A Novel Hybrid Model Combining Feature Selection and Imbalance Handling for Prediction of Heart Failure Survival

Wial Hanon*[ID], Hasanain Ali Al Essa[ID], Suhad Hatim Jihad[ID]

College of Information Technology, University of Babylon, Babylon 4000, Iraq

Corresponding Author Email: wailh@uobabylon.edu.iq

ABSTRACT

The imbalance and irrelevant or redundant features often hinder the performance of accurate prediction models. This can aid healthcare professionals in the early detection and intervention of heart failure (HF). HF is a critical cardiovascular condition that poses a significant risk to human health, frequently resulting in high mortality rates if not diagnosed and managed early. In this paper, we propose an enhanced prediction approach for HF survivors based on a hybrid model (hybrid HF) that integrates feature selection techniques with the resampling technique—synthetic minority oversampling technique combined with edited nearest neighbors (SMOTEENN). SMOTEENN effectively addresses the class imbalance problem by synthesizing new minority class instances and improving noisy samples after selecting the most relevant features from the data, thereby enhancing the quality of the training data. Additionally, applied feature selection to identify the most relevant predictors, reducing model complexity and enhancing interpretability. Experimental results demonstrate that our hybrid HF model outperforms existing methods by uniquely integrating SMOTEENN with feature selection to achieve 0.9315 accuracy in predicting a heart patient's survival, improving model accuracy by 8% compared to the baseline methods for the RF algorithm. Finally, improving classification, sensitivity, and overall model robustness compared to the baseline method.

1. INTRODUCTION

Heart failure (HF), also known as congestive heart failure (CHF), is a chronic condition in which the heart cannot pump blood effectively to meet the body's needs [1]. To prevent cardiovascular disease (CVD), it is essential to recognize that it is caused by multiple contributing factors, including excessive blood pressure, cholesterol, irregular pulse rate, and others; distinguishing between heart disease and CVD can be challenging [2]. Numerous approaches have been investigated by researchers to anticipate heart illnesses; however, early disease prediction is ineffective for a variety of reasons, such as approach complexity, execution time, and accuracy [3, 4]. Therefore, several lives can be saved by appropriate diagnosis and treatment [5].

Many data-driven and method-driven challenges frequently confront machine learning (ML) [6]. On the other hand, it is characterized by its robustness and reliability in real-world settings, such as patient self-assessment or as an aid for clinical support and decision-making [7].

The main challenge in predicting critical cases or survivors of heart failure is the imbalance in data and the bias of most models toward the majority class [8]. This underrepresented dataset, particularly relevant in classification issues. ENN stands for Edited Nearest Neighbors, while SMOTE stands for Synthetic Minority Over-sampling Technique [9]. By interpolating between current minority samples and their

closest neighbors, SMOTE creates artificial examples of the minority class. ENN reduces noise and borderline examples by eliminating examples that are incorrectly classified by their k-nearest neighbors [10].

Ensemble methods significantly improve the performance of conventional ML classifiers when applied to class imbalance techniques, and ensemble ML classifiers to predict HF [11]. Noting a set of benefits of medical dataset balancing [12-15]:

Keeping the distribution of classes balanced and assisting models in discovering patterns associated with both outcomes (survival or death).

By balancing the data, the minority class's models perform better and can produce more accurate forecasts.

1. For the minority class, which is frequently the more clinically significant class in healthcare settings, balancing improves memory, precision, F1-score, and Accuracy.
2. Improves the model capacity to identify patients who are danger of dying by lowering bias toward the majority class and assisting in treating both classes more fairly.
3. Improves generalization instead of memorization of particular instance. Because it offers a wider variety of training samples, this is particularly useful for small dataset.
4. Prevents customization by generating new synthetic instances based on interpolation. As opposed to simple

oversampling.

This paper highlights a critical research hiatus in developing an integrated, clinically viable framework that balances data enhancement and feature reduction for more accurate and practical HF survival predication. Several studies have explored the use of either SMOTE for class imbalance correction or feature selection techniques or enhance model interpretability and performance [16-18]. Moreover, existing methods often neglect clinical constraints such as data sparsity, limited feature availability, and the need for explainable outcomes in medical decision-making [19-21].

Most existing studies apply these techniques sequentially without optimizing their interaction, potentially leading to overfitting, information leakage, or suboptimal feature relevance [22]. This lack of a unified, clinically aware framework restricts the development of robust and generalizable models for HF survival under real-word clinical constrain.

This paper, we propose a novel hybrid model (hybrid HF) that combines feature selection MLs and an imbalance technique (SMOTEENN) to enhance predictive data from the training dataset, and then trains the model to obtain the best accuracy for disease survivors. The paper demonstrate that ensemble methods are becoming increasingly state-of-the-art solutions for addressing multiple challenges encountered with ML algorithms, including overfitting, computational intensity, underfitting, and representation.

This hybrid model objectives to enhancing predictive model performance, lessening overfitting, and better feature selection strategies, addressing the class imbalance, and ensuring that only the most pertinent features are used for model training. As a result, achieving more balanced and effective ML results from this model's efficient handling of both class imbalance and irrelevant features [23].

To demonstrate its effectiveness, we employ nine ML models: Random Forest (RF), Extra Tree Classifier (ETC), Extreme Gradient Boosting (XGBoost), Gaussian Naive Bayes classifier (G-NB), Decision Tree (DT), Logistic Regression (LR), Gradient Boosting classifier (GBM), K-nearest neighbor (KNN), and Support Vector Machine (SVM) [11, 24, 25].

