Development and Validation of a Class Imbalance-Resilient Cardiac Arrest Prediction Framework Incorporating Multiscale Aggregation, ICA and Explainability

Iffa Afsa C M, Mohammed Yusuf Ansari, Santu Paul, Osama Halabi, Ezzedin Alataresh, Jassim Shah, Afaf Hamze, Omar Aboumarzouk, Abdulla Al-Ansari, and Sarada Prasad Dakua*

***Abstract*—Objective: Despite advancements in artificial intelligence (AI) for predicting cardiac arrest (CA) with multivariate time-series vital signs data, existing models continue to face significant problems, particularly concerning balance, efficiency, accuracy, and explainability. While neural networks have been proposed to extract multiscale features from raw data in various applications, to our knowledge, no work has utilized multiscale feature extraction, specifically for diagnostic CA prediction. This paper presents a new framework that tackles these difficulties by utilizing multiscale feature aggregation via Independent Component Analysis (ICA). Methods: We present the Pareto optimal StrataChron Pyramid Fusion Framework (SPFF) that improves temporal vital signs statistics by capturing long-term dependencies using multi-scale temporal feature aggregation. SPFF integrates with ICA to eliminate information redundancy and enhance model efficiency. We have developed and validated the approach using the public MIMIC IV dataset. Results: The proposed model demonstrates resilience to data imbalance and enhances explainability through SHapley Additive exPlanations (SHAP), Partial Dependence Plots (PDP), and Individual Conditional Expectation (ICE) across varying time windows. The multilayer perceptron (MLP) using SPFF and ICA achieves an accuracy of 0.982, precision of 0.969, recall of 0.989, F1-score of 0.979, and AUROC of 0.998. Conclusion: The proposed method effectively predicts CA across varying time windows, offering a robust solution to the challenges of efficiency, accuracy, and explainability in current models. Significance: This method is significant to biomedical research as it provides superior performance in CA prediction while capturing both short- and long-term dependencies in patient data, potentially improving patient outcomes and guiding clinical decision-making.**

Index Terms—**Artificial Intelligence, Cardiac Arrest, Multi-Scale Temporal Aggregation, Independent Component Analysis, Explainable AI**

We would like to thank the Clinical Academic Department for their support and guidance. This work did not involve human subjects or animals in its research.

Iffa Afsa C M is a Research Assistant in Hamad Medical Corporation, Doha (Qatar). Iffa is pursuing her MSc in Qatar University, Doha (Qatar).
Mohammed Yusuf Ansari is a PhD student at Texas A&M University, College Station (USA).
Santu Paul is a Clinical Research Specialist in Hamad Medical Corporation, Doha (Qatar).
Osama Halabi is an Associate Professor in Qatar University, Doha (Qatar).
Ezzedin Alataresh and Jassim Shah are Consultant Cardiologists in Heart Hospital, Hamad Medical Corporation, Doha (Qatar).
Afaf Hamze, Omar Aboumarzouk, Abdulla Al-Ansari, and Sarada Prasad Dakua serve as Medical Resident, Vice Chairman of Surgery for Research, Chief Medical Officer, and Senior Research Scientist in Hamad Medical Corporation, Doha (Qatar) respectively. Additionally, Afaf, Sarada, and Omar serve in the Clinical Advancement Department in Hamad Medical Corporation, Doha (Qatar). Furthermore, Sarada and Omar serve as Clinical Assistant Professor and Clinical Professor in Qatar University, respectively. (e-mail: SDakua@hamad.qa)

I. Introduction

The World Heart Report (2023) indicates cardiovascular diseases (CVDs) account for approximately one-third of global mortality, with an estimated 17.9 million deaths annually [1], [2], [3], [4]. Sudden cardiac arrest (SCA) refers to the unexpected cessation of cardiac activity, with resuscitation efforts defining the attempt to restore circulation; failure of these efforts results in sudden cardiac death (SCD) [3], [5], [6]. The International Liaison Committee on Resuscitation has classified cardiac arrests (CA) based on medical or external causes, including drug overdose, drowning, asphyxia, electrocution, and trauma [7].

Although the literature has several approaches to predict CA such as by determining left ventricle segmentation, ejection fraction, etc. [8], [9], [10], [11], [12], [13], [14], detecting CA is often difficult due to its subtle onset and late diagnosis, complicating timely intervention. Thus, healthcare systems certainly require effective tools that could predict CA early to lower the mortality rates [1].

There have been some key indicators already outlined, including: a) multivariate time-series vital signs data, which includes diastolic blood pressure, systolic blood pressure, heart rate, oxygen saturation, respiratory rate, and temperature; and b) baseline demographic indicators which include gender and age. Any abnormal fluctuations in these indicators are clinically associated with the onset of CA [15]. AI-based algorithms have already been tried using various machine learning (ML) [16], [17], [18] and deep learning (DL) capabilities to analyze trends from retrospective EHR data [19]. Citing the fact that AI algorithms need significant computing power [20], it is imperative to decrease the number of features while retaining class-discriminative context both on local and global scales. Moreover, it is crucial to preserve information over different time periods to capture the key details, especially their fluctuations. Therefore, an efficient framework is required to aggregate temporal features at multiple scales to capture long-term relationships and shifting distributions of vital signs.

Transparency and explainability are crucial for implementing AI models in clinical practice, as inaccurate predictions

might have serious implications. The clinicians must comprehend the AI model by identifying the key pieces of the model that could enhance their understanding [21]. Nevertheless, it can be argued that ML models provide a certain degree of explainability when compared to DL models, known as "black boxes." Although ML models are comprehensible, they are less accurate than DL models; thus, we integrate Pareto optimality into the proposed methodology. It may be noted that the collection of decisions representing the best possible outcomes, when there are multiple candidates for the optimal solution, is referred to as the Pareto front [22]. Incorporating Pareto optimality seeks to create AI algorithms that are both computationally efficient and highly accurate while possessing a degree of explainability.

This paper proposes a novel fused Pareto optimal CA prediction model using multivariate time-series vital signs and demographic EHR data. The model uses a few temporal features and computational resources to achieve state-of-the-art performance.

Fundamentally, we contribute the following:

1) We propose a StrataChron Pyramid Fusion Framework (SPFF) with a multi-scale temporal feature aggregation to capture long-term dependencies and changing distributions of the vital signs, thereby improving the prediction performance.
2) We employ Independent Component Analysis (ICA) on the aggregated features to reduce the amount of information that is duplicated across time scales, thereby improving the computational efficiency and efficacy of ML models.
3) We simulate and validate that the multi-scale temporal feature aggregation framework and ICA are robust to data imbalance.
4) We empirically evaluate the effectiveness of ICA with other dimensionality reduction techniques and conduct a comparative analysis of the proposed SPFF and ICA method with some recent studies in the literature.
5) We provide a comprehensive explainability layer by employing SHapley Additive exPlanations (SHAP), Partial Dependence Plots (PDP), and Individual Conditional Expectation (ICE) to provide insights about the model's decision-making capabilities.

To the best of our understanding, this study is the first to integrate multi-scale temporal aggregation, Independent Component Analysis (ICA), and combine explainability methods including SHAP, PDP, and ICE on temporal vital signs data for predicting CA. This methodology could diminish data complexity, emphasize the essential characteristics, and offer insights at both global and individual levels. The approach streamlines the model by concentrating on a limited number of features, enhancing its efficiency, usability, and interpretability, all while preserving high accuracy.

The rest of the paper is organized as follows: The Section II discusses the literature. In Section III, we describe the methodology that incorporates the SPFF, ICA, and explainability. Section IV describes the experimental setup, and Section V presents the outcomes, empirical investigations, explainability, and comparison with previous studies. Lastly, Section VI summarizes the study and concludes the paper.

II. Related Works

In recent years, several ML and DL models have been proposed to predict in-hospital CA (IHCA); most of the models utilize vital signs, demographic characteristics, laboratory measurements, ECG, or EKG from EHR data. Kim et al. [23] present a clinically interpretable ensemble method that combines sliding window-based multiresolution statistical features and cosine similarity-based features using multivariate time-series vital signs up to 6 hours in advance. The approach achieves an overall area under the receiver operating characteristic curve (AUROC) of 0.86 and an area under the precision-recall curve (AUPRC) of 0.58. Despite utilizing SHAP values for feature evaluation, the approach fails to incorporate optimal feature selection and achieve high precision. Hong et al. [24] propose a model using sequential characteristics of vital signs and demographic features; it uses three ML models: random forest (RF), recurrent neural network (RNN), and logistic regression (LR). The RF method demonstrates superior performance, achieving AUROC and AUPRC of 0.97 and 0.86, respectively. A chronological depiction of features presents the model's explainability. Layeghian et al. [25] present an ISAF framework for predicting CA in sepsis patients using textual EHR data; the stacking model can predict 85% of CA cases one hour prior to the event (with a sensitivity of at least 0.85) and 73% of CA cases 25 hours before they happen (with a sensitivity of at least 0.73). Yijing et al. [26] present CAPI, a real-time and interpretable extreme gradient boosting (XGB) model using vital signs that predicts CA from bedside vital signs with an AUROC of 0.94. The vital signs are analyzed using differential features and statistical features based on 2-hour sliding windows. However, the model exhibits alarm fatigue by having average sensitivity (0.86) and specificity (0.85) scores, a very low AUPRC (0.12), and an F1-score (0.05). Lee et al. [27] develop an explainable ML model using 33 HRV parameters that are derived from ECG data, where the light gradient boosting machine (LGBM) model achieves an AUROC of 0.881.

