

International Journal of Information Technology, Research and Applications (IJITRA)

K. Manohari and S. Manimekalai (2023). An Efficient Exploration on Big Data Analysis in Adolescent Diabetic Prediction with Deep Learning Techniques, 2(2), 33-40.

ISSN: 2583 5343

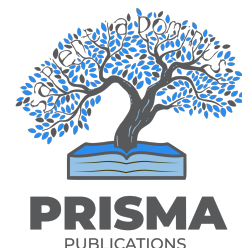
DOI: 10.59461/ijitra.v2i2.51

The online version of this article can be found at:
<https://www.ijitra.com/index.php/ijitra/issue/archive>

Published by:
PRISMA Publications

IJITRA is an Open Access publication. It may be read, copied, and distributed free of charge according to the conditions of the Creative Commons Attribution 4.0 International license.

International Journal of Information Technology, Research and Applications (IJITRA) is a journal that publishes articles which contribute new theoretical results in all the areas of Computer Science, Communication Network and Information Technology. Research paper and articles on Big Data, Machine Learning, IOT, Blockchain, Network Security, Optical Integrated Circuits, and Artificial Intelligence are in prime position.



<https://www.prismapublications.com/>

Journal homepage: <https://ijitra.com>

An Efficient Exploration on Big Data Analysis in Adolescent Diabetic Prediction with Deep Learning Techniques

¹K. Manohari and ²S. Manimekalai

¹ Ph.D Research Scholar, Department of Computer Science, Affiliated to Thiruvalluvar University
Theivanai Ammal College for Women, Autonomous, Villupuram, Tamil Nadu, India

² HOD, Department of computer science, Affiliated to Thiruvalluvar University
Theivanai Ammal College For Women, Autonomous, Villupuram, Tamil Nadu, India
¹hmano730@gmail.com, ²mamekaroshni@gmail.com

Article Info

Article history:

Received: Mar 23, 2023

Accepted: May 02, 2023

Published: June 22, 2023

Keywords:

Big Data Analysis
Automatic Systems
Healthcare Services
Diabetes
Artificial Intelligence
Deep Learning

ABSTRACT

The foundation of big data analysis is a massive volume of data. Diabetes is caused by an excess of sugar collected in the blood. Diabetes is one of the most serious chronic health issues. Diabetes sufferers' eyes, hearts, kidneys, and nerves may be damaged if they go undiagnosed. Humans can benefit from automated technologies to assist them in managing their hectic schedules. It inspires us to create a diabetes management scheme for patients that uses an IoT device to track their blood sugar, blood pressure, sports activities, nutrition plan, oxygen level, and ECG data. Machine learning has risen to prominence in healthcare services (HCS) due to its potential to enhance disease prediction. AI and ML approaches have already been used in the HCS field. We give a complete review of DL applications in diabetes in this study. We did a thorough literature review as well as discovered three key areas where this method is used: diabetes diagnosis, glucose control, and diabetes-related complication diagnosis. The search yielded 40 original research articles, from which we summarized essential data on used learning methods, development methods, main outcomes, and performance evaluation baseline techniques. According to the Reviewed Literature, various DL techniques and frameworks attained state-of-the-art performance in many diabetes-related tasks by outperforming conventional ML methods.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

K. Manohari
Ph. D Research Scholar,
Department of Computer Science, Affiliated to Thiruvalluvar University
Theivanai Ammal College for Women, Autonomous, Villupuram, Tamil Nadu, India
hmano730@gmail.com

1. Introduction:

Diabetes is a category of chronic metabolic disease caused by insufficient insulin production or insulin activity. International Diabetes Federation estimates that 463 million individuals worldwide have diabetes in 2019, half of whom are undiagnosed due to disease's complicated pathophysiology [1]. Furthermore, the global prevalence of diabetes is expected to rise dramatically in the future decade. As a result, diabetes prevention and treatment is significant financial burden on national economies, healthcare and personal medical spending, particularly in low- and middle-income nations [2]. Type 1 diabetes (T1D), type 2 diabetes (T2D), and gestational diabetes mellitus (GDM) are three primary clinical types of diabetes, according to etiopathology [3]. Adult-onset latent autoimmune diabetes and young-onset maturity-onset diabetes are two further kinds with particular aetiology. T1D is caused by the immune system attacking the pancreas' insulin-secreting β cells [4]. People with T1D lack endocrine insulin produced by pancreatic β - cells; hence they must rely on exogenous insulin.

