Feature Selection, Importance and Missing Value Imputation in Diabetes Mellitus Prediction

Ovass Shafi Zargar
Department of Computer Applications
Lovely Professional University
Phagwara, India
owaisfour03@gmail.com

Avinash Bhagat
Department of Computer Applications
Lovely Professional University
Phagwara, India
avinash.baghat@lpu.co.in

Tawseef Ahmed Teli
Department of Computer Applications
Govt. Degree College Anantnag
Higher Education Department, J&K, India
mtawseef805@gmail.com

***Abstract*— With the increasing rate of diabetes globally, it has become one of the major concerns for public health. To keep the disease under control there is a need to make reasonable changes in the common man's lifestyle along with safe medication. Diabetes mellitus can be of two types. While there are many classifiers for predicting diabetes, all techniques rely on the quality of the dataset and preprocessing methods. In this study, feature importance, feature selection methods, missing value handling and the effect of these techniques on the accuracy of results are discussed. The experiment is done using Random Forest and a Deep Neural Network (DNN). The results show that the classifiers perform better when feature selection and missing value imputation techniques are implemented in tandem with the RF and DNN which highlights its importance, usage and significance in the process of prediction. The results also provide a clear insight that the DNN-based classifier has significant performance improvement over the RF classifier.**

Keywords*— *Diabetes, missing value, mellitus prediction, feature selection, DNN, RF, classifier.

I. Introduction

With the change in the lifestyle of modern man, several diseases have created serious consequences and diabetes mellitus is one among them. Currently, almost 422 million persons have been the victim of diabetes mellitus which is much higher as compared to diabetes cases diagnosed in 1980. Besides almost 30 to 80 per cent of the total diabetes patients remain undiagnosed [1]. With the modern lifestyle, there has been a significant increase in obesity and due to less physical activity, the diabetes risk has increased many folds. If diabetes is left untreated, it may create some serious problems like cardiovascular diseases, kidney failures, eyesight problems and many more. These problems can be avoided if diabetes can be prevented in its early stages however the main issue is that this disease is usually asymptomatic.

Two types of the disease are Type 1 (T1DM) and Type 2 (T2DM). T1DM is not a common form of diabetes and affects less than 10% of the total diabetic cases. It is the T2DM that is widespread, affecting more than 90% of the total diabetes patients. There are multiple risk factors associated with the cause of diabetes mellitus and the most prominent ones are smoking, high cholesterol level, less physical activity, obesity, high BP, and raised blood glucose level, however, the actual links between these risk factors and T2DM is still unclear.

The rate at which diabetes is increasing worldwide is a point of concern mostly in case of youth having an age below 20 years. The most serious issue is the development of diabetes at young ages as it is very difficult to detect diabetes in young patients. Medical practitioners are facing multiple challenges in diagnosing diabetes in young generations. Another problem faced by medical practitioners around the globe is the diagnosis of pediatric diabetes which is still poor amid the advances in medical sciences. Minority groups and some tribal and ethnic groups are unequally affected by suboptimal glucose control and are exposed to an increased risk of serious and long-lasting problems of diabetes. Diabetes can be managed and kept under control by well-timed and exact classification of the disease however in the case of children it is not so easy due to increasing obesity and unstable demographic structure. Due to the increasing cases of obesity worldwide, the classification based on obesity's effects on diabetes is a big challenge. In the case of racial/tribal groups, the migration between counties or regions results in changes in the diabetes distribution in various geographical areas and hence leads to the difficult classification of the disease.

Artificial intelligence has played a crucial role in the early diagnosis of various diseases in humans and can be used as a useful tool in predicting diabetes mellitus before its onset. Extensive research has already been carried out to find the possibility of using Machine Learning (ML) and Deep Learning (DL) for predicting and diagnosing diabetes mellitus during an early stage. Machine learning algorithms have proven helpful to some extent to detect diabetes mellitus at its early stages. The ensemble of various algorithms performs even more accurately and provides promising results in the early diagnosis of various diseases. This study focuses on the application of a hybrid DL model with a significant focus on feature importance and selection as well as missing value imputation in the early detection of diabetes mellitus. This research utilised a dataset known as PIMA. This is a widely-used reference dataset for research on diabetes mellitus detection.

