Explainable Artificial Intelligence Based Framework for Non-Communicable Diseases Prediction

KHISHIGSUREN DAVAGDORJ[1], **JANG-WHAN BAE**[2], **VAN-HUY PHAM**[3],
NIPON THEERA-UMPON[4,5], **(Senior Member, IEEE),**
AND KEUN HO RYU[3,4], **(Life Member, IEEE)**

[1]School of Electrical and Computer Engineering, Chungbuk National University, Cheongju 28644, Republic of Korea
[2]College of Medicine, Chungbuk National University, Cheongju 28644, Republic of Korea
[3]Faculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City 700000, Vietnam
[4]Biomedical Engineering Institute, Chiang Mai University, Chiang Mai 50200, Thailand
[5]Department of Electrical Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai 50200, Thailand

Corresponding author: Keun Ho Ryu (khryu@tdtu.edu.vn; khryu@ieee.org)

ABSTRACT The rapid rise of non-communicable diseases (NCDs) becomes one of the serious health issues and the leading cause of death worldwide. In recent years, artificial intelligence-based systems have been developed to assist clinicians in decision-making to reduce morbidity and mortality. However, a common drawback of these modern studies is related to explanations of their output. In other words, understanding the inner logic behind the predictions is hidden to the end-user. Thus, clinicians struggle to interpret these models because of their black-box nature, and hence they are not acceptable in the medical practice. To address this problem, we have proposed a Deep Shapley Additive Explanations (DeepSHAP) based deep neural network framework equipped with a feature selection technique for NCDs prediction and explanation among the population in the United States. Our proposed framework comprises three components: First, representative features are done based on the elastic net-based embedded feature selection technique; second a deep neural network classifier is tuned with the hyper-parameters and used to train the model with the selected feature subset; third, two kinds of model explanation are provided by the DeepSHAP approach. Herein, (I) explaining the risk factors that affected the model's prediction from the population-based perspective; (II) aiming to explain a single instance from the human-centered perspective. The experimental results indicated that the proposed model outperforms various state-of-the-art models. In addition, the proposed model can improve the medical understanding of NCDs diagnosis by providing general insights into the changes in disease risk at the global and local levels. Consequently, DeepSHAP based explainable deep learning framework contributes not only to the medical decision support systems but also can provide to real-world needs in other domains.

INDEX TERMS Non-communicable diseases, explainable artificial intelligence, deep shapley additive explanations, feature selection, deep neural network, prediction.

I. INTRODUCTION

NCDs are the major global health issues confronting humankind. According to the NCDs global status report by the World Health Organization, NCDs are the leading cause of death accounting for 41 million people die each year. It is equal to 71% of the 57 million deaths globally [1].

The associate editor coordinating the review of this manuscript and approving it for publication was Li He.

In particular, NCDs are the leading cause of 15 million premature deaths that occurred in adults between 30 and 70 years old annually. The main varieties of NCDs are cardiovascular diseases, diabetes mellitus, cancers and chronic respiratory disease. Remarkably, cardiovascular disease is assumed first with the highest number of 17.9 million, followed by cancers of 9 million, respiratory diseases of 3.9 million, and then diabetes handles 1.6 million people for deaths due to NCDs annually [2]. NCDs are driven by forces of modifiable and

non-modifiable risk factors. As we know well that modifiable risk factors include harmful use of tobacco and alcohol, environmental factors, unhealthy diets, and physical inactivity, which lead to obesity, hypertension and raised cholesterol. On the contrary, non-modifiable risk factors consist the age, sex and genetics [3], [4]. Fortunately, about 80% of all heart disease, stroke, diabetes and 40% of cancers can be prevented if the major risk factors were eliminated [5], [6]. Most NCDs are often diagnosed at a late stage. If NCDs can be predicted before it occurs, healthcare actions can be taken by individuals and reduce patient harm. Hence, there remains a need to construct a decision support system, serving the progression of NCDs for detecting patients with high risk and minimizing the death rate.

Recently, the use of artificial intelligence in the healthcare industry has been rapidly increasing. However, decision-making responses suffer from various problems. To deal with these problems, advanced data-driven and machine learning approaches have been regularly developed in recent researches [7]–[13]. Furthermore, many systems cannot handle high dimensional datasets and select significant features to compute a general weight for them based on their significance due to the lack of a smart framework [7]–[9]. In a study [9], a smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion was proposed. The authors combined the extracted features from both sensor data and electronic medical records using the feature fusion method. After that, irrelevant and redundant features were eliminated based on features information gain technique, and the conditional probability approach computed a specific feature weight for each class, which further improved system performance. Finally, the ensemble deep learning model was trained for heart disease prediction; also, it obtained higher accuracy than other compared methods. Therefore, most of the existing literature considered the accuracy and false-positive rate for assessing the performance of classification algorithms. The absence of other performance measures, such as model build time, misclassification rate, and precision, should be considered the major limitation for classifier performance evaluation [10].

Moreover, time-series data analysis is essential in the healthcare industry for the management of chronic diseases to medical experts could analyze the patient's history when making a progression diagnosis. However, data are usually either limited or not available to use because of their cost, especially in developing countries. In a study [11], authors used a collection of cost-effective time-series features including patient's comorbidities, cognitive scores, medication history, and demographics to predict Alzheimer's disease progression using support vector machine (SVM), RF, k-nearest neighbor (KNN), logistic regression (LR), and decision tree techniques. In their results, the early fusion of comorbidity and medication features with other features revealed significant predictive power with all models. The RF model reached the best predictive performance. Likewise,

complex models such as deep learning and ensemble techniques contribute superior performance for enhancing the diagnosis and treatment of various chronic diseases [12], [13].

However, most of the existing studies have limited explanations of their results. In general, accurate prediction performance and explainability are two dominant criteria of the best decision support system [14]. Accurate prediction performance during testing may set up some degree of trust in the model. Despite their prediction accuracy, a common drawback of these modern studies is related to their black-box model. Thus, understanding the inner logic behind the predictions is hidden to the end-user, also it is challenging to apply them in real-world health care applications. Mostly, medical experts do not trust decisions yielded by black-box models without any explanations [14]–[16]. In the meantime, the selection of a suitable feature set is crucial to remove redundant features that contribute varied benefits such as improve learning accuracy and alleviate better readability and understanding. In healthcare applications, a set of significant feature selections is still a challenging procedure. Numerical studies proposed feature selection methods such as information gain, gain ratio, and correlation coefficients. Nevertheless, these techniques do not consider the interactions among the features and are not suitable for applying them directly to healthcare applications [17]. Furthermore, a limited number of studies have concentrated on optimizing the parameters of machine learning models in order to improve performance.

To address the above problems, we propose the DeepSHAP based deep neural network (DNN) equipped with a feature selection technique to construct an accurate and explainable decision support system. In this study, the National Health and Nutrition Examination Survey (NHANES) dataset is used to build the predictive and explainable decision support model of NCDs. The proposed framework comprises three components. In the first component, representative features are done based on the data cleaning and elastic net (EN) based embedded feature selection technique. In the second component, the DNN classifier is tuned with the hyper-parameters and used to train the model with the selected feature subset. In the last component, two kinds of model explanations are provided by the DeepSHAP approach: (I) explaining the risk factors that affected the model's prediction from the population-based perspective; (II) aiming to explain a single instance from the human-centered perspective. The entire process of modelling, including feature selection, training, hyper-parameter tuning, model evaluation and explanations is considered.

Furthermore, our proposed model is contrasted against state-of-the-art baseline models. For constructing decision support models of NCDs, three different sets of significant features are generated from the NHANES dataset. The generated features are selected by support vector regression-based recursive feature elimination (SVR-RFE), sequential backward feature selection with random forest (SBFS-RF), and proposed EN feature selection techniques. In order to find the optimal combination of features in each subset, features

are kept when it maximizes the model performance. After-wards, these feature subsets used for the model training of all classifiers, namely SVM, KNN, RF, multilayer perceptron (MLP), extreme gradient boosting (XGBoost) and proposed DNN, respectively. This would suggest that machine-learning techniques used in the proposed framework could be interchanged with other efficient techniques of experimental design depending on the domain.

Meanwhile, classifiers are tuned with their corresponding hyper-parameters in order to improve prediction performance and avoid the over-fitting problem. Finally, a comparison of experimental results is conducted between the proposed framework and state-of-the-art baseline models in validation and test datasets for NCDs. The accuracy, specificity, recall, precision, f-scores and area under the curve (AUC) are exploited to evaluate prediction model performances.

The major contributions of this study are:

- We propose DeepSHAP based explainable deep learning framework, which is incorporated with a feature selection approach for early prediction of non-communicable diseases.
- We appraise the effectiveness of the proposed framework on a real-world non-communicable diseases dataset from the NHANES, which was collected among the population of the United States. The results from the empirical study confirm that the proposed model outperforms various state-of-the-art baseline models.
- We define the several feature subset of non-communicable diseases using embedded and model-based feature selection techniques. It helps improve computational speed and prediction accuracy and allows domain experts to understand predictions.
- The proposed framework provides global and local level explanations of the complex deep neural networks model along with population and human-centered perspectives. Experimental findings are better able to represent the decision process of the model and allow giving personalized health recommendations to patients.
- The proposed framework contributes not only to the health care applications of non-communicable diseases but also provide to other domains.

The rest of this paper is structured as follows: In Section II, we present the literature review related to the subject of this research paper. Section III describes the proposed framework and its main three components. In Section IV, we introduce the dataset, experimental setup, detailed process of experimental design. Section V discusses the overall experimental results that were accomplished by the proposed framework and experimental design. Finally, Section VI ends with a conclusion of the current study and some notes on directions of future enhancement.

II. RELATED WORK

The health care industry has been greatly benefited from the advantages of modern technological advances. Machine-learning techniques achieve the remarkable prospect of transformation for the diagnosis and treatment of various chronic diseases. In this section, we discuss the related work in two parts: A) Machine-learning techniques for NCDs; B) Explainable artificial intelligence in health care applications.

A. MACHINE LEARNING TECHNIQUES FOR NON-COMMUNICABLE DISEASES

Various studies have focused on the accuracy enhancement of NCDs diagnostic models concerning feature selection techniques and refined machine-learning classifiers [18]–[29].

According to the study [18], authors proposed to construct the prediction model for multiple diseases applying metagenome data from 1,079 individuals collected among the healthy group and patients with one of six diseases. Authors built prediction models based on LogitBoost, SVM, KNN and logistic model tree classifiers using forward selection and backward elimination techniques. In their comparative results, the LogitBoost performed the highest accuracy of 98.1 among four classifiers. In addition, they suggested the optimal feature subsets at the genus level obtained by backward elimination. Similarly, authors studied multi-label neural network method to predict chronic diseases combining neural network and multi-label learning technology based on cross entropy lost function and backward propagation algorithm [19]. They utilized 19,773 patients with 10 chronic diseases extracted from MIMIC-II database in order to identify the types of chronic diseases.

The authors aimed to build a heart disease prediction model applying feature selection and machine-learning techniques namely, Naive Bayes, generalized linear model, linear regression, deep learning, decision trees, RF, gradient boosted tree and SVM in their study [20]. They used the hearth disease dataset, which included 13 features and 303 patient records, from the UCI repository. Experimental results showed an enhanced performance with an accuracy level of 88.7 through the prediction model for heart disease with the hybrid RF with a linear model. In another study [21], a chronic kidney disease dataset from the Department of Nephrology, Huadong Hospital, and Shanghai Fudan University Affiliated Hospital was utilized to develop a prediction model of chronic kidney disease progression. The authors compared machine learning classifiers including LR, EN, LASSO, ridge, SVM, RF, KNN, NN and XGBoost, also analyzed the importance of variable factors in each predictive model. The empirical results from their experiments indicated that EN, LASSO regression, ridge regression and LR showed the highest overall predictive power, with an average AUC and precision above 0.87 and 0.80, respectively.