T2D is most common type of diabetes, accounting for over 90% of cases. It is caused by insulin resistance or inadequate insulin synthesis. GDM can develop during pregnancy as well as may necessitate lifestyle changes as well as exogenous insulin delivery to avoid issues in baby. The early identification and classification of diabetes are typically challenging due to rising heterogeneity and a lack of continuous

surveillance [5]. To perform like a human, AI and ML systems require causability as well as Explainability, as many of best-performing ML and AI techniques are least transparent. Capacity of AI methods to explain themselves helps doctors have more faith in future AI methods. Causability is based on a causal process evaluated in terms of efficiency, effectiveness in terms of causal comprehension, and user transparency.

This review article's key contribution is that it considers both ML as well as AI methods in DM detection, diagnosis, self-management, and personalization. As far as we know, this is first review study to include both machine learning as well as artificial intelligence in detection, diagnosis, and self-management of diabetes, as well as the personalization of DM therapy. Review papers are important because they provide a complete summary of current research in a specific field. Furthermore, the authors have only looked at ML procedures, leaving out important aspects of the field such as databases, pre-processing techniques, feature extraction and selection methodologies utilised to find DM as well as AI answers to demand for intelligent DM assistants.

2. Diabetes Mellitus:

Diabetes Mellitus (DM), sometimes known as diabetes, is a term that refers to a group of diseases that affect how body transforms food into energy. When food is consumed, body converts it into glucose, a sugar that is then transported to bloodstream. Insulin is a hormone produced by pancreas that helps glucose move from bloodstream to cells that need it for energy. If you don't take medicine for diabetes, your body won't create as much insulin as it should. A condition known as high blood sugar occurs when the body retains a large amount of glucose. It has the potential to cause serious or life-threatening health problems. DM manifests itself in various ways, depending on the source [7].

2.1. Prediabetes

Prediabetes develops when blood sugar levels rise over what they should but are not high enough for a doctor to diagnose diabetes. As a result, prediabetes increases risk of type II diabetes as well as heart disease. These risks can be reduced by increasing exercise and losing extra weight (usually less than 5%-7% of body weight) [8].

2.2. Type-I diabetes

Insulin-dependent diabetes mellitus (IDDM) is used to describe type I diabetes. It's also known as juvenile-onset diabetes because it usually starts in childhood. Type I diabetes is an autoimmune disease. Antibodies threaten the pancreas, causing this to happen. Organ has deteriorated to the point where it no longer produces insulin. A person's genes can cause this type of diabetes. Complications with insulin-producing cells in the pancreas may potentially cause this. Many health problems associated with type I diabetes are caused by narrowing blood arteries in the kidneys, eyes, and nerves. Type-I people are at a higher risk of heart failure and stroke [9].

2.3. Type-II diabetes

Non-insulin-dependent diabetes, often known as adult-onset diabetes, is type-II diabetes. Type-II diabetes affects around 90% of individuals. When you have type II diabetes, your pancreas produces some insulin. However, either that isn't enough, or body isn't utilising it to its full potential. Type-II diabetes is also less severe than type-I diabetes. However, it has the potential to create serious health issues, particularly in small blood vessels of the nerves, kidneys, and eyes. Type-II diabetes also increases risk of stroke as well as heart failure. Insulin resistance is a side effect of obesity, and the pancreas has to work harder to create more insulin [10].

2.4. Other forms of diabetes

Other causes are causes of diabetes in 1% to 5% of people. Those who suffer from pancreatic diseases, as well as other operations and medications, as well as illnesses. In these circumstances, the doctor may wish to monitor blood sugar levels. While extensive DM research has yielded a wealth of material on a) etiopathology, b) diagnosis and c) illness detection, diagnosis, and treatment in recent decades, there is still much more to be discovered, unraveled, explained, and delineated. In this effort, relying on a large, rapidly developing body of clinical data and research provides a solid foundation for effective examination and follow-up. As a result, machine learning and artificial intelligence (AI) look to be essential technologies that

will greatly impact clinical decision-making. As a result, the goal is to link data analysis with treatment and intelligent decision-making in drug implementation and application [11].

