

Article

Forecasting Heart Disease Risk with a Stacking-Based Ensemble Machine Learning Method

Yuanyuan Wu ¹, Zhuomin Xia ¹, Zikai Feng ^{1,*}, Mengxing Huang ¹, Huizhou Liu ¹ and Yu Zhang ²

¹ College of Information and Communication Engineering, Hainan University, Haikou 570228, China; wyuanyuan82@163.com (Y.W.); xzmm0308@163.com (Z.X.); huangmx09@163.com (M.H.); huizhou_liu@hainanu.edu.cn (H.L.)

² College of Computer Science and Technology, Hainan University, Haikou 570228, China; yuzhang2015@hainanu.edu.cn

* Correspondence: zkfeng@hainanu.edu.cn

Abstract: As one of the main causes of sickness and mortality, heart disease, also known as cardiovascular disease, must be detected early in order to be prevented and treated. The rapid development of computer technology presents an opportunity for the cross-combination of medicine and informatics. A novel stacking model called SDKABL is presented in this work. It uses three classifiers, namely K-Nearest Neighbor (KNN), Decision Tree (DT), and Support Vector Machine (SVM) at the base layer and the Bidirectional Long Short-Term Memory based on Attention Mechanisms (ABiLSTM) model at the meta layer for the ultimate prediction. For lowering the temporal complexity and enhancing the model's accuracy, the dimensionality reduction approach is seen to be crucial. Principal Component Analysis (PCA) was utilized in SDKABL to minimize dimensionality and facilitate feature fusion. Using several performance measures, including precision, F1-score, accuracy, recall, and Receiver Operating Characteristic (ROC) score, the performance of SDKABL was compared to that of other independent classifiers. The experimental findings demonstrate that our proposed model combining individual classifiers with the stacking method helps improve the prediction model's accuracy.

Keywords: heart disease; risk forecasting; machine learning



Citation: Wu, Y.; Xia, Z.; Feng, Z.; Huang, M.; Liu, H.; Zhang, Y. Forecasting Heart Disease Risk with a Stacking-Based Ensemble Machine Learning Method. *Electronics* **2024**, *13*, 3996. <https://doi.org/10.3390/electronics13203996>

Academic Editor: Alejandro L. Borja

Received: 10 September 2024

Revised: 28 September 2024

Accepted: 8 October 2024

Published: 11 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Non-communicable illnesses currently account for 41 million annual fatalities, or 74% of all deaths worldwide [1]. With 17.9 million fatalities annually, cardiovascular illnesses represent the greatest percentage of noncommunicable disease (NCD)-related mortality [1]. One of the main causes of sickness and mortality is heart disease, often known as cardiovascular disease [2].

Data storage, processing, and transmission capabilities have significantly improved due to the rapid development of computer techniques such as the big data [3] and the Internet of Things (IoT) [4]. This has facilitated the development of health big data and created opportunities for the cross-integration of informatics and medicine [5,6].

When people seek medical treatment, a large amount of data will be generated, which is converted into digital format and stored in computers or databases and becomes Electronic Health Records (EHRs) [7]. The EHRs not only record important patient data but also hide hidden knowledge, including the relationship between data and conditions. In-depth analysis of these medical data based on data mining technology can extract useful information from huge datasets and make decision support for individual disease diagnosis methods and treatment programs [8,9].

However, there are a few significant drawbacks to conventional machine learning (ML) approaches, such as [10,11]:

- Model training requires a lot of resources and high-quality datasets.
- Overfitting, the model lacks generalization and robustness.

In ensemble learning, a more efficient and complex model is created by combining various ML algorithms to build a unique, more powerful model, thus harnessing their combined strengths [12]. Compared with a single model [13], this cooperative method can upgrade the performance and generality of forecasts. Ensemble learning has been used to detect, diagnose, and predict different diseases [14,15].

There are several integrated learning strategies, such as stacking, augmentation, voting, and bagging [16,17]. The model used in this paper adopts the idea of superposition. Stacking requires multiple learning stages to improve overall performance by combining multiple basic models [18].

The main principle is:

First, the training dataset is divided into subsets that are used to train different basic models.

These basic models are then used to make predictions, and these predictions are used as new features.

Finally, a meta model (usually a simple linear model) is used to learn how these new features relate to the real tags.

This paper aims to generate a disease prediction model using the stacking method to enhance the prediction accuracy. This article makes the following specific contributions:

- Three classifiers at the base layer and one at the meta layer are used to create a stacked model that is intended to predict heart disease.
- Exploratory data analysis and preprocessing are used to enhance the standard of the dataset.
- The contribution of each feature to the disease is determined, and dimensionality reduction techniques are utilized to obtain valuable information from the dataset.
- Extensive experiments are conducted, and the model is optimized in combination with hyperparameter tuning.

The rest of the article is organized as follows: Section 2 provides a review of this work. Section 3 discusses the methodology used. Section 4 provides the experimental results and comparative analysis. Section 5 summarizes the conclusions of the research and outlines future directions.

2. Related Work

In recent years, medical systems have used IoT technology to collect data to diagnose diseases and prognoses. An automatic prediction model of heart disease with three main phases was proposed in [19]. The input data are modified by Z-score standardization as a preprocessing step. An improved quantum CNN (IQCNN) makes the prediction based on extracted characteristics. Compared with other traditional methods, the proposed IQCNN model achieves 0.91 with a learning rate of 70% and thus has better performance in predicting heart disease.

A heart disease prediction system based on hierarchical Bidirectional Long Short-Term Memory with Guard dog Hunting Optimization (GdHO-BiLSTM) was proposed in [20]. The GdHO-BiLSTM model utilizes optimization techniques to enhance its capability in data recognition. By using the fusion layered BiLSTM model, the time dependence in the sequence data is obtained. The findings suggest that the proposed model has greater predictive capabilities and provides a reference for clinicians and healthcare providers.

Three enhancement algorithms were used to predict heart disease in [21]. The gradient boosting method has the best accuracy of all the algorithms. With gradient augmentation, the suggested model outperforms the others regarding accuracy, recall, and F1-score.

Similarly, researchers have utilized ensemble learning, particularly the stacking method, because it has been effective in forecasting other illnesses.

In [22], a novel stacking model called PaRSEL was put forward. At the meta layer, LogitBoost is used to carry out the final prediction after four classifiers are used at the base layer. In PaRSEL, several dimensionality reduction techniques are also used, including Linear Discriminant Analysis (LDA) and Recursive Feature Elimination (RFE). Furthermore, a variety of methods are employed to address the imbalance of datasets. Our proposed PaRSEL model achieves the highest accuracy compared with other independent classifiers. This suggests that PaRSEL is better at predicting heart disease than other independent classifiers.

A Statistical Feature Selection (SFS) stacking framework was proposed in [23]. It selects the best features from a data collection using four feature techniques. In addition, the results are predicted using two stacking methods. Both models achieved significant performance indicators on all three datasets.

Two datasets were used to conduct experiments in the literature [24]. The best performance was achieved when a stackable integrated classifier was used, where support vector machines and gradient enhancement were used to extract features, and logistic regression was used to predict Parkinson's disease. The stacked integrated classifier achieved 94.87% accuracy and a 90.00% Area Under Curve (AUC) in the first dataset, and 96.18% accuracy and a 96.27% AUC in the second dataset. The final results indicate the proposed stacking model's validity, which helps improve the overall diagnostic outcome.

A new heart disease prediction model based on a two-stage stacked method was proposed in [25]. For higher-performance models such as random forests, extreme gradient enhancement, and decision trees, additional improvements were made using stack integration techniques. The stacked model leveraged the power of all three models to achieve 96% accuracy, 98% recall, and 96% ROC scores. Under the same experimental settings, the accuracy of the model always reached 96.88%.

An integrated model is proposed in [26], which consists of three stages: feature extraction, stack feature set creation, and final prediction. Convolutional and deep neural networks were employed to extract the dataset's features. According to the simulation findings, the suggested model is more precise than both single-mode and homogenous multi-mode frameworks.

