A PERSONALIZED MULTI-VIEW ENSEMBLE FRAMEWORK FOR HYPERKALEMIA DETECTION USING WEARABLE SINGLE-LEAD ECG

[1]CHUAN-SHENG HUNG, [2]CHUN-HUNG RICHARD LIN, [3]SHI-HUANG CHEN

[1,2]Department of Computer Science and Engineering, National Sun Yat-Sen University, Kaohsiung, Taiwan
[3]Department of Computer Science and Information Engineering, Shu-Te University, Kaohsiung, Taiwan
E-mail: cshung@g-mail.nsysu.edu.tw, lin@cse.nsysu.edu.tw, shchen@stu.edu.tw

Abstract - Chronic kidney disease (CKD) patients are prone to hyperkalemia, which can trigger fatal arrhythmias and poses a substantial risk to life. Current potassium monitoring largely relies on blood draws, making it difficult to support home-based use and high-frequency follow-up. In this study, we develop a personalized hyperkalemia detection approach using wearable single-lead ECG and formulate the task as a time-series binary classification problem. Specifically, each heartbeat is segmented by R-peak alignment into fixed-length windows (100 samples) and then decomposed into physiologically meaningful P/QRS/T components. We further construct seven waveform-band combinations to obtain a multi-view representation of cardiac dynamics. We then establish an algorithm pool of eight time-series classifiers. At the "patient × waveform-band" level, we select the optimal model via a weighted scoring function that integrates Precision, Recall, F1-score, and AUC, and then fuse predictions using voting- and stacking-based ensemble strategies. Experiments using ECG recordings from eight MIMIC-III patients show that our approach yields clear improvements over a baseline XGBoost model for most patients, particularly in AUC and Precision/Recall. Hard voting is the most stable ensemble overall, and in certain cases achieves AUC ≈ 0.99 with Precision/Recall > 0.95. These results indicate strong potential for deployment on wearable and edge-device platforms.

Keywords - Electrocardiography (ECG), Hyperkalemia, Time-Series Classification, Ensemble Learning

I. INTRODUCTION

Time series classification (TSC) learns discriminative morphological and dynamic patterns directly from continuous waveforms and has been widely adopted in biomedical signal analysis, industrial anomaly detection, and financial forecasting. Systematic reviews further highlight the growing maturity and cross-domain importance of TSC methodologies [1]–[5]. In clinical care, patients with chronic kidney disease (CKD) frequently develop electrolyte disturbances. Among these, hyperkalemia is common and can be fatal; accordingly, multiple reviews and consensus statements emphasize hyperkalemia risk management as a key component of CKD care [6], [7]. Because abnormal serum potassium can induce characteristic ECG changes (e.g., T-wave and QRS alterations), ECG has long served as an important complementary signal for assessing potassium abnormalities in practice [8]–[10]. However, potassium assessment still relies predominantly on invasive blood tests, making high-frequency, real-time, long-term monitoring difficult and limiting early warning and personalized management in home settings.

To enable noninvasive monitoring, recent studies have explored AI models that infer hyperkalemia risk from ECG, including single-lead deep-learning screeners, clinically oriented screening systems, and architectures that jointly detect hypo- and hyperkalemia [11]–[13]. While these efforts demonstrate feasibility, reliable long-term home monitoring and stable individual-level inference remain challenging. First, most approaches train a single population-level model on heterogeneous cohorts, implicitly assuming consistent potassium-related ECG manifestations across individuals, despite substantial inter-individual variability driven by CKD history, medications, demographics, and baseline rhythm. Second, deep models often impose non-trivial computational and memory demands, which can hinder real-time and sustained operation on wearable or edge platforms. Third, many methods rely on full-length segments or limited handcrafted/statistical features, with limited emphasis on interpretable, compositional representations grounded in heartbeat physiology and waveform bands, making outputs more sensitive to signal-quality variation, heart-rate fluctuations, and cross-subject morphology.

We therefore formulate wearable single-lead ECG hyperkalemia detection as a time-series classification task and propose a personalized inference framework. Continuous ECG is segmented into fixed-length, R-peak–aligned single-beat subsequences, each decomposed into P/QRS/T components and recombined into seven complementary band views to form multi-view inputs. An algorithm pool of time-series classifiers is trained, band-wise best models are selected, and band-specific predictions are fused via ensemble inference (with hard voting typically most stable). We further introduce a time-series similarity–based incremental strategy to reduce update and inference costs, improving feasibility for wearable/edge deployment and long-term home monitoring.

The remainder of this paper is organized as follows. Section II reviews related work and discusses its limitations. Section III presents the proposed method and overall framework, including ECG heartbeat segmentation, waveform-band representations, multi-model ensembling, and the similarity-based incremental strategy. Section IV describes the dataset and reports the experimental results. Finally, Section V concludes the paper and outlines directions for future work.

