

RESEARCH ARTICLE

OPEN ACCESS

Manuscript received February 10, 2024; revised February 26, 2024; accepted February 26, 2024; date of publication February 27, 2024
Digital Object Identifier (DOI): <https://doi.org/10.35882/jeeemi.v6i2.359>

Copyright © 2024 by the authors. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License ([CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/)).

How to cite: Arman Ghavidel, Pilar Pazos, Rolando Del Aguila Suarez, and Alireza Atashi, Predicting the Need for Cardiovascular Surgery: A Comparative Study of Machine Learning Models, Journal of Electronics, Electromedical Engineering, and Medical Informatics, vol. 6, no. 2, pp. 92-106, April 2024.

Predicting the Need for Cardiovascular Surgery: A Comparative Study of Machine Learning Models

Arman Ghavidel¹, Pilar Pazos¹, Rolando Del Aguila Suarez², and Alireza Atashi³

¹ Department of Engineering Management and Systems Engineering, Old Dominion University, Norfolk, USA

² School of Nursing, Hampton University, Hampton, VA, USA

³ Department of Digital Health, Tehran University of Medical Sciences, Tehran, Iran

Corresponding author: Arman Ghavidel (aghav001@odu.edu)

ACKNOWLEDGMENT: I am sincerely thankful to Dr. Pilar Pazos for her invaluable guidance and unwavering support during this research

ABSTRACT This research examines the efficacy of ensemble Machine Learning (ML) models, mainly focusing on Deep Neural Networks (DNNs), in predicting the need for cardiovascular surgery, a critical aspect of clinical decision-making. It addresses key challenges such as class imbalance, which is pivotal in healthcare settings. The research involved a comprehensive comparison and evaluation of the performance of previously published ML methods against a new Deep Learning (DL) model. This comparison utilized a dataset encompassing 50,000 patient records from a large hospital between 2015-2022. The study proposes enhancing the efficacy of these models through feature selection and hyperparameter optimization, employing techniques like grid search. A novel aspect of this research was the comparison of a newly developed DNN model with existing ensemble models based on similar cardiovascular datasets. The results indicated the DNN model's superior predictive accuracy, demonstrating an Area Under the Curve (AUC) of 74%, alongside notable precision (68%) and recall (72%) for the minority class, which indicates patients requiring surgery. The model further achieved a 70% F1-Score and a balanced accuracy rate of 72%, significantly outperforming the existing ensemble models in every key performance metric. The study underscores the transformative potential of DNNs in predictive modeling for cardiovascular care and highlights the importance of integrating advanced ML techniques into clinical workflows. Future research should delve into the practical application and integration of these models.

INDEX TERMS: clinical decision-making, deep neural networks, health feature selection, resampling methods, imbalanced data, machine learning.

I. INTRODUCTION

In today's world, where electronic health data is widely collected and stored, there's an increasing conviction that Artificial Intelligence (AI), particularly its branch of Machine Learning (ML) could revolutionize healthcare practices. The utilization of machine learning (ML) to support understanding of cardiovascular disease represents a burgeoning field, with potential to enhance diagnostic precision and tailor treatment strategies [1]. Incorporating ML to support cardiovascular care is a prime example of how these advanced technologies can help inform healthcare practices, especially in critical surgical settings [2]. These sophisticated systems can help inform diagnostic methods

and personalize treatment plans. However, compared to traditional medical diagnostics, its effectiveness remains a subject of ongoing research. Studies have shown that ML can be particularly valuable in areas rich with complex and large datasets. The utilization of ensemble machine learning models and deep learning techniques in healthcare data analysis represents a significant paradigm shift in predictive modeling, particularly in critical clinical decision-making contexts. Ensemble models, which combine the predictive power of multiple base models, and deep neural networks (DNNs), a specialized category of machine learning algorithms, are pivotal in this transformation [3]. These complementary approaches empower healthcare organizations to make more accurate and informed decisions based on the intricate nuances of the data that support an

TABLE 1
 A summarized list of ensemble and deep learning models used for cardiovascular-related prediction.

Article	Ensemble Model						Validation			
	Bagging	Stacking	Boosting	RF	Resampling	Deep Learning	Output variables			
							Hold-Out	K-Fold CV	Feature Selection	
[11]	*	*	*	*	*		Mortality in ICU of cardiac surgery	*	*	*
[15]		*	*	*			Mortality in ICU of cardiac surgery	*	*	
[16]		*	*	*	*		Mortality in ICU of cardiac surgery	*	*	*
[17]		*			*		Major Adverse Cardiovascular Events in Patients with Acute Coronary Syndrome	*		*
[18]	*	*	*	*	*		Bariatric surgery complications.	*	*	*
DL						*	Cardiovascular surgical intervention	*		*

advancement in the field [4]. Predicting the need for surgery is a critical area, and ML can profoundly impact clinical settings due to the cost and risks associated to cardiovascular surgery [5]. ML employs algorithms that dynamically adapt to data, creating models capable of efficiently processing and understanding complex data. Machine learning is increasingly important in cardiovascular surgery, offering the potential to reduce costs. Machine learning can optimize surgical and postoperative processes by analyzing data to improve support treatment strategies and patient care. This is crucial in cardiovascular care, where surgery costs are high. Effective use of machine learning can lead to more efficient surgical planning, fewer complications, and better patient outcomes, thereby reducing the financial burden of cardiovascular surgeries [6-8]. While ensemble models excel at capturing patterns and insights from diverse and extensive healthcare datasets, DNNs, with their remarkable capacity to extract intricate patterns in rich and large healthcare datasets, also play a central role. Traditional statistical methods struggle with the multi-dimensionality and volume of clinical data, whereas ensemble models and DNNs offer a more nuanced and accurate approach. Machine Learning (ML) techniques, including ensemble and Deep Neural Networks (DNNs), excel at analyzing vast and intricate datasets, demonstrating benefits in forecasting clinical results [9, 10]. The challenge of class imbalance in surgical datasets, where non-surgical cases often outnumber surgical cases, is a significant hurdle in predictive modeling. Methods such as the Synthetic Minority Over-sampling Technique (SMOTE) and Adaptive Synthetic Sampling (ADASYN) are employed to mitigate this challenge, successfully equalizing datasets to enhance the model's forecasting accuracy [11]. Feature selection is also crucial in determining the efficacy of DNN

models. Techniques like SelectKBest and Random Forest reduce data dimensionality, enhancing model efficiency and interpretability by selecting the most influential features on the target variable prediction [12]. This research aims to compare a Deep Neural Network (DNN) and several previously published ensemble ML models to forecast the need for surgical intervention using a dataset of 50,000 patient records from an Iranian hospital. The data was collected between 2015 to 2022. This study also addresses critical issues in clinical datasets like class imbalance and feature selection. The suggested method utilizes a grid search algorithm for optimizing hyperparameters, providing a methodical way of adjusting parameters that dictate machine learning models' learning procedures and structures. Differing from model parameters, hyperparameters are predetermined before the training phase and have a notable impact on the model's effectiveness. The rationale behind selecting these specific hyperparameters was to balance the model's accuracy and generalizability to unseen data and ensure equitable prediction performance for both minority and majority classes in our imbalanced dataset [13, 14]. In this study, the selection of Deep Neural Networks (DNNs) over Convolutional Neural Networks (CNNs) is strategically justified by the nature of our data and the specific objectives of our research. Unlike CNNs, which excel in extracting features from image-based data through their specialized architecture, DNNs are more adept at processing the structured, high-dimensional datasets commonly found in healthcare analytics. This capability is particularly crucial for our focus on cardiovascular disease prediction, where the data encompasses a wide array of variable types—numeric, categorical, and binary. DNNs leverage their deep, layered architecture to uncover complex patterns within such

