Grenze International Journal of Engineering and Technology, June Issue

Diabetes Disease Prediction Review on Machine Learning Techniques

Kumar Gaurav[1]and Dr.Kakoli Banerjee[2]
[1-2]JSS Academy of Technical Education/CSE, Noida, India
E mail: believegaurav22@gmail.com,kakoli.banerjee@jssaten.ac.in

***Abstract*—Study into decision support applications in healthcare, such as those dealing to diagnosis, forecasting, treatment planning, and various aspects of patient care, among other things, have seen a significant uptick in interest over the past few years. This surge in interest has been accompanied by a substantial increase in funding. This expansion has taken place as a consequence of developments in artificial intelligence and machine learning , as well as an increase in the availability of computer resources, as well as an increase in the quantity of data that is readily available. Every day, examples of studies that are both very intriguing and have been peer-reviewed are published in journals, which demonstrates the promise of this topic. At the same time, however, there are certain expectations and assumptions regarding the development, validation, and acceptance of such techniques that are unrealistic and at times overly optimistic. These expectations and assumptions exist at the same time as the unrealistic and overly optimistic ones. There are in-depth lectures provided on the topics of machine learning and deep learning, both of which are often used for health recommender systems, as well as how they are applied in the field of health informatics. The purpose of this study is to investigate many approaches that have been used in the area of health informatics in the past.**

***Index Terms*—Dataset, Healthcare, Machine Learning Algorithms, Random Forest, KNN, logistic Regression, Naïve Bayes, SVC, Diabetes Disease Prediction .**

I. INTRODUCTION

The area of computer science referred to as machine learning (ML) is the one that is rising at the one that is expanding at the one that is increasing at the one that is developing at the quickest speed, and the field of health informatics may include one of the most challenging challenges to tackle. The purpose of the field of study known as machine learning is to produce software applications for computers that are capable of learning new information and enhancing their capabilities over the course of time. In addition to this, the goal of machine learning is to create computer programs that may be used for the purpose of creating forecasts. The great majority of researchers working in the field of machine learning are concentrating their efforts on automatic machine learning, which is sometimes referred to as AutoML. This is an area of study that has seen substantial progress achieved in a number of different domains, including voice recognition, recommender systems & autonomous cars, amongst others [1]. When employing automated processes, it is likely to be to one's advantage to have access to substantial quantities of data as well as a sizeable number of different training sets. In spite of the fact that the health industry often works with occasional events and relatively small data sets, there are times when machine learning algorithms are constrained because there are not enough instances for training. This is particularly the case in the area of the economy that deals with health care [2].

