Multi-Disease Risk Prediction Using Clinical–Lifestyle Feature Fusion and Deep Tabular Neural Networks

Shaik Lathifbasha[1], Morla Avinash[2], Anusha[3]
[1,2] Department of Computer Science and Engineering Vignan's Foundation for Science, Technology and Research Guntur, India
[3]Assistant Professor, Department of Computer Science and Engineering Vignan's Foundation for Science, Technology and Research Guntur, India

***Abstract*—The rising trends of non-communicable diseases such as cardiovascular disease and diabetes are some of the reasons why reliable early risk prediction systems are required. Physiological conditions and lifestyle behaviours influence these diseases, but most of the existing machine learning approaches focus on disease prediction in isolation and do not sufficiently factor in the interactions between the clinical and lifestyle factors. The current paper introduces a risk prediction framework of multiple diseases combining both clinical and lifestyle features by using a systematic feature fusion strategy combined with deep tabular neural networks. The assessment of clinical indica- tors is done through threshold based abnormality scoring, and lifestyle factors are normalised and synthesised to demonstrate behavioural risk patterns. To describe nonlinear interactions between clinical and lifestyle factors, a sigmoid-gated fusion mechanism is presented, which generates a better feature representation to make predictions.**

The feature set is combined and then a deep learning model is trained using TabNet which is able to select the features selectively and learn using structured healthcare data. The system classifies the forecasts in diabetes and heart diseases into three levels of risks namely low, moderate and high. It does this by the possibility-based classification. This helps people to make decisions on how to take care of them so that they can avoid falling sick. The determination of the model performance is done by the use of the F1-score, the confusion matrix and the ROC- AUC.

Experiments indicate that TabNet-based explainable artificial intelligence can be used to accurately predict the probability of various diseases and yet is simple to interpret. The importance of global features and patient-level explanations are considered to identify the most common clinical and lifestyle factors influencing the disease prediction. The suggested framework improves predictive stability, transparency, and scalability. This is why it is suitable in clinical decision support and proactive health monitoring systems.

***Index Terms*—Multi-Disease Prediction, Clinical–Lifestyle Feature Fusion, TabNet, Deep Tabular Learning, Healthcare Ana- lytics, Risk Stratification, and Explainable Artificial Intelligence (XAI)**

I. INTRODUCTION

Two of the most prevalent causes of death and long-term disability in the world are diabetes and heart disease. These conditions have become very common especially with the rapid urbanization, sedentary living, poor dietary patterns, stress, smoking, and alcohol use. Preventive healthcare, lower cost of treatment and better patient outcome in the long run depend on early identification of high-risk individuals Traditional approaches to diagnosis are based mostly on laboratory tests and clinical skills. Although effective, these methods tend to pick up diseases at later stages when they have already acquired complications. As more and more structured healthcare data is becoming available, machine learning and deep learning methods have become the promising directions of early disease prediction [5], [6]. These methods are capable of examining complicated correlations among the features of patients and revealing concealed patterns that are not readily discovered by use of conventional statistical techniques [15]. A number of studies have implemented machine learning algorithms, including Support Vector Machines, Random For- est, and Gradient Boosting to predict diseases like diabetes and heart disease [1]–[3]. Moreover, deep neural networks have shown better performance because they extract nonlinear feature correlation relationships [6]. Nevertheless, the majority of the available studies are based on single-disease prediction and tend to isolate clinical and lifestyle aspects of the problem [10], [16]. The interaction of lifestyle behaviours with clinical abnormalities is little studied in a systematic way in a multi- disease model with a unified framework of interaction models[14].

The real-life healthcare setting is closely interconnected between clinical indicators and lifestyle behaviours. On a signatory note, obesity and inactivity are key determinants that affect the cardiovascular and glucose regulation. Such interactions may not be paid enough attention to, which can reduce the strength and interpretability of predictive models. It, therefore, follows that the integration of clinical and lifestyle data on the basis of application of a structured feature fusion mechanism may augment risk modeling and generalization of different diseases [7], [14].

