ISSN: 2321-2152

IJMECE

International Journal of modern
electronics and communication engineering

E-Mail
editor.ijmece@gmail.com
editor@ijmece.com

www.ijmece.com

ISSN2321-2152

www.ijmece .com

Vol 12, Issue 2, 2024

Diabetes Disease Prediction Using Machine Learning Algorithms

[1]MR.CH SURESH, [2]KOMMULA LAKSHMAN, [3]ADI VENKAT ESWAR

[1](Assistant Professor), MCA, Swarnandhra College

[23]MCA, scholarS, Swarnandhra College

ABSTRACT

Even though diabetes is one of the most prevalent illnesses globally, it is curable and preventable if caught in its early stages. Based on certain diagnostic metrics in the dataset, we create a model to predict whether a patient will acquire diabetes. We next investigate several methods to enhance the model's performance and accuracy. This article primarily makes use of logistic regression and conducts its analysis utilizing Python integrated development environments (IDEs). A dataset from Vanderbilt based on a study of rural African Americans in Virginia and the PIMA Indians Diabetes dataset from the National Institute of Diabetes and Digestive and Kidney Diseases are the primary data sources used in the experiment. There are two distinct approaches to function selection. To top it all off, we apply aggregation approaches, which boost speed by making more accurate predictions using only one model. The original datasets and datasets created later using feature selection and aggregation approaches are both documented in terms of accuracy and runtime. In addition, every example is accompanied with a comparison. For dataset 1, the highest accuracy achieved was about 78% when the aggregation approach Maximum Voting was used. For dataset 2, the highest accuracy was around 93% when the combined procedures of maximum polling and stacking were used. When it comes to developing predictive models, logistic regression is among the most successful algorithms.

1.INTRODUCTION

Diabetic, a condition characterized by elevated blood glucose levels due to insufficient or nonexistent insulin, is one of the most infamous illnesses that has recently swept the globe. Predicting and detecting this illness may be time-consuming and frustrating since there are so many factors that need to be considered for a person to be infected. The good news is that early detection is well within the realm of possibility. Federal - Israel Defense Forces. In nations where the median income was between 0 and 7 percent, 79 percent of the adults were residing. Diabetes will affect almost 700 million people by 2045, according to estimates (IDF).

Due to both hereditary and environmental causes, the prevalence of diabetes is steadily rising over the globe. There are a lot of reasons why these figures are quickly increasing, such as people eating unhealthy food and not getting enough exercise. In diabetes, a hormonal illness, a person's blood glucose levels rise because their body is unable to make insulin, which leads to improper sugar metabolism. Severe thirst, intense hunger, and the need to urinate often are among the noticeable symptoms. The condition is influenced by several risk factors, which include age, body mass index (BMI), glucose levels (GLUT), blood pressure (BP), and so on.

Every year, we see an increase in the number of cases, and the rate of new cases is also increasing. Concern about this matter is of the utmost importance, since diabetes has rapidly emerged as a leading cause of death worldwide.

The widespread availability of big data and the subsequent need to extract useful insights from it have contributed to machine learning's recent meteoric rise in popularity. While there are many other kinds of Machine Learning algorithms, the two most common are

Non-Trained and Unlabeled Data: This is the Domain of Unsupervised Learning. To discover trends, if any, we simply put the data into action here.

In supervised learning, we use preexisting labels to train the model, and then we use those labels to assess how well the model performs on fresh data.

There have been major challenges with its identification in the past, but with the advent of Machine Learning and related techniques, these issues may be simplified while still producing thorough and precise results. It is now known that Machine Learning along with the medical field has grown even more useful and successful. By analyzing a person's traits, machine learning may soon allow for the early diagnosis of a disease. Such preliminary efforts may lead to disease inhibition and prevent the illness from progressing to a critical stage. Early detection and treatment of diabetes is the goal of the study detailed in this article, which use machine learning algorithms to make such predictions.