2. RELATED WORKS

In this section, review a group of studies that have contributed to improving heart failure prediction, as well as the most important techniques and algorithms used in the prediction process. Most studies have not extensively discussed the misclassification of inliers vs. outliers, which could introduce biases and degrade model performance. Additionally, heart failure datasets are often imbalanced, which leads to a consequence of lower recall for the minority, which is often more important. Reduced robustness when deployed on heterogeneous or real patient populations because the model may become too specialized in detecting patterns in "clean" or ideal data, and generalize poorly to more complex, noisy, or real-world clinical data.

Improvements to the conventional RF method were proposed to enhance heart failure prognostics [26]. Particularly in clinical datasets where outliers are prevalent, traditional RF model's sensitivity to outliers might reduce their predictive accuracy. The author used the funnel technique, which proactively detects and eliminates outliers before training. Training performance on benchmark medical

datasets improved after this funnel technique was put into practices. Especially when used in patients' health records for heart failure. According to some results the funnel RF approach provides improved, explain ability and performance, which makes it a useful tool for clinical decision support in predicting the prognosis of heart failure.

The study [22] focused on applying ML techniques to improve the predication accuracy of patient survival in HF. By using SMOTE, it tackles the problem of class imbalance, where survivors greatly outnumber non-survivors. To balance the dataset and enhance ML model performance, SMOTE create fake samples for the minority class. The research concludes that by precisely identifying patients with high-risk heart failure, integrating SMOTE with strong ML can improve clinical decision-making.

ML techniques were employed to analyze and predict heart failure in patients [27]. In order to determine the best predictors of HF outcome. Uses a number of models, including LR, DT, RF, KNN, and SVM, using a clinical dataset that includes patient health metrics (e.g. age blood pressure, serum creatinine, and ejection fraction). Performance parameters such as accuracy, precision, recall, and F1-score are used to assess the models. The findings show that when compared to individual classifiers, ensemble models such as RF performed better in terms of prediction.

Sophisticated ML algorithms were explored for CVD risk detection and control [28]. Assesses the effectiveness of several MLs, such as RF, SVM, Gradient boosting, and neural networks, in identifying high-risk individuals based on variables by utilizing extensive clinical dataset. Shows that in terms of accuracy, sensitivity, and specificity, ML models perform better than conventional risk-scouring methods. Furthermore, to improve clinical trust and applicability, interpretability technique like SHAP (Shapley Additive explanations) are used to offer insights into model judgments.

A hybrid ML algorithm combining two methods was proposed to enhance lung cancer prediction accuracy [29]. The most pertinent feature is chosen from high-dimensional clinical datasets using Recursive Feature Elimination (RFE) in conjunction with SVM. Second, the Extreme Gradient Boosting (XGBoost) classifier, a derivative-free technique useful for navigating intricate search spaces, is employed for prediction in this step, which also lowers noise and enhances model interpretability and performance. In comparison to conventional techniques, the suggested model greatly increases predication accuracy by fusing a robust tuned classifier with strong feature selection. The method is performing better in terms of F1-score, recall, accuracy, and precision.

An innovative XGBoost-based approach was proposed to predict stroke risk, aiming to improve early detection and prevention [30]. With focus on the importance of ML in medical diagnostics, the outcomes show that XGBoost is a dependable technique for assessing stroke risk since it produces high predictive performance.

3. THE PROPOSED HYBRID HF MODEL

This section details the methodologies employed in the study, providing an in-depth exposition of the proposed framework for breast cancer diagnosis and prediction. The performance of the Hybrid HF model is rigorously evaluated using standard metrics, including Precision, Accuracy, Recall and F1-score.

3.1 Proposed model

The hybrid HF model is enhanced prediction for survivors from heart failure disease by integrating feature selection techniques with resampling technique (SMOTEENN) to create a novel hybrid HF model. Applying SMOTEENN after feature selection to create highly correlated synthetic samples for the minority class. Additionally, feature selection first guarantees the model-selected features are balanced, have a high impact on the accuracy of the results, and robustly correlated with each other.

To balance the class, the hybrid HF model applies SMOTEENN to the smaller dataset after performing features selection to eliminate superfluous or irrelevant features. The ML model should be trained. The hybrid HF model following benefits:
1. Lowers computing costs by removing extraneous features before creating synthetic samples.
2. Reduces the possibility of adding noisy by assisting SMOTEENN in working on only the most essential elements.
3. Prevents the overfitting and improves generalization.