The present paper suggests a framework of the multi-disease risk prediction algorithm based on combining clinical and lifestyle characteristics through a gated feature fusion ap- proach in a deep tabular neural network. Normalized lifestyle behaviours are used to measure behavioural risk and clinical features are assessed using threshold-based scoring of abnor- malities. A TabNet model of nonlinear learning and attentive feature selection is used to process the fused representation to achieve this task [8]. Probability-based classification and risk stratification are done in the system into Low, Moderate and High categories, where explainable AI methods give global and patient-level interpretability of predictions [9], [12], [13].

II. LITERATURE SURVEY

Machine learning in the recent past has significantly con- tributed to research on disease prediction. Some of the most widely used traditional classification algorithms to predict long term illnesses such as diabetes and heart disease include Sup- port Vector Machines (SVM), Random Forest (RF), K-Nearest Neighbours (KNN), and XGBoost. Numerous studies show encouraging classification results, but issues with dataset bias, restricted model generalization, and decreased interpretability still exist.

Deep learning methods, specifically, Deep Neural Networks (DNN) have also improved predictive power by identifying nonlinear associations between clinical variables. These mod- els are able to determine trends associated with disease pro- gression by analyzing high-dimensional healthcare datasets. Regardless of such developments, most of the frameworks are seldom structured to model heterogeneous healthcare charac- teristics and focus on prediction of a single disease.

Health outcome of patients is influenced in the real-world healthcare environment by more than by physiological mea- surements. The factors of lifestyle such as diet, exercise, stress, smoking, and drinking contribute also. However many existing prediction systems consider both clinical and lifestyle factors independently of each other and, therefore, it is difficult to predict how the behavioural and physiological risk factors combine to cause disease. Ensemble and hybrid machine learning models have been proposed to improve the accuracy of prediction, though they often increase computational costs and reduce interpretability. Moreover, there is little research that examines the structured feature fusion mechanisms that combine clinical abnormalities with lifestyle risk factors in a unified multi-disease prediction system.

A detailed review of 20 reference papers explains the ben- efits and limitations of existing disease prediction techniques. Prof. K.K. Ingole and others (2025) created a multiple human disease predictor using Support Vector Machine (SVM) on conditions such as diabetes and heart disease and an accuracy of about 77 per cent was achieved in predicting diabetes using Kaggle datasets. But, performance varied across datasets, which points towards challenges in model generalisation. Hosam El-Sofany et al. (2024) designed a system to predict heart disease with various machine learning algorithms. The most precise one was XGBoost with an accuracy rate of approximately 89.2 percent, despite the lack of a particularly representative dataset.

A detailed review of 20 reference papers explains the ben- efits and limitations of existing disease prediction techniques. Prof. K.K. Ingole and others (2025) created a multiple human disease predictor using Support Vector Machine (SVM) on conditions such as diabetes and heart disease and an accuracy of about 77 per cent was achieved in predicting diabetes using Kaggle datasets. But, performance varied across datasets, which points towards challenges in model generalisation. Hosam El-Sofany et al. (2024) designed a system to predict

heart disease with various machine learning algorithms. The most precise one was XGBoost with an accuracy rate of approximately 89.2 percent, despite the lack of a particularly representative dataset.

The ensemble learning techniques have been examined in many studies to enable predictions that are more precise. Single classifiers may not do as well as ensemble models, but cannot be applied in practice since they cannot be scaled or the data is not balanced. Disease prediction frameworks designed using deep learning are highly effective in modeling nonlinear relationships between healthcare features and require a sub- stantial amount of data and computational power. Others have proposed the application of hybrid machine learning models that can be used to integrate several classifiers to ensure that diagnoses are more accurate. However, these models make the system more complicated and take longer to train.

Comparative studies reveal that ensemble models usually outperform traditional classifiers in predicting chronic dis- eases. Nevertheless, model stability is affected by preprocess- ing challenges, feature selection issues, and dissimilar medical data. Deep neural-based AI-based diagnostic systems also have the danger of overfitting and lack of interpretability. The schol- ars have examined cloud-based and federated learning models to enable healthcare systems to be more scalable and private yet latency, security, and diagnostic accuracy issues remain. Recent studies have applied Explainable Artificial Intelligence techniques, such as SHAP, to provide greater transparency to medical predictions; the integration of explainability into systematic multi-disease prediction systems remains limited.