Grenze ID: 01.GIJET.9.2.127

The practice of medicine is entering a new era, one in which the plethora of biological data will play an ever more vital role, and this shift is already well on its way to being completed. The practice of medicine is about to enter a whole new period as a result of this new age. For the purpose of demonstration, the goal of precision medicine is "to guarantee that the appropriate therapy is administered to the appropriate patient at the appropriate time."
This is achieved by taking into consideration a broad range of elements linked to the patient's data, such as the variability in the patient's genetic characteristics, in addition to environmental factors, electronic health records (EHRs), and lifestyle choices [3]. One example of this kind of factor is the patient's molecular trait variability. The patient's molecular attributes are one example of this sort of factor, and they might differ from case to case.
The field of study that focuses on health care provides an overwhelming number of possible opportunities and challenges as a direct result of the large amounts of biological data that are now accessible. The production of reliable medical tools that are based on data-driven methodologies and machine learning techniques involves a number of obstacles. These tools must be able to learn from their own experience. Exploration of the links between all of the multiple bits of information that are present within these data sets is one of the most major of these problems. These data sets contain a lot of information. Earlier research [4] have made efforts to attain this aim by building linkages across different data sources with the goal of developing common knowledge bases that may be utilized for predictive analysis and discovery. This has been done with the intention of creating shared knowledge bases that might be used for these purposes. The links were built in these research with the goal of building common knowledge bases that could be utilized for these many reasons. In spite of the fact that the models that are already in use have a sizeable amount of potential, the use of medical prediction tools that are based on the ideas of machine learning has not yet attained widespread adoption.
Diabetes is a chronic metabolic illness that is characterized by high blood glucose (or blood sugar) levels. Diabetes causes long-term damage to the cardiovascular system, including the heart, blood vessels, eyes, kidneys, and nerves. When your blood glucose level, which is sometimes referred to as blood sugar, is too high, you are at risk for developing the illness known as diabetes. Diabetes mellitus type 1, diabetes mellitus type 2, and diabetes caused by pregnancy are the most frequent forms of the disease.
Diabetes type 1 is distinguished from other forms of the disease by the body's inability to produce insulin. Insulin-producing cells in your pancreas are targeted by your immune system, which leads to their death. Despite the fact that diabetes may manifest at any age, type 1 diabetes is more often seen in children and young adults. Insulin is essential to the survival of people with type 1 diabetes on a daily basis.
Insulin production and use are both hampered in people with type 2 diabetes. It is possible to acquire type 2 diabetes at any age, including in young children. On the other hand, the risk of developing this kind of diabetes is highest in persons who are in their forties and fifties. The great majority of people who are diagnosed with diabetes have type 2.
There is a possibility that some pregnant women may acquire gestational diabetes. In most cases, diabetes of this kind disappears following the birth of the baby. On the other hand, if you have gestational diabetes, you have a greater chance of getting type 2 diabetes in the years to come. There are several instances in which type 2 diabetes is diagnosed throughout the course of a pregnancy. Two types of diabetes that are far less prevalent include monogenic diabetes, which is caused by a genetic mutation, and diabetes that is associated with cystic fibrosis.
A variety of novel technologies are being utilised to foretell diabetic illness. Among these technologies are: AI : Artificial intelligence (AI) is being utilised to create algorithms that can examine vast volumes of data to find trends that could be related to diabetes. One research, for instance, utilised AI to analyse data from more than 100,000 individuals and was able to pinpoint a number of variables that were linked to a higher risk of acquiring type 2 diabetes. Obesity, high blood pressure, and high cholesterol were some of these contributing variables.
wearable technology : Data on a person's physical activity, sleep habits, and heart rate may be gathered via wearable technology, such as fitness trackers and smartwatches. This information can be utilised to spot diabetes early warning signals such variations in blood sugar levels. Smartphones: Smartphones may be used to gather information on a person's eating patterns, exercise routines, and medication compliance. Personalised diabetes management strategies may be created using this data.

II. Literature Review

According to Moreb et al. [1] researchers and developers are permitted to make use of the SEMLHI framework in order to organize activities inside the framework by using a SEMLHI approach. This permission extends to the general public. Software, artificial learning, artificial learning algorithms, and fitness informatics data are the four individual components that come together to make the overall framework. These modules not only provide software developers with a new road map for designing fitness applications that incorporate both system functions

and software implementations, but they also make it possible for software developers and researchers to carry out an analysis of fitness informatics software from an engineering point of view. This is a significant contribution to the field of fitness informatics.

Beam et al. [2] The influence that big data and machine learning are having on almost every facet of modern life is leading people to adjust their lives in order to make way for the breakthroughs that are coming about as a result of these two fields. As a direct result of this, the contemporary world is growing more complicated. Netflix and Google are both aware of the movies that users choose to watch on their respective platforms because of the history of people's search queries. In addition, Netflix and Google are both aware of the information that people are interested in learning more about but don't know how to access.

Ferdous et al. [3] In this specific piece of work, there is no preference exhibited for any one particular strategy that was identified during the course of the research that was carried out. Instead, there is a preference demonstrated for all of the approaches taken together. The author has not too long ago investigated the algorithms that were used, in addition to the degree of accuracy that was provided by the various ways. After everything is said and done, the authors of this study have high hopes that the findings of this research will make it simpler for both medical professionals and academic researchers to select the algorithm that will be the most effective in predicting the diseases that the authors of this study have described. This hope is based on the fact that the authors of this study have described these diseases.