The literature review is discussed in section II, section III is dedicated to feature selection methods. Section IV discusses missing value imputation followed by the proposed methodology in section V. The experimentation is discussed in section VI and concluding remarks are given in Section VII.

II. Related Work

To improve the health conditions of the common man, researchers have been working hard to make it possible to diagnose various diseases at an early stage so that the control measures could be taken well in advance and the disease could be kept under control. The researchers in [1] worked on a cross-sectional work. The authors developed a DL-based model for the early prediction of diabetes mellitus. The authors developed the DLM using seven non-invasive variables. After proper comparison, the study shows that the proposed model can be used to diagnose diabetes at an early stage with more reliability as compared to other ML techniques. The authors in [2] studied the use of retina image data for diagnosing diabetes mellitus before its onset. The authors carried out a study [3–6] to use a combination of models from ML, DNN and ensemble techniques to develop a model for early diagnosis of diabetes disease. They worked on a dataset of female patients and tried to predict the possibility of the occurrence of diabetes in females during pregnancy commonly known as gestational diabetes. They used the DNA sequence to highlight the chances of occurring high glucose levels in patients in future. Disease prediction can be more effective and reliable if the feature selection method can be applied appropriately. To classify the presence or absence of diabetes mellitus, researchers in a recent study [7] developed a thorough diagnosis model based on a mix of feature selection techniques, including the Chi-Square test and recursive feature reduction approaches. The authors used only significant features that are more significant towards the detection of diabetes mellitus. The elimination of unnecessary features results in time efficiency and more accurate classification. The deadliest type of diabetes is Type 2 diabetes and many researchers have studied and contributed to the detection of TYPE 2 diabetes. The authors in [8-11] studied risk factors of type 2 diabetes. The authors explored abdominal CT biomarkers in a dataset using deep learning for the diagnosis of T2DM. Some authors [12-13] carried out the research study and tried to find the risk factor of diabetes in loss of eyesight. Due to long-lasting diabetes, the patient may develop some defects in the retina which result in permanent loss of eyesight. The authors used a dataset of retina images and applied an enhanced CNN with a GA technique to detect the early onset of diabetes mellitus. Prolonged diabetes may also cause some serious consequences and one among them is depression in adult persons [14-15]. The authors studied the commonness and health factors associated with depression among older persons. The results showed a high probability of depression in older age due to the occurrence of diabetes mellitus. In study [16-17], the authors analysed T2DM and thyroid disorders. The authors found that thyroid disorders become prevalent in patients after having diabetes for 5 years with more chances in the case of female patients. The authors [18-20] provided a review of commonly used ML algorithms for the detection of DM and also proposed an ensemble of ML algorithms for improving accuracy, reliability and efficiency in the classification of DM. In a review study [21] the authors analysed the work carried out by several authors and compared the results obtained by them to provide the comparative analysis of ML algorithms, ensemble algorithms and DL algorithms in the early prediction of DM. The study depicts that ensemble and DL algorithms provide much higher accuracy as compared to ML algorithms. The role of metabolomics has been studied by various authors for predicting diabetes mellitus. In a research study [16] authors used longitudinal metabolomics for facilitating the early prediction of diabetes mellitus. There are multiple biomarkers for diseases related to diabetes. The authors in [22-25] worked on the Advanced Glycation End Products as a diagnostic tool for the early prediction of diabetes. The authors gave a non-invasive mechanism to control diabetes at an early stage and improve the living conditions of humans. The dataset often contains missing values and outliers that may lead to biased classification. The authors in [26-27] carried out research work and used various pre-processing techniques like normalization, and data integration to remove the various anomalies from the dataset before applying various machine learning algorithms to it. They also provided a comparison of various ML algorithms with a focus on NN for the prediction of diabetes mellitus at an early stage. ML techniques encompass application areas including cryptographic networks [28-31], navigation [32] and most importantly healthcare. Deep learning techniques that have applications in secure healthcare [33-37] perform well in predictive analysis.

