Ensemble Learning for Diabetes Prediction: An Integration of TabNet and Neural Oblivious Decision Ensembles (NODE)

Majid Rahardi
Department of Informatics, Faculty of Computer Science, Universitas Amikom Yogyakarta, Sleman, Indonesia
majid@students.amikom.ac.id (corresponding author)

Ferian Fauzi Abdulloh
Department of Informatics, Faculty of Computer Science, Universitas Amikom Yogyakarta, Sleman, Indonesia
ferian@amikom.ac.id

Ahlihi Masruro
Department of Informatics Engineering, Faculty of Computer Science, Universitas Amikom Yogyakarta, Sleman, Indonesia
ahlihi.m@amikom.ac.id

Bima Pramudya Asaddulloh
Department of Informatics Engineering, Faculty of Computer Science, Universitas Amikom Yogyakarta, Sleman, Indonesia
bima@students.amikom.ac.id

Afrig Aminuddin
Department of Information Systems, Faculty of Computer Science, Universitas Amikom Yogyakarta, Sleman, Indonesia
afrig@amikom.ac.id

Nafiatun Sholihah
Department of Informatics, Faculty of Computer Science, Universitas Amikom Yogyakarta, Sleman, Indonesia
nafiatun@amikom.ac.id

Received: 28 September 2025 | Revised: 16 October 2025 and 27 October 2025 | Accepted: 29 October 2025

ABSTRACT

The accurate prediction of diabetes risk is paramount for advancing healthcare and personalized medicine. This study presents a comparative analysis of advanced deep learning models for structured data, focusing on two novel architectures, Neural Oblivious Decision Ensembles (NODE) and TabNet. The method encompasses comprehensive data preprocessing, including a critical technique to address the imbalanced nature of the dataset (oversampling). Finally, a combined modeling approach (a soft-voting ensemble) was implemented to combine the predictive probabilities from the trained individual models. The soft-voting ensemble demonstrated strong performance, achieving a validation accuracy of 93.55, a precision of 92.60, a recall of 94.58, and an F1-score of 93.58. These findings underscore the potential of advanced deep learning techniques, especially when combined in an ensemble, to provide highly reliable and accurate diabetes risk prediction from complex tabular data.

Keywords-ensemble learning; deep learning; diabetes prediction; TabNet; NODE

I. INTRODUCTION

Diabetes is a critical chronic metabolic disorder that requires early and precise prediction to improve patient outcomes. The need for a timely diagnosis is increased due to complex complications, such as retinopathy [1], and the specific requirements for managing gestational diabetes [2]. Consequently, the development of robust machine learning prediction frameworks has become crucial to optimize healthcare resources [3]. Deep learning has emerged as an advanced non-linear prediction method, demonstrating outstanding success in various fields, such as speech analysis, aiding in the diagnosis of neurological conditions such as Alzheimer's [4], and clinical informatics, supporting complex Natural Language Processing (NLP) tasks through the development of specialized corpora [5]. Specifically, Convolutional Neural Networks (CNNs) have been utilized for the automatic detection of diabetes.

Despite promising initial efforts, the field of diabetes prediction still faces significant challenges to achieve reliable clinical generalization. Early work utilizing conventional machine learning models demonstrated strong potential: In [6, 7], algorithms such as Multilayer Perceptron (MLP), Random Forest (RF), and Support Vector Machines (SVM) achieved high classification accuracies (up to 89%), demonstrating the utility of feature selection techniques. However, a major common limitation across these foundational studies, including [8], is the heavy reliance on a single, often imbalanced dataset (e.g., PIMA Indian Diabetes dataset), which severely restricts the generalizability and robustness of the findings in diverse clinical populations. More specialized studies have explored focused prediction, such as a stacking model for Gestational Diabetes Mellitus (GDM) in [9] and the investigation of regional body composition in [10]. Although these efforts highlight the potential for personalized risk assessment, they are often constrained by limited sample sizes or the presence of measurement bias, further complicating the path to large-scale clinical deployment.

Despite promising results, there is still a significant research gap in the development of highly robust and generalizable deep learning models for the prediction of diabetes in large and complex tabular datasets. Previous studies have been limited by small sample sizes or a lack of advanced techniques to effectively address class imbalance and complex feature interactions. The literature also reveals two critical limitations in current health predictive models: over-reliance on data-specific models results in poor generalization across different clinical settings, and the persistent lack of interpretability in conventional deep learning models hinders trust and clinical adoption.

