THE ANATOLIAN JOURNAL OF CARDIOLOGY

Official journal of the TURKISH SOCIETY OF CARDIOLOGY

Integrative Identification of Frailty Status in Elderly Patients with Atrial Fibrillation Using a Noise-Resilient Multimodal Machine Learning Approach

ORIGINAL INVESTIGATION

ABSTRACT

Background: Frailty is common yet underdiagnosed in elderly patients with atrial fibrillation (AF), worsening outcomes and complicating treatment. Traditional assessments are time-consuming and subjective, while physiological monitoring offers potential for automated detection. Combining machine learning (ML) with multimodal data, especially electrocardiogram (ECG) and clinical features, may improve frailty identification, even in noisy real-world conditions. To develop a noise-resilient ML framework for identifying frailty status in elderly patients with AF using multimodal clinical and ECG data and to compare traditional and deep learning models under varying signal conditions.

Methods: This retrospective study included 110 patients aged ≥65 with documented AF. Frailty was assessed via a composite of the Fried Phenotype and Clinical Frailty Scale. Electrocardiogram data (resting 12-lead and Holter) were processed through a denoising and segmentation pipeline. Over 180 ECG features and clinical parameters were used as inputs. Five models (Random Forest, Extreme Gradient Boosting, 1-dimensional convolutional neural network, bidirectional long short-term memory-attention, SiamAF) were trained and evaluated using an 80:20 split. Performance metrics included accuracy, F1-score, receiver operating characteristic-area under the curve (ROC-AUC), and Brier score; robustness was tested with synthetic noise.

Results: Frailty and pre-frailty prevalence were 41.82% and 34.55%. The SiamAF achieved the best performance (accuracy 90.00%, F1-score 89.75%, ROC-AUC 93.00%, Brier score 0.084), maintaining robustness under noise. Deep learning models also showed strong performance (ROC-AUC > 90%). Key predictors included standard deviation of normal-to-normal (NN) intervals, corrected QT interval, N-terminal pro-brain natriuretic peptide, and grip strength.

Conclusion: Multimodal ML effectively identifies frailty status in elderly patients with AF. The SiamAF model demonstrated strong discriminatory performance and noise robustness in this single-center cohort; however, these findings should be considered hypothesis-generating and require external validation in larger, multicenter AF populations before clinical adoption. Given the limited sample size relative to the high-dimensional feature space, these results should be interpreted as exploratory and hypothesis-generating. The present framework primarily leverages predefined ECG-derived features rather than fully end-to-end raw-signal learning.

Keywords: Atrial fibrillation, ECG, frailty, machine learning, noise-resilient models

Shangying Hu[1]

Chao Feng[2]

[1]*Department of Nursing, The Fourth Affiliated Hospital of School of Medicine, and International School of Medicine, International Institutes of Medicine, Zhejiang University, Yiwu, China*
[2]*Department of Cardiology, The Fourth Affiliated Hospital of School of Medicine, and International School of Medicine, International Institutes of Medicine, Zhejiang University, Yiwu, China*

Corresponding author:
Chao Feng
✉ Shang89935901@hotmail.com

Received: December 11, 2025
Accepted: March 31, 2026
Available Online Date: June 15, 2026

Cite this article as: Hu S, Feng C. Integrative identification of frailty status in elderly patients with atrial fibrillation using a noise-resilient multimodal machine learning approach. *Anatol J Cardiol.* 2026;XX(X):X-X.

DOI: 10.14744/AnatolJCardiol.2026.5642

INTRODUCTION

The main sustained cardiac arrhythmia of atrial fibrillation (AF) is currently a major public health problem, especially among older populations. It has been reported that the epidemiological data show that the prevalence of AF is more than 10% in persons above the age of 80 years, and the prevalence has been increasing along with the increase in life expectancy. The condition is not linked to increased risks of stroke and systemic embolism only, as they are accompanied by an increased susceptibility to functional impairment, falls, and dependence; in particular, they become more serious in terms of functional impairment when combined with frailty. The syndrome of frailty, with its physiologic deficit being cumulative and stress tolerance minimal, contributes significantly to both morbidity and mortality related to AF. The interaction of these 2 leads to a vicious cycle in which AF increases frailty via compromised hemodynamics, neurohormonal imbalance, and polypharmacy, whereas frailty predisposes to poor outcomes of procedures and arrhythmia complications. Although there is increased awareness of this complicated interaction, it is poorly represented in the majority of clinical predictive algorithms, which are very specific toward cardiovascular risk without considering the limitations of physiological reserve or functional capacity. Thus, constructing integrative assessment models, including multimodal data, primarily, electrocardiogram (ECG)-based signals, vital trends, and other clinical indicators, is essential. This strategy presents an opportunity for timely and individualized intervention that encourages long-term outcomes and quality of life in aging patients with AF.[1-3]

Frailty itself is a multidimensional syndrome characterized by reduced physiologic reserve, which includes tissue deficit in strength, mobility, cognitive function, nutrition, and homeostasis. It has conventionally been assessed on the basis of phenotype models such as the Fried criteria or index models such as the Rockwood frailty index.[4-6] Although these tools are well validated, they rely heavily on in-person physical evaluations, grip strength, and detailed questionnaires; thus, they may not act as practical tools for frequent use in most clinical environments. Additionally, these tools not sensitive enough to detect the minor physiological alterations in the early phases; therefore, opportunities for proactive intervention may be missed. In this regard, the emergence of machine learning (ML) techniques, in a specific deep learning architecture able to adapt to complex time-based data, has opened the possibility of automating fractional detection with passive physiological monitoring. The adaptation of the ECG-based parameters, such as heart rate variability (HRV), P-wave morphology, QT dynamics, and signal entropy, which reflect autonomic tone and cardiovascular resilience, is among the most promising approaches. These biomarkers tend to show premature variations in frail people even prior to the obvious functional decline. Machine learning models can construct a high-dimensional representation of frailty risk when combined with structured clinical parameters such as age, comorbidity burden, medication use, and functional scores. Longitudinal studies have demonstrated that this integrated approach significantly improves the predictive performance of frailty models over conventional tools, allowing earlier identification and stratification of at-risk elderly populations.[7-10]

HIGHLIGHTS

- Frailty and pre-frailty were highly prevalent (76.36%) among elderly patients with atrial fibrillation (AF) using a composite of Fried Phenotype and Clinical Frailty Scale.
- A multimodal machine learning framework integrating >180 electrocardiogram (ECG)-derived features with clinical, laboratory, echocardiographic, and functional data was developed for frailty risk prediction.
- The SiamAF dual-stream ECG + photoplethysmography model achieved the best performance (accuracy 90.00%, receiver operating characteristic-area under the curve 93.00%, Brier score 0.084) and maintained superior robustness under increasing synthetic noise levels.
- Key predictors of frailty included heart rate variability indices (e.g., standard deviation of normal-to-normal intervals), corrected QT interval duration, N-terminal pro-brain natriuretic peptide, serum albumin, and grip strength, linking autonomic dysfunction, cardiac strain, and physical frailty.
- The proposed noise-resilient models support scalable, wearable, and remote frailty screening in real-world AF populations and warrant validation in larger multicenter cohorts.

Nevertheless, one of the critical technical challenges is still present in spite of these achievements: the ECG records are subjected to noise, which most ML algorithms do not tolerate well, at least not when applied in real-world (or ambulatory or wearable) conditions. In addition, motion, signal dropout, lead displacements, electromyographic interference, and baseline drift are common causes of artifacts in outpatient and home ECG recordings, unlike the well-controlled clinical setting. Such distortions have the capability to puzzle feature extraction and augment the performance of even the most advanced ML models. This is of particular concern in geriatrics in which tremor (frailty associated), postural instability, and arrhythmic symptoms may additionally increase the variability of the signal. To cope with this, most modern research has introduced robust algorithmic methods that seek to filter and correct the noisy data before they are ingested by the model.[11-13] These can be the adaptive filtering techniques to dynamically adapt to nonstationary noise and wavelet-based denoising that have the ability to leave the morphology of the signal intact, and the attention masking methods that ignore segments that are less confident during training.[14-16] New model architectures have been formulated in parallel with specific goals of increasing robustness to noisy conditions. Interestingly, ECGNET employs attention-weighted filtering based to reject any irrelevant noise, ArNet-ECG employs the residual learning of the hierarchical architectures to preserve the relevant features, and SiamAF employs contrastive learning, the parallel inputs of the ECG and photoplethysmography (PPG) to impose noise-invariant representations.[15,16,17] Individually as well as mutually, these strategies are essential steps on the path toward the development of frailty-AF risk predictors applicable in practical situations and capable of acting consistently in various clinical and community contexts.

The realization of hemodynamic interaction and vulnerability within the systems is another major aspect in the integration of frailty and AF. The unrelated behavior of the ventricular response in AF creates inconsistencies in stroke volume, which causes subsequent poor perfusion in the vital organs, such as the brain, muscles, and kidneys.[18-21] This adds to tiredness, sarcopenia, orthostatic intolerability, and mental misfunctioning, characteristics of the frailty phenotype. At the same time, baroreflex problems that are associated with frailty also increase the vulnerability of aged people to the symptomatic and hemodynamic impacts of AF.[22] As an example, weak people can have an unusual course of symptoms including delirium or falls instead of palpitations, which delay the diagnosis and treatment. Further, they have an increased risk of procedural complications, bleeding caused by anticlotting drugs, and nonadherence to therapeutic recommendations caused by polypharmacy or changes in cognition.[23-25] This abidirectional relationship in terms of pathophysiology calls into the need for the creation of predictive models, which would not be limited by 1 single domain input. Good models need to combine the hemodynamic parameters, signs of frailty, patterns of comorbidities with real-time biosignals to reflect the dynamic interaction between vagaries of cardiac rhythm and loss of systemic functionality. To this end, multidomain data containing ECGs, echocardiograms, blood biomarkers, history of medication, and patient-reported functions are indispensable. Also, the use of the expertise of geriatrics, cardiology, neurology, and biomedical engineering leads to the interdisciplinary design of models that are more realistic of the true complex nature of physiology in the aging man.[26-28] With the help of such integrative models, adverse outcomes can not only be predicted more reliably but also entail the development of more individualized methods of management (because they depend on the peculiarities of needs and restrictions of frail and elderly patients with AF).

