UDC 004.032.26:616.12-073.97

AUTOMATED CARDIAC RHYTHM CLASSIFICATION FROM 12-LEAD ECG USING ATTENTION-ENHANCED DEEP NEURAL NETWORKS

N.L. Kisialeu
Student, Faculty of Applied Mathematics and Computer Science, Belarusian State University, Republic of Belarus
nkisialeu@gmail.com

A.M. Nedzved
Head of the Department, Associate Professor, Doctor of Technical Sciences, Belarusian State University, Republic of Belarus
Anedzved@bsu.by

N.L. Kisialeu
Student at the Faculty of Applied Mathematics and Computer Science, Belarusian State University. Research interests: deep learning, medical signal processing, computer vision, time series classification, automated machine learning.

A.M. Nedzved
Head of the Department, Associate Professor, Doctor of Technical Sciences, Belarusian State University. Research interests: image processing, computer vision, pattern recognition, mathematical morphology, biomedical image analysis.

Abstract. This paper presents a comprehensive deep learning framework for automated multi-class ECG rhythm classification using 12-lead electrocardiogram recordings from the Chapman–Shaoxing database (10,615 records, 8 rhythm classes). A 1D Squeeze-and-Excitation ResNet-18 (SE-ResNet18) architecture is proposed for temporal feature extraction. Seven experimental configurations are systematically evaluated, investigating the impact of clinical metadata (14 features), NeuroKit2 [1] morphological features, and demographic inputs on classification performance. The best configuration achieves test accuracy of 94.2%, macro AUC of 0.956, and macro F1-score of

0.851. A universal lead-agnostic model (UniversalECGNet) with shared single-channel backbone, learnable lead embeddings, and attention pooling enables deployment across devices with 1–12 leads without retraining.

Keywords: ECG classification, deep learning, SE-ResNet18, squeeze-and-excitation, cardiac arrhythmia, 12-lead electrocardiogram, clinical metadata, universal ECG model, attention pooling, lead dropout.

Introduction. Electrocardiography (ECG) remains the most widely used non-invasive diagnostic tool in cardiology, providing critical information about cardiac rhythm, conduction abnormalities, and myocardial ischemia. Clinical ECG interpretation is performed by cardiologists who visually analyze waveform morphology, intervals, and rhythm regularity – a process that is time-consuming, subject to inter-observer variability, and dependent on specialist availability. Automated rhythm classification from ECG recordings is therefore a critical component of computer-aided diagnosis systems, enabling timely detection of arrhythmias including atrial fibrillation (AFIB), sinus bradycardia (SB), sinus tachycardia (ST), atrial flutter (AF), and other clinically significant conditions [2].

According to the World Health Organization, cardiovascular diseases account for approximately 19.8 million deaths in 2022, representing 32% of all global deaths [3]. Early and accurate detection of cardiac rhythm abnormalities through ECG analysis plays a critical role in timely intervention and treatment planning. However, manual ECG interpretation is subject to considerable inter-observer variability, particularly for complex arrhythmias, which motivates the development of automated classification systems.

Traditional rule-based approaches to ECG classification rely on hand-crafted feature extraction pipelines: QRS complex detection, P-wave and T-wave delineation, interval measurement (PR, QT, QRS), and axis calculation, followed by threshold-based or decision-tree classification. While interpretable, these methods lack generalization across diverse patient populations, recording conditions, and device configurations. Deep learning methods, particularly convolutional neural networks (CNNs), have demonstrated superior performance by learning discriminative features directly from raw signal data, eliminating the need for manual feature engineering [4]. Hannun et al. demonstrated cardiologist-level arrhythmia detection using a 34-layer CNN trained on 91,232 single-lead ECG recordings [2]. Ribeiro et al. extended this to 12-lead ECGs using a residual network, achieving diagnostic accuracy comparable to cardiology residents [5].