The study by [27] proposed an enhanced method for the early prediction of chronic diseases. This method used RFE in combination with SVM to reduce complexity. The improved dataset is then fed into a robust XGBoost classifier. Model performance is improved by hyperparameter tuning using Bayesian optimization.

Ref. [28] used the extremely imbalanced diabetes dataset to propose two integrated approaches: hybrid and blending. For this imbalance problem, the Proximity-Weighted Synthetic Oversampling (ProWSyn) technique is implemented. For the precise and early identification of diabetes, we suggest the Hi-Le model, a mix of the Highway and LeNet models. The Hi-Le model outperforms its individual model in precision, recall, accuracy, and F1-score. Additionally, we suggest a hybrid model dubbed HiTCLe that predicts diabetes by combining Temporal Convolutional Networks (TCN), LeNet, and Highway. Of all the methods compared, HiTCLe performed best.

In [29], the risk of heart disease was predicted using an enhanced ML method. This algorithm uses an average-based block algorithm to divide the sample randomly. Classification and Recovery Tree (CART) is used to model each part. Finally, the time classifier integration method of weights is used to integrate multiple CART patterns. Finally, a precision-based weighted burn-in classifier integration is used to create uniform integrations from different CART models. The usefulness of the suggested strategy for estimating the risk of heart disease was shown by experiments conducted on the two datasets.

3. Proposed Methodology

3.1. Dimensionality Reduction

Dimensionality reduction is the process of feature dimensionality reduction on a given dataset, which can eliminate redundancy and noise and improve model performance and visualization.

This paper mainly uses the PCA algorithm, the main principle of which is to find orthogonal bases sequentially from the original space. The rule is the direction of the maximum difference in the primitive data, and the number of orthogonal bases to be selected is determined by the cumulative contribution rate, and then the original data are projected onto this set of orthogonal bases; that is, a new set of features is obtained, while the maximum information of the original data is retained [30].

In this paper, PCA is used for data reduction and feature extraction. By setting the `n_components` parameter, you can select the features that are most associated with disease. The optimum value of `n_components` is 4, according to the experimental results. By setting `n_components = 4`, the four features most related to heart disease are selected, which can improve the model's prediction accuracy.

3.2. Machine Learning Techniques

3.2.1. Support Vector Machine (SVM)

The SVM is a powerful supervised learning system that can handle problems related to regression and classification. Its core goal is to construct a decision model by training the label information of the sample, which effectively classifies the unknown data points into two preset categories. Unlike general linear classifiers, SVM is unique in that it is non-probabilistic and focuses on finding an optimal hyperplane that maximizes the degree of separation between different classes of data points to form clear boundaries. SVM can also be used to classify nonlinear data. The key to this transformation lies in the kernel function, which can cleverly map the original low-dimensional data to a high-dimensional space, thus giving the ability to accurately classify nonlinear data [31].

3.2.2. Decision Tree (DT)

A decision tree is a tree-like data structure used to represent decision rules and class results. It is an inference-based method used to transform the seemingly chaotic known information into a tree structure that can predict the unknown information. Every path that connects the root node to the leaf node is used by the algorithm as a decision rule.

3.2.3. K-Nearest Neighbor

K-Nearest Neighbor (KNN) is a typical classification algorithm in the ML field. The core concept of the algorithm can be summarized as "birds of a feather flock together, people flock together"; that is, using close examples to predict category belonging. It involves comparing a new sample with a set of samples of an existing known class, first measuring the distance between the new sample and all known samples, and then selecting the k samples that are closest. Based on the rule of majority decision, the new sample will be assigned to the category with the most occurrences among the k samples, which is mathematically expressed as follows:

$$o(x_1, x_2) = \sqrt{\sum_{i=1}^n (x_{1,i} - x_{2,i})^2}$$

where n represents the number of features of the sample points, and x_1 and x_2 represent two different sample points.

3.2.4. ABiLSTM

In 1990, Elman first proposed the concept of Recurrent Neural Networks (RNNs). RNNs show remarkable ability in processing language sequence data. Since RNNs consider the state of the previous time point when processing data, when the network input is a long sequence of data, the distant gradient will tend to 0, the near-distance gradient will not disappear, and it is easy to produce the phenomenon of gradient disappearance during the training process.

The LSTM and RNN are structurally consistent in that both combine the output of the previous unit with the current input. The key difference, however, is that LSTM can use

the sigmoid function to selectively retain information through the forget gate. BiLSTM is a stack of forward and backward LSTMs. Compared with one-way LSTM, the transitivity of the flow sequence in both directions can be considered simultaneously.

Attention mechanisms are inspired by the human visual system, which allows neural networks to focus on specific inputs or features. Bahdanau and his team [32] employed an attention mechanism to weed out a large amount of irrelevant information by sifting information from top to bottom and adding weights to key features to assess their importance.

Therefore, by integrating the attention mechanism into the BiLSTM network (ABiLSTM), the output of the bidirectional LSTM layer is given different levels of attention, which can better focus on the identification of the core features of the patient's disease and can effectively focus on the key feature information of the patient's disease, thus improving the accuracy of the model classification.

3.3. Ensemble Model Named SDKABL

Numerous ML methods have been developed to forecast heart disease in individuals. The suggested stacking model makes an important contribution to heart disease prediction. Figure 1 displays the flow chart for the suggested model, called SDKABL.

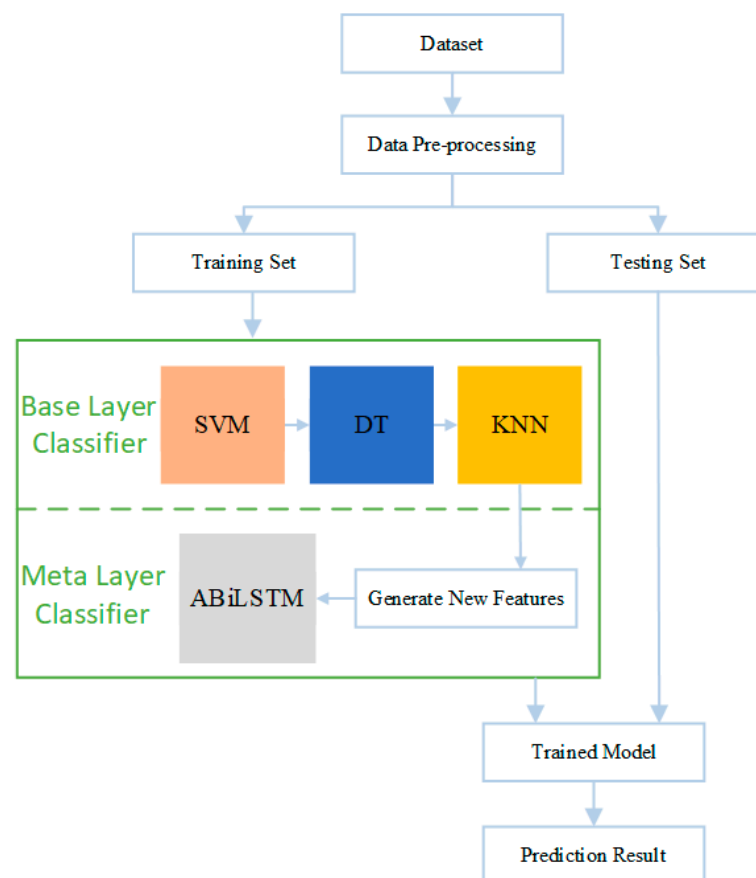


Figure 1. Flow chart of SDKABL model.

As shown in Figure 1, the SDKABL model is mainly composed of three parts, namely data preprocessing, the base layer, and the meta layer. In the data preprocessing module, the SDKABL model first processes the original data, mainly including processing outliers and dividing the training set and the test set. Then, feature importance analysis is carried out to screen out the few features most relevant to the prediction results. Finally, the PCA method is used for feature fusion. The SDKABL model uses the SVM, DT, and KNN ML methods for feature fitting and selection in the base classifier. Finally, the meta-layer learner

uses the features generated above to make the final prediction. Theoretical analysis is carried out below.