3. Related works:

As data have become an integral part of any organization and analyzing it to discover hidden knowledge has become inevitable for improvement in services, same is true for medical field where predictive data mining is used for prognosis of disease at an early stage to pre-empt its effects and to aid physicians in developing contingency plan. The available literature reveals majority of the work that has been carried out on diabetes has focused mainly on developing the methods for prognosis or diagnosis of type II diabetes to reduce its complications, in majority of the cases Pima India dataset has been used for experimentation though methods and tool used have varied. To detect diabetes at initial stage a multi-agent system was designed by [12] called "CoLe" the system was not just a combinational agent but had multiple data miners. The intermingling of learning to depict information with alternate viewpoints was its fundamental goal and was achieved to large accuracy. To predict diabetes [13] applied, associative rule mining to discretize continuous-valued attributes an equal interval binning technique was used and for diabetic classification Apriori algorithm was applied and at the end association rules were generated for understanding relationship among measured fields used in prediction. In [14], Fuzzy ID3 combined with EM (Estimation Maximization) was used for diabetic prediction. The model was called Hybrid Classification System, it was a 2-phasic system in initial phase EM algorithm that was fed with cleaned data for clustering data and in second phase adaption rules essential for diabetic prognosis were obtained using ID3 algorithm. The models accuracy was about 91.32%. An expert system for diabetic prognosis was developed by [15]. The system used an extended classifier system (XCS)-a learning agent in artificial intelligence with a greater accuracy than others. The system is composed of a simple set of if-else rules. The achieved accuracy was 91.3%. Their research work [16] used novel artificial bee algorithm to predict diabetes. To improve prediction, they used a fake mutation operator. The model achieved an accuracy of 84.21%. Joint implementation of two most widely used algorithms, Support Vector Machines and Naïve Bayes was used by [17] for prognosis of diabetes. The system had an accuracy of 97.6%. Hybrid prediction model for prognosis of type II diabetes was proposed by [18], the system was the combination of two data mining classifiers K-means clustering as well as C4.5 decision tree method, K-fold cross-validation was applied for validation, system achieved an accuracy of 92.38%. To predict risk of heart attack in a diabetic patient [19] applied Naïve Bayes classifier on diabetic data with an accuracy of 74%. Ensembling methods were combined with the J48 decision tree algorithm to design a technique that could be applied to diabetic data set to predict the chances a diabetic patient has of diabetic neuropathy, diabetic nephropathy, and cardiovascular disease [20].

Researchers have begun to recognise the DL approaches' handling of huge datasets. As a result, DL approaches have also been used to predict diabetes. In last six years, seven research have been published. In addition, [21] employed a DNN. MLP, GRNN and RBF make up DNN's architecture. Pima Indian dataset was used to evaluate method. Because DNN can filter data as well as develop biases, authors did not pre-process dataset on purpose. Dataset is divided into two parts: 192 samples for testing as well as rest for training. Authors reported an accuracy percentage of 88.41 percent. Another study [22] employed two DL approaches to increase diabetes prediction accuracy. The performance of CNN and CNN-LSTM was calculated utilizing a private dataset called Electrocardiograms. There were 142,000 samples and eight qualities in total. Dataset was separated into training as well as testing sets using five-fold cross-validation. Because of DNN's self-learning, the authors did not pre-process data or utilise a feature selection strategy. Models produced accuracy rates of 90.9 percent and 95.1 percent, respectively. Logistic regression was employed as a baseline for MPNNs and CNN in study [23]. The goal was to use CGM signal dataset to identify diabetes patients. The dataset contains nine individuals, each of whom had 10,800 days of CGM data for 97,200 simulated CGM days. There was no discussion of qualities employed in this investigation.

Furthermore, [24] proposed the Deep Patient framework, an unsupervised DNN framework. Framework made use of a database of 704,857 patients' electronic health records. The scientists did not explain which features were employed in this dataset, but they did say that it may be used to forecast various diseases. The authors divided data into 5000 for validation, 76,217 for testing, and rest for training during validation phase. Area Under Curve (AUC) was used to determine accuracy, 0.91. To improve the prediction performance, the authors suggested pre-processing the dataset. Before running the DL, they proposed utilising PCA to extract relevant attributes. On a manually gathered dataset from a regional Australian hospital, work [25] used three distinct DL approaches. The dataset consists of 12,000 samples (patients) with a male-to-female ratio of 55.5 percent. To clean as well as decrease samples to 7191 patients, certain pre-processing procedures were used. The dataset was divided into two thirds for training, one-sixth for validation, and one-

sixth for testing for validation. LSTM, Markov, and Plain RNN were approaches used. Precision value was utilised to compare approaches' performance. Using LSTM, best precision value of 59.6 percent was reached. Recurrent Neural Network (RNN) was also utilised by [26] to predict the two forms of diabetes.