III. Feature Selection

Machine learning algorithms are based on the axiom "Garbage in, garbage out," where "garbage" is defined as unwanted information or data. The model is trained using a massive dataset for improved machine learning. There is a lot of noise in this massive dataset, and some of the attributes in the dataset aren't useful for categorization. Large amounts of data can stifle the model's ability to learn, and including unnecessary data can lead to an erroneous prediction.

Feature Selection is a mechanism that shrinks the input vector of our model by eliminating noisy data from the dataset and trains the model on significant attributes only. It automatically removes the irrelevant features from the dataset that do not contribute towards the classification process so that the resulting model is more accurate and learns in less time. The concept of feature selection is shown in Fig. 1.

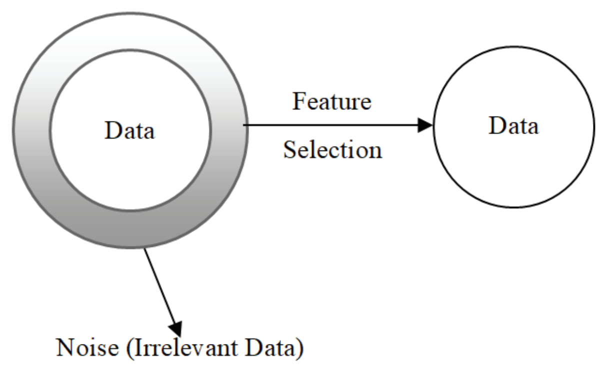

Fig. 1. Feature Selection (FS)

A. *Feature Selection Models*

The FS models are supervised and unsupervised (Fig. 2). The supervised feature selection uses labels for the output. In this method, the target variable is used to find the attributes that can enhance the performance. In the case of unsupervised feature selection models, there is no need for the output label class. The feature selection is done based on unlabelled data.

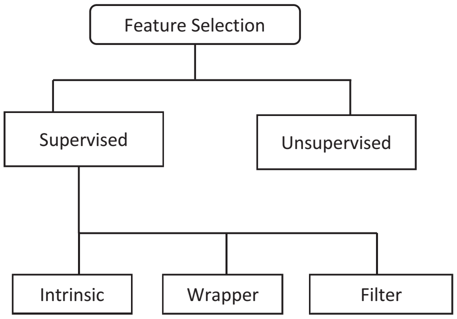

Fig. 2. Feature Selection Models

1) Filter Method

This method of feature selection works on the concept of correlation (Fig. 3). The method calculates the correlation between the dependent and independent features. The correlation checks which features are positively related to the output variables and which features are negatively related to the output variable. Based on this the positively related features are considered and negatively related features are dropped from the dataset. Some of the filter-based feature selection methods include information gain and chi-square etc.

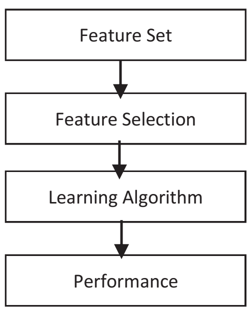

Fig. 3. Filter Feature Selection

2) Wrapper Method

In this feature selection method (Fig. 4), the data is divided into a subset and this subset is used to train the model. Based on the output of the learning model, the features are added or subtracted and the training process is restarted. This is the greedy approach for feature selection that evaluates the performance of all the possible combinations of features. Examples of Wrapper methods of feature selection are Forward Selection and Backward Elimination.

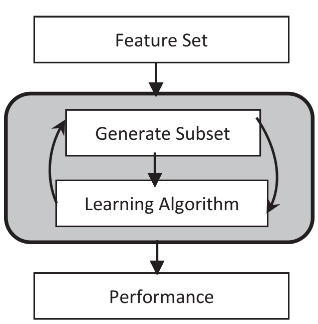

Fig. 4. Wrapper Feature Selection

3) Intrinsic Method

The Intrinsic method is the combination of Filter and Wrapper methods and utilizes both mechanisms to create the best subset of data for the learning algorithm (Fig. 5). The Intrinsic method trains the model iteratively while trying to maintain the computation cost at a minimum. Examples of intrinsic methods are Lasso and Ridge Regression