The PhysioNet/Computing in Cardiology Challenge 2020 [6] established standardized benchmarks for 12-lead ECG classification across multiple diagnostic categories. The top-performing solutions employed deep residual networks augmented with attention mechanisms, achieving challenge scores above 0.5 on a weighted multi-label metric. Notably, the 2nd-place solution combined ResNet-18 with Squeeze-and-Excitation (SE) blocks [7], demonstrating that channel-wise attention recalibration significantly improves feature discrimination for ECG signals. However, most existing approaches are designed for a fixed lead configuration (typically 12-lead) and cannot adapt to variable input channels. This is a significant limitation in clinical practice, where ECG acquisition devices range from single-lead wearable monitors (Apple Watch, AliveCor KardiaMobile) to standard 3-lead, 6-lead, and 12-lead hospital systems.

This work addresses three research objectives: (1) developing an SE-ResNet18 architecture optimized for multi-class ECG rhythm classification with systematic investigation of input modalities across seven experimental configurations; (2) analyzing the contribution of clinical metadata, morphological features extracted via NeuroKit2, and minimal demographic information to classification quality; and (3) proposing a universal lead-agnostic model (UniversalECGNet) that operates on arbitrary subsets of 1–12 ECG leads using a shared backbone with attention-based lead aggregation and lead dropout augmentation.

The remainder of this paper is organized as follows. Section 2 describes the dataset characteristics, class distribution, and preprocessing pipeline. Section 3 presents the SE-ResNet18 architecture with detailed description of the squeeze-and-excitation mechanism. Section 4 introduces the UniversalECGNet architecture with its shared backbone, lead embeddings, and

attention pooling mechanism. Section 5 details the training configuration including optimizer, scheduler, loss function, and augmentation strategies. Section 6 presents experimental results across seven configurations with comparative tables. Section 7 provides per-class analysis with confusion patterns. Section 8 discusses the implications of the findings, and Section 9 concludes with a summary and future directions.

Dataset and preprocessing. The Chapman–Shaoxing 12-lead ECG database from PhysioNet [8, 9] is used in this study. The database contains 12-lead ECG recordings from patients at Shaoxing People's Hospital, sampled at 500 Hz with 10-second duration (5,000 samples per channel per recording). Each recording is accompanied by diagnostic labels and clinical metadata provided in the accompanying metadata files. After filtering rare rhythm classes with insufficient representation for reliable training (AVNRT – atrioventricular nodal reentrant tachycardia, AVRT – atrioventricular reentrant tachycardia, SAAWR – sinus atrium to atrial wandering rhythm), 10,615 records across 8 rhythm classes are retained.

The class distribution exhibits significant imbalance (Table 1). The ratio between the largest (SB) and smallest (AT) classes is 32:1, presenting a challenge for model training. Stratified splitting with random seed 42 preserves class proportions across train/validation/test sets of 7,430/1,592/1,593 records.

Table 1. Class distribution in the Chapman–Shaoxing dataset

Class	**Abbr.**	**Total**	**Train**	**Val**	**Test**
Sinus Bradycardia	SB	3,889	2,722	583	584
Sinus Rhythm	SR	1,826	1,278	274	274
Atrial Fibrillation	AFIB	1,780	1,246	267	267
Sinus Tachycardia	ST	1,568	1,098	235	235
Supraventr. Tachycardia	SVT	587	411	88	88
Atrial Flutter	AF	445	311	67	67
Sinus Arrhythmia	SA	399	279	60	60
Atrial Tachycardia	AT	121	85	18	18

Signal preprocessing includes per-channel normalization to zero mean and unit variance using training set statistics. Clinical metadata comprises 14 features: PatientAge, Gender (one-hot encoded as 2 binary features), VentricularRate, AtrialRate, QRSDuration, QTInterval, QTCorrected, RAxis, TAxis, QRSCount, QOnset, QOffset, TOffset, and POnset. All metadata features are standardized using z-score normalization. Additionally, morphological features are extracted using the NeuroKit2 library: R-peak amplitudes and locations, P/Q/S/T wave boundary positions, RR interval statistics, and heart rate variability (HRV) metrics – 28 features in total. Feature extraction is precomputed and cached in a CSV file. Analysis revealed that 10–50% of NeuroKit2 features contain fallback values due to detection failures on certain recordings, motivating both unfiltered and filtered (>5% fallback threshold) experimental variants.