Various gaps in research are identified in the literature review. The existing systems focus on prediction of single diseases rather than multi-disease modelling. In addition, explicit modelling of structured interactions between clinical indicators and lifestyle behaviours is rarely done. Although ensemble models and deep learning models are more accurate in making predictions, they tend to complicate things without simplifying them. Risk prediction in numerous studies con- tinues to be confined to binary classification and is devoid of probability-based risk stratification. Besides, limited studies have integrated organized feature fusion procedures with deep tabular neural network plans designated to healthcare records. In order to address these limitations, the present study presents a multi-disease risk prediction paradigm, which com- bines clinical abnormality analysis and standardised lifestyle modelling using a gated feature fusion mechanism alongside deep tabular neural networks. The objective of the proposed framework is to ensure that the proposed real-world healthcare decision support systems become more reliable, understand-able, and scalable.

III. PROPOSED METHODOLOGY

The proposed framework integrates clinical and lifestyle variables based on a structured feature fusion approach and deep tabular neural networks to estimate the risk of various diseases. The whole process involves data preprocessing, feature categorisation, score computation, gated feature fusion, model training, risk stratification and explainability analysis.

A. Diagram

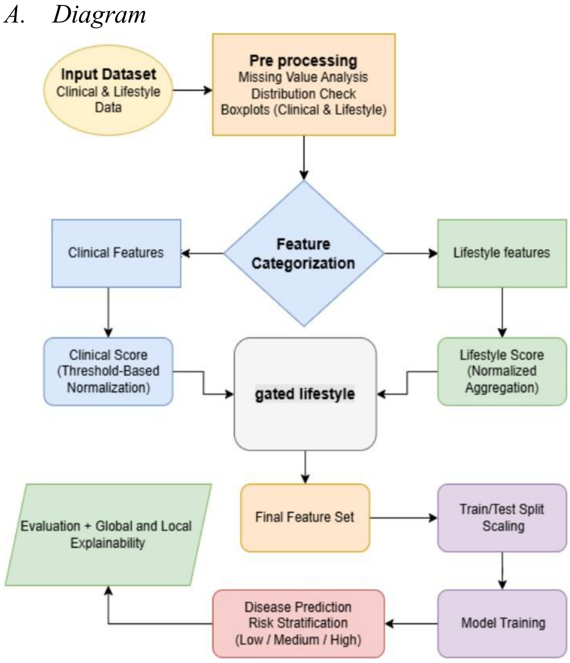

Fig. 1: Proposed Framework for Multi-Disease Risk Prediction

B. Input Dataset

The data in this research will be structured healthcare attributes, which are classified as clinical and lifestyle factors. The clinical characteristics age, blood pressure, glucose level, cholesterol level, insulin level and body mass index (BMI) are common factors used in the prediction of cardiovascular dis- ease and diabetes [4], [6]. They are widely applied in machine learning-based healthcare models since these characteristics are highly correlated with the risk of chronic diseases. Besides, there are lifestyle factors like smoking status, alcohol use, level of physical activity, sleep duration, stress level, dietary and sedentary time to include behavioural risk factors that have a significant impact on disease development and progression [7], [14]. The experiment is based on publicly accessible benchmark datasets on heart disease and diabetes prediction acquired in the UCI Machine Learning Repository [21], [22]. Prior to the model development, the dataset passes through preprocessing steps such as the processing of missing values, coding categorical variables and scaling features using stan- dard normalization procedures to enhance data consistency and performance of the models used to predict the data set [15]. An 80 percent 20 percent split is then used to split the processed dataset into training and testing. The model parameters are learnt using the training set and the testing set is to test how well the model can generalize, which provides an impartial and solid test of the suggested model [15].

C. Feature Categorization

All input attributes are divided into two main groups:

- Clinical Features: Physiological and laboratory indica- tors that show how sick a patient is right now.
- Lifestyle Features: Behavioural and environmental de- terminants affecting long-term disease risk.

This structured categorisation allows for the independent modelling of physiological anomalies and behavioural risk patterns prior to feature integration.

D. Clinical Abnormality Scoring

We use medically defined threshold conditions to figure out clinical risk. If the value of a clinical feature is higher than a certain clinical threshold, it adds to the abnormality score.