Jang et al. [28] present an early prediction model using three artificial neural network (ANN) models (multilayer perceptron (MLP), long short-term memory (LSTM), and hybrid) that are trained using demographic information, chief complaints, vital signs, and degree of awareness. The hybrid ANN achieves an AUROC of 0.936 and is proven to be better than the Modified Early Warning Score (MEWS). However, the proposed methodology lacks model explainability and timely assessment. Raheem et al. [29] propose an ANN model for early prediction of major adverse cardiac events (MACE) in ED triage. The model demonstrates better performance in predicting in-hospital mortality as compared to RF and LR models, with a sensitivity and specificity of 94.6% and 93.3%, respectively. To reduce false alarms and provide better specificity-sensitivity, Baral et al. [30] propose using a hybrid model that combines an MLP and an enhanced Bidirectional LSTM (Bi-LSTM) to process time series vital signs. However,

the method is capable of predicting only 6 hours prior to the CA with an AUC of 0.94 for an hour time window. Kwon et al. [31] propose a DL-based Early Warning System (DEWS) to predict CA using only four vital signs (heart rate, body temperature, systolic blood pressure, and respiratory rate). Although the model outperforms MEWS, RF, and LR, it exhibits an average AUROC of 0.850 and lacks model explainability.

Smith et al. [32] compare the National Early Warning Score (NEWS) to 33 other Early Warning Scores (EWS) using a large vital signs database; NEWS can identify patients at risk within 24 hours with an AUROC of 0.722. However, its statistical performance with respect to operational efficacy poses a concern. The superiority of DEWS over MEWS for IHCA prediction is validated by Lee et al. [33]. Shamout et al. [34] present another DEWS model that achieves an AUROC of 0.880 and is proved to be better than NEWS (AUROC of 0.866).

Explainability and efficiency remain two major concerns in most of the aforementioned methods. Despite the emphasis on improving models in most studies, the complexity of EHR data has received little attention. The data is multimodal, high-dimensional, semantically heterogeneous, biased, and confounding. We believe that preprocessing and feature extraction could probably play a critical role in preparing EHR data for ML analysis. Effective preprocessing and feature extraction techniques could be essential for handling the complexity, heterogeneity, and nuances inherent in EHR data, ultimately enabling more accurate, reliable, and actionable insights. ICA is considered valuable for simplifying complex data such as EHRs; it can reduce the dimensionality, identify the independent components, remove the redundancy and noise, discover the hidden relationships, extract the features, and better interpret the results [35], [36], [37]. To the best of our understanding, our present study is probably the first attempt that employs ICA with time-series vital signs data.

III. Proposed Methodology

In this section, we present the technical details and insights of the pyramid fusion framework, along with feature reduction using ICA and ML explainability.

A. StrataChron Pyramid Fusion Framework

The proposed framework employs Pareto optimality and uses a pyramid design to integrate the multi-scale temporal context of multivariate time-series vital signs data, incrementally accumulating layers or hours of time. By doing this, an optimal equilibrium between the accurate prediction ability in the short term and long term is obtained while minimizing the loss of temporal information across multiple time scales. Given a time-series dataset $D = \{d_1, d_2, \ldots, d_n\}$, where each d_i represents data collected during the i^{th} hour. We aim to perform a pyramid-style, multi-scale temporal feature aggregation. This process involves aggregating features across all possible contiguous intervals within the n-hour period, using an aggregation function $\Phi = \{mean, median, min, max\}$.

The hierarchical structure of the aggregation process begins with the most granular level (individual hours) and progresses to the most aggregated level (the entire dataset) The process is defined as follows:

1) Level 1 (Base Level): Aggregate features for individual hours.
$$F_{i,i} = \Phi(d_i) \quad \text{for each hour } i \tag{1}$$
2) Subsequent Levels of Aggregation: Continue this process for all consecutive k-hour intervals, where k ranges from 2 to n.
$$F_{i,i+k-1} = \Phi(\{d_i, d_{i+1}, \ldots, d_{i+k-1}\}) \quad \text{for each level } k \text{ and starting hour } i \quad \text{where } i + k - 1 \leq n \tag{2}$$
3) Final Level Aggregation (Entire n-Hour Period): Aggregate features for the entire n-hour period.
$$F_{1,n} = \Phi(\{d_1, d_2, \ldots, d_n\}) \tag{3}$$

The pyramid-style multi-scale temporal feature aggregation can be compactly represented as:

$$F_{i,j} = \begin{cases} \Phi(\{d_i\}) & \text{if } j = i, \\ \Phi(\{F_{i,k}, F_{k+1,j}\}) & \text{for } i < j \leq n \text{ and some } k \text{ where } i \leq k < j \end{cases} \tag{4}$$

In this formula, $F_{i,j}$ represents the aggregated feature over the interval from hour i to hour j, inclusive. The first case applies, when the interval is just one hour, and the second case represents the recursive aggregation from i to j, facilitating the hierarchical aggregation process as detailed in Algorithm. 1.

Algorithm 1 Pyramid-Style Multi-Scale Temporal Feature Aggregation

Input: Dataset $D = \{d_1, d_2, \ldots, d_n\}$, where each d_i is data from the i-th hour, Aggregation Function Φ (mean, median, min, max)
Output: Aggregated features for all possible contiguous intervals
$AggregatedFeatures = []$ ▷ Empty list to store results
for $k \in \{1, 2, \ldots, n\}$ **do**
  if $k = 1$ **then**
    for $i \in \{1, 2, \ldots, n\}$ **do**
      Compute $F_{i,i} = \Phi(d_i)$
      Append $F_{i,i}$ to $AggregatedFeatures$
    end for
  else
    for $i \in \{1, 2, \ldots, n - k + 1\}$ **do**
      Compute $F_{i,i+k-1} = \Phi(\{d_i, \ldots, d_{i+k-1}\})$
      Append $F_{i,i+k-1}$ to $AggregatedFeatures$
    end for
  end if
end for
return $AggregatedFeatures$

We employ the proposed SPFF for a 3-hour time window of vital signs data and derive multi-scale temporal aggregated

features. To reinforce the application of SPFF for multi-scale temporal aggregation, an empirical study is conducted to examine the feature distribution of the dataset based on the occurrence of CA events in both CA patients and non-CA patients. A normal 2-hour sliding window-based aggregation technique is utilized for the same purpose inspired by the studies [23] and [26]. The objective is to demonstrate that SPFF offers greater performance in terms of differentiating between individuals at risk and those with no risk of CA while yielding promising results.

B. Transforming Complexity: Leveraging ICA

ICA has previously been utilized with waveform data, such as ECG [35], EEG [36], and fMRI [37]. This study is probably the first to employ ICA in analyzing aggregated multivariate time-series vital signs data. ICA is a statistical method that can be used as an alternative to PCA. Unlike PCA, ICA is specifically designed to separate multivariate signals into independent components, even if the signals are not normally distributed [38]. Therefore, our objective is to utilize ICA in conjunction with aggregated multivariate time-series data to create features that possess independent components with the highest amount of variance which can aid in detecting concealed patterns within these aggregated features. This process has the potential to unveil underlying processes or sources of variation that may not be evident solely from the aggregated data.

We systematically vary the number of components and assess the models' performance to determine the optimal number of components for ICA that could maximize the variance and minimize the noise. Using an iterative procedure, we determine that the optimal number of components is 18. An empirical study is done to analyze the feature distribution of the dataset based on CA events in patients with and without CA using data subset 1 to demonstrate the use of ICA for feature engineering and dimensionality reduction. For this, PCA and kernel PCA are used; kernel PCA uses RBF, poly, and sigmoid kernels. The objective is to show that ICA is better at distinguishing CA-risk patients from those without CA risk. An empirical study examines how the proposed SPFF in conjunction with ICA affects the class imbalance. Various levels of imbalance are simulated for the same. The objective is to demonstrate the resistance of the proposed methodology to class imbalance.

Despite evaluating both PCA and kernel PCA, we ultimately choose ICA, because neither PCA nor kernel PCA is designed to perform source separation along with other ICA advantages. For instance, ICA is more effective in extracting features from data where the underlying signals or components are non-Gaussian, like our data which includes heart rate & heart rate variability, blood pressure, respiratory rate, and oxygen saturation. These non-Gaussian vital signs also exhibit non-linear correlations, and ICA is proficient at isolating statistically separate components from this data [39]. PCA, although effective for maximizing variance, operates under the presumption of linear correlations, which may not be optimal in this context [40].

Our empirical studies demonstrate that ICA outperforms both PCA and kernel PCA in differentiating between patients at risk of CA and those who were not. This is because ICA emphasis on identifying independent sources of variation, facilitating the isolation of significant signals, such as alterations in systolic and diastolic blood pressure, from background noise. Consequently, ICA yields more interpretable and therapeutically pertinent characteristics for forecasting CA.

