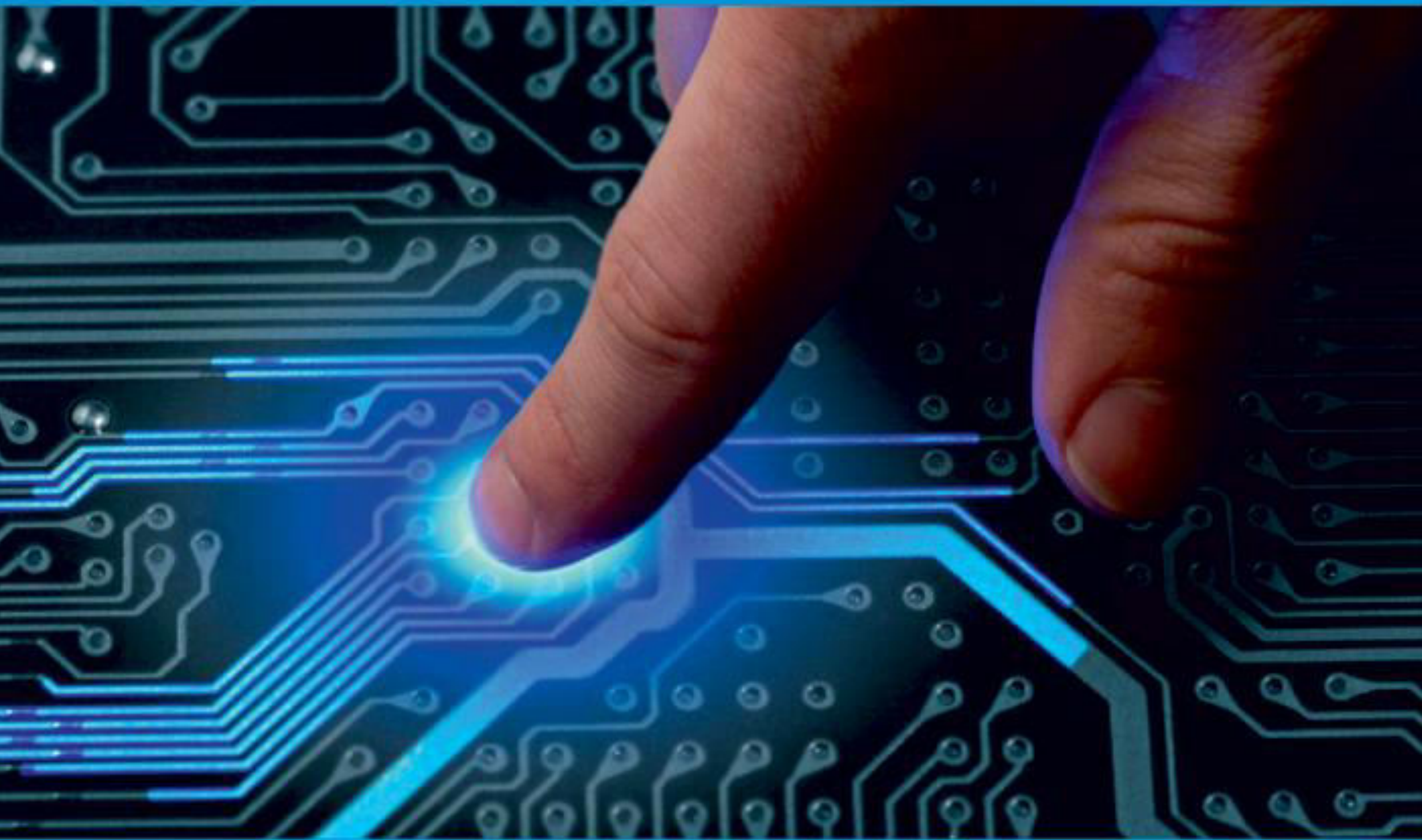
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 9, September 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.625

© 2024 IJIRCCE | Volume 12, Issue 9, September 2024| DOI: 10.15680/IJIRCCE.2024.1209028

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

A Review on Diabetes Prediction using Machine Learning Analytics

Dr. Dinesh D. Patil, Mayuri R. Chandratre

HOD, Department of Computer Engineering, Shri Sant Gadge Baba College of Engineering & Technology, Bhusaval, India