Rahman et al. [4] Researchers, scientists, and industry specialists who specialize in the study of fitness care and health have been digging deeply into the analysis of both the technologies that are now available in this sector and those that are still in the process of being produced in this sector. This has been done in order to get a better understanding of both sets of technologies. Because of the conditions, it is required to establish a clinical decision support system in order to offer medical practitioners with the tools they need in order to make an accurate diagnosis of this ailment in its early stages. This is because an accurate diagnosis of this condition in its early stages is essential to the successful treatment of the condition. It is imperative that this step be taken in order to provide patients the highest level of care that is humanly feasible.

[5] Patel et al. explain that the process of illness prediction involves the use of a variety of different algorithms for supervised artificial learning to be carried out. This learning system is completely man-made. In the coming years, accurate medical data analysis will, without a doubt, be of assistance in the early detection of illnesses and in the treatment of patients as a result of the development of big data in the fields of biomedicine and health care. This is without a doubt going to be the case as a direct result of the application of big data to the fields of biomedicine and health care. This is just one of the numerous ways that big data may be put to use.

Along with the others, Shams [6] In this research, the Support Vector Machine, often known as an SVM, is used to produce estimates about both the effect that the COVID-19 pandemic will have on a diet and the number of individuals who would die as a direct consequence of the pandemic. These projections are made using the data collected from the previous study. In addition to this, the SVM is used in order to make projections about the total number of individuals who will get infected with the pandemic.

Hari kumar et al. [7] Forecasting sales is one of the most prevalent uses, along with predicting sicknesses, increasing sales, improving water irrigation systems, and increasing profits. Other typical applications include optimizing water irrigation systems, increasing sales, and enhancing sales. These are some examples of other common applications: Some of the further uses of this technology include the administration of portfolios, the monitoring of security, the identification of fraudulent behavior, and the management of inventories.

Mienye et al. [8] The key emphasis of the research that is being carried out right now has been on the creation of enhanced ways of artificial learning, in particular pre-processing pipelines and feature learning algorithms, in order to get around these difficulties. This research is being carried out in order to overcome the challenges that are now being faced. The purpose of doing this study is to identify a way to overcome the obstacles that have been put before us.

Fetaji et al. [9] Many different machine learning algorithms, such as J48, Random Forest, and Nave Bayes, have been shown to be effective after being tested, and the evaluation process has concluded that these algorithms have been successful as a result of their effectiveness. The process of analyzing and generating conclusions from data sets that are just too big for people to manage on a day-to-day basis is referred to as "artificial learning," which is a term that refers to a methodology. The term "machine learning" is often abbreviated with the letter "ML."

Khan et al. [10] In order to produce projections about software flaws for the purpose of this study, seven separate datasets were used, along with a range of artificial learning methodologies. This was done in order to help answer the research question.

Lalmuanawma et al. [11] This study aims to provide an in-depth analysis of the role that artificial intelligence and artificial learning play as a crucial tool in the areas of screening, predicting, forecasting, contact tracing, and

therapeutic development for SARS-CoV-2 and its associated pandemic. This will be accomplished by examining the role that artificial intelligence and artificial learning play as a crucial tool in these areas. The purpose of this research is to do a comprehensive examination of the part that machine learning and artificial intelligence play as an important instrument in the world today. The goal of this study is to conduct an in-depth investigation of the role that machine learning and artificial intelligence play as significant tools in the world as it exists now.

Scott et al. [12] The use of algorithms is not devoid of the possibility of difficulties that may follow their employment. The author checklist of ten questions, which clinicians may use to evaluate the reasons provided in favor of a particular method, does not require clinicians to exhibit understanding of what may be quite difficult statistical and computational concepts. However, this understanding would be required of non-experts in the field if they were to use the checklist. Clinicians may use the checklist to evaluate the reasons provided in favor of a particular method.

