Submitted: 30 May, 2024 Accepted: 16 August, 2024 Published: 30 October, 2024 DOI:10.22514/jomh.2024.166

ORIGINAL RESEARCH

JOMH Journal of Men's Health

An explainable hybrid deep learning model for prediabetes prediction in men aged 30 and above

Hung Viet Nguyen[1], Younsung Choi[2,*,†], Haewon Byeon[2,*,†]

[1]Department of Digital Anti-Aging Healthcare (BK21), Inje University, 50834 Gimhae, Republic of Korea
[2]Department of AI-Software, Inje University, 50834 Gimhae, Republic of Korea

***Correspondence**
cys2020@inje.ac.kr
(Younsung Choi);
bhwpuma@naver.com
(Haewon Byeon)

[†] These authors contributed equally.

Abstract

Men diagnosed with type 2 diabetes at age 30 have a significantly reduced life expectancy compared to their non-diabetic peers. Prediabetes, like diabetes, not only heightens the risk of microvascular complications but also significantly increases the likelihood of cardiovascular diseases and overall mortality. Early detection and intervention in prediabetes can prevent these complications and delay or avoid the progression to diabetes. However, prediabetes prediction remains a challenging area due to biased accuracy and a lack of explainability in many existing machine learning methods. To address these issues, this study aims to develop a novel hybrid model, HyperTab-LIME, which combines a hypernetwork-based deep learning approach designed for small tabular datasets (HyperTab) with the Local Interpretable Model-Agnostic Explanations (LIME) framework to predict prediabetes in men aged over 30. We employed data from 1527 male participants aged 30 and above from the 2022 Korean National Health and Nutrition Examination Survey to train and test our model. The performance of HyperTab-LIME was evaluated against several baseline models. The results demonstrated that the HyperTab-LIME model excelled, achieving the Area Under the Receiver Operating Characteristic Curve (AUC) scores of 0.8429 on the validation set and 0.8300 on the hold-out set, thereby outperforming other baseline models included in this study. To provide a dependable and interpretable diagnostic tool for healthcare professionals, the optimized HyperTab model was integrated with the LIME framework. This innovative approach ensures detailed and precise explanations of predictive outcomes, greatly enhancing the medical community's ability to understand and trust the model's decisions, thereby facilitating earlier interventions in the treatment of prediabetes.

Keywords

Explainable AI; HyperTab; LIME; Prediabetes

1. Introduction

In recent years, there has been a notable escalation in the incidence of diabetes worldwide [1–3]. According to the International Diabetes Federation (IDF) Diabetes Atlas [1], approximately 463 million adults aged between 20 and 79 years were living with diabetes in 2019, constituting 9.3% of the global adult population. This number increased to 536.6 million (10.5% of the population) by 2021. The proportion of adults diagnosed with diabetes has more than doubled from 4.7% in 1980 to 10.5% in recent years [2, 3]. Moreover, it is estimated that over half of the 463 million individuals diagnosed with diabetes are not aware of their condition [1].

The significant prevalence of diabetes and its complications pose serious challenges to global health. For instance, diabetes ranked as the eighth leading cause of daily mortality worldwide in 2019 [4]. In Korea, it was the fifth leading cause of death [5], with age-standardized mortality rates from diabetes higher than the average among other Organization for Economic Co-operation and Development (OECD) countries [6]. By 2018, the prevalence of type 2 diabetes in South Korean adults aged 30 and older had escalated, affecting one in every seven adults [7]. The condition is more prevalent in men, with a rate of 15.9% compared to 11.8% in women, underscoring the critical need for targeted interventions that address the specific triggers of type 2 diabetes in the male population [7, 8]. In 2020, about 15.83 million Korean adults over 30 (44% of the population), comprising 46.9% of men and 41.2% of women, were identified with prediabetes [9]. In addition, data from the USA population indicate that men diagnosed with type 2 diabetes at the ages of 30, 40 and 50 experience reduced life expectancies by approximately 14, 9 and 5 years, respectively, compared to those without the condition [10]. These statistics underline the critical need for early detection and preventive strategies against diabetes and prediabetes, particularly in men starting from the age of 30.

First identified in 1997 by the Expert Committee on Diagnosis and Classification of Diabetes Mellitus, prediabetes is recognized as a crucial diagnostic category indicating a significantly elevated risk for the eventual development of diabetes [11]. This condition serves as an intermediary phase between normal glucose tolerance and diabetes mellitus, characterized by either impaired fasting glucose or impaired glucose tolerance (IGT) [12]. Prediabetes is viewed as a condition with a high propensity for progression to diabetes, with annual conversion rates to diabetes ranging from 5% to 10% [13]. In 2021, it was estimated that 10.6% of the global population, or approximately 541 million people, were affected by IGT, with projections suggesting an increase to 730 million (11.4%) by 2045 [14]. Prediabetes is diagnosed through several parameters: a fasting plasma glucose (FPG) level of 100–125 mg/dL (5.6–6.9 mmol/L), impaired glucose tolerance (IGT) indicated by a 2-hour oral glucose tolerance test (OGTT) result of 140–199 mg/dL (7.8–11.0 mmol/L), or a glycated hemoglobin (HbA1c) level ranging from 5.7%–6.4% (39–46 mmol/mol) [12]. Like diabetes, prediabetes increases the risk of microvascular complications [13] and significantly elevates the risk for cardiovascular disease and overall mortality—nearly doubling it [15, 16]. Early detection and intervention in prediabetes can not only prevent these complications but also delay or avert the progression to diabetes, proving to be cost-effective [17, 18].

Machine learning, a subfield of artificial intelligence (AI), leverages historical data to enhance decision-making, employing statistical, probabilistic and optimization algorithms to classify new data inputs [19]. Traditionally, robust statistical methods such as multivariate regression and correlation analysis were employed to develop models by linearly integrating relevant variables [20]. However, the digitization of medical records has amassed extensive multidimensional data within healthcare databases, presenting a prime opportunity for sophisticated machine learning techniques in pattern recognition and prediction [21].

Machine learning diverges from conventional statistical methods by utilizing a diverse array of parameters, including Boolean logic, absolute constraints, conditional probabilities, and innovative optimization techniques for classification, mimicking human-like decision-making processes. Although rooted in statistical and probabilistic principles, machine learning has emerged as a superior tool for classification, capable of deriving decisions or inferences from datasets beyond the scope of traditional statistical methods [22]. While machine learning has significantly enhanced the screening and prediction of diabetes, there is a noticeable scarcity of studies applying these techniques to the diagnosis of prediabetes [23]. This gap underscores the need for further research into the application of machine learning in the early detection and diagnosis of prediabetes.

To facilitate the broader integration of machine learning in healthcare, enhancing the explainability of these models is crucial. Implementing explainable AI (XAI) techniques improves the transparency of machine learning operations [24]. XAI enables healthcare professionals to predict and comprehend the factors contributing to increased disease risk, a critical factor for machine learning's acceptance in medical settings. One effective XAI technique is Local Interpretable Model-Agnostic Explanations (LIME) [25], which is widely used to elucidate predictions from opaque machine learning models. The strength of LIME lies in its model-agnostic nature, allowing its application across various machine learning models regardless of their underlying architecture. This method focuses on identifying and highlighting the significance of specific data features on an individual patient level, thereby enhancing the interpretability of the model's outcomes. This feature-specific insight is pivotal for clinicians to trust and effectively use machine learning predictions in their decision-making processes [26].

Deep learning, a branch of machine learning, has achieved notable success across various domains including reinforcement learning [27], audio analysis [28], natural language processing [29], and computer vision [30]. Despite these advancements, deep learning has not been as effective or popular for analyzing tabular data, where it generally underperforms compared to conventional tree-based models such as Random Forests, Decision Trees and XGBoost in real-world applications [31, 32]. This is particularly true in scenarios involving small tabular datasets, where traditional deep learning approaches often fall short [32]. Nonetheless, a recent innovation in this area is the HyperTab [33] model, a hypernetwork-based deep learning strategy designed specifically for small tabular datasets. HyperTab combines the strengths of Random Forests and neural networks to create an ensemble of neural networks. Each network within the ensemble is tailored to handle a specific lower-dimensional slice of the dataset. This novel approach has shown superior performance over traditional machine learning models on various small tabular datasets [33–35], highlighting its potential to enhance data analysis in contexts where deep learning has previously underachieved.

In this research, we introduce a novel hybrid model, HyperTab-LIME, which integrates the HyperTab model with the LIME framework to predict prediabetes in men aged over 30. We utilized data from 1527 male participants over the age of 30 from the 2022 Korean National Health and Nutrition Examination Survey (KNHANES), and evaluated the performance of our proposed model against several baseline models. To establish a reliable and interpretable diagnostic tool for prediabetes for healthcare professionals, we combined the optimized HyperTab model with LIME. This innovative hybrid approach provides detailed and precise explanations of its predictive outcomes, enhancing the medical community's ability to understand and trust the decisions made by the model.

2. Literature review

Research on the application of machine learning to prediabetes is still relatively sparse compared to its use in diabetes prediction. To date, only a handful of studies have explored machine learning techniques for predicting prediabetes. One notable study by Choi *et al.* [36] utilized data from the Korean National Health and Nutrition Examination Survey (KNHANES), specifically selecting individuals without diabetes. They used data from 2010 (n = 4685) for training and internal validation and data from 2011 (n = 4566) for external validation. They

created two models, an artificial neural network (ANN) and a support vector machine (SVM), to screen for prediabetes and compared their efficacy with a previously developed logistic regression-based screening score model. The SVM model demonstrated superior performance with an AUC of 0.731 on the external dataset, outperforming both the ANN model and the logistic regression-based model.

Kushwaha *et al.* [37] focused on an Indian national dataset covering children and adolescents aged 5–19 years to craft a non-invasive screening model for prediabetes using machine learning. The dataset included 26,567 participants, categorized into normal (n = 23,777) and prediabetes (n = 2790) based on HbA1c levels. The XGBoost model achieved the highest cross-validation accuracy at 90.13%, significantly outperforming other models like Random Forest (RF), Gradient Boosting (GB), SVM, Decision Tree (DT) and Extra Trees (ET).

Silva *et al.* [38] analyzed data from the National Health and Nutrition Examination Survey 2013–2014, involving 6346 participants. The data were divided into training (n = 3174) and internal validation sets (n = 3172), with an external validation set prepared from the 2011–2012 survey data (n = 3000). They applied four machine learning algorithm (Logistic Regression, ANN, RF and GB) across 46 exposure variables to develop a prediabetes prediction model. The models were also tested using different resampling methods to balance class distribution. The RF model, particularly when using minority class oversampling, achieved the highest performance, demonstrating an AUC of 71.58% on internal validation data and 70.01% on external validation data.

Recently, XAI methodologies have been extensively applied in numerous studies focusing on machine learning applications for diabetes prediction tools. For example, Tasin *et al.* [39] created an automated diabetes prediction system using various machine learning classifiers, including XGBoost paired with Adaptive Synthetic (ADASYN) sampling to address data imbalances. This approach achieved a high level of performance, recording an accuracy of 81% and an AUC of 0.84. They further enhanced the system's transparency by integrating XAI tools such as LIME and Shapley Additive Explanations (SHAP), which elucidate the model's predictive decisions.

Similarly, Kim *et al.* [40] employed XAI, specifically LIME, to investigate key factors influencing comorbid conditions associated with diabetes, including cancer, heart disease and mental health issues. Their findings highlighted the critical roles of body mass index, socioeconomic status, life satisfaction, and family support in managing diabetes and its related health complications. In another study, Aelgani *et al.* [41] introduced an ensemble explainable model designed to predict diabetes. Their approach utilized a soft voting classifier that combines five different machine learning algorithm: Logistic Regression, K-Nearest Neighbors, DT, RF and AdaBoost. This ensemble method achieved an accuracy of 81% on the Pima Indian diabetes dataset, surpassing traditional predictive models. They also applied LIME to enhance the interpretability of the ensemble model, aiding medical experts in understanding the predictions more clearly.