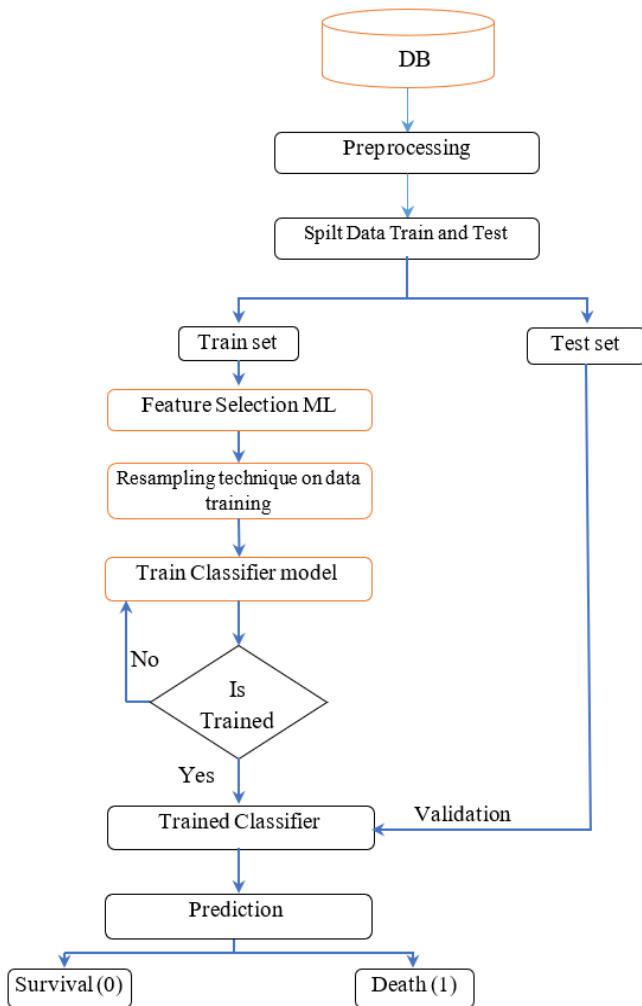The flowchart of the proposed methodology is presented in Figure 1.

Figure 1. The hybrid HF model

3.2 Description of the methodology

Class imbalance is a dataset that contains significantly fewer samples than the rest of the data. The effect on the model is a bias toward the majority of the data classes. The model is biased toward predicting the majority class most of the time, achieving high accuracy, and showing poor performance on the minority class, which is often the most important to detect.

In feature selection issues, the number of possible feature subsets increases exponentially as the number of features increases. In addition, there are numerous problems with feature selection. Therefore, it is not feasible to carry out a thorough search to find the optimal answer, even with low-dimensional data.

SMOTEENN is chosen over other resampling techniques such as SMOTE-Tomek, due to its enhanced ability to address both class imbalance and noise in the clinical dataset. While SMOTE generates synthetic minority class samples to balance the dataset, ENN goes further by removing ambiguous or mislabeled instances from both classes, which helps reduce overlapping and improves class separation [31]. In contrast, SMOTE-Tomek focuses on eliminating overlapping samples but retains noisy data points that may negatively impact model performance [32]. Therefore, SMOTEENN provides a more aggressive and effective cleansing mechanism, leading to improved classification accuracy and robustness in real-world clinical environments [9].

Predictive mode performance can be enhanced by integrating feature selection MLs with SMOTEENN, particularly when working with unbalanced datasets. In order to train a classifier on the resampled data, we first apply feature selection to the training data, then SMOTEENN to the training data, and evaluate on the unaltered test set. To prevent information from the synthetic sample from leaking and to make sure that irrelevant feature selection is completed before to resampling.

4. EXPERIMENTATION

4.1 Dataset set overview

This paper utilizes the Heart Failure Clinical Records dataset (HFCR), is a widely used dataset for predicting mortality in patients with HF. The dataset includes 233 patients' records with a mean age of 69.5 years [5, 22, 26]. Among the patients, 156 (67%) are alive (class 0) and 77 (33%) are deceased (class 1). The overview of the data is illustrated in Table 1.

4.2 Evaluation matrices

There are various techniques for assessing an ML model's [33]. It is anticipated that the combination of several assessment instruments will support the advancement of analytical research [7], which will compare ML based on algorithms using four fundamentals measures accuracy (ac) Eq. (1), precision (pr) Eq. (2), recall (re) Eq. (3), and F1-Score (fs) Eq. (4). We compute all four measures with the use of a confusion matrix [4, 34]. The most important predication is whether the data is connected to a medical false negative.

$$ac = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$pr = \frac{TP}{TP + FP} \quad (2)$$

$$re = \frac{TP}{TP + FN} \quad (3)$$

$$fs = \frac{2 \times TP}{2TP + FP + FN} \quad (4)$$

Table 1. HFCR dataset overview

Column Name	Description Column
Age (years)	Age of the patient
Anaemia (binary:0=no,1=yes)	Decrease of red blood cells or hemoglobin
Creatinine_phosphokinase CPK mcg/L	Level of the CPK enzyme in the blood
Diabetes (binary)	If the patient has diabetes (1=Yes,0=No)
Ejection_fraction (percentage each time)	Percentage of blood leaving the heart at each contraction
High blood pressure (binary)	If the patient has hypertension (1=Yes,0=No)
Platelets (kilo platelets/mL)	Platelets count in blood
Serum creatinine (mg/dL)	Level of serum creatinine in the blood
Serum sodium (mEq/L)	Level of serum sodium in the blood
Sex (binary)	Gender of the patient (1=Male, 0=Female)
Smoking (binary)	Whether the patient smokes (1=Yes,0=No)
Time (follow-up in day)	Follow-up period (how long the patient was observed)
Death event (target)	Whether the patient died during the follow-up (1=Yes,0=No)

5. WORKING OF THE PROPOSED METHOD

The hybrid HF framework decreased the initial feature set from original number to reduced number crucial predictors by integrating feature selection-based methods (such as RF, ETC, DT, and mutual information) with class imbalance techniques. Both clinical markers (e.g., time, ejection fraction, serum creatinine, serum sodium, age, platelets, and creatinine phosphokinase) were included in the characteristics that were chosen, highlighting their combined influence on prediction. The nine ML algorithms- RF, ETC, KNN, SVM, XGBoost, DT, LR, GBM, and G-NB—that were trained and assessed are used in the hybrid HF model. Other models were routinely outperformed by RF.