Let C_i stand for the i^{th} clinical trait. To find the Clinical Abnormality Score (CAS), do the following:

$$CAS = \frac{1}{n}\sum_{i=1} f(C_i) \qquad (1)$$

where $f(C_i)$ is a binary indicator of an abnormality defined as:

$$f(C) = \begin{cases} 1, & \text{if the clinical feature exceeds its abnormal threshold} \\ 0, & \text{otherwise} \end{cases}$$

This normalisation makes sure that the clinical score is between 0 and 1.

E. Lifestyle Risk Aggregation

Lifestyle characteristics are consolidated to measure be- havioural risk exposure. Let L_j stand for the j^{th} lifestyle trait. The Lifestyle Risk Score (LRS) is the average of lifestyle factors that have been normalised:

$$LRS = \frac{1}{m}\sum_{j=1}^{m} L_j \qquad (2)$$

where m is the total number of lifestyle traits. This normal- isation makes sure that the lifestyle score stays between 0 and 1.

F. Gated Feature Fusion

A sigmoid-based gating mechanism is used to model how physiological problems and behavioural risk factors work together. The gating function determines how much lifestyle risk affects clinical abnormalities based on how serious they are.

The gating function is defined as:

$$g = \sigma(\alpha \cdot CAS) \quad (3)$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$ is the sigmoid function and α controls how sensitive the gating mechanism is.

The last gated lifestyle representation is calculated as:

$$GL = LRS \cdot g \quad (4)$$

This mechanism makes sure that lifestyle risk has a bigger effect when clinical abnormality levels go up.

G. Training a Model with Deep Tabular Networks

The derived features, consisting of clinical attributes, the clinical abnormality score, and the gated lifestyle representation are then trained in a TabNet-based deep learning model. TabNet is designed to work on structured tabular data and applies sequential attention to select the most significant aspects of training.

This model involves binary classification to determine whether one is a heart disease or diabetes patient. The prob- ability scores in the output layer have a range of 0 to 1 and indicate the likelihood of occurrence of a disease.

H. Risk Stratification

Predicted probabilities are also classified into useful risk levels in order to simplify the results. The thresholds are in percentiles and divide predictions into three categories: low risk, moderate risk, and high risk. This is as opposed to using raw probability values. This grouping simplifies the results and their usage. By putting patients into groups according to their estimated risk, healthcare professionals are able to compare their conditions and make improved decision regarding monitoring, prevention and clinical management.

I. Model Explainability

In order to further clarify medical predictions, the explain- ability mechanism provided by TabNet model is employed to introduce explainable artificial intelligence. The feature attribution masks provided by TabNet enable one to perceive something both at a global and local scale.

Global feature importance analysis finds the most important clinical and lifestyle factors that affect predictions across the whole dataset. Additionally, patient-level explanations are presented in a manner that aims at identifying the influence of each feature on a particular prediction. This explainability level increases the transparency of this model and aids in the clinical validation of the predictive system.

IV. RESULTS AND DISCUSSION

We test the clinical-lifestyle feature fusion framework and a deep tabular neural network with the help of standard classification metrics to predict several diseases. To evaluate the effectiveness of the model, we consider its Accuracy, Precision, Recall, F1-score, Confusion Matrix, Receiver Op- erating Characteristic (ROC) curves, and Area Under the Curve (AUC). Additionally, explainable artificial intelligence procedures that make use of TabNet feature attribution are used to analyse feature contributions of the entire population and individual patients.

A. Classification Performance

Table I presents the predictive accuracy of the classification of heart disease. The TabNet-based model is able to predict well since the precision and recall values of the model are not skewed. The overall accuracy of the model is 0.90 and it functions well in both classes indicating that it has the ability to distinguish between disease and non-diseases cases.

TABLE I: Heart Disease Classification Performance

Class	Precision	Recall	F1-score	Support
0 (No Disease)	0.91	0.88	0.90	1020
1 (Disease)	0.88	0.91	0.90	980
Accuracy	0.90			2000
Macro Avg	0.90	0.90	0.90	2000
Weighted Avg	0.90	0.90	0.90	2000

B. Model Comparison

We compare the proposed TabNet model to baseline models like Logistic Regression and Deep Neural Networks (DNN). Tables II and III show the results of the comparison for predicting heart disease and diabetes, respectively.