II. RELATED WORK

A. ECG correlates of potassium abnormalities

Early emergency medicine and clinical studies have systematically summarized the characteristic ECG manifestations associated with hyperkalemia and outlined key considerations and high-risk scenarios for bedside interpretation, thereby establishing the medical basis that potassium abnormalities may be reflected in ECG waveforms [8]. Subsequent work, often framed from educational and clinical alerting perspectives, further emphasized practical cues and diagnostic reasoning for recognizing ECG changes across different potassium levels, underscoring ECG as an important complementary indicator of potassium dysregulation [9]. In addition, reviews situated within the broader context of electrolyte disturbances have discussed ECG presentations more generally and highlighted that clinicians must account for multiple physiological and comorbid factors that can confound ECG findings, cautioning against over-inference from any single sign [10].

However, this body of clinical evidence primarily focuses on observable phenomena and interpretive principles rather than developing deployable automated models. Moreover, ECG responses to potassium abnormalities can vary substantially across individuals due to differences in medical history, medication use, comorbidities, and baseline rhythm. Consequently, population-level, rule-based interpretation alone is unlikely to support long-term, continuous, and personalized home monitoring. These limitations have motivated a recent shift toward data-driven approaches for building noninvasive detection models.

B. Data-driven ECG-based hyperkalemia detection: deep learning, personalization, and classical machine-learning pathways

Within data-driven approaches, a common direction is to develop deep-learning screeners that infer hyperkalemia directly from ECG signals. Galloway et al. developed and validated a deep-learning model for ECG-based hyperkalemia screening, demonstrating feasibility in real-world clinical data [12]. Lin et al. proposed ECG12Net, which can concurrently identify hypokalemia and hyperkalemia, indicating that deep networks can learn discriminative patterns linked to electrolyte disturbances [13]. Urtnasan et al. focused on noninvasive screening using single-lead ECG and emphasized usability and practical deployment considerations [11].

Nevertheless, two limitations remain. First, most studies train a single population-level model on cross-subject data; if potassium-related ECG manifestations vary substantially across individuals, patient-level stability and generalizability may degrade. Second, deep models often impose non-trivial computational and memory demands, and their real-time feasibility on wearable/edge platforms requires careful quantification. To address inter-individual differences, Chiu et al. explored transfer-learning personalization by fine-tuning a general model with subject-specific data, but this may introduce adaptation overhead and risk overfitting when per-subject data are limited [15]. Beyond deep learning, Taguelmimt et al. proposed transforming ECG into informative representations followed by conventional machine learning, offering a more controllable deployment pathway [16]. Moreover, Quintero et al. reported heterogeneous real-world ECG abnormalities in hyperkalemia [14]. Motivated by these challenges, we propose a heartbeat-centric multi-band, multi-view framework with ensemble fusion and similarity-driven incremental classification to improve stability and reduce long-term maintenance/inference costs.

III. METHOD

We model personalized ECG-based hyperkalemia detection for chronic kidney disease (CKD) patients as a time-series binary classification problem. The overall pipeline consists of three major components. First, we preprocess the wearable-recorded ECG and perform heartbeat-level segmentation to obtain time-series samples suitable for modeling. Second, we develop personalized time-series classification models to identify potassium abnormalities. Third, to accommodate deployment constraints on wearable and edge devices, we propose a time-series similarity–driven classification strategy that reduces inference cost while preserving detection capability, enabling real-time home monitoring and early warning.

In the experimental evaluation, we compare the performance of the personalized time-series models and the time-series similarity–based classification method on the potassium abnormality detection task using metrics such as Accuracy, Precision, and Recall. Let the training dataset be denoted as $\mathcal{D} = \{d_1, d_2, \dots, d_n\}$, where d_i is the i-th univariate time-series sample composed of $k$uniformly sampled points:

$$d = \{x_1^{(i)}, x_2^{(i)}, \dots, x_j^{(i)}\} \quad \forall i \in \{1,2,\dots,n\} \qquad (1)$$