Mijwil et al. [13] There is a diverse range of ailments, each of which may present itself in a unique way on a variety of distinct people. There is a chance that some people are conscious of their circumstances, while others may be unaware of the dangers of ignoring their health, which might result in death or make the sickness chronic for the rest of their lives. There are a great number of diseases, and each one, when it manifests in the body of a certain person, may do so in a way that is unique to that particular condition.

Zhao et al. [14] Deep learning, on the other hand, provides crucial abilities for analyzing and interpreting the huge volumes of data that are produced by industrial equipment. This information may be accessible in a number of different file formats. The major objective of this investigation is to provide both a review and a synopsis of the most current deep learning research that has been published on the topic of artificial fitness monitoring. This information will be gathered in order to fulfill the aforementioned objectives.

Kumar et al. [15] The approach that has been suggested takes use of an artificial learning model that is already pretrained and that was trained on the same dataset before being posted on firebase. This model was trained on the dataset before it was placed on firebase. After that, the analysis is carried out using this model, and the outcomes of the analysis are afterwards shown inside the android application.The information pertaining to the model that is being suggested is submitted into the system via the use of an android application. Following that, the findings of the research are displayed within the android application.

Arvanitis et al. [16] As a potentially useful alternative that may speed up the spread of digital applications, the generation of synthetic datasets is now taking place as a step in the right direction. These datasets are constructed with the purpose of testing digital applications in mind, so that they may be as realistic as possible and better ease the testing process.

Kaissis et al. [17] In the last portion of the investigation, it is shown that the trained model was successfully applied to an end-to-end encrypted remote inference scenario by using secure multi-party computing. The secrecy of the data and model in issue is the key concern of this application's design and implementation. It is quite likely that acquiring access to medical data in order to train artificial intelligence applications would be difficult due to concerns about the privacy of patients and the property rights of a number of different organizations. This would make it difficult to train artificial intelligence applications on medical data.

Ojo et al. [18]. The Random Forest model was used to develop a feature significance in order to identify the most significant characteristics that were responsible for determining whether or not a hospital offered maternal fitness delivery services. This was done in order to determine the most significant characteristics that were responsible for determining whether or not a hospital provided maternal fitness delivery services. This was done in order to discover the most relevant criteria that were responsible for determining whether or not a hospital offered services related to maternal fitness delivery.

Rahman et al. [19] These methods have the potential to be used to databases that deal with health and fitness care in order to achieve the goal of knowledge discovery in a number of unique fitness care domains that have been discovered. The use of artificial neural networks is one of the data mining strategies that is briefly covered in this article, which provides an overview of a number of other data mining strategies. Throughout the entirety of this article, a particular emphasis is placed on how the aforementioned strategies influence medical diagnosis.

There are a variety of ethical issues that need to be taken into account throughout the process of designing and implementing such systems, as well as while obtaining judgments from them. In spite of the fact that artificial intelligence is positioned to play a large part in the healthcare business, the fact that these issues are crucial should not be discounted as a result of this reality. The current state of affairs is like way despite the fact that AI is capable of playing a significant part in the healthcare industry.

The accountability and openness of judgments made by such systems, the potential for collective harm coming from algorithmic prejudice, and the professional obligations of doctors are just a few examples of the ethical challenges that need to be addressed by academics. There are a great number of other ethical issues that need to

be resolved. Integrity problems among medical professionals are yet another topic that has to be discussed and resolved.

III. COMPARISONS

After an extensive literature review, it was concluded that the following are the technologies that have been used for Prediction Diabetes Disease , Advantages and Disadvantages of the technologies that have been used for Prediction Diabetes Disease are shown in Table 1 :