Pima Indian dataset, which had 768 samples as well as eight attributes, was used by researchers. According to their research, "Glucose, BMI, Age, Pregnancies, Diabetes Pedigree Function, Blood Pressure, Skin Thickness, and Insulin" are most important qualities. Dataset was split 80 percent for training and 20% for testing to validate the study. Type 1 diabetes was correctly predicted 78% of the time, whereas type 2 diabetes was correctly predicted 81%. Work [27] employed a one-dimensional modified CNN to predict diabetes based on breath signals and other investigations. The researchers gathered a breath signal dataset from 11 healthy patients, nine types 2 diabetes patients, and five types 1 diabetic patient. The dataset was not pre-processed in any way. The authors employed Leave-One-Out Cross-Validation for validation method. ROC curve, which achieved 0.96, was used to assess the performance.

3.1 Diagnosis of Adolescent Diabetes using Hybrid Architecture of Neural Network

Diabetes Mellitus (DM) is a prevalent chronic condition that can lead to serious health consequences and even death. The early detection of diabetes is critical, and significant complexity needs to occur countered. Many research studies on diabetes diagnosis have to obtain, the majority of which do base on a single data set, the Pima Indian diabetes data set. As a result, early detection and treatment are essential for illness prevention. Machine Learning (ML) approaches are self-handled to make more accurate predictions and improve performance. Hadoop/Map Reduce system was employed and a predictive analysis method to anticipate the forms of diabetes that are prevalent, the difficulties related with it and type of therapy that will deliver. This approach, according to the report, provides significant way to cure and care for patients with superior outcomes such as affordability as well as availability. Hadoop Distributed File System (HDFS) is used to store a large set of data and Map Reduce programming model is used to Analyze the massive amount of patient data sets in parallel programming method. Using this analysis it is possible to identify patients likely to suffer from diabetic risk and diagnose the patient at the earliest. Once the diabetes range is detected, for those patients with an abnormal diabetic range from their pre-historic medical data will obtain analysis for predicting the risk of cardiac attack. The data has been collected as images and classified using the hybrid architecture of the Convolutional neural network (CNN) that comprises VGG-19 and inception V3 algorithm.

3.2 Map reduce algorithm with Hadoop K-means clustering in Hadoop/ Map reduce environment for predictive analysis of Adolescent Diabetes:

In this system architecture, Hadoop clustering and hybrid architecture of neural network which comprises of VGG-19 and inception V3 algorithm are used, which enhances data processing and collection with higher efficiency. Hadoop is an open-source application of Map Reduce parallel processing model where processing details are hidden which includes distribution of data to the processing nodes, Subtasks and collection of computational results. In this framework, developers can focussing on computational problems in the series of Parallelization. DFS does use in Hadoop. Hadoop works on the master-slave framework along with Hadoop Distributed File System (HDFS) has two nodes that are Name Node (acts as a master) and Data Node (acts as a slave)., which works with identical patterns.

3.3 Pattern discovery:

For diabetic treatment it is necessary to test the patterns like, plasma glucose concentration, serum insulin, diastolic blood pressure, diabetes pedigree, Body Mass Index (BMI), age, number of times pregnant.

The pattern discovery of predictive analysis must include the following:

- Association rule mining- Association between diabetic type and pages viewed (e.g. laboratory results)
- Clustering- clustering of similar patterns of usage, etc.
- Classification- Classification of health risk value by the level of patient health condition.
- Usage of statistics

3.4 Predictive – Pattern Matching System Predictive Pattern Matching:

Whenever the warehoused dataset was sent to Hadoop system, immediately the map reduce task is performed. In mapping phase, the Master Node splits large data into smaller tasks for numerous Worker

Nodes. It deploys the exact operation of predictive pattern matching system. The Master node is one consists of Name Node (NN) and Job Tracker (JT), which always employs the map and reduce task. The Worker Node or Slave Node receives the order from the Master Node, process the pattern matching task for diabetes data with the help of Data Node – Same Machine (DN) and Task Tracker (TT). The predictive matching is the process of comparing the analyzed threshold value with the obtained value. If the pattern matching process was completed by all Worker Nodes based on the requirement, it was stored in intermediate disks. This process is known as local write. If the reduce task was initiated by Master Node, all other allocated Worker Nodes will read the processed data from intermediate disks. Based on the query received from Client through Master Node, the reduce task will be performed in Worker Node. The results obtained from the reduce phase will be distributed in various servers.

