

Predicting Diabetes Using External Factors

Pratiksha Dhawas
Student of B.E.[C.S.E]
Tulsiramji Gaikwad Patil
College of Engineering and
Technology, Nagpur.
pratikshadhawas10@gmail.com

Sonali Meshram
Student of B.E.[C.S.E]
Tulsiramji Gaikwad Patil
College of Engineering and
Technology, Nagpur.
meshramsonalu811@gmail.com

Siddhant Malkhede
Student of B.E.[C.S.E]
Tulsiramji Gaikwad Patil
College of Engineering and
Technology, Nagpur.
siddantmalkhede143@gmail.com

Gunjan Sahare
Student of B.E.[C.S.E]
Tulsiramji Gaikwad Patil
College of Engineering and
Technology, Nagpur.
gunjansahare@gmail.com

Arjun Virkhade
Student of B.E.[C.S.E]
Tulsiramji Gaikwad Patil
College of Engineering and
Technology, Nagpur.
arjunvirkhade1@gmail.com

Prof. Neha Mogre
Assistant Professor [C.S.E]
Tulsiramji Gaikwad Patil
College of Engineering and
Technology, Nagpur.
neha.cse@tgpcet.com

Abstract:

Diabetes is one of the chronic diseases in the world and many people are suffering from this. In many middle class families, the load of such expensive test affects their economy. There are several factors which are the main causes of diabetes such as consumption of bad meal, not having any physical exercise, overweight, inherited diabetes, way of person's lifestyle, hypertension. This is very critical disease as it gives rise to many other dangerous diseases like cardiovascular diseases, renal failures, brain problems, eye problem, peripheral neuropathy and many other diseases. Nowadays, many doctors are researching in this field, they have found so many methods like medication, vaccination, AI pancreas technology, testing regimes, etc. Basically diabetes is only the condition of body which is related to improper energy levels, it is also related to Sugar level. As we know that the science plays very important role in healthcare domain. Hospitals and medicals maintain the large amount of databases for future research purpose. By using data science technique, we can analyze major data sets and we can also discover cryptically data and design that will help us to find required information from the huge information and foretell an outcome. In the prevailing models, the diabetes is not getting predicted with high accuracy. In the following paper, we have introduced model for prediction of having diabetes which is preferable for the structured classification of this diabetes that involve some regular external elements. Those elements are responsive for diabetes are Level of glucose, Body Mass Index, age, body Insulin level, number of pregnancies. A pipeline model for diabetes prediction is implemented to improve the reliability of categorization and classing.

Keywords

Diabetic Mellitus (DM), diagnosis, Non-Communicable Disease, Insulin-Dependent Diabetes Mellitus, Gestational Diabetes, Cardiovascular, Peripheral Neuropathy.

I. Introduction

Health service centres has tremendous data. All these records carry history data of the patients treatment which is useful in their future treatment also. This data may be structured, unstructured or semi structured. Diabetes is becoming one of the severe disease mostly in the underdeveloped and developing countries. Many diabetic patients die due to not getting proper diagnosis at a time. It is Non Communicating Disease so does not affect others. Every year, 25 lakh people are dying due to diabetes.

Doctors often use full name for diabetes which is termed as diabetes mellitus.

It is classified as in three types which are as follows-

- IDDM (insulin dependent diabetes mellitus): It is also called as type-1. The reason behind this type of diabetes is the inability of the human's body to produce adequate insulin. Therefore, there is need to give them an external insulin.[1]
- NIDDM (non insulin dependent diabetes mellitus): It is also called as type-2. In this type of diabetes, the body cells of a patient are unable to use the insulin in proper way.[1]
- Gestational diabetes: It is also known as type-3. This type diabetes is caused due to high blood sugar level during pregnancies, where diabetes is not diagnosed earlier.[1]

For a diabetic person, there are too much ventures regarding health issues.