The further development of remote sensing, eHealth devices, and mobile ECG recorders has further democratized the collection of data, realizing in real time the possibility to screen frailty outside the clinical environment. Continuous monitoring studies have been able to find fine variations in the dynamics of the heart rate and abnormalities in the heart rhythm that herald the occurrence of frailty transitions.[29-30] The combination of ECG and the assessment of blood pressure variability, sleep, and activity allows a more comprehensive physiological picture to be created, making it possible to actively plan care in high-risk elderly patients.[31-34] Noteworthy, an example of AF detection and risk stratification tools based on ML is PhysioNet-derived or commercial wearable datasets, which are being translated to use in concurrent populations of frailty risk profiling.[35] When trained using heterogeneous data sets that include elderly people living in community settings, these models hold potential to classify who is resilient versus pre-frail versus frail with respect to their trajectories, thus intervening in a timely fashion. In this context, ECG-derived features are explored as complementary physiological markers that may enhance scalability and automation of frailty screening, rather than replacing established clinical, laboratory, or echocardiographic assessments.

Despite increasing recognition of the bidirectional relationship between AF and frailty, frailty assessment remains inconsistently integrated into routine AF management, largely due to time constraints and reliance on in-person evaluations. Automated, physiology-based frailty detection using multimodal ECG and clinical data may offer a scalable alternative, particularly in elderly patients frequently monitored for arrhythmia care. However, real-world ECG recordings in AF are inherently noisy, irregular, and heterogeneous, posing challenges for conventional ML approaches. The present study therefore aims to develop and evaluate a noise-resilient multimodal ML framework for frailty prediction in elderly AF patients. This work is intended as a hypothesis-generating investigation, assessing feasibility, robustness, and clinical interpretability rather than definitive clinical deployment.

METHODS

Study Design

This was a retrospective, observational, ML-based diagnostic modeling study conducted over a period of 18 months at a tertiary care academic medical center. The study was approved by the Institutional Ethics Committee, and written informed consent was obtained from all participants or their legally authorized representatives prior to enrollment. The research adhered to the principles outlined in the Declaration of Helsinki. A total of 110 elderly patients aged ≥65 years with a documented diagnosis of AF were consecutively enrolled from the outpatient and inpatient cardiology and geriatrics units.

Inclusion Criteria

- Age ≥65 years
- Confirmed diagnosis of AF on 12-lead ECG or Holter monitoring
- Ability to undergo frailty assessment and ECG testing
- Availability of recent clinical and laboratory records

Exclusion Criteria

- Advanced cognitive impairment (e.g., Mini-Mental State Examination <18) limiting reliable frailty assessment
- Terminal illness with life expectancy <6 months
- Presence of implanted pacemaker or defibrillator interfering with ECG readings
- Incomplete data records or withdrawal of consent

Frailty Assessment

The frailty status of elderly patients enrolled was closely measured at baseline with a 2-method study that combined the calculation of both objective phenotypic and a scale with clinical comprehension to create better diagnostic sensitivity and detail. The physical frailty indicator—a validated and well-used model Fried Frailty Phenotype was utilized to determine the physical indicators of frailty. This phenotype consists of 0-5 core items including (1) unintentional weight loss, which is more than 4.5 kg or an increase of over 5% of body weight over the last 12 months, (2) Canadian Experts Consensus Definition self-reported fatigue, which is measured using validated items of the Center for Epidemiologic Studies Depression scales, (3) low physical activity, which is measured through the Minnesota Leisure Time Activity Questionnaire, (4) slowness, which is ascertained using 4-meter gait speed, and (5) muscle weakness. The participants who fulfilled 1 or more of the above rules became frail, those with 2 or 3 criteria were pre-frail, and those with none were robust.

To expand the phenotypic classification and represent a more comprehensive manifestation of functional decline, there was also a modified presence of Clinical Frailty Scale (CFS). A 9-point ordinal scale assesses frailty based on clinical assessment of mobility, burden of comorbidity, cognitive ability, and activity of the day in terms of dependability. Trained geriatric clinicians who were blinded to ECG data and ML results used the scale to eliminate bias. Patients with CFS scores of 1-3 (very fit to managing well), 4 (vulnerable), and 5 or higher (mildly to severely frail) were classified as robust, pre-frail, and frail, respectively. When the Fried phenotype and CFS scores did not agree in classification, the discrepancy was resolved through consensus adjudication by 2 senior clinicians. The hybrid assessment approach resulted in a more objective ground truth for frailty to be used in the subsequent training and testing of ML models.

Electrocardiogram Data Collection and Processing

Each of the enrolled participants was subjected to a typical 12-lead resting ECG recording. The duration of the recorded data was at least 10 seconds and measured in a supine position in calm and ambient surroundings to guarantee permanence of the signal and limit contamination. A high-resolution digital electrocardiograph that is set to a 500 Hz sampling rate with 16-bit amplitude resolution was used to capture the ECG data and provides a high fidelity of the waveform data that is ideally computed. The placement of the electrodes was according to standard clinical practices, thus there would be reproducibility among the participants and there would be less variation about the placement of the leads.

A subsample of 70 patients was used to facilitate 24 hours of Holter monitoring of dynamic cardiac behavior via compact and portable ECG monitors. The long recordings were able to record HRV beat to beat, circadian autonomic patterns, and paroxysmal arrhythmias which would not be recorded in long resting ECGs. A combination of short-term as well as long-duration ECG data offered a more complete description of electrical cardiac performance in older patients with AF (Figure 1).

Signal Preprocessing Pipeline

A multistage signal preprocessing framework was designed to ensure the integrity and analytical value of ECG signals used in the ML pipeline. In the first place, high-pass filtering with a cutoff frequency of 0.5 Hz rectified baseline wander that usually occurs during respiration or movement of a patient. Then, power line interference caused by environmental electrical noise around 50/60 Hz was reduced through a narrow-band notch filter without loss of important spectral elements of the heart signal. In order to increase the level of signal clarity further, processes of de-noising were also provided, which were performed using wavelet-based techniques. This included multilevel discrete wavelet transforms filters to appropriately filter high-frequency (HF) noise except for the structures' morphology important in diagnosis (i.e., P wave, QRS complex, and T wave). After denoising, R-peaks were identified and the heartbeat segmented employing the Pan-Tompkins algorithm, which is a proven method due to its reliability and is computationally efficient. The presence of R-peaks was identified as anchors to slice and synchronize heartbeat windows as standardized inputs of the same model.

The Data Quality Index (DQI) was subsequently calculated based on a product of signal-to-noise ratio, RR interval consistency, and waveform completeness, which were were evaluated for each segmented ECG window. The model training and evaluation datasets comprised only segments with a DQI that was beyond the currently specified threshold. The segments that do not match the quality requirement have also been discarded to avoid weakening the performance of the conventional and noise-immune ML models. Such a demanding preprocessing pipeline was necessary to ensure that the ECG data used as inputs, both in clinic and ambulatory settings, were of adequate quality to support predictive modeling in elderly AF patients.

Multimodal Feature Extraction

In this work, a multimodal extensive feature set was created to detect both electrophysiological and clinical factors associated with frailty in elderly patients with AF. Besides the parameters extracted based on the ECG, the participants were assessed using clinically relevant data based on their electronic health records. Among them were the core demographic parameters, including age, sex, and body mass index (BMI), which are known to be stable factors of cardiovascular risk stratification. The presence of such comorbidities as hypertension, diabetes mellitus, chronic kidney disease (CKD), chronic obstructive pulmonary disease (COPD), along with a history of stroke, were recorded systematically. Medical histories and their medication profiles were also assessed, paying special concern to the medications known to affect cardiovascular and autonomic activity, namely, anticoagulants, beta-blockers, angiotensin-converting enzyme (ACE) inhibitors, and diuretics (Figure 2).

The feature set included laboratory tests conducted earlier than 2 weeks post-enrollment, which included hemoglobin levels, serum creatinine, N-terminal pro-brain natriuretic peptide (NT-proBNP), and serum albumin, all of which are associated with frailty or other cardiovascular degradation. Also, the echocardiographic index, like the size of the left atrium, left ventricular (LV) ejection fraction, and LV mass index, were included to represent structural and functional heart remodeling. Out of the ECG recordings, over 180 quantitative measures were retrieved through a combination of nonlinear, time-domain, and frequency-domain signal processing procedures. The time-domain HRV measures were the standard deviation of NN intervals (SDNN), the root mean square of successive differences (RMSSD), and the ratio of the low-frequency (LF) to HF power distribution. Electrophysiological dispersion was captured by morphological parameters that were calculated: QRS duration, P-wave duration, and corrected QT interval (QTc). Additional nonlinear indices, such as sample entropy and fractal dimension, were calculated to evaluate the complexity and irregularity of the signal that demonstrated association, in the past, with autonomic dysfunction and with the loss of signal complexity with age.