TABLE I TECHNIQUE COMPARISONS

S.no.	Paper	Techniques	Dataset	Methodology	Advantages	Disadvantages
1.	[1]	KNN, SVM, Naïve Bayes	Health dataset prediction output.	In this SEMLHI methodology is used	Simple Prediction algorithm.	Low accuracy.
2.	[2]	Bigdata based advanced Machine learning	Google Dataset.	In this instead of having the behaviour explicitly written, machine learning was initially defined as a programme that learns to carry out a task or make a judgement automatically from data.	Preprocessing to reduce complexity levels	High complexity due to Large Dataset.
3.	[4]	CSDC- SVM .	15 fold – Cross-Validation Strategy Data	Cloud based smart detection Algorithm is used.	Better accuracy and efficiency	Complexity in cross -fold validation testing
4.	[5]	Support Vector Machine	Deep Regression model Dataset	In this system uses various Supervised learning algorithm for evaluating the model. The system will predict disease based on vivid symptoms.	Low RMSE	Low Accuracy
5.	[6]	Naïve bayes	Diabetes Dataset.	The suggested model in this case relies on SVM with DL for regression and both RBF and poly kernels.	ML algorithms to Predict Diabetes is good	Complex Calculation .
6.	[8]	SVM and KNN	Cleveland dataset	An unsupervised sparse autoencoder (SAE) and artificial neural network were used in this technique to predict the risk of heart disease..	Better performance compared to other ML algorithms	Computation is expensive
7.	[9]	Naive Bayes, Random Forest, and J48.	Datasets is from public domain	Data that is just too large for humans to analyse is utilised for machine learning (ML) to assess and make predictions.	Simplicity and lack of bias	Only works in sample frame
8.	[20]	Support vector machine and random forest (RF) based on decision trees.	Pima Indian Diabetes Database	Using machine learning, a special intelligent diabetes mellitus prediction framework (IDMPF) is created in this study.	Provides 83% accuracy with the lowest possible error rate.	Absent hypothetical diabetes monitoring system.

9.	[21]	ANN model	The National Institute of Diabetes and Digestive and Kidney Diseases.	To anticipate the diabetes diagnosis, a practical framework has been put out.	More Efficient	Complex calculation
10.	[22]	Logistic regression (LR)	Survey of National Health and Nutrition.	In this classifier uses naive Bayes (NB), decision tree (DT), Adaboost (AB), and random forest (RF) to predict diabetes patients	Performance is good	Very complicated
11.	[23]	Decision Trees, Random Forest, AdaBoost, Naive Bayes, and XGBoost	Pima Indian Diabetes Dataset	The data standardisation, feature selection, K-fold cross-validation, and several Machine Learning (ML) classifiers are included in this comprehensive system for diabetes prediction.	Precise prediction	Cannot be used to create a web application.
12.	[24]	K-means clustering, random forest, and artificial neural network (ANN).	National Institute of Diabetes and Digestive and Kidney Diseases	For grouping, predicting, and association rule mining for diabetes, a variety of methods are employed..	Accuracy of 74.7%	A structured dataset has been chosen.
13	[25]	Logistic regression, gradient boosting, Ada boosting, Naive Bayes, K-NN, SVM, Random Forest, and artificial neural networks.	Pima Indians Diabetes Database	The accuracy, precision, F-measure, and recall of each of the eleven algorithms are investigated in this evaluation. The top three machine learning algorithms are chosen, followed by cross-validation and hyper-tuning, then employed in Ensemble Voting Classifier	Accuracy Of 86%	Extensive process
14	[26]	Decision Tree	Real-time data are gathered from the UCI repository.	In this, it is suggested that the polygenic disorder dataset from the UCI repository be used to create call tree models for categorization mistreatment LAD, NB, and Genetic J48 trees.	Swiftness process	No substitute call support systems that use test parameters from clinical trials

IV. DISCUSSION

All of the methods that have been proposed so far for diabetes prediction have concentrated on feature selection strategies and a few machine learning methods such as random forest, naive Bayes, support vector machine,

ANN,K-NN, and decision trees, with only a few features to be selected for the purposes of prediction. This is because there are only a small number of features that can accurately predict diabetes.
Throughout the course of our research into all of these articles, we have identified the following research gaps: (i) The most significant problem in the field of research is that no one has yet developed a hybrid model for predicting diabetes illness. Such a model, which will improve accuracy and make diabetes disease prediction simpler, has not been built. (ii) The vast majority of the authors eliminated missing values from the standard dataset. Because of this, the findings may be affected when the size of the dataset decreases. (iii) Based on the aforementioned investigations, a smaller number of authors have employed ANN (artificial neural network) principles, but they have achieved a lower level of accuracy. Therefore, it is necessary to construct a system that is capable of producing data that is more accurate, can be processed more quickly, and is more helpful for making predictions.