Comparing with diabetes, the prediction of prediabetes remains a challenging area due to biased accuracy and a lack of explainability in many existing machine learning methods. Hence, our research aims to address these issues by enhancing the interpretability and predictive performance of machine learning models for prediabetes, proposing models that not only predict effectively but also provide reliable and comprehensible insights into disease risks.

3. Materials and methods

3.1 Data source and study workflow

This research leveraged data from the 2022 Korea National Health and Nutrition Examination Survey (KNHANES) [42], administered by the Korea Centers for Disease Control and Prevention under the auspices of the Korean Ministry of Health and Welfare. KNHANES is a comprehensive national survey designed to collect a wide array of data spanning social, economic, nutritional and health dimensions. This includes information gathered through health interviews, behavior surveys, clinical examinations, and nutritional assessments, covering aspects such as healthcare utilization, anthropometric measurements, biochemical profiles, quality of life and health behaviors. The overview of KNHANES dataset is described in Table 1. The original 2022 KNHANES dataset contained 6265 samples. For the purposes of this study, we focused on male participants, reducing the sample size to 2797. We further refined the dataset by excluding males with missing values in the target column "HE_DM_HbA1c", resulting in 2177 samples. The analysis was limited to men aged 30 and above, hence narrowing down the sample to 1909. After excluding men under the age of 30 and those without diagnosed diabetes, the final dataset comprised 1527 records of men aged 30 or older with prediabetes or normal state. This subset was used to develop a precise model for predicting prediabetes, as depicted in Fig. 1.

Fig. 2 depicts the flow of the study. The dataset was divided into training, validation and hold-out sets through a random stratified split, adhering to a 70:15:15 distribution. The training set, comprising 1069 participants, was employed to train all models. The validation set, consisting of 229 participants, served to assess model performance and facilitate the optimization of each model's hyperparameters. The hold-out set was used to evaluate the models' performance on previously unseen data, providing an unbiased assessment of their effectiveness. Before training, a data preprocessing step was implemented to ensure the data was suitable for training models. This was followed by a feature selection phase using the minimum Redundancy Maximum Relevance (mRMR) technique to decrease the dataset's dimensionality and isolate the most predictive features for prediabetes. Several classification models were then trained and their prediction performances evaluated in order to find the best performance model. Finally, the LIME framework was integrated with the best model to extract insights from its predictions and improve interpretability. For the implementation of this study, the computational environment was configured using Google Colab. The experiments were conducted using Python 3.10.12 (Python Software Foundation, Wilmington, DE, USA), Pytorch 2.3.0 (Meta AI, Astor Place, New York City, NY, USA)

TABLE 1. Overview of KNHANES dataset.

Category	Description
Survey Components	- Health Interview: Addresses socioeconomic factors, lifestyle choices like smoking and alcohol use, physical activity, mental well-being, and chronic health conditions. - Health Examination: Includes body measurements, blood pressure readings, clinical tests, pulmonary function testing, dental health assessments, visual and auditory examinations. - Nutrition Survey: Collects data on dietary habits, food intake, nutritional supplements, and food security.
Survey Methods	- Health Interview: Conducted face-to-face or via self-administered questionnaires in mobile centers. - Health Examination: Performed in mobile examination centers. - Nutrition Survey: Face-to-face interviews at the participant's residence.
Indicators	Encompasses approximately 250 health indicators related to lifestyle behaviors, nutritional status, and chronic disease conditions.

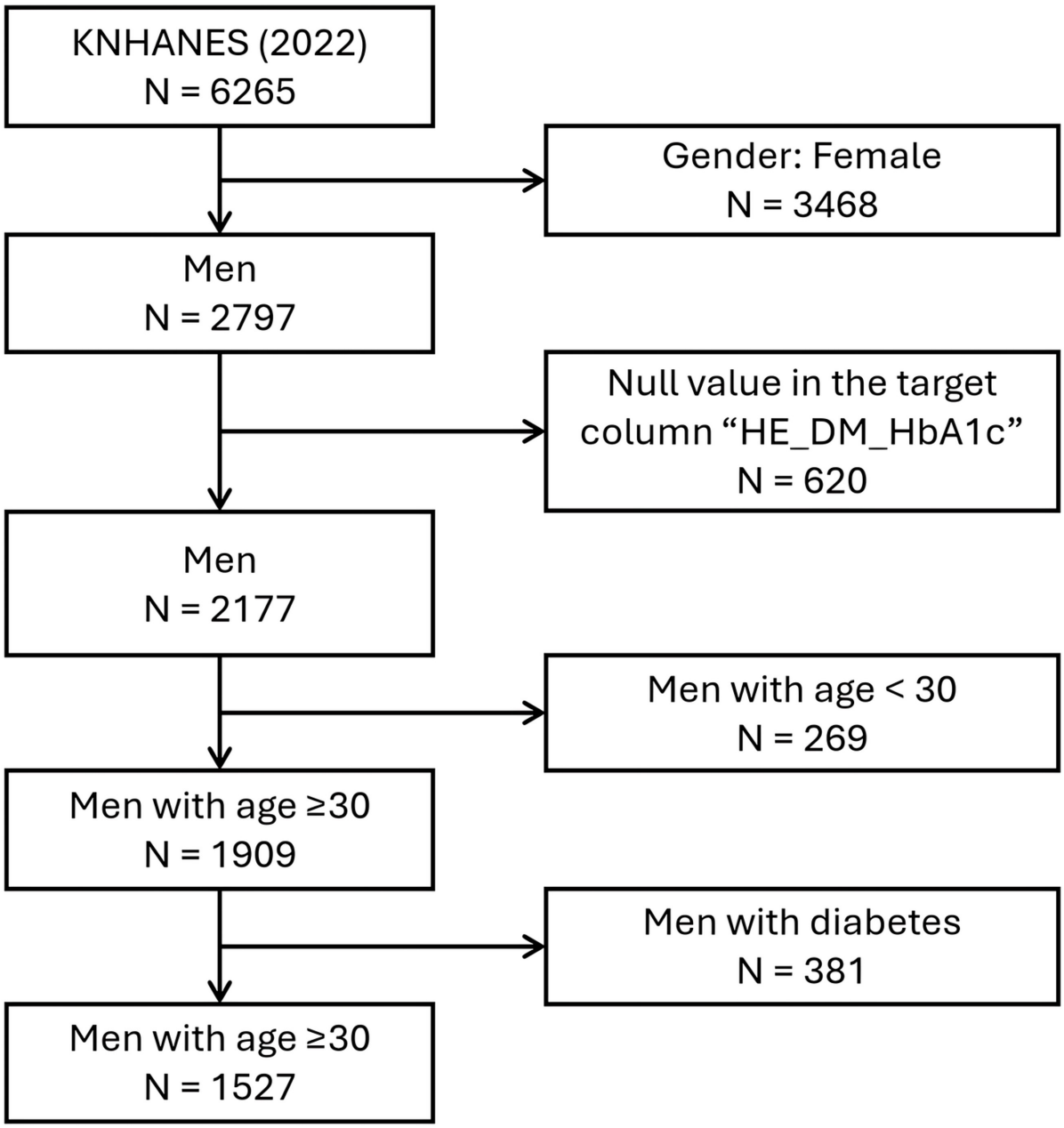

FIGURE 1. Population selection flowchart. KNHANES: Korean National Health and Nutrition Examination Survey; HE_DM_HbA1c: the status of diabetes prevalence (comprising two categories: normal and prediabetes).

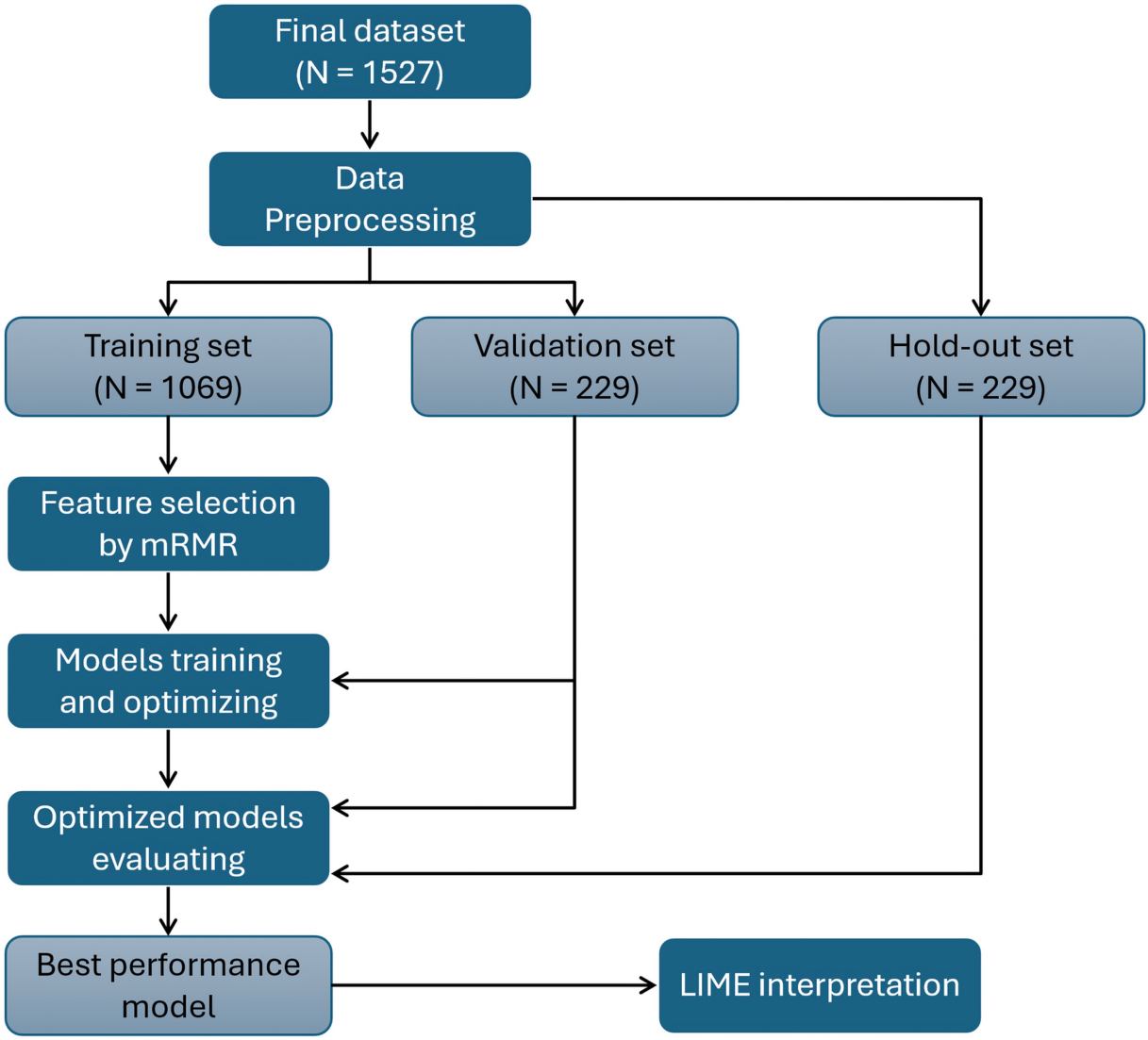

FIGURE 2. Study workflow. mRMR: minimum Redundancy Maximum Relevance; LIME: Local Interpretable Model-Agnostic Explanations.

and sklearn 1.2.2. Additional libraries required for specific tasks or analyses can be installed as needed.