Suppose the original dataset is $D = \{(x_i, y_i)\}_{i=1}^n$, where $x_i \in R^d$ represents the eigen-vector of the i th sample and $y_i \in \{0, 1\}$ represents the label of the i th sample. While $y_i = 1$ means that the individual has heart disease, and $y_i = 0$ means that the individual does not have heart disease. We then split the dataset into an 80% training set and a 20% test set, i.e., $D_{train} = \{(x_i, y_i)\}_{i=1}^{0.8n}$, $D_{test} = \{(x_i, y_i)\}_{i=0.8n+1}^n$. Data preprocessing is performed on the dataset, including feature selection, outlier processing, and PCA. A tree-based model is used to analyze the importance of features, and the feature importance score is assumed to be $f_{importance}$. Each characteristic value x_{ij} is judged as to whether it is an outlier. If it is an outlier, it is filled or deleted. A new feature space $z_i = PCA(x_i)$, $z_i \in R^k, k < d$ is obtained by PCA. Multiple base learners (SVM, DT, KNN) are used to train the training set, and the prediction probability or category is output as a new feature. For each base learner M_j , the output is $f_{M_j} = M_j(D_{train})$, $f_{M_j} \in R^m$. Where $j \in \{1, 2, 3\}$ represents the output of base learners, such as SVM, DT, and KNN, respectively, which are combined into a new matrix $F = [f_{M_1}, f_{M_2}, f_{M_3}]$, $F \in R^{m \times 3}$. On the basis of the newly generated features, the ABiLSTM model is used for training and prediction. The input of ABiLSTM is a new eigen-matrix F, and its process can be described as $h_t = BiLSTM(F_t, h_{t-1})$. Where h_t indicates the hidden state of step t . On the basis of BiLSTM, an attention mechanism is added to assign the importance of different time step features. It can be expressed by the following formula: $\alpha_t = \frac{\exp(h_t^T W h_T)}{\sum_{t=1}^T \exp(h_t^T W h_T)}$. The final predicted output is $\hat{y} = \sigma \left(W_o \left(\sum_{t=1}^T \alpha_t h_t \right) + b_o \right)$. Where σ represents the sigmoid function, and W_o and b_o are the weight matrix and bias of the output layer.

The structure diagram of the ABiLSTM model is listed in Figure 2.

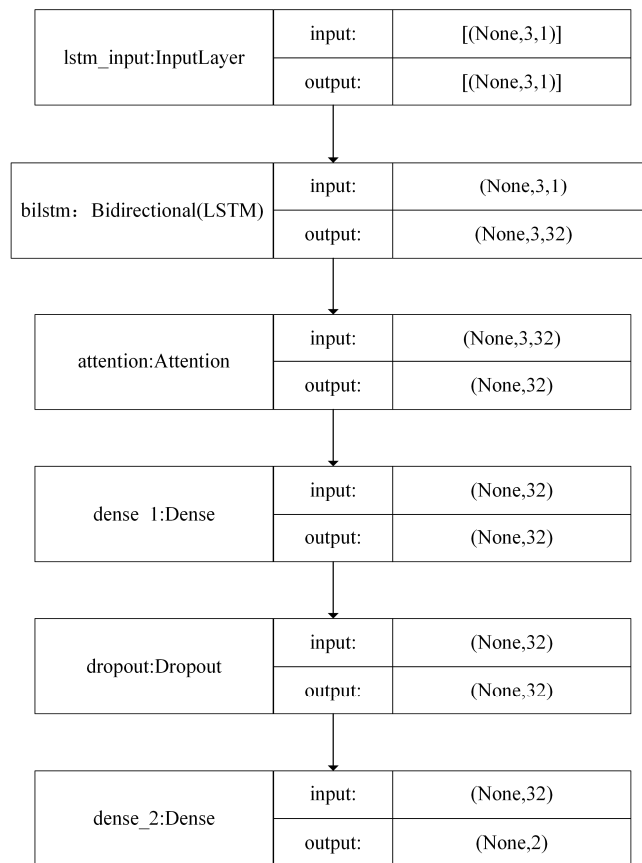


Figure 2. Structure diagram of ABiLSTM model.

3.4. Performance Metrics

In the classification system of this study, the data are divided into two main categories: one is the actual situation about whether an individual suffers from heart disease, which is defined as the true category of the data, the other is the disease condition predicted by the model, which is called the expected state. Based on these two pieces of information, we are able to construct a Confusion Matrix (CM). CMs, also called error matrices, are a common tool for assessing the capability of data mining models. In this matrix, each row represents the actual class of the sample, while each column shows the output result predicted by the model. The details are shown in Table 1.

Table 1. CM.

		Predict	
		1	0
True	1	TP	FN
	0	FP	TN

(1) Accuracy (Acc)

Accuracy is the proportion of model classification results that are correct.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

(2) Precision (Pre)

Precision is the proportion of true positive class in the data, the model classification results of which are positive class.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

(3) Recall (Rec)

The recall rate is the proportion of all positive data that is successfully predicted to be positive.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

(4) F1-score (F1)

The parameter closely related to the accuracy rate and recall rate is the F1-score, which is a comprehensive index to reconcile the accuracy rate and recall rate. This index is more objective and fair for model evaluation, and its calculation method is as follows:

$$\text{F1} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

(5) Area Under Curve (AUC)

The AUC of a curve represents the size of the area covered by the ROC curve. The corresponding calculation formula is as follows:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

Generally, if the AUC index of the ROC curve exceeds 0.5, it behaves as a convex curve, which indicates that the model is performing well. On the contrary, if it is lower than this

value, it implies that the classification performance is not satisfactory. Therefore, selecting a classifier with a high AUC value is the key to optimizing the model evaluation of machine learning algorithms.

3.5. Hyperparameter Optimization

The setting of hyperparameters greatly affects the performance of the model. Grid-SearchCV and other technologies are used to explore the hyperparameter space effectively, and the optimal parameters of the SDKABL model are determined. The SDKABL model's parameters are shown in Table 2.

Table 2. SDKABL model's parameters.

Model	Parameter	Value
SVM	C	1
	kernel	RBF
	gamma	scale
	degree	3
	tol	0.001
DT	max_depth	10
	min_samples_leaf	1
	min_samples_split	2
	criterion	gini
	splitter	best
KNN	n_neighbors	5
	weights	uniform
	algorithm	auto
ABiLSTM	bilstm node amount	32
	dense_1 node amount	32
	dense_1 activate function	relu
	dense_2 node amount	2
	dense_2 activate function	sigmoid
	learning_rate	0.001
	loss	binary_crossentropy

3.6. Experimental Environment

For data processing, we used the numpy and pandas libraries. We used seaborn and matplotlib for data visualization and result evaluation. In addition, we used a range of functions from sklearn, including machine learning models, PCA methods, and model evaluation functions. Finally, we used the tensorflow and keras libraries to build a neural network model. The deployment environment for the above code was built using Anaconda 2022.05, and the IDE tool was Pycharm 2022.3. All experiments in this paper are completed on Lenovo Rescuer R9000P 2021. The Rescuer is made in China and sourced from Haikou, China. The Rescuer consists of an AMD Ryzen 7 5800H developed by Advanced Micro Devices, Inc. (Santa Clara, CA, USA), memory from Samsung and an NVIDIA GeForce RTX 3070 Laptop made by Nvidia (Santa Clara, CA, USA).

4. Experiments and Results

In this section, the simulation results of the SDKABL model are evaluated and discussed. At the base level, three different ML classifiers, namely SVM, DT, and KNN, are implemented. ABiLSTM is deployed on the meta-layer of the proposed model. In order to obtain accurate and efficient results, we used PCA technology to perform feature fusion.