3.5 Processing Analyzed Reports

After the analysis of large diabetic data went through Hadoop, the final results are distributed over various server and replicated through several nodes depending on the geographical area. By employing proper electronic communication technology to exchange the information of individual patients among health care centers will leads to get proper treatment at right time in remote locations at low cost.

Table-3.1 Summary of DL method related works in Diabetes predictive analysis

Reference	Technique used	Dataset	Results	Limitation
[15]	Extended classifier system	CPCSSN Database	Accuracy-91.3%	Need to minimize the error of desired output
[16]	Artificial bee algorithm	Chinese Gut Microbiota Datasets	Accuracy-84.21%	Accuracy of this method was needed to be improved
[17]	Support Vector Machines and Naïve Bayes	Pima Indians Diabetes Database	Accuracy-97.6%	This model was verified with only one dataset and not with multiple datasets.
[18]	K-means clustering as well as C4.5 decision tree method	Electronic Health Records	Accuracy-92.38%	The main limitations of the model are deficiency of the training dataset
[19]	Naïve Bayes classifier	Pima Indians Diabetes Dataset	Accuracy- 74%	performed well but not up to the mark considering all the parameters
[21]	DNN	Pima Indians Diabetes Dataset	Accuracy rate reported by authors was 88.41%	Accuracy in disease prediction needs to be enhanced
[23]	MPNN and CNN	Private Dataset (Collected Manually)	CNN attained highest accuracy of 77.5%	This method uses very complicated procedure and accuracy has been enhanced

3.6 Benefits of this predictive analysis system

The diabetes may associate with severe diseases such as heart attacks, strokes, eye diseases and kidney diseases, etc. Analyzing the risk value by the level of patient health condition using above results of can be used by the physicians at remote locations to serve the people. Detecting diseases at earlier stages can help to be treated more easily and effectively. In developing countries such as India, it is mandatory to manage specific individual and population health and detecting health care fraud more quickly. The middle-income families can be with the high availability of medical facility at minimum cost. This system leads to the improved focus on every individual patient health. Thereby we can reduce and save our next generation from diabetic mellitus.

4. Results and Discussion:

Although DL has advanced state of the art in various areas of diabetes, healthcare applications must be solid, dependable, and persuasive to avoid safety concerns as well as deliver significant treatment support. In this context, several restrictions and obstacles must be overcome before DL may be used in therapeutic settings. First, due to human errors and sensor abnormalities, data obtained from people with diabetes in real-world circumstances are prone to being imperfect. The process of gathering real data can be costly and time-consuming. Sharing datasets among research teams might be problematic due to data privacy policies. As a result of these issues, many studies use a limited and sometimes insufficient amount of data.

Furthermore, deep learning models are opaque. Why methods produce output for a specific input situation is crucial to physicians, especially for some critical decision-making applications. Intricate structures of DNN layers significantly learn patterns from non-linear data but impair interpretability of method. When looking into DL for diabetes, it's critical to think about trade-off between performance as well as interpretability. Finally, new algorithmic and hardware advances are likely to improve the efficiency of training DL methods.

5. Applications:

DL is a hot topic in the AI era, and it's worth noticing that the majority of the papers chosen are from the last two years, indicating that this is a new technique. As a result, there is much room to improve the present diabetic applications. First, multimodal methods with wearables as well as smartphone applications are progressively collecting digital data and vital indicators. The majority of this information can be easily uploaded to centralised systems or cloud storage. Data volumes, as well as variability of data sources, are expected to enhance efficiency in numerous healthcare applications, including diabetes care, as IoT, as well as 5G networks, become more popular. Many low-quality data samples are filtered out as well as deleted from training sets as data volume grows, and advancements in wearables can effectively reduce measurement mistakes. DL is well-suited to handle such a surge in data availability.