3.2 Data preprocessing

Initially, our dataset comprised 1527 records of men aged over 30, containing 623 variables. We eliminated columns with more than 70% missing values and irrelevant columns such as date, time, personal identifiers and address. Additionally, we removed all variables associated with medically diagnosed diabetes and treatment information. These exclusions reduced the dataset to 297 variables (detailed in **Supplementary Table 1**) devoid of missing values.

Following this initial filtering, the refined dataset featured 212 categorical variables and 85 numerical variables. We applied Label Encoding to the categorical variables and normalized the numerical variables using a Standard Scaler to standardize the data for model training. The target variable, "HE_DM_HbA1c", which indicates the status of diabetes prevalence, comprised two categories: normal and prediabetes. These were encoded using Label Encoding, with "normal" assigned a value of 0 and "prediabetes" a value of 1. The distribution of these classes within the target feature was nearly balanced at 769 for class 0 and 758 for class 1. The processed dataset was subsequently divided into training, validation, and hold-out sets, with the specific numbers of samples in each detailed in Fig. 2.

3.3 Feature selection

After preprocessing, our dataset comprised 297 variables, necessitating the application of a feature selection technique to minimize noise variables, prevent overfitting, and reduce the time required to train models. Feature selection techniques are commonly categorized into three types: filter, wrapper and embedded methods [43]. For this study, we chose the filter strategy due to its computational efficiency and broad generalizability across various machine learning models [43]. Specifically, we employed the "Minimum Redundancy Maximum Relevance" (mRMR) [44] method. The mRMR method has demonstrated robust generalization capabilities and has been extensively validated across various healthcare datasets.

For instance, Özyurt *et al.* [45] utilized the mRMR feature selection method to enhance a fused Convolutional Neural Network (CNN) model for analyzing white blood cell tests. Similarly, Eroglu *et al.* [46] developed a hybrid CNN model that employs mRMR for the classification of Alzheimer's disease. Kumar *et al.* [47] applied mRMR as a feature selection filter in diagnosing heart disease using Electrocardiogram (ECG) data. Furthermore, the mRMR method has been successfully implemented in several diabetes-related machine learning models [48–50], underscoring its effectiveness in enhancing model accuracy in diverse medical applications.

The rationale behind selecting the mRMR method is its effectiveness in eliminating redundant features while retaining those most relevant to the model. It is recognized that the m best features are not the best m feature, due to the presence of correlation and redundancy among key features. mRMR addresses this by selecting features based on their relevance to the target variable and their redundancy relative to previously selected features. At each iteration, the relevance of a feature f is determined by the F-statistic between the feature and the target. The redundancy is calculated as the mean Pearson correlation between the feature and all previously selected features. Thus, a score is computed for each feature under consideration at iteration i, as detailed in Eqn. 1 below:

$$score_i\,(f) = \frac{F(f, target)}{\frac{1}{|S|}\sum_{s\in S}|corr(f,s)|} \tag{1}$$

Where i represents the iteration number, S is the subset of features selected until $i-1$, $|S|$ is size of subset S, f is the feature being evaluated, F is the F-statistic, and $corr$ is the Pearson correlation.

In our study, we employed the mRMR feature selection method using the mrmr_selection package (accessible via https://github.com/smazzanti/mrmr) to identify the optimal number of features (k) that enhance prediction accuracy for our target variable. Initially, we conducted training and validation across all models with default hyperparameters, using the complete set of 297 variables to find a baseline model with the best performance. Following the selection of the best base model, we implemented a loop to vary k from 10 to 100. During each iteration of the loop, we applied the mRMR method to select the top k features for that iteration and retrained the best base model using only these selected features. We recorded the performance metrics (accuracy, precision, recall and AUC) for each iteration. The optimal k value was determined as the one where the selected k features yielded the maximum AUC score compared to other k values. The process for this selection is outlined in the code shown in Fig. 3.

3.4 Classification models

3.4.1 Proposed model: HyperTab

HyperTab represents an innovative approach to constructing an ensemble of neural networks, specifically designed for handling small tabular datasets, as illustrated in Fig. 4. This method integrates an ensemble strategy with a feature subsetting augmentation mechanism, which significantly expands the training dataset and allows for the application of more complex network architectures. This augmentation ensures that the class labels remain unchanged, maintaining the integrity of the dataset while facilitating dimensional reduction.

The architecture of HyperTab includes a primary hypernetwork and a series of target networks. The hypernetwork is responsible for generating the weights for each target network based on a given augmentation identifier, which defines a subset of features. This setup enables each target network to process data represented only by these selected features, effectively increasing the dataset's variability by treating each (augmentation, example) pair as a unique training instance. Notably, the target networks' parameters are not independently optimized but are derived from the hypernetwork, reducing the number of parameters that need to be trained directly. The only parameters trained directly are those of the hypernetwork itself, which streamlines the learning process and reduces computational overhead.

The HyperTab model not only increases the effective number of training examples through augmentation but also allows for the automated design of network architectures tailored to specific feature subsets. Each target network processes a lower-dimensional view of the data, enabling more efficient learning from small datasets. HyperTab thus offers a potent and adaptable tool for deep learning on challenging small tabular data cases. The flexibility of this model is further enhanced by its ability to customize various parameters through its implementation package (can be accessed at https://github.com/wwydmanski/hypertab). This customization includes settings for the train-test split ratio, computational device, number of test nodes, training epochs, and dimensions of hidden layers, providing a comprehensive tool for tackling classification tasks in tabular datasets.

3.4.2 Baseline models

In this research, we evaluated the performance of the HyperTab model against several established machine learning algorithms, such as AdaBoost [51], Gradient Boost [52], XGBoost [53], LightGBM [54], Random Forest [55], and Logistic Regression [56], which have previously been utilized in the development of prediabetes or diabetes prediction models [23, 36–41]. Here is a concise overview of each model:

- AdaBoost: This model employs an ensemble strategy known as "adaptive boosting" to convert a series of weak classifiers into a strong classifier. It sequentially adds classifiers to the ensemble, correcting the predictions of prior classifiers until the training set is accurately classified or a predetermined limit of classifiers is reached. This method was incorporated into an explainable ensemble model for diabetes prediction in research by Aelgani *et al.* [41].
- Gradient Boost: Known for its speed and efficiency, Gradient Boost reduces bias errors through a method where decision trees are incrementally added to minimize the model's prediction errors. It operates under a greedy algorithm that can potentially lead to overfitting if not properly managed. Depending on its application as a regressor or classifier, it uses Mean Square Error (MSE) or Log loss as its cost function respectively. Kushwaha *et al.* [37] and Silva *et al.* [38] have

Optimal Feature Selection by mRMR

Input: Training and testing datasets (X_train, Y_train, X_test, Y_test)

Output: Optimal *k*-value and corresponding performance metrics

1. Initialize arrays for storing:
   - *k*-values from 10 to 100
   - AUC score for each *k*

2. FOR each *k* in the range 10 to 100:
   2.1. Perform feature selection using the mRMR method with parameter *k*
      - Selected_features ← mrmr_classif(X_train, Y_train, *k*)

   2.2. Subset the training dataset to include only the selected features
      - X_train_mrmr ← X_train[;, Selected_features]

   2.3. Apply a predefined model function with the subsetted training and testing datasets
      - Model_output ← Model(X_train_mrmr, X_test[:, Selected_features], Y_train, Y_test)
      - Extract performance metric (AUC) from Model_output

   2.4. Store each metric in its respective array

3. Determine the optimal *k*-value that maximizes the AUC metrics
   - Best_*k* ← *k*-values[IndexOfMaximumValue(AUC_array)]

4. Output the optimal *k*-value and its corresponding performance metrics
   - Print("Optimal *k*-value:", Best_*k*)

FIGURE 3. Pseudo code for finding optimal *k* value. mRMR: minimum Redundancy Maximum Relevance; AUC: Area Under the Receiver Operating Characteristic Curve.

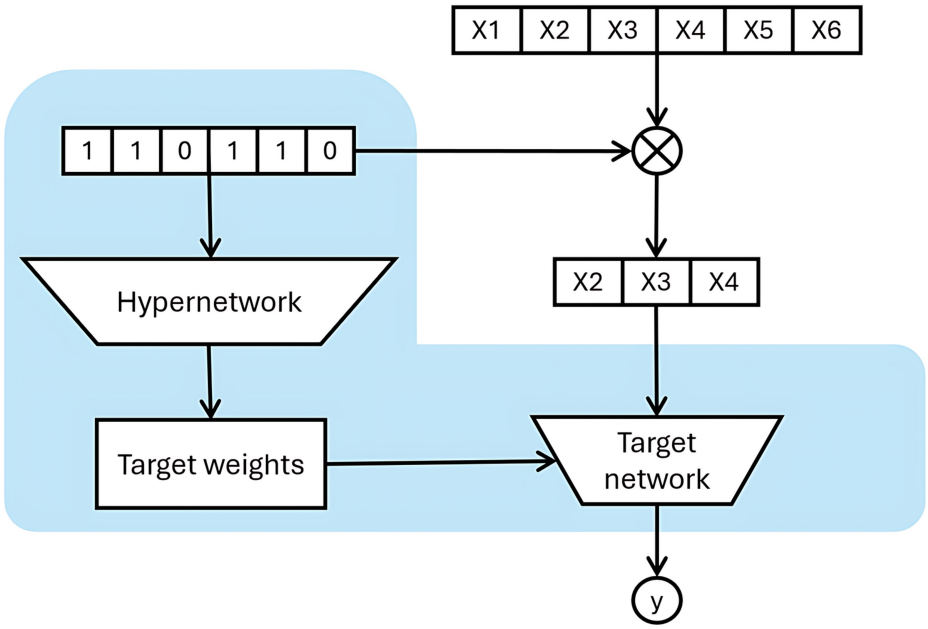

FIGURE 4. Architecture of HyperTab model. This configuration involves a hypernetwork that utilizes a binary mask to select a subset of features, thus defining the augmentation. The hypernetwork then produces the weights for the target network, which processes the data in a reduced dimensionality as specified by this mask. The output of the HyperTab model is derived from the collective responses of multiple target networks, each corresponding to different individual augmentations.

used this model to benchmark prediabetes prediction efforts.

• XGBoost: This is an advanced implementation of gradient boosting that has gained popularity for its predictive accuracy in various competitions. XGBoost is a scalable and efficient framework that emphasizes performance and speed in tree boosting, enhancing the predictive power by combining multiple weak classifiers into a robust model. Kushwaha *et al.* [37] highlighted its efficacy in predicting prediabetes among children and adolescents.

• LightGBM: As a gradient boosting framework, LightGBM builds the model by adding weak learners using a gradient descent algorithm. It stands out for its efficiency in memory usage and speed, primarily due to its innovative Gradient-based One-Side Sampling (GOSS) technique. Yuk *et al.* [57] applied this model effectively in prediabetes prediction.

• Random Forest: Frequently used in various predictive tasks, Random Forest constructs multiple decision trees during training and outputs the mode of the classes (classification) or mean prediction (regression) of the individual trees. It randomly selects subsets of features and samples to build each tree, thereby increasing diversity in the model and reducing overfitting. This model was noted for its superior performance in prediabetes prediction in research by Silva *et al.* [38].

• Logistic Regression: This is a statistical method for binary classification that provides probabilities for the outcomes. It estimates the probability of a binary response based on one or more predictor variables, making it suitable for situations where the response variable is categorical. This model was applied as part of an explainable ensemble model for diabetes prediction by Aelgani *et al.* [41] and it served as a comparative baseline in studies by Choi *et al.* [36] and Silva *et al.* [38].