TABLE II: Heart Disease Model Comparison

Model	Accuracy	AUC	Precision	Recall	F1-score
Logistic Regression	0.6650	0.7469	0.6677	0.6296	0.6481
DNN	0.8805	0.9475	0.8716	0.8867	0.8791
TabNet	0.9285	0.9820	0.9257	0.9286	0.9272

TABLE III: Diabetes Disease Model Comparison

Model	Accuracy	AUC	Precision	Recall	F1-score
Logistic Regression	0.7680	0.8502	0.7756	0.7271	0.7505
DNN	0.9245	0.9755	0.9140	0.9302	0.9220
TabNet	0.9580	0.9937	0.9451	0.9688	0.9568

The findings indicate that the TabNet model substantially surpasses both Logistic Regression and the conventional Deep Neural Network model on all evaluation metrics. The improve- ment is especially clear in ROC-AUC scores, which show that structured healthcare data can be very useful for making decisions.

C. Confusion Matrix Analysis

Figures 2 and 3 show confusion matrices for predicting heart disease and diabetes. The results show that the suggested model makes a lot of correct classifications with very few false positives and false negatives.

The confusion matrix shows that the positive and negative classes are very different from each other, which shows that the feature fusion strategy works well for predicting heart disease. The model still makes reliable predictions for diabetes, even though the classification boundary is a little more complicated.

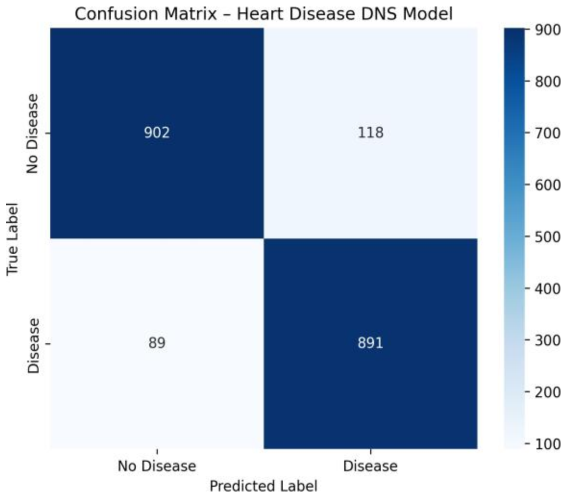

Fig. 2: Confusion Matrix for Heart Disease

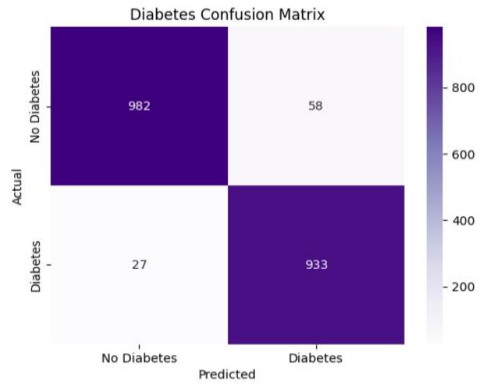

Fig. 3: Confusion Matrix for Diabetes

D. ROC Curve and AUC Analysis

Figures 4 and 5 show the ROC curves for predicting heart disease and diabetes. The ROC curves show how the true positive rate and the false positive rate change when the classification threshold is changed.