V. Conclusion

In the last section of this essay, we will highlight the fact that the use of machine learning in the field of health informatics is still in the formative stages of its development. We are able to report that very little work has been done in the area of diabetes illness prediction utilizing artificial neural networks, and there has also been very little development of hybrid models for diabetic disease prediction. In light of this, the goal of this research was to collect a variety of classification approaches and integrate them in order to achieve a higher level of accuracy, specificity, and sensitivity.

Acknowledgement

I would like to thank Dr. Kakoli Banerjee for her support and help throughout this journey.

References

[1] M. Moreb, T. A. Mohammed and O. Bayat, "A Novel Software Engineering Approach Toward Using Machine Learning for Improving the Efficiency of Health Systems," in IEEE Access, vol. 8, pp. 23169-23178, 2020, doi: 10.1109/ACCESS.2020.2970178.
[2] Beam, Andrew & Kohane, Isaac. (2018). Big Data and Machine Learning in Health Care. JAMA. 319. 10.1001/jama.2017.18391.
[3] Ferdous, Munira & Debnath, Jui & Chakraborty, Narayan. (2020). Machine Learning Algorithms in Healthcare: A Literature Survey. 1-6. 10.1109/ICCCNT49239.2020.9225642.
[4] Rahman, Atta & Sultan, Kiran & Naseer, Iftikhar & Majeed, Rizwan & Musleh, Dhiaa & Gollapalli, Mohammed & Chaabani, Sghaier & Ibrahim, Nehad & Siddiqui, Shahan & Khan, Muhammad. (2021). Supervised Machine Learning-based Prediction of COVID-19. Computers, Materials and Continua. 69. 21-34. 10.32604/cmc.2021.013453.
[5] Patel, Yash. (2021). Machine Learning in HealthCare .
[6] Shams, Mahmoud & Elzeki, Omar & Abdelfatah, Mohamed & Abou El-Magd, Lobna & Darwish, Ashraf & Hassanien, Aboul Ella. (2021). Impact of COVID-19 Pandemic on Diet Prediction and Patient Health Based on Support Vector Machine. 10.1007/978-3-030-69717-4_7.
[7] Harikumar & Mustafa, Malik & Sanchez, D. & Sajja, Guna & Gour, Sanjeev & Naved, Mohd & Jawarneh, Malik. (2021). IMPACT OF MACHINE learning ON Management, healthcare AND AGRICULTURE. Materials Today: Proceedings. 10.1016/j.matpr.2021.07.042.
[8] Mienye, Domor & Sun, Yanxia & Wang, Zenghui. (2021). Improved Machine Learning Algorithms with Application to Medical Diagnosis.
[9] Fetaji, Bekim & Fetaji, M. & Ebibi, Mirlinda & Ali, Maaruf. (2021). Predicting Diabetes Using Diabetes Datasets and Machine Learning Algorithms: Comparison and Analysis. 10.1007/978-3-030-90016-8_13.
[10] Khan, Bilal & Naseem, Rashid & Shah, Muhammad Arif & Wakil, Karzan & Khan, Atif & Uddin, M. Irfan & Mahmoud, Marwan. (2021). Software Defect Prediction for Healthcare Big Data: An Empirical Evaluation of Machine Learning Techniques. Journal of Healthcare Engineering. 2021. 1-16. 10.1155/2021/8899263.
[11] Lalmuanawma, Samuel & Hussain, Jamal. (2020). Applications of Machine Learning and Artificial Intelligence for Covid-19 (SARS-CoV-2) pandemic: A review. Chaos Solitons & Fractals. 10.1016/j.chaos.2020.110059.
[12] Scott, Ian & Carter, Stacey & Coiera, Enrico. (2021). Clinician checklist for assessing suitability of machine learning applications in healthcare. BMJ Health & Care Informatics. 28. e100251. 10.1136/bmjhci-2020- 100251.
[13] Mijwil, Maad & Salem, Israa & Abttan, Rana. (2021). Utilisation of Machine Learning Techniques in Testing and Training of Different Medical Datasets. Asian Journal of Computer and Information Systems.9. 29-34. 10.24203/ajcis.v9i4.6765.
[14] Zhao, Rui & Yan, Ruqiang & Chen, Zhenghua & Mao, Kezhi & Wang, Peng & Gao, Robert. (2019). Deep learning and its applications to machine health monitoring. Mechanical Systems and Signal Processing.115. 10.1016/j.ymssp.2018.05.050.

