima Indians Diabetes Database (PIDD) which is sourced from UCI AI storehouse. The exhibitions of the multitude of three calculations are assessed on different estimates like Precision, Accuracy, F-Measure, and Recall. Exactness is estimated over effectively and erroneously ordered cases. Results got show Naive Bayes beats with the most elevated precision of 76.30% relatively different calculations. These outcomes are confirmed utilizing Receiver Operating Characteristic (ROC) bends in a legitimate and methodical way.

Muhammad Azeem Sarwar et al. [2], there are a few AI strategies that are utilized to perform prescient investigation over enormous information in different fields. Prescient examination in medical services is a difficult undertaking in any case can help professionals settle on huge information educated ideal choices about tolerant's wellbeing and therapy. This paper talks about the prescient investigation in medical care, six diverse AI calculations are utilized in this examination work.

For analyze reason, a dataset of patient's clinical record is gotten and six distinctive AI calculations are applied on the dataset. Execution and precision of the applied calculations is examined and analyzed. Correlation of the diverse AI methods utilized in this examination uncovers which calculation is most appropriate for forecast of diabetes. This paper intends to help specialists and experts in early expectation of diabetes utilizing AI strategies.

Rao et al. [3], extraction of complex head and hand developments alongside their continually changing shapes for acknowledgment of communication via gestures is viewed as a troublesome issue in PC vision. This paper proposes the acknowledgment of Indian communication through signing motions utilizing an incredible man-made consciousness device, convolution neural organizations (CNN). Selfie mode ceaseless gesture based communication video is the catch technique utilized in this work, where a conference weakened individual can work the SLR versatile application freely.

Because of non-accessibility of datasets on portable selfie communication through signing, we started to make the dataset with five distinct subjects performing 200 signs in 5 diverse survey points under different foundation conditions. Each sign involved for 60 edges or pictures in a video. CNN preparing is performed with 3 distinctive example

estimates, each comprising of different arrangements of subjects and survey points. The leftover 2 examples are utilized for testing the prepared CNN. Distinctive CNN models were planned and tried with our selfie gesture based communication information to acquire better precision in acknowledgment. We accomplished 92.88% acknowledgment rate contrasted with other classifier models wrote about the equivalent dataset.

Reddy S.S. et al. [7], diabetic retinopathy (DR) is the leading cause of blindness worldwide, and therefore its early detection is important in order to reduce disease-related eye injuries. DR is diagnosed by inspecting fundus images. Since micro aneurysms (MA) are one of the main symptoms of the disease, distinguishing this complication within the fundus images facilitates early DR detection. In this paper, an automatic analysis of retinal images using convolutional neural network (CNN) is presented.

Kavakiotis, I et al. [8], the astounding advances in biotechnology and wellbeing sciences have prompted a critical creation of information, for example, high throughput hereditary information and clinical data, produced from huge Electronic Health Records (EHRs). To this end, use of AI and information mining strategies in biosciences is as of now, like never before previously, imperative and essential in endeavours to change shrewdly all accessible data into significant information. Diabetes mellitus (DM) is characterized collectively of metabolic issues applying huge tension on human wellbeing around the world.

Broad exploration in all parts of diabetes (determination, etiopathophysiology, treatment, and so on) has prompted the age of tremendous measures of information. The point of the current examination is to direct a methodical survey of the uses of AI, information mining strategies and instruments in the field of diabetes research regarding a) Prediction and Diagnosis, b) Diabetic Complications, c) Genetic Background and Environment, and e) Health Care and Management with the main class having all the earmarks of being the most famous.

A wide scope of AI calculations was utilized. As a rule, 85% of those utilized were portrayed by directed learning draws near and 15% by solo ones, and all the more explicitly, affiliation rules. Backing vector machines (SVM) emerge as the best and broadly utilized calculation. Concerning the kind of information, clinical datasets were predominantly utilized. The title applications in the chose articles project the convenience of removing important information prompting new speculations focusing on more profound arrangement and further examination in DM.

Majid Ghonji Feshki et al. [9], the impressive developing of cardiovascular malady what's more, its belongings and inconveniences and additionally the high expenses on society influences restorative group to look for answers for counteractive action, early recognizable proof and viable treatment with bring down expenses. In this manner, profitable information can be built up by utilizing counterfeit insight and information mining; they found learning makes enhance the nature of administration.