4.1. Dataset Description

For this study, we used the Cleveland dataset, the Framingham dataset, and the Z-Alizadeh Sani dataset. The Cleveland and Z-Alizadeh Sani datasets were obtained from the

UCI (The University of California, Irvine) machine learning repository [33]. The Cleveland dataset contains 303 instances, with 14 different attributes (13 predictors; 1 class), such as exang, oldpeak, slope, etc. (Table 3). The Z-Alizadeh Sani dataset contains 303 instances with 56 different attributes (55 predictors; 1 class) [34]. The Framingham dataset was obtained from the Kaggle website [35], and consists of 4240 instances with 16 attributes (15 predictors, 1 class) (Table 4).

Table 3. Cleveland dataset’s details.

Sr. No.	Attributes	Description	Range
1	Age	Age	[29, 77]
2	Sex	Sex (1 = Male, 0 = Female)	{0, 1}
3	Cp	Chest pain type	{0, 1, 2, 3}
4	Trestbps	Resting blood pressure	[94, 200]
5	Chol	Serum cholestoral	[126, 564]
6	Fbs	Fasting blood sugar > 120 mg/dL	{0, 1}
7	Restecg	Resting electrocardiogram results	{0, 1, 2}
8	Thalach	Maximum heart rate achieved	[71, 202]
9	Exang	Exercise induced angina	{0, 1}
10	Oldpeak	ST depression induced by exercise relative to rest	[0, 6.2]
11	Slope	The slope of the highest moving ST-segment	{0, 1, 2}
12	Ca	Number of major vessels (0–3) colored by flourosopy	{0, 1, 2, 3, 4}
13	Thal	Blood disease	{0, 1, 2, 3}
14	HeartDisease	1 = YES, 0 = NO	{0, 1}

Table 4. Framingham dataset’s details.

Sr. No.	Attributes	Description	Range
1	male	1 = YES, 0 = NO	{0, 1}
2	age	patient age	[32, 70]
3	education	patient’s education level	{0, 1, 2, 3, 4}
4	currentSmoker	current smoker	{0, 1}
5	cigsPerDay	number of cigarettes smoked per day	[0, 70]
6	BPMeds	whether to take blood pressure medication	{0, 1}
7	prevalentStroke	whether you have had a stroke	{0, 1}
8	prevalentHyp	whether you have high blood pressure	{0, 1}
9	diabetes	whether you have diabetes	{0, 1}
10	totChol	total cholesterol level	[107, 696]
11	sysBP	systolic blood pressure	[83.5, 295]
12	diaBP	diastolic blood pressure	[48, 142.5]
13	BMI	body mass index	[15.54, 56.8]
14	heartRate	heart rate	[44, 143]
15	glucose	glucose level	[40, 394]
16	HeartDisease	1 = YES, 0 = NO	{0, 1}

The Framingham dataset included missing attributes, and attributes with missing data included education, cigsPerDay, BPMeds, totChol, BMI, heartRate, and glucose. There is no direct relationship between the education attribute and the heart disease prediction, so the blank value of the education attribute is treated with all zeros. The cigsPerDay attribute is the number of cigarettes smoked in a day; the remaining data average is taken as 9 and filled in. The number of 0 value data of the BPMeds attribute is 4063, and the number of 1 value data is 124. The number of 0 value data is much greater than that of 1 value data, so we have blank data with 0 values. The BMI attribute contains 19 blank pieces of data. Take the average of the remaining 4221 pieces of data and fill in the blank with 25.8. The heartRate attribute has only one blank data block, filled with an average of 76 of the remaining 4239 data. The number of blank data in the glucose attribute is 388, and the average value of the remaining 3852 data is filled in. It is important to note that the unit of the age attribute is the number of days and can be converted to the number of years. Since the Z-Alizadeh Sani dataset contains a large number of continuous variables, it is necessary to normalize some of its features, including features FBS, TG, LDL, WBC, and PLT. Since the Z-Alizadeh Sani dataset contains a large number of continuous variables, it is necessary to normalize some of its features, including features FBS, TG, LDL, WBC, and PLT.

In each dataset, 20% of the samples were used for testing and 80% were used for training. The Z-Alizadeh Sani dataset has too many features, so it is not included in this list and heat map.

The heatmap function in the seaborn library is used to display the correlation between each feature. The correlation between the two characteristics is stronger for deeper grid colors; a positive correlation is shown by colors near red, and a negative correlation is indicated by colors near blue. Conversely, the lighter the grid color, the lower the correlation between the corresponding two features. The heat maps of the Cleveland and Framingham datasets are shown in Figures 3 and 4.

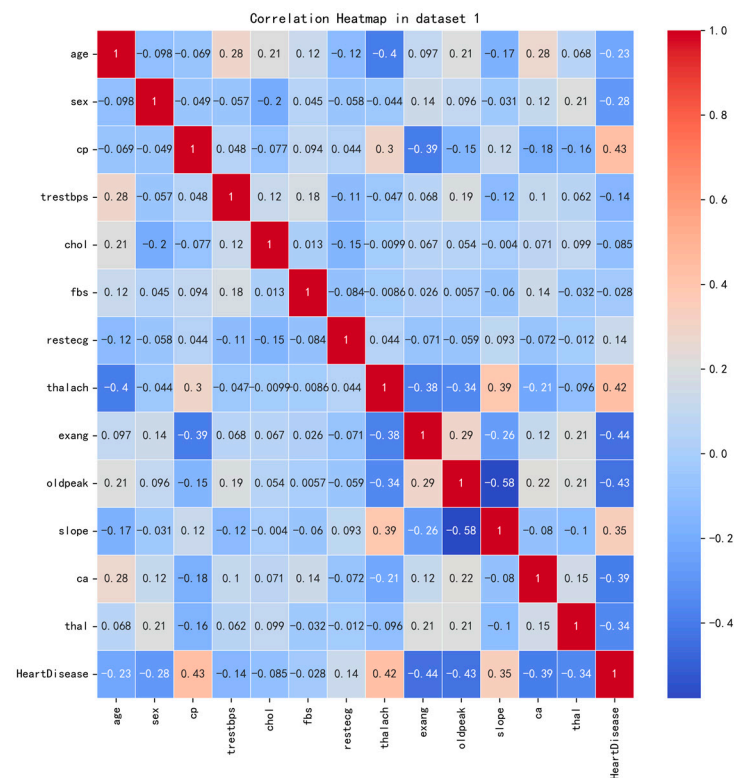


Figure 3. Correlation heatmap in the Cleveland dataset.

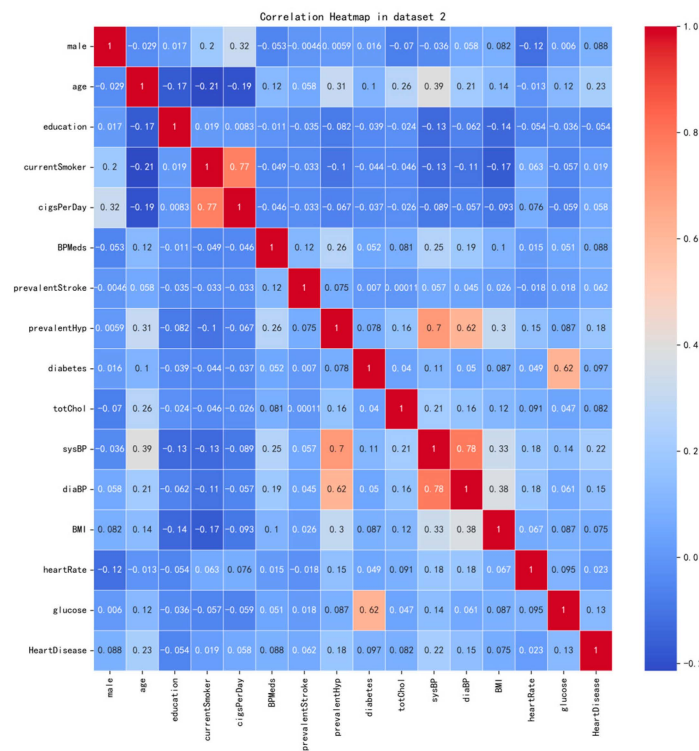


Figure 4. Correlation heatmap in the Framingham dataset.