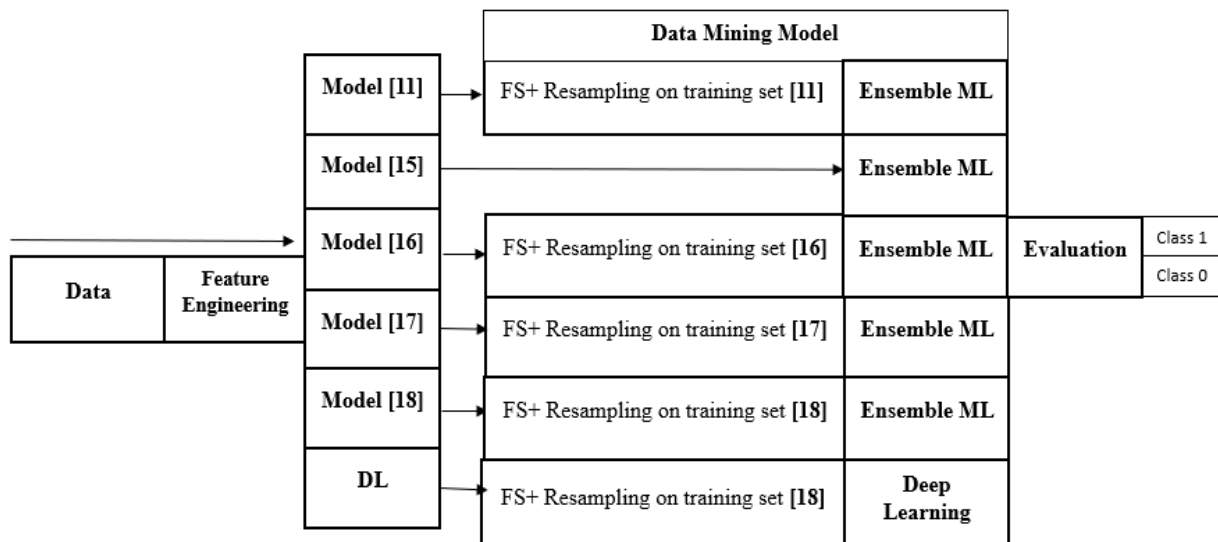


FIGURE 1. Various phases of the suggested approach.

datasets, negating the need for manual feature extraction and thus aligning perfectly with our methodology [19].

Research question: How can machine learning models, including deep neural networks, improve the prediction of surgical interventions for cardiovascular disease patients?

The distinctive contributions and significant methodologies employed in this study, when compared to prior approaches, encompass:

- 1) A new hybrid model incorporating feature selection, resampling, and a DNN.
- 2) A comparison between the proposed DNN and previously published ensemble models used on cardiovascular disease datasets with similar variables.
- 3) Feature engineering and data augmentation on high dimensional data.
- 4) Applying a grid search approach, enabling precise adjustment of hyperparameters, class weights, and decision-making thresholds enhancing the model's ability to accurately predict outcomes for both minority and majority classes.
- 5) Enhancing the recall for the minority class (the important class based on the problem definition) while keeping the precision high.

The following sections describe the dataset and the preprocessing methods, including data feature engineering, and feature selection for dimensionality reduction, feature scaling, and handling imbalanced data problems. TABLE 1 describes the studies selected for this comparative research and the methods that were used in them.

II. METHODOLOGY

In this study, we strategically selected a subset of ensemble models previously applied to cardiovascular-related datasets. Our focus was on models that demonstrated potential in handling large, imbalanced datasets, as this aligns with our study's context and objectives. We did not include all models historically used in this domain but chose representative ones that offered varied cardiovascular disease outcome prediction approaches. This selection was intended to provide a comprehensive overview of different methodologies and their performance in similar scenarios. The comparison is based on several performance indexes crucial for evaluating models on imbalanced datasets, commonly seen in disease prediction, including precision, recall, F1-score, AUC (Area Under the Curve), and balanced accuracy. These metrics collectively provide a comprehensive assessment of model performance, ensuring robustness and reliability in predicting cardiovascular disease outcome. FIGURE 1 shows the research methodology diagram, including different steps towards comparing the models. The aim of this research is to enhance the predictive performance metrics for the minority class without compromising the metrics for the majority class. All the utilized models are coded in Python and carried out in google Colab.

A) DATASET INFORMATION

The dataset chosen for this study was gathered and documented from hospitals connected to Shahid Beheshti University of Medical Sciences and Health Services in Iran during the period 2015 to 2021. The original dataset contains 134 demographical, clinical, and lifestyle variables recorded for 50060 patients. Given the high dimensionality of the data, this research employs the Random Forest feature importance to identify the most influential features for predicting

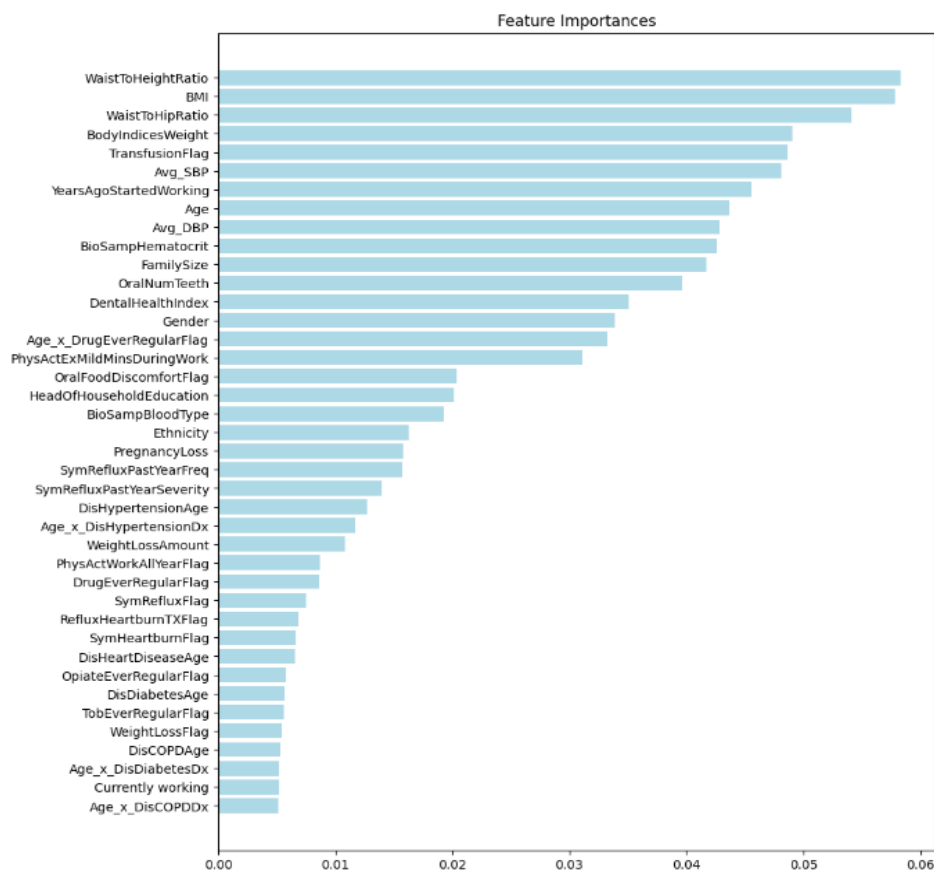


FIGURE 2. Variables extracted by RF feature importance.

cardiovascular surgery intervention. TABLE 2 and FIGURE 2 indicate these features.

B) DATA PREPROCESSING

Data preprocessing is a critical task in data mining. The data must be of optimal quality, making it appropriate for use with various ML models. Enhanced data quality directly correlates with improved forecasting accuracy of ML models. High-quality data typically refers to accurate, complete, relevant, and consistently formatted information in healthcare settings [20]. The utilized preprocessing approaches shown in Figure 1 are discussed as follows:

1. FEATURE ENGINEERING

This study employed feature scaling on continuous variables to enhance the predictive performance of the ML models. This process is crucial given the diverse nature of the utilized clinical data containing a mix of categorical, numeric, and binary variables. In datasets where numeric features have different ranges or units of measurement, feature scaling becomes a critical preprocessing step. This process involves adjusting the scales of various numeric features to a uniform

range, ensuring that no single feature disproportionately influences the model's performance due to its scale. This study used Standard Scaler from Scikit-Learn for the numeric variables in our dataset. Standard Scaler transforms the data such that the distribution of each feature has a mean of zero and a unit variance, not a unit standard deviation. This means that while the variability of each feature is maintained, the values are rescaled to ensure that each feature contributes equally to the analysis. This proves especially advantageous for algorithms that react to the scale of input variables, like neural networks. This normalization process was applied uniformly across both the training and test datasets to maintain consistency in data processing [21, 22]. This study employed preprocessing techniques that are suitable for structured clinical and demographic data, rather than methods traditionally used in signal processing such as Root Mean Square (RMS), Mean Absolute Value (MAV), or Slope Sign Changes (SSC). This strategy, aimed at addressing the dataset's high dimensionality and diversity, did not specifically engage in time, frequency, or wavelet domain feature extractions, focusing instead on streamlining the dataset to emphasize the most informative features for

TABLE 2
 Main variables of the dataset.