As of not long ago, extraordinary investigates have been completed to foresee coronary illness in light of information mining strategies, for example, characterization and grouping strategies; be that as it may, what has been less seen is the correct conclusion of infection with the most reduced cost and time.

In this paper, by utilizing highlight positioning on successful variables of malady identified with Cleveland facility database and by utilizing Particle Swarm Streamlining and Neural Network Feed Forward Back Propagation, powerful factors decreased to 8 enhanced highlights as far as cost and exactness. The evaluation of chose highlights of arranged strategies additionally demonstrated that PSO strategy alongside Neural Networks of Feed Forward Back-Propagation has the best precise criteria of the rate of 91.94% on these highlights.

L. Hermawanti et al. [10], diabetes happens when a body can't create or react legitimately to insulin which is expected to control glucose. Other than adding to coronary illness, diabetes likewise builds the dangers of creating kidney ailment, visual deficiency, nerve harm, and vein harm. Diabetes illness determination by means of legitimate elucidation of the diabetes information is an essential order issue. In this examination, a relative Pima diabetes infection conclusion was figured it out. For this reason, a multilayer neural system structure which was prepared by Levenberg–

Marquardt (LM) calculation and a probabilistic neural system structure were utilized.

The consequences of the examination were contrasted and the aftereffects of the past investigations announced concentrating on diabetes infection finding and utilizing the same UCI machine learning database. A similar report on Pima Indian diabetes sickness is symptomatic by utilizing multilayer neural system which was prepared by LM calculation and probabilistic neural system.

III. MAJOR CAUSES OF DIABETES

- The use of Prednisone (synthetic cortisone).
- The creation of balance cortisone, which be found in all hypothyroid people.
- Stress, which exacerbates hypothyroidism and suppresses the immune system in general
- Genetic defect (juvenile diabetes) - much more rare than hypothyroidism - induced high cortical.
- Certain viral infections
- Chemical harming from substances, for example, pesticides, and certain synthetic added substances, for example, alloxan, which is available in dyed white flour and is utilized as a part of business wheat items also.
- Allergic triggers.

Specifically, diabetics are lacking in lipase, which is required for ideal cell penetrability, which incorporates the creation of insulin and its vehicle inside the phone. Insulin is smothered in extent to the measure of undigested fat (triglycerides) in the blood. On the off chance that fat can be processed, insulin digestion can be moved forward. Lipase inadequacy (fat bigotry) prompts the failure to use glucose.

As above, more cases of Type II diabetes can be controlled through diet, and many cases of Type I will benefit from scheduled diet and nutrition as well. First, let's discuss foods for diabetics to avoid: white flour and all other flour products (including bleached and unbleached).

Expelled white flour is a less reason for nourishment. Not just have the grain and germ been expelled away, however wash out flour likewise contains a substance from the flour blanch (alloxan) which causes diabetes in creatures.

Furthermore, if creatures get diabetes from white flour, I beyond any doubt wouldn't eat it. Stay away from unbleached white flour also. It's as yet white flour, and is without supplements.

Stay away from white sugar, and every single refined sugar, which incorporate white sugar, fructose, corn syrup, and dextrose. Diabetics should by and large be watchful about their utilization of normal sweeteners also, including nectar, maple syrup, and molasses. Keep away from simulated sugars at all costs. NutraSweet is particularly terrible for insulin-subordinate diabetics since it influences controlling blood to sugar more troublesome.

Additionally dodge saccharin, which has been appeared to cause tumors in creatures. This implies barring beverages, for example, consistent and slim down soft drinks, juices with included sugar, and thought nourishments, for example, dried natural products.

IV. METHODOLOGY

Neural networks are the data handling frameworks, which are built and executed to show the human cerebrum. The main object of the neural network analysis is to develop a process of computing device for modeling the brain to execute various process of computing tasks at a faster rate than the traditional systems.

Artificial Neural Networks execute various tasks such pattern matching and classification, optimization function and data clustering. These errands are exceptionally troublesome for conventional PCs, which are quicker in algorithmic procedure of registering undertakings and exact number juggling tasks.