This study aimed to address these limitations, proposing a comparative analysis and soft voting ensemble between Neural Oblivious Decision Ensembles (NODE), known for its robust performance and resistance to overfitting on tabular data, and TabNet, a new architecture that inherently provides feature importance and interpretability through a sequential attention mechanism. The combination of these two leading architectures aims to achieve the improved generalization, stability, and clinical interpretability required for next-generation health predictive models. The main contribution of this study is the implementation and comparative analysis of two state-of-the-art tabular deep learning models on a large-scale dataset. The study also utilizes a soft voting ensemble strategy to improve prediction performance and model stability, offering a new, highly accurate, and generalizable framework for the prediction of diabetes risk.

II. PROPOSED METHOD

A. Dataset

The dataset contains 100.000 instances and 9 attributes representing clinical variables such as hypertension history, smoking history, body mass index, blood glucose level, HbA1c level, history of heart disease, gender, and age [11]. The target variable is a binary indicator of diabetes diagnosis. Table I shows a sample of the dataset, delineating the structure and attributes of the dataset. The proposed model is evaluated using this diabetes prediction dataset [11], a collection of medical and demographic data from patients, along with their diabetes status (positive or negative). The proposed method integrates conventional techniques with modern advances in tabular deep learning to yield accurate and dependable predictive results. Figure 1 presents the overall research flow.

TABLE I. DIABETES CLASSIFICATION DATA SAMPLE

No	diabetes	Gender	Age	hypertension	smoking_history	…	bmi	blood_glucose_level
0	No	Female	80	No	never	…	25.19	140
1	No	Female	54	No	No Info	…	27.32	80
2	No	Male	28	No	never	…	27.32	158
3	No	Female	36	No	current	…	23.45	155
4	No	Male	76	Yes	current	…	20.14	155
…	…	…	…	…	…	…	…	…
99995	No	Female	80	No	No Info	…	27.32	90
99996	No	Female	2	No	No Info	…	17.37	100
99997	No	Male	66	No	former	…	27.83	155
99998	No	Female	24	No	never	…	35.42	100
99999	No	Female	57	No	current	…	22.43	90

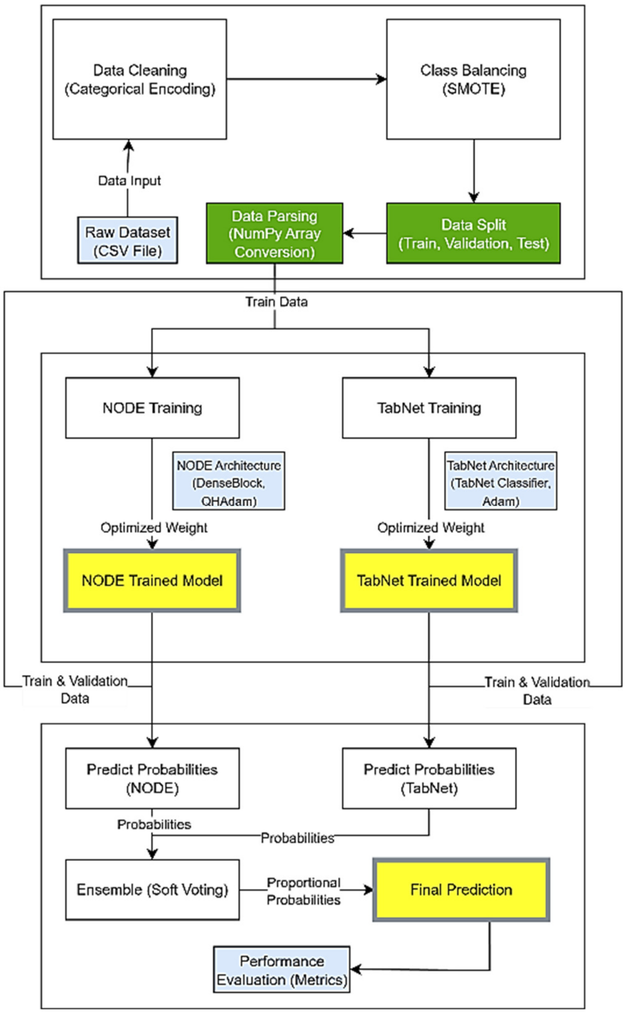

Fig. 1. The research flow diagram.