In addition to evaluating various traditional machine learning models, our study also examines the effectiveness of the HyperTab model against another deep learning model tailored for small tabular datasets, known as TabPFN [58]. TabPFN represents a significant innovation as a deep learning transformer specifically optimized for rapid supervised classification tasks on small datasets. Remarkably, it operates efficiently without the need for extensive hyperparameter tuning and compares favorably with leading classification techniques. TabPFN employs in-context learning (ICL), where it processes and makes predictions based on sequences of labeled example pairs (x, f(x)) provided directly in the input. This method allows TabPFN to generate predictions without the need for additional parameter adjustments post-training. The model architecture of TabPFN is designed to handle both training and test data in a set-valued input format, enabling it to output predictions for the entire test set in a single forward pass. Functioning as a Prior-Data Fitted Network (PFN), TabPFN is pretrained once and uses this training to perform what approximates Bayesian inference on synthetic datasets that are generated based on a predefined prior. This prior is influenced by principles of causal reasoning, favoring structural causal models that are not only comprehensive but also inherently simpler in their construction. This approach allows TabPFN to integrate and leverage causal relationships effectively within its predictive framework.

3.5 Performance metrics

In this research, the efficacy of various prediction models was evaluated using several key performance indicators, including the Area Under the Receiver Operating Characteristic Curve (AUC), accuracy, recall and precision. The calculations for these metrics are based on the following equations:

$$AUC = \int_{1}^{0} TPR\left(t_i\right) d(FPR\left(t_i\right)) \tag{2}$$

$$accuracy = \frac{True_{positive} + True_{negative}}{Total\ predictions} \tag{3}$$

$$recall = \frac{True_{positive}}{True_{positive} + False_{negative}} \tag{4}$$

$$precision = \frac{True_{positive}}{True_{positive} + False_{positive}} \tag{5}$$

Here, $FPR(t_i)$ and $TPR(t_i)$ represent the false positive rate and true positive rate corresponding to threshold t_i. In addition, $True_{negative}$ and $True_{positive}$ refer to accurate predictions for the "normal" (class 0) and "prediabetes" (class 1) categories, respectively. Conversely, $Flase_{negative}$ and $Flase_{positive}$ represent incorrect predictions for class 0 and class 1, respectively.

In our analysis, we assumed that the model demonstrating the highest AUC value possessed the superior predictive capability. In cases where the AUC values were identical, the model with the highest accuracy was considered optimal.

3.6 Optimizing hyperparameters

In this study, with the exception of the TabPFN model, all models underwent hyperparameter optimization using the Optuna framework [59]. Optuna is a versatile and open-source hyperparameter optimization framework in Python that facilitates automated and parallel optimization processes, enhancing both the flexibility and interpretability of the tuning efforts [59]. For the TabPFN model, there is a single parameter named "N_ensemble_configurations" that was systematically trained across a range from 2 to 200 to determine the optimal setting that maximizes the AUC score. The specific ranges for hyperparameter tuning of the HyperTab model and other baseline models are detailed in Table 2.

3.7 LIME: local interpretable model-agnostic explanations

LIME, developed by Ribeiro *et al.* [25], employs a surrogate model to deliver localized explanations of model predictions. It elucidates how a model's predictions vary in response to modifications in the input data, creating a new dataset of perturbed instances alongside their corresponding outputs from the black box model. This altered dataset facilitates the training of an interpretable model, where the weighting is adjusted

TABLE 2. Hyperparameter tuning ranges.

Model	Hyperparameters	Range or Values
HyperTab		
	Subsample (subsample)	0.3 to 0.8 (step = 0.1)
	Hidden Dimensions (hidden_dims)	2 to 20
	Test Nodes (test_nodes)	2 to 64
	Learning Rate (lr)	$(3 \times 10^{-5}, 3 \times 10^{-4}, 3 \times 10^{-3}, 3 \times 10^{-2}, 3 \times 10^{-1})$
	Epochs (epochs)	(20, 30, 50, 70, 100)
LightGBM		
	Number of Leaves (num_leaves)	20 to 120
	Number of boosting rounds (n_estimators)	50 to 500
	Learning Rate (learning_rate)	0.001 to 0.2 (log scale)
	Max Depth (max_depth)	3 to 10
Gradient Boost		
	Number of boosting rounds (n_estimators)	50 to 500
	Learning Rate (learning_rate)	0.001 to 0.2 (log scale)
	Max Depth (max_depth)	3 to 10
XGBoost		
	Number of Leaves (max_leaves)	20 to 120
	Number of boosting rounds (n_estimators)	50 to 500
	Learning Rate (learning_rate)	0.001 to 0.1 (log scale)
	Max Depth (max_depth)	3 to 12
AdaBoost		
	Number of Estimators (n_estimators)	50 to 500
	Learning Rate (learning_rate)	0.001 to 1 (log scale)
Random Forest		
	Number of Trees (n_estimators)	100 to 1000
	Max Features (max_features)	(“auto”, “sqrt”, “log2”)
	Max Depth (max_depth)	5 to 30
Logistic Regression	Inverse of Regularization Strength (c)	0.01 to 100 (log scale)
TabPFN	N_ensemble_configurations	2 to 200

HyperTab: hypernetwork-based deep learning approach designed for small tabular datasets; TabPFN: HyperTab model against another deep learning model tailored for small tabular datasets.

according to how closely related the instances are to the focal instance. The constraints of these local surrogate models, which aim to maintain simplicity while maximizing explanatory power, can be mathematically expressed as shown in Eqn. 6:

$$Exp\,(x) = argmin_{m \in M} \mathcal{L}\,(f, m, \pi_x) + \Omega(m) \qquad (6)$$

In this formulation, Exp(x) denotes the explanatory function for a particular instance x, striving to minimize the loss $\mathcal{L}$, which represents the mean squared error. The objective is to align the surrogate explanation model m as closely as possible with the predictions of the original model f, while keeping the complexity of m, denoted as Ω(m), to a minimum. The term π_x quantifies the proximity of the data points in the vicinity of instance x, indicating the breadth of the neighborhood considered for the explanation. In this study, we chose two particular cases to illustrate the functionality of the LIME model in conjunction with our most effective model for predicting prediabetes in men aged 30 and older. We employed the LIME package, available at https://github.com/marcotcr/lime, to provide interpretative insights into the model’s predictive decisions.

4. Results

4.1 Feature selection results

Initially, we conducted training and validating all models using their default hyperparameters across the entire dataset of 297 variables to establish a performance benchmark. As outlined in Table 3, the evaluation of all models prior to feature selection indicated that the Gradient Boost model achieved the highest

TABLE 3. Performance of models with default settings before feature selection.

Model	AUC	Accuracy	Recall	Precision
Gradient Boost	0.8223	0.7249	0.6491	0.7629
AdaBoost	0.8070	0.7424	0.7105	0.7570
Random Forest	0.7908	0.6943	0.6754	0.7000
XGBoost	0.7886	0.7118	0.6316	0.7500
LightGBM	0.7871	0.7424	0.6316	0.8090
Logistic Regression	0.7568	0.6900	0.7105	0.6807
HyperTab	0.7566	0.6507	0.6491	0.6491
TabPFN	0.6269	0.5939	0.6316	0.5854

AUC: Area Under the Receiver Operating Characteristic Curve; HyperTab: hypernetwork-based deep learning approach designed for small tabular datasets; TabPFN: HyperTab model against another deep learning model tailored for small tabular datasets.

AUC score of 0.8223 with default hyperparameters. Consequently, this model was selected as the foundational model for further exploration of the optimal number of features (k) using the mRMR feature selection method.

Fig. 5 illustrates the performance of this base model across different values of k, ranging from 10 to 100. In this figure, the AUC scores for the base model are depicted in blue. The analysis revealed that the optimal number of features, k, is 38. These 38 features, which identified through the mRMR method, significantly enhanced model performance as shown in Table 4. Descriptions of these features are detailed in Table 5.

Analyzing the results from Table 4, which details the performance of various predictive models post feature selection, we observed a notable improvement in model performance metrics across several models when compared to their performance in Table 3. TabPFN exhibits the most significant improvement. Its AUC improved dramatically from 0.6269 to 0.8344, showcasing a major enhancement in the model's predictive capabilities and overall accuracy, which also increased substantially. HyperTab also shows remarkable improvement, with its AUC jumping from 0.7566 to 0.8293. This suggests that the model responds well to a more focused set of features, enhancing both its accuracy and its ability to correctly classify positive cases. The Gradient Boost model maintained its strong performance, slightly enhancing its AUC and showing a notable increase in accuracy. This model remains one of the top performers, particularly excelling in accuracy among the leading models. The remaining models also showed improvements in their AUC scores though not as dramatically as TabPFN or HyperTab. Upon examining the results from Table 4 in comparison with the initial performances listed in Table 3, it becomes evident that feature selection has significantly influenced the performance of the models evaluated.

4.2 Evaluating optimized models

Excluding the TabPFN model, HyperTab and other conventional machine learning models were subject to hyperparameter optimization using the Optuna framework, with each model undergoing 100 iterations to refine settings. The specific hyperparameters fine-tuned during this process are catalogued in Table 6. The performance outcomes post-optimization are enumerated in Table 7. From these results, HyperTab emerged as the standout model, recording the highest AUC of 0.8429, which underscores its superior predictive effectiveness. Furthermore, HyperTab showcased strong and balanced performance in terms of accuracy, precision and recall, indicating its efficacy across various metrics.

Following AUC performance of the HyperTab model, the TabPFN model exhibited an AUC of 0.8344, which was not improve but its accuracy was increased into 0.7555. The AdaBoost model, with an accuracy equal to HyperTab at 0.7686 and an impressive AUC of 0.8340, was particularly notable for its high precision of 0.8020. The remaining models also displayed improvements in their AUC scores following the hyperparameter adjustments. It was observed that the boosting models, while achieving high precision, tend to have lower recall values. In contrast, models like HyperTab, TabPFN and Logistic Regression maintained a more balanced performance between recall and precision, highlighting their capability to provide consistent and reliable predictions across different evaluation metrics.

Fig. 6 presents the receiver operating characteristic (ROC) curve for all optimized models evaluated on a hold-out set. In this assessment, the proposed HyperTab model also excelled among all models, achieving an AUC of 0.8300. This performance underscores the capability of the HyperTab model to effectively predict outcomes on data it has not previously encountered, suggesting its practical applicability in real-world scenarios. The consistent performance of the optimized HyperTab model on both the validation and hold-out sets reaffirms its robustness and reliability in predicting prediabetes, making it a valuable tool for clinical and healthcare analytics.

4.3 HyperTab-LIME evaluation

To enhance the interpretability of predictions made by the HyperTab model and to facilitate its use in a healthcare assistance tool, the LIME framework was integrated with the optimized HyperTab model. This integration was aimed at elucidating how the hybrid HyperTab-LIME model predicts prediabetes in males over the age of 30. For detailed analysis, two specific instances were chosen, as illustrated in Fig. 7.