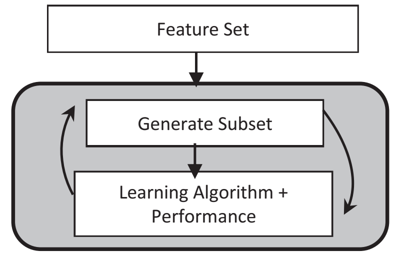

Fig. 5. Intrinsic Feature Selection

B. Choosing Feature Selection Model

The feature selection model to be used depends on whether the features are of a numerical or categorical type and is chosen as given in Table I:

TABLE I. FEATURE SELECTION LOOKUP

In Parameter	Out Parameter	Model
N	N	• Pearson's • Spearman's
N	C	• ANOVA (linear) • Kendall's (nonlinear)
C	N	• Kendall's (linear) • ANOVA (nonlinear)
C	C	• Chi-Squared test • Mutual Information

C = Categorical, N = Numerical

In this research paper, we use Spearman's rank coefficient for the feature selection.

C. Spearman's Rank Coefficient

It is based on correlation and is a nonparametric measure that is it works on finding statistical dependence of ranking between two variables and is denoted by ρ.

If the variables are related linearly then Pearson's Coefficient is more suitable for feature selection and if the variables are monotonically related then Spearman's Rank coefficient is utilized for feature selection. Spearman's Rank coefficient is calculated as given in equation 1:

$$\rho = 1 - \frac{6\sum_i b_i^2}{n(n^2-1)} \qquad (1)$$

where,

n = Count of data points between two variables

b_i = the rank change of i^{th} variable

ρ's values are in the interval [+1, -1] where +1 represents a seamless relationship, 0 represents no relationship and -1 represents a seamless negative relationship between ranks.

For Spearman's Rank Coefficient to work on must be able to rank the data. On increasing one value, the value of another variable must follow a monotonic relation.

IV. Missing Value Imputation

An experiment or observation is a statistical technique for optimizing the performance of a system with some input variables. This experiment starts with an investigational design for a trial plan with all known factors affecting the system's result. The input data collected in a well-planned experiment and under a completely controlled manner may still contain some missing data that can affect the performance of the system significantly and reduce the statistical power of the system and the system may produce biased results or sometimes inaccurate results. Dealing with the missing values is the most important challenge during an experiment. The easiest method to deal with missing values is to delete the records altogether. However, this may result in the loss of other significant information if the percentage of the missing value in the dataset is large. Another most commonly used method to handle the problem of missing values is imputation in which the values are replaced by substituted values and the system may produce more accurate and efficient results. Several imputation methods can be used for substituting missing data. Commonly used imputation methods are:

- Mean Imputation
- Regression
- K-NN
- Hot-deck

A. Mean Imputation

It is a method in which the missing value of an attribute in a specific record uses the mean of all the other values for that parameter. The mean imputation preserves the sample size of the dataset and is easy to use. The dark side of this method is that this method may reduce the variability in data which results in underestimated standard deviation and variance is given in equation (2) as:

$$\hat{y}_i = \overline{y}_h \tag{2}$$

where, $\hat{y}_i =$ imputed value of record i and $\overline{y}_h =$ sample mean of respondent data within some class.

B. Regression

The drawback of replacing missing values with some common value may cause over or under-estimated values that add bias in the classification. To avoid this problem there is an alternate method known as regression imputation. The general formula for missing values is given in equation (3):

$$\hat{y}RI(m) = d_0 + \sum_{i=1}^{2} d_i x_i + \sum_{i=1}^{2} d_{ii} x_i^2 + \sum_{i<j} \sum d_{ij} x_i x_j + \varepsilon \tag{3}$$

C. K-Nearest Neighbors Imputation

Using the K-Nearest Neighbours imputation we use the K-NN algorithm to predict the missing value using the concept of feature similarity. The missing values are predicted by searching for closed neighbours and imputing them based on K-Nearest available values.

V. Proposed Methodology

The flowchart, shown in Fig. 6, shows the proposed methodology for a DNN-based model.