C. Unpacking the Predictors: Exploring Explainability

We use SHAP, PDP, and ICE to comprehensively explain ML models enhanced with SPFF and ICA. The SHAP technique employs game theory to explain model output, assigning a prediction importance rating to each feature. SHAP values quantify how features affect the complex model output, simplifying the interpretation. It is useful since it can be used with multiple models and provides consistent and localized explanations [41]. However, SHAP primarily focuses on feature importance rather than the nature of the relationship between characteristics and outcomes. To complement SHAP and comprehend the impact of individual characteristics on the model's behavior, we require PDP and ICE. PDP demonstrates the variation in average predictions based on a certain feature while accounting for the influence of all other features. Thus, these interpreters help clarify the influence of one or two variables on the given output of an ML model, providing a comprehensive overview of feature impacts throughout the entire dataset. [42], [43]. Conversely, ICE provides detailed perspective by illustrating how the model's predictions fluctuate for individual patients as a particular attribute changes, thereby capturing heterogeneity among diverse patient profiles [44]. ICE plots enable the identification of patient-specific interactions that PDP, which averages predictions, might neglect. As a result, while SHAP provides a global understanding of feature importance to explain each prediction, PDP focuses on the average impact of a feature across the dataset. ICE highlights how individual responses to a feature can differ, thereby revealing the model's behavior throughout the entire feature space.

The three tools discussed above fully explain our process for predicting CA from multivariate time-series EHR data. SHAP gives both local and global insights, helping clinicians to visualize the effect of vital signs on a patient's risk. PDP shows the broader trends across all patients, and ICE highlights the differences at the individual level. Together, these tools complement each other and make the model or process more understandable.

To better understand the model's behavior and identify the most critical features in the dataset, we provide a visual representation of SHAP values for the top five ICA-generated components. We do this for both, the most effective ML model and the DL model. To better understand the components, we plot the PDP and ICE analyzing their distribution. We plot the heat map for the most influential ICA component and the aggregated features that contribute to it using a back projection technique. In this manner, a strong and comprehensible layer utilizes back projection to obtain global and specific characteristics, enhancing the model's prediction and effectiveness [45].

IV. Experimental Setup

Here, we present the details about the dataset, processing, analysis, implementation, and evaluation metrics.

A. Dataset and Pre-processing

The proposed SPFF in conjunction with ICA are evaluated using the Medical Information Mart for Intensive Care IV (MIMIC-IV v2.2) database, which is created by the MIT lab [46]. The database comprises anonymized patient data obtained from MetaVision bedside monitors at an academic medical institution in Boston, MA. The data includes information from over 50,000 adult patients between 2008 and 2019 [46]. The criteria for including patients are illustrated in Fig. 1. The database is refined by excluding patients from the ICU who have no recorded charted data. Subsequently, the patients who do not have any significant record of vital signs of interest are eliminated. Next, the database is divided into two distinct groups: the CA cohort (suffering from CA), and the non-CA cohort (not suffering from CA) while in the ICU. Finally, the CA patients without pre-CA vital signs are eliminated.

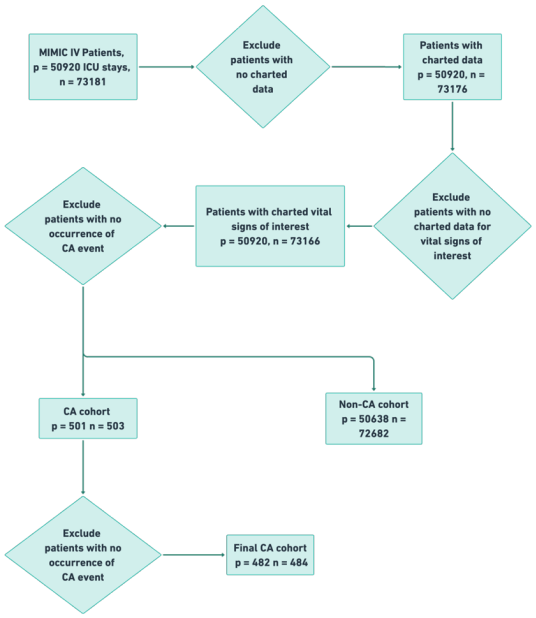


Fig. 1: Patient inclusion criteria depicting creating of CA and non-CA cohort.

Vital signs like diastolic and systolic blood pressure, heart rate, oxygen saturation, respiration rate, and temperature are the physiological indicators that we have used as raw features. We have also considered the demographic indicators like gender and age, since these are linked to the start of CA [15]. We remove the outliers and fill the missing values round-robin using the iterative imputation technique. To understand the feature distribution of the CA and non-CA cohorts, we perform a statistical evaluation using a T-test for the numerical values and a chi-square test for the categorical values, along with an aggregation summary using mean, median, minimum, and maximum. The summary of the same is provided in Table I.

TABLE I: Characteristics of CA and non-CA patients.

Characteristic	CA	Non-CA	p-value
Diastolic Blood Pressure, dbp			
min, max	15, 140	24, 126	**<0.001**
mean, median	57.8, 55	63.5, 62	**<0.001**
Heart Rate, hr			
min, max	27, 199	41, 155	**<0.001**
mean, median	97, 95	84.2, 82	**<0.001**
Oxygen Saturation, SpO2			
min, max	2, 100	69, 100	**<0.001**
mean, median	94.8, 97	97.2, 98	**<0.001**
Respiratory Rate, rr			
min, max	0, 71	0, 75	**<0.001**
mean, median	22.7, 22	19.4, 18	**<0.001**
Systolic Blood Pressure, sbp			
min, max	4, 270	58, 206	**<0.001**
mean, median	104.4, 102	118.4, 116	**<0.001**
Temperature, temp			
min, max	30.7, 40.4	33.2, 39.2	0.431
mean, median	36.5, 36.7	36.6,36.6	0.431
Age			
min, max	19, 91	19, 91	0.361
mean, median	64, 67	63.8, 67	0.361
Gender			0.003
Female (%)	34.7	39	
Male (%)	65.4	61	

For a thorough investigation, we try various experimental subsets; the literature extensively documents alterations in vital signs before clinical deterioration, emphasizing the crucial importance of prompt identification of such preventable outcomes for appropriate intervention [47]. However, the clinicians usually witness abnormal vital indicators 1-4 hours prior to the actual CA [48]. Therefore, it is crucial to extend the period during which the deterioration can be diagnosed using state-of-the-art methods. We create the experimental subsets with progressively longer time periods providing a significant interval. These subsets are based on a 3-hour time window for collecting data, which occurs i hours before CA. Here, i represents values 3, 6, 12, and 15. For the non-CA cohort, subsets of data are created for a matching time window using charted data. Consequently, four subsets of data are created, each consisting of vital signs collected before the CA event for the CA cohort and within a comparable time window for the non-CA cohort. In addition to creating balanced datasets for the main experiments, we specifically design four imbalanced subsets to evaluate the model's performance under real-world conditions, where CA is a rare event. The imbalanced classes have a non-CA cohort with a size exceeding the CA cohort by 100% (CA cohort + 1 x CA cohort), 150% (CA cohort + 1.5 x CA cohort), 200% (CA cohort + 2 x CA cohort), 250% (CA cohort + 2.5 x CA cohort), and 500% (CA cohort + 5 x CA cohort). The various data subsets are provided in the Table II. To handle class imbalance in these subsets, undersampling is used to avoid over-representing the majority non-CA cohort.

In our study, multi-scale temporal aggregation is performed using the equation 4. Since we consider a standard time window of n = 3, we start with aggregating data for the base level, where a sliding window of k = 1 is considered. In the following step, we increase the window size to k = 2 and slide it to aggregate the data. Lastly, we compute the final level of aggregation by increasing the window size to k = 3, which

matches the entire n hour time window. We then use ICA for feature engineering and tune the number of components to 18.

TABLE II: Experimental subsets where n is the number of patients.

Subset	Hours prior CA	CA (n)	Non-CA (n)	Imbalance (degree)
1	3	356	356	0%
2	6	319	319	0%
3	12	279	279	0%
4	15	269	269	0%
5	3	356	712	100%
6	3	356	890	150%
7	3	356	1068	200%
8	3	356	1246	250%
9	3	356	2136	500%

B. Exploratory Data Analysis

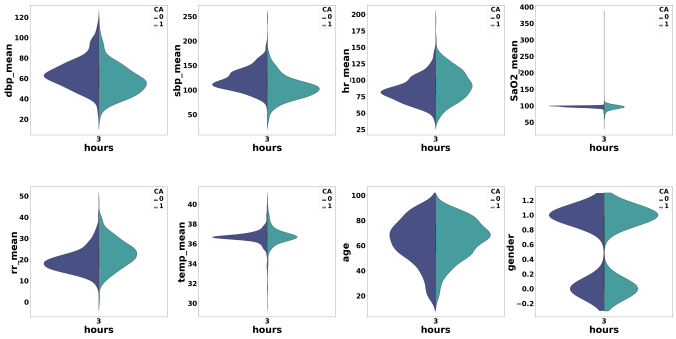


Fig. 2: EDA to view the feature distribution in CA and non-CA cohort at 3 hours prior to the CA event.