5.1 Standard scaler

In ML, the Standard Scaler is a data preprocessing tool that normalizes features (columns) by scaling to unit variance and eliminating the mean. Make use of it to improve algorithm performance, accelerate convergence, and prevent bias. Standardization guarantees that every feature contributes equally and that no feature is dominant just due to its size.

5.2 Feature selection

The process of choosing the variables required to improve accuracy is known as feature selection. To explain the hidden patterns found within the dataset, data visualization can be used. Using a set of feature selection ML algorithms, the best dataset to be used in the data imbalance process is selected.

Figure 2 illustrates the steps in the feature selection process from a set of algorithms. This will be used in the next stage to remove noise and balance the minority data classes present in the medical data.
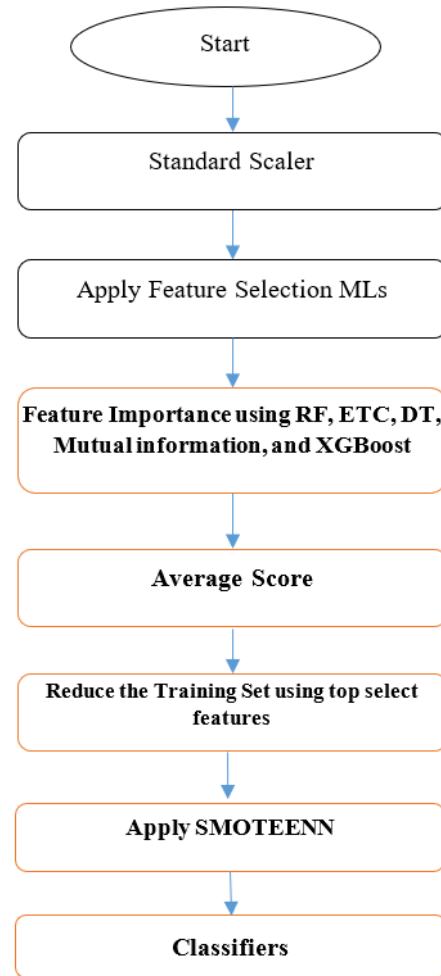

Figure 2. The feature importance models

5.3 Feature importance and average score

The flowchart focuses on enhancing model performance and efficiency by offering an organized method for feature selection ML algorithms. The procedure finds and keeps just the most important features by using a variety of importance algorithms (such as RF, ETC, DT, Mutual Information, and XGBoost). As a result, a fine tuned dataset with carefully chosen features is produces, which can increase model accuracy, decrease overfitting, and boost computing efficiency in later training and forecasting.

5.4 Reduce training set

This step, which follows the use of several ML feature selection techniques, is essential to maximizing the performance of the suggested model. Sub setting the training dataset to only contain those chosen top features is what this step entails. The step is critical. By removing superfluous or redundant features, it lessens overfitting, enhances model performance, and expedites training. Reduced features improve interpretability and speed up training.

Acquisition of a more qualitative understanding of the dataset is facilitated by visualizing its features. To implement feature ranking, use ETC. Table 2 shows the expected feature

significance of the ETC. Time, ejection fraction, serum creatinine, serum sodium, age, platelets, and creatinine phosphokinase are the most crucial factors, according to ETC.

ETC is selected for feature importance analysis due to its computational efficiency and effectiveness in handling high-dimensional clinical data. ETC, as an ensemble of randomized decision trees, provides robust, impurity-based feature rankings by averaging across multiple models, making it less sensitive to variance and more stable than single-tree approaches [35]. While SHAP values offer advanced interpretability and local feature contributions, they are computationally intensive, especially on large datasets and may not be practical for rapid experimentation or resource-constrained clinical environments [36]. Therefore, we preferred ETC because a scalable, fast, and reliable method to identify the most relevant features for HF survival prediction [37].

Table 2. Feature significance of the ETC

No.	Column Name	Mean Score
1	time	1
2	Ejection fraction	0.476161
3	Serum creatinine	0.402527
4	Serum sodium	0.252218
5	age	0.312342
6	platelets	0.240603
7	Creatinine phosphokinase	0.250505

5.5 Handling class imbalance

After extracting the most closely related classes by applying the algorithm ETC, the data with high convergence will be used to remove any noise or imbalance in the classes.

Figure 3. The imbalanced and balanced data

The class imbalance using SMOTEENN technique, a prevalent problem in medical dataset, assessed for the training set after the feature selection procedure on the HFCR Dataset was finished. Applied SMOTEENN solely to the training set to solve this problem and enhance the performance of the prediction models. Applying SMOTEEN after feature selection, the structure of the chosen feature space preserved, and the resampling procedure is guide by the most pertinent variables. To assist in the creation of more reliable and broadly applicable models, this stage contributes to the creation of a training set that is cleaner and more balanced. Crucially, to avoid data and preserve the integrity of the assessment procedure on the test set, only used SMOTEENN on the training data. Figure 3 illustrates the features before and after applying SMOTEEN technique. Where one (1) represents death and zero (0) survivors of the HF.