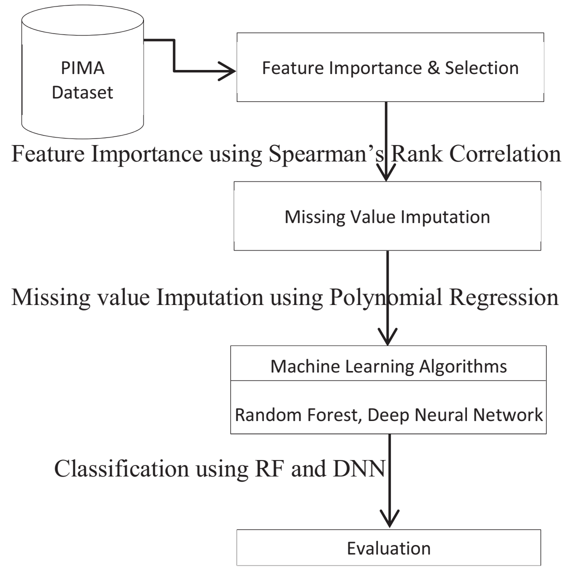

Fig. 6. Proposed Methodology

The dataset used in this study has been downloaded from Kaggle and is freely available. The dataset contains nine attributes all of which are numerical. Out of nine attributes, eight are independent and the last attribute namely class is dependent. From the dataset, the non-significant features are eliminated using the feature importance and selection method. The method used for the selection of features is Spearman's Rank Correlation. After feature selection, the dataset is observed for missing values and the missing values are imputed using the Polynomial Regression Imputation method. After the imputation of missing values, the dataset is used for the classification of diabetes mellitus using the RF and DNN. The results obtained are evaluated based on accuracy measures like F1 score, Specificity, Accuracy etc.

VI. Experiment and Results

The experiment was done on the PIMA dataset. TensorFlow and Keras library was used for the experimentation. In the first part of the experiment, random forest (RF) and the proposed DNN model was trained with no feature selection (NFS). Then the same models were trained with selected features using spearman's rank coefficient to obtain and compare the results as shown in Table II.

TABLE II. Performance (%) with and without Feature Selection

Parameter	RF	DNN	RF	DNN
	NFS	***NFS***	***WFS***	***WFS***
Precision	88.65	96.21	96.65	97.35
Recall	92.50	96.00	96.50	96.67
F1 Score	90.43	96.05	95.98	96.97
Train Acc	92.00	100.00	97.38	98.71
Test Acc	92.50	96.00	96.50	96.67

As is evident from Fig. 7., DNN based model performs better than RF for both the cases NFS and WFS. It is also clear that feature selection models have higher performance than NFS models.

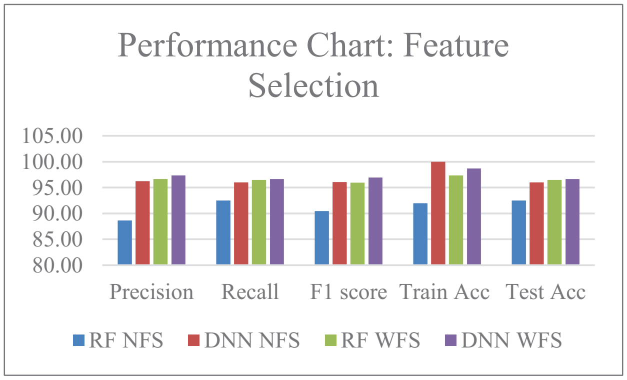

Fig. 7. Feature Selection Performance (%)

In the second part of the experiment, the proposed DNN model was trained with feature selection and the missing value imputation methods were compared to obtain and compare the results as shown in Table III.

TABLE III. PERFORMANCE (%) WITH MEAN, MEDIAN AND POLYNOMIAL REGRESSION

Parameter	Mean	Median	Polynomial Regression
Precision	97.05	97.05	98.12
Recall	96.75	96.75	97.93
F1 Score	96.76	96.76	97.95
Train Acc	98.21	98.21	98.62
Test Acc	96.75	96.75	97.93