II. Literature Review

K. Vijiya Kumar Et Al. worked on diabetes prediction system that predicts the chances of having diabetes using machine learning models. He has used various models in his diabetes prediction system which has given results with effectiveness.[10]

Nonso Namako Et Al. worked on diagnosis of diabetes depending on supervised learning. In this project, five different classifiers are used for getting perfect results about prediction of diabetes. He has worked on diabetes onset for prediction.[11]

Talha Mahboob Alam, Muhammad Atif Iqbal has proposed model for early prediction of diabetes. In this model, association rule mining is used for getting required information. In this model, structured dataset is taken. This model is used to detect type 1 and type 2 diabetes.[18]

Charanjeet Dadiyala, Neha Mogre, Priyanka Mogre, S. Kiran Kumar proposed a Data Mining Techniques for Social Media Analysis for better prediction on consumption of a product, better marketing strategies, better placement of product and for well forecasting of future trends. The SMA tool can also be used to develop new data mining and new machine learning algorithm for social networks.[4]

Tejas N. Josi Et Al. proposed model for diabetes prediction using three machine learning methods. This project is effective for predicting diabetes before the detection of this disease. In this model, Service Vector Machine is used.[12]

Dr. Kayal Vizhi, Aman Dash presented a project on diabetes prediction using models of machine learning. In this project, cloud computing and internet of thing are also used for the information of their diabetic patients.[20]

Dhiraj shety Et Al. designed a diabetes detection system that gives results on the basis of database of diabetic patients. For the analysis of diabetic patients, they used two algorithms such as Bayesian and k nearest neighbor for analysis of elements they have taken for foretelling chances of having diabetes.[13]

Muhammad faisal faruque, Asad Uzzaman, Iqbal h. sarkar Et Al. propose diabetic study for prediction with different six algorithms for better performance in prediction. In this project classification process is much improved.[14]

III. Methodology

Understanding the Problem: In this step, the main aim is to get the reason behind increasing rate of diabetic patients. The factors which are responsible for having diabetes are physical exercise, bad meal, consumption of adequate amount of sugar. So, the main objective of this project is to develop a model which can predict chances of having diabetes in future.

Collection of Data and its Description: We have collected data of Indian diabetic patient because in India the rate of diabetes is increasing day by day. We have collected data from Kaggle site with eight attributes as input and other one attribute is for output of prediction. We considered data of 769 patients in our database.[1]

Data preparation: The dataset with 769 patient's record is examined first. Then all the noisy data and faulty data are labeled as empty dataset. In data preprocessing, four main steps are involved, which are data integration, data cleaning, data transferring and data reduction are also involved. By data preparation we get required data from huge unstructured dataset.

Data modeling: Data mining is used for better prediction of chances of having diabetes in future and it also improved classification. Here we have used five models to get higher accuracy in prediction. As we use five models, that has increased the effectiveness of this project.

Models used are:

- ❖ **Random Forest Classifier:** it is a supervised learning model and used for both regression and classification. In this classifier, numbers of decision trees are there on data samples and get prediction from each decision tree and then this classifier selects best solution. It is highly flexible and having high accuracy.
- ❖ **Decision Tree Classifier:** It is a tree structured classifier, which contains internal node representing features of data set, branches for decision rules and leaf node for outcome. It has capability of breaking complicated problems into smaller and simpler problems that can be easily solved. It requires less data cleaning and useful for solving decision related problems.
- ❖ **Logistic Regression:** It is based on supervised learning used for calculating probability of a outcome variable in prediction. The nature of this is Dichotomous which means it has two possible classes in outcome.
- ❖ **K Neighbors Classifier:** It is based on supervised learning domain. The accuracy of this classifier increases with the increase in number of dataset in our training data set of diabetic patient.
- ❖ **SVM:** it is a Server vector Machine which is best for segregation of two classes. It integrates two data points using hyper plane in n dimensional space. It is used in medical diagnosis for both the regression and classification.

It is observed that these five models are most suitable for implementation of diabetes prediction system.

IV. Proposed System

For getting more accurate probability of having diabetes, we have proposed a diabetes prediction model which is based on data science techniques which takes eight input values of external factors

which are number of pregnancies, glucose level in mg/dL, blood pressure in mmHg, skin thickness in mm, insulin level in uU/mL, body mass index in kg/m², age and weight.