Feature Name	Description	Type
WaistToHeightRattio	A measure to assess the body fat and obesity	Numeric
WaistToHipRatio	A measure to assess the body fat and obesity	Numeric
BMI	Body Mass Index	Numeric
BodyIndicesWeight	Specifies patient weight	Numeric
TransfusionFlag	If the patient has Transfusion history 1 if not 0	Nominal
YearsAgoStartedWorking	The number of years since an individual started working	Numeric
Age	Age of the patient in years	Numeric
BioSampHematocrit	Specifies patient Hematocrit (HCT)	Numeric
SymRefluxPastYearFreq	If the patient has Reflux symptom in past year 1 if not 0	Nominal
SymRefluxPastYearSeverity	Severity of reflux symptom in past year	Categorical
DisCOPDDx	If a patient has COPD 1, otherwise 0	Nominal
OralNumTeeth	Specifies the patient Total number of teeth	Numeric
Gender	Specified the patient gender, Male:1, and Female:0	Nominal
Family size	Number of people in the family	Numeric
DrugEverRegularFlag	Specifies if the patient uses drug regularly or not	Categorical
Ethnicity	Specifies the patient's Ethnicity	Categorical
Avg DBP	Average diastolic blood pressure	Numeric
Avg SBP	Average systolic blood pressure	Numeric
Pregnancy loss	Number of pregnancy loss for a female	Numeric
PhysActExMildMinsDuringWork	Specifies patient mild sport activity during work (mins)	Numeric
OralFoodDiscomfortFlag	Specifies if the patient had discomfort when eating in past year	Nominal
Education	Specifies the patient's Education Level	Categorical
BioSampBloodType	Specifies patient's blood type	Categorical
Dental Health index	Is the sum of two variables - OralNumTeethDecayed and OralNumTeethMissing where a higher number indicates poorer dental health.	Numeric
PregnancyLoss	If a female patient had Pregnancy loss 1, otherwise 0	Nominal
DisHypertensionDx	If a patient has hypertension 1, otherwise 0	Nominal
WeightLossAmount	the amount of weight a person has lost past year	Numeric
DisCOPDAge	Patient's age in first COPD diagnosis	Numeric
DisDiabetesDX	If the patient has diabetes 1, if not 0	Nominal
SymRefluxFlag	if the patient has Reflux Symptom history 1 if not 0	Nominal
SymHeartburnFlag	If the patient has Heartburn Symptom history 1 if not 0	Nominal
DisHeartDiseaseAge	Patient's age in first heart disease diagnosis	Numeric
OpiateEverRegularFlag	If the patient consumes opiate regularly 1, otherwise 0	Nominal
Currently working	If the patient is currently working 1, if not 0	Nominal

cardiovascular surgery prediction [23]. In this study's feature engineering phase, we employed several techniques to enhance our clinical dataset's analytical utility. Though interconnected, each method serves a distinct analytical purpose. Aggregating related variables created aggregate metrics, offering a holistic view of complex health conditions. For instance, two combined blood pressure metrics were generated called Avg_DBP and Avg_SBP by averaging different diastolic and systolic blood pressure measurements to reduce dimensionality of the data. Lastly, developing indices to quantify health risks involved synthesizing multiple health variables into a single risk score. Our study improved the dataset by creating indices that better represent health risks. The 'DentalHealthIndex' was developed by summing 'OralNumTeethDecayed' and 'OralNumTeethMissing,' providing a simplified dental health measure. We also calculated 'PregnancyLoss' to reflect reproductive health, derived from the difference between 'NumberPregnancies' and 'NumberLiveBirths.' Additionally, we introduced body composition ratios, 'WaistToHipRatio' and 'WaistToHeightRatio,' as indicators of obesity-related

health risks. We removed the original variables that contributed to these aggregates. Feature engineering was taken to streamline the dataset and eliminate redundancy. The new aggregates offered a more comprehensive view of health risks and conditions, optimizing the dataset for effective machine learning analysis by focusing on the most informative and synthesized features. These techniques, while related, uniquely contribute to the depth and accuracy of our analysis [24, 25]. Interactions of disease with age factors have been considered, and One-hot encoding and categorical binning were used to transform categorical variables into a format that can be provided to ML algorithms. An average frequency score for specific symptoms quantifying their occurrence and severity was also calculated. This involved assigning custom weights to different levels of symptom frequency and severity. For instance, reflux frequencies like 'Almost every day' were assigned a higher weight (5) compared to 'Never' (0). Similarly, severity levels ranged from 'Slight' (1) to 'Lack of signage' (5). These methods played a crucial role in refining the raw data, enhancing its suitability for in-depth analysis.

TABLE 3
Machine learning models, along with their distinct parameter configurations

Article	Methods	Parameters
[11]	RF_feature_importance+ Stacking	Base Classifiers: Logistic Regression (LR), Bagging Classifier, Decision Tree (DT), Random Forest (RF), XGBoost (XGB); Meta-classifier: LR; Resampling: SVMSMOTE + EditedNearestNeighbours (ENN)
[15]	Random Forest, AdaBoost, Gradient Boosting, Stacking	Random Forest: n_estimators=200; AdaBoost: n_estimators=40; Gradient Boosting: n_estimators=1000, learning_rate=0.01; Stacking (RF): Base Classifiers - LR, RF, NB, DT, Meta-classifier - RF (n_estimators=3000); Stacking (GLM): Base Classifiers - LR, RF, NB, DT, Meta-classifier - Logistic Regression
[16]	SVM-SMOTE+GA+ Stacking	GA: pop_size=100, num_gen=20, mut_prob=0.2, cross_prob=0.5; Stacking: MLP, KNN, ExtraTrees, SVM as meta-classifier; SVM: kernel='rbf', n_estimators=57
[17]	RF_feature_importance +Stacking+SMOTETomek	Stacking: LR, DT, RF, XGB, KNN, Ada; Meta-classifier: RF; Hyperparameters tuned via GridSearchCV
[18]	CFS+SMOTE+Ensemble	CFS: SelectKBest, top features selected; SMOTE: default settings; Ensemble: RandomForestClassifier (n_estimators=100), AdaBoostClassifier (n_estimators=100), BaggingClassifier (base_estimator=DecisionTreeClassifier, n_estimators=100); Resampling: SMOTE

The prior description encompasses only a portion of the extensive feature engineering performed, with many more modifications undertaken to ensure a thorough and nuanced dataset evaluation.

2. FEATURE SELECTION

In the domain of machine learning, selecting features is a crucial procedure that boosts the efficacy of models by pinpointing and employing the key features from the data collection [26]. The utilized feature selection method in each of the selected models is shown in TABLE 3. Model [15] is the only model for which feature selection is not used. The DL model in this study, similar to model [11] and [17] employs Random Forest feature importance, a robust method that evaluates and ranks features based on their contribution to model accuracy. This approach not only streamlines the model by reducing dimensionality and computational complexity but also improves predictive accuracy by focusing on significant, non-redundant features [12, 27]. By effectively eliminating irrelevant and redundant data, Random Forest feature importance aids in mitigating overfitting and enhancing the generalizability of the model to new data [28]. Figure 2 illustrates a bar chart that identifies the most influential factors in predicting the necessity for cardiovascular surgery intervention. This step revealed the relevance of obesity and body composition indicator features as the most important variables. This corresponds with

established studies that associate obesity, a significant marker of cardiovascular health, with a heightened risk of cardiovascular disease (CVD) and overall mortality. Additionally, the variable 'PhysActExMildMinsDuringWork', indicative of mild physical activity levels during work hours, emerges as a significant predictor. This finding highlights the role of occupational physical activity, or lack thereof, in cardiovascular health and supports existing studies that associate sedentary behavior with heightened risks for CVD and mortality [29-31].

Age is the other most important feature, which is consistent with extensive medical literature that recognizes age as a fundamental risk factor for cardiovascular diseases (CVD). As age increases, the risk of CVD typically increases due to various factors like arterial stiffness, hypertension, and endothelial dysfunction [32]. The bar chart prominently displays reflux symptom indicators as significant features, suggesting that symptoms of reflux, possibly related to gastroesophageal reflux disease (GERD), occurring frequently over the years, have a substantial impact on the model's predictions. This aligns with medical findings that GERD can be associated with respiratory conditions and might indirectly influence cardiovascular health [33]. "OralNumTeeth" and "DentalHealthIndex" suggest an interesting connection between dental health and

cardiovascular risk, a relationship that has been explored in epidemiological studies.