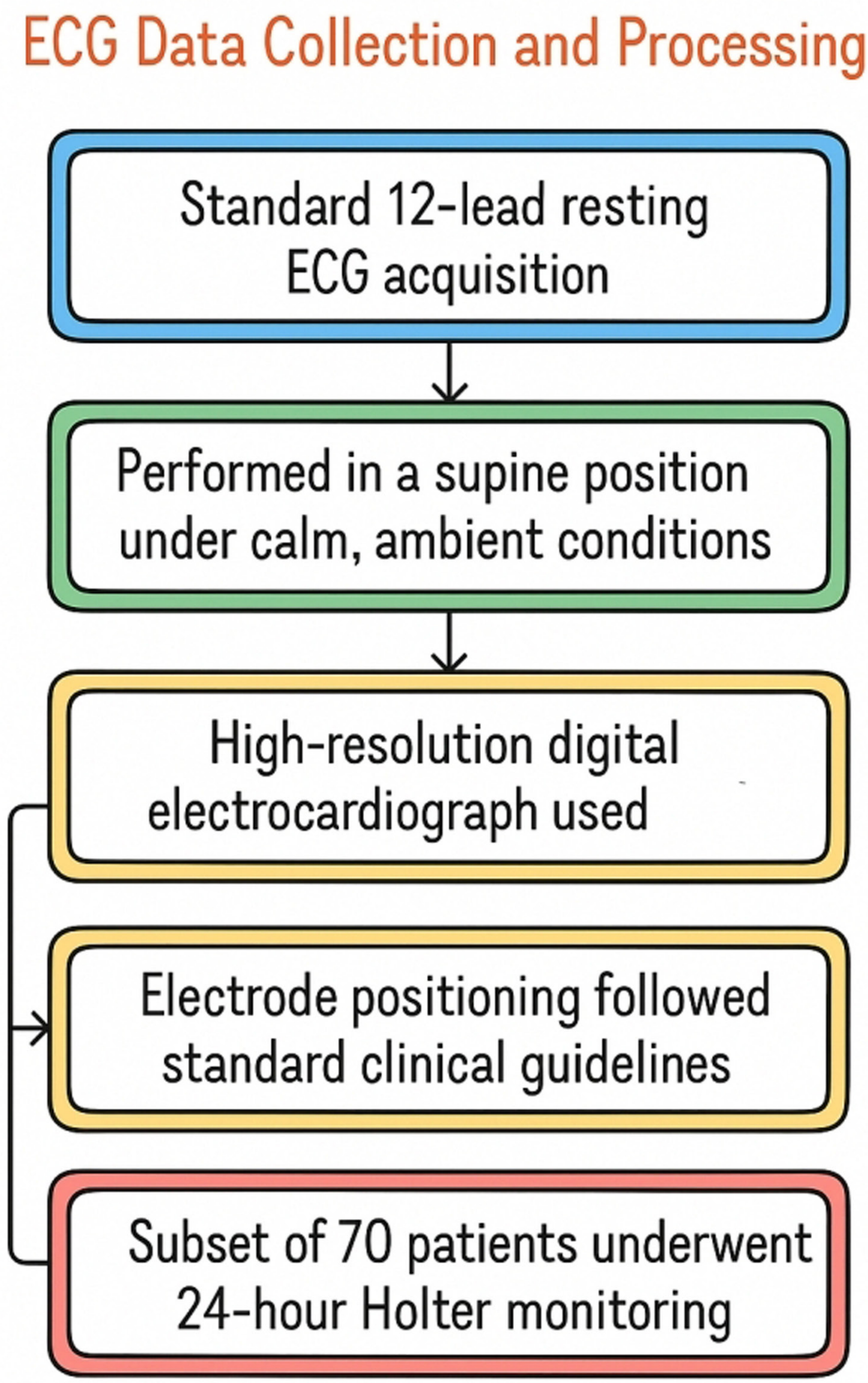

Figure 1. Electrocardiogram data collection and processing.

To mitigate the risk of overfitting associated with high-dimensional feature spaces, feature importance ranking and model-internal regularization were applied during training. No manual feature selection was performed prior to model development, and all features were treated as candidate predictors within regularized learning frameworks.

This study employed predefined ECG-derived features based on established time-domain, frequency-domain, and nonlinear metrics. End-to-end learning directly from raw ECG waveforms was not the primary objective of the present work. Consequently, the model may not capture latent or previously unrecognized ECG patterns associated with

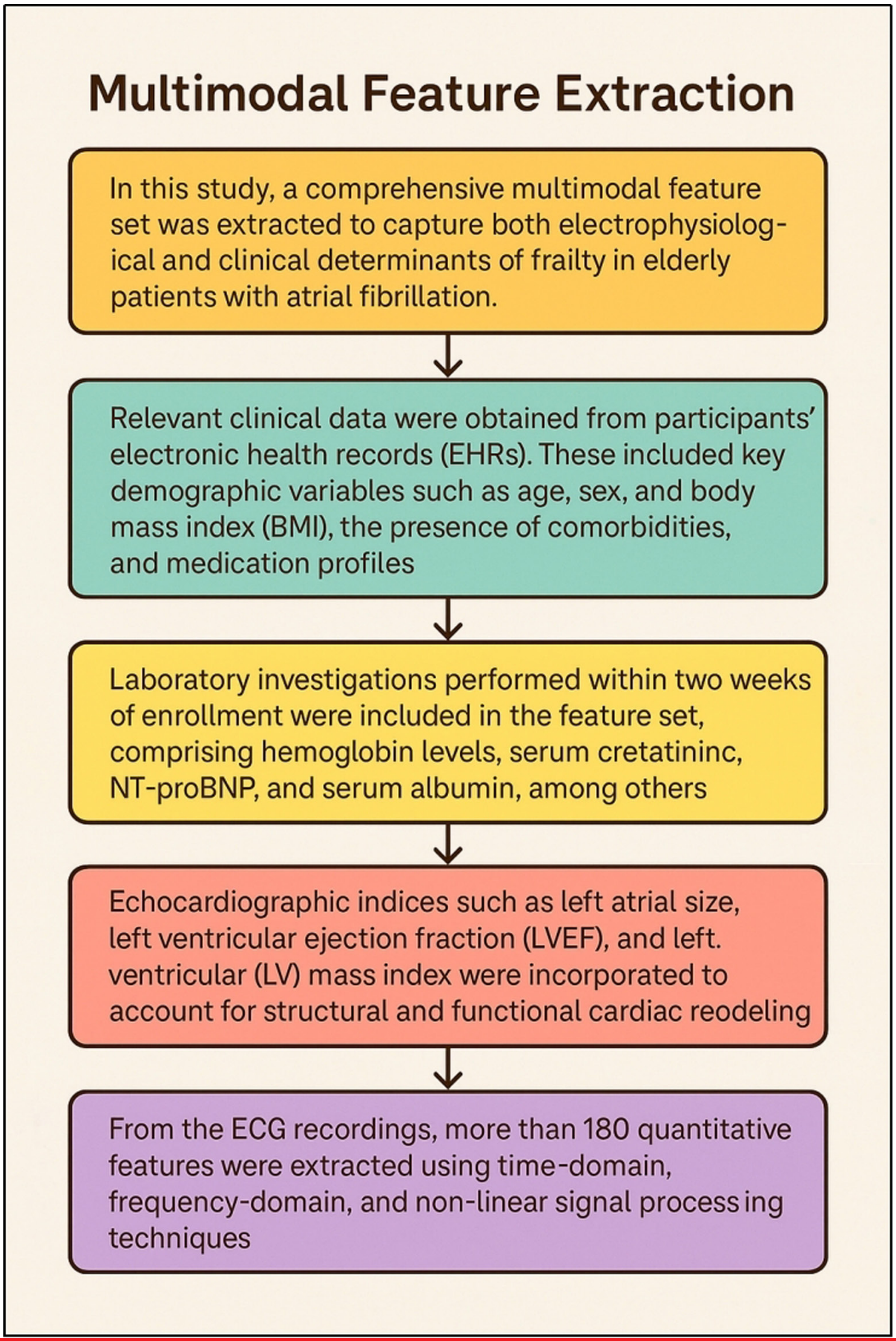

Figure 2. Multimodal feature extraction.

frailty that could potentially be learned through fully data-driven deep learning approaches.

Machine Learning Model Development

The remaining dataset of multimodal features was further randomly divided into a training set (80%, n=88) and a test set (20%, n=22), and the stratified sampling was done to maintain the proportions of frailty status in both sets. A number of supervised ML models were used to classify frailty, both conventional and deep learning ones. These were Random Forest (RF), Extreme Gradient Boosting (XGBoost) in the form of ensemble-based tree modeling, 1-dimensional convolutional neural network (1D-CNN) in the form of processing the temporal specifics of the ECG, and a bidirectional long short-term memory (BiLSTM) network with attention modules in the form of extracting sequential dependencies. A noise-robust variant of SiamAF, a dual-stream neural network based on ECG-PPG fusion, was also conceived and evaluated in a sub-cohort of patients with concurrent PPG data. In order to enhance the model resistance to the label noise, particularly in the case of frailty categories, which are prone to being assessed subjectively, multiple noise-suppression methods were combined. Such features included generalized cross-entropy loss functions to stabilize training in the context of uncertain labels, mixup data augmentation to regularize feature space, label smoothing to avoid overfitting, and early stopping of training based on the performance of the validation area under the curve (AUC). The grid search and 5-fold cross-validation in the training dataset were used to tune model hyperparameters. The unobserved test set was used to test the final model performance to guarantee generalizability. Despite these measures, the ratio of predictors to sample size remains a limitation of the present study and may affect model stability and reproducibility. Although convolutional and recurrent architectures were evaluated, these models were applied to processed ECG representations rather than fully raw, unsegmented ECG waveforms.

Outcome Measures

The major measure of model development was the binary classification of frailty, and therefore frailty was classified as frail or non-frail, depending on the composite frailty measure built earlier. The primary objective of this study was cross-sectional classification of frailty status at the time of assessment, rather than prediction of future frailty progression or clinical outcomes. The secondary outcomes were multiclass frailty prediction in 3 categories alone; namely, robust, pre-frail, and frail. Frailty was analyzed primarily as a binary outcome (frail vs. non-frail) to ensure adequate statistical power and model stability given the limited sample size and class imbalance. A secondary multiclass analysis distinguishing robust, pre-frail, and frail states was additionally performed to explore clinical gradations of vulnerability. Moreover, the installation of synthetic ECG noise on the model performance was also investigated to estimate the possibility of real-life deployment to ambulatory environments. Interpretability of the models was evaluated with the help of Shapley Additive Explanations (SHAP) that indicated the influence of each feature on final model predictions.

Statistical Analysis

All statistical analyses were performed using IBM Statistical Package for the Social Sciences Statistics for Windows, Version 27.0 (IBM Corp., Armonk, NY, USA) and R software, Version 4.3.1 (R Foundation for Statistical Computing, Vienna, Austria). Machine learning model development and evaluation were conducted using Python Version 3.9 within the Anaconda distribution (Anaconda Inc., Austin, TX, USA), utilizing the following libraries:

- Scikit-learn (Version 1.2.2; Google LLC, Mountain View, CA, USA) for classical ML models and performance metrics
- TensorFlow (Version 2.11; Google Brain Team, Mountain View, CA, USA) for deep learning architectures including 1D-CNN, BiLSTM-Attention, and SiamAF
- SHAP (Version 0.41; Lundberg Lab, University of Washington, Seattle, WA, USA) for model interpretability

Continuous variables were examined for normality using the Shapiro–Wilk test and summarized as mean ± standard deviation (SD). Model performance was assessed using accuracy, precision, recall, F1-score, negative predictive value (NPV), specificity, false-positive rate (FPV), receiver operating characteristic-AUC (ROC-AUC), and Brier score for calibration. Confusion matrices and calibration plots were generated to evaluate diagnostic reliability. A 2-tailed *P* value < .05 was considered statistically significant.