SE-ResNet18 architecture. The core classification architecture is a 1D adaptation of ResNet-18 [10] augmented with Squeeze-and-Excitation (SE) blocks [7], inspired by the 2nd-place solution of the PhysioNet 2020 Challenge. All 2D convolutions in the original ResNet-18 are replaced with 1D convolutions (Conv1d) with kernel size 7 and stride 2 in the initial layer. The network processes 12-channel input tensors of shape (12, 5,000) through an initial convolutional layer (64 filters) followed by batch normalization, ReLU activation, and max pooling (kernel 3, stride 2). Four residual stages with [2, 2, 2, 2] basic blocks contain 64, 128, 256, and 512 filters respectively. Each residual block consists of two Conv1d layers with batch normalization and ReLU, with a skip connection adding the input to the output.

Each SE block enhances the residual block output through channel-wise attention recalibration. First, global average pooling along the temporal dimension produces a channel descriptor vector of size C (where C is the number of filters). This descriptor passes through a two-

layer fully connected bottleneck: FC(C → C/r) with ReLU activation, followed by FC(C/r → C) with sigmoid activation, where r = 16 is the reduction ratio. The resulting sigmoid-activated vector provides channel-wise scaling factors that multiply the original feature map, enabling the network to emphasize diagnostically relevant temporal patterns (e.g., P-wave morphology, QRS duration, T-wave alternans) while suppressing uninformative channels. Adaptive average pooling reduces the temporal dimension to 1, producing a 512-dimensional feature vector for the classification head.

For metadata integration (SE-ResNet18-Meta variant), clinical features are processed through a separate MLP branch: Linear(N_features → 64) → ReLU → Dropout(0.3), producing a 64-dimensional metadata embedding. This embedding is concatenated with the 512-dimensional CNN feature vector, forming a 576-dimensional combined representation that passes through the final classification layer Linear(576 → 8). The MLP branch input dimension varies by experiment: 14 for full clinical metadata, 36 for metadata + all NeuroKit2 features, 27 for metadata + filtered NeuroKit2 features, 3 for age + gender only, and 16 for age + gender + filtered NeuroKit2 features.

Universal lead-agnostic model (UniversalECGNet). A key contribution of this work is UniversalECGNet –a single model that operates on any subset of 1–12 ECG leads without retraining. The architecture (9.18M parameters) employs a shared single-channel SE-ResNet18 backbone that processes each lead independently. The processing pipeline consists of seven steps:

Step 1. Input tensor of shape (B, L, 5000), where B is batch size and L is the number of available leads (1 to 12), along with lead index tensor (B, L) identifying which leads are present (0=I, 1=II, 2=III, 3=aVR, 4=aVL, 5=aVF, 6=V1, ..., 11=V6). Step 2. Reshape to (B×L, 1, 5000), treating each lead as an independent single-channel sample. Step 3. Process through shared SE-ResNet18(in_channels=1) with gradient checkpointing for memory efficiency, producing per-lead features of shape (B×L, 512). Step 4. Reshape back to (B, L, 512). Step 5. Learnable lead embedding layer maps lead indices to 64-dimensional vectors encoding each lead's identity and spatial characteristics: Embedding(12, 64). Step 6. Concatenate CNN features with lead embeddings: (B, L, 576), then project via Linear(576 → 512): (B, L, 512). Step 7. Attention pooling with a learned 512-dimensional query vector computes attention scores over leads via dot product, applies masked softmax (masking absent leads), and produces weighted sum: single 512-dimensional patient representation regardless of input lead count. Final classification: Linear(512 → 8).

Lead dropout augmentation is critical for training robustness. During training, 50% of batches see all 12 leads, while the remaining 50% see a random subset of 1–11 leads (uniformly sampled count, then random lead selection). Masked leads are zeroed out and excluded from attention computation via the mask. This augmentation forces the model to extract useful representations from arbitrary lead combinations rather than overfitting to the full 12-lead configuration. Ablation experiments confirm that without lead dropout, the universal model maintains 12-lead performance but degrades sharply (−10–15% accuracy) when evaluated on reduced lead configurations, demonstrating the augmentation's essential role in achieving lead-agnostic generalization.