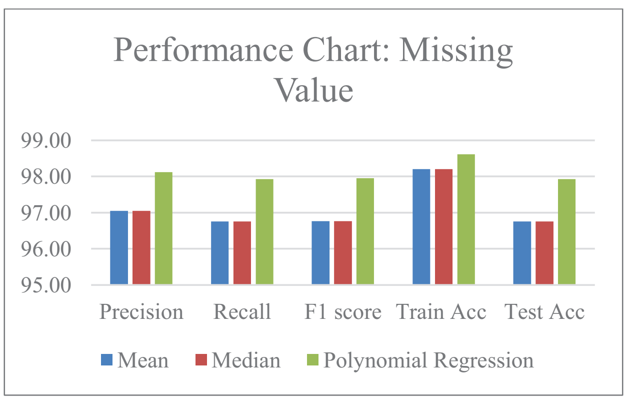

Fig. 8. Missing Value Performance (%)

DNN-based model WFS performs better with polynomial Regression, as summarized in Fig. 8 than any of the other missing value imputation techniques.

VII. CONCLUSION

The study focussed on the importance of feature selection in the diagnosis of DM. A DNN-based model was proposed encompassing all the necessary techniques at different stages of the prediction process, making it a robust and reliable framework. Random Forest and the proposed model were implemented and trained for obtaining as well as comparison of the results. It was observed that the DNN-based model performed better without and with feature selection attaining accuracy of 96.00% and 96.67% respectively. As far as the missing value imputation is concerned, the model achieves an accuracy of 97.93% using polynomial regression. This study signifies the use of feature selection and missing value imputation techniques for an early diagnosis of Diabetes Mellitus.

REFERENCES

[1] K. S. Ryu, S. W. Lee, E. Batbaatar, J. W. Lee, K. S. Choi, and H. S. Cha, "A deep learning model for estimation of patients with undiagnosed diabetes," Applied Sciences (Switzerland), vol. 10, no. 1, Jan. 2020, doi: 10.3390/app10010421.

[2] B. T. Jadhav, K. P. Mali, B. T. Jadhav, and I. K. Mujawar, "Study of Diabetic Retinopathy Detection Using Deep Learning Techniques Delopment of expert system for diabetes patients View project Study of Diabetic Retinopathy Detection Using Deep Learning Techniques." [Online]. Available: RFhttps://www.researchgate.net/publication/359760459.

[3] K. M. Kuo, P. Talley, Y. H. Kao, and C. H. Huang, "A multi-class classification model for supporting the diagnosis of type II diabetes mellitus," PeerJ, vol. 8, 2020, doi: 10.7717/peerj.9920.

[4] N. Ali, W. Khan, A. Ahmad, M. M. Masud, H. Adam, and L. A. Ahmed, "Predictive Modeling for the Diagnosis of Gestational Diabetes Mellitus Using Epidemiological Data in the United Arab Emirates," Information, vol. 13, no. 10, p. 485, Oct. 2022, doi: 10.3390/info13100485.

[5] N. E. El-Attar, B. M. Moustafa, and W. A. Awad, "Deep learning model to detect diabetes mellitus based on dna sequence," Intelligent Automation and Soft Computing, vol. 31, no. 1, pp. 325–338, 2022, doi: 10.32604/IASC.2022.019970.

[6] J. Juan et al., "Progression to type 2 diabetes mellitus after gestational diabetes mellitus diagnosed by IADPSG criteria: Systematic review and meta-analysis," Front Endocrinol (Lausanne), vol. 13, Oct. 2022, doi: 10.3389/fendo.2022.1012244.

[7] S. Hameed and F. Al-Khalidi, "Diagnosis of Diabetes Mellitus based Combined of Feature Selection Methods," 2022. [Online]. Available: www.jatit.org.

[8] K. H. Ismail, Y. Adil Hasan, and K. Hassan Ismail, "Prevalence of Type Two Diabetes Mellitus and Risk Factors in Erbil City among Adult Population Prevalence of Type Two Diabetes Mellitus and Risk Factors in Erbil City among Adult Population", doi: 10.14704/nq.2022.20. 10.NQ55221.

[9] H. Tallam, D. C. Elton, S. Lee, P. Wakim, P. J. Pickhardt, and R. M. Summers, "Fully Automated Abdominal CT Biomarkers for Type 2 Diabetes Using Deep Learning," Radiology, vol. 304, no. 1, pp. 85–95, Jul. 2022, doi: 10.1148/radiol.211914.