2.LITERATURE SURVEY

Primary Stage of Diabetes Prediction using Machine Learning Approaches—IEEE International Conference on Artificial Intelligence and Smart Systems Issue Date: 12.April, 2021 [1]A Machine Learning Approach to Diabetes Prediction

[2] Thanishka 4, Dhanush Murthy 3, Viswanatha V1, and Ramachandra A.C. The author is an associate professor at India's Nitte Meenakshi Institute of Technology in the field of electrical and communication engineering in Bangalore. 2. Professor at the Nitte Meenakshi Institute of Technology in Bangalore, India, Department of Electronics and Communication Engineering 3,4Student at India's Nitte Meenakshi Institute of Technology (EMIT) in Bangalore, studying electrical and communication engineering

1Referring to this work: Viswanatha V.

Here is the source: [1] https://www.researchgate.net/publication/350849420.
The second source is the following:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10107388/...
Review of Related Literature

Millions of individuals all around the globe deal with the long-term effects of diabetes. Treatment efficacy depends on prompt diagnosis. Diabetes prediction systems based on machine learning have been the subject of prior study. In order to diagnose diabetes, Singh et al. [4] used Random Forest, Multilayer Perceptron, and Naive Bayes. Researchers Kavakiotis et al. [5] looked at data mining techniques as they pertained to diabetes research. Diabetes was detected using an artificial neural network model by Jerjawi et al. [6]. Using ML algorithms, Maniruzzaman et al. [7] were able to forecast and categorize cases of diabetes.

Our research expands upon these previous efforts by zeroing in on diabetes risk assessment in its early stages. We preprocess data using a dataset obtained from Bangladesh's Sylhet Diabetes Hospital. Eleven classifiers, such as Support Vector Machine, Random Forest, and Logistic Regression, are trained and tested. Random Forest outperforms all other methods with an AUC of 98% and an accuracy of 98%. Important factors for diabetes prediction may be better identified according to this work.

To sum up, forecasting diabetes risk is greatly assisted by machine learning technologies. Feature selection methods and ensemble classifiers may be investigated in future studies to enhance performance.

Minhaz Uddin Emon, Maria Sultana Keyat, Md. Salman Kaiser, Md. Ariful Islam, Tabassum Tanha, and Md. Sabab Zulfiker's paper titled "Primary Stage of Diabetes Prediction using Machine Learning Approaches" describes how machine learning techniques can be used to predict the onset of diabetes in its early stages. Eleven distinct machine learning classifiers were used to construct prediction models in this work, which makes use of patient data collected from Sylhet Diabetes Hospital in Bangladesh. After calculating the Area Under the Curve (AUC), the Random Forest classifier came out on top with a score of 98%.

An overview of diabetes as a common chronic illness is given in the paper's beginning. In order to effectively treat the condition and avoid consequences, early diagnosis is crucial. The study's overarching goal is to enhance patient outcomes by applying machine learning to make early diabetes predictions. Related efforts in the area of diabetes prediction using machine learning are also reviewed in the publication. Research on diabetes diagnosis using machine learning algorithms is cited. These algorithms include Random Forest, Multilayer Perceptron, Naive Bayes, and others. The review emphasizes the promising future of these methods for diabetes prediction and calls for further study into them.

Methodology details the steps used to gather data, prepare it for analysis, divide it into smaller pieces, tune hyperparameters, apply classifiers, and assess accuracy. Classifiers were trained using data collected from 520 patients, which included information on symptoms and other variables. For all metrics

measured, the Random Forest classifier performed best. This includes accuracy, area under the curve (AUC), precision, recall, and f1 score.
This literature review summarizes the paper's main points, which center on the use of machine learning techniques for the prediction of diabetes in its early stages. It summarizes the study's relevance, discusses relevant literature, and describes the study's methodology and findings in depth**.**

3. EXISTING SYSTEM

They utilized the data analytics tool WEKA to forecast the occurrence of diabetes using healthcare Big Data. Various machine learning classifiers were applied to the publicly accessible dataset from UCI. They used the following classifiers: Naive Bayes, SVM, Random Forest, and Simple CART. They began by gaining access to the dataset, which they then preprocessed using the Weka tool. They then used a 70:30 train/test split to apply several machine algorithms. They skipped the cross-validation stage, which is crucial for getting accurate and optimum findings.