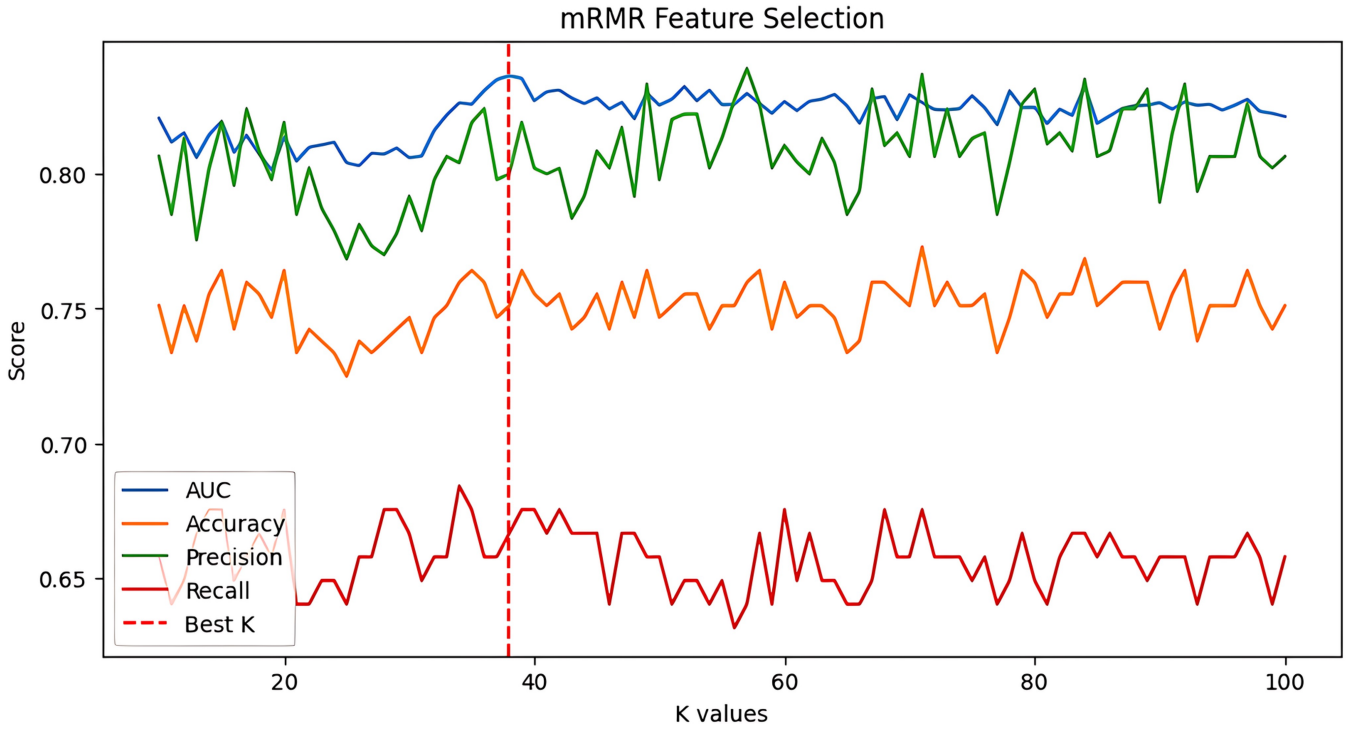

FIGURE 5. Performance of the default gradient boost model during feature selection process. mRMR: minimum Redundancy Maximum Relevance; AUC: Area Under the Receiver Operating Characteristic Curve.

TABLE 4. Performance of models with default settings after feature selection.

Model	AUC	Accuracy	Recall	Precision
TabPFN	0.8344	0.7424	0.7281	0.7477
HyperTab	0.8293	0.7205	0.7368	0.7119
Gradient Boost	0.8286	0.7522	0.7105	0.7714
LightGBM	0.8221	0.7478	0.6842	0.7800
AdaBoost	0.8075	0.7435	0.7018	0.7619
Random Forest	0.8149	0.7336	0.6579	0.7732
XGBoost	0.8062	0.7293	0.6667	0.7600
Logistic Regression	0.7795	0.7087	0.7281	0.6975

AUC: Area Under the Receiver Operating Characteristic Curve; HyperTab: hypernetwork-based deep learning approach designed for small tabular datasets; TabPFN: HyperTab model against another deep learning model tailored for small tabular datasets.

Fig. 7A presents a case study of a 70-year-old male predicted to have prediabetes, with the model assigning an 84% probability to this outcome. The visualization highlights influential features in orange, emphasizing their role in this particular prediabetes prediction. Among the total of 38 variables, the following 10 features were identified as having the most significant impact on the prediction:

- age = 70. Age is 70.
- DI1_ag = 888. Hypertension diagnosis period is not physically diagnosed.
- DI1_2 = 5. Taking Blood Pressure Regulators: Not physically diagnosed.
- HE_sbp3 = 170. Third systolic blood pressure: 170 mmHg.
- N_EN = 1410.29. Daily energy intake: 1410.29 Kcal.
- BM13_5 = 2. Reasons for tooth damage: Other—Not applicable (no experience of tooth damage).
- incm5 = 4. Income quintile (individual): High.
- HE_nc = 40.20. Neck circumference: 40.20 cm.
- HE_alt = 21. Aspartate transaminase (ALT): 21 IU/L.
- BM14_1 = 1. Reasons for not receiving dental treatment: Because the symptoms are mild (I think it will get better over time).

Fig. 7B illustrates a case study involving a 34-year-old male predicted by the model to be in a normal health state, with an 82% probability of accuracy. This section of the visualization uses shades of blue to emphasize the features that significantly influence this prediction of normalcy, aiding in the interpretive process of the model's output. Out of the total 38 variables considered by the model, the following 10 features

TABLE 5. Selected features and their description.

Variable	Description	Type	Values and description
age	Age	Continuous	1–79: 1–79 years old 80. “80 years of age or older”
incm5	Income quintile (individual)	Categorical	0. Low 1. Low-mid 2. Middle 3. Middle-high 4. High
ainc_1	Gross household income	Continuous	~ ×10,000 KRW 200. Less than 2 million won per year More than 200 to less than 18,000 (continuous income amount) 18,000. More than 180 million won per year In the case of question 6-1-2 17. less than 170,000 won per month More than 17 to less than 1500 (continuous income amount) 1500. More than 15 million won per month 9,999,999. No response
DI1_ag	Hypertension Diagnosis Period	Continuous	~ years old 0–79: 0–79 years old 80: 80 years of age or older 888: Not physically diagnosed 999: No response
DI1_pt	Treatment for high blood pressure	Categorical	0. None 1. Yes 2. Not physically diagnosed 3. No response
DI1_2	Taking Blood Pressure Regulators	Categorical	0. Taking daily 1. Taking more than 20 days a month 2. Taking more than 15 days a month 3. Take less than 15 days a month 4. Don’t take it 5. Not physically diagnosed 6. No response
DI2_pt	Treatment of dyslipidemia	Categorical	0. None 1. Yes 2. Not physically diagnosed 3. No response
DI2_2	Dosing of dyslipidemia	Categorical	0. Taking daily 1. Taking more than 20 days a month 2. Taking more than 15 days a month 3. Take less than 15 days a month 4. Don’t take it 5. Not physically diagnosed 6. I don’t know. No response
BD1_11	The frequency of drinking for a year	Categorical	0. I haven’t drank at all in the last year 1. less than once a month 2. About once a month 3. 2–4 times a month 4. 2–3 times a week 5. More than four times a week 6. Not applicable 7. No response

TABLE 5. Continued.

Variable	Description	Type	Values and description
BD7_61	Damage caused by other peoples drinking for a year: disturbance	Categorical	0. No, it's not 1. Yes 2. Not applicable 3. No response
BS2_1	Smoking General Tobacco (Combustion) Start Age	Continuous	~ years old 0–79: 0–79 years old 80: 80 years of age or older 888. Not applicable 999. No response
BE3_94	Physical activity time (minutes): location movement	Categorical	~ minutes 0: 0 1: 10 2: 13 3: 15 4: 20 5: 22 6: 25 7: 26 8: 30 9: 34 10: 35 11: 38 12: 40 13: 41 14: 50 15: Not applicable 16: No response
BE3_88	Medium-intensity physical activity time (minutes): leisure	Categorical	~ minutes 0: 0 1: 10 2: 13 3: 15 4: 20 5: 25 6: 30 7: 35 8: 40 9: 50 10: Not applicable 11: No response
HE_fst	An empty stomach	Continuous	Hungry hours (h)
HE_sbp1	Primary systolic blood pressure	Continuous	Primary systolic blood pressure (mmHg)
HE_dbp1	Primary diastolic blood pressure	Continuous	Primary diastolic blood pressure (mmHg)
HE_sbp2	Secondary systolic blood pressure	Continuous	Secondary systolic blood pressure (mmHg)
HE_sbp3	Third systolic blood pressure	Continuous	Third systolic blood pressure (mmHg)
HE_ht	Kidney	Continuous	Kidney (cm)
HE_wc	waist circumference	Continuous	Waist circumference (cm)
HE_nc	Neck circumference	Continuous	Neck circumference (cm)
HE_BMI	Body mass index	Continuous	Body mass index (kg/m^2)
HE_TG	Neutral fat	Continuous	Neutral fat (mg/dL)
HE_ast	AST	Continuous	(IU/L)
HE_alt	ALT	Continuous	(IU/L)

TABLE 5. Continued.

Variable	Description	Type	Values and description
HE_crea	Blood creatinine	Continuous	(mg/dL)
HE_WBC	White blood cell count	Continuous	White blood cell count (Thous/μL)
HE_Bplt	Platelet count	Continuous	Platelet count (Thous/μL)
HE_Ualb	Urinary albumin	Continuous	Urinary albumin (μg/mL)
HE_Ukal	Potassium	Continuous	Urinary potassium (mmol/L)
BM13_5	Reasons for tooth damage: Other	Categorical	0. No 1. Yes 2. Not applicable (no experience of tooth damage) 3. Don't know
BM14_1	Reasons for Dental Treatment	Categorical	0. Because I don't have time (I can't leave work because it doesn't open when I want, I don't have anyone to watch my kids, *etc*.) 1. Because the symptoms are mild (I think it will get better over time) 2. Economic reasons (because medical expenses are burdensome) 3. Transportation is inconvenient and the distance is long. 4. I don't want to wait a long time at the hospital. 5. It is difficult to make a reservation at a hospital or clinic. 6. Medical treatment (fear of getting tested or treated) 7. Others 8. Not applicable (no, no medical treatment required) 9. Don't know
N_EN	Energy intake (Kcal)	Continuous	Daily energy intake (Kcal)
N_SFA	Saturated fatty acid intake (g)	Continuous	Daily saturated fatty acid intake (g)
N_N3	N-3 Fatty Acid Intake (g)	Continuous	Daily n-3 fatty acid intake (g)
N_CHO	Carbohydrate intake (g)	Continuous	Daily carbohydrate intake (g)
N_VA_RAE	Vitamin A (retinol active equivalent) intake (μgRAE)	Continuous	Vitamin A (retinol active equivalent) intake (μg RAE)
N_VITC	Vitamin C intake (mg)	Continuous	Daily vitamin C intake (mg)

KRW: Korean Won; AST: Aspartate transaminase, ALT: Alanine aminotransferase.

were identified as having the most substantial impact on this prediction:

- age = 34. Age is 34.
- DI1_ag = 888. Hypertension diagnosis period is not physically diagnosed.
- DI1_2 = 5. Taking Blood Pressure Regulators: Not physically diagnosed.
- HE_alt = 12. ALT (SGPT): 12 IU/L.
- N_CHO = 215.22. Daily carbohydrate intake: 215.22 (g).
- N_EN = 1515.87. Daily energy intake: 1515.87 Kcal.
- BE3_88 = 0. Medium-intensity physical activity time (minutes): leisure—0 minute.
- HE_sbp3 = 170. Third systolic blood pressure: 170 mmHg.
- incm5 = 4. Income quintile (individual): High.
- BM13_5 = 0. Reasons for tooth damage: Other—No.
- HE_Ukal = 89. Urinary potassium: 89 mmol/L.