We investigate the data distribution both in the CA cohort and the non-CA cohort with the help of an exploratory data analysis (EDA). To streamline the visualization, the EDA is graphed for a one-hour time frame occurring 3 hours prior to the occurrence of CA. This time frame is selected from subset 1 that includes three hours of data acquired three hours before the CA event. Fig. 2 displays the plot; the plots offer a distinct representation of the closeness of the feature distribution in the dataset across the various cohorts. These findings suggest the need for additional feature engineering techniques to effectively capture both the local and global temporal context of the time series data. This could improve the ability of the model to predict CA using time series information.

C. Implementation Details

The data is split into two sets: the training dataset (70% randomly selected instances) and the testing dataset (30% holdout cases) that are used for model development and validation, respectively. Seven binary classifiers: three ML models, two ensemble stacking models, and two DL models are trained and evaluated. RF, support vector machine (SVM), and XGB are employed in this investigation as ML models. Three XGB models are the first stacking model (XSM); stacked RF, SVM, and XGB (MMS) are the second. MLP and 1D Convolutional Neural Network (1D-CNN) are employed in this study as DL models, in which Python and Sci-kit libraries are used. We have been motivated by Occam's Razor hypothesis to select these simple architectures that favors simpler models achieving par performance [49]. Given the size of the present dataset, MLP has offered strong performance with lower computational costs. Additionally, MLPs are easier to deploy in clinical settings [50].

Multiple strategies were implemented during model training to prevent overfitting. Dropout layers were implemented with a dropout rate of 0.5 to enhance generalization. Furthermore, early stopping was used to terminate the training process when no enhancement in validation loss was observed. Additionally, the default batch size and maximum iterations were adjusted to achieve balanced learning. Ultimately, 5-fold cross-validation was employed to assess the model's performance across each fold, demonstrating consistent results throughout all iterations.

D. Evaluation Metrics

This study considers some popular measures, such as ***accuracy, precision, sensitivity, F1-score, and AUROC***; the accuracy is chosen since it quantifies a model's capacity to detect instances accurately and directly. Similarly, determining the true CA cases requires high sensitivity; the precision is chosen to reduce false positives. Considering high sensitivity and precision, the F1-score is used to evaluate the models. In various binary classification tasks, the AUROC is a reliable metric that measures a model's ability to discriminate between classes. AUROC is reliable for comparing model performance on a dataset. Furthermore, to evaluate the impact on class imbalance, we validated the model's performance on the minority CA cohort utilizing separate accuracy, recall, and F1-score metrics. The focus was on guaranteeing that the model effectively recognized CA cases while minimizing the erroneous classification of non-CA events as CA. This methodology guaranteed that the model retained its robustness and efficacy, even in the presence of significant imbalances.

V. Results and Discussion

In this section, we present the prediction results after employing the pyramid fusion with ICA and display its robustness to class imbalance. Additionally, we compare the ICA to other dimensionality reduction techniques and explain the ML models. Lastly, we show the efficiency of the method by comparing it with other recent studies.

A. Impact of Multi-Scale Temporal Feature Aggregation and ICA

We employ tSNE plots to depict the distribution of multivariate time-series vital signs and visualize distinct groups of CA and non-CA cohorts; Fig. 3 shows no evidence of

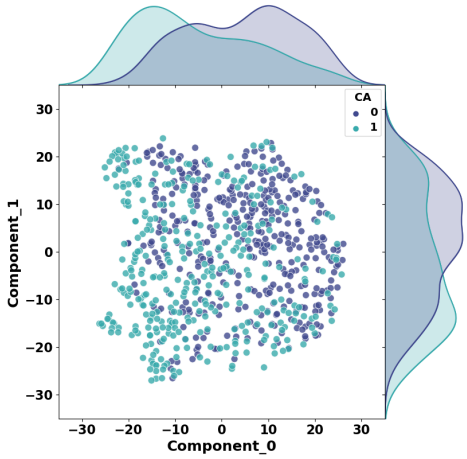
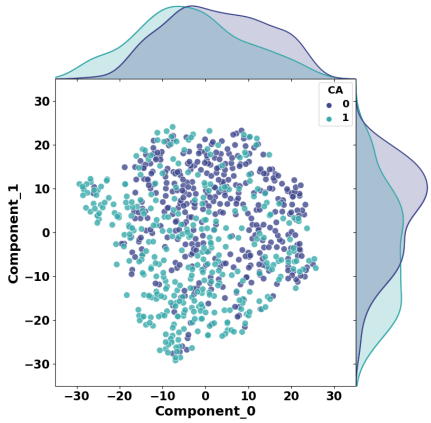
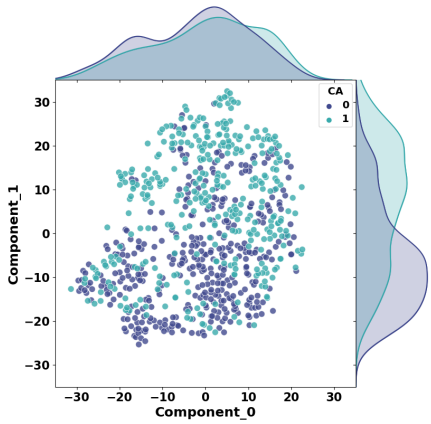


Fig. 3: tSNE projection of the feature distribution of non-aggregated data by CA event. The CA cohort is depicted by CA = 0, and non-CA cohort is depicted by CA = 1.



(a) First-order uni-scale aggregation.



(b) Proposed SPFF aggregation.

Fig. 4: tSNE projection to compare the feature distribution by CA event for (a) first-order uni-scale aggregation and (b) the proposed SPFF framework.

differentiation between the CA group and non-CA cohort, whereas Fig. 4a enhances the distribution of the dataset by CA event to a certain degree after applying standard uni-scale aggregation to each time-series vital sign for every hour within the specified time range. In the proposed SPFF method, we employ phi = mean, median, minimum, and maximum calculations for each vital sign within the specified time window of $n = 3$, $k = 1$, 2, and 3. This shows that the multi-scale temporal aggregation enhances the differentiation between the CA and non-CA cohorts by incorporating both local and global multi-scale temporal features (as illustrated in Fig. 4b). Thus, the SPFF method allows for a more comprehensive understanding of multivariate time series data distribution. However, the SPFF method has yielded results using excessive number of features.

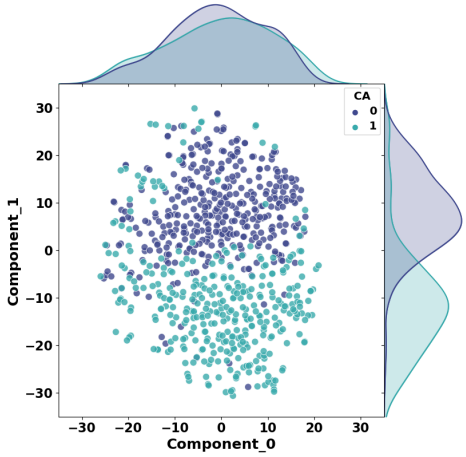


Fig. 5: tSNE projection of the feature distribution by CA event using ICA reduced SPFF features.

TABLE III: Results for CA prediction 3 hours prior to the event using ICA reduced SPFF features.

Model	Accuracy	Precision	Recall	F1 Score	AUROC
RF	0.950	0.911	0.979	0.944	0.992
SVM	0.960	0.913	**1.000**	0.954	0.995
XGB	0.964	0.958	0.957	0.957	0.994
XSM	0.955	0.957	0.936	0.946	0.967
MMS	0.973	0.940	**1.000**	0.969	0.976
MLP	**0.982**	**0.969**	0.989	**0.979**	**0.998**
CNN	0.936	0.877	0.989	0.930	0.993

We use ICA with 18 components on the multi-scale, temporally aggregated features and the demographic data to reduce the computational complexity and prioritize the important features. This is shown in Fig. 5, which shows how the dataset is split into CA and non-CA cohorts. These developed characteristics are later used to train the models, and the scores across them are presented in Table III. The MLP model shows promising performance with accuracy, precision, recall, F1-score, and AUROC of 0.982, 0.969, 0.989, 0.979, and 0.998, respectively. This can be attributed to the non-linear activation function employed by MLP to learn complex intricate non-linear decision boundaries. Furthermore, the nature of the data and the preprocessing steps used might account

for the MLP's stronger performance compared to the CNN model. After applying SPFF and ICA, the features become more abstract and independent, reducing the need for the local pattern detection that CNNs typically excel at. CNNs excel at extracting spatial features from unstructured data, such as images, where neighboring pixels or elements have important relationships [51]. However, in structured tabular data like in this experiment, especially after dimensionality reduction and feature extraction, the data no longer contains local dependencies that CNNs are optimized to detect. In contrast, MLPs are well-suited to tabular data, where the input features are already well-defined and independent. MLPs can more effectively handle structured data, where each feature contributes individually to the model's output. Several studies have shown that MLPs outperform CNNs on tabular datasets for similar reasons, where feature extraction is not needed, and simpler feedforward networks are sufficient to capture relationships between independent variables [52], [53], [54]. Thus, the superior performance of the MLP model in this study is consistent with these findings. However, both the SVM and MMS have a very high recall (1.00), which could be due to the margin maximization [55].