M. Tech Student (Machine Learning), Shri Sant Gadge Baba College of Engineering & Technology, Bhusaval, India

ABSTRACT: Diabetes Mellitus (sugar in blood) , is a disease that courses because of high blood sugar (Glucose in blood) levels over a long period, it requires early diagnosis to reduce its severity significantly. Nowadays, the Machine Learning (ML) community has introduced diabetes prediction and much research has been done for decades for its prediction. Keeping in view the severity of these diseases, the given paper introduces a model, named Diabetes Expert System using Machine Learning Analytics (DESMLA), exploring the diabetes data to predict the disease more effectively. The diabetes dataset is imbalanced. Therefore, the DESMLA model used the 5 most prominent, oversampling techniques namely SMOTE, Borderline SMOTE, ADASYN, KMeans SMOTE, and Gaussian SMOTE to get rid of this class imbalance problem of the diabetes dataset. DESMLA model used a Decision Tree (DT) and Random Forest (RF) as classified along with all the data preprocessing steps for diabetes prediction. The experimental results showed that the DESMLA model with KMeans SMOTE and Gaussian SMOTE performed better.

KEYWORDS: Data Mining, Machine Learning

I. INTRODUCTION

Among the head 5 countries globally, India is second with 69.20 million people with diabetes and another 36.50 million borderline diabetes [1], perilous diabetes, and circulatory system disease. Diabetes mellitus [2] also known as diabetes, is an all-over disease and has no stable treatment. The pancreas [3] produces insulin which has a significant role in regulating the blood sugar plane. There are three significant diabetes mellitus: Type 1[4], Type 2 [5], and gestational diabetes [6] [7]. Diabetes Mellitus signs differ upon how much the glucose is exalted. Type 1 diabetes occurs due to a lack of insulin. Symptoms of Type 1 diabetes are mostly severe, which include increased thirst, frequent urination, starvation, and weight loss. A person suffering from Type 1 diabetes is required to inject insulin one day. Insulin resistance causes Type 2 diabetes and is occasionally combined with an absolute shortage of insulin. Following a healthy lifestyle such as a nutritious diet, and proper exercise, could help to prevent diabetes mellitus. Without a prior diagnosis of diabetes, when pregnant a high sugar blood level then it leads to Type 2 diabetes mellitus.

People could make a preceding decision about diabetes mellitus by Machine Learning(ML) with the use of their everyday physical examination data. The challenges faced by the ML method were how to determine the valuable features and the accurate classifier to get highly correct conclusions. Freshly, for diabetes guesses, various ML algorithms have been used, like RF [12-13], DT [8-11], Support Vector Machine (SVM), Naïve Bayes(NB), etc. DT is one of the trendy ML methods because of its strongest match and appearance. However, RF has a greater classification power compared to DT as it generated a large number of DTs for indicator minimizing the overfitting issue. Henceforth, the model, Diabetes Expert System using Machine Learning Analytics (DESMLA) is planned to explore diabetes data to predict diabetes more effectively. With the high demand for the use of ML techniques in the medical field, an enormous amount of data is collected. The characteristics of the data play a vital part in the performance of ML techniques. Hence the characteristics of the data need to be examined before using any ML techniques. Thus, in the proposed DESMLA, Machine Learning Analytics (MLA) is used to detect diabetes using DT and RF more adeptly. In the proposed DESMLA model, the five most prominent oversampling techniques namely SMOTE [14], Borderline SMOTE [15], ADASYN [16], KMeans SMOTE [17],

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Gaussian SMOTE [18] is used to get rid of the class imbalance problem of the diabetes dataset after which attribute selection is applied using Pearson's Correlation Coefficient (PCC) and then by using DT and RF diabetes is forecasted. Finally from the inspection, it could be concluded that DESMLA with KMeans SMOTE and Gaussian SMOTE worked better than others.

Diabetes is characterized by elevated blood glucose levels, resulting from either insufficient insulin production (Type 1 diabetes), impaired insulin utilization (Type 2 diabetes), or a combination of both. Early diagnosis is crucial because untreated or poorly managed diabetes can lead to a cascade of health issues, significantly reducing the quality of life for affected individuals.

Machine learning has been a game-changer in healthcare due to its ability to analyze vast and complex datasets, extract meaningful patterns, and make accurate predictions. In the context of diabetes detection, ML algorithms can process diverse sources of data, including medical records, laboratory results, genetics, lifestyle factors, and wearable device data, to identify individuals at risk or those already affected by the disease.