As can be seen from Figure 3, feature trestbps, chol, thalach, oldpeak, ca, and HeartDisease are highly correlated with age. The features of trestbps, chol, oldpeak, and ca are positively correlated with age, while the features of thalach and HeartDisease are negatively correlated with age. Generally speaking, a feature increases with the increase in another feature that is positively correlated with it, and decreases with the increase in another feature that is negatively correlated with it, and vice versa. For example, the feature cp obtained from the heat map showed a high positive correlation with the feature HeartDisease. It can be seen from Table 3 that cp is the kind of chest pain; that is, the more severe the type of chest pain, the higher the likelihood of having heart disease, which is the same as common sense would suggest.

It can also be seen from Figure 3 that features of HeartDisease are highly correlated. From high to low, the features are exang, cp (feature oldpeak), thalach, ca, slope, thal, sex, age, trestbps(feature restecg), chol, and fbs.

As can be seen from Figure 4, features cigsPerDay and currentSmoker, features sysBP and diaBP, features prevalentHyp and diaBP, and features prevalentHyp and diaBP are highly correlated. Among them, the feature cigsPerDay represents the number of cigarettes smoked per day, and the feature currentSmoker represents whether or not one smokes, and only then will you obtain the number of cigarettes smoked per day. prevalentHyp, sysBP, and diaBP are all indicators of blood pressure. prevalentHyp represents whether or not you have hypertensive disease, sysBP represents systolic blood pressure, and diaBP represents diastolic blood pressure. People with hypertensive disease have high systolic and diastolic blood pressure, which is what people would expect. Feature glucose represents a person’s glucose level, and when glucose is high, it will lead to diabetes, so feature glucose is related to feature diabetes.

4.2. Experimental Results

In this section, we utilize different machine learning classifiers to build a prediction model and validate the efficacy of our suggested methodology across three datasets.

4.2.1. Experiment Using the Cleveland Dataset

First, Figure 5 and Table 5 compare the influence of the PCA method for feature fusion on prediction accuracy in dataset 1.

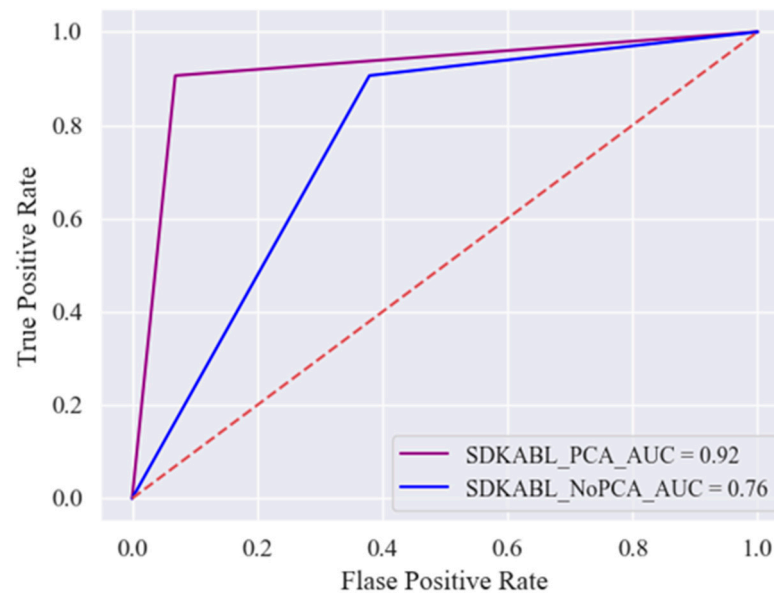


Figure 5. ROC curve of SDKABL model with and without PCA in dataset 1.

Table 5. Performance of SDKABL model with and without PCA in dataset 1.

Methods	Acc	Pre	Rec	F1	CM
SDKABL with PCA	0.918	0.918	0.919	0.918	27 2 3 29
SDKABL without PCA	0.770	0.791	0.763	0.777	18 11 3 29

From Figure 5 and Table 5, the influence of the PCA method on the performance indicators of the SDKABL model, including accuracy, precision, and recall, can be obtained. In addition, CMs are also given to show the performance of the classifier.

What can be obtained is that the data processed by PCA show better performance under the same classifier. The values of these four indexes all reach more than 91%, among which recall increases the most, by 15.6% compared with the data without PCA processing. Although precision shows the least improvement, it improves by 12.7%. In the CM, the data processed by PCA greatly reduces false positives in the classification of the SDKABL model and greatly increases the number of true negative samples. Finally, the AUC values of the two methods were obtained through the ROC curve. The AUC value of the model trained by PCA processing data was 0.92, which was much higher than that of the model trained without PCA processing, while the AUC value of the model trained without PCA processing was only 0.76.

Figure 6 and Table 6 show the comparison between the SDKABL method and the DT, KNN, and SVM. The max_depth parameter of DT is 5, and the n_neighbors parameter of KNN is 7.

From Table 6, the performance of the SDKABL model is superior to the other three comparison models in all aspects, followed by the DT model, the AUC value of which is 0.82, and finally, the KNN and SVM models, the AUC values of which are almost 0.70. Among them, all indicators of the SDKABL model reach more than 90%, and its AUC value is 0.92, which is 0.1 higher than that of the DT model after its second performance and about 0.2 higher than that of the KNN and SVM models.

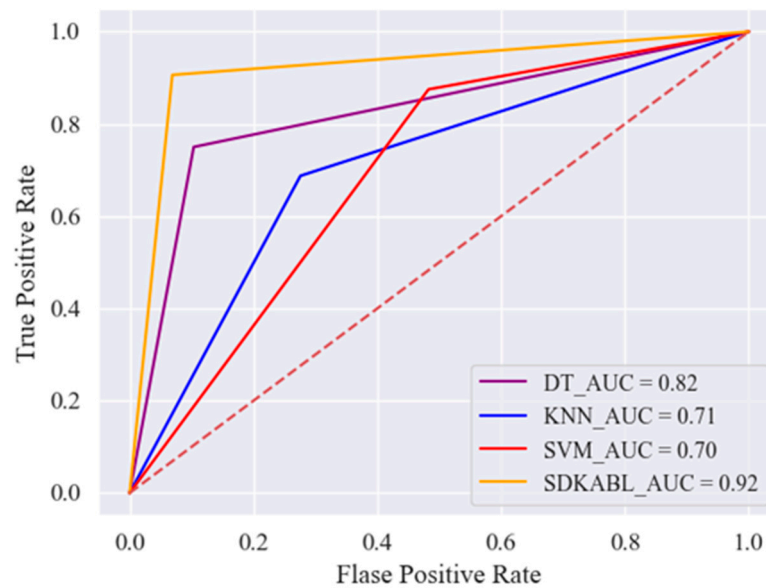


Figure 6. ROC curve of SDKABL model compared with base models in dataset 1.

Table 6. Performance of SDKABL model compared with base models in dataset 1.

Methods	Acc	Pre	Rec	F1	CM
SDKABL	0.918	0.918	0.919	0.918	27 2 3 29
DT	0.820	0.827	0.823	0.825	26 3 8 24
KNN	0.705	0.705	0.706	0.705	21 8 10 22
SVM	0.705	0.728	0.696	0.712	15 14 4 28

4.2.2. Experiment Using the Framingham Dataset

Figure 7 and Table 7 compare the influence of the PCA method for feature fusion on prediction accuracy in dataset 2.

Table 7. Performance of SDKABL model with and without PCA in dataset 2.

Methods	Acc	Pre	Rec	F1	CM
SDKABL with PCA	0.907	0.803	0.895	0.847	661 64 15 108
SDKABL without PCA	0.836	0.702	0.783	0.740	622 103 36 87

Figure 7 and Table 7 show the performance indexes of the PCA method for the SDKABL model using dataset 2.