Poor oral health, particularly periodontal disease, has also been associated with higher rates of CVD [34]. Family-related features like “FamilySize” and “Currently working” suggest socioeconomic factors and family support structures may also play roles, potentially affecting stress levels and access to healthcare, which are known contributors to cardiovascular health [35]. Avg_DBP and Avg_SBP, representing average diastolic and systolic blood pressure, respectively, are key cardiovascular health indicators. Hypertension is a major risk factor for CVD, and its management is necessary in preventing the need for surgical intervention [36]. Drug abuse indicators like “OpiateEverRegularFlag” are also shown as influential features of cardiovascular problems [37, 38]. In conclusion, the feature importance bar chart provides a valuable insight into the range of factors that the machine learning model has identified as significant in predicting the need for cardiovascular surgical intervention. The factors span from direct clinical measures to lifestyle and socioeconomic indicators, offering a multifaceted view of patient health. These findings are largely consistent with established medical knowledge and underscore the complexity of cardiovascular disease risk assessment [39].

3. APPROACH TO IMBALANCED DATA PROBLEM

An imbalanced data problem is a common challenge in predictive models based on clinical datasets because positive cases of a disease or diagnosis are typically part of a minority-class sample [26]. In this case, the majority class samples representing patients with a negative diagnosis are remarkably higher than the minority class samples. Therefore, ML models are more proficient in forecasting outcomes for the majority class compared to the minority class in datasets with imbalances. The clinical data for this study is imbalanced, with 38 percent of the total patients having undergone Cardiovascular Surgery. Multiple strategies have been developed to tackle the challenge of datasets with imbalances, typically applied during the preprocessing stage. Among various strategies to address data imbalance, resampling is widely adopted, particularly in scenarios where collecting or augmenting new data is impractical or impossible. Resampling balances the dataset by modifying the representation of classes [40]. SMOTE and ADASYN are two popular methods used to tackle imbalanced datasets in machine learning. SMOTE enhances data balance by generating artificial samples in the minority class. The process involves selecting a random minority class instance and then identifying its nearest neighbors in the feature space. It then synthesizes new examples by adding a scaled difference between the chosen point and its neighbors,

effectively creating similar yet distinct instances [41]. ADASYN builds upon SMOTE by introducing a more focused approach in generating synthetic samples. It prioritizes those minority class samples that are more challenging for the model to learn, determined by the proximity of majority class samples. ADASYN adapts the number of synthetic samples based on the learning difficulty of each minority class sample, producing more synthetic data in regions where the classifier struggles the most. This strategy leads to a more nuanced balancing of the class distribution, particularly improving model performance in more complex or challenging areas of the data space. Both techniques aim to equalize the representation of classes in imbalanced datasets, enhancing the robustness of machine learning models [42]. The specific resampling methods used in each model is shown in table 3.

C) MACHINE LEARNING MODELS

Various machine learning models are capable of resolving classification issues [43]. This study focuses on evaluating and comparing different ensemble machine learning models, specifically those previously applied to similar datasets, in the context of cardiovascular disease data. Moreover, it proposes a Deep Learning model that targets some of the weaknesses of prior approaches.

1. DEEP LEARNING MODEL

In this research, a deep learning model was created using TensorFlow and Keras, tailored to the specific requirements of our clinical dataset. The dataset, initially preprocessed for data quality enhancement, was split into training and testing sets in an 80-20 ratio. This division occurred before the standardization process, ensuring that the training and test sets were independently standardized. Continuous features in both sets were normalized using the Standard Scaler from Scikit-Learn, which adjusted each feature's distribution to have a mean of zero and unit variance. This normalization was essential for algorithms that are responsive to the scale of input attributes, particularly neural networks. Post-standardization, the training set underwent hybrid resampling using ADASYN (Adaptive Synthetic Sampling) to address class imbalance. This method effectively balanced the dataset by generating synthetic samples in the minority class, enhancing the model's ability to learn from a more data distribution. Class weights were computed to manage the imbalance within the training data further. These weights were used during the model training phase to ensure a balanced representation of both minority and majority classes, allowing for more effective learning. The deep learning model, constructed as a Sequential model in Keras, included multiple dense layers with ELU (Exponential Linear Unit) activations and dropout layers. The model's

architecture and hyperparameters, such as the number of layers, units per layer, dropout rates, and learning rate for the Adam optimizer, were optimized using Keras Tuner's Random Search. This optimization process was vital for tailoring the model to our dataset's specific characteristics. The training involved binary cross-entropy as the loss function, with class weights incorporated to handle the imbalance. Early stopping based on validation loss was implemented to prevent overfitting, ensuring the model ceased training at the optimal point. The model's performance was evaluated on the test set, focusing on metrics such as accuracy, F1 score, balanced accuracy, precision, and AUC score. Particular emphasis was placed on the recall score, given its importance in the context of our problem definition. This comprehensive evaluation provided insights into the model's effectiveness in classification tasks. The flow diagram of the deep learning model, as illustrated in FIGURE 3, outlines this structured process, from data preprocessing to model evaluation. This approach ensured the robustness and reliability of the model, making it a valuable tool in our study's machine-learning arsenal.

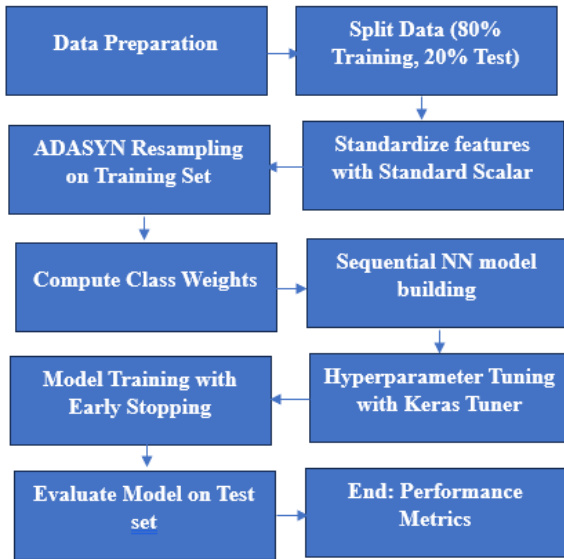


FIGURE 3. The Deep Neural Network model's flowchart

Data Standardization for continuous variables can be formulated as shown in Eq. (1).

$$X_{scaled} = \frac{X - \mu}{\sigma} \quad (1)$$

where X is the input data, μ is the mean, and σ is the standard deviation of the training data features. ADASYN Resampling can be formulated as shown in Eq. (2).

$$x_{new} = x_i + (x_{zi} - x_i) \cdot \lambda \quad (2)$$

where x_{zi} is a randomly selected neighbor from the k nearest neighbors of x_i in the minority class, and λ is a random

number between 0 and 1. This process is repeated until the class distribution is considered balanced, with the number of synthetic samples for each minority class sample being adjusted based on the density distribution of the minority class. Sequential Model with Dense Layers can be encapsulated by Eq. (3).

$$h_i = \text{ELU}(w_i h_{i-1} + b_i) \quad (3)$$

Where h_{i-1} is the previous layer's output or the first layer's input, W_i and b_i are the weights and biases, and ELU is the activation function. Dropout is shown in Eq. (4).

$$h_i' = h_i \odot M, \quad M_j \sim \text{Bernoulli}(p) \quad (4)$$

where \odot denotes element-wise multiplication, and M_j is the j^{th} element of the mask M , indicating whether the j^{th} unit in the layer output h_i is kept or dropped. Binary Cross-Entropy Loss is shown in Eq. (5).

$$L(y, \hat{y}) = -\frac{1}{N} \sum_{i=1}^N [y_i \cdot \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (5)$$

Where y_i is the true label, \hat{y}_i is the predicted probability, and N is the number of samples. Adam Optimizer's procedure can be formulated as shown in Eqs. (6), (7), (8), (9), and (10) respectively.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (6)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (7)$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (8)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (9)$$

$$w_{t+1} = w_t - \frac{\eta \hat{m}_t}{\epsilon + \sqrt{\hat{v}_t}} \quad (10)$$

where g_t is the gradient at time step, m_t and v_t are estimates of the first and second moments, β_1 and β_2 are decay rates for these moment estimates, \hat{m}_t and \hat{v}_t are bias-corrected estimates, η is the learning rate, and ϵ is a small scalar added to improve numerical stability. Hyperparameter Tuning (Random Search) is indicated in Eq. (11).

$$H^* = \underset{H}{\text{argmax}} f(H | D_{train}, D_{val}) \quad (11)$$

where H^* is the set of optimal hyperparameters found, H represents the hyperparameter space, f denotes the evaluation metric (e.g., AUC), and D_{train}, D_{val} are the training and validation datasets.