The attention pooling mechanism constitutes the central innovation of UniversalECGNet. The LeadAttentionPool module implements a two-layer neural network: a linear transformation from the 512-dimensional feature space to a 128-dimensional hidden space with Tanh activation, followed by a projection to a single scalar score per lead. For available leads, these scores are normalized via softmax to produce attention weights summing to 1.0; for absent leads (either masked during lead dropout training or physically unavailable during inference), scores are set to negative infinity before softmax, effectively zeroing their contribution. The final patient-level representation is computed as the attention-weighted sum of per-lead feature vectors. Analysis of the learned attention weights on the test set reveals that lead II consistently receives the highest average attention weight (0.15–0.18), followed by V1 (0.10–0.14) and aVF (0.09–0.12), which aligns with clinical knowledge about the diagnostic value of these leads for rhythm classification.

The shared backbone architecture presents a computational challenge during training: with batch size B and L leads per sample, the effective batch processed by the backbone is B×L, reaching up to 32×12 = 384 forward passes per training step. Gradient checkpointing is applied to the four residual stages to reduce peak GPU memory consumption by approximately 60% at the cost of ~30% additional computation time. The total parameter count of UniversalECGNet is 9,179,977 parameters: the shared SE-ResNet18 backbone accounts for 9,065,728 (98.8%), the learnable lead embedding layer contributes 768 parameters (12 leads × 64 dimensions), the feature projection layer adds 295,424, and the attention pooling module contains 66,049 parameters. During inference, standard forward propagation without checkpointing is used for maximum throughput. The lead embedding vectors are L2-normalized before concatenation with the signal, ensuring that the identity information does not dominate the raw ECG features in magnitude.

Training configuration. All models are trained using the AdamW optimizer [11] with initial learning rate 0.001 and weight decay 0.0001. The learning rate schedule employs CosineAnnealingWarmRestarts with $T_0 = 10$ and T_mult = 2, providing periodic warm restarts that help escape local minima and explore the loss landscape more thoroughly. The loss function is cross-entropy with inverse-frequency class weights (computed from training set class distribution) and label smoothing $\varepsilon = 0.1$, which prevents overconfident predictions and improves calibration. Gradient clipping at max norm 1.0 ensures training stability.

Signal augmentations are applied online during training with per-sample probability of 0.5 each: (1) baseline wander injection –sinusoidal signal at 0.1–1.0 Hz simulating respiratory and movement artifacts; (2) 50 Hz power line noise –additive sinusoidal at exactly 50 Hz with random amplitude; (3) Gaussian noise at SNR 20–40 dB simulating electronic noise; (4) amplitude scaling by a random factor in [0.8, 1.2] simulating gain variations across devices. These augmentations improve model robustness to real-world recording conditions. Early stopping monitors macro F1 on the validation set with patience of 10 epochs (15 for NeuroKit2 variants due to slower convergence). Batch size is 64 for standard models and 32 for the universal model due to higher memory requirements. The complete set of training hyperparameters and augmentation settings is provided in Table 2.

Table 2. Training hyperparameters and augmentation configuration

Parameter	**Value**
Optimizer	AdamW
Learning rate	0.001
Weight decay	0.0001
Scheduler	CosineAnnealingWarmRestarts (T_0=10, T_mult=2)
Loss function	CrossEntropy + class weights + label smoothing (ε=0.1)
Gradient clipping	max norm 1.0
Batch size	64 (standard) / 32 (universal)
Early stopping	patience 10 on val macro F1
Baseline wander	sinusoidal, 0.1–1.0 Hz
Power line noise	50 Hz additive sinusoidal
Gaussian noise	SNR 20–40 dB
Amplitude scaling	±20%
Lead dropout (universal)	p = 0.5

Experimental results. Seven experimental configurations are systematically evaluated on the held-out test set (N = 1,593). All experiments share the same data splits, augmentations, and training hyperparameters (except input dimensions), enabling fair comparison of input modalities.

A summary of all experimental configurations and their test set performance metrics is presented in Table 3.