It can be seen that the PCA data show better performance under the same classifier. Although the values of these four indicators are not as good as those of dataset 1, they all reach more than 80%, with the recall rate improving the most, by 11.2%, compared to the data without PCA processing. Although the improvement in accuracy is the smallest, it is also an improvement of 7.1%. In dataset 2, PCA-treated data similarly reduced false positives in SDKABL model classification and greatly increased the number of TN samples. Finally, the AUC values of the two methods were obtained by the ROC curve. The AUC

value of the PCA-treated model is 0.89, which is much higher than that of the non-PCA-treated model, and the AUC of the non-PCA-treated model is only 0.78.

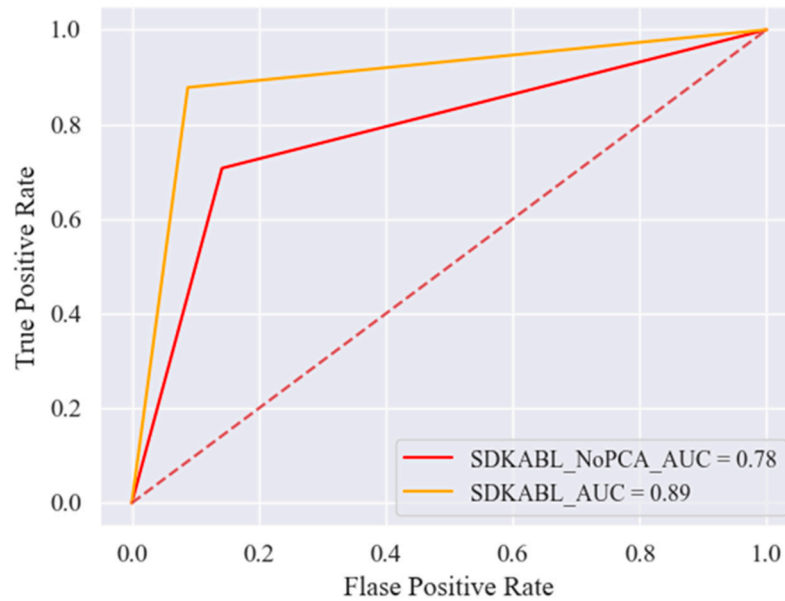


Figure 7. ROC curve of the SDKABL model with and without PCA in dataset 2.

Figure 8 and Table 8 show the SDKABL method compared to DT, KNN, and SVM. The best parameter of DT max_depth obtained by the GridSearchCV method is 7, and the n_neighbors parameter of KNN is 10.

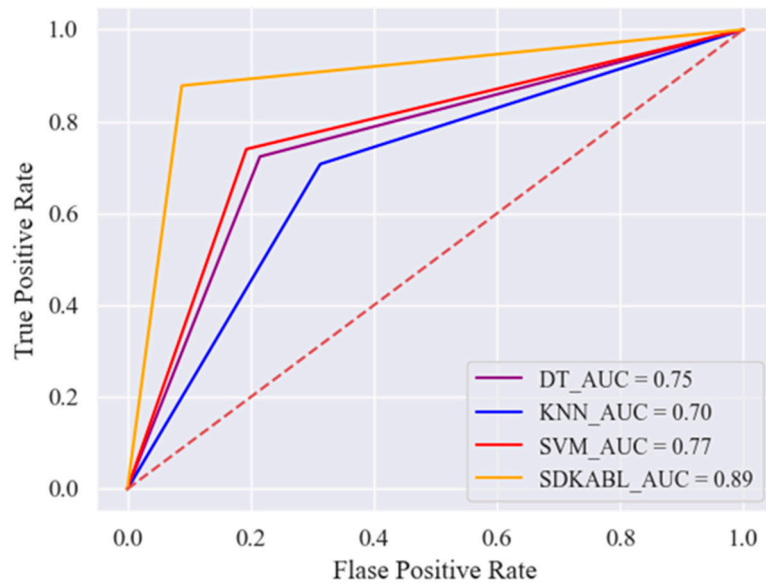


Figure 8. ROC curve of SDKABL model compared with base models in dataset 2.

From Table 8, the performance of the SDKABL model is superior to the other three comparison models in all aspects. The AUC values of the SVM, DT, and KNN models are 0.77, 0.75, and 0.70, respectively. Among them, all indicators of the SDKABL model in dataset 2 reach more than 80%, and the prediction accuracy rate is more than 90%. The AUC of SDKABL is 0.89, which is 0.12 higher than the SVM model, 0.14 higher than the DT model, and nearly 0.2 higher than the KNN model.

Table 8. Performance of SDKABL model compared with base models in dataset 2.

Methods	Acc	Pre	Rec	F1	CM	
SDKABL	0.907	0.803	0.895	0.847	661 15	64 108
DT	0.776	0.653	0.754	0.700	569 34	156 89
KNN	0.690	0.605	0.697	0.648	498 36	227 87
SVM	0.797	0.671	0.773	0.718	585 32	140 91

4.2.3. Experiment Using the Z-Alizadeh Sani Dataset

Figure 9 and Table 9 compare the influence of THE PCA method for feature fusion on prediction accuracy in Z-Alizadeh Sani dataset.

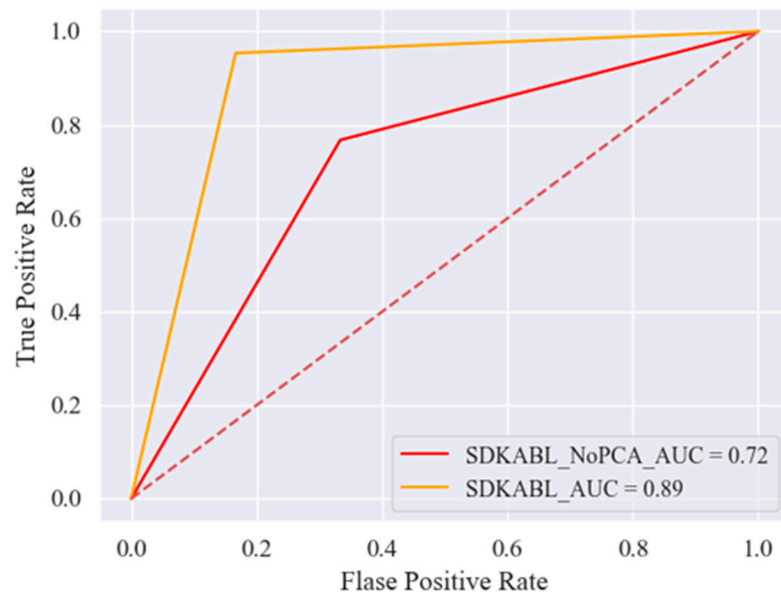


Figure 9. ROC curve of SDKABL model with and without PCA in dataset 3.

Table 9. Performance of SDKABL model with and without PCA in dataset 3.

Methods	Acc	Pre	Rec	F1	CM	
SDKABL with PCA	0.918	0.907	0.893	0.900	15 2	3 41
SDKABL without PCA	0.738	0.696	0.717	0.706	12 10	6 33

Figure 9 and Table 9 also show the performance indexes of the PCA method on dataset 3 of the SDKABL model proposed in this paper.

The classifier trained by PCA-processed data has better performance in classification performance, and has different degrees of improvement in multiple indicators. The performance of SDKABL on dataset 3 is similar to that of dataset 1, with both around 90%. Among them, the improvement in Precision was the largest, increasing by more than 20% compared to data without PCA processing. Although the improvement in accuracy is small, it is still 18%. In dataset 3, PCA-treated data significantly reduced false negatives in SDKABL model classification and significantly increased the number of true positive samples. Finally, the ROC curve obtains the AUC values of the two methods. The AUC

value of the PCA treated model is 0.89, which is much higher than that of the non-PCA treated model. In summary, it can be concluded that the PCA method plays a certain part in extracting key features and can provide certain help in data processing.