B. Data Acquisition and Preprocessing

One fundamental step in data preparation involves the encoding of categorical features into numerical representations. This process is crucial for securing healthcare information and preventing data leakage [12], while the choice of encoding method is paramount for machine learning accuracy, particularly in contexts like quantum computing applications [13]. Addressing the pervasive challenge of class imbalance was a critical component of the preprocessing pipeline. The Synthetic Minority Over-sampling Technique (SMOTE) generates synthetic samples for the minority class [14]. This method's effectiveness in improving predictive performance has been demonstrated across various diagnostic tasks [15]. A new synthetic sample x_{new} is generated from an existing minority sample x_i and one of its k-nearest neighbors x_j using the following formula:

$$x_{\text{new}} = x_i + \delta \times (x_j - x_i) \tag{1}$$

where δ is a random number between 0 and 1.

The meticulously preprocessed dataset was partitioned into training, validation, and test sets using a robust two-step random split strategy [16]. The choice of the data splitting strategy is critical, as demonstrated by its impact on predictive accuracy in various engineering applications [17]. The training set (60%) is utilized for model learning. The validation set (20%) serves as an independent dataset to examine performance during training, and the test set (20%) is reserved exclusively for the final evaluation. All resulting data subsets were explicitly converted to NumPy array format, ensuring seamless consistency and optimal compatibility for subsequent operations within the PyTorch deep learning framework.

C. Model Architectures

The proposed method leverages two distinct deep learning architectures specifically designed for tabular data: NODE and TabNet. These models represent advanced approaches to learning from structured data, differing from conventional neural networks in their ability to process and interpret tabular feature interactions [18]. The NODE architecture draws its inspiration from the principles of gradient-boosted decision trees, aiming to synthesize the interpretability of trees with the powerful predictive capabilities of neural networks [19]. This foundational concept has led to subsequent explorations and developments, including models designed for end-to-end flexible probabilistic regression on tabular data [20]. The final prediction from an individual tree or NODE layer is derived as a weighted linear combination of response tensor entries R with weights dynamically determined by the entries of the choice tensor C, as:

$$c_{l,j}(X) = \sigma_\alpha\left(\frac{g_{l,j}(v) - b_{l,j}}{\tau_{l,j}}\right) \tag{2}$$

TabNet represents another specialized deep learning architecture meticulously crafted for processing tabular data, uniquely combining the strengths derived from both tree-based methods and deep neural networks through a sophisticated sequential attention mechanism [21]. The practical utility of this architecture has been demonstrated in diverse applications, including the prognosis of neurological conditions such as Parkinson's disease or comparative analysis for evaluating landslide susceptibility in engineering geology [22]. The input features X are initially passed through a batch normalization layer, yielding $X' \in R^{w \times z}$, where w denotes the batch size and z represents the feature dimension. This normalized input then proceeds through a series of sequential decision steps.

D. Model Training and Evaluation

The core of the training process for both models hinges on an iterative approach, where data is fed in small, manageable mini-batches. This strategy is implemented via *lib.iterate_minibatches*, a custom utility designed to generate batches of 1024 samples, shuffled for randomness, across an infinite number of epochs (*float('inf')*), signifying that the training loop is designed to continue indefinitely until explicit early stopping criteria are met. For the learning objective, *F.cross_entropy* is employed as the loss function. This function is commonly used for multi-class classification problems, quantifying the dissimilarity between the predicted probability distribution and the true label distribution.

$$L = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{ic}\log(\hat{y}_{ic}) \tag{3}$$

where N is the number of samples in the batch, C is the number of classes, y_{ic} is a binary indicator (1 if sample i belongs to class c, 0 otherwise), and y_{ic} is the predicted probability of sample i belonging to class c. The optimization of the model's parameters is performed by the QHAdam optimizer, an adaptive learning rate optimizer known for its effectiveness in deep learning. Its specific parameters, $nus = (0.7, 1.0)$ and $betas = (0.95, 0.998)$, the recommended default values for this optimizer, are configured to control its momentum and adaptive learning rate components. Beyond loss and error rate, the model's performance is further characterized by Accuracy, which represents the proportion of correctly classified instances. Precision indicates the proportion of positive identifications that were actually correct, while Recall measures the proportion of actual positives that were correctly identified. F1-score provides a balanced measure of the model's performance.