[10] F. Anwar, Qurat-Ul-Ain, M. Y. Ejaz, and A. Mosavi, "A comparative analysis on diagnosis of diabetes mellitus using different approaches – A survey," Inform Med Unlocked, vol. 21, Jan. 2020, doi: 10.1016/j.imu.2020.100482.

[11] V. Mayya, S. Kamath S., and U. Kulkarni, "Automated microaneurysms detection for early diagnosis of diabetic retinopathy: A Comprehensive review," Computer Methods and Programs in Biomedicine Update, vol. 1, p. 100013, 2021, doi: 10.1016/j.cmpbup.2021.100013.

[12] S. Dutta, B. C. S. Manideep, S. M. Basha, R. D. Caytiles, and N. C. S. N. Iyengar, "Classification of diabetic retinopathy images by using deep learning models," International Journal of Grid and Distributed Computing, vol. 11, no. 1, pp. 89–106, 2018, doi: 10.14257/ijgdc.2018.11.1.09.

[13] S. Ayoub et al., "Minimized Computations of Deep Learning Technique for Early Diagnosis of Diabetic Retinopathy Using IoT-Based Medical Devices," Comput Intell Neurosci, vol. 2022, p. 7040141, 2022, doi: 10.1155/2022/7040141.

[14] D. M. B. Andrade, R. M. Rocha, and Í. J. S. Ribeiro, "Depressive symptoms among older adults with diabetes mellitus: a cross-sectional study," Sao Paulo Medical Journal, Oct. 2022, doi: 10.1590/1516-3180.2021.0771.r5.09082022.

[15] A. Hashim, Z. Masood, and F. Amjad, "Frequency of restless legs syndrome in patients with diabetes mellitus and hypertension." [Online]. Available: https://www.researchgate.net/publication/364164864.

[16] M. Zhang and H. Yang, "Perspectives from metabolomics in the early diagnosis and prognosis of gestational diabetes mellitus," Front Endocrinol (Lausanne), vol. 13, Sep. 2022, doi: 10.3389/fendo.2022.967191.

[17] H. A. H. Al-Balushi et al., "Assessment of Thyroid Dysfunction among Patients with Type II Diabetes Mellitus," J Pharm Negat Results, vol. 13, no. S01, Jan. 2022, doi: 10.47750/pnr.2022.13.S01.134.

[18] S. A. Mahmoudinejad Dezfuli, S. R. Mahmoudinejad Dezfuli, S. V. Mahmoudinejad Dezfuli, and Y. Kiani, "Early Diagnosis of Diabetes Mellitus Using Data Mining and Classification Techniques," Jundishapur Journal of Chronic Disease Care, vol. 8, no. 3, Jul. 2019, doi: 10.5812/jjcdc.94173.

[19] D. D. Rufo, T. G. Debelee, A. Ibenthal, and W. G. Negera, "Diagnosis of diabetes mellitus using gradient boosting machine (Lightgbm)," Diagnostics, vol. 11, no. 9, Sep. 2021, doi: 10.3390/diagnostics 11091714.

[20] S. N. H. Bukhari, F. Masoodi, M. A. Dar, N. I. Wani, A. Sajad, and G. Hussain, "Prediction of Erythemato-Squamous Diseases Using Machine Learning,": Applications of Machine Learning and Deep Learning on Biological Data, 2023, CRC Press.

[21] L. Carvalho, A. V. B. Silva, B. J. de A. Danda, E. C. B. F. Freitas, and M. T. de S. Freitas, "Polimorfismos no gene transportador de glicose (GLUT4) associados ao desenvolvimento da Diabetes mellitus tipo 1," Research, Society and Development, vol. 11, no. 13, p. e368111335549, Oct. 2022, doi: 10.33448/rsd-v11i13.35549.

[22] F. Masoodi, M. Quasim, S. Bukhari, S. Dixit, and S. Alam, "Applications of Machine Learning and Deep Learning on Biological Data," 2023, CRC Press.