Upon employing the LIME framework to interpret the predictions for prediabetes prevalence across all hold-out data, a detailed assessment of the factors influencing prediabetes predictions was conducted. Notably, the variable "age" had the most substantial impact, accounting for 15.19% of the prediction's weight. This was followed by the "DI1_ag" (Hypertension Diagnosis Period), which contributed 11.53% to the model's predictions. Other significant factors included "DI1_2" (Taking Blood Pressure Regulators), "HE_alt" (ALT (SGPT)), and "HE_WBC" (White Blood Cell Count), which were responsible for 9.82%, 6.53% and 5.24% of the model's weight, respectively. Collectively, the top five variables accounted for over 48% of the predictive weight. The top 10 influential variables for prediabetes prediction are comprehensively listed and detailed in Fig. 8. The remaining top variables include "N_EN" (Daily energy intake), "N_CHO" (Carbohydrate intake), "HE_sbp3" (Third systolic blood pres-

TABLE 6. Optimized hyperparameters of each model.

Model	Hyperparameters
HyperTab	
	Subsample (subsample) = 0.7
	Hidden Dimensions (hidden_dims) = 20
	Test Nodes (test_nodes) = 53
	Learning Rate (lr) = 0.0003
	Epochs (epochs) = 50
LightGBM	
	Number of Leaves (num_leaves) = 58
	Number of boosting rounds (n_estimators) = 151
	Learning Rate (learning_rate) = 0.02
	Max Depth (max_depth) = −1
Gradient Boost	
	Number of boosting rounds (n_estimators) = 300
	Learning Rate (learning_rate) = 0.05
	Max Depth (max_depth) = 7
XGBoost	
	Number of Leaves (max_leaves)
	Number of boosting rounds (n_estimators) = 300
	Learning Rate (learning_rate) = 0.02
	Max Depth (max_depth) = 10
AdaBoost	
	Number of Estimators (n_estimators) = 400
	Learning Rate (learning_rate) = 0.1
Random Forest	
	Number of Trees (n_estimators) = 800
	Max Features (max_features) = “sqrt”
	Max Depth (max_depth) = 25
Logistic Regression	Inverse of Regularization Strength (c) = 3.79
TabPFN	N_ensemble_configurations = 104

HyperTab: hypernetwork-based deep learning approach designed for small tabular datasets; TabPFN: HyperTab model against another deep learning model tailored for small tabular datasets.

TABLE 7. Performance of hyperparameters optimized models.

Model	AUC	Accuracy	Recall	Precision
HyperTab	0.8429	0.7686	0.7632	0.7565
TabPFN	0.8344	0.7555	0.7281	0.7685
AdaBoost	0.8340	0.7686	0.7105	0.8020
Gradient Boost	0.8338	0.7336	0.6579	0.7732
LightGBM	0.8336	0.7424	0.6579	0.7895
Logistic Regression	0.8288	0.7293	0.7368	0.7241
Random Forest	0.8235	0.7293	0.6579	0.7653
XGBoost	0.8106	0.7380	0.6754	0.7700

AUC: Area Under the Receiver Operating Characteristic Curve; HyperTab: hypernetwork-based deep learning approach designed for small tabular datasets; TabPFN: HyperTab model against another deep learning model tailored for small tabular datasets.

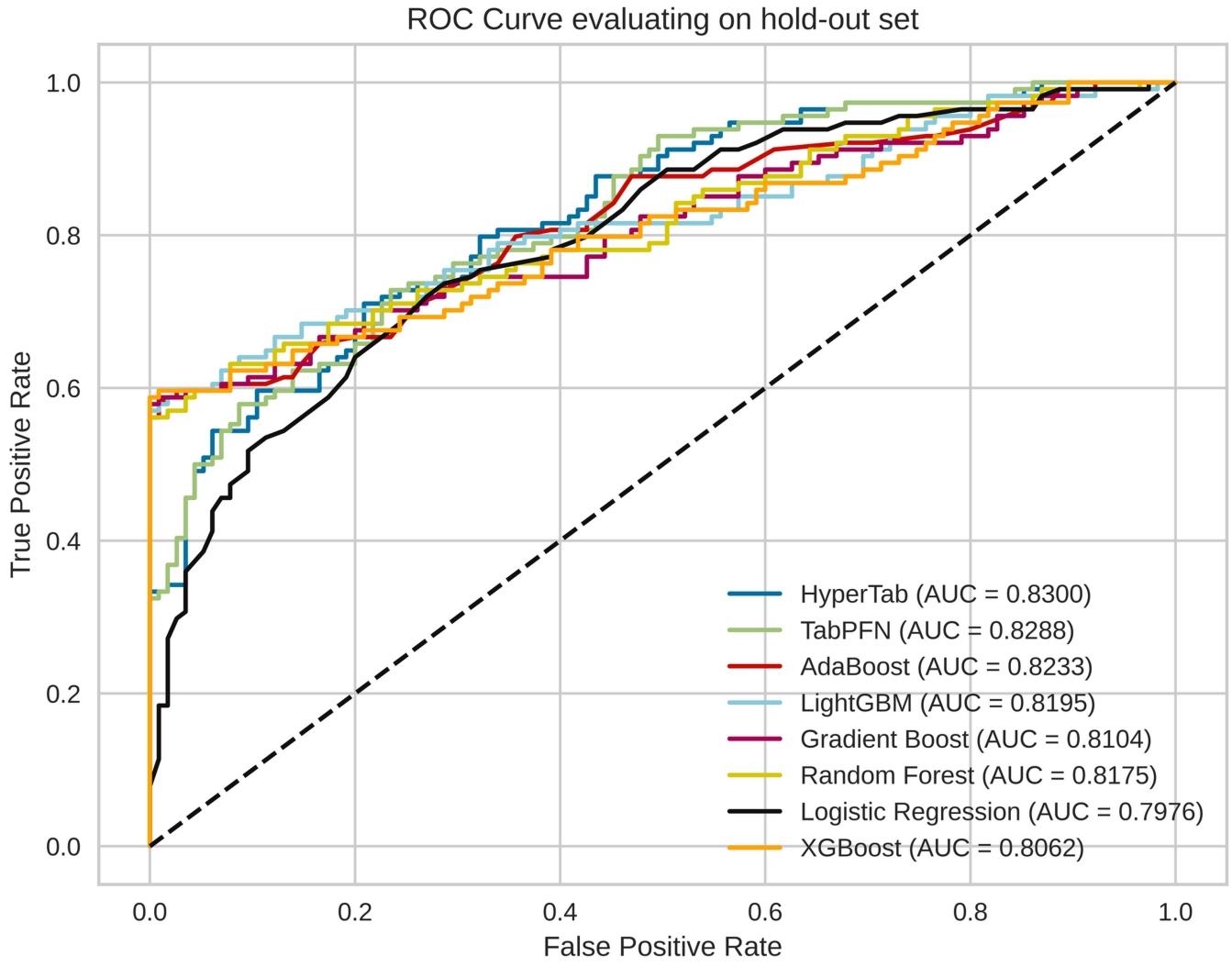

FIGURE 6. ROC curve of optimized models evaluating on hold-out set. ROC: Receiver Operating Characteristic; AUC: Area Under the Receiver Operating Characteristic Curve; HyperTab: hypernetwork-based deep learning approach designed for small tabular datasets; TabPFN: HyperTab model against another deep learning model tailored for small tabular datasets.

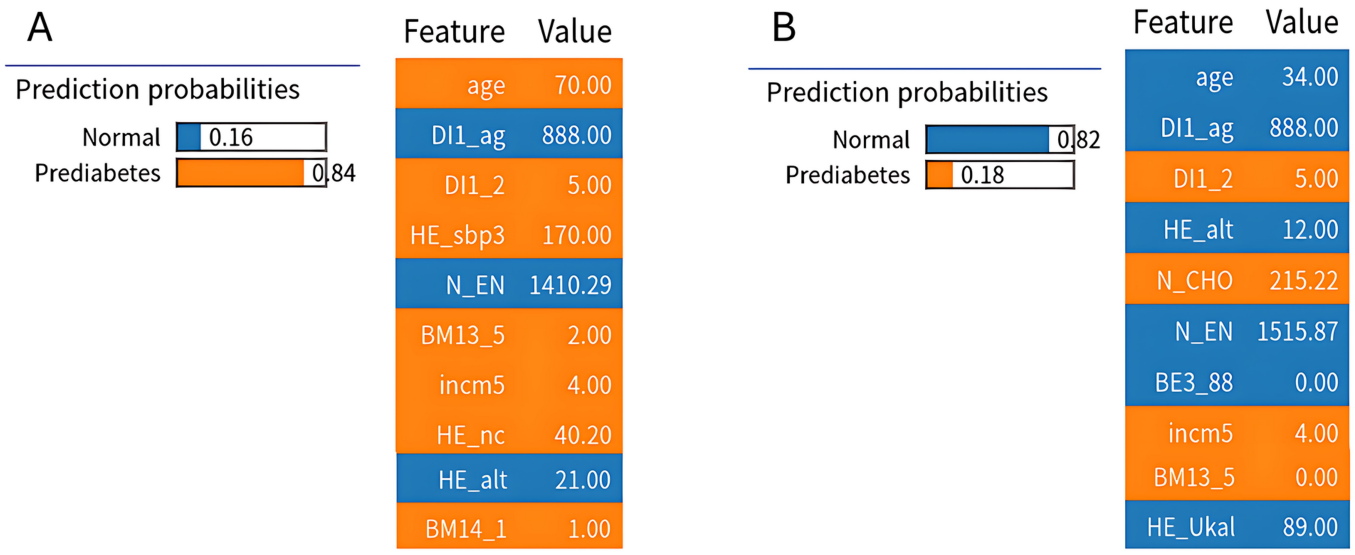

FIGURE 7. HyperTab-LIME explanation examples. (A) Instance with prediabetes state prediction. (B) Instance with normal state prediction.