Assuming a fixed sampling interval Δt, we have:

$$\Delta t = \text{time}\left(x_{j+1}^{(i)}\right) - \text{time}\left(x_j^{(i)}\right) \quad \forall j \in \{1,2,\dots,k-1\} \tag{2}$$

where $\text{time}(\cdot)$ denotes the timestamp function. The corresponding label set is defined as $\mathcal{Y} = \{y_1, y_2, \dots, y_n\} \subseteq \{0,1\}$, where $y_i = 1$ indicates an abnormal potassium condition and $y_i = 0$ denotes normal status. Let $\mathcal{A}: \mathcal{D} \to \mathcal{Y}$ represent the classification model to be learned, and we evaluate its generalization ability on a test set $\{(d_i', y_i')\}$. Our objective is to select the model $\mathcal{A}$ that maximizes the expected performance:

$$\max_{\mathcal{A}} \quad \mathbb{E}\left[\text{Performance}\left(\mathcal{A}(d_i'), y_i'\right)\right] \tag{3}$$

where $\mathbb{E}[\cdot]$ denotes the expectation operator, and $\text{Performance}(\cdot)$ is a performance metric function (e.g., Accuracy, Precision, Recall, F1-score, and AUC) used to quantify the agreement between the model's predictions on the test data and the corresponding ground-truth labels.

A. Data Processing and Heartbeat Segmentation

In time-series classification (TSC), a common practice is to extract representative subsequences (shapelets) from long sequences and use them as model inputs. In this study, we treat wearable ECG as a one-dimensional time series and exploit the inherent periodicity of cardiac signals by considering each heartbeat (beat) as a shapelet. Accordingly, the original long ECG recording is segmented into a set of beat-level samples, which facilitates subsequent hyperkalemia classification for CKD patients.

Because wearable devices (e.g., Apple Watch) typically output ECG as a continuous one-dimensional waveform, we first use the R-peak as the alignment reference for each heartbeat and automatically segment the ECG into multiple independent beats. Let $I(t)$ denote the signal amplitude at time t. We begin by detecting local maxima to obtain a candidate peak set S:

$$S = \{t_i \mid I(t_i) > I(t_{i-1}) \text{ and } I(t_i) > I(t_{i+1})\} \tag{4}$$

where t_i is a time index, and S contains all local maxima satisfying the condition that the signal value is greater than its two neighboring samples. We then compute the mean amplitude of the candidate peaks:

$$\bar{I}_{\text{peak}} = \frac{1}{|S|} \sum_{t_i \in S} I(t_i) \tag{5}$$

To maintain robustness across subjects and varying signal amplitudes, we adopt a statistic-based adaptive threshold and treat peaks whose amplitudes are substantially higher than the mean as R peaks. Let σ denote the standard deviation of the candidate-peak amplitudes. The R-peak set R is defined as:

$$R = \{t_i \mid I(t_i) \geq \bar{I}_{\text{peak}} + \sigma\} \tag{6}$$

Finally, for each detected R peak $t_r \in R$, we extract a fixed-length heartbeat window centered at t_r. In our setting, the ECG sampling rate is 125 Hz, and the average adult cardiac cycle is approximately 0.8 s. Accordingly, we represent a single beat using 100 samples and employ an asymmetric window—40 samples to the left and 60 samples to the right—to capture the complete waveform morphology. The beat-level sample P_{t_r} is defined as:

$$P_{t_r} = \{I(t) \mid t \in [t_r - 40, t_r + 60]\} \tag{7}$$

Through the above procedure, the original long ECG recording is transformed into a set of R-peak–aligned, fixed-length heartbeat-level time-series shapelets, which serve as the foundation for subsequent waveform-band decomposition (P/QRS/T), multi-view representation learning, and personalized classifier training.

B. Waveform-Band Processing

Most prior ECG disease-classification studies use handcrafted features—such as the R–R interval, peak-slope descriptors, or summary statistics—as model inputs. However, under potassium abnormality conditions, these features can become unstable due to pronounced inter-individual differences (e.g., disease history, medication use, comorbidities, and baseline rhythm). Instead, we adopt a band-wise modeling strategy. Using a single heartbeat S_i as the basic unit, we decompose it into three physiologically meaningful segments—P, QRS, and T—and further construct multiple band combinations to form multi-view representations for subsequent personalized classification and ensemble inference.

Let the i-th heartbeat be aligned at the R-peak location R_i, with a heartbeat-window length L (fixed to 100 samples in this study). We define the boundaries of the three segments using proportional coefficients corresponding to the P-wave segment, the QRS complex, and the T-wave segment. To ensure consistency across subjects, these boundaries are specified by relative offsets with respect to R_i, rather than relying on additional peak detection or R–R interval estimation. Specifically, we introduce three constants α, β, and γ (determined based on clinical knowledge and waveform characteristics), such that the P segment covers the slowly varying region preceding R_i, the QRS segment captures the rapid spiking region around R_i, and the T segment corresponds to the repolarization morphology following R_i. Accordingly, the three waveform bands can be formally defined as:

$$P_i = \{x_t \mid t \in [R_i - \lfloor \alpha L \rfloor,\ R_i - \lfloor \beta L \rfloor)\}, \tag{8}$$