E. Ensemble Method

Ensemble methods are well-regarded in machine learning for their capacity to improve generalization capabilities and reduce the risk of overfitting by mitigating the biases or weaknesses of individual models [23]. These methods are frequently deployed in offline-to-online reinforcement learning for task generalization [24] and applied effectively in medical diagnostics using soft voting classifiers [25]. Crucially, the soft voting approach employed here is a true ensemble method, aggregating predictive probabilities from the independently trained NODE and TabNet models, rather than forming a single composite architecture trained end-to-end. The model is evaluated to ensure consistent inference behavior. The softmax function converts a vector of arbitrary real values (logits) into a probability distribution over classes, where each value is in the range (0, 1) and all sum to 1. For a given logit z_k corresponding to class k, the softmax probability $\hat{y}_k$ is calculated as:

$$\hat{y}^k = \frac{e^{z_k}}{\sum_{j=1}^{C} e^{z_j}} \quad (4)$$

where C is the total number of classes.

Once the class probabilities from both individual models (P_{NODE} and P_{TabNet}) are obtained, the soft voting aggregation is performed. This involves computing the element-wise arithmetic mean of the probabilities for each class across the two models. Finally, the performance of the ensemble model is rigorously evaluated by comparing its final predictions against the true labels of the validation set. Standard classification metrics, including accuracy, precision, recall, and F1-score, are calculated for the ensemble.

III. RESULTS AND ANALYSIS

A. Performance of Individual Models

This section details the empirical performance of the NODE and TabNet models, each trained independently on the meticulously preprocessed diabetes prediction dataset. The evaluation encompasses both training and validation metrics, providing a comprehensive understanding of each model's learning trajectory and generalization capabilities.

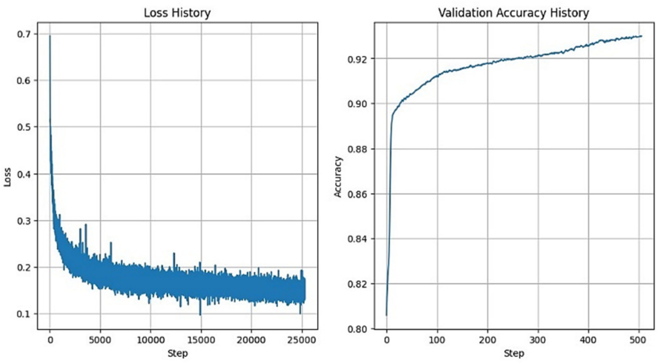

Fig. 2. Learning curves for NODE.

Figure 2 shows the learning curves for NODE. The loss history curve shows a rapid decline in the early steps, indicating that the model is learning aggressively. This decline slows significantly after about 5000 steps and stabilizes at a loss value of 0.15 to 0.20, with minor fluctuations, indicating that the model is converging. Meanwhile, the validation accuracy history curve shows a very rapid increase from an initial value of around 0.80, reaching over 0.90 in a short time. This curve continues to increase slowly, passing 0.92 and approaching 0.94, indicating that the model is not overfitting and its generalization ability continues to improve. Overall, this curve reflects a healthy training process, in which the model effectively learns from the training data and improves its performance on the validation data.

The TabNet architecture is configured with $n\_d = 64$ and $n\_a = 64$, chosen following standard recommendations to effectively balance model complexity and computational efficiency, which determine the dimensions for the decision step layer and the attention layer. The use of $n\_steps = 5$ indicates that the model makes predictions through five sequential decision steps. This allows TabNet to selectively focus its attention on different subsets of features at each step, which is its main advantage in handling tabular data. The parameters $gamma = 1.3$ and $lambda\_sparse = 1e - 3$ are designed to control how the model uses these features, encouraging it to learn from slightly different feature subsets at each step, thereby improving efficiency and interpretability. The training results show that this architecture is very effective for diabetes prediction tasks. Performance improvements are evident in validation recall of 94.20% and validation F1-score of 93.42%, indicating that the model excels at identifying diabetes cases without sacrificing too much precision. The small difference between the training and validation metrics indicates that the model learns well without overfitting, confirming the generalization ability of this architecture.