Table 3. Comparative results across seven experimental configurations (test set, N = 1,593)

#	Configuration	Input	Acc.	P (macro)	F1 (macro)	AUC (macro)
1	Signal only	ECG 12×5000	0.938	0.862	0.859	0.946
2	Signal + metadata	ECG + 14 meta	0.942	0.855	0.851	0.956
3	+ all NeuroKit2	ECG + 36 feat	0.932	0.849	0.850	0.968
4	+ filtered NeuroKit2	ECG + 27 feat	0.923	0.836	0.831	0.966
5	Signal + age/gender	ECG + 3 feat	0.935	0.841	0.843	0.937
6	+ age/gender + NK2	ECG + 16 feat	0.935	0.829	0.846	0.965
7	Universal (12-lead)	1–12 leads	0.936	0.841	0.838	0.953

Experiment 1 (signal only): SE-ResNet18 with raw 12-lead ECG input, no additional features. Accuracy 93.8%, macro precision 0.862, macro F1 0.859, macro AUC 0.946. This establishes a strong baseline demonstrating that SE-ResNet18 effectively extracts discriminative features from raw signals.

Experiment 2 (signal + 14 metadata): SE-ResNet18-Meta with MLP branch (14 → 64) for clinical features. This achieves the best overall performance: accuracy 94.2%, macro precision 0.855, macro recall 0.850, macro F1 0.851, weighted F1 0.942, macro AUC 0.956, Cohen's kappa 0.926, MCC 0.926, log loss 0.466, top-2 accuracy 0.976. The clinical metadata – particularly ventricular rate, QRS duration, QT interval, and patient age – provides complementary information that the CNN cannot fully extract from raw signals alone.

Detailed metrics for the best model (Experiment 2) are presented in Table 4.

Table 4. Global metrics of the best model (Experiment 2, test set)

Metric	Value
Accuracy	0.9422
Precision (macro)	0.8548
Recall (macro)	0.8495
F1-score (macro)	0.8512
F1-score (weighted)	0.9415
AUC (macro, OvR)	0.9562
Cohen's Kappa	0.9259
MCC	0.9260
Log Loss	0.4662
Top-2 Accuracy	0.9755

Experiment 3 (signal + metadata + all NeuroKit2): MLP branch with 36 inputs (14 metadata + 28 morphological features). Accuracy drops to 93.2% while AUC rises to 0.968, suggesting better ranking ability but worse calibration. The 10–50% fallback values in NeuroKit2 features introduce noise that dilutes the useful metadata signal. Experiment 4 (filtered NeuroKit2): columns with >5% fallback are removed, leaving 27 features. Accuracy further drops to 92.3%, AUC 0.966 – even filtered features do not improve over pure metadata.

Experiment 5 (signal + age/gender): minimal demographic features (age + 2 gender one-hot = 3 inputs). Accuracy 93.5%, AUC 0.937. Age and gender provide marginal benefit over signal-only, confirming that clinical metadata's value comes primarily from cardiac-specific measurements (rates, intervals, axes) rather than demographics. Experiment 6 (age/gender + filtered NeuroKit2): 16 features total, accuracy 93.5%, AUC 0.965 – NeuroKit2 features partially compensate for missing clinical metadata but do not exceed the full metadata configuration.

Experiment 7 (UniversalECGNet): the lead-agnostic model is evaluated on four lead configurations without any retraining between configurations. 12-lead: accuracy 93.6%, macro F1 0.838, AUC 0.953. 6-lead (limb leads I, II, III, aVR, aVL, aVF): accuracy 93.7%, F1 0.837 – virtually no degradation, demonstrating that limb leads carry nearly all discriminative rhythm

information. 3-lead (I, II, V1): accuracy 93.8%, F1 0.842 – slightly better than 12-lead, as reduced lead count eliminates noise from redundant leads and allows attention to focus on the most informative subset. Single-lead (II): accuracy 91.6%, F1 0.787, AUC 0.941 – F1 drops by 5 percentage points, but lead II alone still achieves strong AUC, confirming it carries the most rhythm-discriminative information.

Table 5 summarizes the universal model's performance across four lead configurations evaluated without any retraining.