Numerous recent initiatives in AI area is undertaken to improve model transparency as well as comprehend model functionality to interpret DL methods, In particular, SHapley Additive explanations (SHAP) was presented as a unified framework for explaining input features that contribute to final output. This method of ranking input features is also helpful for selecting input features. T-SNE is applied to various CNN applications, such as DR detection, qualitatively assesses the collected feature maps. Furthermore, in terms of glucose-insulin dynamics, a recent study confirmed conformity of NN methods. Analyzing performance of DNNs as well as improving interpretability can be done using similar methods. Rather than relying exclusively on data-driven methods, incorporating expert knowledge into learning process might aid in better understanding underlying causes of a disease like diabetes. There are two viable options in particular. One is to employ expert knowledge as a guide throughout training process, and the other is to incorporate physiological characteristics as an input element of methods. Expert knowledge is also required to create safety constraints as well as evaluate model output confidence. Numerous papers emphasised that their findings needed to be confirmed in real-world situations. In this regard, a Google team has made significant progress. In a human-centred study, they used deep learning to treat diabetic eye problems at 11 clinics. The findings suggest several socio-environmental issues must be addressed before such automated systems can be widely deployed.

6. Conclusion and future scope:

The adoption of Big Data Analytics in Hadoop provides a systematic means to achieve improved outcomes, such as the availability and cost of healthcare services for all populations. NCDs such as diabetes is a major health concern in India. This study will help diabetes patients comprehend difficulties that may emerge by converting diverse health records into relevant analysed results. This paper provides a complete overview of the current trend in DL methods for diabetes research. We conducted a systematic search, chose a set of publications, and summarised major findings in three areas: diabetes diagnosis, glucose management, and diabetes-related problems diagnosis. Many DNN and learning algorithms are used in these areas, and the results have outperformed earlier conventional machine learning approaches in terms of experimental performance. Data availability, feature processing, and model interpretability have all been noted as problems in the literature. Transferring the newest breakthroughs in DL technology into vast multi-modal data of diabetes care has the potential to meet these issues in the future. An anticipate that DL methods are widely used in clinical settings as well as will significantly enhance diabetes management.

References:

- [1]. Maniruzzaman, M., Rahman, M., Al-MehediHasan, M., Suri, H. S., Abedin, M., El-Baz, A., & Suri, J. S. (2018). Accurate diabetes risk stratification using machine learning: role of missing value and outliers. *Journal of medical systems*, 42(5), 1-17.
- [2]. Poss, A. M., Maschek, J. A., Cox, J. E., Hauner, B. J., Hopkins, P. N., Hunt, S. C., ... & Playdon, M. C. (2020). Machine learning reveals serum sphingolipids as cholesterol-independent biomarkers of coronary artery disease. *The Journal of clinical investigation*, 130(3), 1363-1376.
- [3]. De Silva, K., Mathews, N., Teede, H., Forbes, A., Jönsson, D., Demmer, R. T., & Enticott, J. (2021). Clinical notes as prognostic markers of mortality associated with diabetes mellitus following critical care: A retrospective cohort analysis using machine learning and big unstructured data. *Computers in Biology and Medicine*, 132, 104305.
- [4]. Jamuna, A., Priyadarsini, D. R. J., & Titus, D. S. (2020). Survey on Predictive Analysis of Diabetes Disease Using Machine Learning Algorithms. *A. Jamuna et al, International Journal of Computer Science and Mobile Computing*, 9(10), 19-27.
- [5]. Sonar, P., & JayaMalini, K. (2019, March). Diabetes prediction using different machine learning approaches. In *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 367-371). IEEE.
- [6]. Samet, S., Laouar, M. R., & Bendib, I. (2021, October). Use of Machine Learning Techniques to Predict Diabetes at an Early Stage. In *2021 International Conference on Networking and Advanced Systems (ICNAS)* (pp. 1-6). IEEE.
- [7]. Latchoumi, T. P., Dayanika, J., & Archana, G. (2021). A Comparative Study of Machine Learning Algorithms using Quick-Witted Diabetic Prevention. *Annals of the Romanian Society for Cell Biology*, 4249-4259.
- [8]. Rawat, V., & Suryakant, S. (2019). A classification system for diabetic patients with machine learning techniques. *International Journal of Mathematical, Engineering and Management Sciences*, 4(3), 729-744.
- [9]. Mahmud, S. H., Hossin, M. A., Ahmed, M. R., Noori, S. R. H., & 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).
- [10]. Bide, P., & Padalkar, A. (2020, March). Survey on Diabetes Mellitus and incorporation of Big data, Machine Learning and IoT to mitigate it. In *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)* (pp. 1-10). IEEE.
- [11]. Ameena, R. R., & Ashadevi, B. (2020). Predictive analysis of diabetic women patients using R. In *Systems Simulation and Modeling for Cloud Computing and Big Data Applications* (pp. 99-113). Academic Press.
- [12]. Maniruzzaman, M., Rahman, M., Al-MehediHasan, M., Suri, H. S., Abedin, M., El-Baz, A., & Suri, J. S. (2018). Accurate diabetes risk stratification using machine learning: role of missing value and outliers. *Journal of medical systems*, 42(5), 1-17.
- [13]. Poss, A. M., Maschek, J. A., Cox, J. E., Hauner, B. J., Hopkins, P. N., Hunt, S. C., ... & Playdon, M. C. (2020). Machine learning reveals serum sphingolipids as cholesterol-independent biomarkers of coronary artery disease. *The Journal of clinical investigation*, 130(3), 1363-1376.
- [14]. Swapna, G.; Soman, K.P.; Vinayakumar, R. Automated detection of diabetes using CNN and CNN-LSTM network and heart rate signals. *Procedia Comput. Sci.* 2018, 132, 1253–1262.