In addition, the scientists conducted their experiment using the Pima Indians Diabetes Database, a publicly accessible dataset. Dataset selection and pre-processing are the first steps in their approach for making predictions. They used naive Bayes, support vector machine (SVM), and decision tree classification techniques after data preprocessing. They compared the various performance measures and examined the accuracy in a comparative manner because they used several assessment criteria. Their experiment yielded an accuracy of 76.30% at its peak. They have also not used cross-validation, as mentioned in [2].

Using the Indians Pima Diabetes Dataset, the scientists suggested a neural network for diabetes illness prediction. They anticipated the result by using patterns they discovered in the data, which they achieved by using many hidden layers. Adopting a proprietary neural network with numerous partitions and a set of association weights and units, they call their algorithms ADAP. With a sensitivity and specificity crossover point of 0.76, they are now attempting to refine their findings for future us

Disadvantages

1) There are no techniques and models for analyzing large scale datasets in the existing system.

2) Currently, we are unable to work with diabetic datasets in conjunction with hospitals or other medical institutions to improve our outcomes.

3.1 PROPOSED SYSTEM

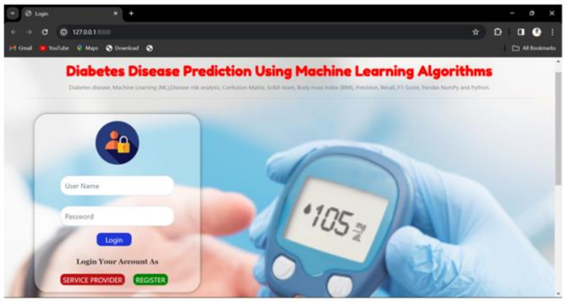
We have used the Pima Indians Diabetes Database, a publically accessible dataset, to conduct our investigation. Several diabetes diagnostic metrics are part of this dataset. The National Institute of Diabetes and Digestive and Kidney Diseases collected the dataset in the first place. Every single case that has been documented involves individuals who are older than 21. The following figure shows the five stages that make up our suggested paradigm.

Advantages

- The system more effective due to fitting datasets for different ML Models by Applying Machine Learning Algorithms.
- Machine learning may examine an individual's features in the suggested system to enable early illness identification.

4. OUTPUT SCREENS

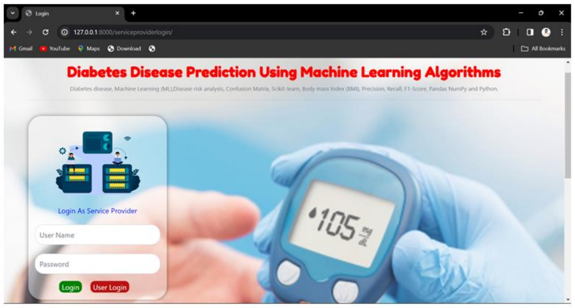
User login:



Diabetes Disease Prediction Using Machine Learning Algorithms

Diabetes disease, Machine Learning (ML),Disease risk analysis, Confusion Matrix, Scikit-learn, Body mass Index (BMI), Precision, Recall, F1-Score, Pandas NumPy and Python.

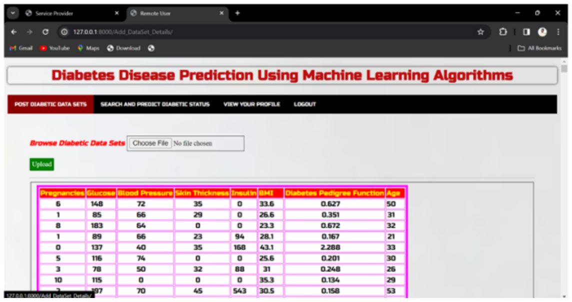
Service Provider Login Page



Diabetes Disease Prediction Using Machine Learning Algorithms

Diabetes disease, Machine Learning (ML),Disease risk analysis, Confusion Matrix, Scikit-learn, Body mass Index (BMI), Precision, Recall, F1-Score, Pandas NumPy and Python.