5.6 Classifiers

Training and assessing several classification algorithms to predict HF after using feature selection to preserve the most informative variables and using the SMOTEENN technique to solve class imbalance. By clearing overlapping majority class instances and synthesizing minority class case, SMOTEENN produced a balanced dataset, while the chosen features guaranteed decreased model complexity and enhanced generalization. RF, ETC, KNN, SVM, XGBoost, DT, LR, GBM, and G-NB were among the classifiers they tested with. To provide reliable performance estimation, stratified cross-validation was used to train each classifier on the processed dataset. Accuracy, precision, recall, and F1-score are used to evaluation measures; accuracy received special emphasis because it is crucial to identify cases of HF.
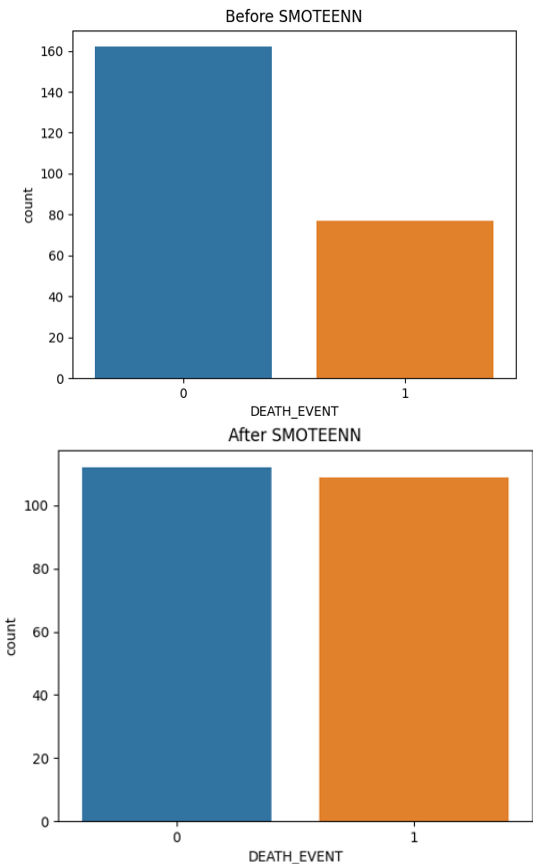
6. DISCUSSION OF RESULTS

In this section, every experiment's design and outcomes are examined to forecast the survival of heat patients. The results with a significant set of features are shown after the results with the full set of features. Thirteen feature of clinical and bodily features are included in the dataset. Anemia, diabetes, blood pressure, smoking, and gender are a few examples of these binary characteristics. In the binary classification task, which determines whether a patient lived or passed away before 130 days of the follow-up period, the death event feature is used as the target class. To balance the dataset, SMOTEEN is used. The balanced dataset is used to train set not on the test set ML models, which were then assessed for accuracy, precision, recall, and F-score. We conducted the first experiments using all the features available in the database HFCR, and the results are shown in Table 3.

The HFCR dataset, while valuable for studying heart survival, is prone to biases such as small size, demographic imbalance, and class skew. Without proper handling, these factors can impair the model's clinical relevance and fairness. Integrating robust validation techniques, applying data augmentation techniques SMOTEENN, to balance minority classes can help mitigate these biases and support more reliable, equitable predictive modeling.

With an accuracy of 88% XGBoost outperformed the other models under evaluation, demonstrating its superior ability to handle the dataset with all of features. Its ensemble nature, regularization methods, and capacity to identify intricate

patterns are probably the cause of this. In the second stage, the best feature selection was applied, and seven distinct and highly convergent features were selected, to which the data-balancing SMOTEENN technique will be applied. The results are shown in Table 3.

After improving their performance with a suggested model and choosing the top seven features, a variety of ML algorithms assessed in this work. The outcome shows how well the suggested fractures selection works to enhance A hybrid HF model performance. With 93% accuracy, RF outperformed all other tested models. ETC came in second with 92% accuracy. The most dependable and well-balanced classifiers in this situation were both models, which also showed excellent precision, recall, and F1-score.

Table 3. The hybrid HF and select features

| ML Name | Seven Features | | | | All Features |
	ac%	pr%	re%	fs%	ac%
RF	**93**	**94**	**91**	**91**	83
ETC	92	93	93	93	85
KNN	75	75	75	75	68
SVM	83	84	83	83	80
XGBoost	86	87	87	86	**88**
DT	81	81	81	81	75
LR	81	81	82	81	75
GBM	86	86	87	86	85
G-NB	83	83	83	82	80

To statistically compare the performance of using seven features vs. All features across multiple classifiers (Table 3), we can apply paired statistical tests to evaluate metrics accuracy (ac %). Then calculate the p-value using the ac of seven features and all features. The steps calculate the p-value by finding the difference in accuracy:
1. Calculate the differences for each ac in the seven and all features.
2. Compute the mean and standard deviation of differences.
3. Calculate the t-statistic.
4. Calculate the p-value.

The p-value is approximately 0.0046, which is less than 0.05, indicating that the improvement in accuracy using the selected features is statistically significant.

The hybrid HF model's application SMOTEENN shows better or equivalent outcomes to the top-performing techniques in the earlier references, especially for ensemble models like RF, ETC, and XGBoost. Overall performance is improved, whereas KNN and LR are less successful, demonstrating the value of SMOTEENN for unbalanced HFCR dataset. The compare is shown in Table 4.

In medical diagnosis, the metric accuracy is most important from training time (TR-time) because the HFCR dataset is used for predicting death events in heart failure patients, a highly complex medical prediction problem. Moreover, accuracy is critical, especially for correctly identifying patients at risk. Impact on the life of patients, where a small drop in accuracy could mean missing a patient who might need life-saving intervention.