$$QRS_i = \{x_t \mid t \in [R_i - \lfloor \beta L \rfloor,\ R_i + \lfloor \gamma L \rfloor)\}, \tag{9}$$

$$T_i = \{x_t \mid t \in [R_i + \lfloor \gamma L \rfloor,\ R_i + \lfloor (1-\alpha) L \rfloor)\}. \tag{10}$$

where x_t denotes the ECG sample at time index t, and $\lfloor \cdot \rfloor$ is the floor (integer-rounding) operator. These definitions ensure a consistent segmentation rule under a fixed window length L and directly preserve the raw morphological information of each physiologically meaningful component. After obtaining the P, QRS, and T segments, we further construct waveform-band combinations to form multi-view inputs. In addition to the individual bands

P_i , QRS_i , and T_i , we create three two-band concatenations and a full-beat representation:

$$(P + QRS)_i = P_i \parallel QRS_i, \quad (11)$$
$$(QRS + T)_i = QRS_i \parallel T_i, \quad (12)$$
$$(P + T)_i = P_i \parallel T_i . \quad (13)$$

$(P+QRS+T)_i = S_i$,where $\parallel$ denotes time-series concatenation. Consequently, each heartbeat yields seven waveform-band representations: $\{P,\ QRS,\ T,\ P + QRS,\ QRS + T,\ P + T,\ P + QRS + T\}$.

During inference, these seven bands are treated as complementary information sources. Each band is fed into a classifier to produce its corresponding prediction, and the final decision on potassium abnormality is obtained via ensemble fusion.

C. Band-Specific Best-Algorithm Selection

Because the proposed ECG preprocessing procedure decomposes each heartbeat into seven waveform-band representations (e.g., P, QRS, T, and their combinations), our preliminary experiments indicate that different bands carry distinct information patterns and discriminative cues. As a result, applying a single model uniformly across all bands does not necessarily yield optimal performance. To this end, we adopt a band-specific model selection strategy: for each waveform band, we train all eight TSC models in the algorithm pool and select the best-performing model for that band based on validation performance.

Let $B_i$denote the i-th waveform band ($i \in \{1,2,\dots,7\}$), and let A_j denote the j-th algorithm ($j \in \{1,2,\dots,8\}$). Let M_{ij} be the model trained using algorithm A_j on band B_i. For each M_{ij} , we compute four commonly used validation metrics—Precision, Recall, F1-score, and AUC—denoted as P_{ij} , R_{ij} , F_{ij} , and AUC_{ij} , respectively.To enable consistent comparison under multiple metrics, we aggregate these four measures into a single composite score S_{ij} via a weighted scheme, defined as follows: $S_{ij} = w_P P_{ij} + w_R R_{ij} + w_F F_{ij} + w_{\mathrm{AUC}}\mathrm{AUC}_{ij}, w_P + w_R + w_F + w_{\mathrm{AUC}} = 1.$ where w_P, w_R, w_F, and $w_{\mathrm{AUC}}$are importance weights assigned to each metric and can be adjusted according to task requirements (e.g., in medical detection settings, higher weights may be placed on Recall or AUC to reduce the risk of missed cases). Finally, for each waveform band B_i, we select the algorithm that maximizes the composite score as the best model for that band:

$$A_i^* = \arg\max_{j \in \{1,\dots,8\}} S_{ij}\ , \quad M_i^* = M_{i,A_i^*}. \quad (14)$$

Accordingly, this procedure yields a band-specific best model for each of the seven waveform representations, resulting in $\{M_1^*, M_2^*, \dots, M_7^*\}$. These band-wise optimal models are subsequently used as candidate inputs for ensemble learning, with the goal of improving the stability and overall detection performance of personalized ECG-based hyperkalemia prediction.

D. Ensemble Strategies Based on Time-Series Algorithms

Leveraging clinically meaningful ECG waveform structure, we innovatively decompose personalized ECG into seven waveform-band combinations and jointly utilize the information across bands to determine whether potassium status is abnormal. After heartbeat segmentation, band decomposition, and feature extraction, the resulting data are used to train the ensemble framework. We further split 20% of the training data as a validation set for model selection and for determining the ensemble strategy.

Next, the training set is partitioned into seven band-specific datasets according to the seven waveform-band representations. For each band, we train eight time-series models from the proposed algorithm pool, yielding a total of 56 candidate models (7×8) . Within the ensemble pipeline, we first perform band-wise best-model selection, i.e., for each band, we select the best-performing model on the validation set, resulting in seven band-specific optimal models. This design is motivated by the fact that different bands encode different physiological information and discriminative characteristics; forcing a single model to be shared across all bands may prevent optimal performance.