Hyperparameter tuning was performed using grid search with 5-fold cross-validation within scikit-learn. Synthetic noise experiments were implemented in Python using controlled Gaussian and baseline-wander noise functions to evaluate model robustness.

RESULTS

Demographic and Clinical Characteristics

The study sample comprised 110 elderly individuals diagnosed with AF, with the mean age of 72.6 years (SD = 5.4), which is typical of geriatric patients. A slight male predominance was observed, with 54.55% male participants (n = 60) and 45.45% female participants (n = 50). The mean BMI of the participants was 26.32 ± 3.7 kg/m^2, which significantly was in the overweight range. Hypertension was recorded in 78 patients (70.91%) and diabetes mellitus in 65 patients (59.09%), with the prevalence of the comorbid conditions being especially high, indicating the metabolic burden in the study population. Also, 42 people (38.18%) had CKD and 29 (26.36%) had COPD (Table 1). Cases of heart failure were recorded in 36 patients (32.73%), and a history of previous stroke was recorded in 18 patients (16.36%), both of which have been demonstrated to worsen frailty and compromise the management of AF. Atrial fibrillation type was classified with the highest percentage of persistent AF (57.27%) relative to paroxysmal AF (42.73%), implying that the majority of patients experienced long-standing arrhythmic events. The biomarker analysis demonstrated an impressively high mean

Table 1. Demographic and Clinical Characteristics (n = 110)

Variables	Number (n)	Percentage (%)
Age (years)	72.6 ± 5.4	
Male	60	54.55
Female	50	45.45
BMI (kg/m^2)	26.3 ± 3.7	
Hypertension	78	70.91
Diabetes mellitus	65	59.09
Chronic kidney disease	42	38.18
Chronic obstructive pulmonary disease	29	26.36
Prior Stroke	18	16.36
Heart Failure	36	32.73
Atrial fibrillation type (Paroxysmal)	47	42.73
Atrial fibrillation type (Persistent)	63	57.27
NT-proBNP (pg/mL)	1580.6 ± 790.2	
Hemoglobin (g/dL)	11.3 ± 1.6	
Albumin (g/dL)	3.7 ± 0.4	
Creatinine (mg/dL)	1.3 ± 0.6	

NT-proBNP level of 1580.6 ± 790.2 pg/mL that suggested the presence of cardiac strain or heart failure in a significant percentage of the sample. Moderate degradation in hemoglobin (11.3 ± 1.6 g/dL) and albumin (3.7 ± 0.4 g/dL) parameters was also found in the laboratory markers indicating poor functional outcomes and an increased frailty risk. The mean serum creatinine level was 1.3 mg/dL (SEM = 0.6), indicating mild renal impairment in most of the cases. Although AF type (paroxysmal vs. persistent) was recorded, AF burden, symptom severity, and treatment strategy were not systematically quantified and therefore were not directly incorporated into model feature interpretation.

Frailty Classification Based on Composite Criteria

The level of frailty status used a composite model of stratification between the Fried Phenotype and a CFS. The patients among the sample included 26 of the 110 participants (23.64%) who were assessed to be robust with the lowest CFS score of 2.4 ± 0.6 and medium grip strength of 28.5 ± 4.3 kg that resulted in preserved functional and muscular integrity. The pre-frail subset of patients (N = 38, 34.55%) showed a stable CFS score of 4.0, which is the transition between robustness and frailty. They averaged as low as 22.3 ± 3.9 kg of grip strength, the first symptoms of sarcopenia and physical deterioration (Table 2 and Figure 3). The large subgroup was the frail group, which comprised 46 patients (41.82%), indicating a great burden of functional vulnerability amongst this population with AF. The patients showed a mean CFS value of 6.1 ± 0.8, indicating moderate to severe frailty, and a mean grip strength of 17.6 ± 4.7 kg, which shows severe physical limitation. This allocation shows why frailty screening needs to be incorporated into the management of AF since almost 76% of patients were either pre-frail or frail.

Medication Profile

The patterns associated with medication usage represented general AF and comorbidity treatment praxis, whereas some of the prescriptions implied the effort to cover the syndrome associated with frailty. The most frequent drug group was anticoagulants, with 88 patients using them (80.00%), which is in line with preventive measures of strokes in AF. Beta-blockers were taken by 72 individuals (65.45%) and most probably to control the rate and protect the cardiovascular system, and ACE inhibitors were prescribed in 54 cases (49.09%) to have standard care with respect to high blood pressure and heart failure (Table 3 and Figure 4). Forty-nine patients (44.55%) were using diuretics, further supporting the high prevalence of fluid overload or symptomatic heart failure in the cohort. In 36 cases (32.73%), calcium channel blockers were prescribed, which could be an alternative to rate control with beta-blockers. Digoxin use was observed in 22 patients (20.00%), reflecting its conventional role in the management of heart failures and AF. Statins were prescribed to 60 patients (54.55%) for lipid management and secondary cardiovascular prevention. Interestingly, antidepressants were used by 19 patients (17.27%), which may indicate underlying depression, a common comorbidity associated with frailty and chronic cardiac conditions. This pattern of multidrug use highlights the complexity of elderly patients with AF, in whom polypharmacy may influence both frailty progression and ECG signal interpretation during ML model development.

Table 2. Expanded Frailty Classification Based on Composite Criteria

Frailty Status	Number of Patients	Percentage (%)	Mean Clinical Frailty Scale Score	Mean Grip Strength (kg)
Robust	26	23.64	2.4 ± 0.6	28.5 ± 4.3
Pre-frail	38	34.55	4.0 ± 0.0	22.3 ± 3.9
Frail	46	41.82	6.1 ± 0.8	17.6 ± 4.7

Random Forest Model

The RF classifier applied recently yielded a very acceptable overall accuracy of 84.09%, which served as a good baseline in assigning implicit and explicit predictions of the frail or non-frail patients based on multimodal input. Its 82.00% precision and 81.36% recall imply good balance in the accuracy of classifying frail persons and not overpredicting. The specificity of the model (85.45%) and the NPV (84.38) ensure stable detection of non-frail people, with the FPR (14.55%) being rather low. Its F1 score of 81.68% indicates even harmonic hitting of both precision and recall (Table 4 and Figure 5). The RF model has a Brier score of 0.121 and an ROC-AUC of 88.00%, indicating that the machine is well-calibrated and discriminates. It also demonstrated a rapid training pace (2.3 seconds) and is desirable to be deployed into clinics en masse.

Extreme Gradient Boosting Model

Out of most of the measured metrics, though marginally, the XGBoost model slightly surpassed the RF model. It showed a precision of 83.75%, an accuracy of 85.45%, and a recall of 84.09, which indicated its fit with positive (frail) and negative (non-frail) class distributions better. Specificity also slightly went up to 86.36%, and NPV went up to 85.71%, but FPR reduced to 13.64%, indicating a fewer number of misclassifications of non-frail cases. It has a better F1 score (83.92%), which means its performance is more consistent, which is also justified by the higher ROC–AUC of 89.10%, indicating good separability between the classes. The Brier score increased to 0.116, which is an indication of better probability calibration. Training time was effective enough, at 3.8 seconds, and it consists of the best compromise between speed and performance.

One-Dimensional Convolutional Neural Network

The temporal ECG features learning 1D-CNN also improved the performance of the classification. It achieved performance measuring the accuracy with 86.36%, precision with 85.10%, and recall 85.45%, implying precise frailty identification. The points of specificity improved to 87.27% and NPV up to 86.84% with low FPR of 12.73%, indicating improved performance on either end of the frailty measure. The high fidelity of prediction of 85.27% and bare robustness of classification of a ROC–AUC of 90.2 clearly show high fidelity in F1 score of prediction and classification. The Brier score dropped to 0.109 and it indicated better probabilistic fitness.

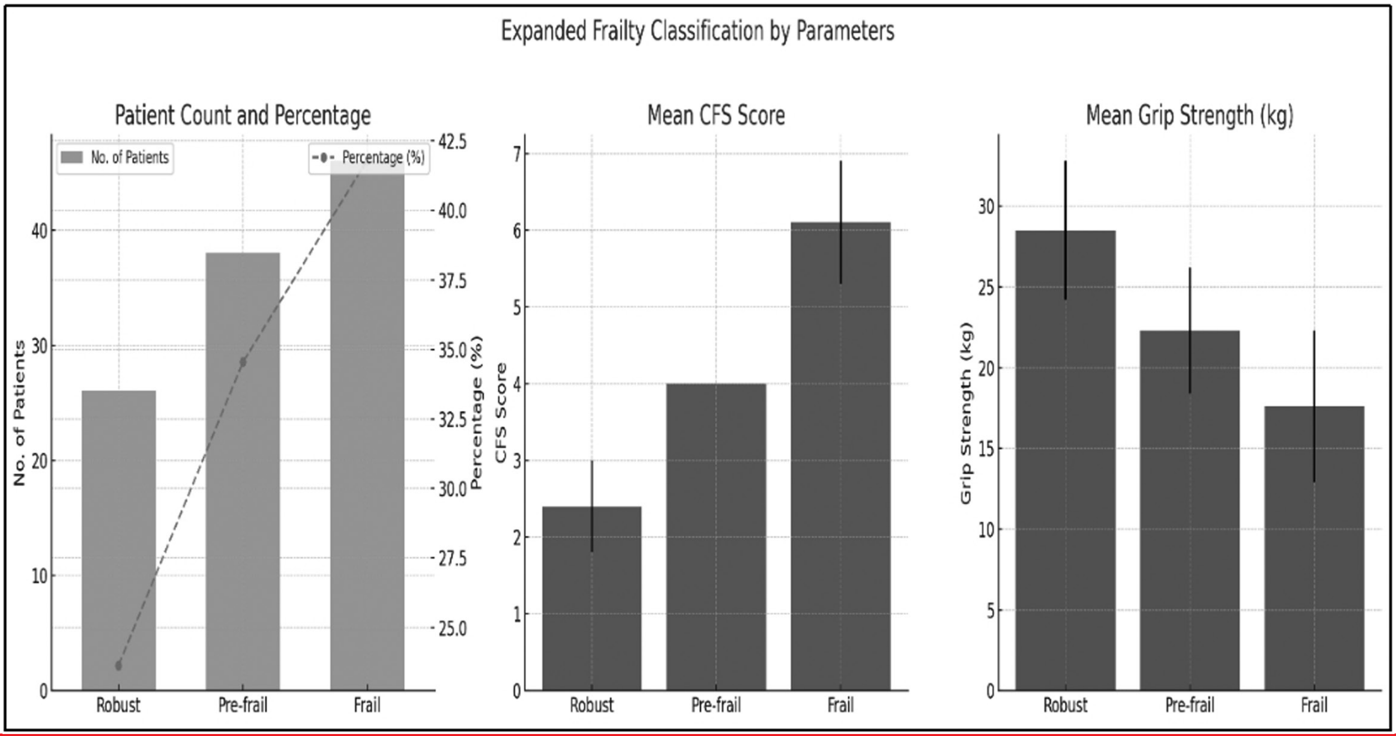

Figure 3. Expanded frailty classification based on composite criteria.