Table 4. Compare the hybrid HF with the other references on the same datasets and ML

ML	[22]	[38]	[39]	[40]	[27]	A Hybrid HF	TR-Time
RF	91		88		81.66	**93**	0.43
ETC	92			91.62		**92**	0.29
KNN		83.50	81.11		68.88	**75**	0.0
SVM	76.22				81.97	**83**	0.0
XGBoost	88				83.92	**90**	0.17
DT	87.33		86		85.33	**86**	0.008
LR	84.28				85.28	**81**	0.026
G-NB	75		77			**81**	0.009
GBM	88					**86**	0.17

RF performed the best. It benefits ensemble learning by averaging multiple decision trees, reducing overfitting. However, this comes at the cost of longer training time, as several trees are built and evaluated.

ETC is similar to RF but with more randomness. Slightly less accurate than RF and marginally faster. Well-performing with moderate cost.

The choosing model with 93% accuracy for RF is better than ETC with 92% accuracy, but faster training time when screening patients. SVM (underperforms due to the default hyper parameters) and KNN (doesn't train, it stores data) may need tuning or are less suited to this dataset.

The hybrid HF model outperforms previous references in general, with an emphasis on the differences in the performance amongst models such as RF, ETC, KNN, etc. A hybrid data balancing method eliminates noisy and borderline cases while increasing the minority class. By managing class imbalance more skillfully than each approach alone, this technique enhances classifier performance.

Traditional classifiers like KNN, SVM, and G-NB were continuously surpassed by ensemble-based models like RF, ETC, XGBoost, and GBM.

This highlights how effective ensemble learning is at managing intricate feature interaction and minimizing overfitting, especially when paired with better feature selection. Figure 4 illustrates the accuracy for ML algorithms with applying the SMOTEEN technique.
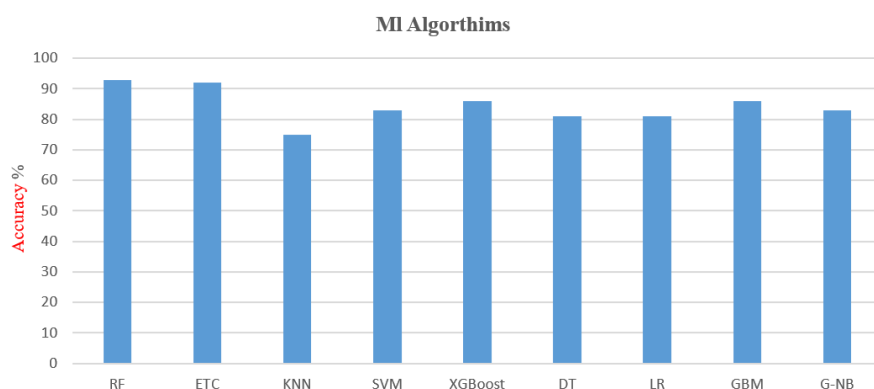

Figure 4. The better ML after apply a hybrid HF model

The performance of multiple classifiers was considerably enhanced by the proposed model and feature selection technique. The best-performing algorithms were RF, ETC, which makes them excellent choices for use in real-world applications that demand high classification accuracy and balanced performance across all assessment variables.

7. CONCLUSION

Integrating ML algorithms into CVS risk assessment will greatly enhance early diagnosis, customize treatment plans, and eventually lessen the prevalence of heart disease worldwide. This paper enhances the prediction of heart failure survival by a hybrid HF model that applies feature selection techniques (RF achieving 93% accuracy) followed by SMOTEENN to address class imbalance using only seven selected features, and demonstrates strong potential to support clinical decisions by minimizing false negatives in high-risk heart failure patients.

Feature selection helped in reducing dimensionality and improving model performance by identifying the most relevant clinical indicators contributing to the heart failure outcome.

The application of SMOTEENN further improved the predictive accuracy by balancing the dataset, combining oversampling of the minority class (survivors or not), and cleaning noisy or misclassified samples. This dual strategy led to more robust and generalizable models, as evidenced by improved evaluation metrics such as accuracy, precision, recall, and F1-score. Overall, the integration of feature selection with SMOTEENN proves to be an effective preprocessing pipeline for developing predictive models in medical datasets, particularly those with imbalanced classes like heart failure survival data. This model can support clinicians in risk stratification and decision-making, ultimately contributing to better patient outcomes. Reducing the misclassification, particularly false negatives, can directly translate to improved patient outcomes, fewer hospital readmissions, and more informed care planning, which is critical in time-sensitive cardiovascular cases.

Ultimately, its application remains limited, particularly in clinically sensitive domains such as heart survival prediction, performance on multi-center datasets remains untested.

REFERENCES

[1] Saqib, M., Perswani, P., Muneem, A., Mumtaz, H., Neha, F., Ali, S., Tabassum, S. (2024). Machine learning in heart failure diagnosis, prediction, and prognosis. Annals of Medicine and Surgery, 86(6): 3615-3623. https://doi.org/10.1097/MS9.0000000000002138

[2] Dimitriadis, N., Tsiampalis, T., Arnaoutis, G., Tambalis, K.D., Damigou, E., et al. (2024). Longitudinal trends in physical activity levels and lifetime cardiovascular disease risk: Insights from the ATTICA cohort study (2002-2022). Journal of Preventive Medicine and Hygiene, 65(2): E134. https://doi.org/10.15167/2421-4248/jpmh2024.65.2.3243

[3] Shenoy, M., Chapman, C.B., Nawaz, M.Z., Sweitzer, N. (2006). Diagnosis and management of stage a heart failure. Congestive Heart Failure, 12(3): 146-152. https://doi.org/10.1111/j.1527-5299.2006.04621.x