Table 5. UniversalECGNet: performance across lead configurations (test set)

Config.	Leads	Accuracy	F1 (macro)	F1 (wtd)	AUC (macro)
12-lead	I, II, III, aVR, aVL, aVF, V1–V6	0.936	0.838	0.936	0.953
6-lead	I, II, III, aVR, aVL, aVF	0.937	0.837	0.936	0.945
3-lead	I, II, V1	0.938	0.842	0.939	0.951
1-lead	II	0.916	0.787	0.915	0.941

Per-class analysis and error patterns. Detailed per-class analysis of the best model (Experiment 2) reveals distinct performance tiers. Well-represented classes achieve excellent classification: SB (sensitivity 0.990, specificity 0.997, F1 0.992), ST (sensitivity 0.975, specificity 0.992, F1 0.964), AFIB (sensitivity 0.959, specificity 0.985, F1 0.943), SR (sensitivity 0.938, specificity 0.988, F1 0.940), and SVT (sensitivity 0.932, specificity 0.995, F1 0.921).

Per-class sensitivity, specificity, positive predictive value (PPV), and F1-score for the best model are shown in Table 6.

Table 6. Per-class metrics of the best model (Experiment 2, test set)

Class	Sens.	Spec.	PPV	F1	N (test)
SB	0.990	0.997	0.995	0.992	584
SR	0.938	0.988	0.941	0.940	274
AFIB	0.959	0.985	0.928	0.943	267
ST	0.975	0.992	0.954	0.964	235
SVT	0.932	0.995	0.911	0.921	88
AF	0.642	0.990	0.741	0.688	67
SA	0.750	0.994	0.818	0.783	60
AT	0.611	0.994	0.550	0.579	18

Minority classes show clinically interpretable error patterns. Atrial flutter (AF, 67 test samples) achieves sensitivity 0.642, specificity 0.990, F1 0.688. Confusion matrix analysis reveals that 12 of 67 AF samples are misclassified as AFIB – a clinically expected confusion since both conditions involve abnormal atrial activity with similar morphological features on ECG, differing primarily in regularity and atrial rate (AF: regular ~300 bpm, AFIB: irregular ~350–600 bpm). Sinus arrhythmia (SA, 60 test samples) achieves F1 0.783, with 13 of 60 samples misclassified as SR – expected since SA is a physiological variant of sinus rhythm where the only distinguishing feature is respiratory-linked heart rate variability. Atrial tachycardia (AT, 18 test samples) shows the lowest F1 of 0.579, primarily due to extreme class imbalance (only 121 training samples) and morphological overlap with SVT and ST.

Discussion. The experimental results yield several important findings. First, the SE attention mechanism provides consistent improvement over standard ResNet-18, with minority class F1 gains of 3–5% – the channel-wise recalibration helps the network focus on subtle morphological differences that distinguish challenging class pairs (AF vs. AFIB, SA vs. SR). Second, clinical metadata from the diagnostic system provides the highest-value supplementary information, with

the 14-feature metadata configuration outperforming all NeuroKit2 variants. This is likely because the diagnostic system's measurements are more reliable than automated feature extraction from raw signals, which suffers from detection failures on certain recording types.

Third, the universal model's near-zero degradation from 12 to 6 leads has significant clinical implications: standard limb leads alone are sufficient for accurate rhythm classification, suggesting that precordial leads (V1–V6) are largely redundant for this task. The counterintuitive improvement at 3 leads can be attributed to the attention mechanism's ability to focus more precisely when fewer leads are available, avoiding noise from redundant channels. The lead dropout augmentation is the key enabler of this generalization – without it, the model memorizes the 12-lead input structure and fails catastrophically on reduced configurations.

Fourth, the primary performance bottleneck remains class imbalance for rare arrhythmias. Atrial Tachycardia with only 121 training samples achieves F1 0.579, while SB with 3,889 samples achieves F1 0.992. Future improvements could employ synthetic oversampling (SMOTE for time series), few-shot learning techniques, or curriculum learning strategies that progressively focus training on difficult minority classes.