Remote User Dashboard

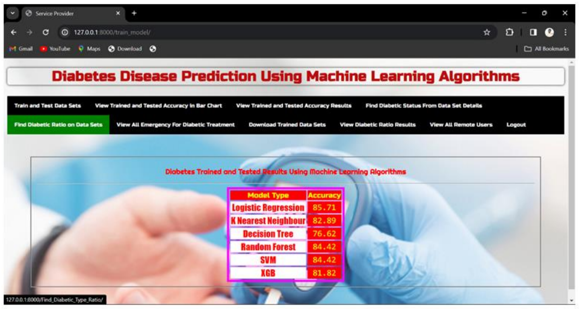


Diabetes Disease Prediction Using Machine Learning Algorithms

POST DIABETIC DATA SETS
SEARCH AND PREDICT DIABETIC STATUS
VIEW YOUR PROFILE
LOGOUT
Browse Diabetic Data Sets
Choose File
No file chosen
Upload
127.0.0.1:8000/Add_DataSet_Details/

Pregnancies	Glucose	Blood Pressure	Skin Thickness	Insulin	BMI	Diabetes Pedigree Function	Age
6	148	72	35	0	33.6	0.627	50
1	85	66	29	0	26.6	0.351	31
8	183	64	0	0	23.3	0.672	32
1	89	66	23	94	28.1	0.167	21
0	137	40	35	168	43.1	2.288	33
5	116	74	0	0	25.6	0.201	30
3	78	50	32	88	31	0.248	26
10	115	0	0	0	35.3	0.134	29
[illegible]	197	70	45	543	30.5	0.158	53

Service Provider Dashboard


Service Provider
127.0.0.1:8000/train_model/
Gmail
YouTube
Maps
Download
All Bookmarks

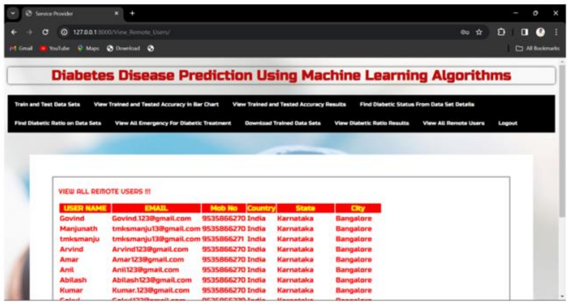
Diabetes Disease Prediction Using Machine Learning Algorithms

Train and Test Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
Find Diabetic Status From Data Set Details
Find Diabetic Ratio on Data Sets
View All Emergency For Diabetic Treatment
Download Trained Data Sets
View Diabetic Ratio Results
View All Remote Users
Logout
127.0.0.1:8000/Find_Diabetic_Type_Ratio/

Diabetes Trained and Tested Results Using Machine Learning Algorithms

Model Type	Accuracy
Logistic Regression	85.71
K Nearest Neighbour	82.89
Decision Tree	76.62
Random Forest	84.42
SVM	84.42
XGB	81.82

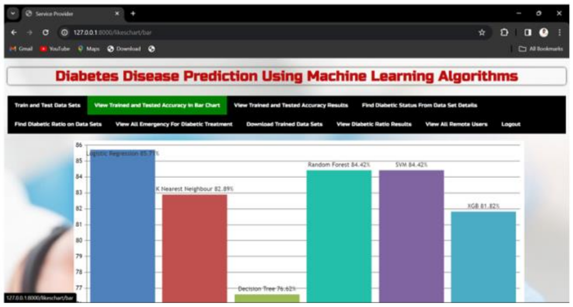
All Remote Users


Service Provider
127.0.0.1:8000/View_Remote_Users/
Diabetes Disease Prediction Using Machine Learning Algorithms
Train and Test Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
Find Diabetic Status From Data Set Details
Find Diabetic Ratio on Data Sets
View All Emergency For Diabetic Treatment
Download Trained Data Sets
View Diabetic Ratio Results
View All Remote Users
Logout