[4] Hanon, W., Salman, M.A. (2024). Smart controller integrated with MQTT broker based on machina learning techniques. Journal Européen des Systèmes Automatisés, 57(1): 87-94. https://doi.org/10.18280/jesa.570109

[5] Budi Susilo, Y.K., Abdul Rahman, S., Amgain, K., Yuliana, D. (2025). Artificial intelligence in heart attack research: A bibliometric analysis of trends, innovations, and future directions. SSRN. http://doi.org/10.2139/ssrn.5177991

[6] Al Essa, H.A., Hanon, W., Wotaifi, T.A., Raheem, A.K.A. (2025). Associative memory for recognition and translating American sign language. Ingénierie des Systèmes d'Information, 30(3): 703-711. https://doi.org/10.18280/isi.300314

[7] Hanon, W., Salman, M.A. (2024). Integration of ML techniques for early detection of breast cancer: Dimensionality reduction approach. Ingénierie des Systèmes d'Information, 29(1): 347-353. https://doi.org/10.18280/isi.290134

[8] Lindenfeld, J., Costanzo, M.R., Zile, M.R., Ducharme, A., Troughton, R., et al. (2024). Implantable hemodynamic monitors improve survival in patients with heart failure and reduced ejection fraction. Journal of the American College of Cardiology, 83(6): 682-694. https://doi.org/10.1016/j.jacc.2023.11.030

[9] Mbunge, E., Sibiya, M.N., Takavarasha, S., Millham, R.C., Chemhaka, G., Muchemwa, B., Dzinamarira, T. (2023). Implementation of ensemble machine learning classifiers to predict diarrhoea with SMOTEENN, SMOTE, and SMOTETomek class imbalance approaches. In 2023 Conference on Information Communications Technology and Society (ICTAS), Durban, South Africa, pp. 1-6. https://doi.org/10.1109/ICTAS56421.2023.10082744

[10] Pandit, R., Pandit, T., Goyal, L., Ajmera, K. (2022). A review of national level guidelines for risk management of cardiovascular and diabetic disease. Cureus, 14(6): e26458. https://doi.org/10.7759/cureus.26458

[11] Chaurasia, V., Chaurasia, A. (2023). Novel method of characterization of heart disease prediction using sequential feature selection-based ensemble technique. Biomedical Materials & Devices, 1(2): 932-941. https://doi.org/10.1007/s44174-022-00060-x

[12] Fernando, C.D., Weerasinghe, P.T., Walgampaya, C.K. (2022). Heart disease risk identification using machine learning techniques for a highly imbalanced dataset: A comparative study. KDU Journal of Multidisciplinary Studies, 4(2): 43-55. https://doi.org/10.4038/kjms.v4i2.50

[13] Wongvorachan, T., He, S., Bulut, O. (2023). A comparison of undersampling, oversampling, and SMOTE methods for dealing with imbalanced classification in educational data mining. Information, 14(1): 54. https://doi.org/10.3390/info14010054

[14] Omotehinwa, T.O., Oyewola, D.O., Dada, E.G. (2023). A light gradient-boosting machine algorithm with tree-structured parzen estimator for breast cancer diagnosis. Healthcare Analytics, 4: 100218. https://doi.org/10.1016/j.health.2023.100218

[15] Rawat, S.S., Mishra, A.K., Motwani, D. (2024). Forecasting the survival and mortality of patients by machine learning trained on heart failure clinical imbalanced data. In 2024 IEEE International Conference on Interdisciplinary Approaches in Technology and

Management for Social Innovation (IATMSI), Gwalior, India, pp. 1-6. https://doi.org/10.1109/IATMSI60426.2024.10502741

[16] Azad, C., Bhushan, B., Sharma, R., Shankar, A., Singh, K.K., Khamparia, A. (2022). Prediction model using SMOTE, genetic algorithm and decision tree (PMSGD) for classification of diabetes mellitus. Multimedia Systems, 28(4): 1289-1307. https://doi.org/10.1007/s00530-021-00817-2

[17] Zheng, Y., Luo, J., Chen, J., Chen, Z., Shang, P. (2023). Natural gas spot price prediction research under the background of Russia-Ukraine conflict-based on FS-GA-SVR hybrid model. Journal of Environmental Management, 344: 118446. https://doi.org/10.1016/j.jenvman.2023.118446

[18] Omondiagbe, D.A., Veeramani, S., Sidhu, A.S. (2019). Machine learning classification techniques for breast cancer diagnosis. In IOP Conference Series: Materials Science and Engineering, Sarawak, Malaysia, p. 012033. https://doi.org/10.1088/1757-899X/495/1/012033

[19] Aziz, R.M., Baluch, M.F., Patel, S., Kumar, P. (2022). A machine learning based approach to detect the Ethereum fraud transactions with limited attributes. Karbala International Journal of Modern Science, 8(2): 139-151. https://doi.org/10.33640/2405-609X.3229

[20] Saoud, H., Ghadi, A., Ghailani, M., Abdelhakim, B.A. (2018). Using feature selection techniques to improve the accuracy of breast cancer classification. In The Proceedings of the Third International Conference on Smart City Applications, Springer, Cham, pp. 307-315. https://doi.org/10.1007/978-3-030-11196-0_28

[21] Bakır, Ç., Öngenli, K. (2025). Prediction and comparative analysis of factors affecting the mathematical achievement of gifted students with machine learning models. International Journal of Adult Education and Technology (IJAET), 16(1): 1-25. https://doi.org/10.4018/IJAET.383672

