A Multi-Scale CNN–BiLSTM Framework for Robust ECG-Based User Authentication

Mohamed Abdalla Elsayed Azab and Victoriia Korzhuk

***Abstract*—Reliable biometric authentication remains a critical challenge for modern security systems, particularly in applications requiring continuous verification and strong resistance to spoofing attacks. Among physiological biometrics, the electrocardiogram (ECG) offers inherent liveness information, subject-specific morphological patterns, and robustness against external forgery. This paper presents a novel deep learning architecture for ECG-based user authentication that jointly models spatial morphology and temporal dynamics through an integrated multi-scale convolutional neural network (CNN) and bidirectional long short-term memory (BiLSTM) framework. The proposed model uses parallel convolutional branches with different kernel sizes to simultaneously capture discriminative characteristics of the P-wave, QRS complex, and T-wave at multiple temporal scales. The extracted multi-scale features are fused and subsequently processed by a BiLSTM layer to exploit both forward and backward temporal dependencies across heartbeat sequences. Extensive experiments are conducted on the publicly available PTB-XL dataset using a subject-independent evaluation protocol. The proposed approach achieves an authentication accuracy of 99.2% and an equal error rate of 0.8%, outperforming conventional CNN, LSTM, and unidirectional CNN–LSTM baselines across all evaluation metrics. The findings indicate a noteworthy aspect of integrating multi-scale morphological research with bidirectional temporal modeling. This combination significantly enhances ECG-based biometric authentication. Enhances its robustness, hence reducing the likelihood of failure in atypical conditions. Moreover, reliability increases, which is essential for security-related matters such as this.**

***Keywords*—ECG biometrics, user authentication, electrocardiogram, liveness-aware authentication, deep learning, multi-scale convolutional neural networks, bidirectional LSTM, hybrid CNN–RNN architecture, temporal sequence modeling, morphological feature extraction, biometric security.**

I. Introduction

Secure and reliable user authentication has become a foundational requirement for modern digital systems, particularly in environments that demand continuous access control, high usability, and strong resistance to spoofing and replay attacks. Conventional authentication mechanisms based on passwords or physical tokens remain vulnerable to theft, duplication, and social engineering, while many widely adopted biometric modalities rely on externally observable traits that can be imitated or captured without user awareness [1]. These limitations motivate the exploration of physiological signals that inherently encode liveness and are closely tied to internal biological processes.

The electrocardiogram (ECG) represents a compelling biometric modality due to its intrinsic association with cardiac activity and its continuous generation by the human body. Unlike surface-level biometric traits, ECG signals arise from the electrical depolarization and repolarization of the heart, providing an inherent guarantee of liveness that is difficult to replicate artificially [2]. In addition to this security advantage, ECG waveforms exhibit subject-specific characteristics shaped by individual cardiac anatomy, electrophysiological properties, and physiological conditions. These properties make ECG-based authentication particularly attractive for applications requiring both strong security guarantees and unobtrusive user interaction.

Even though there are some good points about using ECG for authentication, putting together a solid system like that is not straightforward. The signals from the heart are always changing over time, and they do not stay the same from one beat to the next or even between different times you record them [3]. It seems like the shape of the waves shifts around a lot. Discriminative identity information is distributed across multiple components of the cardiac cycle, including the P-wave, QRS complex, and T-wave, each occupying different temporal scales and exhibiting distinct morphological patterns. Effective authentication therefore requires feature representations capable of capturing fine-grained local structures while simultaneously modeling longer-term temporal relationships across sequential heartbeats.

Deep learning offers a natural framework for addressing these challenges by enabling end-to-end learning directly from ECG signals. Convolutional neural networks (CNNs) are well suited for extracting local morphological patterns, while recurrent architectures are effective for modeling temporal dependencies. However, ECG morphology is inherently multi-scale: sharp, high-frequency transitions characterize the QRS complex, whereas smoother, lower-frequency structures define the P- and T-waves [4]. A single convolutional receptive field may emphasize certain waveform components while neglecting others, leading to incomplete or biased representations. Furthermore, temporal dependencies in ECG sequences are not strictly unidirectional; contextual information from both preceding and subsequent heartbeats can contribute to more stable and discriminative identity modeling [5].

These observations motivate the need for authentication architectures that jointly address two fundamental aspects of ECG biometrics: multi-scale morphological representation and comprehensive temporal modeling. Capturing ECG features at multiple temporal resolutions allows the model to

exploit complementary identity cues embedded across waveform components, while bidirectional temporal analysis enables the incorporation of full contextual information within heartbeat sequences. The integration of these capabilities within a unified architecture has the potential to significantly enhance robustness, discrimination power, and generalization across subjects.