The proposed SPFF, in conjunction with ICA, has yielded promising results for up to 15 hours prior to the CA event, as shown in Fig. 6. The figure shows that accuracy decreases when the prediction window approaches 15. Considering the challenges in predicting the future events, this seems to be a reasonable inference. CNN achieves a good 3-hour accuracy score, but its 12-hour accuracy significantly decreases, while its 15-hour accuracy increases. MLP model accuracy decreases in a similar fashion for 3 hours to 6 hours; however, it remains consistent as the time window increases to 12 hours, and then the accuracy decreases as the time window increases to 15 hours. Models and prediction window periods have significant effects on recall counts. RF and SVM have strong recall rates across all the time frames; on the other hand, the MMS and CNN fluctuate. Extending the time frame to 15 hours causes the MMS and CNN recall values to fluctuate, indicating challenges in identifying actual positive rates at longer intervals from the event. F1 score grows from a 3-hour to a 12-hour projection window but drops at a 15-hour window. The F1 Score of the XSM model is more stable over time than other models. All the models have high AUROC values, indicating that most of the models can accurately differentiate between classes. It is observed that the MMS model has the lowest recall and AUROC ratings at all the time stamps. The consistent decline in performance metrics (as the prediction window grows) suggests that models have more relevant data and can make better predictions as the CA event approaches. However, the models can accurately predict CA 15 hours in advance with accuracy, precision, recall, F1 score, and AUROC of 0.962, 0.988, 0.962, 0.962, and 0.997, respectively. This is consistent with our goal of extending the time frame during which we can detect and use the deterioration in vital signs to make accurate predictions. This supports Smith's findings that suggests a five-step "chain of prevention" to help hospitals organize care procedures to avoid and identify clinical decline [56].

B. Empirical Study of Dimensionality Reduction Techniques

To visualize the feature distribution of the data subset 1, we employ ICA, PCA, rbf kernel PCA, poly kernel PCA, and sigmoid kernel PCA; tSNE plots are thus obtained as depicted in Fig. 7. The PCA and ICA plots exhibit a discernible clustering of data points, indicating that both techniques possess the capability of achieving a certain degree of differentiation among the classes. However, the feature distribution by CA event is better distinguished by ICA as compared to the other methods. Furthermore, the rbf kernel PCA appears to have more overlapped data points, thus, a little differentiated distribution. The poly kernel PCA and sigmoid kernel PCA exhibit distinct levels of segregation and dispersion of data points, with the sigmoid kernel PCA demonstrating a significant convergence of the two classes. These results are supplemented by the scores obtained for each dimensionality reduction technique of the empirical study. It is observed that the method having ICA provides superior performance. While the PCA, rbf kernel PCA, and sigmoid kernel PCA yield acceptable performance, the poly kernel PCA technique generates the lowest scores for all metrics across the models. Poly kernel PCA may introduce a higher-dimensional feature space and variance extraction that does not match the multi-scale temporal data distribution. Thus, the multi-scale temporal features do not appear to be gaining valuable feature insights from employing poly kernel PCA. This affirms the rationale of selecting ICA for feature engineering and dimensionality reduction.

C. Empirical Study to Evaluate Impact on Class Imbalance

Addressing class imbalance is crucial for developing models that generalize well across different populations. We evaluate the efficiency of the proposed SPFF in conjunction with ICA and class imbalance using subset 9 that has 5x class imbalance. The results are detailed in Table IV. Additionally, we compare the results tested on the balanced and imbalanced class data subsets (5-9); they are illustrated in Fig. 8. Although the models exhibit a change in performance owing to the degree of imbalance, the results are close to those of the balanced class. The plots in Fig. 8 show that the class imbalance affects the metrics. For instance, the AUROC of the CNN model remains consistent across various levels of class imbalance, indicating that it may possess greater resilience to class imbalance in comparison to other models. This can be attributed to CNN's capability of hierarchical feature extraction [57]. The F1 Score of SVM exhibits a substantial decline when subjected to a 5x imbalance scenario, suggesting that SVM may be susceptible to class imbalance in relation to F1 score. This could be due to the margin optimization approach employed in SVM; the algorithm may favor the majority class in imbalanced scenarios, resulting in a poor recall for the minority class and, consequently, a lower F1 Score [55]. RF, SVM, XGB, and CNN exhibit superior performance on specific measures. The ensemble feature of RF, which utilizes several decision trees for prediction, can enhance its resilience against class imbalance by mitigating the potential impact of variance and bias on individual trees within an unbalanced dataset [58].

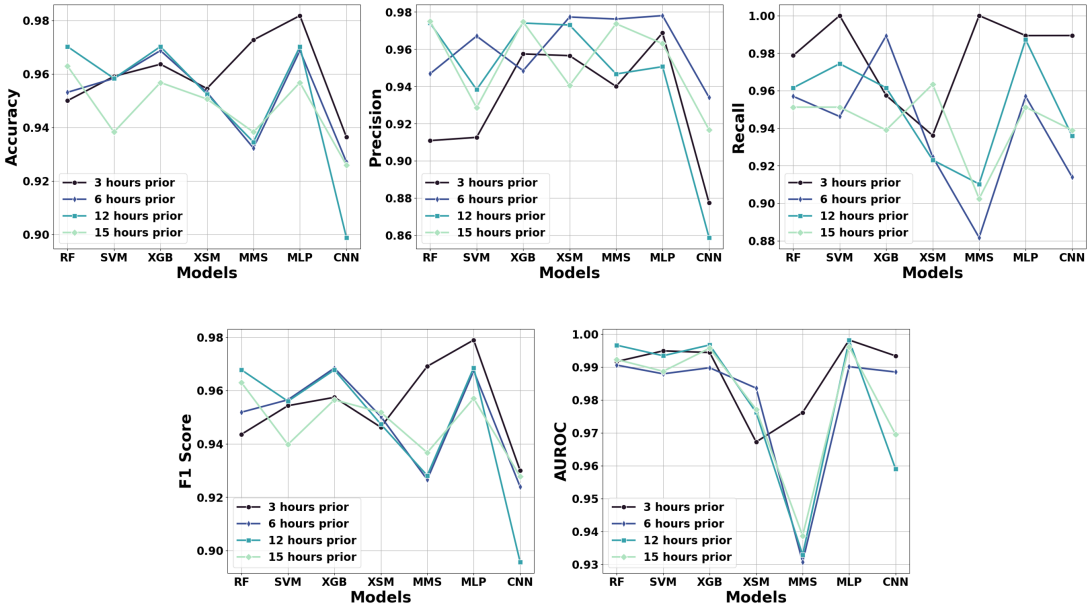


Fig. 6: Results comparing the metrics for all temporal subsets of data for all models.

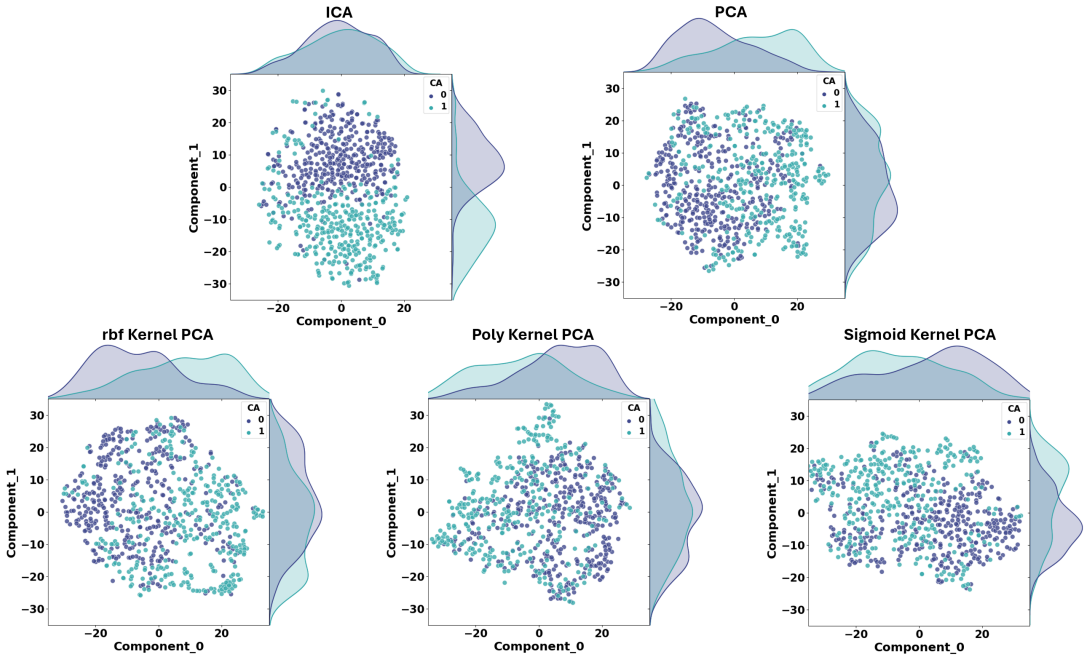


Fig. 7: tSNE projection comparison of the empirical study of dimensionality reduction technqiues.