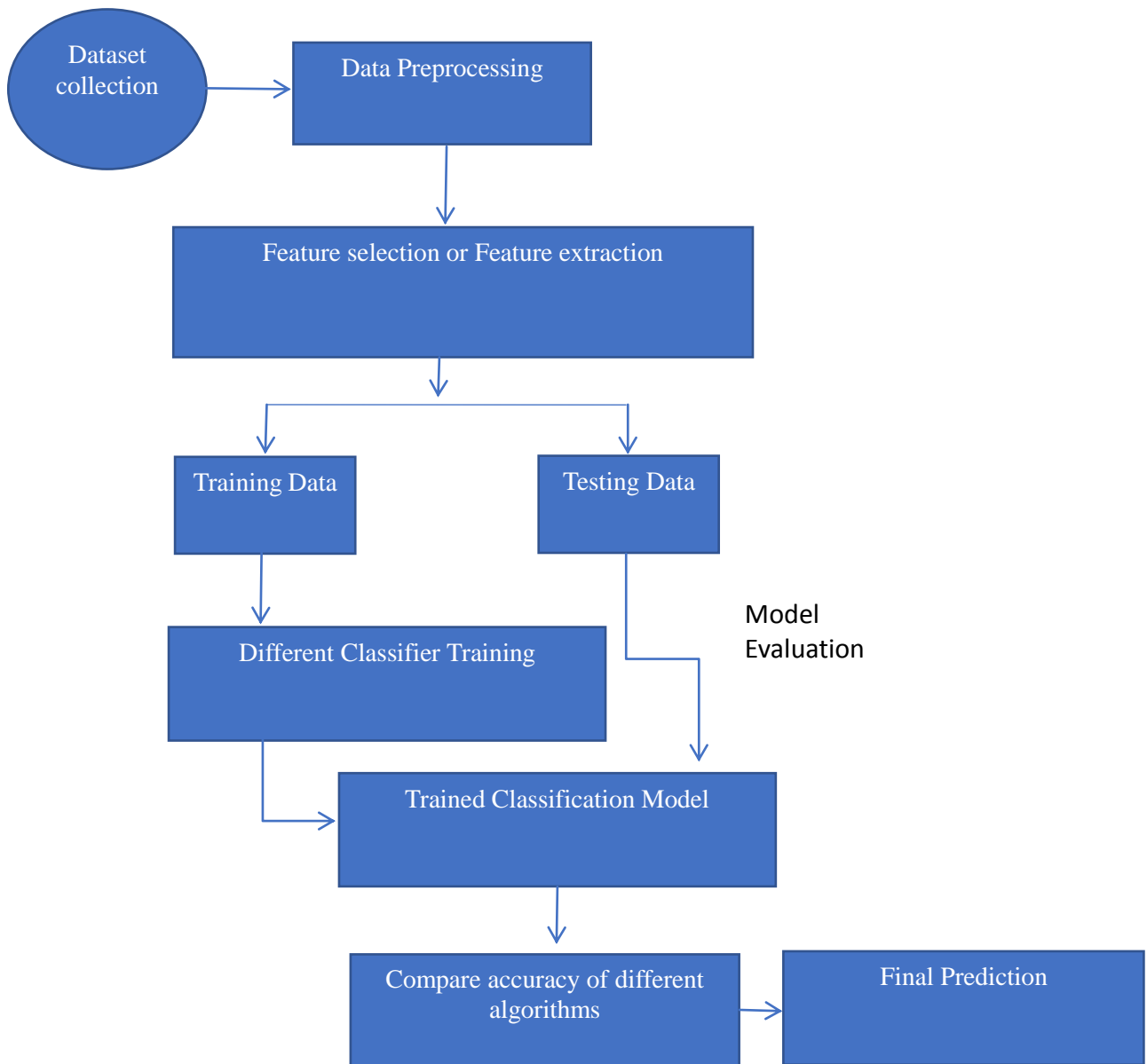
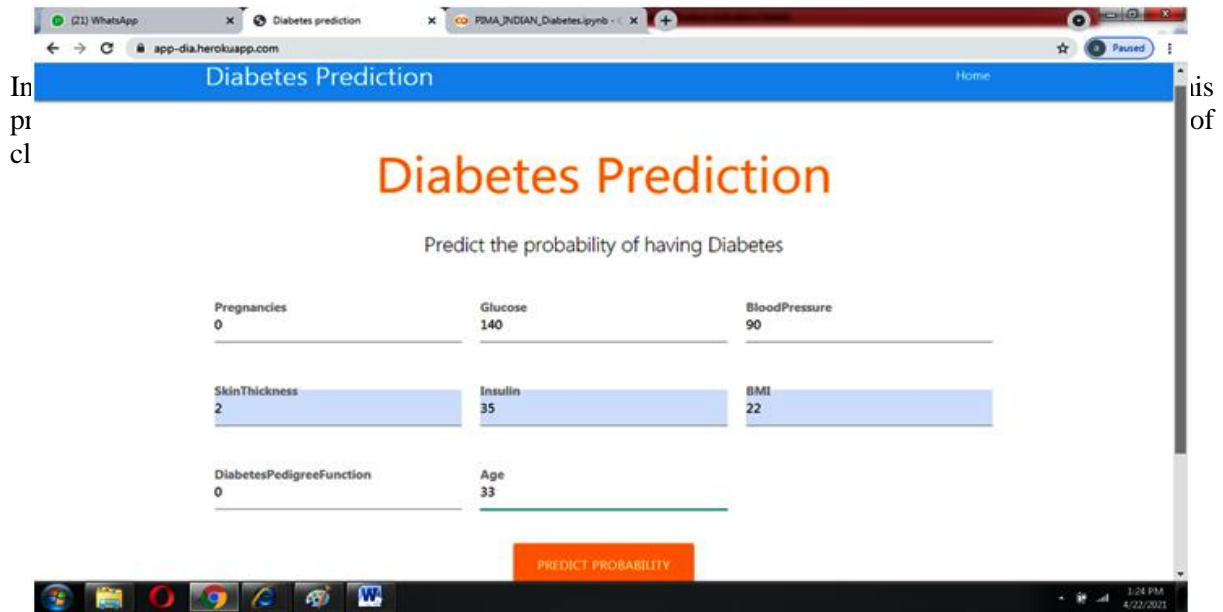