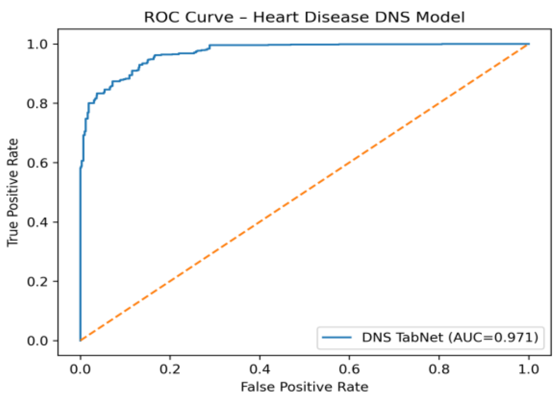

Fig. 4: ROC Curve for Heart Disease

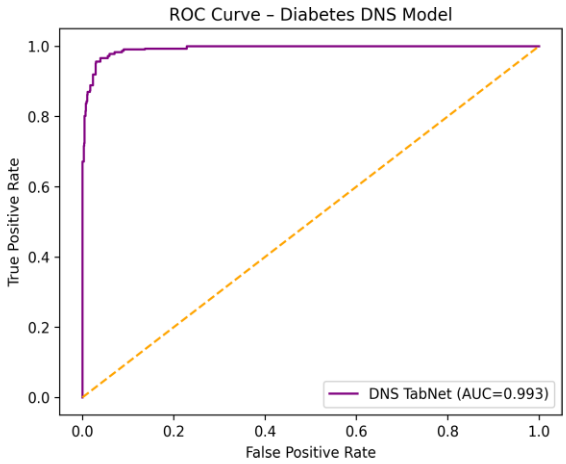

Fig. 5: ROC Curve for Diabetes

The model gets an AUC of 0.982 for predicting heart disease, which means it can tell the difference between things very well. The model has an AUC of 0.9937 for predicting diabetes, which means it can tell the difference between things very well.

E. Explainable AI Analysis

To enhance transparency in medical decision-making, fea- ture attribution analysis was conducted utilising the explain- ability mechanism offered by the TabNet model. This analysis gives both global feature importance and explanations for each patient. Figures 6 and 7 show how important different factors are for predicting heart disease and diabetes around the world. The findings demonstrate that clinical variables, including age, cholesterol level, blood pressure, and insulin, play a significant role in disease prediction. Lifestyle-related factors, including physical activity and gated lifestyle scores, significantly influ- ence overall risk assessment.

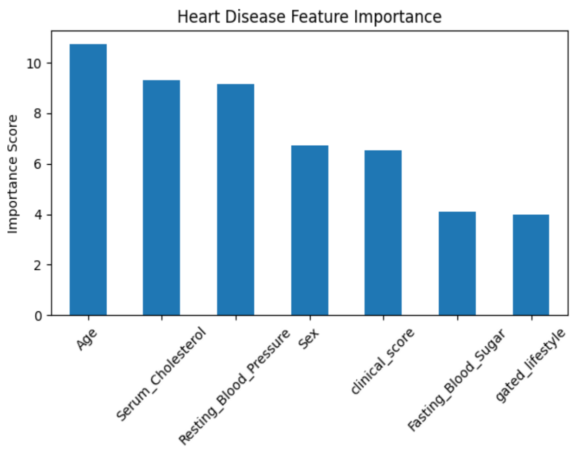

Fig. 6: Global Feature Importance for Heart Disease Prediction

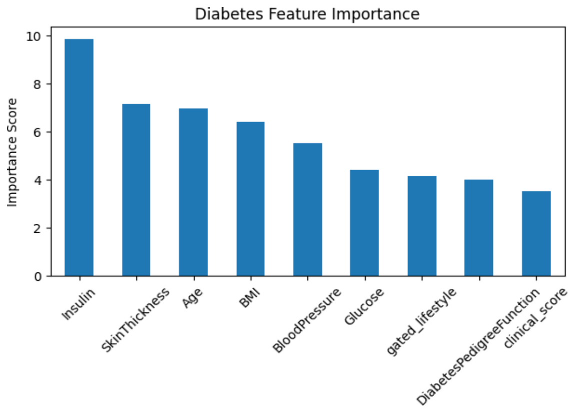

Fig. 7: Global Feature Importance for Diabetes Prediction

Figures 8 and 9 show patient-level explanations. These numbers show how each feature affects the prediction outcome for a certain patient instance. Two things that can help predict heart disease are fasting blood sugar and serum cholesterol. Insulin and glucose levels, on the other hand, are two things that can help you figure out how likely you are to get diabetes.

F. Discussion

The experimental findings indicate that the combination of clinical abnormality evaluation with normalised lifestyle aggregation markedly enhances predictive stability for multi- disease risk forecasting. Clinical factors like cholesterol, glucose, and insulin have a big effect on predicting how diseases will turn out. Lifestyle-related factors make predictions even better.