Finally, we fuse the outputs of the seven selected models using two ensemble paradigms: voting and stacking. In stacking, the predicted class labels/probabilities from the band-specific models are used as second-level inputs to train a meta-learner that produces the final decision. To mitigate overfitting, we adopt lightweight and robust meta-learners, including logistic regression, k-nearest neighbors (KNN), and support vector machines (SVM). The overall ECG-based potassium abnormality detection framework using ensemble learning is illustrated in Fig. 1.

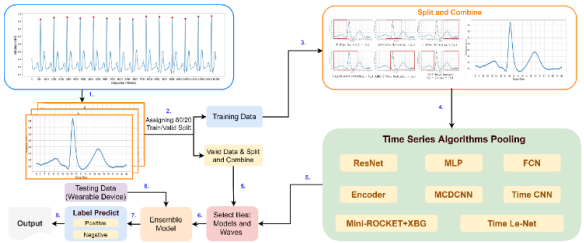


Fig. 1. Ensemble-learning framework for ECG-based potassium abnormality detection

IV. EXPERIMENTAL RESULTS

This section evaluates the proposed personalized ECG-based hyperkalemia detection framework using ECG recordings and laboratory measurements from the MIMIC-III database. The experimental protocol consists of three stages: (i) heartbeat segmentation and

construction of seven waveform-band datasets, (ii) band-specific best-model selection, and (iii) comparison between ensemble methods and the proposed similarity-based classifier.

First, for each patient, we segment the ECG into heartbeat-level samples and decompose each beat into seven waveform-band combinations based on physiological waveform structure. For each band, we construct training and validation sets and train the eight models in the time-series algorithm pool, forming a set of candidates at the "band × model" level. Next, we evaluate these candidates on the validation set and select the best model for each band for subsequent fusion.

In the second stage, we compare different ensemble strategies (e.g., voting and stacking) in terms of test-set performance. In the third stage, we further compare the best-performing ensemble strategy against the proposed similarity-based classification method to assess overall effectiveness under personalization and deployability constraints.

A. Dataset (MIMIC-III)

We use data from MIMIC-III (Multi-Parameter Intelligent Monitoring for Intensive Care III), which contains physiological monitoring signals (including ECG) and in-hospital laboratory test records. The laboratory records provide serum potassium concentrations along with blood-draw timestamps. Using time alignment, we pair ECG segments with their corresponding potassium labels, thereby constructing a binary classification dataset for normal versus abnormal potassium status.

Patients (Subject ID)	Training samples (Normal:Abnormal)	Validation samples (Normal:Abnormal)	Testing samples (Normal:Abnormal)
A	1659 (896 : 763)	414 (224 : 190)	1462 (1175 : 287)
B	682 (199 : 483)	169 (49 : 120)	691 (432 : 259)
C	22722(13786 :8936)	5680 (3446 : 2234)	15509 (10152 :5357)
D	534 (387 : 147)	132 (96 : 36)	957 (683 : 274)
E	23404 (16065 : 7339)	5846 (4012 : 1834)	17801 (11745 :6056)
F	6682 (1911 : 4771)	1663 (472 : 1191)	8884 (2933 : 5951)
G	46568(39726:6842)	11642(10472:1710)	26975 (22534 : 4441)
H	432 (212 : 220)	107 (53: 54)	1040 (782 : 258)

Table 1. Summary of training, validation, and test set statistics for eight MIMIC-III patients.

In MIMIC-III, each hospitalization record can be indexed by a Time ID (representing a specific admission/ICU stay) and includes multiple Chart Time entries (corresponding to individual measurement or documentation time points). Because a patient's physiological condition is typically more consistent within the same Time ID, using data from the same Time ID for both training and testing can lead to overly optimistic evaluation. To better reflect real-world home-monitoring scenarios—where a patient's physiology and signal distribution may shift across time—we perform data splitting at the Time ID level: the training and validation sets are drawn from specific Time IDs, whereas the test set uses data from different Time IDs to assess cross-period generalization. Table 1 summarizes the sample counts (normal/abnormal) for the eight patients across the training, validation, and test sets.

B. Personalized Differences in the Best Models Across Waveform Bands

To capture inter-individual variability in ECG morphology, we evaluate eight time-series models across seven waveform-band representations (P, QRS, T, P–QRS, QRS–T, P–T, and P–QRS–T) using AUC, F1-score, and F2-score for model selection. The best-performing model varies modestly across patients and bands, indicating meaningful personalization in potassium-related ECG patterns. Some patients achieve better discrimination with the full waveform (P–QRS–T) or repolarization-informative combinations (e.g., QRS–T), whereas others align better with localized segments (e.g., P or QRS). These results support a "patient × band" selection strategy rather than a single fixed model. Across patients, ResNet and Encoder are most frequently selected, suggesting stable representation capacity, while MiniROCKET + XGBoost performs competitively for certain band combinations. Overall, band-wise modeling with personalized selection improves robustness while preserving physiologically grounded structure in Table 2.