This manuscript was distributed into five sections. In section 2 a survey on prediction of diabetes is performed. The Machine Learning methodologies namely DT and RF were illustrated in section 3 followed by section 4 which discussed the results, and finally, section 5 concluded.

II. LITERATURE SURVEY

Several researchers used ML methods to predict diabetes.Some of them were mentioned in this section.Alam et al. [19] applied ANN techniques and recorded a correctness of 76.82%. Canadian Primary Care Sentinel Surveillance Network and classifier Bootstrap aggregating, Adaptive Boosting, and DT were used by Perveen et al. [20] and they found that Adaboost could predict diseases and gave better correctness. Sisodia et al. [21] showed the comparison of SVM, NB, and DT using PIDD and finally concluded that NB was the better classified with 76.86% correctness. After reducing of's dimensionality of PIDD, Sivaranjani et al. [22] used SVM and RF to detect diabetes. Tigga et al. [23] used logistic regression on PIDD and, found the count of pregnancies, level of glucose, and BMI as extremely important. In Diwani et al. [24] Naive Bayes and DT are trained by using 10-fold cross-validations. Experimentation showed that NB gave a better performance of 76.30% accuracy. Zou et al. [25] did experimentation on PIDD using RF, DT, and ANN as classifiers and Minimum Redundancy Maximum Relevance (mRMR) and PCA methods as feature reduction procedures. From the experimentation, it is observed that RF with the mRMR feature reduction method is giving the best performance with 77.21% correctness.Kandhasamy et al. [26] compared J48, SVM, RF, and K- Nearest Neighbors (KNN). Inspections were done in two procedures, one by preprocessing and the other without preprocessing using a 5-foldcross-validation. Yuvaraj et al.

[27] Used RF, DT, and the Naïve Bayes for predicting diabetes. After using this Information Gain method, the relevant features, they used the classifier for prediction and found that the RF is giving the highest correctness. Boruah et al.

[28]

[29] Were proposed a way to find risk factors of Parkinson's disease by using DT. The rules generated from DT were processed to find the important factor, which was/were the main cause of the disorder. An enhanced model was forwarded by the new Tafa et al. [29] for predicting diabetes using SVM and NB for the data set acquired from three distinct locations in Kosovo which consisted of 402 patients out of which a total of 80 was diagnosed with diabetes of Type 2 form. The dataset is comprised of eight attributes. The proposed approach has enhanced to 97.6% which was much better than SVM and Naïve Bayes. Khanam et al. [30] used 7 ML algorithms on PIDD to detect diabetes and concluded that Logistic Regression and SVM worked better in prediction. Boruah et al. [31] put forward a methodology to predict Parkinson's disease. In the proposed approach, the dataset was firstly treated for class imbalance problems using Borderline SMOTE, Safe-Level SMOTE, and SMOTE, and then by using DT Parkinson's disease was caught. From the inspection, Borderline SMOTE with DT was given the best accuracy and thus it is further processing to find the risk factor of Parkinson's disease.

© 2024 IJIRCCE | Volume 12, Issue 9, September 2024| DOI: 10.15680/IJIRCCE.2024.1209028

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

III.THE PROPOSED METHOD DESMLA

The proposed model Diabetes Expert System used Machine Learning Analytics (DESMLA) consisting of data preprocessing and classification as 2 of its steps. In data preprocessing, the dataset was first preprocessed then the model was trained using the DT and RF. The workflow associated with the proposed .

DESMLA is shown in the figure 1.

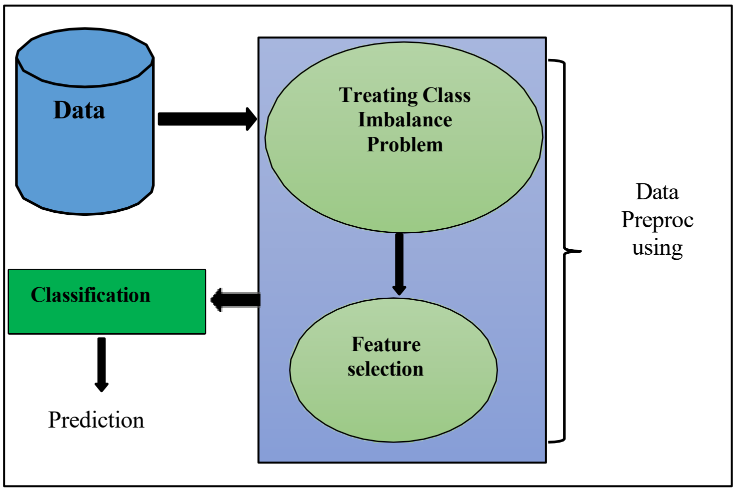


Fig.1]Workflow diagram of DESMLA

A. Data Preprocessing
In this step, the data was analyzed and it was preprocessed to balance the dataset and to select the feature set. This step is subdivided into 2 sub-steps: class balancing and feature selection.