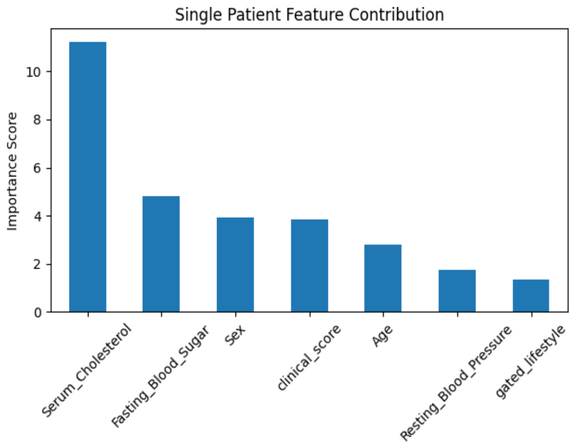

Fig. 8: Single Patient Feature Contribution for Heart Disease

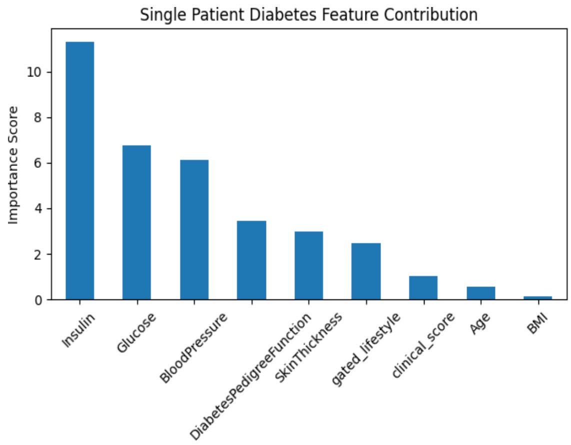

Fig. 9: Single Patient Feature Contribution for Diabetes

The TabNet architecture accurately identifies nonlinear rela- tionships among structured healthcare features while yielding interpretable feature attribution outcomes. Global feature importance analysis shows which risk factors are most important, and patient-level explanations make individual predictions clearer.

The proposed framework shows that using structured clini- cal–lifestyle feature fusion with deep tabular neural networks makes predictions more accurate, easier to understand, and more reliable than traditional machine learning models.

V. CONCLUSION

The paper presented a multi-disease predictive model that combines clinical and lifestyle features using a systematic feature fusion mechanism together with deep tabular neural networks. The suggested approach has the advantage of be- ing able to model complex relations in structured healthcare data using a separate model of physiological indicators and behavioural risk factors and connecting them with the help of a gated interaction mechanism.

Its model based on TabNet had a strong predictive perfor- mance in the classification of heart disease and diabetes. Ex post-analysis showed superior discriminative predictive power on heart diseases and strong predictive power on diabetes as shown by ROC-AUC analysis, classification metrics, and confusion matrix analysis. The introduction of probability- based risk stratification through the classification of patients as Low, Moderate and High risk based on the likelihood further simplified the understanding.

Along with accurate predictions, TabNet's built-in explain- ability feature gave both global and patient-level feature attri- bution analysis. It was possible to identify the most significant clinical and lifestyle factors influencing predictions, illustrat- ing the healthcare decision support applications as trustworthy and transparent.

Overall, the suggested framework demonstrates that a com- bination of structured clinical and lifestyle data with deep tabular neural networks can significantly enhance the capacity to forecast the risk of various diseases and still be able to interpret the findings. Further work will focus on extending the framework to include other chronic illnesses, testing the model on a larger and more diverse healthcare data set, and developing scalable implementation solutions to real-time preventative healthcare systems.

REFERENCES

[1] K. K. Ingole, P. Sharma, and R. Patil, "Multiple Human Disease Prediction Model Using Support Vector Machine," *International Journal of Advanced Computer Science and Applications*, vol. 16, no. 2, pp. 115–122, 2025.

[2] H. El-Sofany, M. A. Elhoseny, and S. M. Mostafa, "Heart Disease Prediction Using XGBoost and Machine Learning Techniques," *IEEE Access*, vol. 12, pp. 45231–45244, 2024.