To construct an accurate early prediction model of cervical cancer, the authors solved outlier and class imbalance problems [22]. First, they used outlier detection methods such as density-based spatial clustering of applications with noise and isolation forest. Then, the synthetic minority over-sampling technique (SMOTE) and SMOTE with Tomek link applied to solve the class imbalance problem. Finally,

a random forest (RF) classifier was used to predict cervical cancer. In the study [23], the authors studied a consolidated decision tree-based intrusion detection system for binary and multiclass imbalanced datasets. In their study, an improved version of the random sampling mechanism called supervised relative random sampling proposed to generate a balanced sample from a high class-imbalanced dataset at the pre-processing stage of detector.

Therefore, wearable sensors play a key role in contributing a new method to collect patient data for efficient healthcare monitoring. However, one effort is accompanied by a large amount of healthcare data generated from wearable sensors and social networking data. Hence, authors [24] introduced a novel healthcare monitoring framework based on the cloud environment and a big data analytics engine to store and analyze healthcare data and to improve classification accuracy. Their proposed big data analytics engine was utilized data mining techniques, ontologies, and bidirectional long short-term memory (Bi-LSTM). Bi-LSTM classifier utilized to predict drug side effects and abnormal conditions in patients. In another study [25], authors designed a computerized process of classifying skin disease through deep learning-based MobileNet V2 and LSTM using a skin disease dataset that contained over 10,000 dermatoscopic images that are collected from different people around the world. Their designed method outperformed several methods such as fine-tuned neural networks, convolutional neural networks, very deep convolutional networks for large-scale image recognition developed by visual geometry group, and convolutional neural network, and also it obtained minimal computational efforts.

As determined machine learning techniques for NCDs, the wrapper, embedded or hybrid feature-selection techniques have been adopted more efficiently to select important feature subsets than common filtered techniques [26], [27]. Moreover, it should be noted that most researchers in previous studies aimed to apply the feature selection techniques for not only higher accuracy but also an improvement in understanding the causes of NCDs. The NCDs predictive results of previous studies implied that DNN, SVM, Ensemble classifiers achieved the best performances when compared with other baseline models [28], [29].

B. EXPLAINABLE ARTIFICIAL INTELLIGENCE IN HEALTH CARE APPLICATIONS

Explainable artificial intelligence is presently becoming popular with multi-disciplines, including health care applications. We have recognized this trend as more crucial for health care experts to overcome several challenges such as readiness of outcomes. Meanwhile, few numbers of studies intended to solve the black-box issue in the health care area [16], [30]–[34].

In a study [16], authors aimed to study the utility of various model-agnostic explanation techniques of machine learning models for predicting individuals at risk of developing hypertension based on cardiorespiratory fitness data.

That data included 23,095 patients who underwent treadmill stress testing by physician referrals and was collected from Henry Ford Affiliated Hospitals between 1991 and 2009. For assisting the better understanding of the outcomes of the prediction, five global (Feature Importance, Partial Dependence Plot, Individual Conditional Expectation, Feature Interaction, Global Surrogate Models) and two local (Local Surrogate Models, Shapley Value) interpretable techniques were applied. Only the RF classifier was utilized to predict the outcome because these authors had already compared with the LogitBoost, Bayesian Network classifier, Locally Weighted Naive Bayes, ANN, SVM and RF in their previous research study [30] on the same dataset. Their comparison results performed the best AUC of 0.93 by RF among other classifiers.

In another study [31], authors developed explainable clinical predictive models for stroke outcome using a dataset of 514 patients accessed from a committee of Charite Universitatsmedizin Berlin. The NN and tree boosting modern classifiers were used for the prediction and explanation of the outcome. In order to explain the outcomes, they used deep Taylor decomposition for MLP and CATBOOST algorithm with SHAP values for tree boosting. In addition, predictive performance and explanations of the generalized linear model, lasso and elastic net were used to compare with NN and tree boosting models.

The researchers studied dynamic and explainable machine learning prediction of mortality in patients in the intensive care units using longitudinal data from patients admitted to four ICUs in the Capital Region, Denmark, between 2011 and 2016 in the study [32]. A recurrent neural network was trained with a temporal resolution. After that, the SHAP algorithm was applied to the prediction model for obtaining explanations of the features that drive patient-specific predictions to mitigate the issue of black-box predictions at any given time point. According to the study [33], the authors developed an accurate and interpretable Alzheimer's disease diagnosis and progression detection model. This model provided physicians with accurate decisions along with a set of explanations for every decision using 11 modalities of 1048 subjects from the Alzheimer's disease Neuroimaging Initiative real-world dataset. For model explainability, authors used global and instance-based explanations of the RF classifier by using the SHAP, significantly.

Therefore, most studies were limited to coverage including both problems of prediction accuracy and explainability in the decision model of NCDs. It was noted about some issues accompanied to global and local level explanations through black-box models. Global explanations can explain the decision-making process of prediction models in general, but it is limited to explain the reason for the individual level. The local approach can explain the conditional interaction between features and classes for a single instance. Local explanations can be more accurate than global explanations [34]. Thus, we propose DeepSHAP based deep learning framework, which incorporated a feature selection technique
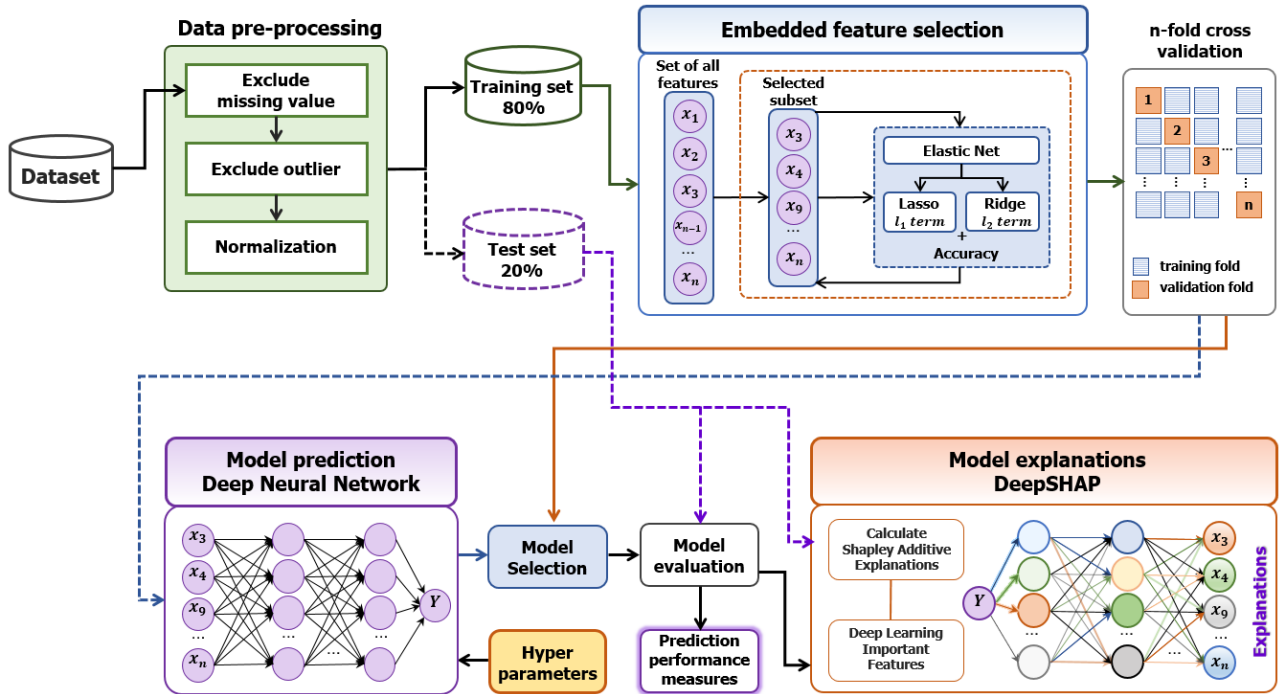
FIGURE 1. Deep shapley additive explanations based deep neural network framework equipped with elastic-net based feature selection technique.

for prediction and interpretation of NCDs in order to solve existing both problems.

III. PROPOSED DEEPSHAP BASED EXPLAINABLE DEEP LEARNING FRAMEWORK

In this paper, we propose the DeepSHAP based DNN framework equipped with an EN feature selection technique to build an accurate and explainable decision support system as illustrated in FIGURE 1. The proposed framework incorporates main three components: data pre-processing with feature selection, DNN prediction model construction and model explanation.

A. DATA-PREPROCESSING
Data pre-processing is a crucial step where the dataset is prepared for training before constructing the classification models. There are several important steps in data pre-processing, such as data cleaning, scaling and feature selection.

Data cleaning is a basic procedure of preparing data for analysis by removing missing values and outliers. The problem of missing values is common in data, which appears when data values are not stored for the features in an observation. In statistics, an outlier refers to the observations that have an abnormal distance from other values on the tails of the distribution. Therefore, the lost data and outliers are excluded in our data analyzing process because it can be the cause of bias in model estimation.

In addition, high-ranged values can intrinsically affect the results of prediction algorithms. Thus, normalization is needed when features have a highly different values range. The normalization is used in the pre-processing step to rescale

the feature values in the interval of 0 to 1. The equation for normalization [35] is derived by initially deducting the minimum value from the variable to be normalized. The minimum value is deducted from the maximum value, and then the previous result is divided by the latter as follows Equation 1:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

where: X is a feature and X_{max} and X_{min} are the maximum and minimum similarity values. X' represents the normalized feature value.

B. ELASTIC-NET BASED EMBEDDED FEATURE SELECTION TECHNIQUE
Regularization methods have become popular, which select the feature subset efficiently and prevent over-fitting issues. Elastic net (EN) is a regularized multiple regression method that combines the l_1 and l_2 penalties of the LASSO (Least Absolute Shrinkage and Selection Operator) and Ridge to solve the high-dimensional feature selection problem [36].

LASSO method executes continuously shrinkage and selects features automatically at the same time. l_1 penalty minimizes the size of all coefficients and allows some coefficients to be minimized to the value zero, which eliminates the predictor from the model. However, LASSO cannot deal with the prediction of features correlations. On the contrary, l_2 penalty of ridge regression is to penalize a model based on the residual sum of squared coefficient values. Ridge regression is efficient where there are dependencies between the features in the model. The ridge cannot produce

a parsimonious model, because it prevents any coefficients from being removed from the model.

Thus, EN method brings the strong advantages of LASSO and ridge regressions, and addresses the drawbacks of both. EN [37] uses a combination of the l_1 (LASSO) and l_2 (Ridge) penalties and it can be defined as Equation 2:

$$\hat{\beta} = arg \min_{\beta}(\|y - X\beta\|^2 + \lambda_2 \|\beta\|^2 + \lambda_1 \|\beta\|_1) \quad (2)$$

where: $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$ are two regularization parameters. Adding a quadratic part to the penalty, EN removes the limitation on the selected features and stabilizes the selection from grouped variables. EN finds an estimator in a two-stage procedure: first, for each fixed λ_2 it finds the ridge regression coefficients and then does a LASSO type shrinkage.