1. Treating Class Imbalance Problem
If one of the classes was extremely high compared to the other classes present in the dependent variable then it was termed as the class imbalance problem in Machine Learning(ML). Which means there was a bias towards the majority class present in the dependent variable. Fraud detection, medical diagnosis, and e-mail classification were areas where such data could be found. Hence, to have a proper prediction of diabetes, class imbalance must be rectified. There was an assumption of even data distribution within classes in ML algorithms. The extensive issue in the class imbalance problem was that the algorithm would not learn the patterns in the minority class as it did not have enough data leading to high misclassification errors for the minority class.

To rectify the class imbalance problem, the proposed DESMLA used SMOTE techniques namely, borderline SMOTE, ADASYN SMOTE, Means SMOTE, and Gaussian SMOTE:

1) **SMOTE:** SMOTE stands for Synthetic Minority Oversampling Technique. The synthetic points were created for data augmentation depending on the original data points. The main advantage of using SMOTE was in the creation of different simulated data points than the original points of data.
2) **Borderline SMOTE:** Borderline-SMOTE generated simulated data between the two classes along the decision boundary.
3) **ADASYN SMOTE:** ADASYN stands for adaptive synthetic oversampling which was another variation from SMOTE. ADASYN creates synthetic data according to the data density.
4) **KMeans SMOTE:** It was an effective and straightforward oversampling method based on k-means clustering and SMOTE that evades noise generation and mitigates imbalanced data in classes.
5) **Gaussian SMOTE:** Gaussian oversampling was based on the Gaussian distribution. The newly generated minority

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

samples were simulated based on the area under the Gaussian density function.

2. Feature Selection
In statistics, PCC is the bivariate correlation. A threshold of
0.08 was used for PCC and thus the attributes with PCC that were less than the threshold were removed from the dataset.

B. Classification
The proposed model DESMLA used 2 classifiers namely DT and RF.

1. **Decision Tree:** DT was a Machine Learning(ML) algorithm with a tree-like structure. The internal nodes were represented by the features while the outcome was by the leaf nodes. Thus the branches of the tree represent the decision rules.
2. **Random Forest:** RF was a collective learning and decision-making algorithm that ensemble multiple DTs from a randomly selected subset of the training set and for prediction it depended on the votes from different DTs.

IV. RESULT AND ANALYSIS

The inspection was done in the PYTHON PLATFORM 3.0 version on the Windows 10 domain. The proposed model DESMLA was used with the Pima-Indians Diabetes Dataset (PIDD), available in the UCI ML repository. A total of 768 patients' information along with their corresponding nine unique attributes were there in the dataset out of which 500 were negative and 268 were positive. After applying, SMOTE, Borderline SMOTE, K-Means SMOTE, ADASYN, and Gaussian SMOTE to the original data set the synthetic instances created were as shown in Table I.

TABLE I.NUMBER OF SYNTHETIC INSTANCES CREATED BY SMOTE, BORDERLINE SMOTE, ADASYN, KMEANS, AND GAUSSIAN SMOTE

Methods	Instances in train set class 0	Instances in train set class 1	Instances synthetically formed in the minority class
DESMLA WITH SMOTE	500	268	232
DESMLA WITH BORDERLINE SMOTE	500	268	232
DESMLA WITH ADASYN	500	268	232
DESMLA WITH KMEANS SMOTE	500	268	232
DESMLA WITH GAUSSIAN SMOTE	500	268	232

Fig.2. show the imbalanced data in the original dataset. Fig.3. shows the balanced data after the treatment of imbalanced data.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

In the next sub-step of the data preprocessing step, the correlation of the features was extracted using the PCC. Table. I showed the Pearson's correlation coefficient between input and output attributes. Depending upon the coefficient, the attributes with a coefficient lower than the threshold are removed from the dataset. Hence, skin thickness, and blood pressure, are removed from the dataset and the remaining 6 attributes are used for prediction.