III. RESULTS

There are several metrics to evaluate the predictive performance of ML models depending on the context and objective of the model. This study uses hold-out strategy and

assigns 80 percent of the data to the training set and 20 percent to the test set. The imbalance in the dataset, characterized by a lower proportion of patients undergoing surgery compared to those who don't, necessitates metrics that can accurately predict the smaller class. Precision measures the proportion of true positive predictions in the positive class, while recall (sensitivity) assesses the model's ability to identify all relevant cases correctly. In medical predictions, high recall might be more crucial as it is essential to identify as many actual cases (patients needing surgery) as possible [44]. The F1 score is calculated as the harmonic mean of precision and recall, which is two times the product of precision and recall divided by their sum, offering a balanced measure of both metrics. Unlike the arithmetic mean, which averages values directly, the harmonic mean in the F1 score penalizes extreme values more, ensuring a balanced account of both false positives and false negatives in model evaluation. It is beneficial in situations of class imbalance, as it considers both false positives and false negatives, providing a balance between precision and recall. Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is valuable for evaluating the classifier's performance across various threshold settings. It is effective in handling imbalanced datasets as it measures the ability of the model to discriminate between the two classes [45]. Balanced accuracy calculates the average of recall obtained in each class. Balanced accuracy is beneficial for imbalanced datasets as it considers the balance of measurements for each class, providing a more fair comparison when the class distribution is skewed [46]. Mathematical equations for evaluation metrics are listed as Equation (12), (13), (14), and (15).

$$\text{Precision} = \frac{TP}{TP+FP} \quad (12)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (13)$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (14)$$

$$\text{Balanced Accuracy} = \frac{1}{2} \left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP} \right) \quad (15)$$

In ensemble machine learning models, a myriad of configurations exists, each with potential applicability in the domain of cardiac and cardiovascular disease prediction. This paper undertakes a comparative analysis of prevalent ensemble models in the context of cardiovascular disease prediction, introducing a deep learning model specifically developed to address specific weaknesses identified in previous approaches. These weaknesses include the limited ability to handle the imbalanced nature of healthcare data effectively, challenges in capturing complex non-linear relationships inherent in clinical information, and difficulties in managing high-dimensional feature spaces. The proposed

deep learning model is designed to target these challenges, offering a more robust tool for predictive analysis in this critical domain. The efficacy of this model, in comparison to established ensemble models, will be systematically evaluated to determine its relative performance. Ghavidel et al. [11] presented a stacking ensemble model, which, in their study, surpassed traditional classification approaches in performance metrics. Similarly, Ghorbani et al. [16], reported an innovative approach that combines a Genetic Algorithm with a stacking ensemble, incorporating a suite of base classifiers Logistic Regression, Decision Tree, Random Forest, and XGBoost to predict early mortality in intensive care units, particularly within highly imbalanced datasets. Additionally, AdaBoost emerged as the top-performing model in the study by Linda Lapp et al [15]. it showed the best overall performance with an AUC of 0.73. However, in terms of sensitivity (recall) considering the importance of minority class prediction Gradient Boosting Model (GBM) and RF performed better than others. The present study implemented all three models on the data and selected the highest performer which was the Stacking (GLM) model from Linda Lapp's study. Next, we compared the ensemble models used in the study conducted by Talayeh Razzaghi et al. [18] and found that AdaBoost outperformed the others. The performance of these models, when applied to the dataset of the current study, is comprehensively detailed in Table 4. In this study's meticulous quest to ascertain the most effective model for predicting cardiovascular surgical interventions, we evaluated numerous ensemble models against a deep learning (DL) model we developed, which is tailored to address the inherent challenges of such predictions. These ensemble models, denoted as [11], [15], [16], [17], and [18], were rigorously analyzed across multiple performance metrics, as illustrated in Figures 4, 5, and 6, and summarized in Table 4. Model [11], while demonstrating high precision (71%) for the majority class, fell short with a recall of only 27% for the minority class, which translates into a critical deficiency in the context of surgical intervention prediction. The capacity to detect true positives is paramount; thus, a model that misses potential surgical cases is less suitable despite its strength in specificity, as evidenced by its performance for the majority class with a recall of 82%. Model [11] achieved an AUC of 73 and a balanced accuracy of 55%, yet the low recall for the minority class is a significant limitation in a clinical setting. Model [15] offered a more balanced performance for class 1 with precision and recall scores of 63% and 50%, respectively. However, in the context of surgical interventions, the recall rate still implies that half of the critical cases could be overlooked, which is far from optimal. For the majority class, it delivered robust precision and recall, culminating in an AUC of 70 and a

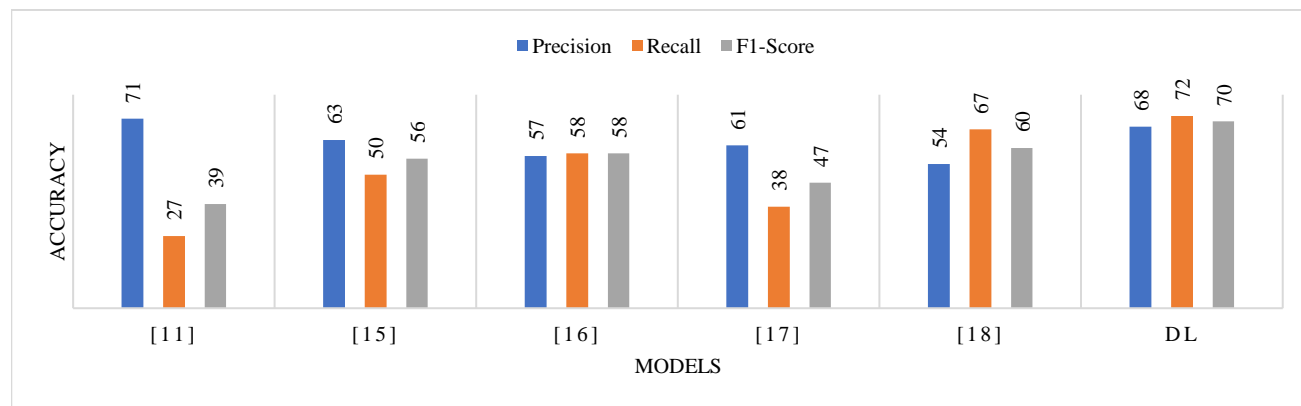


FIGURE 4. Comparison of models' predictive performance on minority class based on precision, recall, and F1-Score.

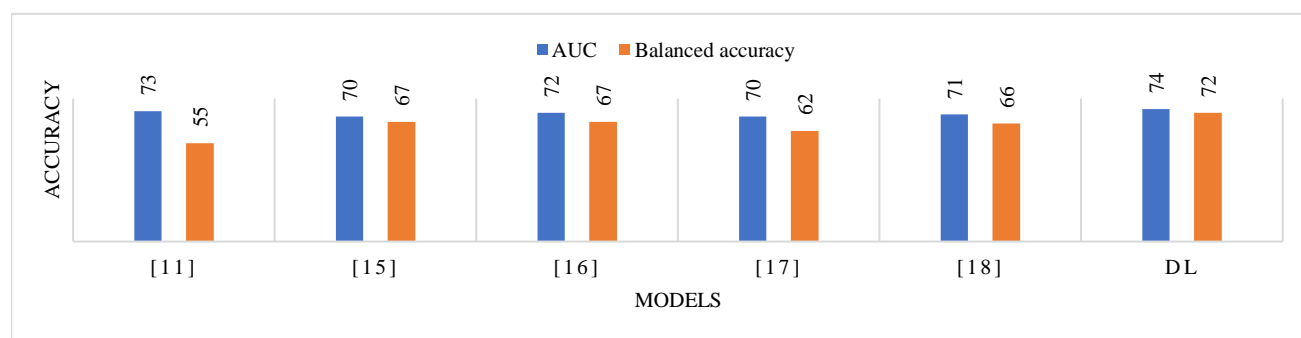


FIGURE 5. Comparison of AUC scores and balanced accuracy of different models on the dataset.