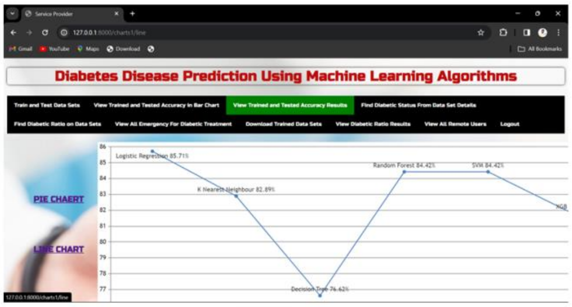
VIEW ALL REMOTE USERS !!!

USER NAME	EMAIL	Mob No	Country	State	City
Govind	Govind.123@gmail.com	9535866270	India	Karnataka	Bangalore
Manjunath	tmksmanju13@gmail.com	9535866270	India	Karnataka	Bangalore
tmksmanju	tmksmanju13@gmail.com	9535866271	India	Karnataka	Bangalore
Arvind	Arvind123@gmail.com	9535866270	India	Karnataka	Bangalore
Amar	Amar123@gmail.com	9535866270	India	Karnataka	Bangalore
Anil	Anil123@gmail.com	9535866270	India	Karnataka	Bangalore
Abilash	Abilash123@gmail.com	9535866270	India	Karnataka	Bangalore
Kumar	Kumar.123@gmail.com	9535866270	India	Karnataka	Bangalore

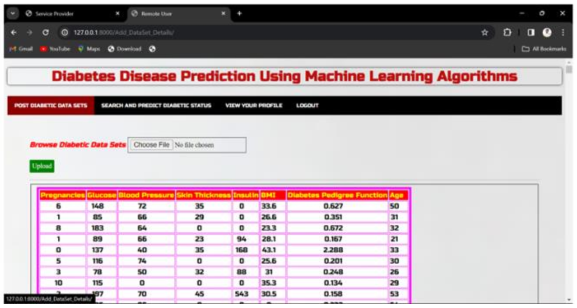
Train and Test Accuracy Bars


Service Provider
127.0.0.1:8000/likeschart/bar
Diabetes Disease Prediction Using Machine Learning Algorithms
Train and Test Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
Find Diabetic Status From Data Set Details
Find Diabetic Ratio on Data Sets
View All Emergency For Diabetic Treatment
Download Trained Data Sets
View Diabetic Ratio Results
View All Remote Users
Logout
86
85
84
83
82
81
80
79
78
77
Logistic Regression 85.71%
K Nearest Neighbour 82.89%
Decision Tree 76.62%
Random Forest 84.42%
SVM 84.42%
XGB 81.82%

Accuracy Results


Service Provider
127.0.0.1:8000/charts1/line
Diabetes Disease Prediction Using Machine Learning Algorithms
Train and Test Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
Find Diabetic Status From Data Set Details
Find Diabetic Ratio on Data Sets
View All Emergency For Diabetic Treatment
Download Trained Data Sets
View Diabetic Ratio Results
View All Remote Users
Logout
PIE CHAERT
LINE CHART
86
85
84
83
82
81
80
79
78
77
Logistic Regression 85.71%
K Nearest Neighbour 82.89%
Decision Tree 76.62%
Random Forest 84.42%
SVM 84.42%
XGB