TABLE II.INPUT - OUTPUT ATTRIBUTE CORRELATION

ATTRIBUTES	CORRELATION COEFFICIENTS
Glucose level	.4666
BMI	.2926
Insulin	.1305
Pregnancies	.2218
Age	.2383
Skin thickness	.0747
Blood pressure	.0650
Diabetes pedigree function	.1738

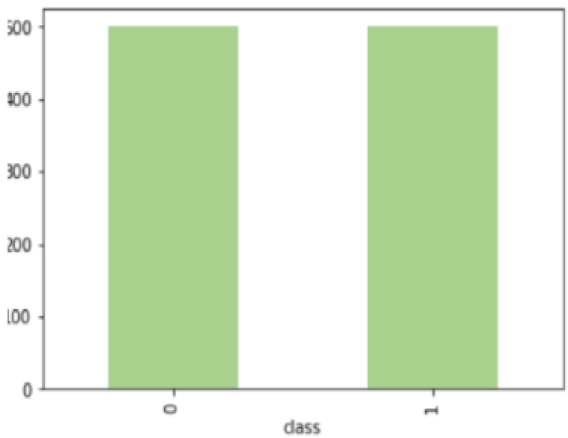


Fig. 2. Class Imbalance Problem in the Original Pima Indians Diabetes Dataset

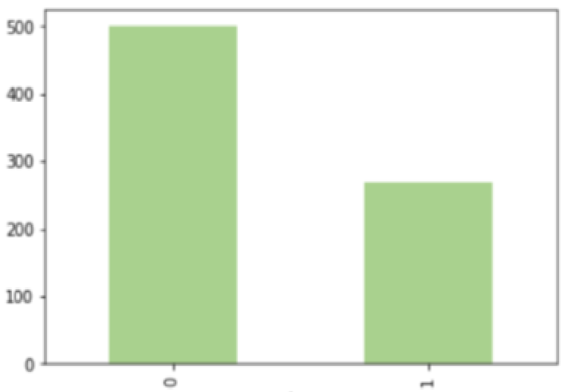


Fig.3. After Applying SMOTE, Borderline SMOTE, ADASYN, KMeans, and Gaussian SMOTE Techniques

The proposed model DESMLA was evaluated using accuracy, recall, precision, and F1 score. Table III shows

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Correctness evaluation of the proposed model DESMLA used DT and RF and by using the oversampling techniques SMOTE, Borderline SMOTE, ADASYN, KMeans, and Gaussian smote with the original data using RF and DT classifiers.

TABLE III.ACCURACY COMPARISONS

METHODS	DT	RF
With original dataset	70.12	**77.27**
DESMLA with SMOTE	65.58	**78.2**
DESMLA with Borderline SMOTE	69.48	**79.87**
DESMLA with ADASYN	67.53	**79.22**
DESMLA with KMEANS SMOTE	72.72	**81.07**
DESMLA with Gaussian SMOTE	75.97	**80.52**

From Table. III, it was seen that DESMLA with RF gives a better projection than DESMLA with DT even for imbalanced data, it was because RF was more robust than a single DT. In addition to this treating the imbalance nature reduces the bias towards the majority class. Further, the proposed DESMLA using RF with KMeans Smote gives the highest accuracy of 81.07%.

TABLE IV.PRECISION COMPARISONS

METHODS	DT	RF
With original dataset	77	**81**
DESMLA with SMOTE	72	**83**
DESMLA with Borderline SMOTE	78	**88**
DESMLA with ADASYN	75	**85**
DESMLA with KMEANS SMOTE	81	**82**
DESMLA with Gaussian SMOTE	85	**86**

TABLE V.RECALL COMPARISONS

METHODS	DTII	RF
With original dataset	76	**84**
DESMLA with SMOTE	77	**83**
DESMLA with Borderline SMOTE	73	**80**
DESMLA with ADASYN	74	**80**
DESMLA with KMEANS SMOTE	74	**90**
DESMLA with Gaussian SMOTE	76	**79**

Moderately less number of false positives and false negatives gave accurate prediction which leads to better precision and recall. Table IV showed that DEMLA with Borderline SMOTE using RF gives a high precision of 88%. Also, table V, shows that t DEMLA with KMeans SMOTE and RF gives a high recall of 90%.