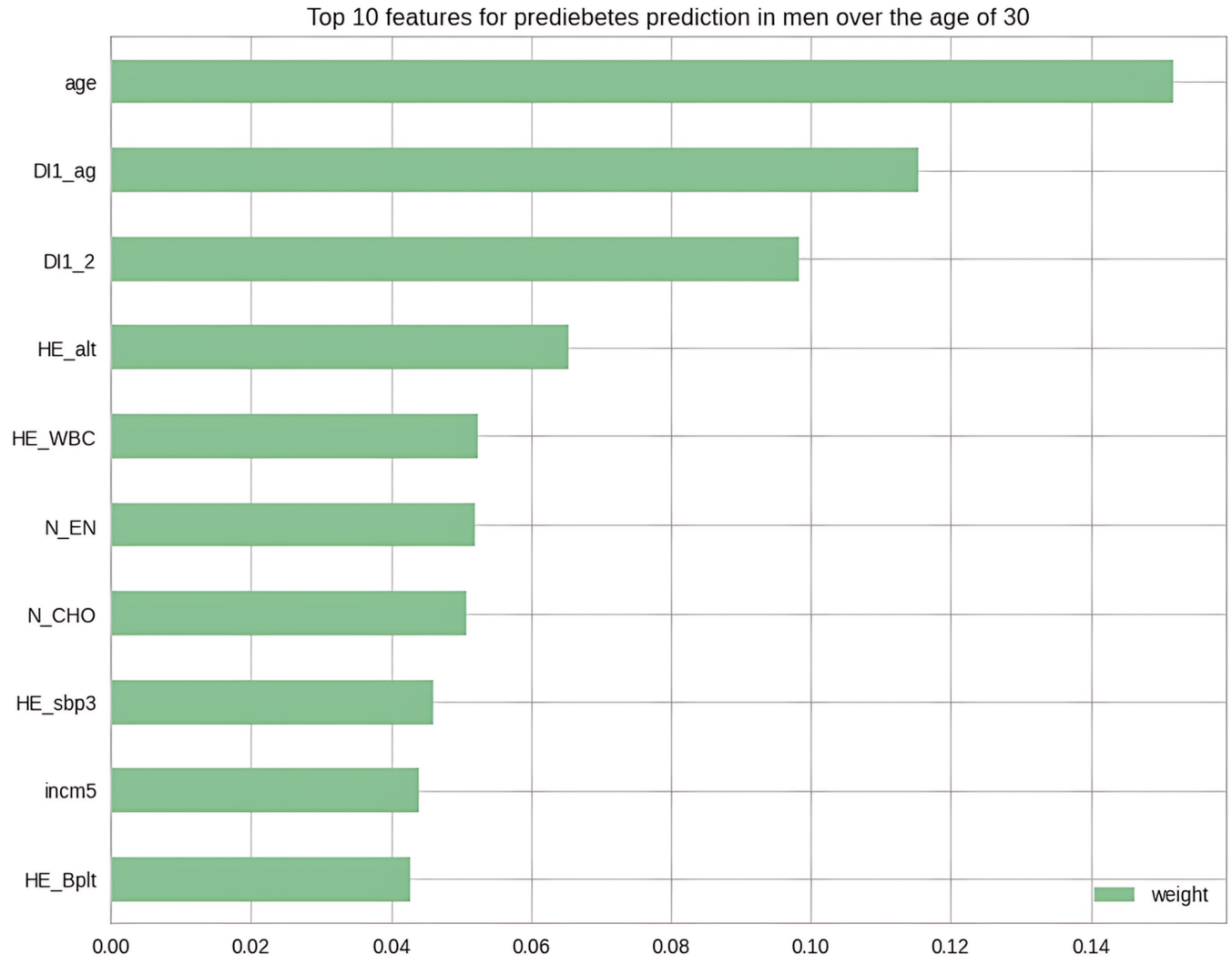

FIGURE 8. Top 10 most impact variables for prediabetes prediction. WBC: White Blood Cell Count; EN: Daily energy intake: CHO: Carbohydrate intake; sbp3: Third systolic blood pressure; incm5: Income quintile (individual); Bplt: Platelet count.

sure), "incm5" (Income quintile (individual)), and "HE_Bplt" (Platelet count).

5. Discussion

This research introduced a hybrid model, HyperTab-LIME, which leverages the HyperTab model to predict prediabetes occurrences in male over the age of 30, with subsequent explanations for each prediction (whether normal or prediabetic) provided by LIME, elucidating the decision-making process of the otherwise opaque "black-box" model. The HyperTab model was trained and optimized to achieve peak predictive performance, evidenced by an AUC of 0.8429 on the validation set and 0.8300 on the hold-out set. To the best of our knowledge, this is the inaugural application of HyperTab to the prediction of prediabetes. Deep learning models typically necessitate a substantial dataset; nevertheless, HyperTab is a deep learning model specifically tailored for small tabular datasets [33]. Hence, this model proved to be appropriate for our small dataset in this study, yielding exceptional performance outcomes. Compared to previous research [36–38], our model significantly improves the interpretability and predictive performance of machine learning models for prediabetes, offering not only effective predictions but also providing detailed and reliable insights into potential prediabetes risks.

This research highlighted the pivotal role of feature selection strategies and hyperparameter optimization in enhancing the performance of models, particularly those based on deep learning architectures. It was observed that once the number of selected features surpassed a specific threshold, the predictive outcomes of all models stabilized. Further addition of features did not augment model accuracy but instead imposed a significant computational load. Given the typically more complex nature of deep learning architectures compared to traditional machine learning architectures, such an increase in computational demands could lead to longer processing times, greater resource consumption, or even diminished model performance due to the inclusion of extraneous variables.

During this study, the mRMR filter selection method was systematically applied to evaluate a range of 10 to 100 top features selected by mRMR on the baseline Gradient Boosting model. The analysis determined that the optimal number of features was 38. Utilizing this optimized set of variables significantly enhanced the performance of all models. Notably,

for the two deep learning models examined, dramatic improvements were observed: the TabPFN model's AUC surged from 0.6269 to 0.8344, and the HyperTab model's AUC increased from 0.7566 to 0.8293. Further optimization of the HyperTab model's hyperparameters raised its AUC performance to 0.8429. Overall, the feature selection and hyperparameter optimization processes boosted the AUC performance of the HyperTab model by nearly 9% and the TabPFN model by nearly 21%. This demonstrated effectiveness of feature selection and hyperparameter optimization in the context of deep learning architecture.

Currently, there is a notable lack of research focusing on predictive models for prediabetes, particularly within the context of the South Korean population. We identified only one previous study that also employed the KNHANES dataset to develop a machine learning model for prediabetes prediction. Specifically, Choi *et al.* [36] used data from KNHANES 2010 (n = 4685) for training and validation, while data from KNHANES 2011 (n = 4566) served for external validation, serving a similar purpose to our hold-out set. In that study, an SVM model was developed to predict prediabetes, achieving AUCs of 0.734 and 0.712 in the internal and external validation sets, respectively. In comparison, our proposed HyperTab model demonstrated superior performance with AUCs of 0.8429 on the validation set and 0.8300 on the hold-out set, significantly outperforming the SVM model's AUC performance reported by Choi *et al.* [36].

Diabetes is more commonly found in men than women, particularly among those aged 35 to 69 years [60]. Further data from the United States indicate that men diagnosed with type 2 diabetes at the ages of 30, 40 and 50 see their life expectancies reduced by approximately 14, 9 and 5 years, respectively, compared to non-diabetic individuals [10]. Given these findings, our HyperTab-LIME model, which specifically targets males aged 30 and over, offers a critical tool for the early detection of prediabetes. This model not only aids healthcare professionals in intervening early in the prediabetes stage to potentially prevent the progression to full-blown diabetes but also enhances the AI model's utility by providing explainable predictions. The integration of the LIME framework into the HyperTab model ensures that each prediction is accompanied by clear, interpretable data that elucidate the factors driving the model's assessments. This explainability is crucial for clinical settings, where understanding the rationale behind diagnostic assessments can significantly influence treatment decisions and improve patient outcomes.

In practical terms, the proposed HyperTab-LIME model can be deployed as part of routine health assessments in clinical environments where early detection is crucial. For instance, during regular health screenings, the model could analyze patient data in real-time to assess the risk of prediabetes based on a variety of factors such as age, blood pressure, and other relevant health metrics. The model's ability to provide explainable predictions is particularly valuable in clinical settings, as it allows healthcare providers to understand the rationale behind each assessment. This transparency is also crucial for patient trust and compliance, as patients are more likely to engage in preventative measures or treatments if they understand the reasons behind the recommendations.

This study has several limitations that warrant mention. Firstly, while the minimum Redundancy Maximum Relevance (mRMR) feature selection method proved effective in our analysis, numerous other feature selection techniques could potentially yield better outcomes. Future research should consider a comprehensive examination of various feature selection methods applied to this high-dimensional dataset to identify the most effective combination of features for enhancing the performance of prediabetes prediction models. Secondly, the scope of our model training and evaluation was confined to eight models: HyperTab, TabPFN, LightGBM, XGBoost, Gradient Boost, AdaBoost, Random Forest and Logistic Regression. Future studies could broaden this range by incorporating additional machine learning models or employing ensemble techniques to improve predictive accuracy. Thirdly, the LIME method, used for explaining individual predictions from the black box machine learning models, does not provide a comprehensive view of how features affect the model's overall predictions. Future work should incorporate other interpretive methods that offer both local and global explanations of the models' decisions. Lastly, the stability and consistency of the explanations provided by LIME might be limited, as these can vary with the specific samples used or the choice of local data points for building the local model. Consequently, the interpretations offered by LIME should be rigorously evaluated and discussed by medical experts to ensure their validity and applicability in clinical settings.

6. Conclusions

Prediabetes, akin to diabetes, significantly escalates the risk of microvascular complications and markedly increases the chances of cardiovascular diseases and overall mortality. Previous research has shown that men diagnosed with type 2 diabetes at age 30 have their life expectancies reduced by an average of 14 years compared to their nondiabetic counterparts. Thus, early detection and intervention in prediabetes among men aged 30 and older are crucial for preventing these health complications and potentially delaying or halting the progression to diabetes, thereby prolonging life. Prediabetes prediction remains a challenging area due to biased accuracy and a lack of explainability in many existing machine learning methods. To address these issues, we introduced a new hybrid model, HyperTab-LIME, which combines the HyperTab model with the LIME framework to predict prediabetes in males over the age of 30. We analyzed data from 1527 male participants aged 30 and older from the 2022 Korea National Health and Nutrition Examination Survey (KNHANES). Our model was compared with several established models, including TabPFN, LightGBM, XGBoost, Gradient Boost, AdaBoost, Random Forest and Logistic Regression. The HyperTab-LIME model demonstrated superior performance, with AUC scores of 0.8429 in the validation set and 0.8300 in the hold-out set, surpassing not only the baseline models in this research but also the performance of an SVM model from a previous study using similar KNHANES data. By integrating the optimized HyperTab model with LIME, we created a reliable and interpretable diagnostic tool for prediabetes. This innovative approach provides clear and

detailed explanations of the predictive outcomes, significantly enhancing the ability of healthcare professionals to understand and trust the model's predictions.

Future research directions should consider the implementation of various feature selection methods and the exploration of alternative classification algorithms to enhance the model's predictive capabilities. Additionally, further research can adopt different methods for model interpretation that provide not only local explanations but also global insights into the workings of the black-box model. These advancements will deepen our understanding of the prediabetes prediction model's decision-making processes and potentially improve its accuracy and reliability.

AVAILABILITY OF DATA AND MATERIALS

The data presented in this study are provided at the request of the corresponding author. The data is not publicly available because researchers need to obtain permission from the Korea Centers for Disease Control and Prevention. Detailed information can be found at: http://knhanes.cdc.go.kr.

AUTHOR CONTRIBUTIONS

HVN and HB—conceptualization, methodology, validation, investigation; HVN—writing–original draft preparation, software, visualization; HB—formal analysis, writing–review and editing, supervision; YC and HB—project administration, funding acquisition. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Prior to the survey, written informed consent was obtained from each study participant. The current study utilized only existing de-identified data. The study was conducted in accordance with the guidelines of the Declaration of Helsinki. The protocol for the 2016–2018 KNHANES was approved by the Institutional Review Board (IRB) of the Korea Centers for Disease Control and Prevention (IRB approval numbers for 2016–2018: 2018-01-03-P-A and 2018-01-03-C-A). Written informed consent was obtained from all participants.

ACKNOWLEDGMENT

Not applicable.

FUNDING

This research Supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-RS-2023-00237287, NRF-2021S1A5A8062526) and local government-university cooperation-based regional innovation projects (2021RIS-003).

CONFLICT OF INTEREST

The authors declare no conflict of interest. Haewon Byeon is serving as a Guest Editor of this journal. We declare that Haewon Byeon had no involvement in the peer review of this article and has no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to XXY.

SUPPLEMENTARY MATERIAL

Supplementary material associated with this article can be found, in the online version, at https://oss.jomh.org/files/article/1851505070884962304/attachment/Supplementary%20material.docx.

REFERENCES

[1] Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, *et al.* IDF diabetes atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Research and Clinical Practice. 2022; 183: 109119.

[2] World Health Organization. Global report on diabetes. 1st edn. World Health Organization: Geneva. 2016.

[3] NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. The Lancet. 2016; 387: 1513–1530.

[4] World Health Organization. Global health estimates: life expectancy and leading causes of death and disability. 2021. Available at: https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates (Accessed: 27 May 2024).

[5] World Health Organization. Global health estimates: leading causes of DALYs. 2021. Available at: https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/global-health-estimates-leading-causes-of-dalys (Accessed: 27 May 2024).