[15] Kumar, Naresh & Das, Nripendra & Gupta, Deepali & Gupta, Kamali & Bindra, Jatin. (2021). Efficient Automated Disease Diagnosis Using Machine Learning Models. Journal of Healthcare Engineering. 2021. 1- 13. 10.1155/2021/9983652.
[16] Arvanitis, Theodoros & White, Sean & Harrison, Stuart & Chaplin, Rupert & Despotou, George. (2021). A Method for Machine Learning Generation of Realistic Synthetic Datasets for Validating Healthcare Applications. 10.1101/2021.02.11.21250741.
[17] Kaissis, Georgios & Ziller, Alexander & Passerat-Palmbach, Jonathan & Ryffel, Théo & Usynin, Dmitrii & Trask, Andrew & Lima, Ionésio & Mancuso, Jason & Jungmann, Friederike & Steinborn, Marc-Matthias & Saleh, Andreas & Makowski, Marcus & Rueckert, Daniel & Braren, Rickmer. (2021). End-to-end privacy preserving deep learning on multi-institutional medical imaging. Nature Machine Intelligence. 3. 1-12. 10.1038/s42256-021-00337-8.
[18] Ojo, Adebola & Olanrewaju, Ahmed. (2019). Knowledge Discovery in Medical Database using Machine Learning Techniques. International Journal of Computer Applications. 178. 14-21.
[19] Rahman, Mohammed & Akhter, Nahida & Saleh, Sabbir. (2017). Health Data Mining: Machine Learning Approach. 3. 89-96.
[20] Raja Krishnamoorthi,Shubham Joshi,Hatim Z. Almarzouki,Piyush Kumar Shukla,Ali Rizwan,C. Kalpana,and Basant Tiwari. (2022). A Novel Diabetes Healthcare Disease Prediction Framework Using Machine Learning Techniques
[21] M. M. Bukhari, B. F. Alkhamees, S. Hussain, A. Gumaei, A. Assiri, and S. S. Ullah, "An improved artificial neural network model for effective dia- betes prediction," Complexity, vol. 2021, Article ID 5525271, 10 pages, 2021.
[22] Md. Maniruzzaman, Md. Jahanur Rahman, B. Ahammed, and Md. Menhazul Abedin, "Classification and Prediction of Diabetes Disease Using Machine Learning Paradigm," Health Information Science and Systems, vol. 8, 2020.
[23] M. K. Hasan, M. A. Alam, D. Das, E. Hossain, and M. Hasan, "Diabetes prediction using ensembling of different machine learning classifiers," IEEE Access, vol. 8, 2020.
[24] T. M. Alama, M. A. Iqbala, Y. Ali et al., "A Model for Early Prediction of Diabetes," Informatics in Medicine Unlocked, vol. 16, Article ID 100204, 2019.
[25] A. Mahabub, "A Robust Voting Approach for Diabetes Prediction Using Traditional Machine Learning Techniques," SN Applied Sciences, Springer, 2019.
[26] K. Dwivedi, "Analysis of decision tree for diabetes prediction," International Journal of Engineering and Technical Research, vol. 9, 2019.