balanced accuracy of 67%. Model [16] presented a near-equilibrium between precision and recall for the minority class, suggesting a harmonious identification process. Nonetheless, the model's modest recall rate of 58% for predicting surgical interventions may lead to a concerning rate of missed diagnoses despite an AUC of 72 and a balanced accuracy of 67%. Model [17] produced a precision of 61% for the minority class, indicating reasonable accuracy, but its lower recall of 38% signifies a considerable risk of missed surgical cases. This model's performance for the majority class was substantially better. However, the emphasis on the minority class in our study's context is critical, and thus, the model's overall utility is diminished. Model [18] showed promise with a recall of 67% for the minority class, suggesting a potent ability to identify a higher proportion of patients requiring surgery. However, its precision of 54% raises concerns about the false positive rate, which, while less critical than false negatives in our study's context, still reflects the need for refinement. With its refined architecture and nuanced learning capabilities, the proposed DL model outperformed the ensemble models by securing the highest AUC score of 74%. It achieved an impressive balance of precision (68%) and recall (72%) for the minority class, signifying a robust capability to detect cases requiring

surgical intervention accurately and comprehensively. This balance is especially crucial in medical diagnostics, where overlooking a single patient in need of surgery could have dire consequences. Moreover, the DL model exhibited solid precision (77%) and recall (71%) for the majority class, underscoring its reliability, and the F1 scores, along with a balanced accuracy of 72%, reinforce its all-encompassing effectiveness. The comparative analysis of the models' effectiveness in classifying the minority class is visually represented in FIGURE 4, while FIGURE 5 provides a comprehensive overview of the overall performance of the models employed. The distinct advantage of the DL model lies in its intricate neural network structure, which effectively captures the complex, nonlinear interactions within the highly dimensional clinical data. The model's hyperparameter optimization, employing Keras Tuner's Random Search, advanced activation functions such as ELU (Exponential Linear Unit), and dropout regularization, contributed significantly to its robustness. Importantly, introducing class weights in the model's loss function counteracted the dataset's imbalanced nature, ensuring an equitable focus on the minority class. This strategic and multifaceted approach resulted in a DL model that excels in predictive accuracy and in delivering equitable performance

across classes, making it a reliable tool for clinical decision-making in cardiovascular care. The performance disparities among the ensemble models and between these models and the proposed DL model underscore the nuanced considerations when choosing an appropriate predictive model for clinical application. Our analysis reveals that while some models may excel in precision, others may prioritize recall, and the DL model's capability to balance these metrics effectively positions it as a superior choice in the context of our study. Based on the calculated confusion matrixes, it can be inferred that the DL model demonstrates superior performance in accurately predicting both positive (surgery) and negative (no surgery) cases compared to the other models [11], [15], [16], [17], and [18] (FIGURE 6).

FIGURE 7 presents the ROC-AUC curves of the three models that achieved the highest AUC metrics, clearly

illustrating the DL model's outperformance relative to the others.

TABLE 5 indicates the result of the Friedman test. The null hypothesis of this test is rejected because the P-value=0 is less than the significance level of 0.05. It means at least one of the models performs differently. This study highlights the significant advantages of Deep Neural Networks (DNNs) in improving the accuracy and efficiency of cardiovascular surgery predictions. The superior performance of DNNs over traditional methods marks a step forward in clinical decision-making, paving the way for their wider application in healthcare. Future research should focus on refining these models and expanding their use to enhance patient outcomes and streamline medical practices. The evolution of machine learning in healthcare promises a new era of personalized and precise medical interventions.

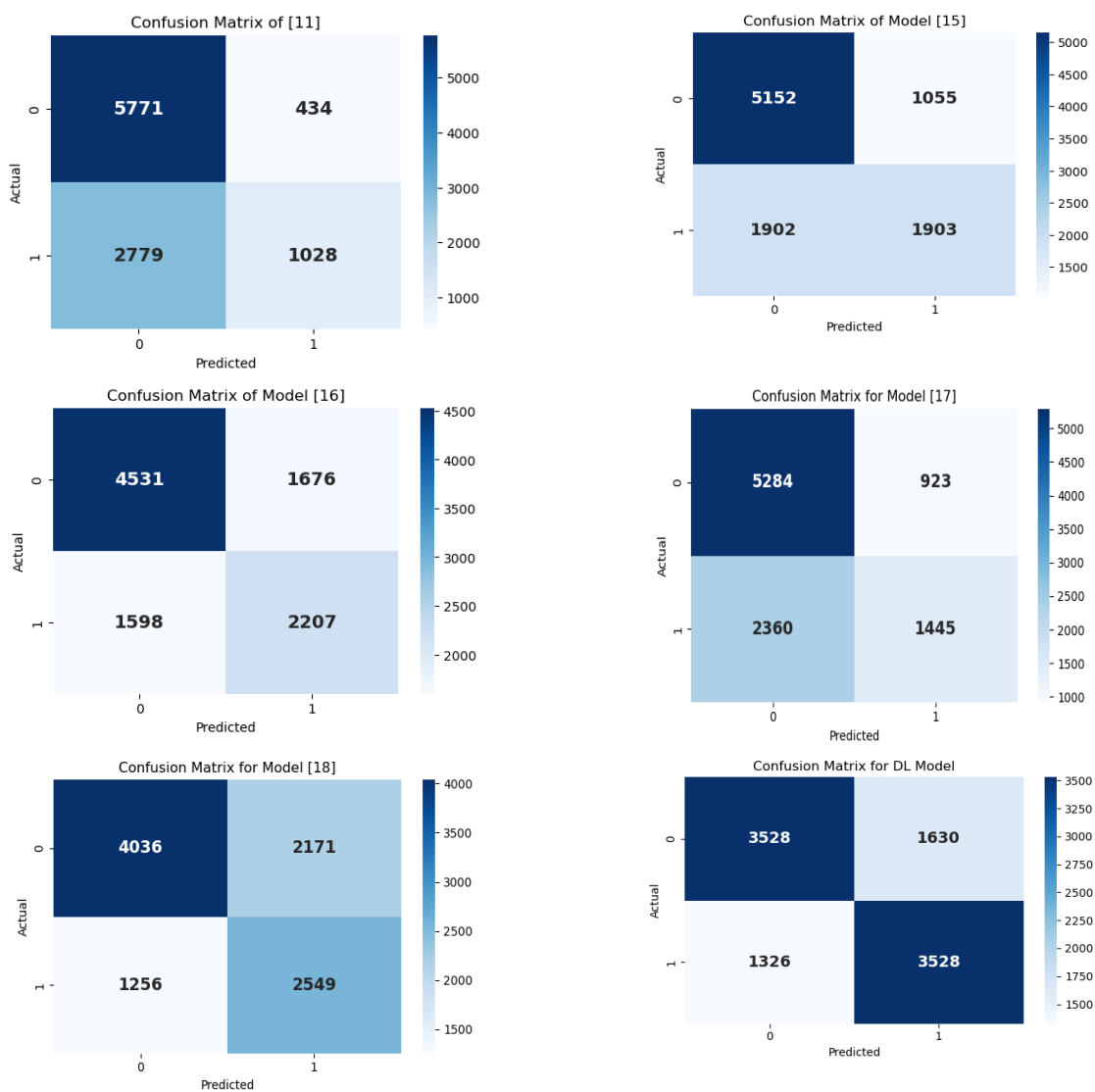


FIGURE 6. Comparison of the models' confusion matrixes

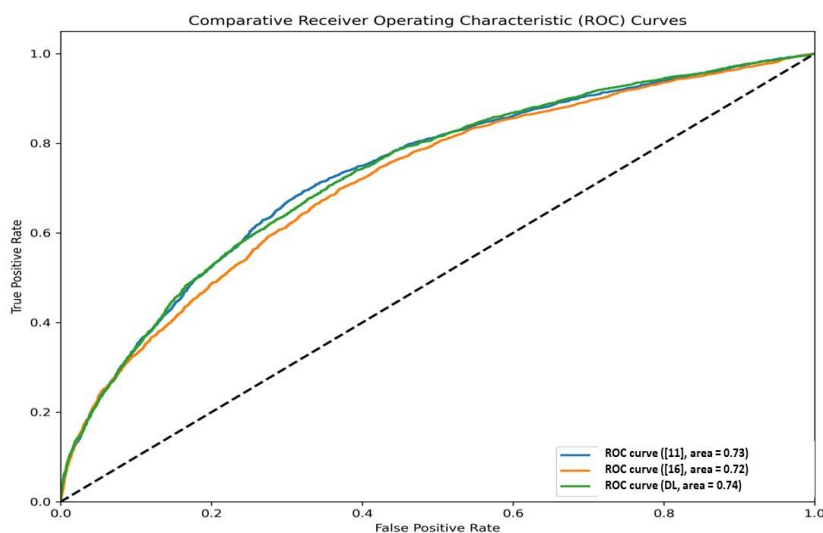


FIGURE 7. Comparison of ROC-AUC curves of the models with highest AUC-scores

IV. DISCUSSION

The present study conducts a detailed comparative analysis of various machine learning models on cardiovascular-related datasets, focusing on the performance of the minority class, a critical aspect in this domain. Our analysis revealed varying degrees of effectiveness among the models in predicting outcomes for the minority class, highlighting the inherent trade-offs in predictive modeling.