[6] Gwon YG, Han SJ, Kim KH. Trends in the quality of primary care and acute care in Korea from 2008 to 2020: a cross-sectional study. Journal of Preventive Medicine and Public Health. 2023; 56: 248–254.

[7] Oh J, Kim S, Lee M, Rhee SY, Kim MS, Shin JY, *et al.* National and regional trends in the prevalence of type 2 diabetes and associated risk factors among Korean adults, 2009–2021. Scientific Reports. 2023; 13: 16727.

[8] Jung CH, Son JW, Kang S, Kim WJ, Kim HS, Kim HS, *et al.* Diabetes fact sheets in Korea, 2020: an appraisal of current status. Diabetes & Metabolism Journal. 2021; 45: 1–10.

[9] Bae JH, Han KD, Ko SH, Yang YS, Choi JH, Choi KM, *et al.* Diabetes fact sheet in Korea 2021. Diabetes & Metabolism Journal. 2022; 46: 417–426.

[10] Kaptoge S, Seshasai S, Sun L, Walker M, Bolton T, Spackman S, *et al.* Life expectancy associated with different ages at diagnosis of type 2 diabetes in high-income countries: 23 million person-years of observation. The Lancet Diabetes & Endocrinology. 2023; 11: 731–742.

[11] American Diabetes Association. Diagnosis and classification of diabetes mellitus. Diabetes Care. 2013; 37: S81–S90.

[12] ElSayed NA, Aleppo G, Aroda VR, Bannuru RR, Brown FM, Bruemmer D, *et al.* 2. Classification and diagnosis of diabetes: standards of care in diabetes—2023. Diabetes Care. 2022; 46: S19–S40.

[13] Tabák AG, Herder C, Rathmann W, Brunner EJ, Kivimäki M. Prediabetes: a high-risk state for diabetes development. The Lancet. 2012; 379: 2279–2290.

[14] Magliano DJ, Boyko EJ, IDF Diabetes Atlas 10th Edition Scientific Committee. IDF diabetes atlas. 10th edn. International Diabetes Federation: Brussels. 2021.

[15] Coutinho M, Gerstein HC, Wang Y, Yusuf S. The relationship between glucose and incident cardiovascular events. A metaregression analysis of

published data from 20 studies of 95,783 individuals followed for 12.4 years. Diabetes Care. 1999; 22: 233–240.

[16] Port SC, Goodarzi MO, Boyle NG, Jennrich RI. Blood glucose: a strong risk factor for mortality in nondiabetic patients with cardiovascular disease. American Heart Journal. 2005; 150: 209–214.

[17] Tian L, Zhu J, Liu L, Liang Y, Li J, Yang Y. Prediabetes and short-term outcomes in nondiabetic patients after acute ST-elevation myocardial infarction. Cardiology. 2013; 127: 55–61.

[18] Herman WH, Hoerger TJ, Brandle M, Hicks K, Sorensen S, Zhang P, *et al.* The cost-effectiveness of lifestyle modification or metformin in preventing type 2 diabetes in adults with impaired glucose tolerance. Annals of Internal Medicine. 2005; 142: 323–332.

[19] Jaiswal V, Negi A, Pal T. A review on current advances in machine learning based diabetes prediction. Primary Care Diabetes. 2021; 15: 435–443.

[20] Heltberg A, Andersen JS, Sandholdt H, Siersma V, Kragstrup J, Ellervik C. Predictors of undiagnosed prevalent type 2 diabetes—the Danish general suburban population study. Primary Care Diabetes. 2018; 12: 13–22.

[21] L AJP, Sengan S, Kamalam, G G K, Vellingiri J, Gopal J, Velayutham P, *et al.* Medical information retrieval systems for e-Health care records using fuzzy based machine learning model. To be published in Microprocessors and Microsystems. 2020. [Preprint].

[22] Kumar U. Applications of machine learning in disease pre-screening. In Information Resources Management Association (ed.) Research Anthology on Artificial Intelligence Applications in Security (pp. 1052–1084). IGI Global: Hershey, PA. 2020.

[23] Nawi AM, Kamaruddin PSNM, Nordin NRM, Soffian SSS, Baharom M. Machine learning models in prediabetes screening: a systematic review. Journal of Clinical and Diagnostic Research. 2022; 16: LE01–LE09.

[24] Antoniadi AM, Du Y, Guendouz Y, Wei L, Mazo C, Becker BA, *et al.* Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: a systematic review. Applied Sciences. 2021; 11: 5088.

[25] Ribeiro M, Singh S, Guestrin C. "Why Should I Trust You?": explaining the predictions of any classifier. Proceedings of the 2016 Conference of the North American chapter of the association for computational linguistics: demonstrations (pp. 97–101). Association for Computational Linguistics: San Diego, California. 2016.

[26] Nguyen HV, Byeon H. Prediction of Parkinson's disease depression using LIME-based stacking ensemble model. Mathematics. 2023; 11: 708.

[27] Wang X, Wang S, Liang X, Zhao D, Huang J, Xu X, *et al.* Deep reinforcement learning: a survey. IEEE Transactions on Neural Networks and Learning Systems. 2024; 35: 5064–5078.

[28] Prashanth A, Jayalakshmi SL, Vedhapriyavadhana R. A review of deep learning techniques in audio event recognition (AER) applications. Multimedia Tools and Applications. 2024; 83: 8129–8143.

[29] Wu S, Roberts K, Datta S, Du J, Ji Z, Si Y, *et al.* Deep learning in clinical natural language processing: a methodical review. Journal of the American Medical Informatics Association. 2020; 27: 457–470.

[30] Chai J, Zeng H, Li A, Ngai EWT. Deep learning in computer vision: a critical review of emerging techniques and application scenarios. Machine Learning with Applications. 2021; 6: 100134.

[31] Shwartz-Ziv R, Armon A. Tabular data: deep learning is not all you need. Information Fusion. 2022; 81: 84–90.

[32] Grinsztajn L, Oyallon E, Varoquaux G. Why do tree-based models still outperform deep learning on typical tabular data? Advances in Neural Information Processing Systems. 2022; 35: 507–520.

[33] Wydmański W, Bulenok O, Śmieja M. HyperTab: hypernetwork approach for deep learning on small tabular datasets. 2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA) (pp. 1–9). IEEE: Thessaloniki, Greece. 2023.

[34] Sachdeva B, Rathee H, Gambhir P, Bansal P. Empirical analysis of machine learning models on Parkinson's speech dataset. 2023 26th Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA) (pp. 1–5). IEEE: Delhi, India. 2023.

[35] Rathee H, Sachdeva B, Gambhir P, Dev A, Bansal P. Advancing Parkinson's disease detection: an empirical evaluation of machine learning models based on speech analysis. 2023 6th International Conference on Contemporary Computing and Informatics (IC3I) (pp. 2339–2344). IEEE: Gautam Buddha Nagar, India. 2023.

[36] Choi SB, Kim WJ, Yoo TK, Park JS, Chung JW, Lee Y, *et al.* Screening for prediabetes using machine learning models. Computational and Mathematical Methods in Medicine. 2014; 2014: e618976.

[37] Kushwaha S, Srivastava R, Jain R, Sagar V, Aggarwal AK, Bhadada SK, *et al.* Harnessing machine learning models for non-invasive pre-diabetes screening in children and adolescents. Computer Methods and Programs in Biomedicine. 2022; 226: 107180.

[38] De Silva K, Jönsson D, Demmer RT. A combined strategy of feature selection and machine learning to identify predictors of prediabetes. Journal of the American Medical Informatics Association. 2020; 27: 396–406.

[39] Tasin I, Nabil TU, Islam S, Khan R. Diabetes prediction using machine learning and explainable AI techniques. Healthcare Technology Letters. 2023; 10: 1–10.

[40] Kim R, Kim CW, Park H, Lee KS. Explainable artificial intelligence on life satisfaction, diabetes mellitus and its comorbid condition. Scientific Reports. 2023; 13: 11651.

[41] Aelgani V, Gupta SK, Narayana VA. Local agnostic interpretable model for diabetes prediction with explanations using XAI. Proceedings of Fourth International Conference on Computer and Communication Technologies (pp. 417–425). Springer Nature: Singapore. 2023.

[42] Yun S, Oh K. The Korea national health and nutrition examination survey data linked cause of death data. Epidemiology and Health. 2022; 44: e2022021.

[43] Khaire UM, Dhanalakshmi R. Stability of feature selection algorithm: a review. Journal of King Saud University—Computer and Information Sciences. 2022; 34: 1060–1073.

[44] Hanchuan Peng, Fuhui Long, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2005; 27: 1226–1238.

[45] Özyurt F. A fused CNN model for WBC detection with MRMR feature selection and extreme learning machine. Soft Computing. 2020; 24: 8163–8172.

[46] Eroglu Y, Yildirim M, Cinar A. mRMR-based hybrid convolutional neural network model for classification of Alzheimer's disease on brain magnetic resonance images. International Journal of Imaging Systems and Technology. 2022; 32: 517–527.

[47] Kumar N, Kumar D. Imgwo based ann: a new heart disease diagnosis model to classify real world dataset. Indian Journal of Computer Science and Engineering. 2021; 12: 1001–1017.

[48] Gayathri S, Gopi VP, Palanisamy P. Automated classification of diabetic retinopathy through reliable feature selection. Physical and Engineering Sciences in Medicine. 2020; 43: 927–945.

[49] Srinivasu PN, Shafi J, Krishna TB, Sujatha CN, Praveen SP, Ijaz MF. Using recurrent neural networks for predicting type-2 diabetes from genomic and tabular data. Diagnostics. 2022; 12: 3067.

[50] Zou Q, Qu K, Luo Y, Yin D, Ju Y, Tang H. Predicting diabetes mellitus with machine learning techniques. Frontiers in Genetics. 2018; 9: 515.

[51] Freund Y, Schapire RE. Experiments with a new boosting algorithm. Proceedings of the Thirteenth International Conference on International Conference on Machine Learning (pp. 148–156). Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA. 1996.

[52] Friedman JH. Greedy function approximation: a gradient boosting machine. The Annals of Statistics. 2001; 29: 1189–1232.

[53] Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). Association for Computing Machinery: New York, NY, USA. 2016.

[54] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, *et al.* LightGBM: a highly efficient gradient boosting decision tree. Proceedings of the 31st International Conference on Neural Information Processing Systems (pp. 3149–3157). Curran Associates Inc.: Red Hook, NY, USA. 2017.

[55] Breiman L. Random forests. Machine Learning. 2001; 45: 5–32.

[56] Bisong E. Logistic regression. Building machine learning and deep learning models on google cloud platform: a comprehensive guide for beginners (pp. 243–250). Apress: Berkeley, CA. 2019.

[57] Yuk H, Gim J, Min JK, Yun J, Heo TY. Artificial intelligence-based

prediction of diabetes and prediabetes using health checkup data in Korea. Applied Artificial Intelligence. 2022; 36: 2145644.

[58] Hollmann N, Müller S, Eggensperger K, Hutter F. TabPFN: a transformer that solves small tabular classification problems in a second. 2023. Available at: https://arxiv.org/abs/2207.01848 (Accessed: 05 May 2024).

[59] Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: a next-generation hyperparameter optimization framework. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 2623–2631). Association for Computing Machinery: New York, NY, USA. 2019.

[60] Tramunt B, Smati S, Grandgeorge N, Lenfant F, Arnal JF, Montagner A, *et al.* Sex differences in metabolic regulation and diabetes susceptibility. Diabetologia. 2020; 63: 453–461.

How to cite this article: Hung Viet Nguyen, Younsung Choi, Haewon Byeon. An explainable hybrid deep learning model for prediabetes prediction in men aged 30 and above. Journal of Men's Health. 2024; 20(10): 52-72. doi: 10.22514/jomh.2024.166.