The comparison with related work provides additional context for the achieved results. Hannun et al. [2] reported sensitivity of 0.79 and specificity of 0.97 for arrhythmia detection on single-lead ambulatory ECG using a 34-layer CNN. The present work achieves comparable specificity (0.99 average) with higher macro sensitivity (0.85) on 8-class multi-label classification from standard 12-lead recordings. Ribeiro et al. [5] achieved 80% F1-score on 6 ECG abnormalities using a ResNet architecture; our SE-ResNet18 with metadata achieves 85.1% macro F1 on 8 rhythm classes, suggesting that the squeeze-and-excitation mechanism and clinical metadata provide meaningful improvements. The universal model's ability to maintain 91.6% accuracy on a single lead has practical implications for wearable ECG devices (smartwatches, portable monitors) that typically provide only lead I or lead II, enabling clinically useful rhythm screening outside hospital settings.

Several limitations should be noted. The dataset originates from a single hospital (Shaoxing People's Hospital), which may limit generalizability to other populations and recording equipment. The 8-class rhythm taxonomy excludes many clinically important conditions (myocardial infarction, bundle branch blocks, ventricular tachycardia). The NeuroKit2 feature extraction demonstrated high failure rates (10–50% fallback values), suggesting that production deployment should rely on the signal-only or signal+metadata configurations. The universal model was trained exclusively on Chapman–Shaoxing data; cross-dataset validation on CPSC-2018, PTB-XL, or MIT-BIH databases would strengthen the generalization claims.

To facilitate practical deployment, a REST API inference service has been developed using the FastAPI framework. The service accepts raw ECG signal data in JSON format, performs preprocessing (per-channel normalization using precomputed training statistics), runs inference through the trained model, and returns the predicted rhythm class along with confidence scores for all eight classes. The API supports both the standard 12-lead SE-ResNet18 model and the universal lead-agnostic model, automatically detecting the number of input leads and routing to the appropriate inference pathway. Response latency on a single NVIDIA RTX 3060 GPU averages 47 milliseconds per recording (including preprocessing), enabling real-time classification in clinical monitoring workflows.

An exploratory experiment was conducted to evaluate the capability of multimodal large language models for ECG rhythm classification [12]. GPT-4o Vision was presented with rendered 12-lead ECG images in standard clinical layout and prompted to classify the rhythm into one of eight categories. On a random subset of 100 test recordings, GPT-4o achieved 42% accuracy compared to 94.2% for the specialized SE-ResNet18 model. While the LLM demonstrated rudimentary medical knowledge by correctly identifying sinus bradycardia and tachycardia at above-chance rates, it failed almost entirely on minority classes such as atrial flutter and atrial tachycardia. This

comparison underscores the continued necessity of purpose-built deep learning architectures for safety-critical diagnostic tasks.

Conclusion. A comprehensive deep learning framework for multi-class ECG rhythm classification has been developed and evaluated across seven experimental configurations on the Chapman–Shaoxing 12-lead ECG database. The principal results are:

1. SE-ResNet18 with 14 clinical metadata features achieves the best classification performance: accuracy 94.2%, macro AUC 0.956, macro F1 0.851, Cohen's kappa 0.926 across 8 rhythm classes, outperforming signal-only, NeuroKit2, and demographic-only variants.

2. Systematic ablation reveals that clinical metadata provides the optimal accuracy–complexity trade-off, while NeuroKit2 morphological features with high fallback rates degrade performance despite higher AUC.

3. UniversalECGNet with shared backbone, lead embeddings, attention pooling, and lead dropout enables a single 9.18M-parameter model to operate across 1–12 lead configurations: 93.6% accuracy on 12-lead, 93.8% on 3-lead, 91.6% on single-lead, with negligible degradation from 12 to 6 leads.

4. Clinically interpretable error patterns – AF/AFIB and SA/SR confusion – align with known morphological similarities between these arrhythmia pairs.

Future work includes extending classification to the full diagnostic category set, applying transformer architectures for long-range temporal dependencies, implementing federated learning for privacy-preserving multi-hospital training, investigating few-shot learning for rare arrhythmias, and conducting clinical validation in real-time patient monitoring scenarios. The FastAPI-based inference service developed alongside the models enables straightforward integration into clinical decision support systems.

Reference list

[1] Makowski D., Pham T., Lau Z., Brammer J., Lespinasse F., Pham H., Schölzel C., Chen S. NeuroKit2: A Python toolbox for neurophysiological signal processing. Behavior Research Methods. 2021; 53(4):1689–1696. DOI: 10.3758/s13428-020-01516-y.