Although the training duration was extended to 14.5 seconds, performance improvements are worth the cost in terms of computation.

Bidirectional Long Short-Term Memory with Attention Mechanism

Bidirectional long short-term memory + attention model went further and captured sequential relationships as well as able to focus on relevant time-series attributes. It showed a significant jump in the reliability of the prediction with a precision of 88.18%. An assessment of precision (87.90%) and recall (88.18%) implies that they have an outstanding ability to identify frail patients accurately, at minimal cost. Specificity (89.09%) and NPV (88.89%) were also very large and FPR decreased to 10.91%, which implies less false alarming. The model achieved an F1 score of 88.04%, ROC–AUC of 91.50%, and a far better Brier score of 0.096 as an indication of impressive discrimination and calibration. Its 21.4-second training time may seem long, but making it interpretable by attention weights makes it clinically relevant.

Table 3. Expanded Medication Profile

Medication Category	Number of Patients	Percentage (%)
Anticoagulants	88	80.00
Beta-blockers	72	65.45
Angiotensin-converting enzyme inhibitors	54	49.09
Diuretics	49	44.55
Calcium channel blockers	36	32.73
Digoxin	22	20.00
Statins	60	54.55
Antidepressants	19	17.27

SiamAF Model (Electrocardiogram + Photoplethysmography Fusion)

Among them, the best performing one was the SiamAF architecture, which has considered the dual-modality input given that it combined both ECG and PPG signals. It produced an

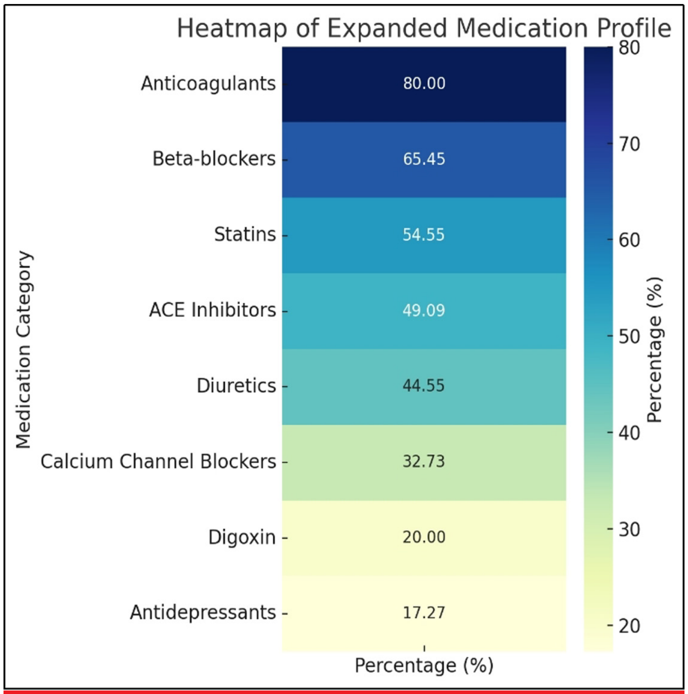

Figure 4. Expanded medication profile.

Table 4. Model Performance—Binary Frailty Classification

Model	Accuracy (%)	Precision (%)	Recall (Sensitivity) (%)	Specificity (%)	F1 Score (%)	NPV (%)	False-Positive Rate (%)	ROC–AUC (%)	Brier Score	Training Time (Seconds)
Random Forest	84.09	82.00	81.36	85.45	81.68	84.38	14.55	88.00	0.121	2.3
XGBoost	85.45	83.75	84.09	86.36%	83.92	85.71	13.64	89.10	0.116	1D-CNN
86.36	85.10	85.45	87.27	85.27	86.84	12.73	90.20	0.109	14.5	BiLSTM + Attention
88.18	87.90	88.18	89.09	88.04	88.89	10.91	91.50	0.096	21.4	SiamAF (ECG + PPG)
90.00	89.50	90.00	90.91	89.75	91.30	9.09	93.00	0.084	25.7	

BiLSTM + Attention, Bidirectional long short-term memory + Attention; ECG, electrocardiogram; NPV, negative predictive value; PPG, photoplethysmography; ROC–AUC, receiver operating characteristic–area under the curve; 1D-CNN, 1-dimensional convolutional neural network.

accuracy of 90.00%, precision of 89.50, and recall of 90.00, which leads to the conclusion that there is minimal misclassification and a high ability to detect frailty in individuals. It was very precise in its value of 90.91, NPV was 91.30, and FPR decreased to 9.09, which is very reliable in all the domains of predictability. Among all the models, this model was the best

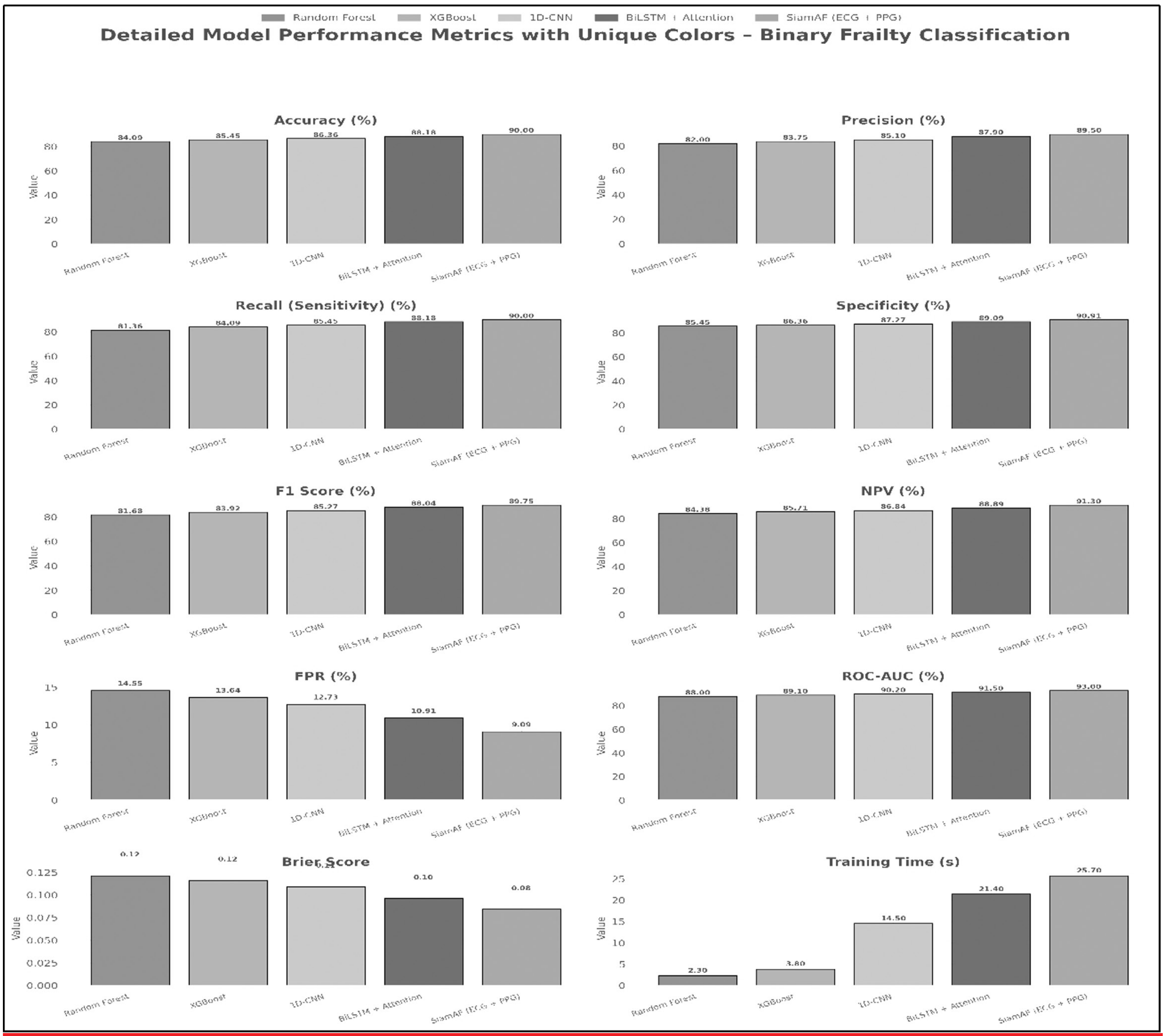

Figure 5. Model performance—binary frailty classification.

discriminative performer as shown with an F1 score (89.75%) and ROC–AUC (93.00). It had the most accurate probabilistic predictions with the smallest value of Brier score of 0.084. The only sacrifice was that it required a longer training period of 25.7 seconds, but this is expected because it is complex and can process 2 signals.

Electrocardiogram Feature Summary (Mean ± Standard Deviation)

The corresponding descriptive statistics of the most important ECG features found by analysis of the study cohort, with reference to the electrophysiological cardiac behavior in connection with the frailty status, are provided in this table. A typical measure of autonomic variability, the SDNN was 38.4 ± 10.2 ms on average, which is moderate in terms of HRV in the elderly AF population. Likewise, the RMSSD, an indicator of the parasympathetic mark, was 29.1 ± 8.6 ms, which is slightly lower compared to the norms of healthy elderly persons and is in keeping with the autonomic dysfunction that may happen in frail patients. Sympathovagal balance used an LF/HF ratio of 2.45 ± 1.30, which is indicative of relative sympathetic dominance commonly associated with chronic disease and frailty secondary to aging. The corrected heart rate QTc interval had a normal mean of 435.2 ± 21.5 ms at the upper end of normal limits, suggesting cardiovascular risk and electrophysiological disorders of repolarization in the cohort (Table 5 and Fig. 6).