XGB prioritizes the sequential correction of prior errors and possesses mechanisms that can partially address imbalanced data [59], [60]. Although the models exhibit a decline in performance for subset 9, it still matches the scores in the literature. We find the F1 Score and AUROC on subset 9 to have yielded good results, i.e., the F1-score ranges from 0.846 in the MMS model to 0.917 in the XGB model, and the AUROC ranges from 0.896 in the MMS model to 0.993 in the RF model. Our analysis demonstrates that the proposed SPFF framework, in conjunction with ICA, remains resilient across various levels of imbalance, as evidenced by consistent F1-scores and AUROC values. This robustness ensures the model's applicability to real-world clinical environments where CA cases are rare. The performance across different imbalanced subsets supports the framework's generalizability and reliability for broader clinical deployment.

D. Explainability

SHAP provides insights into the predictions made by different models. For the purpose of simple illustration, we have illustrated the SHAP output for the best-performing ML model and DL model, i.e., the MMS model as depicted in Fig. 9, and the MLP model as depicted in Fig. 10.

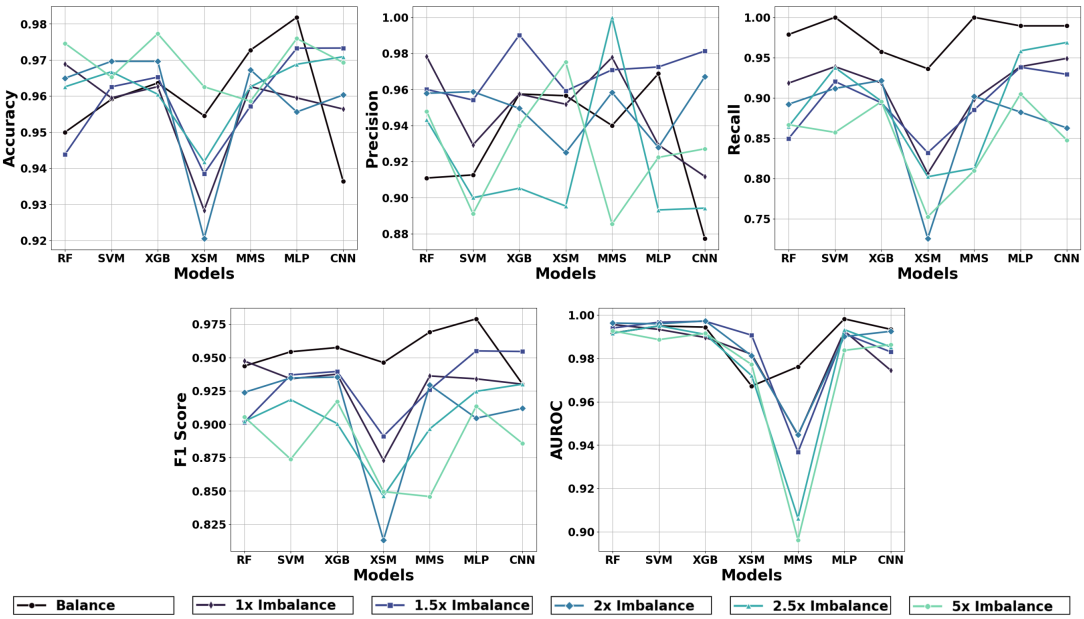


Fig. 8: Comparison of the results obtained using subset 5 -9 to visualize the impact the proposed SPFF and ICA has on class imbalance.

TABLE IV: Results for CA prediction 3 hours prior to the event using ICA reduced SPFF features of an imbalanced class (subset 9).

Model	Accuracy	Precision	Recall	F1 Score	AUROC
RF	0.975	0.948	0.867	0.905	**0.993**
SVM	0.965	0.891	0.857	0.874	0.989
XGB	0.977	0.940	0.895	**0.917**	0.992
XSM	0.963	**0.975**	0.752	0.849	0.977
MMS	0.960	0.885	0.810	0.846	0.896
MLP	**0.980**	0.922	**0.905**	0.913	0.984
CNN	0.969	0.927	0.848	0.886	0.986

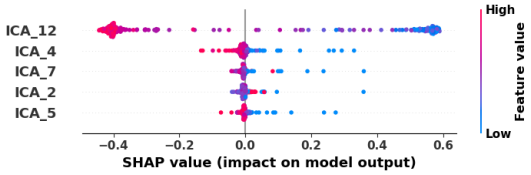


Fig. 9: SHAP output for MMS model.

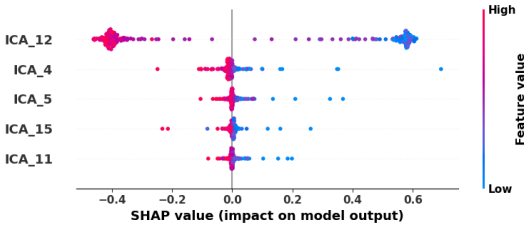


Fig. 10: SHAP output for MLP model.

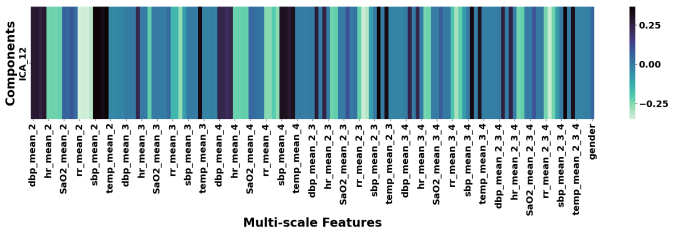


Fig. 11: Heat-map for ICA component 12.

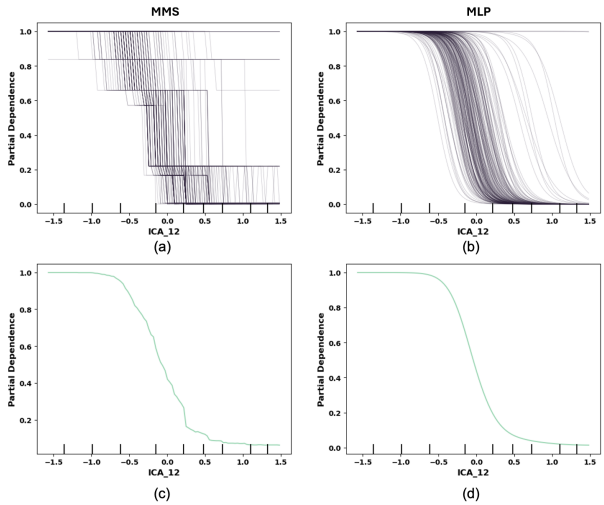


Fig. 12: PDP plots (bottom) and ICE plots (top) from MMS model (left) and MLP model (right).

As we can see from both Fig. 9 and Fig. 10, a single uniform ICA component, ICA 12, consistently contributes the most to the prediction in both models. To evaluate and understand the features that contribute to the ICA 12 component, we have

illustrated a heatmap, as shown in Fig. 11. As we can see from Fig. 11, the multi-scale temporal aggregated features contributing to the component have varying scales of temporal data. For instance, the multi-scale aggregation of systolic blood pressure on a scale of 1, 2, and 3 positively influences ICA 12. The same observation applies to diastolic blood pressure. Nevertheless, heart rate and respiratory rate have a detrimental impact on ICA 12. The occurrence and consequences of the abnormal fluctuation of these essential vital signs before CA have been extensively investigated and documented in literature [48], [31]. The back projection heatmap confirms that ICA component 12 consistently contributes the most to the prediction in both models, which comprise multi-scale temporal features such as systolic and diastolic blood pressure positively, and heart rate along with respiratory rate negatively. Therefore, it is clear that gathering information at multiple scales of measurement reflects the variation of vital signs, which is important in the clinical context. Thereby, reaffirming the importance of local as well as global context of data variability in multivariate time-series data. Additionally, the PDP and ICE plots are plotted for ICA 12 for the MMS model and the MLP model as shown in Fig. 12 to analyze the results obtained through SHAP. The PDP plots in Fig. 12 illustrate the consistent and gradual impact of ICA 12, while taking into account the average values of all other features, on the average anticipated outcome. The PDP displayed in the bottom right (d) corner exhibits a consistently declining correlation, indicating that greater values of ICA 12 are linked to lower projected results. This suggests that there is an inverse relationship between the value of ICA 12 and CA, meaning that when the value of ICA 12 grows, the probability of CA occurrence drops. The PDP in the bottom left (c) corner exhibits a step-like pattern, indicating that the association between ICA 12 and the CA event alters at distinct intervals or values of ICA 12. The ICE plots in Fig. 12 exhibit an identical correlation as PDPs, but they provide a visual representation of the dispersion around the average impact for individual data. The plot in the upper left corner (a) exhibits numerous intersecting lines, which suggests that the CA prediction trajectory of ICA 12 undergoes significant and potentially nonlinear changes for each patient in the dataset. There is a significant amount of variation in the model predictions among different occurrences. The plot in the top right (b) corner exhibits ICE lines with smoothness and forms a cluster of lines that provide the illusion of a surface. Changes in ICA 12 influence the individual CA predictions, but their effect is more consistent and predictable than in the left plot.