C. DEEP NEURAL NETWORK AND HYPER-PARAMETERS

A deep neural network (DNN) is an artificial intelligence-based technique, which inspired by the structure of the human brain function. NN is a group of interconnected neurons that learn together to perform a particular function. The common configuration of NN is to employ three or more layers namely input, hidden and output. FIGURE 2 shows the neural network architecture with one hidden layer. First, the n number of input nodes $\{x_1, x_2, x_3, \ldots, x_{n-1}, x_n\}$ at the input layer is determined by features. Then, these input nodes are transferred to the hidden layer, and the corresponding n number of weights $\{w_1, w_2, w_3, \ldots, w_{n-1}, w_n\}$ multiplies each connection between nodes during training. At least one hidden layer contains the weighted nodes. Nodes in neighboring layers are interconnected, but nodes in the same layer are not. If a network does not have enough hidden nodes, the input and output mappings cannot learn well. Finally, the output of hidden nodes depends on the classes' number, which is estimated by applying an activation function and make final decisions Y based on optimized weights by minimizing the error between predicted and actual values [38], [39].

The performance of the DNN is highly associated with the selection of hyper-parameters. Neural networks can be learned via an efficiently one-hidden-layer MLP, however, learning with two or more hidden layers of perceptrons can produce much better performance than a single one. Thus, in the proposed framework, DNNs consist to turn the {2-10}
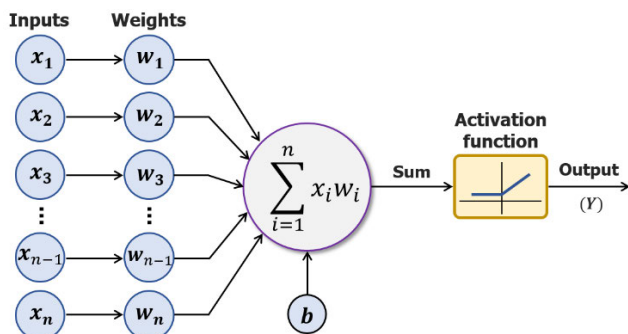

FIGURE 2. Neural network architecture.

hidden layers. Here, per hidden layer varies the number of nodes {3, 5, 10, 15, 20, 25} was tested to compare the performance between three activation functions: Rectified linear units (ReLU), Sigmoid (sigm) and Tanh (tanh). In the experimental result, we provide the optimal hidden layers and their nodes regarding the highest predictive performance. Also, these models are optimized by Adam, and the learning rate is set at 0.001 while retaining the rest hyper-parameters fixed. Regularization methods are employed to reduce the likelihood of over-fitting. Moreover, l_2 with 0.0001 value of regularization parameter added to the loss function that shrinks model parameters to prevent over-fitting [40].

D. DEEP SHAPLEY ADDITIVE EXPLANATIONS

For constructing the accurate and explainable framework for NCDs, this study is motivated by Deep Shapley Additive Explanations (DeepSHAP) [41] for mixed model types, a framework for layer-wise propagation of SHAP [42] values that build upon deep learning important features (DeepLIFT) [43].

SHAP is a unified framework based on the Shapley value. SHAP is used to explain the predictions of an instance x by evaluating the contribution of each feature to the prediction. Shapley value is utilized in cooperative game theory to estimate the contribution of each player in a coalition game. The main idea is that determining all possible different permutations in a player's contribution in each game by using marginal contribution. After that, the average of these contributions of each player is calculated in Equation 3 as follows [42]:

$$\phi_i\left(f, x^i\right) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(M-|S|-1)!}{M!} * \left[f_{(S \cup \{i\})} - f_{(S)}\right] \quad (3)$$

where: M is the number of features and the sum extends over all subsets S of N not containing feature i; $f(S)$ is the prediction for feature values in the set S; here, i^{th} feature is excluded, then simulated random values of the ith feature from the dataset. However, the exact evaluation of the Shapley value is computationally expensive because of the estimation of contributions among many features contributions for prediction. SHAP assigns each feature an importance value for a particular prediction.

Moreover, local explanations inspired by local surrogate models and global explanations based on aggregations of Shapley values can be provided by SHAP for predictions.

IV. EXPERIMENTAL DESIGN FOR NON-COMMUNICABLE DISEASES
A. EXPERIMENTAL ENVIRONMENT

All experiments were performed on a PC with 3.50 GHz, Intel Core i5-6600K, and 32GB RAM using Microsoft Windows 10 operating system. Python programming language is utilized in the experimental analysis with the open libraries and packages such as Pandas [44], Numpy [45], Matplotlib [46], Scikit-learn [47], Statsmodels [48], XGBoost package [49], DeepSHAP [42] and so on.
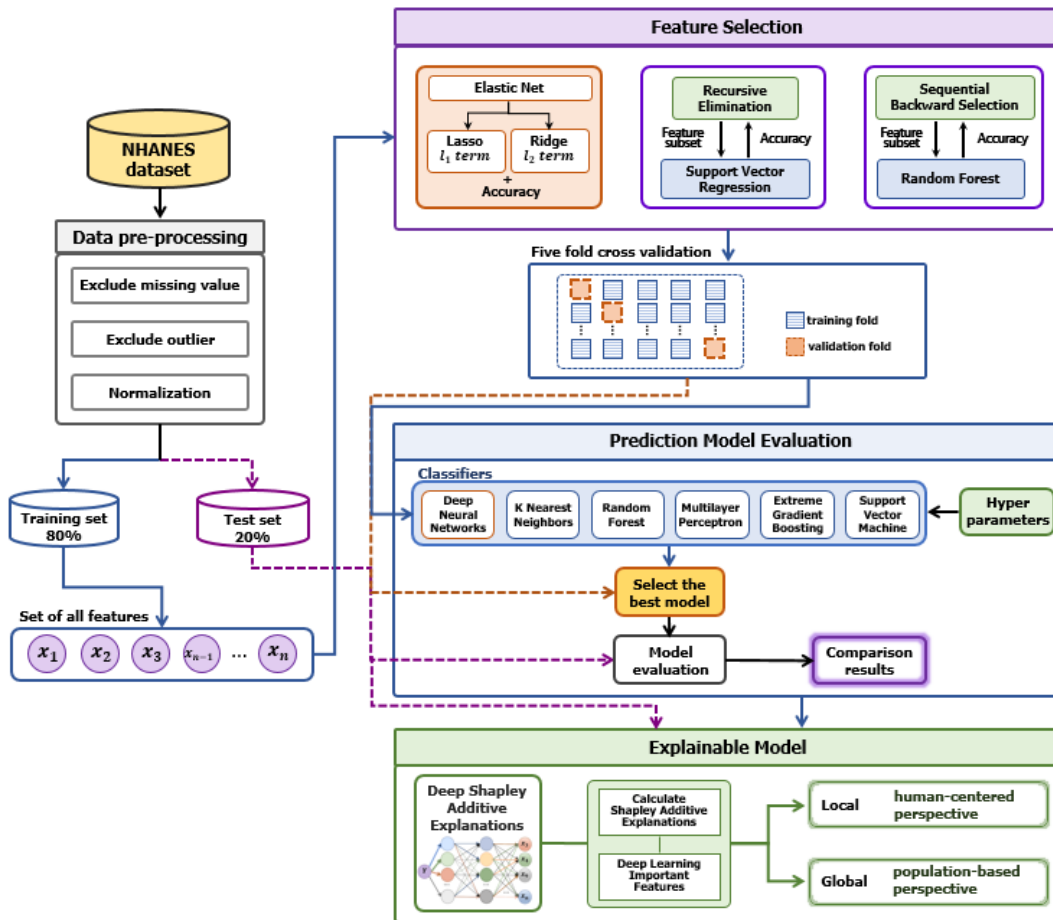
FIGURE 3. Experimental design for non-communicable diseases prediction and explanation.

B. NHANES DATASET

The National Health and Nutrition Examination Survey (NHANES) dataset is used to construct the predictive and explainable decision support models of NCDs. NHANES designed to assess the health and nutrition status of the general population in the United States. This nationwide survey is a major program of the National Center for Health Statistics that is part of the Centers for Disease Control and Prevention (https://www.cdc.gov/nchs/nhanes).

In general, this survey examines approximately 5000 people each year across the United States. The NHANES interview includes demographic, socioeconomic, dietary, and health-related questions. The examination component consists of medical, dental, and physiological measurements and laboratory tests that were administered by medical personnel. Notably, the NHANES dataset is used to determine the prevalence of major diseases and their risk factors in epidemiological studies and health sciences research.

C. BASELINE METHODS
1) WRAPPER FEATURE SELECTION TECHNIQUES

In the proposed framework, we aim to apply the elastic-net based embedded feature selection technique to find out the essential and applicable features, besides improving the prediction performance. The wrapper technique unifies a supervised machine-learning algorithm in the procedure of feature selection in order to find the optimal combination that maximizes model performance. In wrapper techniques, a search strategy iteratively adds and/or removes features from the dataset. The most commonly used search strategies are forward, backward and recursive selections.

Thus, support vector regression-based recursive feature elimination (SVR-RFE) [50], sequential backward feature selection with random forest (SBFS-RF) [51] wrapper feature selection techniques are used to compare to the proposed technique in part of empirical comparison analysis.

2) BASELINE CLASSIFICATION ALGORITHMS AND PARAMETERS

In this study, we compare the proposed model to the following various classification algorithms for NCDs prediction.

- **Support Vector Machine (SVM):** The SVM classifier is a supervised machine-learning model that can be used for solving classification and regression problems [52]. We tested the parameters on SVM as a kernel (linear, poly, RBF), gamma (scale) and tolerance (0.001).

- **K-nearest neighbors (KNN):** KNN is a machine-learning algorithm that can solve the classification task [53]. In the classification phase, instances are classified to the class most frequently occurring amongst the neighbors which measured by the distance function. In parameter setting, we have turned the values of k number 3, 6, 9; weight options (uniform, distance) and metric (Minkowski).
- **Random Forest (RF):** RF is a parallel structured ensemble tree-based method that utilizes bagging to aggregate multiple decision tree classifiers [54]. We have configured the number of estimators in the random forest as 250, 500, 750, 1000, 1250, and 1500, also the quality of split-measured criteria has selected by ''gini'' for the Gini impurity and ''entropy'' for the information gain, respectively.
- **Multilayer Perceptron (MLP):** MLP is a class of feed-forward neural network. MLP has an input layer, one hidden layer, and one output layer [55]. In this experiment, the single hidden layer of MLP consists of several nodes {3, 5, 10, 15, 20, 25}. Moreover, we use Adam optimizer for training and set the learning rate is 0.001 with the Rectified linear units (ReLU), Sigmoid (sigm) and Tanh (''tanh'') activation functions.
- **Extreme Gradient Boosting (XGBoost):** XGBoost is an efficient and scalable machine learning classifier, which was popularized by [49]. Gradient boosting decision tree is the original model of XGBoost, which combines multiple decision trees in a boosting way. For hyper-parameters, gamma (0, 0.001), column subsample ratio (0.9, 0.95, 1), subsample ratio (0.9, 0.95, 1), maximum delta step (0.4, 0.6, 0.8, 1), number of boost (60), learning rate (0.1, 0.01, 0.001), early stop round (100), minimum child weight (2, 3, 4, 5) and maximum tree depth (2, 4, 6, 8).

D. EVALUATION METRICS

For evaluating the NCDs prediction model, evaluation metrics are determined using four values, namely True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). A total number of (TP+TN) can interpret correct predictions and incorrect predictions built by the model is (FN+FP), respectively. The evaluation metrics of accuracy, specificity, recall, precision, f-score and, the area under curve have been utilized as model comparisons and model selection. The evaluation metrics [35,56] were defined as:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (4)$$

$$Specificity = \frac{TN}{TN + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$F - score = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (8)$$

V. EXPERIMENTAL RESULT AND ANALYSIS

The general architecture of experimental design is illustrated in FIGURE 3. The overall experimental results and analysis are provided in this section.