The deep learning (DL) approach proposed in this study demonstrated superior performance, exhibiting high precision and recall for the minority class. This balance is particularly significant in medical diagnostics, where the ability to accurately identify true positive cases (high precision) and minimize false negatives (high recall) is paramount. Our DL model's balanced enhancement of these metrics underscores its reliability and effectiveness for clinical decision-making. In contrast to our DL model, ensemble models, including [11], [15], [16], [17], and [18], exhibited notable weaknesses in their performance metrics, particularly for the minority class. Model [11], while precise for the majority class, significantly lacked recall for the minority class, posing a risk of overlooking crucial cases. Model [15] offered a better balance but failed to detect critical surgical cases optimally. Model [16], despite its balanced approach, struggled to capture all positive instances effectively. Model [17] showed moderate precision but a marked deficiency in recall, potentially leading to missed surgical interventions. Model [18], with its higher recall for the minority class, indicated a stronger ability to identify true positives, but its lower precision raised concerns about an

increased rate of false positives. These models, while effective in certain aspects, underscore the complexities and trade-offs inherent in predictive modeling for medical diagnostics.

The study focused on feature selection to improve predictions for cardiovascular surgical interventions. It identified obesity, body composition, and age as top predictive factors for cardiovascular disease (CVD), which is supported by existing medical literature. Other significant predictors include gastroesophageal reflux disease (GERD), dental health, and socioeconomic factors such as family size and employment status. Additionally, blood pressure levels and drug abuse history emerged as important indicators. These findings underscore the complex interplay of medical, lifestyle, and socioeconomic factors in assessing the risk of CVD and the need for surgical intervention. This research underscores the importance of precision and recall in predicting the minority class in cardiovascular datasets. The DL model demonstrates proficiency in accurately identifying true positives (with a recall of 72% for class 1, indicating successful identification of 72% of true positive cases) and effectively minimizing false positives, as evidenced by its precision metric. This balance, alongside the model's high balanced accuracy and AUC score, reinforces its suitability in scenarios that demand a nuanced understanding of both classes. Our DL model effectively navigates these challenges, achieving a superior balance of precision and recall, a critical factor in the high-stakes realm of cardiovascular surgical intervention prediction. This study highlights the significant advantages of Deep Neural Networks (DNNs) in improving the accuracy and efficiency

of cardiovascular surgery predictions. The superior performance of DNNs over traditional methods marks a step forward in clinical decision-making, paving the way for their wider application in healthcare. Future research should focus on refining these models and expanding their use to enhance patient outcomes and streamline medical practices. The evolution of machine learning in healthcare promises a new era of personalized and precise medical interventions.

TABLE 5
 Shapiro normality test result

Friedman test	Degrees of freedom	Chi-square	P-value
7		135	0.00

The study acknowledges limitations regarding the generalizability of findings. The effectiveness of the DL model may vary across different datasets or clinical scenarios, and its performance is inherently tied to the data quality and representativeness. Furthermore, the complexity of the DL model's architecture may pose challenges in certain clinical settings, necessitating ongoing monitoring and recalibration. Another limitation of the present study's predictive modeling approach is its inability to account for the dynamic clinical status of patients. The models generate risk scores based on data from a single time point, overlooking ongoing physiological changes. This static approach may not accurately reflect the evolving nature of a patient's health condition, which is a critical aspect of real-time clinical decision-making. Future enhancements to our models should aim to incorporate dynamic data analysis, allowing for more accurate and timely predictions that align with the rapidly changing clinical scenarios often encountered in healthcare.

V. CONCLUSION

The aim of this study was to conduct a comparative analysis of machine learning models on cardiovascular-related datasets, focusing on predicting outcomes for surgical interventions. Our findings highlight the superior performance of the deep learning (DL) approach, which demonstrates a balanced enhancement of precision and recall for the minority class. Future research should focus on integrating these ML techniques into clinical workflows, incorporating dynamic data analysis for more accurate predictions, and addressing limitations related to data quality and representativeness.

Disclosure of interest:

The authors report there are no competing interests to declare.

Consent to Publish

Consent for publication was obtained from all participants for whom identifying information is included in this article.

Data Availability:

Due to the sensitive nature of the medical data used in this study and in strict compliance with patient confidentiality and privacy regulations, the dataset cannot be publicly shared.

REFERENCES

- [1] Park, J. and P.N. Bonde, *Machine Learning in Cardiac Surgery: Predicting Mortality and Readmission*. Asaio j, 2022. **68**(12): p. 1490-1500.
- [2] Song, Z., et al., *Machine learning in predicting cardiac surgery-associated acute kidney injury: A systemic review and meta-analysis*. Frontiers in Cardiovascular Medicine, 2022. **9**.
- [3] Musikant, D., et al., *Host Epa1 is required for cAMP-mediated invasion by Trypanosoma cruzi*. Molecular and Biochemical Parasitology, 2017. **211**: p. 67-70.
- [4] Fu, M., et al. *An Ensemble Machine Learning Model For the Early Detection of Sepsis From Clinical Data*. in *2019 Computing in Cardiology (CinC)*. 2019.
- [5] Anderson, J.G., et al., *Modeling the costs and outcomes of cardiovascular surgery*. Health Care Management Science, 2002. **5**: p. 103-111.
- [6] Park, D.J., et al., *Development of machine learning model for diagnostic disease prediction based on laboratory tests*. Scientific reports, 2021. **11**(1): p. 7567.
- [7] Vervoort, D., et al., *Cardiac surgery in low-and middle-income countries: a state-of-the-art review*. The Annals of Thoracic Surgery, 2021. **111**(4): p. 1394-1400.
- [8] Vervoort, D., et al., *Global cardiac surgery: access to cardiac surgical care around the world*. The Journal of thoracic and cardiovascular surgery, 2020. **159**(3): p. 987-996. e6.
- [9] Huang, Y., et al., *Application of machine learning in predicting survival outcomes involving real-world data: a scoping review*. BMC Medical Research Methodology, 2023. **23**(1): p. 268.
- [10] Zhou, P.-Y. and A.K.C. Wong, *Explanation and prediction of clinical data with imbalanced class distribution based on pattern discovery and disentanglement*. BMC Medical Informatics and Decision Making, 2021. **21**(1): p. 16.
- [11] Ghavidel, A., R. Ghousi, and A. Atashi, *An ensemble data mining approach to discover medical patterns and provide a system to predict the mortality in the ICU of cardiac surgery based on stacking machine learning method*. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 2023. **11**(4): p. 1316-1326.
- [12] Ghavidel, A., P. Pazos-Lago, and R. Ghousi, *Proceedings of the IISE Annual Conference & Expo 2023*, in *Institute of Industrial and Systems Engineers (IISE)*, B.E. K. Babski-Reeves, D. Hampton, eds., Editor. 2023, IISE: New Orleans, Louisiana.
- [13] Sayilar, B.C. and O. Ceylan. *Grid Search Based Hyperparameter Optimization for Machine Learning Based Non-Intrusive Load Monitoring*. in *2023 58th International Universities Power Engineering Conference (UPEC)*. 2023. IEEE.
- [14] Triwiyanto, T., et al., *How Deep Learning and Neural Networks can Improve Prosthetics and Exoskeletons: A Review of State-of-the-Art Methods and Challenges*. Journal of Electronics, Electromedical Engineering, and Medical Informatics, 2023. **5**(4): p. 277-289.
- [15] Lapp, L., et al. *Evaluation of random forest and ensemble methods at predicting complications following cardiac surgery*. in *Conference on Artificial Intelligence in Medicine in Europe*. 2019. Springer.