The SHAP values provide a detailed breakdown of each ICA-transformed feature contributing to the prediction, offering both local and global interpretability. When combined with PDP and ICE, these techniques provide a comprehensive understanding of the model's behavior. PDP shows that ICA 12 affects CA risk across the entire patient population, while ICE reveals how the individual patient trajectories vary, highlighting critical variations in risk assessment. Moreover, the persistent emphasis on systolic blood pressure, heart rate, and oxygen saturation corresponds with recognized clinical standards and documented risk factors for cardiac arrest. The model's predictions quite match with the literature [48], [31].

When SHAP, PDP, and ICE are combined, a more nuanced and patient-specific insights could be well anticipated as against other methods, such as global feature importance or chronological feature depiction, which could only provide general feature rankings or localized explanations [24]. This is crucial in clinical settings, where understanding the individual risk factors for each patient can make a significant difference in clinical decision-making. For instance, when the model identifies a pronounced elevation in heart rate or a decrease in oxygen saturation as critical risk factors for CA, doctors might implement specific interventions. This depth of understanding could enable the physicians to respond quickly and efficiently. We believe that by implementing these techniques, the predictions could probably be more interpretable and trustworthy.

E. Comparative Analysis

The proposed class-imbalance-resistant methodology is compared to the state-of-the-art methods from literature. The AUROC scores for the state-of-the-art models range from 0.82 [25] to 0.97 [24] that reflect their ability to capture temporal dependencies in vital sign data. The proposed SPFF and ICA framework, on the other hand, has a much higher AUROC value of 0.998, showing it capability for multi-scale temporal aggregation and ability to capture both short- and long-term dependencies. Thus, the SPFF and ICA combination is found effective for timely prediction of CA occurrences, as shown in TableV, where the time-series vital signs and demographic data are well considered. Although Hong et al. [24], Jang et al. [28], and Baral et al. [30], Layeghian et al. [25], Kwon et al. [31], Yijing et al. [26], and Kim et al. [23] show decent results, they lack in multi-scale temporal context of vital signs and explainability, raising suspicion on the usefulness of the approaches. The study by Lee et al. [27] uses ECG measures without vital signs and examines multi-scale contexts; however, only key indicators are monitored in resource-constrained circumstances. Our proposed method, combining SPFF with ICA considers the multi-scale temporal context and extracts the features determining the hidden patterns. SHAP, PDP, and ICE add a complete and intelligible layer to facilitate and comprehend the explainability of the model.

Alongside the seven models already evaluated, including three ML models (RF, SVM, and XGB), two ensemble stacking models (XSM and MMS), and two DL models (MLP and 1D-CNN), we have compared the performance of the proposed SPFF and ICA framework to baseline models like Naive Bayes and K-Nearest Neighbors (KNN). These baseline models provided lower performance metrics, with Naive Bayes and KNN achieving AUROC scores below 0.75, as compared to the higher performance metrics achieved by our proposed framework. This further substantiates the improvements brought by SPFF and ICA in capturing complex temporal dependencies and handling class imbalance. Furthermore, the tSNE plots, as shown in Fig. 3, Fig. 4, and Fig. 5, help visualize the improvements by the proposed SPFF and ICA framework.

TABLE V: Comparative analysis with previous work found in literature.

Reference	Classifier	Feature	Multi-scale	ICA	Explainable	AUROC
Kim et al. [23]	XGB	Multivariate vital signs	Yes	No	Yes (SHAP)	0.87
Hong et al. [24]	RF	Vital signs, demographic data, chief concerns	No	No	Yes (chronological feature depiction)	0.97
Jang et al. [28]	Hybrid ANN	Vital signs, demographic data, chief complaints, level of consciousness	No	No	No	0.936
Layeghian et al. [25]	Stacking XGBoost	Textual EHR data	No	No	No	0.82
Yijing et al. [26]	XGB	Vital signs	Yes	No	Yes (SHAP)	0.94
Baral et al. [30]	MLP, Bi-LSTM	Vital signs, demographic data, chief complaints	No	No	No	0.94
Kwon et al. [31]	RNN	Vital signs	No	No	No	0.85
Lee et al. [27]	LGBM	ECG (heart rate variability)	Yes	No	Yes (SHAP)	0.881
Our method	Multiple models	Multivariate vital signs, demographic data	Yes	Yes	Yes (SHAP, PDP, ICE)	0.998 (MLP)

VI. Conclusion

In this paper, we introduce SPFF, a novel technique that uses multi-scale temporal feature aggregation to accurately capture long-term dependencies and changing distributions in vital signs data. We could simplify the computational process and maximize the effectiveness of the models by integrating ICA. The current study showcases SPFF's resilience to data imbalance when combined with ICA. Additionally, we have incorporated an explainability layer using SHAP, PDP, and ICE to enhance transparency and interpretability. The implementation of AI in healthcare has ethical obligations, especially in guaranteeing model fairness and handling of patient data. To address this issue, we have cross-validated and employed balanced datasets to minimize bias, ensuring the model does not disproportionately impact underrepresented demographic groups. We believe the current approach could be valuable in a clinical decision-support set-up, aiding clinicians in identifying high-risk CA patients while upholding their ultimate decision-making authority. By modifying the feature sets, the SPFF could further be extended to other time-series prediction problems (beyond CA), such as sepsis or stroke. This is because it utilizes multi-scale temporal aggregation for predicting time-sensitive health conditions when early diagnosis is crucial. In the future, we intend to investigate the impact of shorter time intervals (e.g., seconds or minutes) on prediction accuracy. Furthermore, we aim to ascertain its generalizability across clinical environments by investigating the usefulness of federated learning on larger datasets obtained from multiple sources.

Acknowledgement

This publication was made possible by MRC-01-23-367 from the Medical Research Center, Hamad Medical Corporation, Doha, Qatar. The findings herein reflect the work, and are solely the responsibility of the authors. All authors declare that they have no known conflicts of interest in terms of competing financial interests or personal relationships that could have an influence or are relevant to the work reported in this paper. The Open Access funding was provided by the Qatar National Library.

References

[1] "World heart report 2023: Confronting the world's number one killer," *World Heart Federation*, 2023.

[2] L. e. a. Priya, "Early prediction model for coronary heart disease using genetic algorithms, hyper-parameter optimization and machine learning techniques," *Health and Technology*, vol. 11, no. 1, pp. 63–73, 2021.

[3] C. W. T. et al., "Heart disease and stroke statistics—2023 update: A report from the american heart association," *Circulation*, vol. 147, no. 8, pp. e93–e621, 2023. [Online]. Available: https://www.ahajournals.org/doi/abs/10.1161/CIR.0000000000001123

[4] S. P. Dakua and J. S. Sahambi, "Lv contour extraction from cardiac mr images using random walks approach," in *2009 IEEE International Advance Computing Conference*, 2009, pp. 228–233.

[5] ——, "Detection of left ventricular myocardial contours from ischemic cardiac mr images," *IETE Journal of Research*, vol. 57, no. 4, pp. 372–384, 2011. [Online]. Available: https://www.tandfonline.com/doi/abs/10.4103/0377-2063.86338

[6] ——, "Automatic left ventricular contour extraction from cardiac magnetic resonance images using cantilever beam and random walk approach," *Cardiovascular Engineering*, vol. 10, no. 1, pp. 30–43, 2010. [Online]. Available: https://doi.org/10.1007/s10558-009-9091-2

[7] G. D. Perkins and I. G. e. a. Jacobs, "Cardiac arrest and cardiopulmonary resuscitation outcome reports: update of the utstein resuscitation registry templates for Out-of-Hospital cardiac arrest: a statement for healthcare professionals from a task force of the international liaison committee on resuscitation (american heart association, european resuscitation council, australian and new zealand council on resuscitation, heart and stroke foundation of canada, InterAmerican heart foundation, resuscitation council of southern africa, resuscitation council of asia); and the american heart association emergency cardiovascular care committee and the council on cardiopulmonary, critical care, perioperative and resuscitation," *Circulation*, vol. 132, no. 13, pp. 1286–1300, Nov. 2014.

[8] S. P. Dakua and J. S. Sahambi, "Modified active contour model and random walk approach for left ventricular cardiac mr image segmentation," *International Journal for Numerical Methods in Biomedical Engineering*, vol. 27, no. 9, pp. 1350–1361, 2011. [Online]. Available: https://www.onlinelibrary.wiley.com/doi/abs/10.1002/cnm.1430

[9] S. Mohanty and S. P. Dakua, "Toward computing cross-modality symmetric non-rigid medical image registration," *IEEE Access*, vol. 10, pp. 24 528–24 539, 2022.

[10] S. P. Dakua, "Use of chaos concept in medical image segmentation," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 1, no. 1, pp. 28–36, 2013. [Online]. Available: https://doi.org/10.1080/21681163.2013.765709

[11] ——, "Lv segmentation using stochastic resonance and evolutionary cellular automata," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 29, no. 03, p. 1557002, 2015. [Online]. Available: https://doi.org/10.1142/S0218001415570025

[12] ——, "Towards left ventricle segmentation from magnetic resonance images," *IEEE Sensors Journal*, vol. 17, no. 18, pp. 5971–5981, 2017.

[13] ——, "Annularcut: a graph-cut design for left ventricle segmentation from magnetic resonance images," *IET Image Processing*, vol. 8, pp. 1–11(10), 2014. [Online]. Available: https://digital-library.theiet.org/content/journals/10.1049/iet-ipr.2013.0088

[14] ——, "Performance divergence with data discrepancy: A review," *Artificial Intelligence Review*, vol. 40, pp. 429–455, 2013.