TABLE II. PERFORMANCE OF NODE, TABNET, AND ENSEMBLE MODELS ON TEST & VALIDATION DATASET

Metric	NODE		TabNet		Soft-voting
	Test	Val	Test	Val	Val
Loss	14.03	07.00	11.76		-
Acc	93.24	93.00	94.44	93.40	93.55
Prec	92.43	91.85	93.86	92.65	92.60
Rec	94.23	94.29	95.12	94.20	94.58
F1	93.32	93.05	94.49	93.42	93.58

Table II presents a comprehensive performance analysis of the individual NODE and TabNet models. Both architectures demonstrated robust performance, with NODE achieving a validation accuracy of 93.00% and an F1-score of 93.05%. Similarly, the TabNet model exhibited slightly superior performance, with a validation accuracy of 93.40% and an F1-score of 93.42%. The consistently high metrics for both models on their respective training and validation sets suggest effective learning from the data without overfitting. This marginal lead of TabNet suggests that its sophisticated sequential attention mechanism may be slightly more adept at capturing complex feature interactions within this specific tabular dataset. The application of a soft voting approach leverages the collective strengths of both models to achieve an F1 validation score of 92.14% and a recall rate of 93.86%.

B. Comparative Analysis and Discussion

Comparing the performance of the individual models, the NODE model demonstrated slightly stronger overall performance, particularly in its F1-score of 0.9260 and Recall of 0.9341 on the test set. The TabNet model, while highly capable, showed a test F1-score of 0.9224 and a Recall of 0.9317. This marginal difference suggests that, for this specific diabetes dataset, NODE was slightly more effective in capturing the underlying complex patterns. Nevertheless, both architectures individually exhibited robust performance, consistently delivering validation accuracies and F1-scores above 92%, underscoring their viability for critical prediction tasks.

The soft voting ensemble strategy proved effective in harnessing the collective intelligence of both NODE and TabNet. Interestingly, the ensemble's F1-score directly matched that of the individual NODE model and slightly surpassed TabNet's individual F1-score. Although the ensemble's F1-score did not significantly exceed the best individual model in this particular instance, its Recall (0.9365) remained exceptionally high. This outcome demonstrates the ensemble approach's capacity to maintain strong detection capabilities for the positive class while consistently delivering a reliable and stable performance by effectively combining the strengths of its constituent models.

TABLE III. COMPARISON OF THE PROPOSED AND EXISTING METHODS IN DIABETES CLASSIFICATIONS

Scheme	Best model	Acc	Prec	Rec	F1
[6]	MLP	86.08	86.60	85.10	-
[7]	KNN	88.00	87.00	90.00	88.00
[8]	Ensemble	92.91	-	-	-
[9]	Ensemble	88.83	87.34	92.13	89.56
[10]	XGBoost	89.96	90.20	89.65	89.91
Proposed	NODE+TabNet	93.55	92.60	94.58	93.58

The proposed method integrates several technical advantages that directly address the limitations of previous research. This method utilizes advanced deep learning architectures, NODE and TabNet, which are specifically designed for tabular data. Unlike conventional models such as MLP or XGBoost, TabNet uses a sequential attention mechanism to selectively process feature subsets at each step, which improves efficiency and interpretability. This combined, highly accurate, and interpretable model is ideal for clinical deployment scenarios, such as integration into Electronic Health Records (EHR) or real-time Internet of Things (IoT) healthcare pipelines, providing physicians with immediate, trustworthy risk assessments and feature-based explanations for proactive intervention.

This study addressed the challenge of class imbalance commonly found in medical datasets using SMOTE. Preventing model bias towards the majority class directly improves the model's ability to identify actual diabetes cases, as reflected in its exceptional Recall (94.58%). The strategy of mitigating the individual weaknesses of each model and creating a more reliable model is evidenced by the F1-score (93.58%) and Accuracy (93.55%) that exceed the benchmarks set by other stacked ensemble models. The collective advantages, as shown in Table III, highlight the effectiveness of the proposed approach in uniting the complementary strengths of deep learning architectures for stronger and more accurate predictions.

IV. CONCLUSION

This study demonstrated the efficacy of applying advanced deep learning architectures, namely NODE and TabNet, to the critical task of diabetes prediction on a large-scale tabular dataset. The application of SMOTE to address class imbalance proved to be highly effective, enabling both models to achieve a validation accuracy and an F1-score exceeding 93%. The proposed soft-voting ensemble method further enhanced predictive performance, reaching an outstanding accuracy of 93.55% and an F1-score of 93.58% on the validation set, thereby setting a new benchmark compared to existing schemes. This research underscores the significant potential of combining state-of-the-art deep learning models to create more robust and reliable predictive systems for healthcare applications.