- [16] Ghorbani, R., et al., *A New Hybrid Predictive Model to Predict the Early Mortality Risk in Intensive Care Units on a Highly Imbalanced Dataset*. IEEE Access, 2020. **8**: p. 141066-141079.
- [17] Zheng, H., S.W.A. Sherazi, and J.Y. Lee, *A Stacking Ensemble Prediction Model for the Occurrences of Major Adverse Cardiovascular Events in Patients With Acute Coronary Syndrome on Imbalanced Data*. IEEE Access, 2021. **9**: p. 113692-113704.
- [18] Razzaghi, T., et al., *Predictive models for bariatric surgery risks with imbalanced medical datasets*. Annals of Operations Research, 2019. **280**(1): p. 1-18.
- [19] Shams Shirband, S., et al., *A review on deep learning approaches in healthcare systems: Taxonomies, challenges, and open issues*. Journal of Biomedical Informatics, 2021. **113**: p. 103627.
- [20] Kotsiantis, S.B., D. Kanellopoulos, and P.E. Pintelas, *Data preprocessing for supervised learning*. International journal of computer science, 2006. **1**(2): p. 111-117.
- [21] Géron, A., *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. 2022: " O'Reilly Media, Inc."
- [22] Javadnejad, F., et al., *Analyzing incentives and barriers to electric vehicle adoption in the United States*. Environment Systems and Decisions, 2023.
- [23] Iklima, Z., T.M. Kadarina, and R. Priambodo, *Reduction of feature extraction for COVID-19 CXR using depthwise separable convolution network*. Journal of Electronics, Electromedical Engineering, and Medical Informatics, 2022. **4**(4): p. 204-209.
- [24] Zheng, A. and A. Casari, *Feature engineering for machine learning: principles and techniques for data scientists*. 2018: " O'Reilly Media, Inc."
- [25] Basiri, M., F. Javadnejad, and A. Saedi. *Forecasting crude oil price with an artificial neural network model based on a regular pattern for selecting training and testing sets using dynamic command-line functions*. in *24th International Mining Congress & Exhibition of Turkey (IMCET)*. 2015.
- [26] Ghavidel, A. and P. Pazos, *Machine learning (ML) techniques to predict breast cancer in imbalanced datasets: a systematic review*. Journal of Cancer Survivorship, 2023.
- [27] Ansyari, M.R., et al., *Implementation of Random Forest and Extreme Gradient Boosting in the Classification of Heart Disease Using Particle Swarm Optimization Feature Selection*. Journal of Electronics, Electromedical Engineering, and Medical Informatics, 2023. **5**(4): p. 250-260.
- [28] Ghavidel, A., R. Ghousi, and A. Atashi, *An ensemble data mining approach to discover medical patterns and provide a system to predict the mortality in the ICU of cardiac surgery based on stacking machine learning method*. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 2022: p. 1-11.
- [29] Veerman, J.L., et al., *Television viewing time and reduced life expectancy: a life table analysis*. British journal of sports medicine, 2012. **46**(13): p. 927-930.
- [30] Powell-Wiley, T.M., et al., *Obesity and cardiovascular disease: a scientific statement from the American Heart Association*. Circulation, 2021. **143**(21): p. e984-e1010.
- [31] Nugraha, A.T., *Design and build a distance and heart rate monitoring system on a dynamic bike integrated with power generating system*. Journal of Electronics, Electromedical Engineering, and Medical Informatics, 2022. **4**(4): p. 210-215-210-215.
- [32] Lakatta, E.G., *Age-associated cardiovascular changes in health: impact on cardiovascular disease in older persons*. Heart failure reviews, 2002. **7**: p. 29-49.
- [33] Lee, A.L. and R.S. Goldstein, *Gastroesophageal reflux disease in COPD: links and risks*. International journal of chronic obstructive pulmonary disease, 2015: p. 1935-1949.
- [34] Herrera, D., et al., *Association between periodontal diseases and cardiovascular diseases, diabetes and respiratory diseases: Consensus report of the Joint Workshop by the European Federation of Periodontology (EFP) and the European arm of the World Organization of Family Doctors (WONCA Europe)*. Journal of Clinical Periodontology, 2023. **50**(6): p. 819-841.
- [35] Rosengren, A., et al., *Socioeconomic status and risk of cardiovascular disease in 20 low-income, middle-income, and high-income countries: the Prospective Urban Rural Epidemiologic (PURE) study*. The Lancet Global Health, 2019. **7**(6): p. e748-e760.
- [36] Fuchs, F.D. and P.K. Whelton, *High blood pressure and cardiovascular disease*. Hypertension, 2020. **75**(2): p. 285-292.
- [37] Akasaki, Y. and M. Ohishi, *Cerebrovascular and cardiovascular diseases caused by drugs of abuse*. Hypertension Research, 2020. **43**(5): p. 363-371.
- [38] Masoudkabar, F., et al., *Does opium consumption have shared impact on atherosclerotic cardiovascular disease and cancer?* Archives of Iranian Medicine, 2022. **25**(1): p. 50-63.
- [39] Jortveit, J., et al., *Incidence, risk factors and outcome of young patients with myocardial infarction*. Heart, 2020.
- [40] Erlanlita, N., et al., *Implementation of SMOTE and whale optimization algorithm on breast cancer classification using backpropagation*. Journal of Electronics, Electromedical Engineering, and Medical Informatics, 2023. **5**(4): p. 303-313-303-313.
- [41] Chawla, N.V., et al., *SMOTE: synthetic minority over-sampling technique*. Journal of artificial intelligence research, 2002. **16**: p. 321-357.
- [42] He, H., et al. *Adaptive synthetic sampling approach for imbalanced learning*. *IEEE international joint conference on neural networks. in IEEE World Congress On Computational Intelligence*. 2008.
- [43] Mishra, S., *Artificial intelligence: A review of progress and prospects in medicine and healthcare*. Journal of Electronics, Electromedical Engineering, and Medical Informatics, 2022. **4**(1): p. 1-23.
- [44] Davis, J. and M. Goadrich. *The relationship between Precision-Recall and ROC curves*. in *Proceedings of the 23rd international conference on Machine learning*. 2006.
- [45] Fawcett, T., *An introduction to ROC analysis*. Pattern recognition letters, 2006. **27**(8): p. 861-874.
- [46] Carrington, A.M., et al., *Deep ROC analysis and AUC as balanced average accuracy, for improved classifier selection, audit and explanation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022. **45**(1): p. 329-341.

AUTHOR BIOGRAPHY



Arman Ghavidel is currently pursuing his Ph.D. in Engineering Management and Systems Engineering at Old Dominion University, located in Norfolk, VA, USA. He specializes in the intersection of healthcare and technology, and his research is focused

on leveraging Machine Learning, Deep Learning, and Data Mining techniques to address critical challenges in healthcare. Arman is dedicated to advancing the field through innovative research and is actively involved in the academic community, contributing to the development of cutting-edge solutions that promise to transform healthcare delivery and patient outcomes.



Pilar Pazos is a Professor in the Department of Engineering Management and Systems Engineering at Old Dominion University. She received a B.S. and M.S. Industrial Engineering from the University of Vigo, Spain, a M.S. in Systems Engineering, and PhD in Industrial Engineering from Texas Tech University. She has published

studies in process improvement, modeling and simulation, and data analytics in health care.



Dr. Rolando Del Aguila Suarez is a distinguished professional with a Ph.D. in Healthcare Administration and a Master of Nursing, among other qualifications. He has worked as a biomedical researcher, healthcare administrator, and nurse, and has

extensive experience managing medical and educational facilities. Currently, he serves as an Assistant Professor at Hampton University School of Nursing, where he integrates his vast experience into his teaching, with a focus on promoting mental health awareness in the community.



ALIREZA ATASHI is the Head of the Cancer Informatics Department at Motamed Cancer Institute, Tehran, Iran, and an Assistant Professor in medical informatics at the e-Health Department at Tehran University of Medical Sciences. His research interests include medical data mining, clinical informatics, patient registries,

and telemedicine, especially in cancer and intensive care fields, where he has done several studies.