ANNs gangs substantial number of exceedingly interconnected preparing components called hubs or unit or neuron, which more often than not work in parallel and are arranged in standard models [8]. Every neuron is associated with the other by an association interface. Every association interface is related with weights, which contain data about the info flag. This data is required by neuron net to take care of a specific issue. ANN, s aggregate conduct is portrayed by their capacity to learn, review and sum up preparing examples or information like that of human mind.

1. Learning:
The main property of an ANN is its capability to learn. Learning or preparing is a procedure by methods for which a neural system adjusts to a boost by making legitimate parameter modifications, bringing about the generation of wanted reaction. Learning in an ANN is chiefly ordered into two classes as [9].

- Supervised learning
- Unsupervised learning

2. Supervised Learning:
On the off chance that a named reaction variable is accessible then the order has a place with the measurably managed learning topic. Regulated learning is two stage forms, in the initial step: a model is fabricated depicting a foreordained arrangement of information classes or ideas. The model developed by investigating database tuples portrayed by traits. Each tuple is expected to have a place with a predefined class, as dictated by one of the qualities, called to have a place with a reclassified class, as controlled by one of the traits called the class name characteristic. The information tuple are dissected to fabricate the model all things considered from the preparation dataset.

3. Unsupervised learning:
It is the kind of learning in which the class mark of each preparation test isn't knows, and the number or set of classes to be scholarly may not be known ahead of time. The prerequisite for having a named reaction variable in preparing information from the administered learning system may not be fulfilled in a few circumstances.

Data mining field is a highly efficient techniques like association rule learning. Data mining performs the interesting machine-learning algorithms like inductive-rule learning with the construction of decision trees to development of large databases process. Data mining techniques are employed in large interesting organizations and data investigations. Many data mining approaches use classification related methods for identification of useful information from continuous data streams.

4. Nearest Neighbors Algorithm:
The Nearest Neighbor (NN) rule differentiates the classification of unknown data point because of closest neighbor whose class is known. The nearest neighbor is calculated based on estimation of k that represents how many nearest neighbors are taken to characterize the data point class. It utilizes more than one closest neighbor to find out the class where the given data point belong termed as KNN. The data samples are required in memory at run time called as memory-based technique.

The training points are allocated weights based on their distances from the sample data point. However, the computational complexity and memory requirements remained key issue. For addressing the memory utilization problem, size of data gets minimized. The repeated patterns without additional data are removed from the training data set.

5. Naive Bayes Classifier:
Naive Bayes Classifier technique is functioned based on Bayesian theorem. The designed technique is used when dimensionality of input is high. Bayesian Classifier is used for computing the possible output depending on the input. It is feasible to add new raw data at runtime. A Naive Bayes classifier represents presence (or absence) of a feature (attribute) of class that is unrelated to presence (or absence) of any other feature when class variable is known.

Naïve Bayesian Classification Algorithm was introduced by Shinde S.B and Amrit Priyadarshi (2015) that denotes statistical method and supervised learning method for classification. Naive Bayesian Algorithm is used to predict the heart disease. Raw hospital dataset is employed. After that, the data gets preprocessed and transformed. Finally by using the designed data mining algorithm, heart disease was predicted and accuracy was computed.

6. Support Vector Machine:
SVM are used in many applications like medical, military for classification purpose. SVM are employed for classification, regression or ranking function. SVM depends on statistical learning theory and structural risk minimization principal. SVM determines the location of decision boundaries called hyper plane for optimal separation of classes as described in figure 1.4.

Margin maximization through creating largest distance between separating hyper plane and instances on either side are employed to minimize upper bound on expected generalization error. Classification accuracy of SVM not depends on dimension of classified entities.

The data analysis in SVM is based on convex quadratic programming. It is expensive as quadratic programming methods need large matrix operations and time consuming numerical computations.

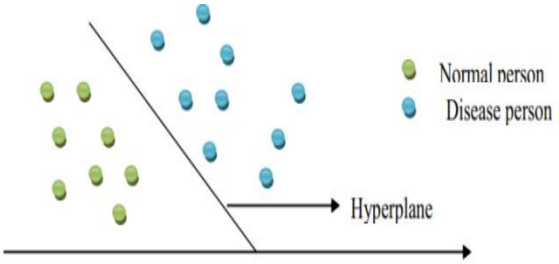

Fig 1. Support Vector Classification.