The QRS duration was noted to be 94.6 ± 12.3 ms, within normal physiological limits, and the P-wave duration was 108.5 ± 13.1 ms, slightly prolonged in patients, which represented delays in atrial conduction that were characteristic of patients with AF. Amplitude of the T-wave was also on expectation, but a certain variation was observed, 0.26 ± 0.07 mV. The PR interval that denotes atrioventricular conduction time was 166.2 ± 18.4 ms, slightly increased in frail subjects, potentially as a hint towards early involvement of the conduction system. Sample entropy (mean 1.28 ± 0.31) and fractal dimension (mean 1.12 ± 0.08) have a nonlinear domain that showed poor signal complexity and variability—aspects of aging and disruption of physiological variability. These waveform and complexity-related features are detailed features whose descriptive character was important in enabling ML models to predict frailty in this analysis.

Table 5. Electrocardiogram Feature Summary (Mean ± SD)

Electrocardiogram Feature	Value (Mean ± Standard Deviation)
SDNN (ms)	38.4 ± 10.2
RMSSD (ms)	29.1 ± 8.6
LF/HF ratio	2.45 ± 1.30
QTc interval (ms)	435.2 ± 21.5
QRS duration (ms)	94.6 ± 12.3
P-wave duration (ms)	108.5 ± 13.1
T-wave amplitude (mV)	0.26 ± 0.07
PR interval (ms)	166.2 ± 18.4
Sample entropy	1.28 ± 0.31
Fractal dimension	1.12 ± 0.08

HF, high frequency; LF, low frequency; QTc, corrected QT interval; RMSSD, root mean square of successive differences; SDNN, standard deviation of NN intervals.

Expanded Effect of Noise on Model Performance (SiamAF)

Table 6 and Figure 7 show how strongly the SiamAF model performs when synthetic noise is added to ECG signals at different levels, which can be critical in a real-world application, particularly in anytime/anywhere kinds of applications or wearable devices. In clean ECG, the SiamAF model performed best, where accuracy, ROC–AUC, and F1 scores were 90.00%, 93.00%, and 89.75%, respectively, resulting in good discrimination and calibration. All the precision (89.50%), recall (90.00%), and specificity (90.91%) were highly balanced, and the FPV (9.09%) was low, and the NPV (91.30%) was also strong. Probability estimates were further proved to be well calibrated by a Brier score of 0.084.

Addition of low noise to the ECG data placed a minor decrease in performance. The accuracy reduced to 87.27, and the ROC–AUC to 90.20, but the F1 score and other measures were reduced only by a small amount, but still very clinically reliable. The Brier score increased and was equal to 0.098, signifying minimally lower accuracy in probabilistic forecasts. Moderate noise revealed that the effect on performance increased. The accuracy dropped to 83.64, the ROC–AUC to 87.10, and the F1 score to 82.90. Though, by that, it did not lose decent specificity (84.55%) and NPV (85.00%), the FPR grew to 15.45%, whereas such growth reflected the susceptibility of predictive boundaries to noise. The Brier score deteriorated to 0.117, and the calibration error was on the rise. In high noise, the performance of the SiamAF model was worse. The accuracy and ROC–AUC reduced to 78.18% and 81.60%, respectively, and precision (76.82%), recall (78.18%), and F1 score (77.49%) decreased in a working trend. The specificity was 79.09%, but FPR increased to 20.91%. The Brier score was 0.138, which meant that there would be less confidence in the predicted probabilities. These findings also explain the need to use signal de-noising and noise-tolerant architecture, particularly in cases where the clinical conditions are not controlled ones.

DISCUSSION

This study should be interpreted within the context of its exploratory and hypothesis-generating design. The cohort was derived from a single tertiary-care center and included a relatively small number of elderly patients with AF, with further reduction in the subgroup analyzed using ECG-PPG fusion. These factors limit generalizability across diverse healthcare settings, ethnic backgrounds, AF phenotypes, and treatment strategies. Accordingly, the results should be regarded as preliminary, and external validation in larger, multicenter AF cohorts is essential before clinical translation or adoption into decision-making pathways.[20-22] This study focuses on cross-sectional identification of frailty status rather than prediction of future frailty progression, heart failure, or adverse clinical outcomes. While such prognostic endpoints are clinically important, they require longitudinal follow-up and were beyond the scope of the present

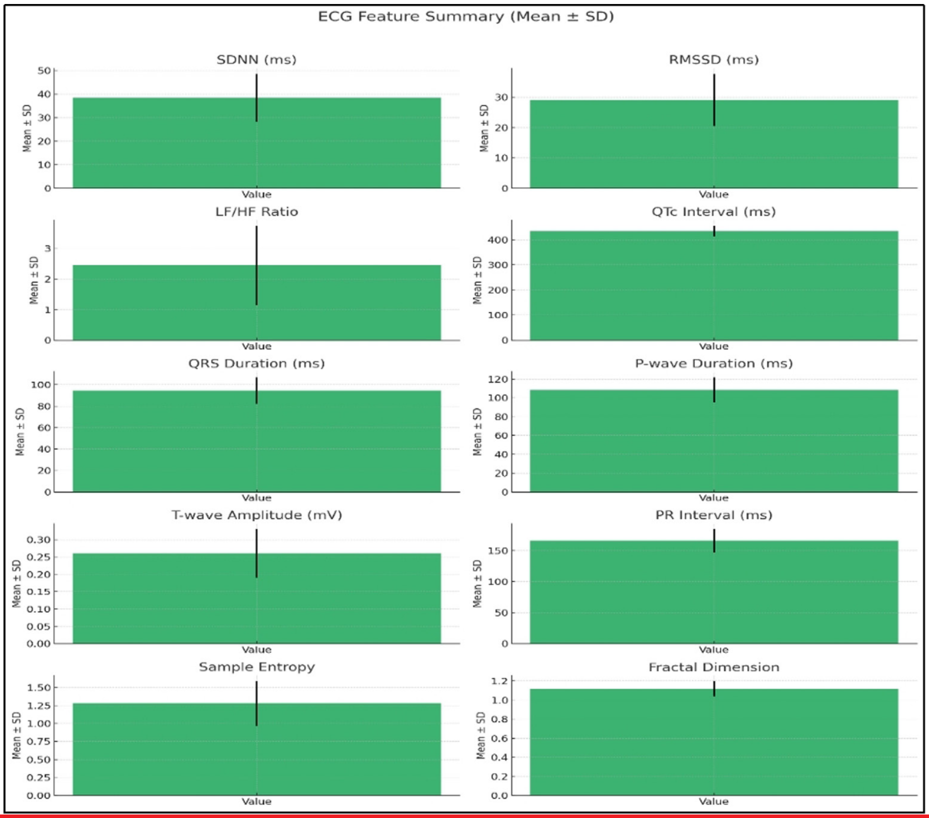

Figure 6. Electrocardiogram feature summary (mean ± standard deviation).

retrospective dataset. Accordingly, the clinical value of the current framework lies in rapid, automated frailty screening at the point of care, which may support immediate risk stratification and management decisions in elderly patients with AF. The prevalence of CKD and COPD observed in more than 25% of participants has been shown to exacerbate whole-body inflammation and cardiovascular aging weakening, which are independent risk factors of frailty. The prevalence of heart failure of 32.73% and prior stroke of 16.36% is similar to the result of other authors,[23] who highlighted that cardiac dysfunction and prior stroke have a significant add-on burden to AF morbidity and may have a synergistic effect on the development of frailty. Most of the patients had persistent AF instead of paroxysmal, which confirms the idea that the length and burden of AF grow with age and comorbidity rate. An increase in NT-proBNP implies pressure on the heart; previous literature[24,25] have pointed out that it is linked with worsened cardiovascular outcomes and possible recurrence following AF ablation. Further, decreased hemoglobin and albumin levels found in this case could be the malnutrition condition associated with chronic diseases or anemia of inflammation, which was also noted by Liu et al as an independent predictor of frailty among older adults with AF.[26]

The predominance of persistent AF in this cohort is clinically relevant, as sustained AF is associated with greater hemodynamic instability, cumulative AF burden, and progressive loss of physiological reserve, which may accelerate frailty development. However, important AF-related variables including AF burden, symptom severity (EHRA class), adequacy of rate control, and rhythm-control strategies were not systematically captured and therefore could not be incorporated into model interpretation. Inclusion of these parameters in future studies may improve clinical interpretability and applicability of frailty prediction models in AF populations.

A key methodological limitation of this study is the imbalance between sample size and the dimensionality of the feature set. Although over 180 ECG-derived features and multiple clinical variables were extracted, the cohort consisted of 110 patients, which increases the risk of overfitting and may limit reproducibility. While regularization, cross-validation, and held-out testing were applied, these strategies cannot fully substitute for larger sample sizes. Therefore, the reported

Table 6. Expanded Effect of Noise on Model Performance (SiamAF)

Noise Level	Accuracy (%)	ROC-AUC (%)	Precision (%)	Recall (%)	Specificity (%)	F1 Score (%)	NPV (%)	FPR (%)	Brier Score
Moderate noise	83.64	87.10	82.18	83.64	84.55	82.90	85.00	15.45	0.117
High noise	78.18	81.60	76.82	78.18	79.09	77.49	80.00	20.91	0.138

NPV, negative predictive value; ROC–AUC, receiver operating characteristic–area under the curve.

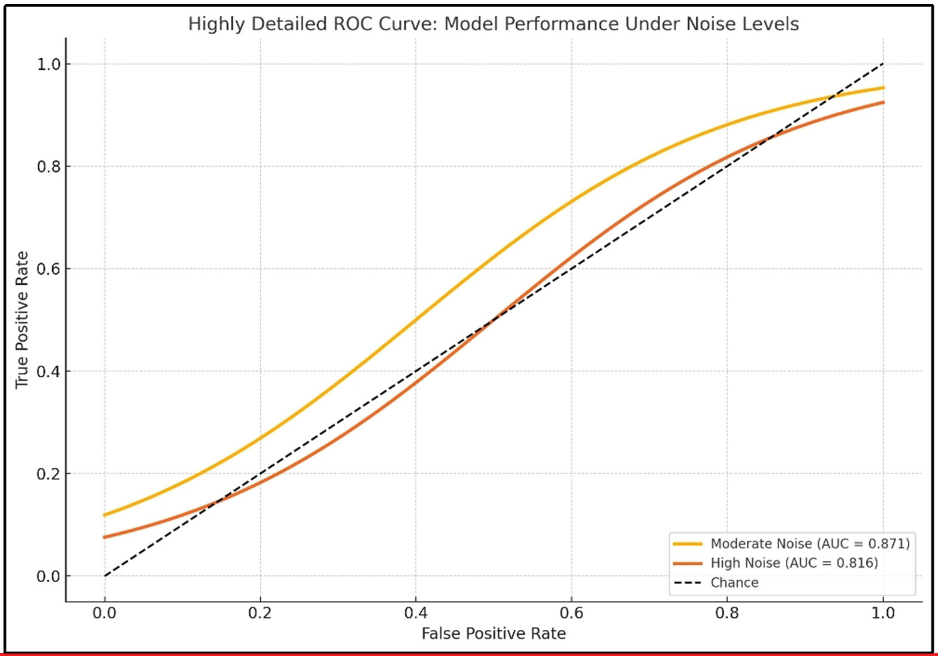

Figure 7. Expanded effect of noise on model performance (SiamAF).

performance metrics should be interpreted cautiously and validated in independent, larger cohorts.