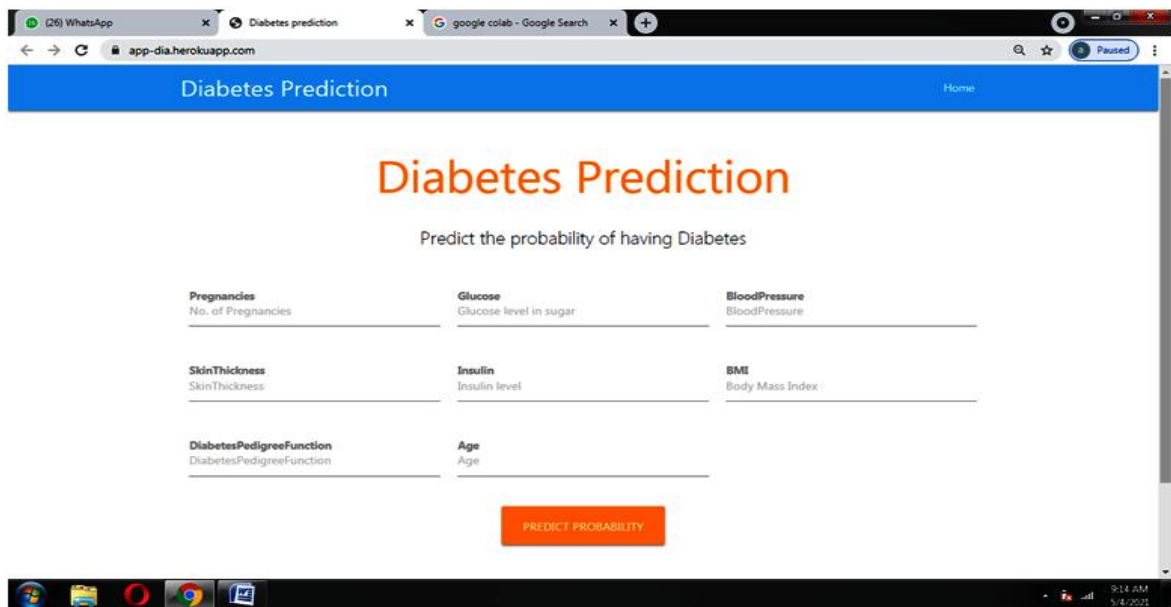