ACKNOWLEDGMENT

The Ministry of Higher Education, Science, and Technology of the Republic of Indonesia funded this research under Fundamental Research Grants No. 126/C3/DT.05.00/PL/2025, 0498.21/LL5-INT/AL.04/2025, and 004/KONTRAK-LPPM/AMIKOM/VI/2025.

REFERENCES

[1] J. Niu, Y. Liu, H. Peng, J. Chen, and L. Chen, "Early-stage diabetic retinopathy: gut-metabolic triggers, immune-neurodegeneration, and interventions," *Graefe's Archive for Clinical and Experimental Ophthalmology*, vol. 263, no. 10, pp. 2723–2736, Oct. 2025, https://doi.org/10.1007/s00417-025-06906-6.

[2] R. Candido *et al.*, "Retrospective cohort study on treatment outcomes of early vs late onset gestational diabetes mellitus," *Acta Diabetologica*, vol. 62, no. 6, pp. 881–889, Nov. 2024, https://doi.org/10.1007/s00592-024-02405-y.

[3] S. S. Bhat, M. Banu, G. A. Ansari, and V. Selvam, "A risk assessment and prediction framework for diabetes mellitus using machine learning algorithms," *Healthcare Analytics*, vol. 4, Dec. 2023, Art. no. 100273, https://doi.org/10.1016/j.health.2023.100273.

[4] M. Kumar, Sushant, and A. K. Yadav, "Speech signal's phase information based Alzheimer's disease detection using deep learning," *International Journal of Speech Technology*, vol. 28, no. 2, pp. 397–410, Jun. 2025, https://doi.org/10.1007/s10772-025-10193-1.

[5] L. E. S. E. Oliveira *et al.*, "SemClinBr - a multi-institutional and multi-specialty semantically annotated corpus for Portuguese clinical NLP tasks," *Journal of Biomedical Semantics*, vol. 13, no. 1, Dec. 2022, Art. no. 13, https://doi.org/10.1186/s13326-022-00269-1.

[6] U. M. Butt, S. Letchmunan, M. Ali, F. H. Hassan, A. Baqir, and H. H. R. Sherazi, "Machine Learning Based Diabetes Classification and Prediction for Healthcare Applications," *Journal of Healthcare Engineering*, vol. 2021, pp. 1–17, Sep. 2021, https://doi.org/10.1155/2021/9930985.

[7] H. Kaur and V. Kumari, "Predictive modelling and analytics for diabetes using a machine learning approach," *Applied Computing and Informatics*, vol. 18, no. 1/2, pp. 90–100, Mar. 2022, https://doi.org/10.1016/j.aci.2018.12.004.

[8] K. Oliullah, M. H. Rasel, Md. M. Islam, Md. R. Islam, Md. A. H. Wadud, and Md. Whaiduzzaman, "A stacked ensemble machine learning approach for the prediction of diabetes," *Journal of Diabetes & Metabolic Disorders*, vol. 23, no. 1, pp. 603–617, Nov. 2023, https://doi.org/10.1007/s40200-023-01321-2.

[9] H. Zaky *et al.*, "Machine learning based model for the early detection of Gestational Diabetes Mellitus," *BMC Medical Informatics and Decision Making*, vol. 25, no. 1, Mar. 2025, Art. no. 130, https://doi.org/10.1186/s12911-025-02947-3.

[10] M. A. Nematollahi *et al.*, "A cohort study on the predictive capability of body composition for diabetes mellitus using machine learning," *Journal of Diabetes & Metabolic Disorders*, vol. 23, no. 1, pp. 773–781, Nov. 2023, https://doi.org/10.1007/s40200-023-01350-x.

[11] M. Mustafa, "Diabetes prediction dataset." Kaggle, [Online]. Available: https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset.

[12] A. Thakur, T. Zhu, V. Abrol, J. Armstrong, Y. Wang, and D. A. Clifton, "Data encoding for healthcare data democratization and information leakage prevention," *Nature Communications*, vol. 15, no. 1, Feb. 2024, Art. no. 1582, https://doi.org/10.1038/s41467-024-45777-z.