A. DATA PRE-PROCESSING AND FEATURE SELECTION ANALYSIS

The National Health and Nutrition Examination Survey (NHANES) dataset is utilized to construct the decision support system of NCDs. For NHANES, 9,971 individuals in 2015-2016 and 9,254 individuals in 2017-2018 were involved. In this experiment, we have collected a total of 19,225 individuals' data from 2015 to 2018. There are several methods to deal with missing data and produce complete data for analysis. To prevent model complication, we have excluded all the missing values and outliers. Interquartile range was used in outlier elimination. Moreover, individuals under 20 years of age were eliminated from our further analysis, because most NCDs were diagnosed by individuals aged equal to or more than 20 years old in the examination survey of NHANES. Thus, 13,014 individuals were excluded due to missing values, outliers and selection criteria of age, which were stored in the dataset. Totally 6,211 individuals' records have remained after the data cleaning phase.

This study was designed to consist of experimental and control groups. The experimental group was characterized by individuals with one of the NCDs diagnosed in their medical history, including diabetes, prediabetes, asthma, heart failure, coronary heart disease, heart attack, stroke, hypertension, kidney failure, and angina. On the other hand, the control group was defined by normal individuals who had not been diagnosed with NCDs. Consequently, NCDs diagnosed 2,107 (34%) individuals in the experimental group and 4,104 (66%) individuals in the control group were kept for further analysis. The statistics about this process are shown in FIGURE 4.

The dataset contains 51 features including information about demographic, socioeconomic, main vital signs, health-related questions, medical, dental, and physiological
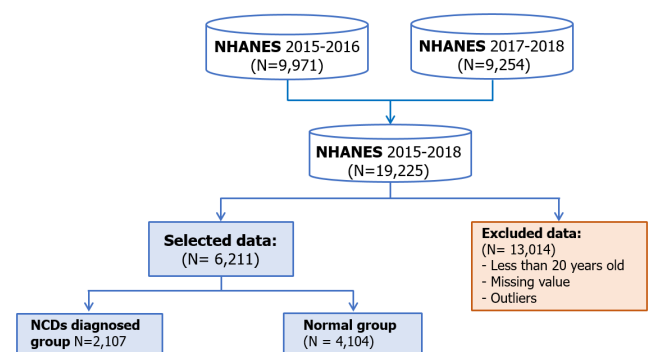

FIGURE 4. Sample selection procedure of national health and nutrition examination survey dataset.

Features	Description
RIAGENDR	Gender
BMXBMI	Body mass index (kg/m^2)
(BPXSY1; BPXDI1)	Blood pressure (mm Hg)
RIDAGEYR	Age in years at screening
DMDEDUC2	Education level
DMDMARTL	Marital status
DMDHHSIZ	Total number of people in the household
DIQ170	Health risk for diabetes (among family history)
DIQ050	Taking insulin now
BPXPULS	Pulse regular or irregular
DID250	Past year how many times seen doctor
DBD100	How often add salt to food at table
DR1TSUGR	Total sugars (gm)
DR1TALCO	Alcohol use (gm)
ALQ120U	How many days drink alcohol in per week, month, year
BPQ020	Ever told you had high blood pressure
HSD010	General health condition
DPQ020	Feeling down, depressed, or hopeless
DPQ040	Feeling tired or having little energy
DPQ050	Poor appetite or overeating
MCQ080	Doctor ever said you were overweight
HUQ051	How many times receive healthcare over past year
INDFMMPC	Monthly poverty level of family
MCQ365b	Doctor told you to exercise
OCQ260	Description of job/work situation
SLQ050	Ever told doctor had trouble sleeping
SMD460	Number of people who live here smoke tobacco
SMD470	Number of people who smoke inside this home
SMQ856	During the last 7 days worked at job not at home
SMQ858	During the last 7 days at job someone smoked indoors
SMQ878	During the last 7 days in other smoked indoor area
BPQ080	Doctor told you - high cholesterol level
SMQ020	Smoked at least 100 cigarettes in life
SLD010H	How much sleep do you get (hours)
SLQ060	Ever told by doctor have sleep disorder
ALQ101	Had at least 12-alcohol drinks/1 year
ALQ130	Average number of alcoholic drinks/day in past 12 months
ALQ151	Ever have 4/5 or more drinks every day
MCQ370a	Are you now controlling or losing weight
MCQ370b	Are you now increasing exercise
MCQ370c	Are you now reducing salt in diet
MCQ370d	Are you now reducing fat in diet
DPQ010	Have little interest in doing things
DPQ030	Trouble sleeping or sleeping too much
DPQ060	Feeling bad about yourself
DPQ080	Moving or speaking slowly or too fast
PAQ620	Moderate work activity
PAQ635	Moderate work activity: walk or bicycle
PAQ710	Hours watch TV or computer use past 30 days
WHQ030	Body weight

measurements, as well as laboratory tests administered by medical personnel as shown in Table 1.

In the proposed framework, the EN regularization method is utilized to eliminate redundant and irrelevant features from the feature space. First, we trained the EN model with initial features. Then, it derives feature importance scores from the predictive model that has been fit on the training dataset. Inspecting the importance score provides insight into that elastic net model and which features are the most important and least important to the predictive model when making a prediction. In the final stage, the EN feature selection model rejects the irrelevant and useless features using the derived feature importance scores. As a result, BPX-PULS, DMDHHSIZ, DIQ050, DID250, DPQ040, DPQ050, SMQ856, SMQ858, SMQ878, SLQ060, ALQ130, ALQ151, MCQ370d, DPQ080 and DPQ060 features were eliminated. At the end of the EN feature selection procedure, totally 35 features were remained.

B. COMPARISON RESULTS OF PREDICTION MODELS

For evaluating the prediction models, we split the data into 80% for the training set and 20% for the test set. Cross-validation with the 5-fold splitting procedure is applied to the training set. In 5-fold cross-validation, the training dataset is randomly partitioned into five folds; where 4 folds are used for training models, and 1 fold is used for validation while tuning model hyper-parameters [57]. The cross-validation process has ten times repetitions to decrease the unlikely chance of getting over-optimistic results. Consequently, after configuring the best setting of hyper-parameters, we evaluate the final model on the test set.

The statistical significance test of performance comparison to find difference in NCDs predictive models is summarized in Table 2. In our analysis, the various machine-learning classifiers were compared in terms of accuracy metric across validation and test sets. Statistical test was calculated with ρ-value of $1.27 \cdot 10^{-8}$ (ρ-value < 0.001) and rejected the null hypothesis at a 99% significance level. Accuracy is the proportion of the total number of correct predictions. The error rate of accuracy is when the classifier predicts positive instances as negative and negative instances as positive. The accuracy metric determines which model is best at identifying healthy and unhealthy individuals in input samples. For the superior predictive model, accuracy should be nearer to 1. The highest accuracies of validation and test sets are marked in bold.

A careful look at the accuracy results of the validation set, the highest performance of 0.9351 (95% Confidence Interval (CI), 0.9241–0.9481) was scored by our proposed DNN with EN, following by 0.9193 (0.9083–0.9313) was reached by XGBoost with EN model. Moreover, DNN and XGBoost classifiers with different feature selection techniques performed higher accuracy compared with other classifiers. DNN with SVR-RFE and SBFS-RF achieved accuracy scores of 0.8219 (95% CI, 0.8079–0.8379) and 0.9102 (95% CI, 0.8962–0.9242), respectively. In case of the XGBoost classifier, 0.8365 (95% CI, 0.8235–0.8505) with SVR-RFE and 0.9014 (95% CI, 0.8864–0.9174) with SBFS-RF higher accuracy scores were reached.

Among the test set, the DNN classifier exhibited the best capability of prediction accuracy of 0.9501 (95% (CI), 0.9391–0.9621) when it incorporated with EN embedded technique. Following by it, 0.9411 (95% CI, 0.9281–0.9551), 0.9256 (95% CI, 0.9106–0.9406) and 0.9074 (95% CI, 0.8944–0.9214) accuracy scores were achieved by the XGBoost with EN, DNN with SBFS-RF and MLP with SBFS-RF, respectively. It is clearly shown that SVR-RFE

TABLE 2. Statistical significance test of the overall accuracy evaluation for non-communicable diseases prediction models.

Feature selection techniques	Classifier	Validation set		Test set	
		Accuracy	CI 95 %	Accuracy	CI 95 %
SVR-RFE	SVM	0.7805	0.7655 − 0.7965	0.8012	0.7872 − 0.8172
	KNN	0.7763	0.7633 − 0.7913	0.7905	0.7775 − 0.8055
	RF	0.7903	0.7763 − 0.8043	0.8315	0.8185 − 0.8455
	MLP	0.8105	0.7985 − 0.8235	0.8288	0.8178 − 0.8418
	XGBoost	0.8365	0.8235 − 0.8505	0.8485	0.8325 − 0.8625
	DNN	0.8219	0.8079 − 0.8379	0.8498	0.8338 − 0.8668
SBFS-RF	SVM	0.8424	0.8294 − 0.8574	0.8827	0.8677 − 0.8987
	KNN	0.8452	0.8322 − 0.8592	0.8454	0.8324 − 0.8604
	RF	0.8801	0.8681 − 0.8931	0.9037	0.8897 − 0.9177
	MLP	0.8801	0.8661 − 0.8951	0.9074	0.8944 − 0.9214
	XGBoost	0.9014	0.8864 − 0.9174	0.9006	0.8886 − 0.9126
	DNN	0.9102	0.8962 − 0.9242	0.9256	0.9106 − 0.9406
EN	SVM	0.8541	0.8391 − 0.8691	0.8624	0.8484 − 0.8754
	KNN	0.8598	0.8468 − 0.8748	0.8504	0.8344 − 0.8644
	RF	0.9001	0.8861 − 0.9151	0.8793	0.8633 − 0.8963
	MLP	0.8898	0.8768 − 0.9038	0.8832	0.8682 − 0.8992
	XGBoost	0.9193	0.9083 − 0.9313	0.9411	0.9281 − 0.9551
	DNN	**0.9351**	0.9241 − 0.9481	**0.9501**	0.9391 − 0.9621

technique based models performed the lowest accuracy results of 0.8012 (95% CI, 0.7872–0.8172) by SVM; 0.7905 (95% CI, 0.7775–0.8055) by KNN; 0.8315 (95% CI, 0.8185–0.8455) by RF; 0.8288 (95% CI, 0.8178–0.8418) by MLP; 0.8485 (95% CI, 0.8325–0.8625) by XGBoost; and 0.8498 (95% CI, 0.8338–0.8668) by DNN.

FIGURE 5 illustrates accuracy result charts of NCDs predictive models in the test set. The evidence from these results suggests that the proposed EN based classifiers indicated better performances, followed by SBFS-RF based classifiers achieved the second-best accuracy performances than other predictive models. Contrary, KNN and SVM classifiers with SVR-RFE achieved lower accuracies of 0.7905 and 0.8012, respectively. The difference of average accuracies between EN based DNN classifier and that with the worst model of SVR-RFE based KNN was 15.96%.

Different feature selection techniques and classifiers enabled us to discover the best predictive model of NCDs and enhance the overall accuracies. As seen in Table 3, comparison results of NCDs predictive models in validation and test sets are provided. The highest performances of evaluation metrics are marked in bold.

For validation set, among feature subset of SVR-RFE, the highest specificity of 0.7919, recall of 0.8014, AUC of 0.7967 were outperformed by DNN classifier, and XGBoost classifier reached the highest precision of 0.8547 and f-score of 0.8265, significantly. Considering these evaluation performances of high-performing models based on SVR-RFE, DNN and XGBoost classifiers indicated the best classifiers than other baseline classifiers. However, it should be noted that overall evaluation metric scores were decreased significantly when compared with other feature selection techniques in the validation set. Meanwhile, SVM with the SVR-RFE model reached the worst predictive results

such as specificity of 0.6801, recall of 0.7305, the precision of 0.8032, f-score of 0.7651 and AUC of 0.7053, when compared with the other SVR-RFE based classifiers.