Fig 1. System Architecture



In
pr
cl

is
of



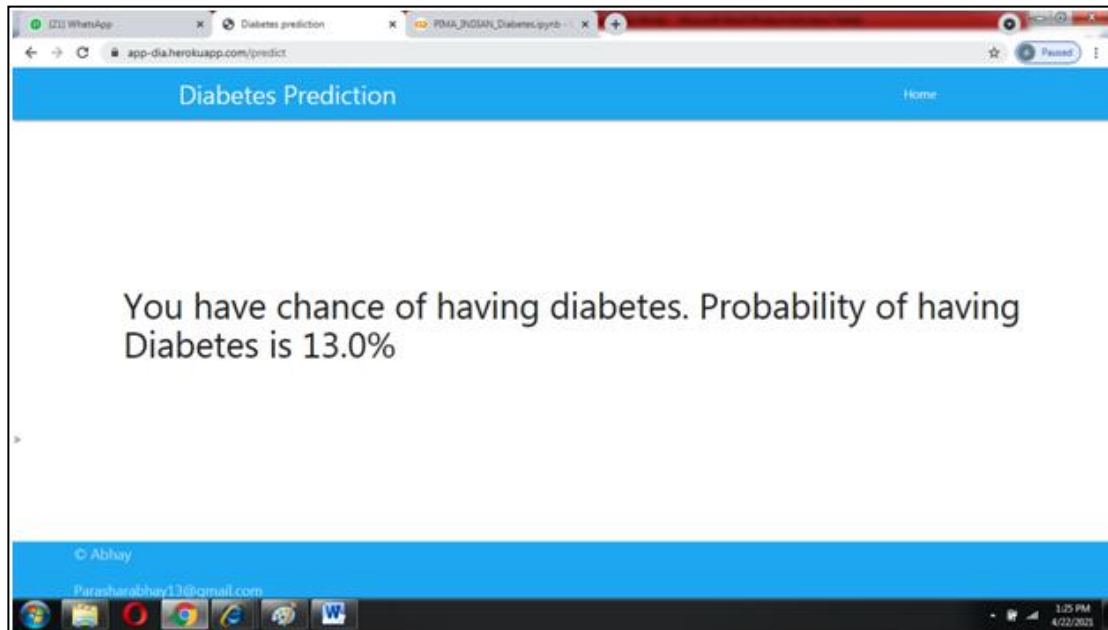


Fig 3. After prediction

VI. Advantages

1. It requires less time to get the probability of having diabetes to a person.
2. It is beneficial for the patient as it is not expensive.
3. In this project, various algorithms are used for high accuracy and to make this model better than all the existing models.
4. It helps to detect diabetes before it gets severe.
5. There is no need to wait for reports and waste whole day.

VII. Conclusion and Future Work

In this project, we have introduced a prediction model for knowing the chances of having diabetes using general regular factors. As we all know, there is a shortage of doctors in rural areas and due to this, so many patients do not get a diagnosis of diabetes. In the traditional diabetes diagnosis process, the patient has to visit a clinic to do the tests, and they have to wait a whole day for getting reports. So, our main purpose of this project is to avoid such expensive tests for such checkups.

For getting accurate decision, we have imposed pipeline and five models due to which we are getting accuracy in classification and also in prediction.

Data mining is used to deal with huge data of patients collected from hospitals for future treatment of same person and for research purpose. This model has highest efficiency, accuracy and specificity. It requires least time to get result of prediction.

In future, we can extent this project to get idea about how much chances are there to have diabetes in next years to any non-diabetic person.