This work proposes a hybrid deep learning architecture for ECG-based user authentication that combines multi-scale convolutional feature extraction with bidirectional LSTM temporal modeling. Parallel convolutional branches with heterogeneous kernel sizes capture complementary morphological characteristics of ECG waveforms. The fused features are processed by a BiLSTM to model bidirectional temporal dependencies across heartbeat sequences.

II. Literature Review

Deep learning has really pushed things forward lately in how we use ECG for biometrics, like authenticating people based on their heart signals. It covers both identifying who someone is and just verifying if its them [6]. A prominent line of work adopts Siamese or metric-learning frameworks that directly optimize similarity between ECG segments from the same subject [7]. The EDITH framework exemplifies this approach, demonstrating low equal error rates while reducing the number of heartbeats required for reliable verification, highlighting the effectiveness of discriminative representation learning in ECG biometrics [8]. Building on this idea, ensemble-based Siamese architectures further improve robustness by combining multiple pairwise learners, yielding strong performance on standard ECG datasets and demonstrating resilience to session variability and noise [9].

In parallel, end-to-end classification and sequence-modeling approaches have been explored. CNN-based models effectively capture morphological characteristics of ECG waveforms, while LSTM-based architectures exploit temporal dependencies across heartbeats [10]. Hybrid CNN–LSTM pipelines report high authentication and identification accuracy and confirm the suitability of deep models for ECG biometrics. However, many such approaches rely on single-scale convolutional processing and are sensitive to segmentation strategies, potentially limiting their ability to capture complementary waveform features across different temporal resolutions [11]. Single-beat ECG authentication has gained attention due to its low-latency advantages [12]. Recent 1D-CNN frameworks demonstrate that carefully segmented and normalized single heartbeats can support reliable identification. While attractive for rapid authentication, purely convolutional designs may underutilize inter-beat temporal cues that can enhance robustness when morphological information alone is insufficient [13].

Closely related work in user identification and de-duplication within ECG systems further informs authentication research [14]. Lead-agnostic self-supervised learning methods have been proposed to address variability in ECG lead configurations, introducing ECG-specific augmentations and demonstrating strong performance on the PTB-XL dataset [15]. Similarly, compact verification models designed for gallery–probe matching in PTB-XL emphasize efficiency and generalization, reflecting deployment constraints common to biometric systems.

Collectively, the literature indicates that CNNs are effective for morphological feature extraction, while recurrent and pairwise learning strategies enhance verification performance. Nevertheless, many ECG authentication models employ single-scale convolutions or unidirectional temporal modeling, and evaluation protocols remain heterogeneous across studies [16]. In contrast, multi-scale convolutional architectures and bidirectional recurrent models are well established in diagnostic ECG analysis, where they consistently improve classification performance by capturing features across multiple temporal resolutions and exploiting full contextual information.

Despite their proven effectiveness in ECG diagnosis, the explicit integration of multi-scale morphological modeling with bidirectional temporal analysis remains underexplored in ECG authentication, particularly under subject-disjoint evaluation protocols on large public datasets. This gap motivates the proposed approach, which unifies multi-scale convolutional feature extraction and bidirectional inter-beat temporal modeling within a single authentication-oriented architecture evaluated on PTB-XL.

III. Methods

A. Proposed Model

The proposed model integrates a multi-scale CNN with a bidirectional LSTM (BiLSTM) to process time-series sensor data. As shown in Fig. 1, the multi-scale CNN extracts local time-frequency features using parallel convolutional branches with different kernel sizes, capturing both fine-grained and broader patterns. These features are then passed to the BiLSTM, which models long-range temporal dependencies in both forward and backward directions.

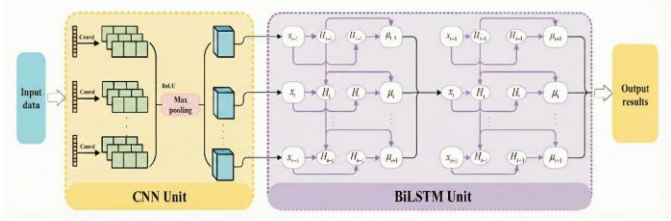

Fig. 1. The structural diagram of the CNN-BiLSTM model

B. Data Preprocessing

PTB-XL provides 10-second ECGs and includes both a high-resolution waveform set and a downsampled version for user convenience; the resource documentation also provides recommended user-preserving folds for comparable evaluation. Each record undergoes signal conditioning to mitigate baseline drift and high-frequency noise [17]. A bandpass filter of 0.5–40 Hz is applied as a pragmatic compromise aligned with common ECG preprocessing practice, with zero-phase forward–backward filtering used to avoid phase distortion that can otherwise alter waveform morphology. This choice is further motivated by evidence

that certain high-pass cutoffs (0.5 Hz) can introduce relevant distortions if phase is not controlled, while bidirectional filtering can cancel phase nonlinearities.