[15] G. B. Smith, "Vital signs: Vital for surviving in-hospital cardiac arrest?" *Resuscitation*, vol. 98, pp. A3–4, Oct. 2015.

[16] M. Y. e. a. Ansari, "Dense-psp-unet: A neural network for fast inference liver ultrasound segmentation," *Computers in Biology and Medicine*, vol. 153, p. 106478, 2023.

[17] ——, "Re-routing drugs to blood brain barrier: A comprehensive analysis of machine learning approaches with fingerprint amalgamation and data balancing," *IEEE Access*, vol. 11, pp. 9890–9906, 2022.

[18] V. e. a. Chandrasekar, "Investigating the use of machine learning models to understand the drugs permeability across placenta," *IEEE Access*, 2023.

[19] K. J. W. Tang and C. K. E. A. et al., "Artificial intelligence and machine learning in emergency medicine," *Biocybernetics and Biomedical Engineering*, vol. 41, pp. 156–172, 1 2021.

[20] M. Y. e. a. Ansari, "A lightweight neural network with multiscale feature enhancement for liver ct segmentation," *Scientific Reports*, vol. 12, no. 1, p. 14153, Aug 2022. [Online]. Available: https://doi.org/10.1038/s41598-022-16828-6

[21] S. M. L. et al., "Explainable artificial intelligence model to predict acute critical illness from electronic health records," *Nature Communications*, vol. 11, 12 2020.

[22] M. e. a. Farid, "Scheduling scientific workflow in multi-cloud: A multi-objective minimum weight optimization decision-making approach," *Symmetry*, vol. 15, no. 11, p. 2047, 2023, copyright - © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/). Notwithstanding the ProQuest Terms and Conditions, you may use this content in accordance with the terms of the License; Last updated - 2023-11-25. [Online]. Available: https://www.proquest.com/scholarly-journals/scheduling-scientific-workflow-multi-cloud/docview/2893318747/se-2

[23] Y. K. K. et al., "Explainable artificial intelligence warning model using an ensemble approach for in-hospital cardiac arrest prediction: Retrospective cohort study," *Journal of Medical Internet Research*, vol. 25, 2023.

[24] S. H. et al., "Prediction of cardiac arrest in the emergency department based on machine learning and sequential characteristics: Model development and retrospective clinical validation study," *JMIR Medical Informatics*, vol. 8, 8 2020.

[25] S. L. Javan and M. M. Sepehri, "A predictive framework in healthcare: Case study on cardiac arrest prediction," *Artificial Intelligence in Medicine*, vol. 117, 7 2021.

[26] L. Y. et al., "Prediction of cardiac arrest in critically ill patients based on bedside vital signs monitoring," *Computer Methods and Programs in Biomedicine*, vol. 214, 2 2022.

[27] H. L. et al., "Real-time machine learning model to predict in-hospital cardiac arrest using heart rate variability in icu," *npj Digital Medicine*, vol. 6, 12 2023.

[28] D. H. J. et al., "Developing neural network models for early detection of cardiac arrest in emergency department," *American Journal of Emergency Medicine*, vol. 38, pp. 43–49, 1 2020.

[29] A. R. et al., "Prediction of major adverse cardiac events in the emergency department using an artificial neural network with a systematic grid search," *International Journal of Emergency Medicine*, vol. 17, 12 2024.

[30] S. B. et al., "A novel solution of using deep learning for early prediction cardiac arrest in sepsis patient: enhanced bidirectional long short-term memory (lstm)," *Multimedia Tools and Applications*, vol. 80, pp. 32 639–32 664, 9 2021.

[31] J. M. K. et al., "An algorithm based on deep learning for predicting in-hospital cardiac arrest," *Journal of the American Heart Association*, vol. 7, 7 2018.

[32] G. B. S. et al., "The ability of the national early warning score (news) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death," *Resuscitation*, vol. 84, pp. 465–470, 4 2013.

[33] Y. J. L. et al., "A multicentre validation study of the deep learning-based early warning score for predicting in-hospital cardiac arrest in patients admitted to general wards," *Resuscitation*, vol. 163, pp. 78–85, 6 2021.

[34] F. E. S. et al., "Deep interpretable early warning system for the detection of clinical deterioration," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, pp. 437–446, 2 2020.

[35] V. e. a. Gupta, "R-peak detection based chaos analysis of ecg signal," *Analog Integrated Circuits and Signal Processing*, vol. 102, no. 3, p. 479 – 490, 2020, cited by: 60. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85074443638&doi=10.1007%2fs10470-019-01556-1&partnerID=40&md5=c66164700c046267c79d59de01abde66

[36] M. H. e. a. Othman, "Resting-state nirs–eeg in unresponsive patients with acute brain injury: A proof-of-concept study," *Neurocritical Care*, vol. 34, no. 1, p. 31 – 44, 2021, cited by: 29. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85084141011&doi=10.1007%2fs12028-020-00971-x&partnerID=40&md5=02209ad3b03cc322c7fbb7edfcb009b5

[37] L. e. a. Norton, "Disruptions of functional connectivity in the default mode network of comatose patients," *Neurology*, vol. 78, no. 3, p. 175 – 181, 2012, cited by: 148. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84863267602&doi=10.1212%2fWNL.0b013e31823fcd61&partnerID=40&md5=84a0bec95d15e9cfbe89eea6462634a3

[38] A. e. a. Singh, "Investigating tattoo pigments composition with uv-vis and ft-ir spectroscopy supported by chemometric modelling," *Current Analytical Chemistry*, vol. 20, pp. 1–17, 07 2024.

[39] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, no. 4, pp. 411–430, 2000. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608000000265

[40] M. e. a. Greenacre, "Principal component analysis," *Nature Reviews Methods Primers*, vol. 2, no. 1, p. 100, 11 2022.

[41] G. e. a. Van den Broeck, "On the tractability of shap explanations," *Journal of Artificial Intelligence Research*, vol. 74, pp. 851–886, 2022.

[42] X. Z. et al., "Factors affecting traffic risks on bridge sections of freeways based on partial dependence plots," *Physica A: Statistical Mechanics and its Applications*, vol. 598, p. 127343, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378437122002709

[43] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[44] A. G. et al., "Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation," *Journal of Computational and Graphical Statistics*, vol. 24, no. 1, pp. 44–65, 2015. [Online]. Available: https://doi.org/10.1080/10618600.2014.907095

[45] A. V. e. a. Singh, "Advancing predictive risk assessment of chemicals via integrating machine learning, computational modeling, and chemical/nano-quantitative structure-activity relationship approaches," *Advanced Intelligent Systems*, vol. 6, no. 4, p. 2300366, 2024. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/aisy.202300366

[46] A. e. a. Johnson, "Mimic-iv (version 1.0)," 2020.

[47] I. J. e. a. Brekke, "The value of vital sign trends in predicting and monitoring clinical deterioration: A systematic review," *PLoS One*, vol. 14, no. 1, p. e0210875, Jan. 2019.

[48] L. W. e. a. Andersen, "The prevalence and significance of abnormal vital signs prior to in-hospital cardiac arrest," *Resuscitation*, vol. 98, pp. 112–117, Sep. 2015.

[49] P. Domingos, "The role of occam's razor in knowledge discovery," *Data mining and knowledge discovery*, vol. 3, pp. 409–425, 1999.

[50] J. S. et al., "Ctmlp: Can mlps replace cnns or transformers for covid-19 diagnosis?" *Computers in Biology and Medicine*, vol. 159, p. 106847, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482523003128

[51] R. Shwartz-Ziv and A. Armon, "Tabular data: Deep learning is not all you need," *Information Fusion*, vol. 81, pp. 84–90, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253521002360

[52] M.-H. G. et al., "Can attention enable mlps to catch up with cnns," *Computational Visual Media*, vol. 7, no. 3, pp. 283–288, 2021.

[53] I. Shavitt and E. Segal, "Regularization learning networks: deep learning for tabular datasets," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[54] A. e. a. Botalb, "Contrasting convolutional neural network (cnn) with multi-layer perceptron (mlp) for big data analysis," in *2018 International Conference on Intelligent and Advanced System (ICIAS)*, 2018, pp. 1–5.

[55] S. Alwajidi and L. Yang, "Multiresolution hierarchical support vector machine for classification of large datasets," *Knowledge and Information Systems*, vol. 64, no. 12, pp. 3447–3462, Dec. 2022.

[56] G. B. Smith, "In-hospital cardiac arrest: Is it time for an in-hospital 'chain of prevention'?" *Resuscitation*, vol. 81, no. 9, pp. 1209–1211, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0300957210002455

[57] S. e. a. Albawi, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*, 2017, pp. 1–6.

[58] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[59] Z. e. a. Ali, "Exploring the power of extreme gradient boosting algorithm in machine learning: a review," vol. 12, pp. 320–334, 05 2023.

[60] M. e. a. Bukowski, "Custom loss functions in XGBoost algorithm for enhanced critical error mitigation in Drill-Wear analysis of Melamine-Faced chipboard," *Sensors (Basel)*, vol. 24, no. 4, Feb. 2024.