The F1 score is considered a performance metric whenever there is a class imbalance problem in the dataset. The reason

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

behind it was, that model predicts correctly for a majority class (no diabetes in this case). That is why the F1 score is used as the evaluation metric.

TABLE VI.F1 SCORE COMPARISONS

METHODS	DT	RF
With original dataset	76	**84**
DESMLA with SMOTE	77	**83**
DESMLA with Borderline SMOTE	74	**83**
DESMLA with ADASYN	75	**84**
DESMLA with KMEANS SMOTE	77	**86**
DESMLA with Gaussian SMOTE	80	**82**

From Table VI, is seen that DEMLA with KMeans SMOTE using RF also hasa better F1 score measure.

V. CONCLUSION

The proposed method DESMLA was to boost the accuracy of the model by using various sampling techniques to rectify the class imbalance problem of the dataset. DESMLA first treated the class imbalance problem by using SMOTE, Borderline SMOTE, ADASYN, KMeans, and Gaussian smote, and then by using DT and RF diabetes was predicted.
The proposed procedure performs better for PIDD but has not considered other crucial factors related to gestational diabetes, like family history, metabolic syndrome, the habit of smoking, some dietary patterns, lazy routines, etc. Hence in the future, more advanced classifiers could be used to produce better results using more relevant and location-oriented data.

REFERENCES

[1]Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes—2018 American Diabetes Association Diabetes Care 2018; 41(Supplement 1): S13–S27. https://doi.org/10.2337/dc18-S002
[2]K. G. Alberti and P. Z. Zimmet, "Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation", Diabet Med., vol. 15(7), pp. 539–53, 1998.
[3]J. S. Kaddis, B. J. Olack, J. Sowinski, J. Cravens, J. L. Contreras, and
J. C. Niland, "Human pancreatic islets and diabetes research", JAMA, vol. 301(15), pp. 1580–1587,2009, https://doi.org/10.1001/jama.2009.482
[4]O. Raha, S. Chowdhury, S. Dasgupta, P. Raychaudhuri, B. N. Sarkar,
P. V. Raju and V. R. Rao, "Approaches in type 1 diabetes research: A status report", International journal of diabetes in developing countries, vol. 29(2), pp. 85–101, 2009. https://doi.org/10.4103/0973-3930.53126
[5]L. Bellamy, J. P. Casas, A. D. Hingorani and D. Williams, "Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis", Lancet, vol. 373, pp. 1773–1779 2009. doi: 10.1016/S0140-6736(09)60731-5. PMID: 19465232.
[6]J. B. Meigs, R. B. D'Agostino, P. W. Wilson, L. A. Cupples, D. M. Nathan and D. E. Singer, "Risk variable clustering in the insulin resistance syndrome: The Framingham Ofspring Study", Diabetes, vol. 46, pp. 1594–1600, 1997. doi: 10.2337/diacare.46.10.1594.
PMID: 9313755.
[7]V. Anna, H. P. van der Ploeg, N. W. Cheung, R. R. Huxley and A. E. Bauman, "Socio-demographic correlates of the increasing trend in the prevalence of gestational diabetes mellitus in a large population of women between 1995 and 2005", Diabetes Care, vol. 31(12), pp. 2288–2293, 2008.
[8]P. Swain and H. Hauska, "The Decision Tree Classifier: Design and Potential." IEEE Transactions on Geoscience Electronics, vol. 15(3),