References

- [1] Associated Machine Learning Techniques based On Diabetes Based Predictions (2019) :- <https://ieeexplore.ieee.org/document/9065411>.
- [2] Prediction of Diabetes Risk using Machine Learning Techniques (2020) :- <https://ieeexplore.ieee.org/document/9160276>.
- [3] American Diabetes Association. (2019). Diagnosis and classification of diabetes: standards of medical care in diabetes—2019. *Diabetes Care*, 42(Supplement 1), S13-S28.
- [4] Charanjeet Dadiyala, Neha Mogre, Priyanka Mogre, S. Kiran Kumar (2019) Data Mining Techniques For Social Media Analysis: A Review. *IOSR Journal of Engineering (IOSRJEN)*. International Conference on Management Practices, Innovation & Research 2019. ISSN (e): 2250-3021, ISSN (p):2278-8719, December 2019, PP 21-26.
- [5] Yu, K. H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature biomedical engineering*, 2(10), 719.
- [6] Worachartcheewan A, Shoombuatong W, Pidetcha P, Nopnithipat W, Prachayasittikul V, Nantasenamat C. Predicting metabolic syndrome using the random forest method. *ScientificWorldJournal*. 2015;2015:581501. doi: 10.1155/2015/581501.
- [7] B.M. Patil, Durga Toshniwal and R.C. Joshi, "Association Rule for Classification of (Non-Insulin-Dependent Diabetes Mellitus) Diabetic Patients", *ICMLC '10 Proceedings of the 2010 Second International Conference on Machine Learning and Computing*, February 09 - 11, 2010.
- [8] VeenaVijayan.V, Anjali.C, "Decision Support Systems for Predicting Diabetes Mellitus –A Review", *Proceedings of 2015 Global Conference on Communication Technologies(GCCT 2015)*, 978-1-4799-8553-1/15/\$31.00 © s2015 IEEE.
- [9] DeepikaVerma , Dr.Nidhi Mishra, "Analysis and prediction of breast cancer and diabetes disease dataset using data mining classification techniques", *Proceedings of the International Conference on Intelligent Sustainable Systems (ICISS 2017),IEEE Xplore Compliant - Part Number:CFP17M19-ART*, ISBN:978-1-5386-1959
- [10] K.VijjiyaKumar, B.Lavanya, I.Nirmala, S.Sofia Caroline, "Random Forest Algorithm for the Prediction of Diabetes ".*Proceeding of International Conference on Systems Compu- tation Automation and Networking*, 2019.
- [11] Nonso Nnamoko, Abir Hussain, David England, "Predicting Diabetes Onset: an Ensemble Supervised Learning Approach ". *IEEE Congress on Evolutionary Computation (CEC)*, 2018.
- [12] Tejas N. Joshi, Prof. Pramila M. Chawan, "Diabetes Prediction Using Machine Learning Techniques".*Int. Journal of Engineer- ing Research and Application*, Vol. 8, Issue 1, (Part -II) January 2018, pp.-09-13

- [13] Deeraj Shetty, Kishor Rit, Sohail Shaikh, Nikita Patil, "Diabetes Disease Prediction Using Data Mining". International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), 2017.
- [14] Md. Faisal Faruque, Asaduzzaman, Iqbal H. Sarker, "Performance Analysis of Machine Learning Techniques to Predict Diabetes Mellitus". International Conference on Electrical, Computer and Communication Engineering (ECCE), 7-9 February, 2019.
- [15] Charanjeet Dadiyala, Neha Mogre, Priyanka Mogre (2019). Recommendation Systems and its Preliminary Engine Algorithms Walk-Through: A Survey. IOSR Journal of Engineering (IOSRJEN). bISSN (e): 2250-3021, ISSN (p): 2278-8719, PP 09-22.
- [16] Charanjeet Dadiyala, Neha Mogre, Priyanka Mogre (2020). Analytical Study of Various Machine Learning Algorithms for Sentiment analysis of Social Media Data. Vol. 13, No. 2, (2020), pp 2057-2068.
- [17] Chen, P. H. C., Liu, Y., & Peng, L. (2019). How to develop machine learning models for healthcare. Nature materials, 18(5), 410.
- [18] Talha Mahboob Alam, Muhammad Atif Iqbal (2019) A model for early prediction of diabetes. Informatics in Medicine Unlocked, Volume 16, 2019, ISSN-100204
- [19] Nitin P. Jagtap, Kanchan A. Patil, Shaziya Sayyed Shakil and Nitin S. Ingle (2015), Mobile Activity Monitoring System Using Android Spy, International Journal of Advanced Research in Computer and Communication Engineering.
- [20] Dr. Kayal Vizhi, Aman Dash. (2020). Diabetes Prediction Using Machine Learning. International Journal of Advanced Science and Technology (IJAST), volume- 29 no.-6, ISSN (e) 2842, ISSN (p)- 2852.