[23] D. K. Karumanchi, E. R. Gaillard, and J. Dillon, "Early Diagnosis of Diabetes through the Eye," Photochem Photobiol, vol. 91, no. 6, pp. 1497–1504, Nov. 2015, doi: 10.1111/php.12524.

[24] S. Hasib et al., "Development of a Deep Learning Model for Dynamic Forecasting of Blood Glucose Level for Type 2 Diabetes Mellitus: Secondary Analysis of a Randomized Controlled Trial (Preprint)", doi: 10.2196/preprints. 14452.

[25] M. Koga et al., "The glycated albumin to HbA1c ratio is elevated in patients with fulminant type 1 diabetes mellitus with onset during pregnancy," Journal of Medical Investigation, vol. 60, no. 1–2, pp. 41–45, 2013, doi: 10.2152/jmi.60.41.

[26] O. Shafi, J. S. Sidiq, T. Ahmed Teli, and K. -, "Effect of Pre-Processing Techniques in Predicting Diabetes Mellitus with Focus on Artificial Neural Network," Advances and Applications in Mathematical Sciences, vol. 21, no. 8, pp. 4761-4770, 2022.

[27] O. Shafi, S. J. Sidiq, T. A. Teli, and M. Zaman, "A Comparative Study on Various Data Mining Techniques for Early Prediction of Diabetes Mellitus," in Proc. of the Global Emerging Innovation Summit (GEIS-2021), 2021, pp. 51–61. doi: 10.2174/ 9781681089010121010009.

[28] T. A. Teli, R. Yousuf, and D. A. Khan, "MANET Routing Protocols, Attacks and Mitigation Techniques: A Review," International Journal of Mechanical Engineering , vol. 7, no. 2, ppp. 1468-1478, 2022.

[29] T. A. Teli, F. Masoodi, and R. Yousuf, "Security Concerns and Privacy Preservation in Blockchain-based IoT Systems: Opportunities and Challenges," in Proc. of the International Conference on IoT based Control Networks and Intelligent Systems (ICICNIS 2020), 2020.

[30] F. Masoodi, A. M. Bamhdi, and T. A. Teli, "Machine Learning for Classification analysis of Intrusion Detection on NSL-KDD Dataset," 2021.

[31] T. A. Teli, F. S. Masoodi, and A. M. Bahmdi, HIBE: Hierarchical Identity-Based Encryption. 2021. In: Ahmad, K.A.B., Ahmad, K., Dulhare, U.N. (eds) Functional Encryption. EAI/Springer Innovations in Communication and Computing. Springer, Cham. doi: 10.1007/978-3-030-60890-3_11.

[32] T. A. Teli and M. A. Wani, "A fuzzy based local minima avoidance path planning in autonomous robots," International Journal of Information Technology, vol. 13, no. 1, pp. 33–40, Feb. 2021, doi: 10.1007/s41870-020-00547-0.

[33] S. Mushtaq, A. Roy, and T. A. Teli, "A Comparative Study on Various Machine Learning Techniques for Brain Tumor Detection using MRI," in Proc. of the Global Emerging Innovation Summit (GEIS-2021), 2021, pp. 125–137. doi: 10.2174/9781681089010121 010016.

[34] T. Ahmed Teli and F. Masoodi, "Blockchain in Healthcare: Challenges and Opportunities," SSRN Electronic Journal, 2021, doi: 10.2139/ssrn.3882744.

[35] O. Shafi, J. S. Sidiq, and T. A. Teli, "Effect of Pre-Processing Techniques in Predicting Diabetes Mellitus with Focus on Artificial Neural Network," Advances and Applications in Mathematical Sciences, vol. 21, no. 8, pp. 4761-4770, 2022.

[36] T. A. Teli and R. Yousuf, "Deep Learning for Bioinformatics," in Applications of Machine Learning and Deep Learning on Biological Data, Boca Raton: Auerbach Publications, 2023, pp. 181–196. doi: 10.1201/9781003328780-11.

[37] T. A. Teli, F. S. Masoodi, and Z. Masoodi, "Application of ML and DL on Biological Data," in Applications of Machine Learning and Deep Learning on Biological Data, Boca Raton: Auerbach Publications, 2023, pp. 159–180. doi: 10.1201/9781003328780-10.