[22] Ishaq, A., Sadiq, S., Umer, M., Ullah, S., Mirjalili, S., Rupapara, V., Nappi, M. (2021). Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques. IEEE Access, 9: 39707-39716. https://doi.org/10.1109/ACCESS.2021.3064084

[23] Spencer, R., Thabtah, F., Abdelhamid, N., Thompson, M. (2020). Exploring feature selection and classification methods for predicting heart disease. Digital Health, 6: 2055207620914777. https://doi.org/10.1177/2055207620914777

[24] Yang, J., Guan, J. (2022). A heart disease prediction model based on feature optimization and smote-Xgboost algorithm. Information, 13(10): 475. https://doi.org/10.3390/info13100475

[25] Aggarwal, R., Kumar, S. (2022). HRV based feature selection for congestive heart failure and normal sinus rhythm for meticulous presaging of heart disease using machine learning. Measurement: Sensors, 24: 100573. https://doi.org/10.1016/j.measen.2022.100573

[26] Parisi, L., Manaog, M.L. (2025). Funnel Random Forest: Inliers-focused ensemble learning for improved prognostics of heart failure. https://doi.org/10.21203/rs.3.rs-5784003/v1

[27] Sanni, R.R., Guruprasad, H.S. (2021). Analysis of performance metrics of heart failured patients using Python and machine learning algorithms. Global

Transitions Proceedings, 2(2): 233-237. https://doi.org/10.1016/j.gltp.2021.08.028

[28] Gandhimathi, S.K., Rajesh, S.M., Ghamya, K., Lohitha, A.L., Vinitha, A., Reddy, C.S.P. (2024). Machine learning mastery in cardiovascular risk assessment. In 2024 1st International Conference on Advances in Computing, Communication and Networking (ICAC2N), Greater Noida, India, pp. 700-705. https://doi.org/10.1109/ICAC2N63387.2024.10895421

[29] Al-Jamimi, H.A., Ayad, S., El Kheir, A. (2025). Integrating advanced techniques: RFE-SVM feature engineering and Nelder-mead optimized XGBoost for accurate lung cancer prediction. IEEE Access, 13: 29589-29600. https://doi.org/10.1109/ACCESS.2025.3536034

[30] Bhuria, R., Sudheer, K. (2025). Leveraging XGBoost for accurate brain stroke prediction: An innovative approach to data-driven healthcare solutions. In 2025 International Conference on Pervasive Computational Technologies (ICPCT), Greater Noida, India, pp. 144-148. https://doi.org/10.1109/ICPCT64145.2025.10939264

[31] Husain, G., Nasef, D., Jose, R., Mayer, J., Bekbolatova, M., Devine, T., Toma, M. (2025). SMOTE vs. SMOTEENN: A study on the performance of resampling algorithms for addressing class imbalance in regression models. Algorithms, 18(1): 37. https://doi.org/10.3390/a18010037

[32] Zaimi, R., Hafidi, M., Lamia, M. (2024). A deep learning mechanism to detect phishing URLs using the permutation importance method and SMOTE-Tomek link. The Journal of Supercomputing, 80(12): 17159-17191. https://doi.org/10.1007/s11227-024-06124-7

[33] Sumayli, A. (2023). Development of advanced machine learning models for optimization of methyl ester biofuel production from papaya oil: Gaussian process regression (GPR), multilayer perceptron (MLP), and K-nearest neighbor (KNN) regression models. Arabian Journal of Chemistry, 16(7): 104833. https://doi.org/10.1016/j.arabjc.2023.104833

[34] Naji, M.A., El Filali, S., Aarika, K., Benlahmar, E.H., Ait Abdelouhahid, R., Debauche, O. (2021). Machine learning algorithms for breast cancer prediction and diagnosis. Procedia Computer Science, 191: 487-492. https://doi.org/10.1016/j.procs.2021.07.062

[35] Wang, H., Liang, Q., Hancock, J.T., Khoshgoftaar, T.M. (2024). Feature selection strategies: A comparative analysis of SHAP-value and importance-based methods. Journal of Big Data, 11(1): 44. https://doi.org/10.1186/s40537-024-00905-w

[36] Lee, Y.G., Oh, J.Y., Kim, D., Kim, G. (2023). Shap value-based feature importance analysis for short-term load forecasting. Journal of Electrical Engineering & Technology, 18(1): 579-588. https://doi.org/10.1007/s42835-022-01161-9

[37] Guo, F., Zou, F., Luo, S., Liao, L., Wu, J., Yu, X., Zhang, C. (2022). The fast detection of abnormal ETC data based on an improved DTW algorithm. Electronics, 11(13): 1981. https://doi.org/10.3390/electronics11131981

[38] Souza, V.S., Lima, D.A. (2025). Cardiac disease diagnosis using k-nearest neighbor algorithm: A study on heart failure clinical records dataset. Artificial Intelligence and Applications, 3(1): 56-71. https://doi.org/10.47852/bonviewAIA42022045

[39] Reddy, D.K.K., Behera, H.S., Ding, W. (2022). Advance

machine learning and nature-inspired optimization in heart failure clinical records dataset. In Nature-Inspired Optimization Methodologies in Biomedical and Healthcare, Springer, Cham, pp. 221-246. https://doi.org/10.1007/978-3-031-17544-2_10

[40] Oladimeji, O.O., Oladimeji, O. (2020). Predicting survival of heart failure patients using classification algorithms. Journal of Information Technology and Computer Engineering, 4(2): 90-94. https://doi.org/10.25077/jitce.4.02.90-94.2020