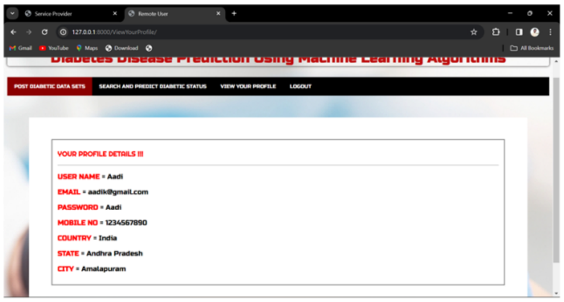
Uploaded Dataset


Service Provider
Remote User
127.0.0.1:8000/Add_DataSet_Details/
Diabetes Disease Prediction Using Machine Learning Algorithms
POST DIABETIC DATA SETS
SEARCH AND PREDICT DIABETIC STATUS
VIEW YOUR PROFILE
LOGOUT
Choose File
Upload

Browse Diabetic Data Sets No file chosen

Pregnancies	Glucose	Blood Pressure	Skin Thickness	Insulin	BMI	Diabetes Pedigree Function	Age
6	148	72	35	0	33.6	0.627	50
1	85	66	29	0	26.6	0.351	31
8	183	64	0	0	23.3	0.672	32
1	89	66	23	94	28.1	0.167	21
0	137	40	35	168	43.1	2.288	33
5	116	74	0	0	25.6	0.201	30
3	78	50	32	88	31	0.248	26
10	115	0	0	0	35.3	0.134	29
[illegible]	197	70	45	543	30.5	0.158	53

User Profile


Service Provider
Remote User
127.0.0.1:8000/ViewYourProfile/
Diabetes Disease Prediction Using Machine Learning Algorithms
POST DIABETIC DATA SETS
SEARCH AND PREDICT DIABETIC STATUS
VIEW YOUR PROFILE
LOGOUT

YOUR PROFILE DETAILS !!!

USER NAME = Aadi

EMAIL = aadik@gmail.com

PASSWORD = Aadi

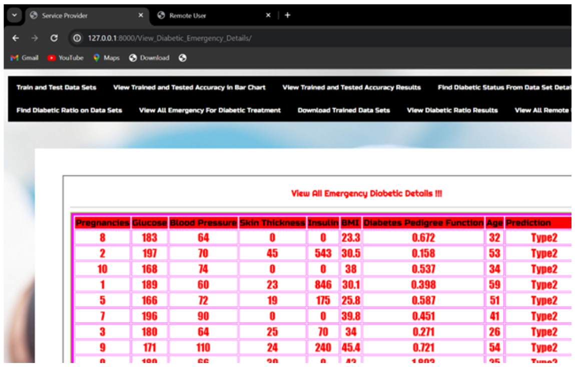
MOBILE NO = 1234567890

COUNTRY = India

STATE = Andhra Pradesh

CITY = Amalapuram

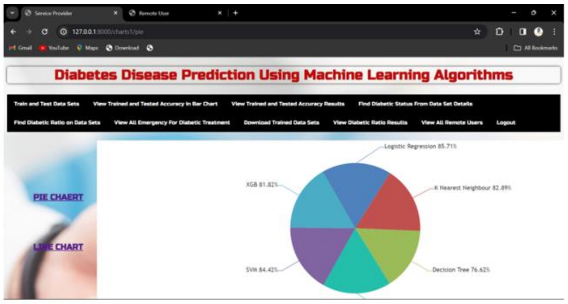
Emergency Patients Report


Service Provider
Remote User
127.0.0.1:8000/View_Diabetic_Emergency_Details/
Train and Test Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
Find Diabetic Status From Data Set Detai
Find Diabetic Ratio on Data Sets
View All Emergency For Diabetic Treatment
Download Trained Data Sets
View Diabetic Ratio Results
View All Remote

View All Emergency Diabetic Details !!!