-
- [15]. Rahimloo, P.; Jafarian, A. Prediction of Diabetes by Using Artificial Neural Network, Logistic Regression Statistical Model and Combination of Them. Bull. Société R. Sci. Liège 2016, 85, 1148–1164.
- [16]. Gill, N.S.; Mittal, P. A computational hybrid model with two level classification using SVM and neural network for predicting the diabetes disease. J. Theor. Appl. Inf. Technol. 2016, 87, 1–10.
- [17]. Sun, Y.L.; Zhang, D.L. Machine Learning Techniques for Screening and Diagnosis of Diabetes: A Survey. Teh. Vjesn. 2019, 26, 872–880.
- [18]. Choudhury, A.; Gupta, D. A Survey on Medical Diagnosis of Diabetes Using Machine Learning Techniques. In Recent Developments in Machine Learning and Data Analytics; Springer: Singapore, 2019; pp. 67–68.
- [19]. Meherwar, F.; Maruf, P. Survey of Machine Learning Algorithms for Disease Diagnostic. J. Intell. Learn. Syst. Appl. 2017, 9, 1–16.
- [20]. Avinash Golande, Pavan Kumar T ,—Heart Disease Prediction Using Effective Machine Learning Techniques I, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-1S4, June 2019.
- [21]. S.Pitchumani Angayarkanni ,—Predictive Analytics of Chronic Kidney Disease using Machine Learning AlgorithmI, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-2, July 2019.
- [22]. Aditi Gavhane, Gouthami Kokkula ,I Prediction of Heart Disease Using Machine LearningI 2nd International conference on Electronics, Communication and Aerospace Technology, IEEE (ICECA 2018)
- [23]. D.Ramana kumar, Dr. S.Krishnamohan Rao,I Health care System: Stream Machine Learning Classifier for features Prediction in Diabetes TherapyI International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 1 (2018)
- [24]. Mrs. Ashwini Abhale ,Shruti Gulhane, —Predictive analysis of Diabetic Patient Data Using Machine Learning and Big Datal, International Journal of Research in Advent Technology (IJRAT) Special Issue E-ISSN: 2321-9637, 27th February 2019.
- [25]. B. Suvarnamukhi, M. Seshashayee ,— Big Data Processing System for Diabetes Prediction using Machine Learning TechniqueI, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-12, October 2019.
- [26]. Thakkar, H., Shah, V., Yagnik, H., & Shah, M. (2021). Comparative anatomization of data mining and fuzzy logic techniques used in diabetes prognosis. *Clinical eHealth*, 4, 12-23.
- [27]. Sharma, M., Singh, G., & Singh, R. (2018). Accurate Prediction of Life Style Based Disorders by Smart Healthcare Using Machine Learning and Prescriptive Big Data Analytics. *Data Intensive Computing Applications for Big Data*, 29, 428