In the feature subset of SBFS-RF, the best model was distinguished by DNN in terms of the recall, precision, f-score and AUC, which reached 0.9025, 0.9236, 0.9129 and 0.8833, respectively. Considering the recall, XGBoost and DNN models performed higher scores of 0.8659 and 0.8641, respectively.

Furthermore, the worst results such as specificity of 0.7431, recall of 0.8413, precision of 0.8226, f-score of 0.8318 and AUC of 0.7922 were reached by KNN with the SBFS-RF model. Regarding the EN based feature subset; the best specificity of 0.8657 was achieved by the XGBoost model, also DNN model performed better prediction ability than other models, with a recall of 0.9231, the precision of 0.8795, f-score of 0.9008 and AUC of 0.8912, significantly. Concerning the specificity, recall and AUC evaluation metrics, EN with the KNN model presented the lowest scores of 0.7974, 0.8657 and 0.8316, respectively.

For comparison results of the predictive models in the test set, in terms of the SVR-RFE, the best results for recall of 0.8293, the precision of 0.8528, and f-score of 0.8305, were determined by MLP, SVM, and DNN classifiers, respectively. The XGBoost classifier produced the best specificity of 0.8345 and AUC of 0.8242. By contrast, whilst using SVR-RFE, lower values for recall of 0.7443, f-score of 0.7887 and AUC of 0.7431 were identified by KNN, moreover, specificity of 0.7188 and precision of 0.8257 were reached by SVM and RF, respectively.

In SBFS-RF, the best results for the specificity of 0.8997, recall of 0.9166, the precision of 0.9158, f-score of 0.9162 and AUC of 0.9082 were performed by DNN when ReLU activation, 4 hidden layers with 10 neurons, "Adam"
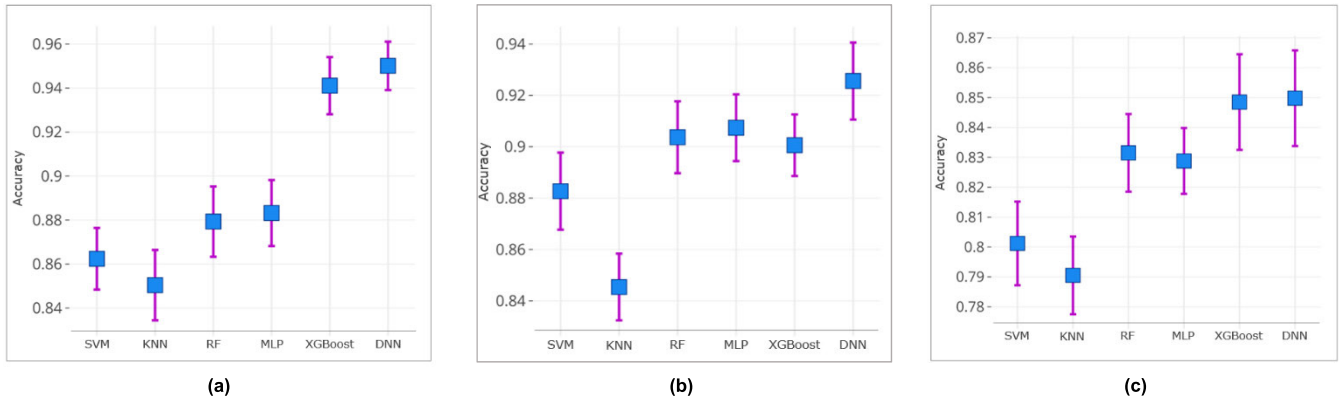
FIGURE 5. Statistical significance test of non-communicable diseases prediction models along with accuracy scores. (a) models equipped by support vector regression-based recursive feature elimination; (b) models equipped by sequential backward feature selection with random forest; (c) models equipped by elastic net.
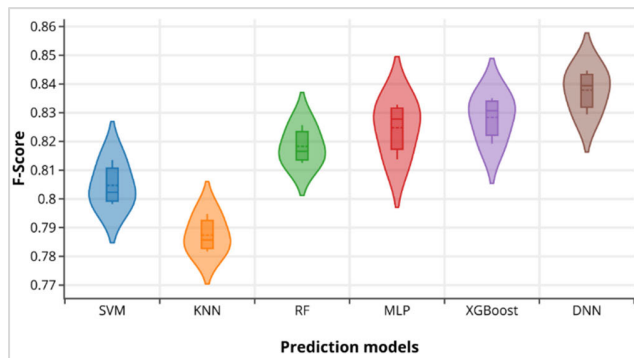

FIGURE 6. Box plot of the f-scores comparison of non-communicable diseases prediction models equipped by support vector regression-based recursive feature elimination.
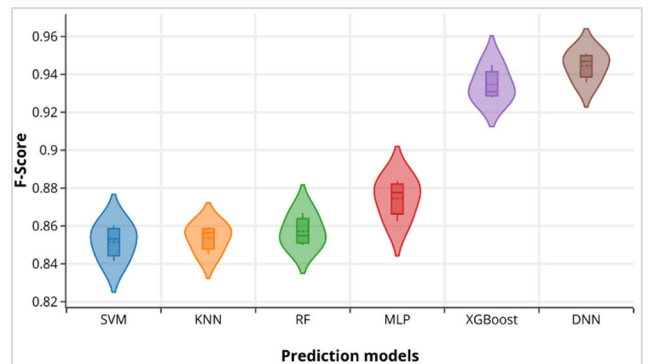

FIGURE 8. Box plot of the f-scores comparison of non-communicable diseases prediction models equipped by elastic net.
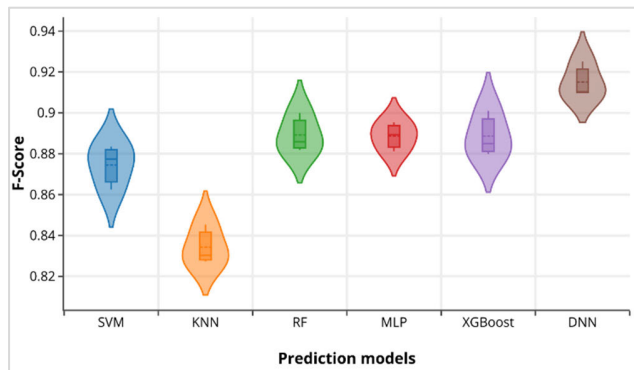

FIGURE 7. Box plot of the f-scores comparison of non-communicable diseases prediction models equipped by sequential backward feature selection with random forest.

optimizer and 0.001 learning rate were used. Contrary, the KNN reached the lowest specificity, recall, precision, f-score and AUC of 0.7551, 0.8278, 0.8329, 0.8303 and 0.7915, respectively.

Considering EN based predictive models, the DNN model achieved promising performances across evaluation metrics for the specificity of 0.8988, recall of 0.9573 precision of 0.9368, f-score of 0.9469 and AUC of 0.9281, when it used

"Adam" optimizer, ReLU activation function, 3 hidden layers with 10 neurons, and 0.001 learning rate. As can be seen, EN based SVM and KNN models performed lower results compared with EN based other models, but these models reached slightly better result across the other feature selection techniques based on SVM and KNN models. Among EN based models, the lowest NCDs predictive performances were reached as following models such as SVM with a recall of 0.8493, KNN with the specificity of 0.8062, AUC of 0.8302, and RF with the precision of 0.8215 and f-score of 0.8498. With regard to the most evaluation metrics, EN based predictive models reached the highest results in the test set.

It is well known that the f-score is a commonly used evaluation metric in the prediction task, which balances both the concerns of precision and recall into a single score.

Thus, concerning the f-score, FIGURES 6 to 8 illustrate the box plots of the prediction models in the test set. In these figures, the x-axis denotes the f-score and the y-axis presents the developed predictive models on NCDs. The f-score verified that the proposed EN with the DNN model was better at NCDs early diagnosis among classes, achieving the f-score of 0.9469, which outperformed the lowest scored SVR-RFE based KNN model by 15.82%.

TABLE 3. Statistical significance test of the overall accuracy comparison for non-communicable diseases prediction models.

Feature selection technique	Classifier	Validation set					Test set				
		Specificity	Recall	Precision	F-score	AUC	Specificity	Recall	Precision	F-score	AUC
SVR-RFE	SVM	0.6801	0.7305	0.8032	0.7651	0.7053	0.7188	0.7699	0.8528	0.8092	0.7444
	KNN	0.7015	0.7316	0.8165	0.7717	0.7166	0.7419	0.7443	0.8388	0.7887	0.7431
	RF	0.7545	0.7731	0.8316	0.8013	0.7638	0.7988	0.8057	0.8257	0.8156	0.8023
	MLP	0.7871	0.7817	0.8262	0.8033	0.7844	0.8015	0.8293	0.8303	0.8298	0.8154
	XGBoost	0.7884	0.8001	0.8547	0.8265	0.7943	0.8345	0.8158	0.8366	0.8261	0.8252
	DNN	0.7919	0.8014	0.8416	0.8210	0.7967	0.8295	0.8199	0.8413	0.8305	0.8247
SBFS-RF	SVM	0.7593	0.8738	0.8358	0.8544	0.8166	0.7889	0.8856	0.8695	0.8775	0.8373
	KNN	0.7431	0.8413	0.8226	0.8318	0.7922	0.7551	0.8278	0.8329	0.8303	0.7915
	RF	0.8347	0.8859	0.8603	0.8729	0.8603	0.8494	0.8983	0.8738	0.8859	0.8739
	MLP	0.8322	0.8739	0.8747	0.8743	0.8531	0.8672	0.8825	0.8882	0.8853	0.8749
	XGBoost	**0.8659**	0.8959	0.8924	0.8941	0.8809	0.8944	0.8838	0.8963	0.8900	0.8891
	DNN	0.8641	0.9025	**0.9236**	**0.9129**	0.8833	**0.8997**	0.9166	0.9158	0.9162	0.9082
EN	SVM	0.7991	0.8762	0.8421	0.8588	0.8377	0.8206	0.8493	0.8635	0.8563	0.8350
	KNN	0.7974	0.8657	0.8359	0.8505	0.8316	0.8062	0.8542	0.8486	0.8514	0.8302
	RF	0.8482	0.8962	0.8401	0.8672	0.8722	0.8546	0.8802	0.8215	0.8498	0.8674
	MLP	0.8445	0.8705	0.8296	0.8496	0.8575	0.8637	0.8879	0.8675	0.8776	0.8758
	XGBoost	0.8657	0.9018	0.8681	0.8846	0.8838	0.8839	0.9345	0.9273	0.9309	0.9092
	DNN	0.8592	**0.9231**	0.8795	0.9008	**0.8912**	0.8988	**0.9573**	**0.9368**	**0.9469**	**0.9281**

Essentially, the DNN model with each SBFS-RF and SVR-RFE technique performed the highest f-scores compare with other classifiers in each of the corresponding feature selection techniques. Among overall DNN based predictive models, f-scores were slightly lower than proposed EN by 11.64% in SBFS-RF and 3.07% in SVR-RFE, significantly.

FIGURES 9 to 11 illustrate the ROC curves for the NCDs predictive models on the four kinds of feature selection techniques. In ROC curves analysis, we can demonstrate the separation and discrimination ability of the predictive models.

The ROC curve is plotted with the measurements of true positive (sensitivity) along with the y-axis, and false positive (1-specificity) along with the x-axis. The sensitivity and specificity are crucial measures to classify an individual as having the disease or not having the disease. For the NCDs prediction model, the false-positive error occurs when unhealthy individuals are misclassified as healthy, whereas false-negative error appears when healthy individuals are misclassified as not having a disease. Hence, the false-positive error is worse than a false negative error [17].