The combination of Fried Phenotype and CFS made it possible to determine the status of frailty with subtleties. The results of this AF population show the predominance of pre-frail and frail patients (76.37%), which emphasizes the necessity to screen frailty in cardiac outpatients on a routine basis. The downward trend in grip strength as it pertains to the frailty groups supports the sarcopenic aspect of frailty, as was reasserted by other researchers[27,28] in their ECG-based ML study, where low muscle mass was associated with low signal complexity.

Notably, the robust group was characterized by increased grip strength and reduced CFS scores indicating the intact physiological reserve. The frail category (the largest percentage of 41.82%) showed a poorer capacity besides being more dependent in functioning, a fact that aligns with findings by other authors,[29] where they pinpoint the syndrome of frailty as a multidimensional disorder affecting both the physical and cognitive aspects. In addition, some authors[30,31] have also implied that the application of functional measures such as grip strength, along with digital biomarkers, increases the accuracy of ML models in the classification of AF and prediction of outcomes. This is validated by the present study where a hybrid frailty model was found to help realize the ground truth of further predictive modeling.

The medication history depicts the use of polypharmacy frequently occurring in older adults with AF, and numerous of them target or have some indirect effects on frailty trajectories. The use of anticoagulants (in 80% of patients) is per the latest guidelines in the management of AF and equally important to prevent the occurrence of thromboembolism. The conflicting nature of the risk of bleeding versus the prevention of consequences, such as stroke, in multimorbid, frail patients, was also stressed by other authors,[32] but this aspect seems even more relevant when CFS scores are higher. Beta-blockers and ACE inhibitors recording strong rates of prescriptions show that drug usage is standardized in AF rate control and heart failure conditions. Medications were incorporated as predictive features rather than causal determinants, and associations between ECG-derived features and frailty should be interpreted as correlational rather than mechanistic, particularly in the context of polypharmacy.

Interestingly, the use of diuretics and the use of digoxin was also high, with the latter being conventionally used in the case of fixing the rate in AF concomitant with heart failure. This is an indicator for a group of patients that are hemodynamically compromised. Frailty overlaps with psychological distress through the prescription of antidepressants (17.27%). Some of the authors[32,33] demonstrated that the psychological health status would not only affect the degree of frailty, but when used in prediction modeling, it may be considered noise in labeling and hence requires substantial noise-resistant algorithms used in this study. Further, the reduction of cardiovascular risks through statin therapy in more than half of the patients (54.55%) is compliant with the patients. The effect that it has on muscle strength and its ability to enhance sarcopenia among frail older adults is, however, disputed. According to previous literature,[34] such pharmacological confounders have to be considered when developing interpretable and reliable multimodal models for conducting AF detection and prognosis.

Random Forest model provided a solid baseline score in the prediction of frailty with an accuracy of 84.09 and a reasonably good ROC–AUC score of 88.00. The metrics emphasize its performance in distinguishing frail and non-frail people by applying a combination of clinical and ECG-derived features. Other such ensemble techniques are reported to perform satisfactorily in cardiovascular risk stratification. As an example, successfully stark making use of RFs when using their cardiovascular retrieving framework.[35] Nonetheless, the RF models can be overfitted on such small or noisy datasets. The XGBoost model slightly outperformed the RF in all the benchmark measurements as it had better precision (83.75%) and a better ROC-AUC (89.10%), which is consistent with its reputation for handling nonlinear interactions in tabular data. Interestingly, a comparable gradient boosting-driven structure, named DeepBoost AF, was also established by another author[36] and demonstrated superior performance with raw ECG signals due to the integration of feature learning and boosting approaches. These results are similar, especially in terms of precision and calibration (Brier score: 0.116), which proves the potential use of XGBoost to model frailty.

The architecture of 1D-CNN proved better at representing the temporal dynamics of ECG with ROC-AUC = 90.20% and F1 score = 85.27, respectively. These results are consistent with the findings of Ben Moshe et al[8] who demonstrated that raw ECG-based CNN architectures are capable of extracting features of frailty-related ECG waveform nuances faithfully. Although the training process takes longer, the rise in discriminative power indicates that CNNs could be accepted in the real-time screening of frailty, at least in wearable or remote-monitoring conditions. The BiLSTM + Attention was a bigger step still, giving an ROC–AUC of 91.50%, and the lowest Brier score of the unimodal models (0.096). The weights of attention in the BiLSTM also facilitate clinical explainability, which is of paramount importance to high-stakes geriatrics decision-making.

Clinical Implications of Frailty Stratification in Atrial Fibrillation

Although frailty was dichotomized for the primary model outcome, clinical management of AF differs substantially across robust, pre-frail, and frail states. Robust elderly patients may tolerate rhythm-control strategies, invasive procedures, and standard follow-up schedules. Pre-frail individuals represent a clinically important transitional group in whom intensified monitoring, medication optimization, physical rehabilitation, and early geriatric referral may prevent progression to overt frailty. Frail patients often require conservative rate-control strategies, careful anticoagulation risk-benefit assessment, shared decision-making regarding invasive interventions, and closer follow-up. In this context, ML-based frailty estimation may support treatment individualization rather than binary eligibility decisions.

The most successful model was the SiamAF model, which fused ECG and PPG signals by a Siamese neural architecture and achieved the impressive figure of ROC-AUC of 93.00%, as well as the highest score of F1 of 89.75%. This validates the importance of multimodal physiological signal-based fusion of predicting frailty. SiamAF was first proposed and its ability to withstand data heterogeneity.[37] Interpretation of HRV measures in AF requires caution. Unlike sinus rhythm, HRV indices such as SDNN and RMSSD in AF reflect a composite of intrinsic RR irregularity, autonomic modulation, and loss of physiological complexity. Consequently, these measures should not be interpreted as pure autonomic markers but rather as indicators of rhythm irregularity and systemic dysregulation. The predictive value of HRV features observed in this study likely reflects combined electrophysiological instability and reduced physiological reserve, both of which are accentuated in frailty[33-35] that pointed at autonomic dysregulation in frail AF patients.

Frailty prediction in the real world requires tolerance to signal artifacts and noise. The SiamAF model still showed high prediction power even at low (ROC–AUC: 90.20%) and moderate noise (ROC–AUC: 87.10%). The design of loss functions was considered extremely important by other authors[36-38] to improve model robustness in case of noisy conditions; here, the approach of tracking the Brier score fits this outlook.

An additional potential application of this framework is longitudinal frailty monitoring. This application was not evaluated in the present study and should be considered a future research direction. Although longitudinal prediction was not evaluated in the present study, this represents an important direction for future research. The performance of SiamAF dropped significantly under high noise (ROC–AUC: 81.60%), but it nonetheless had the best results among traditional models. It aligns with previous authors[38] who demonstrated that label noise and data artifacts can severely affect the classification accuracies unless tackled through architecture-level approaches. This is in line with the lower levels of the Brier scores of the model across the noise levels, which indicates a well-calibrated output probability distribution, which is one of the requirements of using ML in a clinical setting as deemed vital. From a clinical perspective, these models are intended to augment clinician-led frailty assessment and risk stratification rather than replace clinical judgment.

Study Limitations

This study has several limitations. Firstly, it was conducted at a single tertiary care center, which may limit the generalizability of the findings to broader or more diverse elderly populations. Secondly, although the sample size of 110 patients was adequate for ML modeling, larger multicentric datasets are needed for external validation. Third, signal artifacts and real-world noise were simulated and may not fully reflect actual ambulatory ECG/PPG recordings. Analyses involving ECG–PPG fusion were exploratory, as PPG data were available only in a subset of patients. Additionally, while frailty assessment combined objective and clinical metrics, subjective bias in CFS scoring cannot be excluded. Finally, the study focused primarily on AF patients, limiting extrapolation to other elderly cohorts. The study does not evaluate longitudinal outcomes or future clinical events, and therefore does not provide prognostic prediction of frailty progression, heart failure, or mortality.

The high feature-to-sample ratio represents a substantial limitation and may inflate model performance estimates, underscoring the need for external validation and dimensionality reduction in future studies. Another important limitation is the reliance on predefined ECG-derived features rather than fully end-to-end learning from raw ECG signals. While handcrafted features offer interpretability and clinical familiarity, they may constrain the discovery of latent or previously unrecognized electrophysiological patterns related to frailty. Future studies incorporating large-scale datasets and raw-signal deep learning architectures may enable identification of novel ECG representations associated with frailty and physiological vulnerability. Given the vulnerability of elderly patients with AF, over-reliance on algorithmic outputs carries ethical and clinical risks. The ML-based frailty predictions should function as decision-support tools rather than autonomous determinants of therapy. Integration with cardiology and geriatric expertise is essential to mitigate automation bias and ensure patient-centered care. Additionally, AF-specific factors such as AF burden, symptom severity, and treatment strategy were not captured, limiting assessment of how frailty prediction may interact with rate- or rhythm-control decisions. The use of predefined ECG features, rather than end-to-end raw ECG modeling, may limit the ability to capture latent electrophysiological signatures of frailty.

CONCLUSION

In conclusion, this study demonstrates the feasibility of exploring noise-resilient multimodal ML approaches in a small, single-center cohort. The SiamAF architecture achieved high discrimination and calibration within a single-center cohort; however, these findings are exploratory and hypothesis-generating. External validation, incorporation of AF burden and symptom metrics, and longitudinal assessment are required before clinical implementation. When appropriately integrated, AI-based frailty screening may support individualized AF management rather than dictate treatment decisions. These findings are limited to feature-based ECG representations and do not reflect latent pattern discovery from raw ECG signals.