[13] M. Rath and H. Date, "Quantum data encoding: a comparative analysis of classical-to-quantum mapping techniques and their impact on machine learning accuracy," *EPJ Quantum Technology*, vol. 11, no. 1, Dec. 2024, Art. no. 72, https://doi.org/10.1140/epjqt/s40507-024-00285-3.

[14] R. S. Selina, M. Rahardi, A. Aminuddin, F. F. Abdulloh, H. Badi, and B. P. Asaddulloh, "Optimizing Diabetes Diagnosis Using Machine Learning With SMOTE and Feature Selection," in *2025 International Conference on Computer Sciences, Engineering, and Technology Innovation (ICoCSETI)*, Jakarta, Indonesia, Jan. 2025, pp. 647–652, https://doi.org/10.1109/ICoCSETI63724.2025.11020043.

[15] M. Rahardi, A. Aminuddin, F. F. Abdulloh, B. P. Asaddulloh, H. R. Enriquez, and K. Kusnawi, "Analyzing the Impact of Data Resampling on Stroke Prediction using Machine Learning," *Engineering, Technology & Applied Science Research*, vol. 15, no. 2, pp. 20790–20797, Apr. 2025, https://doi.org/10.48084/etasr.9736.

[16] A. K. Salih, A. K. Faraj, M. A. Ahmed, and A. N. A. Al-Hasnawi, "The Impact of Data Splitting Strategy on Drilling Rate Prediction in the Rumaila Oil Field," *Petroleum Chemistry*, vol. 64, no. 7, pp. 781–786, Jul. 2024, https://doi.org/10.1134/S0965544124050025.

[17] H. Babaei, M. Zamani, and S. Mohammadi, "The impact of data splitting methods on machine learning models: A case study for predicting concrete workability," *Machine Learning for Computational Science and Engineering*, vol. 1, no. 1, Jun. 2025, Art. no. 21, https://doi.org/10.1007/s44379-025-00021-3.

[18] Y. Fan and P. Waldmann, "Tabular deep learning: a comparative study applied to multi-task genome-wide prediction," *BMC Bioinformatics*, vol. 25, no. 1, Oct. 2024, Art. no. 322, https://doi.org/10.1186/s12859-024-05940-1.

[19] S. Popov, S. Morozov, and A. Babenko, "Neural Oblivious Decision Ensembles for Deep Learning on Tabular Data." arXiv, Sep. 19, 2019, https://doi.org/10.48550/arXiv.1909.06312.

[20] P. Wielopolski, O. Furman, and M. Zięba, "NodeFlow: Towards End-to-End Flexible Probabilistic Regression on Tabular Data," *Entropy*, vol. 26, no. 7, Jul. 2024, Art. no. 593, https://doi.org/10.3390/e26070593.

[21] T. Kumar and R. L. Ujjwal, "TabNet unveils predictive insights: a deep learning approach for Parkinson's disease prognosis," *International Journal of System Assurance Engineering and Management*, Jul. 2024, https://doi.org/10.1007/s13198-024-02450-4.

[22] S. Yingze, S. Yingxu, Z. Xin, Z. Jie, and Y. Degang, "Comparative analysis of the TabNet algorithm and traditional machine learning algorithms for landslide susceptibility assessment in the Wanzhou Region of China," *Natural Hazards*, vol. 120, no. 8, pp. 7627–7652, Jun. 2024, https://doi.org/10.1007/s11069-024-06521-4.

[23] A. Rezaeezade and L. Batina, "Regularizers to the rescue: fighting overfitting in deep learning-based side-channel analysis," *Journal of Cryptographic Engineering*, vol. 14, no. 4, pp. 609–629, Nov. 2024, https://doi.org/10.1007/s13389-024-00361-5.

[24] C. Wang, X. Yu, C. Bai, Q. Zhang, and Z. Wang, "Ensemble successor representations for task generalization in offline-to-online reinforcement learning," *Science China Information Sciences*, vol. 67, no. 7, Jul. 2024, Art. no. 172203, https://doi.org/10.1007/s11432-023-4028-1.

[25] K. Akyol, E. Uçar, Ü. Atila, and M. Uçar, "An ensemble approach for classification of tympanic membrane conditions using soft voting classifier," *Multimedia Tools and Applications*, vol. 83, no. 32, pp. 77809–77830, Feb. 2024, https://doi.org/10.1007/s11042-024-18631-z.