[3] A. Yadav, S. Kumar, and R. Mehta, "Intelligent Multi-Disease Diagnosis System Using KNN and SVM," *Procedia Computer Science*, vol. 215, pp. 102–110, 2025.

[4] T. Liu, Y. Chen, and X. Wang, "Machine Learning Models for Cardio- vascular Disease Prediction: A Systematic Review and Meta-Analysis," *Artificial Intelligence in Medicine*, vol. 148, pp. 102678, 2025.

[5] R. Jayaraj and S. Narayanan, "A Multiple Disease Prediction System Using Advanced Machine Learning Techniques," *Journal of Healthcare Engineering*, vol. 2025, Article ID 4589213, 2025.

[6] S. Sharma and A. Gupta, "Deep Learning-Based Diabetes Prediction Using Clinical Data," *Computers in Biology and Medicine*, vol. 162, pp. 107053, 2024.

[7] M. Patel, D. Shah, and N. Desai, "Lifestyle-Oriented Disease Risk Prediction Using Neural Networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 29, no. 3, pp. 1456–1465, 2025.

[8] S. O. Arik and T. Pfister, "TabNet: Attentive Interpretable Tabular Learning," in *Proc. AAAI Conf. Artificial Intelligence*, vol. 35, no. 8, pp. 6679–6687, 2021.

[9] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 4765–4774, 2017.

[10] P. Singh and A. Kaur, "Multi-Label Disease Prediction Using Hybrid Machine Learning Models," *Expert Systems with Applications*, vol. 240, pp. 121342, 2025.

[11] D. Rao and K. Srinivas, "Early Detection of Cardiovascular Diseases Using Ensemble Learning," *Applied Soft Computing*, vol. 148, pp. 110821, 2024.

[12] A. Kumar, R. Verma, and S. Joshi, "Risk Stratification in Chronic Disease Prediction Using Deep Neural Networks," *IEEE Access*, vol. 13, pp. 21567–21578, 2025.

[13] F. Ahmed and M. Hossain, "SHAP-Based Interpretation for Clinical Risk Prediction Models," *Artificial Intelligence in Healthcare*, vol. 7, no. 2, pp. 95–108, 2024.

[14] Y. Chen and L. Zhang, "Integrated Clinical and Lifestyle Data Modeling for Disease Forecasting," *IEEE Transactions on Computational Biology and Bioinformatics*, vol. 22, no. 1, pp. 342–351, 2025.

[15] B. Thomas and R. George, "Comparative Analysis of Machine Learning Models for Multi-Disease Prediction," *Biomedical Signal Processing and Control*, vol. 91, pp. 105623, 2025.

[16] A. Rehman, S. Singh, V. Singh, and B. Hazela, "Multiple Disease Prediction System Using Machine Learning," *Journal of Informatics Electrical and Electronics Engineering*, vol. 5, no. 2, pp. 1–13, 2024.

[17] S. Kumar and N. Khan, "Machine Learning-Based Multiple Disease Prediction System," *International Journal of Trend in Scientific Research and Development*, vol. 9, no. 3, pp. 468–472, 2025.

[18] A. Vaish, B. Sihi, J. M. H., and K. Kumar, "Multiple Disease Prediction Model Using Machine Learning," *International Journal of Research and Analytical Reviews*, vol. 12, no.

1, pp. 299–305, 2025.
[19] A. Ul Haq, M. Yaseen, A. Khan, F. Din, and A. Ali, "Multi-Disease Prediction Using Machine Learning and Deep Learning Models," *Kashf Journal of Multidisciplinary Research*, vol. 1, no. 12, pp. 1–12, 2024.
[20] Y. Guo et al., "MDP-GRL: Multi-Disease Prediction by Graph-enabled Representation Learning," *IEEE Journal of Biomedical and Health Informatics*, 2025.
[21] D. Dua and C. Graff, "UCI Machine Learning Repository," University of California, Irvine, 2019. [Online]. Available: https://archive.ics.uci.edu/ml
[22] V. Sigillito, S. Wing, L. Hutton, and K. Baker, "Pima Indians Diabetes Database," UCI Machine Learning Repository, 2019.