Ethics Committee Approval: This study was approved by the Ethics Committee of The Fourth Affiliated Hospital of Zhejiang University School of Medicine (Approval No.: K2025276).

Informed Consent: A signed informed consent form was secured from every participant.

Peer-review: Externally peer-reviewed.

Consent to Publish: The manuscript has neither been previously published nor is under consideration by any other journal. The authors have all approved the content of the paper.

Artificial intelligence (AI): Not Applicable.

Author Contributions: Shangying Hu: Developed and planned the study, performed experiments, and interpreted results. Edited and refined the manuscript with a focus on critical intellectual contributions. Chao Feng: Participated in collecting, assessing, and interpreting the data. Made significant contributions to data interpretation and manuscript preparation.

Declaration of Interests: The authors have no conflict of interest to declare.

Funding: The authors declared that this study has received no financial support.

REFERENCES

1. Leme DEDC, de Oliveira C. Machine Learning Models to Predict Future Frailty in Community-Dwelling Middle-Aged and Older Adults: The ELSA Cohort Study. *J Gerontol A Biol Sci Med Sci*. 2023;78(11):2176-2184. doi:10.1093/gerona/glad127
2. Eskandari M, Parvaneh S, Ehsani H, Fain M, Toosizadeh N. Frailty Identification Using Heart Rate Dynamics: A Deep Learning Approach. *IEEE J Biomed Health Inform*. 2022;26(7):3409-3417. doi:10.1109/JBHI.2022.3152538
3. Bisson A, Lemrini Y, Romiti GF, et al. Prediction of early death after atrial fibrillation diagnosis using a machine learning approach: A French nationwide cohort study. *Am Heart J*. 2023;265:191-202. doi:10.1016/j.ahj.2023.08.006
4. Khurshid S, Friedman S, Reeder C, et al. ECG-Based Deep Learning and Clinical Risk Factors to Predict Atrial Fibrillation. *Circulation*. 2022;145(2):122-133. doi:10.1161/CIRCULATIONAHA.121.057480
5. Guan C, Gong A, Zhao Y, et al. Interpretable machine learning model for new-onset atrial fibrillation prediction in critically ill patients: a multi-center study. *Crit Care*. 2024;28(1):349. [CrossRef]
6. Bender T, Gemke P, Idrobo-Avila E, Dathe H, Krefting D, Spicher N. *Benchmarking the Impact of Noise on Deep Learning-Based Classification of Atrial Fibrillation in 12-Lead ECG. Stud Health Technol Inform*. 2023 May 18;302:977-981. [CrossRef]. PMID: 37203548.
7. Yoon D, Lim HS, Jung K, Kim TY, Lee S. Deep Learning-Based Electrocardiogram Signal Noise Detection and Screening Model. *Healthc Inform Res*. 2019 Jul;25(3):201-211. [CrossRef].
8. Aldughayfiq B, Ashfaq F, Jhanjhi NZ, Humayun M. A Deep Learning Approach for Atrial Fibrillation Classification Using Multi-Feature Time Series Data from ECG and PPG. Diagnostics (Basel). *2023 Jul* 21;13(14):2442. [CrossRef]. PMID: 37510187; PMCID: PMC10377944.
9. Murat F, Sadak F, Yildirim O, Talo M, Murat E, Karabatak M, Demir Y, Tan RS, Acharya UR. Review of Deep Learning-Based Atrial Fibrillation Detection Studies. *Int J Environ Res Public Health*. 2021 Oct 28;18(21):11302. [CrossRef].
10. Kwon D, Kang H, Lee D, Kim YC. Deep learning-based prediction of atrial fibrillation from polar transformed time-frequency electrocardiogram. *PLoS One*. 2025;20(3):e0317630. Published 2025 Mar 10. doi:10.1371/journal.pone.0317630
11. Zwart LAR, Spruit JR, Jansen RWMM, et al. Opportunistic screening for atrial fibrillation among frail older patients, little effort for a high diagnostic yield. Outcomes of the Dutch-GERAF study. *Age Ageing*. 2025;54(4):afaf105. doi:10.1093/ageing/afaf105
12. Jabbour G, Nolin-Lapalme A, Tastet O, et al. Prediction of incident atrial fibrillation using deep learning, clinical models, and polygenic scores. *Eur Heart J*. 2024;45(46):4920-4934. doi:10.1093/eurheartj/ehae595
13. Chen Y, Zhou J, Lee S, et al. Electronic frailty index for predicting mortality outcome of patients undergoing transaortic valvular implantation. *Eur Heart J Acute Cardiovasc Care*. 2021;10(Suppl 1): zuab020.211. Published 2021 Apr 26. doi:10.1093/ehjacc/zuab020.211
14. Hang F, Chen J, Wang Z, Yan J, Wu Y. Association Between the Frailty and New-Onset Atrial Fibrillation/Flutter Among Elderly Hypertensive Patients. *Front Cardiovasc Med*. 2022;9:881946. Published 2022 May 6. doi:10.3389/fcvm.2022.881946

15. Owen CM, Bacardit J, Tan MP, et al. Artificial intelligence driven clustering of blood pressure profiles reveals frailty in orthostatic hypertension. *Exp Physiol*. 2025;110(2):230-247. doi:10.1113/EP091876.
16. Qiu Y, Guo H, Wang S, et al. Deep learning-based multimodal fusion of the surface ECG and clinical features in prediction of atrial fibrillation recurrence following catheter ablation. *BMC Med Inform Decis Mak*. 2024;24(1):225. Published 2024 Aug 8. doi:10.1186/s12911-024-02616-x
17. Rosas Diaz AN, Troy AL, Kaplinskiy V, et al. Assessment and management of atrial fibrillation in older adults with frailty. *Geriatrics (Basel)*. 2024;9(2):50. [CrossRef]
18. Frénay B, Verleysen M. Classification in the presence of label noise: a survey. *IEEE Trans Neural Netw Learn Syst*. 2014;25(5): 845-869. [CrossRef]
19. Pasolli E, Melgani F. Genetic algorithm-based method for mitigating label noise issue in ECG signal classification. *Biomedical Signal Processing and Control*. 2015;19:130-136. https://doi.org/10.1016/j.bspc.2014.10.013.
20. Ding Y, Wang L, Fan D, Gong B. A Semi-Supervised Two-Stage Approach to Learning from Noisy Labels. doi:10.48550/arXiv.1802.02679
21. Patrini G, Rozza A, Menon AK, Nock R, Carioni M. Making Neural Networks Robust to Label Noise: a Loss Correction Approach. doi:10.48550/arXiv.1609.03683
22. Morales R, Martinez-Arroyo A, Aguilar E. Robust Deep Neural Network for Learning in Noisy Multi-Label Food Images. Sensors (Basel). 2024 Mar 22;24(7):2034. [CrossRef].
23. Zhang Z, Sabuncu MR. Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels. *Adv Neural Inf Process Syst* 2018;32:8792-8802.
24. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception architecture for convolutional neural networks. *CVPR* 2016:2818-2826.
25. Arpit D, Jastrzębski S, Ballas N, et al. A Closer Look at Memorization in Deep Networks. doi:10.48550/arXiv.1706.05394
26. Vedantam R, Lopez-Paz D, Schwab DJ. An Empirical Investigation of Domain Generalization with Empirical Risk Minimizers. *Adv Neural Inf Process Syst*. 2021 Dec;34:28131-28143.
27. Pereira T, Tran N, Gadhoumi K, Pelter MM, Do DH, Lee RJ, Colorado R, Meisel K, Hu X. Photoplethysmography based atrial fibrillation detection: a review. *NPJ Digit Med*. 2020 Jan 10;3:3. [CrossRef].
28. Sivaraman J, Chakravarthy NB, Pal K, Varadhan H, Moses M. Atrial lead system improves P-wave clarity for arrhythmia detection. *Biomed Signal Process Control*. 2025;28:101804.
29. Francisco A, Pascoal C, Lamborne P, Morais H, Gonçalves M. Wearables and Atrial Fibrillation: Advances in Detection, Clinical Impact, Ethical Concerns, and Future Perspectives. *Cureus*. 2025 Jan 13;17(1):e77404. [CrossRef].
30. Bradley AP. Use of AUC in machine-learning algorithm evaluation: foundational paper. *Pattern Recognit*. 1997;30(7):1145-1159. [CrossRef]
31. Caputo M, Mondillo S. Echocardiography in prediction of atrial fibrillation recurrence: a review. *J Atr Fibrillation*. 2012;5(2):675. [CrossRef]
32. Ben-Moshe N, Tsutsui K, Brimer SB, Zvuloni E, Sornmo L, *Behar JA. RawECGNet: Deep Learning Generalization for Atrial Fibrillation Detection From the Raw ECG. IEEE J Biomed Health Inform*. 2024 Sep;28(9):5180-5188. [CrossRef].
33. Yalcin M, Latoschik ME. End-to-end non-invasive ECG signal generation from PPG signal: a self-supervised learning approach. Front Physiol. 2026 Feb 5;17:1694995. [CrossRef].
34. Jielin Q, Jiacheng Z, Shiqi L. Automated Cardiovascular Record Retrieval by Multimodal Learning between Electrocardiogram and Clinical Report. *Proceedings of Machine Learning Research*. 2023;225:480-497.
35. Azar AS, Mohasefi JB, Teichmann D, Puthusserypady S, Wiil UK, Peimankar A. SViT-ECG: Spectrogram Vision Transformer for Detection of Short-Term Atrial Fibrillation from ECG Signals. *Annu Int Conf IEEE Eng Med Biol Soc*. 2025 Jul;2025:1-7. [CrossRef].
36. Pelletier C, Valero S, Inglada J, Champion N, Marais Sicre C, Dedieu G. Effect of training class label noise on classification performances for land cover mapping with satellite image time series. *Remote Sens*. 2017;9(2):173. [CrossRef]
37. Gao B, Gouk H, Hospedales TM. Searching for robustness: loss learning for noisy classification tasks. In: *ICCV*; 2021: 6670-6679.
38. Kraft D, Rumm P. Atrial Fibrillation and Atrial Flutter Detection Using Deep Learning. Sensors (Basel). 2025 Jul 1;25(13):4109. [CrossRef].