ECG Waves	P	QRS	T	P-QRS	QRS-T	P-T	P-QRS-T
Patient A							
Best Models	FCN	Time CNN	FCN	MLP	Mini Rocket	MLP	Mini Rocket
AUC	0.85	0.74	0.84	0.95	0.95	0.74	0.88
Patient B							
Best Models	FCN	ResNet	MLP	MLP	MCD CNN	Time Le-Net	MCD CNN
AUC	0.99	0.98	0.99	0.99	0.99	0.99	0.99

Patient C							
Best Models	Encoder	Encoder	Encoder	Time Le-Net	Encoder	Time Le-Net	Encoder
AUC	0.92	0.96	0.96	0.96	0.98	0.96	0.99
Patient D							
Best Models	FCN	Time Le-Net	Time Le-Net	ResNet	ResNet	ResNet	Mini Rocket
AUC	0.99	0.5	0.5	0.5	0.99	0.99	0.5
Patient E							
Best Models	ResNet	ResNet	MLP	ResNet	MCD CNN	ResNet	Encoder
AUC	0.99	0.99	0.98	0.99	0.99	0.99	0.98
Patient F							
Best Models	MCDCNN	Encoder	Mini Rocket	MCDCNN	ResNet	MCDCNN	FCN
AUC	0.97	0.86	0.97	0.96	0.86	0.99	0.85
Patient G							
Best Models	Time Le-Net	FCN	ResNet	ResNet	Time Le-Net	Mini Rocket	Time Le-Net
AUC	0.99	0.96	0.93	0.95	0.99	0.97	0.99
Patient H							
Best Models	MLP	Mini Rocket	Encoder	Mini Rocket	Mini Rocket	Encoder	Mini Rocket
AUC	0.5	0.89	0.5	0.86	0.95	0.55	0.99

Table 2. Band-specific models and AUC values selected for eight MIMIC-III patients.

Based on the summary statistics in Fig. 2, ResNet and Encoder exhibit the most stable performance when modeling ECG data. In particular, across the seven waveform bands for different patients, ResNet and Encoder are frequently selected as the best-performing models, indicating strong generalization capability and flexibility. The Mini-ROCKET + XGBoost pipeline ranks second, following ResNet and Encoder, suggesting that this approach also offers competitive advantages for time-series modeling.

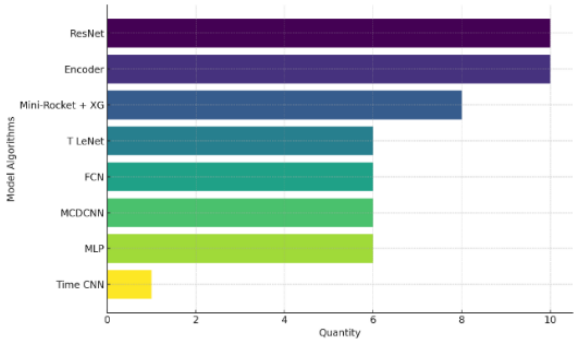


Fig. 2. Frequency of band-specific model selections based on AUC for eight MIMIC-III patients.

C. Comparison of Ensemble Methods Based on Band-Specific Best Models

After selecting the best-performing model for each patient and waveform band, we compared ensemble strategies on the test sets of eight MIMIC-III patients. Table 3 reports results for a baseline XGBoost model and three ensemble categories: hard voting, soft voting, and stacking with logistic regression, SVM, or KNN as the meta-learner. Performance was evaluated using Precision, Recall, AUC, F1-score, and F2-score. Voting-based ensembles were the most stable across patients, consistently outperforming the baseline in AUC and Precision/Recall while remaining interpretable and computationally manageable for wearable or clinical deployment. Stacking achieved comparable or slightly better performance for some individuals, but gains were subject-dependent and likely affected by distribution shifts, class imbalance, and band-specific prediction characteristics, while also introducing additional training and maintenance overhead. Overall, hard voting provided the best aggregate results, delivering balanced Precision/Recall with strong AUC and F1/F2 for most patients. Soft voting was close but occasionally reduced Precision, suggesting higher sensitivity to biased probabilities.