For ROC curve analysis, the DNN classifier equipped with SVR-RFE and EN feature selection techniques reached notable significant performance than other corresponding models. However, in the selected feature subset of SBFS-RF, XGBoost outperformed DNN in terms of ROC value. Besides, enhanced performances were achieved by not only DNN but also XGBoost that reached computable results across each feature selection technique. Therefore, the SBFS-RF with the KNN classifier determined the worst predictive model among overall predictive models, but results were slightly improved when this classifier was combined with SVR-RFE by 4.84% and EN by 8.71%.

It is well recognized that the feature subset selected by EN can improve the predictive performance significantly in most classifiers. EN with the XGBoost showed the second-highest score of 0.9092. Essentially, ROC analysis verified that our proposed DNN classifier incorporated with the EN model was better at NCDs prediction across overall baseline models.

Depends on the collected NCDs data in this study, we did not consider the frequently occurred class imbalance problem in prediction analysis during the experiment. Nonetheless, decision-making responses suffer from the class imbalance problem that also has received much attention from researchers [26], [58]. To deal with this problem, sampling
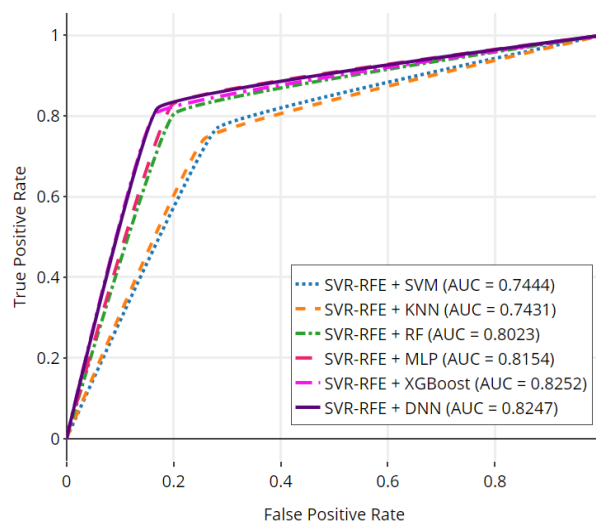

FIGURE 9. Receiver operating characteristic curves of non-communicable diseases prediction models equipped with support vector regression-based recursive feature elimination.
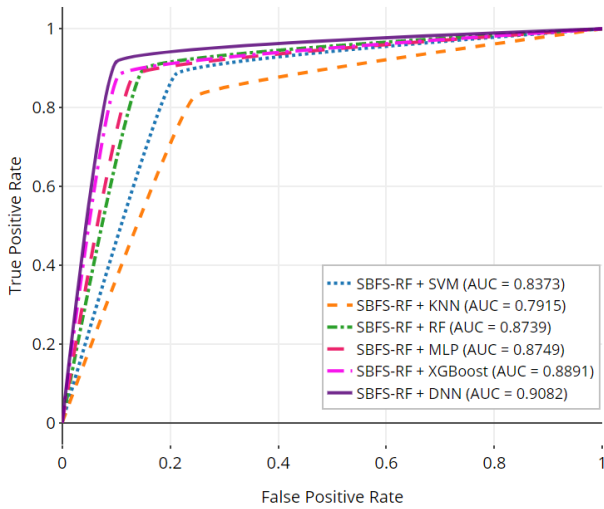
FIGURE 10. Receiver operating characteristic curves of non-communicable diseases prediction models equipped with sequential backward feature selection with random forest.
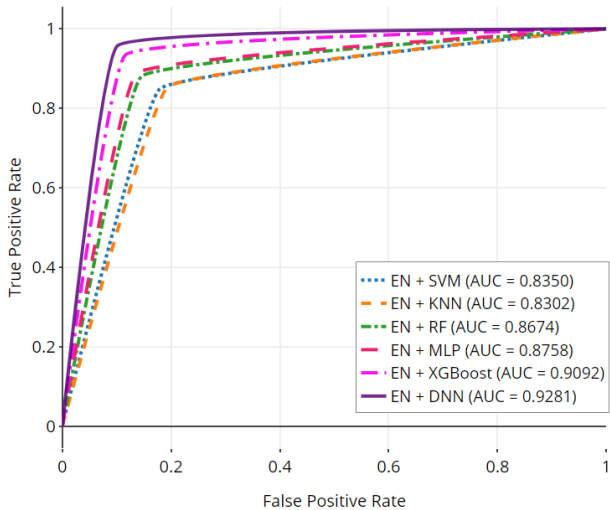

FIGURE 11. Receiver operating characteristic curves of non-communicable diseases prediction models equipped with elastic net.

techniques are investigated to rebalance an imbalanced dataset to alleviate the effect of the skewed class distribution. Broadly, sampling techniques can be classified into two groups, under-sampling and over-sampling. Under-sampling discards the samples from the majority class to make it equal to the minor class. A drawback of under-sampling is the loss of information. On the other hand, over-sampling creates the samples in the minor class to make it equal to the majority class [59]. Moreover, it has a drawback, which is associated with duplicated random records, which can be a cause of over-fitting. Existing studies have exhibited that the SMOTE and adaptive synthetic (ADASYN) over-sampling techniques are used rather than downsizing datasets [13]. To enhance our model consumption, the further solution to the class imbalance problem needs to be reflected.

As shown in Table 4, we compared the execution time of employed machine learning techniques in our analysis. For feature selection techniques, the embedded EN technique

TABLE 4. The comparison of execution time for machine learning techniques.

Machine learning techniques		Execution time (hour : min : sec)
Feature selection	SBFS-RF	0 : 43 : 48.672
	SVR-RFE	0 : 00 : 01.987
	EN	0 : 00 : 01.031
Classifiers	SVM	1 : 56 : 41.391
	KNN	0 : 00 : 53.766
	RF	0 : 17 : 17.438
	MLP	0 : 00 : 43.469
	XGBoost	0 : 42 : 07.219
	DNN	0 : 09 : 25.141

takes the lowest execution time when compare with wrapper feature selection techniques of SBFS-RF and SVR-RFE. As we know well, wrapper techniques are computationally intensive because wrapper techniques train a model for each subset. Therefore, a model and searching strategy are essential to finding the optimal combination. The SBFS-RF takes long execution time than other feature selection techniques. In the case of classification algorithms, we determined the execution time of training data produced by EN. Moreover, only optimized best parameters were used to execute the time of these classifiers.

It is clearly shown that SVM takes a long time when we used the optimal polynomial kernel function. Contrary, KNN and MLP ran faster but could not achieve acceptable prediction accuracies. Among tree-based classifiers, the XGBoost became slower and needs more memory to run than RF, however, we got computable accuracy with the proposed DNN classifier. Even though EN based DNN was performed the fastest one when compare with others, it served our purpose. In addition, the most time-consuming SBFS-RF and SVM techniques may not suit the ongoing trend of healthcare technology.

C. GLOBAL AND LOCAL EXPLANATION RESULTS
The model explanation is an essential task to get a better understanding of the reasoning behind the predictive models. In the second component of the proposed framework in this study, a standard forward pass is applied to DNN and activation at each layer is merged for the prediction task. Thereafter, in the third component, the score obtained at the output of the DNN is propagated backwards in DNN, using the propagation rule of the DeepSHAP approach in order to enhance the interpretability of the NCDs prediction model across the United States population.

The DeepSHAP approach enables two perspectives on the DNN model explanations: population-based and human-centered perspectives. In terms of the population-based perspective, the proposed model is able to explain conditional interactions between risk factors and classes on the training dataset. Moreover, the model can explain the conditional interaction between features and classes for a single individual for human-centered perspective.
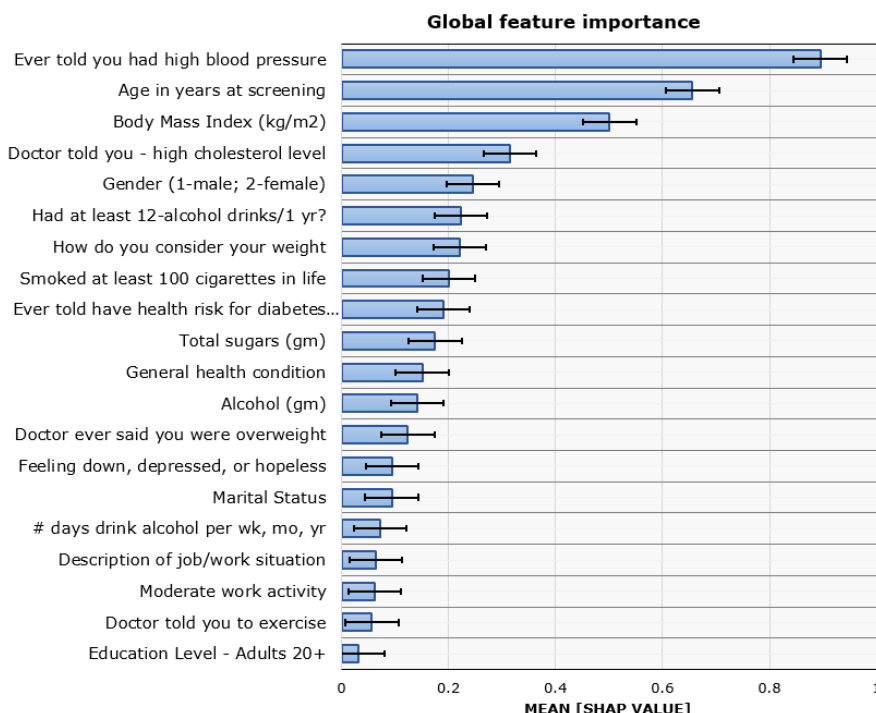
Global feature importance

FIGURE 12. The global explanation for non-communicable diseases among the United States from the population-based perspective.

To ensure model explanation, features were sorted in descending order of their importance scores in DNN model construction. As a selection result of the EN, 35 features were selected to build the prediction model. Among the general population in the United States, the most relevant 20 features of total features with their importance scores are presented in FIGURE 12. For the prediction model of NCDs across the entire population, "Ever told you had high blood pressure", "Age in years at screening", "Body Mass Index (kg/m2)", "Doctor told you - high cholesterol level", "Gender (1-male; 2-female)", "Had at least 12-alcohol drinks/1 year"; "How do you consider your weight"; "Smoked at least 100 cigarettes in life" and "Ever told have health risk for diabetes (among family history)" were maintained most significant features with importance

scores of 0.895, 0.656, 0.501, 0.315, 0.245, 0.223, 0.221, 0.201 and 0.191, respectively.

Therefore, high blood pressure occurred as a highly scored feature in the population of the United States. Similarly, a study [60] found an eminent association between high blood pressure and NCDs among middle-aged and older adults (aged 45 years and older) in China. In addition, approximately 9.4 million deaths are estimated caused by raised blood pressure, which means approximately 40% of adults have hypertension [61]. Overweight, cholesterol level and obesity determined significant features to develop NCDs in our study. Likewise, according to the study [62], the prevalence of increasing body mass index and high cholesterol level increased at more than one-half of the adult population in the United States. Essentially, obesity is a global burden, which has been strongly associated with most NCDs. This result is similar to that study [63], where high rates of overweight and obesity increased the burden of type 2 diabetes, coronary heart disease, and stroke in most countries in the Middle East.