Figure 10 and Table 10 show the SDKABL method compared to decision trees, KNN, and support vector machines.

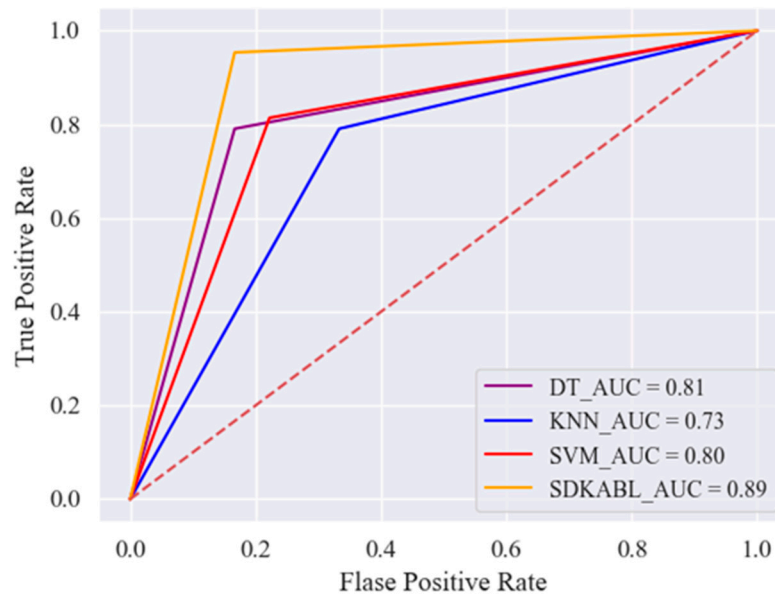


Figure 10. ROC curve of SDKABL model compared with base models in dataset 3.

Table 10. Performance of SDKABL model compared with base models in dataset 3.

Methods	Acc	Pre	Rec	F1	CM
SDKABL	0.918	0.907	0.893	0.900	15 3 2 41
DT	0.803	0.772	0.812	0.792	15 3 9 34
KNN	0.754	0.711	0.729	0.720	12 6 9 34
SVM	0.803	0.767	0.796	0.781	14 4 8 35

From Table 10, the performance of the SDKABL model is superior to the other three comparison models in all aspects. The AUC values of the DT, SVM, and KNN models were 0.81, 0.8, and 0.73, respectively. Among them, the indicators of the SDKABL model in dataset 3 are around 90%, and the prediction accuracy rate is more than 91%. The AUC value is 0.89, which is 0.08 higher than the DT model, 0.09 higher than the SVM model, and 0.16 higher than the KNN model. In summary, the superposition model with DT, KNN, and SVM as the base classifiers and ABiLSTM as the meta-layer classifier greatly improves the model’s performance compared with a single classifier.

4.2.4. Comparative Analysis of Existing and Proposed Methods

Many academics have employed ML methods to forecast heart disease in recent years. We also compared and analyzed the experimental results of the SDKABL model with those in the literature survey, which are shown in Table 11.

Table 11. Comparative analysis of existing and proposed methods.

#Ref	Dataset	Classifiers	Acc	Pre	Rec	F1
[36]	Cleveland dataset	SVM	88.5%	-	-	-
[37]	Cleveland dataset	KNN	88.5%	-	-	-
[38]	Cleveland dataset	Stacking Model	75.1%	-	-	-
[39]	Cleveland dataset	HRFLM	88.5%	90.1%	92.8%	91.4%
Our model	Dataset 1	SDKABL	91.8%	91.8%	91.9%	91.8%
	Dataset 2		90.7%	80.3%	89.5%	84.7%
	Dataset 3		91.8%	90.7%	89.3%	90.0%

Table 11 compares our proposed methods with those of other studies. These methods include SVM and KNN. Other methods include the stacking model involving KNN, random forest, and an SVM classifier [38], as well as the Hybrid Random Forest with Linear Model (HRFLM) [39]. It should be noted that the dataset 1 used in our model is the Cleveland dataset. From Table 11, the SDKABL model achieves considerable performance. Taken together, our proposed method performed best on the Cleveland dataset. On the other two datasets, the SDKABL model we proposed also achieves good results.

It can be seen that the stacking model using the DT, KNN, and SVM models as the base classifier and ABiLSTM as the meta layer classifier greatly improves the model performance compared with a single classifier.

5. Conclusions

This study makes a comprehensive exploration of heart disease prediction using feature selection and a variety of ML classification techniques. A systematic comparison of several ML algorithms focuses on accuracy and recall rate as the primary metrics to assess model performance. The results highlight the importance of feature selection and algorithm selection. This study emphasizes the importance of the PCA method in improving prediction accuracy and reducing dimension. By identifying the optimal configuration of each model, high accuracy is obtained, and the accuracy of the SDKABL model on the three datasets is more than 90%. These results demonstrate how well-tuned machine learning models may improve heart disease early detection. The next step will be to investigate large-scale dataset trials because of the dangers of overfitting and the requirement for strong models on huge datasets.

Author Contributions: Conceptualization, Y.W. and Z.X.; methodology, Z.F.; software, M.H.; validation, H.L.; formal analysis, Y.Z.; investigation, Z.X.; resources, Y.W.; data curation, H.L.; writing—original draft preparation, Z.X.; writing—review and editing, Z.F.; visualization, H.L.; supervision, M.H.; project administration, Y.W.; funding acquisition, Y.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Key R&D Program of China (Grant: 2021ZD0111000) and Key R&D Project of Hainan province (Grant: ZDYF2024SHFZ264).