Patient	Precision	Recall	AUC	F1-score	F2-score
Patient A					
XGBoost (baseline)	0.17	0.82	0.41	0.28	0.46
Hard Vote	0.74	0.97	0.94	0.84	0.91
Soft Vote	0.54	0.99	0.89	0.7	0.85
Stacking (Logit. Reg.)	0.55	0.39	0.66	0.46	0.41
Stacking (SVM)	0.41	0.4	0.63	0.41	0.41
Stacking (KNN)	0.44	0.39	0.64	0.42	0.4
Patient B					
XGBoost (baseline)	0.48	0.9	0.66	0.62	0.78
Hard Vote	0.98	0.99	0.99	0.99	0.99
Soft Vote	0.98	0.99	0.99	0.99	0.99

Stacking (Logit. Reg.)	0.7	0.99	0.87	0.82	0.92
Stacking (SVM)	0.99	0.99	0.99	0.99	0.99
Stacking (KNN)	0.65	0.99	0.84	0.78	0.9
Patient C					
XGBoost (baseline)	0.74	0.68	0.77	0.71	0.69
Hard Vote	0.87	0.97	0.94	0.92	0.95
Soft Vote	0.81	0.88	0.89	0.84	0.86
Stacking (Logit. Reg.)	0.78	0.73	0.81	0.75	0.74
Stacking (SVM)	0.77	0.69	0.79	0.73	0.7
Stacking (KNN)	0.76	0.64	0.77	0.7	0.66
Patient D					
XGBoost (baseline)	0.07	0.07	0.33	0.07	0.07
Hard Vote	0.77	0.99	0.94	0.87	0.94
Soft Vote	0.73	0.99	0.93	0.85	0.93
Stacking (Logit. Reg.)	0.51	0.99	0.81	0.67	0.84
Stacking (SVM)	0.5	0.99	0.8	0.66	0.83
Stacking (KNN)	0.5	0.99	0.8	0.66	0.83
Patient E					
XGBoost (baseline)	0.55	0.78	0.73	0.64	0.72
Hard Vote	0.98	0.82	0.92	0.9	0.85
Soft Vote	0.98	0.82	0.92	0.9	0.85
Stacking (Logit. Reg.)	0.71	0.81	0.82	0.76	0.79
Stacking (SVM)	0.65	0.81	0.79	0.72	0.77
Stacking (KNN)	0.68	0.81	0.81	0.74	0.78
Patient F					
XGBoost (baseline)	0.69	0.97	0.53	0.8	0.9
Hard Vote	0.89	0.99	0.87	0.94	0.97
Soft Vote	0.86	0.98	0.83	0.92	0.96
Stacking (Logit. Reg.)	0.84	0.99	0.8	0.91	0.96
Stacking (SVM)	0.83	0.99	0.79	0.91	0.96
Stacking (KNN)	0.79	0.98	0.72	0.87	0.93
Patient G					
XGBoost (baseline)	0.36	0.35	0.61	0.35	0.35
Hard Vote	0.64	0.48	0.71	0.55	0.5
Soft Vote	0.67	0.47	0.71	0.56	0.5
Stacking (Logit. Reg.)	0.69	0.48	0.72	0.57	0.51
Stacking (SVM)	0.67	0.48	0.72	0.56	0.51
Stacking (KNN)	0.67	0.48	0.72	0.56	0.51
Patient H					
XGBoost (baseline)	0.25	0.99	0.5	0.4	0.62
Hard Vote	0.89	0.99	0.98	0.94	0.98
Soft Vote	0.89	0.99	0.98	0.94	0.98
Stacking (Logit. Reg.)	0.89	0.99	0.98	0.94	0.98
Stacking (SVM)	0.67	0.99	0.92	0.8	0.91
Stacking (KNN)	0.89	0.99	0.98	0.94	0.98

Table 3: Performance evaluation of different ensemble learning methods for eight MIMIC-III patients

V. CONCLUSION

We propose a personalized ECG-based potassium abnormality detection method for patients with chronic kidney disease (CKD). By treating ECG as a time series, we first perform heartbeat-level segmentation and construct seven physiologically grounded waveform bands and band combinations. We then conduct band-level model selection using an algorithm pool comprising eight time-series classifiers, and finally fuse the selected models via ensemble strategies to balance detection accuracy and deployability. This design is well suited for real-time monitoring on wearable and edge platforms. Experimental results show that the proposed framework can accommodate inter-individual ECG variability and build personalized detection models. Among the evaluated ensemble methods, hard voting is the most stable for the majority of patients; in certain cases, it achieves AUC $\approx$ 0.99 with Precision/Recall > 0.95. Future work will focus on further improving inference efficiency and low-power deployment capability, and on extending the framework to other electrolyte disorders.

REFERENCES

[1] M. Ali, A. Alqahtani, M. W. Jones, and X. Xie, "Clustering and classification for time series data in visual analytics: A survey," IEEE Access, vol. 7, pp. 181314-181338, 2019, doi: 10.1109/ACCESS.2019.2958551.