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

pp. 142-147, 1977.
[9]J. R. Quinlan, "Induction of Decision Trees", Machine Learning, 1,
pp. 81-106, 1986.
[10]S. R. Safavian and D. Landgrebe, "A Survey of Decision Tree Classifier Methodology." IEEE Transactions on Systems, Man, and Cybernetics, vol. 21(3), pp. 660-674, 1991.
[11]A. Navada, A. N. Ansari, S. Patil, and B. A. Sonkamble, "Overview of the use of decision tree algorithms in machine learning", IEEE Control and System Graduate Research Colloquium (ICSGRC), pp. 37-42 (2011). https://doi.org/10.1109/ICSGRC.2011.5991826
[12]L. Breiman, "Bagging predictors", Machine Learning, vol. 24(2), pp. 123–140, 1996.
[13]L. Breiman, "Random Forest", Machine Learning, vol. 45, pp. 5–32, 2001.
[14]N. V. Chawla, K. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique", Journal of Artificial Intelligence Research, vol. 16(1), pp. 321-357, 2002.
DOI:10.1613/jair.953
[15]H. Han, W. Y. Wang, B. H. Mao, "Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning." In: Huang DS, Zhang XP., Huang GB. (eds) Advances in Intelligent Computing. ICIC 2005. Lecture Notes in Computer Science, vol 3644. Springer,
Berlin, Heidelberg. https://doi.org/10.1007/11538059_91
[16]H. Haibo, Y. Bai, E. A. Garcia, and L. Shutao, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), pp. 1322-1328, 2008, DOI: 10.1109/IJCNN.2008.4633969.
[17]F. Last, G. Douzas, F. Bacao, "Oversampling for Imbalanced Learning Based on K-Means and SMOTE", Information Sciences, 465, pp. 1-20, 2017.
[18]P. Tingting, Z. Junhong, J. Y. Wei Wu, " Learning imbalanced datasets based on smote and gaussian distribution" [online]: available- https://www.sciencedirect.com/science/article/abs/pii/S002002551931 0187
[19]T.M. Alam, M.A. Iqbal, Y. Ali, A. Wahab, S. Ijaz, T.I. Baig, et al., "A model for early prediction of diabetes", Inform. Med. Unlocked, vol. 16, p. 100204, 2019. https://doi.org/10.1016/j.imu.2019.100204
[20]S. Perveen, M. Shahbaz, A. Guergachi and K. Keshavjee, "Performance analysis of data mining classification techniques to predict diabetes", Procedia Computer Science, vo. 82, pp. 115-121, 2016.
[21]D. Sisodia, D.S. Sisodia, "Prediction of diabetes using classification algorithms", Procedia Comput. Sci., vol 132, pp.1578–1585, 2018.
[22]S. Sivaranjani, S. Ananya, J. Aravinth and R. Karthika, "Diabetes Prediction using Machine Learning Algorithms with Feature
[23]Selection and Dimensionality Reduction," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), pp. 141-146, 2021. doi: 10.1109/ICACCS51430.2021.9441935.
[24]N. P. Tigga and S. Garg, "Predicting type 2 Diabetes using Logistic Regression", In Nath, V., Mandal, J.K. (eds) Proceedings of the Fourth International Conference on Microelectronics, Computing and Communication Systems. Lecture Notes in Electrical Engineering, vol 673. Springer, Singapore. https://doi.org/10.1007/978-981-15- 5546-6_42
[25]S. A Diwani and A Sam, "Diabetes forecasting using supervised learning techniques", Adv. Comput. Sci.: Int. Journal, vol. 3(5), 2014.
[26]Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju, and H. Tang, "Predicting Diabetes Mellitus with Machine Learning", Front. Genet., vol.9, 2018.
[27]J. P Kandhasamy and S. Balamurali, "Performance Analysis of Classifier Models to Predict Diabetes Mellitus", Procedia Comput. Sci., vol. 47, pp. 45-51, 2015.
[28]N. Yuvaraj and K. R. SriPreethaa, "Diabetes prediction in healthcare systems using machine learning algorithms on Hadoop cluster", Cluster Computing, vol.22, pp. 1-9, 2017.
[29]A. N. Boruah, S. Kumar Biswas, S. Bandyopadhyay and S. Sarkar, "Expert System to Manage Parkinson Disease by Identifying Risk Factors: TD-Rules-PD," 2020 International Conference on Computational Performance Evaluation (ComPE), 2020, pp. 001- 006, doi: 10.1109/ComPE49325.2020.9200075.
[30]Z. Tafa, N. Pervetica, and B. Karahoda, "An intelligent system for diabetes prediction," 2015 4th Mediterranean Conference on Embedded Computing (MECO), 2015, pp. 378-382, doi: 10.1109/MECO.2015.7181948.
[31]J. J. Khanam and S. Y. Foo, "A comparison of machine learning algorithms for diabetes prediction", ICT Express, vol. 7(4), pp. 432- 439, 2021.
[32]A. N. Boruah, S. K. Biswas, S.Bandyopadhyay and S. Sarkar, "An Expert System for Identification of Key Factors of Parkinson's Disease: B-TDS-PD," 2020 IEEE India Council International Subsections Conference (INDISCON), 2020, pp. 37-42, doi: 10.1109/INDISCON50162.2020.00020.