V. FPGA IMPLEMENTATION

Field Programmable Gate Arrays (FPGAs) are semiconductor devices that are based around a matrix of configurable logic blocks (CLBs) connected via programmable interconnects. FPGAs can be reprogrammed to desired application or functionality requirements after manufacturing.

This feature distinguishes FPGAs from Application Specific Integrated Circuits (ASICs), which are custom manufactured for specific design tasks. Although one-time programmable (OTP) FPGAs are available, the dominant types are SRAM based which can be reprogrammed as the design evolves. ASIC and FPGAs have different value propositions, and they must be carefully evaluated before choosing any one over the other. Information abounds that compares the two technologies.

While FPGAs used to be selected for lower speed/complexity/volume designs in the past, today's FPGAs easily push the 500 MHz performance barrier. With unprecedented logic density increases and a host of other features, such as embedded processors, DSP blocks, clocking, and high-speed serial at ever lower price points, FPGAs are a compelling proposition for almost any type of design.

VI. CONCLUSION

This thesis provides the need for machine learning techniques in health data mining and management processes. In particular, the dissertation provides special attention to diabetes disease prediction and diagnosis measures.

This due to the reason that diabetes diseases are one of the major concern across several countries and can even lead to death if not treated at an earlier stage.

The major thesis objectives are achieved successfully and specifically we have accomplished three tasks:

- First, we formulate the problems and difficulties of machine learning classification models across healthcare applications.
- We solve the issues across healthcare systems, we formulated algorithms that deal with those issues in an effective manner.
- We evaluated the proposed approach with real-time data.

REFERENCES

[1] Naveen Kishore G, V.Rajesh, A.VamsiAkki Reddy, K.Sumedh, T.Rajesh Sai Reddy, "Prediction of Diabetes using Machine Learning Classification Algorithms", International Journal of Scientific & Technology Research Volume 9, Issue 01, January 2020.

[2] 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 2018.

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

[4] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease Prediction by Machine Learning over Big Data from Healthcare Communities," IEEE Access, vol. 5, no. c, pp. 8869–8879, 2017.

[5] 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.

[6] 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.

[7] Reddy S.S., Suman M., Prakash K.N. ., "Micro aneurysms detection using artificial neural networks", 2018, Lecture Notes in Electrical Engineering, Vol: 434, Issue 3, pp: 409 to 417.

International Journal of Science, Engineering and Technology

An Open Access Journal

[8] Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., Chouvarda, I., "Machine Learning and Data Mining Methods in Diabetes Research", Computational and Structural Biotechnology Journal 15, 104–116, 2017.

[9] Majid Ghonji Feshki and Omid Sojoodi Shijan, "Improving the Heart Disease Diagnosis by Evolutionary Algorithm of PSO and Feed Forward Neural Network", International paper on IEEE 2016.

[10] L. Hermawanti, "Combining of Backward Elimination and Naive Bayes Algorithm To Diagnose Breast Cancer", Momentum, vol. 11, no. 1, pp. 42-45, 2015.

[11] O.S. Soliman, E. Elhamd, "Classification of Diabetes Mellitus using Modified Particle Swarm Optimization and Least Squares Support Vector Machine", IEEE 2014.

[12] K. Saxena, Z. Khan, S. Singh, "Diagnosis of Diabetes Mellitus using K Nearest Neighbour Algorithm", International Journal of Computer Science Trends and Technology (IJCST), 2014.

[13] L. Hermawanti, S.G. Rabiha, "Combining of Backward Elimination and K-Nearest Neighbor algorithms to Diagnose Heart Disease", Prosiding SNST Ke-5 Fakultas Teknik Universitas Wahid Hasyim, pp. 1-5, 2014.

[14] R.A. Vinarti, W. Anggraeni, "Identification of Prediction Factor Diagnosis of Breast Cancer Rates with Stepwise Binary Logistic Regression Method", JurnalInformatik, vol. 12, no. 2, pp. 70-76, November 2014.

[15] Muhammad Waqar Aslam, Zhechen Zhu and Asoke Kumar Nandi, "Feature generation programming with comparative partner selection for diabetes classification", "Expert Systems with Applications", 5402-5412, IEEE 2013.