[2] P. Ivaturi, M. Gadaleta, A. C. Pandey, M. Pazzani, S. R. Steinhubl, and G. Quer, "A comprehensive explanation framework for biomedical time series classification," IEEE J. Biomed. Health Inform., vol. 25, no. 7, pp. 2398-2408, Jul. 2021, doi: 10.1109/JBHI.2021.3060997.

[3] A. Terbuch, P. O'Leary, N. Khalili-Motlagh-Kasmaei, P. Auer, A. Zöhrer, and V. Winter, "Detecting anomalous multivariate time-series via hybrid machine learning," IEEE Trans. Instrum. Meas., vol. 72, pp. 1-11, 2023, Art no. 2503711, doi: 10.1109/TIM.2023.3236354.

[4] H. Li, Y. Cui, S. Wang, J. Liu, J. Qin, and Y. Yang, "Multivariate financial time-series prediction with certified robustness," IEEE Access, vol. 8, pp. 109133-109143, 2020, doi: 10.1109/ACCESS.2020.3001287.

[5] W.-K. Wang et al., "A systematic review of time series classification techniques used in biomedical applications," Sensors (Basel), vol. 22, no. 20, Art. no. 8016, Oct. 2022, doi: 10.3390/s22208016.

[6] E. Morales, P. Cravedi, and J. Manrique, "Management of chronic hyperkalemia in patients with chronic kidney disease: An old problem with new options," Front. Med. (Lausanne), vol. 8, Art. no. 653634, Jun. 2021, doi: 10.3389/fmed.2021.653634.

[7] S. Bianchi, F. Aucella, L. De Nicola, S. Genovesi, E. Paoletti, and G. Regolisti, "Management of hyperkalemia in patients with kidney disease: A position paper endorsed by the Italian Society of Nephrology," J. Nephrol., vol. 32, no. 4, pp. 499-516, Aug. 2019, doi: 10.1007/s40620-019-00617-y.

[8] K. L. Dittrich and R. M. Walls, "Hyperkalemia: ECG manifestations and clinical considerations," J. Emerg. Med., vol. 4, no. 6, pp. 449-455, 1986, doi: 10.1016/0736-4679(86)90174-5.

[9] A. Webster, W. Brady, and F. Morris, "Recognising signs of danger: ECG changes resulting from an abnormal serum potassium concentration," Emerg. Med. J., vol. 19, no. 1, pp. 74-77, Jan. 2002, doi: 10.1136/emj.19.1.74.

[10] D. B. Diercks, G. M. Shumaik, R. A. Harrigan, W. J. Brady, and T.-C. Chan, "Electrocardiographic manifestations: electrolyte abnormalities," J. Emerg. Med., vol. 27, no. 2, pp. 153-160, Aug. 2004, doi: 10.1016/j.jemermed.2004.04.006.

[11] E. Urtnasan, J.-H. Lee, B. Moon, H.-Y. Lee, K. Lee, and H. Youk, "Noninvasive screening tool for hyperkalemia using a single-lead electrocardiogram and deep learning: Development and usability study," JMIR Med. Inform., vol. 10, no. 6, Art. no. e34724, Jun. 2022, doi: 10.2196/34724.

[12] C. D. Galloway et al., "Development and validation of a deep-learning model to screen for hyperkalemia from the electrocardiogram," JAMA Cardiol., vol. 4, no. 5, pp. 428-436, May 2019, doi: 10.1001/jamacardio.2019.0640.

[13] C.-S. Lin et al., "A deep-learning algorithm (ECG12Net) for detecting hypokalemia and hyperkalemia by electrocardiography: Algorithm development," JMIR Med. Inform., vol. 8, no. 3, Art. no. e15931, Mar. 2020.

[14] J. A. Quintero et al., "Electrocardiographic abnormalities in patients with hyperkalemia: A retrospective study in an emergency department in Colombia," Open Access Emerg. Med., vol. 16, pp. 133-144, Jun. 2024, doi: 10.2147/OAEM.S455159.

[15] I.-M. Chiu et al., "Using deep transfer learning to detect hyperkalemia from ambulatory electrocardiogram monitors in intensive care units: Personalized medicine approach," J. Med. Internet Res., vol. 24, no. 12, Art. no. e41163, Dec. 2022, doi: 10.2196/41163.

[16] K. Taguelmimt, J. Wojak, and M. Adel, "Blood potassium concentration estimation on electrocardiogram signals using machine learning techniques," in 2022 7th International Conference on Frontiers of Signal Processing (ICFSP), Paris, France, 2022, pp. 114-119, doi: 10.1109/ICFSP55781.2022.9924675.

★★★