Pregnancies	Glucose	Blood Pressure	Skin Thickness	Insulin	BMI	Diabetes Pedigree Function	Age	Prediction
8	183	64	0	0	23.3	0.672	32	Type2
2	197	70	45	543	30.5	0.158	53	Type2
10	168	74	0	0	38	0.537	34	Type2
1	189	60	23	846	30.1	0.398	59	Type2
5	166	72	19	175	25.8	0.587	51	Type2
7	196	90	0	0	39.8	0.451	41	Type2
3	180	64	25	70	34	0.271	26	Type2
9	171	110	24	240	45.4	0.721	54	Type2

Pie Chart


Service Provider
Remote User
127.0.0.1:8000/charts1/pie
Diabetes Disease Prediction Using Machine Learning Algorithms
Train and Test Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
Find Diabetic Status From Data Set Details
Find Diabetic Ratio on Data Sets
View All Emergency For Diabetic Treatment
Download Trained Data Sets
View Diabetic Ratio Results
View All Remote Users
Logout
PIE CHAERT
LINE CHART
Logistic Regression 85.71%
K Nearest Neighbour 82.89%
Decision Tree 76.62%
SVM 84.42%
XGB 81.82%

5. CONCLUSION

The early diagnosis of an illness, in this instance diabetes, is one of the major hurdles to the advancement of medical technology. But in this work, researchers set out to build a model that can reliably predict when the illness may manifest. We have easily predicted this condition using the tests performed on the Pima Indians Diabetes Database. With a 76% accuracy rate utilizing the K-Nearest Neighbors classifiers, the findings further demonstrated the system's sufficiency. Be that as it may, we hold out hope that we may incorporate this model into a system that can anticipate the onset of other catastrophic illnesses. Automated analysis of diabetes, or any other condition, may one day have opportunity for improvement.

Working with a medical facility to compile a diabetic dataset is something we want to do in the future in the hopes of improving our current findings. To further improve outcomes, we will be increasing the use of ML and DL models.

6.REFERENCES

H [1] P. Saeedi, I. Petersohn, P. Salpea, B. Malanda, S. Karuranga, N. Unwin, S. Colagiuri, L. Guariguata, A. A. Motala, K. Ogurtsova, J. E. Shaw, D. Bright, and R.Williams, “Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045:

Results from the international diabetes federation diabetes atlas, 9th edition,” Diabetes Research and Clinical Practice, vol. 157, p. 107843, 2019.

[2] A. Mir and S. N. Dhage, “Diabetes disease prediction using machine learning on big data of healthcare,” in 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018, pp. 1–6.

[3] D. Sisodia and D. S. Sisodia, “Prediction of diabetes using classification algorithms,” Procedia Computer Science, vol. 132, pp. 1578 – 1585, 2018, international Conference on Computational Intelligence and Data Science.[Online].Available; http://www.sciencedirect.com/science/article/pii/S1877050918308548

[4] J. Smith, J. Everhart, W. Dickson, W. Knowler, and R. Johannes, “Using the adap learning algorithm to forcast the onset of diabetes mellitus,” Proceedings - Annual Symposium on Computer Applications in Medical Care, vol. 10, 11 1988.

[5] P. S. Kohli and S. Arora, “Application of machine learning in disease prediction,” in 2018 4th International Conference on Computing Communication and Automation (ICCCA), 2018, pp. 1–4.

[6] Wes McKinney, “Data Structures for Statistical Computing in Python,” in Proceedings of the 9th Python in Science Conference, St´efan van der Walt and Jarrod Millman, Eds., 2010, pp. 56 – 61.

[7] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del R’ıo, M. Wiebe, P. Peterson, P. G’erard-Marchant, K. Sheppard, T. Reddy, W. Weckesser,H. Abbasi, C. Gohlke, and T. E. Oliphant, “Array programming with NumPy,” Nature, vol. 585, no. 7825, pp. 357–362, Sep. 2020. [Online]. Available: https://doi.org/10.1038/s41586-020-2649-2

[8] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and ´ Edouard Duchesnay, “Scikit-learn: Machine Learning in Python,” Journal of Machine Learning Research, vol. 12, no. 85, p. 28252830, 2011.