[2] Hannun A., Rajpurkar P., Haghpanahi M., Tison G., Bourn C., Turakhia M., Ng A. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature Medicine. 2019; 25(1):65–69. DOI: 10.1038/s41591-018-0268-3.

[3] World Health Organization. Cardiovascular diseases (CVDs): fact sheet. WHO; 2024. Available from: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

[4] LeCun Y., Bengio Y., Hinton G. Deep learning. Nature. 2015; 521(7553):436–444. DOI: 10.1038/nature14539.

[5] Ribeiro A., Stokes M., Meira S., Kroenke T., Paixao A., Nunes M., Ribeiro A. Automatic diagnosis of the 12-lead ECG using a deep neural network. Nature Communications. 2020; 11:1760. DOI: 10.1038/s41467-020-15432-4.

[6] Alday E., Gu A., Shah A., Robber C., Sharma A., Nemati S., Clifford G. Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020. Physiological Measurement. 2020; 41(12):124003. DOI: 10.1088/1361-6579/abc960.

[7] Hu J., Shen L., Sun G. Squeeze-and-Excitation Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018:7132–7141. DOI: 10.1109/CVPR.2018.00745.

[8] Zheng J., Zhang J.,Danioko S., Yao H.,Guo H., Rakovski C.A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients. Scientific Data.2020;7:48.DOI:10.1038/s41597-020-0386-x.

[9] Goldberger A., Amaral L., Glass L., Hausdorff J., Ivanov P., Mark R., Mietus J., Moody G., Peng C., Stanley H. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation. 2000; 101(23):e215–e220. DOI: 10.1161/01.CIR.101.23.e215.

[10] He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition. Proceedings of the IEEE CVPR. 2016:770–778. DOI: 10.1109/CVPR.2016.90.

[11] Loshchilov I., Hutter F. Decoupled weight decay regularization. Proceedings of the 7th International Conference on Learning Representations (ICLR). 2019.

[12] Tohnson K., Torres Soto J., Glicksberg B., Shameer K., Miotto R., Ali M., Ashley E., Dudley J. Artificial intelligence in cardiology. Journal of the American College of Cardiology. 2018; 71(23):2668–2679. DOI: 10.1016/j.jacc.2018.03.521.

Author's contribution

Nikita Kisialeu – problem formulation, SE-ResNet18 and UniversalECGNet architecture design, implementation of all seven experimental configurations, data preprocessing and augmentation pipeline, NeuroKit2 feature extraction, experimental evaluation and ablation study, per-class error analysis, FastAPI inference service development, manuscript preparation.

Alexander Nedzved – scientific supervision, problem formulation, methodology guidance, result validation, manuscript review.

АВТОМАТИЧЕСКАЯ КЛАССИФИКАЦИЯ СЕРДЕЧНОГО РИТМА ПО 12-КАНАЛЬНОЙ ЭКГ С ПОМОЩЬЮ ГЛУБОКИХ НЕЙРОННЫХ СЕТЕЙ С МЕХАНИЗМОМ ВНИМАНИЯ

Н.Л. Киселев
Студент, факультет прикладной математики и информатики, Белорусский государственный университет, Республика Беларусь

А.М. Недзьведь
Заведующий кафедрой, доцент, доктор технических наук, Белорусский государственный университет, Республика Беларусь

Аннотация. Представлена комплексная система глубокого обучения для многоклассовой классификации ритма ЭКГ по 12-канальным записям базы Chapman–Shaoxing (10 615 записей, 8 классов). Предложена архитектура 1D SE-ResNet18 с интеграцией клинических метаданных. Проведено 7 экспериментов. Лучший результат: точность 94,2%, AUC 0,956, macro F1 0,851. Разработана универсальная модель UniversalECGNet для 1–12 отведений без переобучения.

Ключевые слова: классификация ЭКГ, глубокое обучение, SE-ResNet18, сердечная аритмия, 12-канальная ЭКГ, механизм сжатия-возбуждения, клинические метаданные, универсальная модель, механизм внимания, аугментация отведений.