According to the study [64], the risks for severe illness from NCDs increased along with older adults in India. Moreover, they identified that lower socioeconomic status is associated with smoking, alcohol use, low intake of fruit (vegetables), and being underweight, whereas, higher socioeconomic status is associated with greater exposure to obesity, dyslipidemia, diabetes in men, and hypertension in women. Also, the authors highlighted the prevalence of cigarette smoking among men and obesity among women was significantly higher in rural India. In another study [65], authors
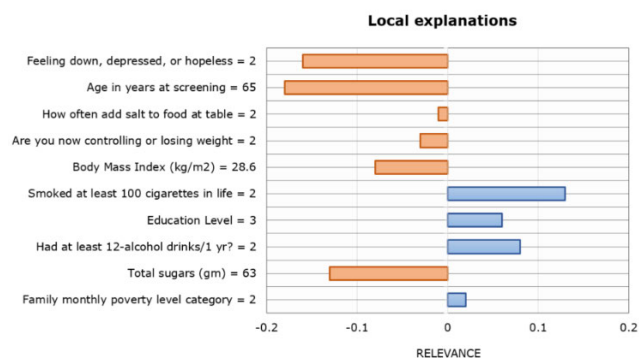
Local explanations

FIGURE 13. Local explanation of a single instance from the human-centered perspective.

determined the prevalence of risk factors of NCDs among rural communities in the Limpopo Province of South Africa. Their results defined that tobacco prevalence, alcohol consumption, and being overweight has a consistently higher association with NCDs among adults.

Therefore, most of the notable risk factors for NCDs were represented as modifiable. It is well known that modifiable risk factors are behaviours and exposures that were highly associated with the risk of developing various diseases. Due to prevent and correct these modifiable risk factors, it is required to take actions such as smoking cessation, alcohol reduction and exercise in public health. The highly scored risk factors enhance the rationale decisions in disease-related health concerns and should be collected in NCDs prediction data.

As a result of the DeepSHAP based global explanation approach, high-scored features are at the top of the list. These importance scores are appropriate for understanding the entire sample of the population, however, not at the individual level. Further, even some features seem to have less impact on the NCDs prediction model among the whole sample, but in some cases, they may have a critical impact on some part of the patients for diagnosing NCDs.

In FIGURE 13, the local explanation of the randomly chosen individual is shown. Local explanation result exhibits the most important 10 features of individual, negative risk factors were colored by orange and positive risk factors were marked by blue. Each value and its detailed descriptions are provided in Appendix. A negative coefficient exhibits an inverse relationship that the event becomes less likely as the predictor increase. The highest negative relationships for preventing NCDs were identified with the randomly chosen individual were "Age in years at screening = 65 (*years old*)", "Feeling down, depressed, or hopeless = 2 (*Several days*)", "Total sugars (gm) = 63", and "Body Mass Index (kg/m2) = 28.6".

On the contrary, "Body Mass Index (kg/m2) = 28.6", "Had at least 12-alcohol drinks/1 yr? = 2 (*no*)" and "Education Level = 3 (*College graduate or above*)" were significantly positively associated with preventing NCDs. As noted, in patient-centered healthcare applications, local explanations can be more accurate than global explanations.

Consequently, domain experts can able to explain the internal behaviour of accurate deep neural networks and know essential reasons to develop NCDs among the entire population and each individual. Furthermore, local explanations are crucial to make personalized health care recommendations.

VI. CONCLUSION AND FUTURE WORK

NCDs lead to the occurrence of premature death and become a significant threat to public health globally. It is essentially needed to reveal a more intelligent model that can assist in an early diagnosis of NCDs in the health care area. However, constructing an accurate and explainable model is challenging in the machine learning community.

Thus, we have proposed DeepSHAP based DNN framework equipped with an EN feature selection for early

prediction of NCDs. The procedure of this framework comprises three components: (I) representative features were selected based on EN; (II) DNN classifier was tuned with the hyper-parameters and used to train the model with the selected feature subset; (III) global and local level explanations were provided by DeepSHAP technique. As a result, the proposed framework emerged as being the best predictive model as it can reach notable superior performance than the other state-of-the-art baseline models. Furthermore, the proposed model provides explanations of NCDs along with information about the entire population as well as each individual in the NHANES dataset. Thus, we are able to explain the internal behaviour of accurate DNN and know exactly why it makes specific decisions.

Despite the potentials, this study has certain limitations. We have used only one cross-sectional dataset because NCDs data are usually not available. In other words, we have done our analysis at only a specific point in time. Nevertheless, the time series forecasting model is significantly crucial to predict the disease progression, estimate early warning scores of critical transitions over time. Thus, we have planned to extend our model to achieve accurate and explainable performances on longitudinal data while handling frequently occurred class-imbalanced problems. The further extended analysis expected to provide more advantages for experts to detect developments or changes in the characteristics of the target.

REFERENCES

[1] *Noncommunicable Diseases Country Profiles*, World Health Organization, Geneva, Switzerland, 2018.

[2] M. Forouzanfar, A. Afshin, L. Alexander, T. H. Anderson, R. Z. Bhutta, A. S. Biryukov, and A. J. Cohen, "Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: A systematic analysis for the Global Burden of Disease Study 2015," *Lancet*, vol. 388, no. 10053, pp. 1659–1724, 2016.

[3] *Noncommunicable Diseases Progress Monitor*, World Health Organization, Geneva, Switzerland, 2017.

[4] S. Kathirvel and J. T. Rapporteurs, "Sustainable development goals and noncommunicable diseases: Roadmap till 2030–A plenary session of world noncommunicable diseases congress 2017," *Int. J. Noncommunicable Diseases*, vol. 3, no. 1, p. 3, 2018.

[5] S. Chen, M. Kuhn, K. Prettner, and D. E. Bloom, "The macroeconomic burden of noncommunicable diseases in the united states: Estimates and projections," *PLoS ONE*, vol. 13, no. 11, Nov. 2018, Art. no. e0206702.

[6] *Action plan for the prevention and control of noncommunicable diseases in the WHO European Region*, World Health Organization, Copenhagen, Denmark, Sep. 2016, pp. 12–15.

[7] K. Davagdorj, S. H. Yu, S. Y. Kim, P. V. Huy, J. H. Park, and K. H. Ryu, "Prediction of 6 months smoking cessation program among women in Korea," *Int. J. Mach. Learn. Comput.*, vol. 9, no. 1, pp. 83–90, 2019.

[8] K. H. Park, H. D. K. Van, L. Munkhdalai, and K. H. Ryu, "A subtype classification of hematopoietic cancer using machine learning approach," in *Proc. ACIIDS*, 2021, pp. 113–121.

[9] F. Ali, S. El-Sappagh, S. M. R. Islam, D. Kwak, A. Ali, M. Imran, and K.-S. Kwak, "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion," *Inf. Fusion*, vol. 63, pp. 208–222, Nov. 2020.

[10] R. Panigrahi, S. Borah, A. K. Bhoi, M. F. Ijaz, M. Pramanik, R. H. Jhaveri, and C. L. Chowdhary, "Performance assessment of supervised classifiers for designing intrusion detection systems: A comprehensive review and recommendations for future research," *Mathematics*, vol. 9, no. 6, p. 690, Mar. 2021.

[11] S. El-Sappagh, H. Saleh, R. Sahal, T. Abuhmed, S. M. R. Islam, F. Ali, and E. Amer, "Alzheimer's disease progression detection model based on an early fusion of cost-effective multimodal data," *Future Gener. Comput. Syst.*, vol. 115, pp. 680–699, Feb. 2021.

[12] K. Davagdorj, V. H. Pham, N. Theera-Umpon, and K. H. Ryu, "XGBoost-based framework for smoking-induced noncommunicable disease prediction," *Int. J. Environ. Res. Public Health*, vol. 17, no. 18, p. 6513, Sep. 2020.

[13] K. Davagdorj, J. S. Lee, K. H. Park, P. Van Huy, and K. H. Ryu, "Synthetic oversampling based decision support framework to solve class imbalance problem in smoking cessation program," *Int. J. Appl. Sci. Eng.*, vol. 17, no. 3, pp. 223–235, 2020.

[14] D. V. Carvalho, M. E. Pereira, and J. S. Cardoso, "Machine learning interpretability: A survey on methods and metrics," *Electronics*, vol. 8, no. 8, p. 832, Jul. 2019.

[15] S. M. Lauritsen, M. Kristensen, M. V. Olsen, M. S. Larsen, K. M. Lauritsen, M. J. Jørgensen, J. Lange, and B. Thiesson, "Explainable artificial intelligence model to predict acute critical illness from electronic health records," *Nature Commun.*, vol. 11, no. 1, pp. 1–11, Dec. 2020.

[16] R. Elshawi, M. H. Al-Mallah, and S. Sakr, "On the interpretability of machine learning-based model for predicting hypertension," *BMC Med. Informat. Decis. Making*, vol. 19, no. 1, p. 146, Dec. 2019.

[17] K. Davagdorj, J. S. Lee, V. H. Pham, and K. H. Ryu, "A comparative analysis of machine learning methods for class imbalance in a smoking cessation intervention," *Appl. Sci.*, vol. 10, no. 9, p. 3307, May 2020.

[18] S. Bang, D. Yoo, S.-J. Kim, S. Jhang, S. Cho, and H. Kim, "Establishment and evaluation of prediction model for multiple disease classification based on gut microbial data," *Sci. Rep.*, vol. 9, no. 1, Dec. 2019, Art. no. 10189.

[19] R. Ge, R. Zhang, and P. Wang, "Prediction of chronic diseases with multi-label neural network," *IEEE Access*, vol. 8, pp. 138210–138216, 2020.

[20] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, pp. 81542–81554, 2019.

[21] J. Xiao, R. Ding, X. Xu, H. Guan, X. Feng, T. Sun, S. Zhu, and Z. Ye, "Comparison and development of machine learning tools in the prediction of chronic kidney disease progression," *J. Transl. Med.*, vol. 17, p. 119, Dec. 2019.

[22] M. F. Ijaz, M. Attique, and Y. Son, "Data-driven cervical cancer prediction model with outlier detection and over-sampling methods," *Sensors*, vol. 20, no. 10, p. 2809, May 2020.

[23] R. Panigrahi, S. Borah, A. K. Bhoi, M. F. Ijaz, M. Pramanik, Y. Kumar, and R. H. Jhaveri, "A consolidated decision tree-based intrusion detection system for binary and multiclass imbalanced datasets," *Mathematics*, vol. 9, no. 7, p. 751, Mar. 2021.

[24] F. Ali, S. El-Sappagh, S. M. R. Islam, A. Ali, M. Attique, M. Imran, and K.-S. Kwak, "An intelligent healthcare monitoring framework using wearable sensors and social networking data," *Future Gener. Comput. Syst.*, vol. 114, pp. 23–43, Jan. 2021.

[25] P. N. Srinivasu, J. G. Sivasai, M. F. Ijaz, A. K. Bhoi, W. Kim, and J. J. Kang, "Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM," *Sensors*, vol. 21, no. 8, p. 2852, Apr. 2021.

[26] K. Davagdorj, J. S. Lee, K. H. Park, and K. H. Ryu, "Cost-sensitive neural network for prediction of hypertension using class imbalance dataset," in *Advances in Intelligent Information Hiding and Multimedia Signal Processing Smart Innovation, Systems and Technologies*, vol. 212, J. S. Pan, J. Li, K. Ryu, Z. Meng, and A. Klasnja-Milicevic, Eds. Singapore: Springer, 2021, p. 44, doi: 10.1007/978-981-33-6757-9_6.

[27] M. F. Pinto, H. Oliveira, S. Batista, L. Cruz, M. Pinto, I. Correia, P. Martins, and C. Teixeira, "Prediction of disease progression and outcomes in multiple sclerosis with machine learning," *Sci. Rep.*, vol. 10, no. 1, Dec. 2020, Art. no. 21038.

[28] G. Fang, P. Xu, and W. Liu, "Automated ischemic stroke subtyping based on machine learning approach," *IEEE Access*, vol. 8, pp. 118426–118432, 2020.