Data Availability Statement: The Cleveland dataset presented in the study are openly available in UCI Machine Learning Repository at <https://archive.ics.uci.edu/dataset/45/heart+disease> (accessed on 22 August 2024). The Framingham dataset presented in the study are openly available in Framingham Heart study dataset at [35]. The Z-Alizadeh Sani dataset presented in the study are openly available in UCI Machine Learning Repository at <https://archive.ics.uci.edu/dataset/412/z+alizadeh+sani> (accessed on 29 August 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Available online: <https://www.who.int/zh/news-room/fact-sheets/detail/noncommunicable-diseases> (accessed on 24 August 2024).
2. Chong, A.Y.; Rajaratnam, R.; Hussein, N.R.; Lip, G.Y. Heart failure in a multiethnic population in Kuala Lumpur, Malaysia. *Eur. J. Heart Fail.* **2003**, *5*, 569–574. [[CrossRef](#)] [[PubMed](#)]
3. McAfee, A.; Brynjolfsson, E.; Davenport, T.H.; Patil, D.; Barton, D. Big data: The management revolution. *Harv. Bus. Rev.* **2012**, *90*, 60–68.
4. Ashton, K. That ‘Internet of Things’ thing. *RFID J.* **2009**, *22*, 97–114.
5. Dimitrov, D.V. Medical Internet of Things and big data in healthcare. *Healthc. Inform. Res.* **2016**, *22*, 156–163. [[CrossRef](#)]
6. Firouzi, F.; Rahmani, A.M.; Mankodiya, K.; Badaroglu, M.; Merrett, G.V.; Wong, P.; Farahani, B. Internet-of-Things and big data for smarter healthcare: From device to architecture, applications and analytics. *Future Gener. Comput. Syst.* **2018**, *78*, 583–586. [[CrossRef](#)]
7. Gunter, T.D.; Terry, N.P. The emergence of national electronic health record architectures in the United States and Australia: Models, costs, and questions. *J. Med. Internet Res.* **2005**, *7*, e383. [[CrossRef](#)] [[PubMed](#)]
8. Gotz, D.; Wang, F.; Perer, A. A methodology for interactive mining and visual analysis of clinical event patterns using electronic health record data. *J. Biomed. Inform.* **2014**, *48*, 148–159. [[CrossRef](#)] [[PubMed](#)]
9. Yadav, P.; Steinbach, M.; Kumar, V.; Simon, G. Mining electronic health records (EHRs): A survey. *ACM Comput. Surv. (CSUR)* **2018**, *50*, 85. [[CrossRef](#)]
10. Onyema, E.M.; Almuzaini, K.K.; Onu, F.U.; Verma, D.; Gregory, U.S.; Puttaramaiah, M.; Afriyie, R.K. Prospects and Challenges of Using Machine Learning for Academic Forecasting. *Comput. Intell. Neurosci.* **2022**, *2022*, 5624475. [[CrossRef](#)] [[PubMed](#)]
11. Ghassemi, M.; Naumann, T.; Schulam, P.; Beam, A.L.; Chen, I.Y.; Ranganath, R. A Review of Challenges and Opportunities in Machine Learning for Health. *AMIA Jt. Summits Transl. Sci. Proc.* **2020**, *2020*, 191–200. [[PubMed](#)]
12. Mohammed, A.; Kora, R. A comprehensive review on ensemble deep learning: Opportunities and challenges. *J. King Saud Univ.—Comput. Inf. Sci.* **2023**, *35*, 757–774. [[CrossRef](#)]
13. Brown, G. Ensemble Learning. In *Encyclopedia of Machine Learning*; Sammut, C., Webb, G.I., Eds.; Springer: Boston, MA, USA, 2011; pp. 312–320.
14. Mahajan, P.; Uddin, S.; Hajati, F.; Moni, M.A. Ensemble learning for disease prediction: A review. *Healthcare* **2023**, *11*, 1808. [[CrossRef](#)]
15. Nguyen, D.-K.; Lan, C.-H.; Chan, C.-L. Deep ensemble learning approaches in healthcare to enhance the prediction and diagnosing performance: The workflows, deployments, and surveys on the statistical, image-based, and sequential datasets. *Int. J. Environ. Res. Public Health* **2021**, *18*, 10811. [[CrossRef](#)] [[PubMed](#)]
16. Sagi, O.; Rokach, L. Ensemble learning: A survey. *WIREs Data Min. Knowl. Discov.* **2018**, *8*, e1249. [[CrossRef](#)]
17. Zhang, C.; Ma, Y. *Ensemble Machine Learning: Methods and Applications*; Springer: New York, NY, USA, 2012.
18. Pamir, J.; Javaid, N.; Akbar, M.; Aldegheishem, A.; Alrajeh, N.; Mohammed, E.A. Employing a Machine Learning Boosting Classifiers Based Stacking Ensemble Model for Detecting Non Technical Losses in Smart Grids. *IEEE Access* **2022**, *10*, 121886–121899. [[CrossRef](#)]
19. Pitchal, P.; Ponnusamy, S.; Soundararajan, V. Heart disease prediction: Improved quantum convolutional neural network and enhanced features. *Expert Syst. Appl.* **2024**, *249*, 123534. [[CrossRef](#)]
20. Chole, V.; Thawakar, M.; Choudhari, M.; Chahande, S.S.; Verma, S.; Pimpalkar, A. Enhancing heart disease risk prediction with GdHO fused layered BiLSTM and HRV features: A dynamic approach. *Biomed. Signal Process. Control* **2024**, *95*, 106470. [[CrossRef](#)]
21. Ganie, S.M.; Pramanik, P.K.D.; Malik, M.B.; Nayyar, A.; Kwak, K.S. An Improved Ensemble Learning Approach for Heart Disease Prediction Using Boosting Algorithms. *Comput. Syst. Sci. Eng.* **2023**, *46*, 3993–4006. [[CrossRef](#)]
22. Noor, A.; Javaid, N.; Alrajeh, N.; Mansoor, B.; Khaqan, A.; Bouk, S.H. Heart Disease Prediction Using Stacking Model With Balancing Techniques and Dimensionality Reduction. *IEEE Access* **2023**, *11*, 116026–116045. [[CrossRef](#)]
23. Mahajan, A.; Kaushik, B.; Rahmani, M.K.I.; Banga, A.S. A Hybrid Feature Selection and Ensemble Stacked Learning Models on Multi-Variant CVD Datasets for Effective Classification. *IEEE Access* **2024**, *12*, 87023–87038. [[CrossRef](#)]
24. Al-Tam, R.M.; Hashim, F.A.; Maqsood, S.; Abualigah, L.; Alwhaibi, R.M. Enhancing Parkinson’s Disease Diagnosis Through Stacking Ensemble-Based Machine Learning Approach. *IEEE Access* **2024**, *12*, 79549–79567. [[CrossRef](#)]
25. Mondal, S.; Maity, R.; Omo, Y.; Ghosh, S.; Nag, A. An Efficient Computational Risk Prediction Model of Heart Diseases Based on Dual-Stage Stacked Machine Learning Approaches. *IEEE Access* **2024**, *12*, 7255–7270. [[CrossRef](#)]
26. Jadoon, E.K.; Khan, F.G.; Shah, S.; Khan, A.; ElAffendi, M. Deep Learning-Based Multi-Modal Ensemble Classification Approach for Human Breast Cancer Prognosis. *IEEE Access* **2023**, *11*, 85760–85769. [[CrossRef](#)]
27. Al-Jamimi, H.A. Synergistic Feature Engineering and Ensemble Learning for Early Chronic Disease Prediction. *IEEE Access* **2024**, *12*, 62215–62233. [[CrossRef](#)]
28. Shaheen, I.; Javaid, N.; Alrajeh, N.; Asim, Y.; Aslam, S. Hi-Le and HiTCL: Ensemble Learning Approaches for Early Diabetes Detection Using Deep Learning and Explainable Artificial Intelligence. *IEEE Access* **2024**, *12*, 66516–66538. [[CrossRef](#)]
29. Mienye, I.D.; Domor, I.; Sun, Y.; Wang, Z. An improved ensemble learning approach for the prediction of heart disease risk. *Inform. Med. Unlocked* **2020**, *20*, 100402. [[CrossRef](#)]

30. Hassan, D.; Hussein, H.; Hassan, M. Heart disease prediction based on pre-trained deep neural networks combined with principal component analysis. *Biomed. Signal Process. Control* **2023**, *79*, 104019. [[CrossRef](#)]
31. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
32. Bahdanau, D.; Cho, K.; Bengio, Y. Neural Machine Translation by Jointly Learning to Align and Translate. *arXiv* **2014**, arXiv:1409.0473v7.
33. Detrano, R.; Jánosi, A.; Steinbrunn, W.; Pfisterer, M.; Schmid, J.; Sandhu, S.; Guppy, K.; Lee, S.; Froelicher, V. *Heart Disease—UCI Machine Learning Repository*; American Journal of Cardiology: New York, NY, USA, 1989. [[CrossRef](#)]
34. Alizadehsani, R.; Habibi, J.; Hosseini, M.J.; Mashayekhi, H.; Boghrati, R.; Ghandeharioun, A.; Bahadorian, B.; Sani, Z.A. A data mining approach for diagnosis of coronary artery disease. *Comput. Methods Programs Biomed.* **2013**, *111*, 52–61. [[CrossRef](#)]
35. Framingham Heart Study Dataset. Available online: <https://www.kaggle.com/datasets/aasheesh200/framingham-heart-study-dataset> (accessed on 29 August 2024).
36. Sun, H.; Pan, J. Heart disease prediction using machine learning algorithms with self-measurable physical condition indicators. *J. Data Anal. Inf. Process.* **2023**, *11*, 1–10. [[CrossRef](#)]
37. Jindal, H.; Agrawal, S.; Khera, R.; Jain, R.; Nagrath, P. Heart disease prediction using machine learning algorithms. *IOP Conf. Ser. Mater. Sci. Eng.* **2021**, *1022*, 012072. [[CrossRef](#)]
38. Shorewala, V. Early detection of coronary heart disease using ensemble techniques. *Inform. Med. Unlocked* **2021**, *26*, 100655. [[CrossRef](#)]
39. Mohan, S.; Thirumalai, C.; Srivastava, G. Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques. *IEEE Access* **2019**, *7*, 81542–81554. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.