[29] E. Batbaatar, K. H. Park, T. Amarbayasgalan, K. Davagdorj, L. Munkhdalai, V.-H. Pham, and K. H. Ryu, "Class-incremental learning with deep generative feature replay for DNA methylation-based cancer classification," *IEEE Access*, vol. 8, pp. 210800–210815, 2020.

[30] S. Sakr, R. Elshawi, A. Ahmed, W. T. Qureshi, C. Brawner, S. Keteyian, M. J. Blaha, and M. H. Al-Mallah, "Using machine learning on cardiorespiratory fitness data for predicting hypertension: The Henry Ford ExercIse Testing (FIT) Project," *PLoS ONE*, vol. 13, no. 4, Apr. 2018, Art. no. e0195344.

[31] E. Zihni, V. I. Madai, M. Livne, I. Galinovic, A. A. Khalil, J. B. Fiebach, and D. Frey, "Opening the black box of artificial intelligence for clinical decision support: A study predicting stroke outcome," *PLoS ONE*, vol. 15, no. 4, Apr. 2020, Art. no. e0231166.

[32] H.-C. Thorsen-Meyer, A. B. Nielsen, A. P. Nielsen, B. S. Kaas-Hansen, P. Toft, J. Schierbeck, T. Strøm, P. J. Chmura, M. Heimann, L. Dybdahl, L. Spangsege, P. Hulsen, K. Belling, S. Brunak, and A. Perner, "Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: A retrospective study of high-frequency data in electronic patient records," *Lancet Digit. Health*, vol. 2, no. 4, pp. e179–e191, Apr. 2020.

[33] S. El-Sappagh, J. M. Alonso, S. M. R. Islam, A. M. Sultan, and K. S. Kwak, "A multilayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer's disease," *Sci. Rep.*, vol. 11, no. 1, Dec. 2021, Art. no. 2660.

[34] L. Munkhdalai, K. H. Ryu, O.-E. Namsrai, and N. Theera-Umpon, "A partially interpretable adaptive softmax regression for credit scoring," *Appl. Sci.*, vol. 11, no. 7, p. 3227, Apr. 2021.

[35] P. N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. London, U.K.: Pearson, 2016.

[36] A. B. Owen, "A robust hybrid of lasso and ridge regression," *Contemp. Math.*, vol. 443, no. 7, pp. 59–72, 2007.

[37] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. Roy. Statist. Soc., B Stat. Methodol.*, vol. 67, no. 2, pp. 301–320, 2005.

[38] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11–26, Apr. 2017.

[39] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. Mohamed, and H. Arshad, "State-of-the-art in artificial neural network applications: A survey," *Heliyon*, vol. 4, no. 11, Nov. 2018, Art. no. e00938.

[40] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[41] H. Chen, S. Lundberg, and S.-I. Lee, "Explaining models by propagating Shapley values of local components," 2019, *arXiv:1911.11888*. [Online]. Available: http://arxiv.org/abs/1911.11888

[42] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 4765–4774.

[43] A. Shrikumar, P. Greenside, and A. Kundaje, "Learning important features through propagating activation differences," 2017, *arXiv:1704.02685*. [Online]. Available: http://arxiv.org/abs/1704.02685

[44] W. McKinney, "Pandas: A foundational Python library for data analysis and statistics," *Python High Perform. Sci. Comput.*, vol. 14, no. 9, pp. 1–9, 2011.

[45] T. E. Oliphant, *A Guide to NumPy*, vol. 1. New York, NY, USA: Trelgol, 2006, p. 85.

[46] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, May/Jun. 2007.

[47] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, and J. Vanderplas, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.

[48] S. Seabold and J. Perktold, "Statsmodels: Econometric and statistical modeling with Python," in *Proc. 9th Python Sci. Conf.*, 2010, p. 61.

[49] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jun. 2016, pp. 785–794.

[50] X. Lin, F. Yang, L. Zhou, P. Yin, H. Kong, W. Xing, X. Lu, L. Jia, Q. Wang, and G. Xu, "A support vector machine-recursive feature elimination feature selection method based on artificial contrast variables and mutual information," *J. Chromatography B*, vol. 910, pp. 149–155, Dec. 2012.

[51] F. Ruan, F. Qi, J. Cyan, H. Tang, T. Zhang, and H. Li, "Quantitative detection of harmful elements in alloy steel by LIBS technique and sequential backward selection-random forest (SBS-RF)," *J. Anal. At. Spectrometry*, vol. 32, no. 11, pp. 2194–2199, 2017.

[52] D. Meyer, F. Leisch, and K. Hornik, "The support vector machine under test," *Neurocomputing*, vol. 55, nos. 1–2, pp. 169–186, Sep. 2003.

[53] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Ann. Transl. Med.*, vol. 4, no. 11, p. 218, Jun. 2016.

[54] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.

[55] P. J. G. Lisboa, "A review of evidence of health benefit from artificial neural networks in medical intervention," *Neural Netw.*, vol. 15, no. 1, pp. 11–39, 2002.

[56] K. J. van Stralen, V. S. Stel, J. B. Reitsma, F. W. Dekker, C. Zoccali, and K. J. Jager, "Diagnostic methods I: Sensitivity, specificity, and other measures of accuracy," *Kidney Int.*, vol. 75, no. 12, pp. 1257–1263, Jun. 2009.

[57] T. Fushiki, "Estimation of prediction error by using K-fold cross-validation," *Statist. Comput.*, vol. 21, no. 2, pp. 137–146, Apr. 2011.

[58] L. Munkhdalai, T. Munkhdalai, and K. H. Ryu, "GEV-NN: A deep neural network architecture for class imbalance problem in binary classification," *Knowl.-Based Syst.*, vol. 194, Apr. 2020, Art. no. 105534.

[59] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, "Learning from class-imbalanced data: Review of methods and applications," *Expert Syst. Appl.*, vol. 73, pp. 220–239, May 2017.

[60] Y. You, J. Wang, W. Teng, G. Ma, and P. Liu, "Blood pressure and noncommunicable diseases in middle-aged and older adults in China," *PLoS ONE*, vol. 13, no. 11, Nov. 2018, Art. no. e0206635.

[61] K. Davagdorj, M. Li, and K. H. Ryu, "Local interpretable model-agnostic explanations of predictive models for hypertension," in *Advances in Intelligent Information Hiding and Multimedia Signal Processing. Smart Innovation, Systems and Technologies*, vol. 212, J. Pan, J. Li, K. Ryu, Z. Meng, and A. Klasnja-Milicevic, Eds. Singapore: Springer, 2021, doi: 10.1007/978-981-33-6757-9_53.

[62] C. D. Brown, M. Higgins, K. A. Donato, F. C. Rohde, R. Garrison, E. Obarzanek, N. D. Ernst, and M. Horan, "Body mass index and the prevalence of hypertension and dyslipidemia," *Obesity Res.*, vol. 8, no. 9, pp. 605–619, Dec. 2000.

[63] F. Kilpi, L. Webber, A. Musaigner, A. Aitsi-Selmi, T. Marsh, K. Rtveladze, and A. Brown, "Alarming predictions for obesity and non-communicable diseases in the Middle East," *Public Health Nutrition*, vol. 2014, 17, pp. 1078–1086.

[64] S. Kinra, L. Bowen, T. Lyngdoh, D. Prabhakaran, K. Reddy, L. Ramakrishnan, and S. Ebrahim, "Sociodemographic patterning of non-communicable disease risk factors in rural India: A cross sectional study," *Bmj*, vol. 27, p. 341, Sep. 2010.

[65] E. Maimela, M. Alberts, S. E. P. Modjadji, S. S. R. Choma, S. A. Dikotope, T. S. Ntuli, and J.-P. Van Geertruyden, "The prevalence and determinants of chronic non-communicable disease risk factors amongst adults in the Dikgale Health Demographic and Surveillance System (HDSS) Site, Limpopo Province of South Africa," *PLoS ONE*, vol. 11, no. 2, Feb. 2016, Art. no. e0147926.

VAN-HUY PHAM received the M.S. degree in computer science from the University of Sciences, Ho Chi Minh City, Vietnam, in 2007, and the Ph.D. degree in computer science from the University of Ulsan, South Korea, in 2015. Since 2015, he has been a Lecturer and a Researcher with the Faculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City. His main research interests include artificial intelligence, image processing, computer vision, and deep learning applications.

NIPON THEERA-UMPON (Senior Member, IEEE) received the B.Eng. degree (Hons.) from Chiang Mai University, the M.S. degree from the University of Southern California, and the Ph.D. degree from the University of Missouri-Columbia, all in electrical engineering. He was the Associate Dean of engineering and the Chairman of graduate study in electrical engineering and graduate study in biomedical engineering. He has been with the Department of Electrical Engineering, Chiang Mai University, since 1993, where he is currently serving as the Director for the Biomedical Engineering Institute. He has published more than 200 full research articles in international refereed publications. His research interests include pattern recognition, machine learning, artificial intelligence, digital image processing, neural networks, fuzzy sets and systems, big data analysis, data mining, medical signal, and image processing. He is a member of IEEE-IES Technical Committee on Human Factors. He is also a member of the Thai Robotics Society, the Biomedical Engineering Society of Thailand, and the Council of Engineers, Thailand. He has been bestowed several royal decorations and won several awards. He has served as the Vice President for the Thai Engineering in Medicine and Biology Society and Korea Convergence Society. He has served as an editor, a reviewer, the general chair, the technical chair, and a committee member for several journals and conferences.

KHISHIGSUREN DAVAGDORJ received the M.Sc. degree from the School of Industrial Technology and Design, Mongolian University of Science and Technology, Mongolia, in 2013, and the Ph.D. degree from the School of Electrical and Computer Engineering, Chungbuk National University, South Korea, in 2021. She is currently working as a Postdoctoral Researcher with Chungbuk National University. Her main research interests include explainable artificial intelligence, deep neural networks, machine learning, data mining, and recommendation systems in the healthcare domain.

JANG-WHAN BAE received the Bachelor of Medical Science and M.D. degrees from Chungbuk National University, Cheongju, South Korea, in 1994, and the Post Graduate Diploma and Ph.D. degrees in medicine from Chungbuk National University, in 2004. He is currently a Professor in medicine (cardiology subspecialty) with the Department of Internal Medicine, School of Medicine, Chungbuk National University. His research interests include ischemic heart disease, lipid metabolism, and stem cell therapy.

KEUN HO RYU (Life Member, IEEE) received the Ph.D. degree in computer science and engineering from Yonsei University, South Korea, in 1988. He has served at Reserve Officers' Training Corps (ROTC) of the Korean Army. He was with The University of Arizona, Tucson, AZ, USA, as a Postdoctoral Researcher and a Research Scientist, and also with the Electronics and Telecommunications Research Institute, South Korea, as a Senior Researcher. He is currently a Professor with the Faculty of Information Technology, Ton Duc Thang University, Vietnam, as well as an Emeritus and the Endowed Chair Researcher with Chungbuk National University, South Korea, and also an Adjunct Professor with Chiang Mai University, Thailand. He is also an Honorary Doctorate of the National University of Mongolia. He has been not only the Director of the Database and Bioinformatics Laboratory, South Korea, since 1986, but also the Co-Director of the Data Science Laboratory, Research Group, Ton Duc Thang University, since March 2019. He is also the former Vice-President of the Personalized Tumor Engineering Research Center. He has published more than 1000 refereed technical articles in various journals and international conferences, in addition to authoring a number of books. His research interests include databases, spatiotemporal databases, big data analysis, data mining, deep learning, biomedical informatics, and bioinformatics. He has been a member of ACM, since 1983. He has served on numerous program committees, including roles as the Demonstration Co-Chair of the VLDB, the Panel and Tutorial Co-Chair of the APWeb, and the FITAT General Co-Chair.

● ● ●