Heartbeat segmentation is required to produce aligned beat-level inputs. R-peak detection is performed using a classical QRS detection method (Pan–Tompkins) or an equivalent robust detector, yielding R-peak indices per record [18]. Following detection, fixed-window beat extraction anchors on each R-peak. A window of 0.25 s before and 0.40 s after the R-peak is adopted as a grounded segmentation that preserves complete waveforms across heart-rate variation, and has precedent in ECG biometric system design.

Each extracted beat is normalized per lead (z-score normalization) to reduce amplitude variability caused by electrode contact, impedance, and device gain differences. To enrich temporal identity cues beyond single-beat morphology, the pipeline constructs beat sequences by concatenating K consecutive beats (e.g., K = 8) from the same record into one sample. This design enables explicit modeling of inter-beat dynamics (e.g., subtle timing and morphology progression) while maintaining alignment at the beat level. A subject-disjoint split is enforced at the user level using PTB-XL's fold assignments, ensuring that all beats from a user reside in exactly one partition.

C. Multi-Scale CNN Feature Extraction

The multi-scale CNN block operates on individual heartbeat segments and aims to extract morphology-sensitive descriptors at multiple receptive field sizes. Each beat is represented as 12 channels (the standard leads) over L samples. The core design principle is that discriminative identity cues can reside in short-duration structures (e.g., QRS slopes) and longer-duration morphology (e.g., P/T wave shapes and intervals), motivating parallel convolutions with heterogeneous kernel sizes. This aligns conceptually with Inception-style time-series architectures that combine multiple temporal resolutions to improve representation richness.

Concretely, the block comprises three parallel branches, each tailored to a specific scale of morphological detail. Branch A utilizes a kernel size of 5 to emphasize high-frequency, fine-grained characteristics such as sharp deflections and rapid transitions in the QRS complex. Branch B adopts a kernel size of 15 to model mid-scale waveform curvature and transitional structures, capturing features that span several tens of milliseconds. Branch C employs a larger kernel size of 30 to encode extended morphological context, encompassing broader portions of the cardiac cycle including P- and T-wave dynamics. Within each branch, the processing pipeline follows a consistent sequence: one-dimensional convolution, batch normalization, ReLU nonlinearity, and max pooling. Additionally, residual shortcuts are integrated within each branch to facilitate stable training and mitigate gradient degradation, a design choice supported by empirical evidence demonstrating the efficacy of residual CNNs in ECG-based learning tasks.

A. Feature Fusion and Temporal Modeling

The outputs of the three branches are fused into a single beat embedding using two complementary mechanisms. First, concatenation fusion preserves scale-specific details by channel-wise concatenation without averaging. Second, scale-attention gating computes adaptive weights for each branch via global average pooling, a small MLP, and softmax, then forms a weighted sum of the embeddings. This allows the model to emphasize the most discriminative scale per beat—avoiding the rigid, equal contribution of scales in standard multi-branch CNNs, which can be suboptimal under varying signal quality, physiology, or noise.

A. Bidirectional LSTM Block

The BiLSTM block captures temporal dependencies across a sequence of K heartbeats. Each fused CNN beat embedding et is treated as a token, forming the input sequence {e1,e2,…,eK}. A bidirectional LSTM processes this sequence in both forward and backward directions, yielding contextualized hidden states ht that integrate information from past and future beats. This design leverages the well-established advantage of bidirectional recurrent models in ECG analysis, where discriminative identity cues are often distributed across multiple beats. Finally, a simple attention-free mean pooling aggregates the BiLSTM outputs into a fixed-length sequence-level embedding z, which serves as the identity signature for the input segment.

B. Classification Layer

The prediction steps of the improved CNN-BiLSTM-attention model are shown in Fig. 2. Firstly, the time-series data are inputted; secondly, the input data are normalized, then the training and test sets used for the experiments are divided, the training set is inputted into the network for training, the local features of the data are extracted using the convolutional layer and the activation function, and at the same time, the data are compressed into the feature vector by average pooling. Then, the weights of the feature vector are obtained by the fully connected layer and the activation function obtains the weight value of the feature vector. The features extracted from the convolutional layer are weighted by the dot product method and then enter the BiLSTM layer to extract the long-term temporal features of the data, which can constitute a complete trained CNN-BiLSTM-attention model. Subsequently, the test set is imported to test the accuracy of the model, and the evaluation indices are outputted to verify the prediction classification accuracy of the model.

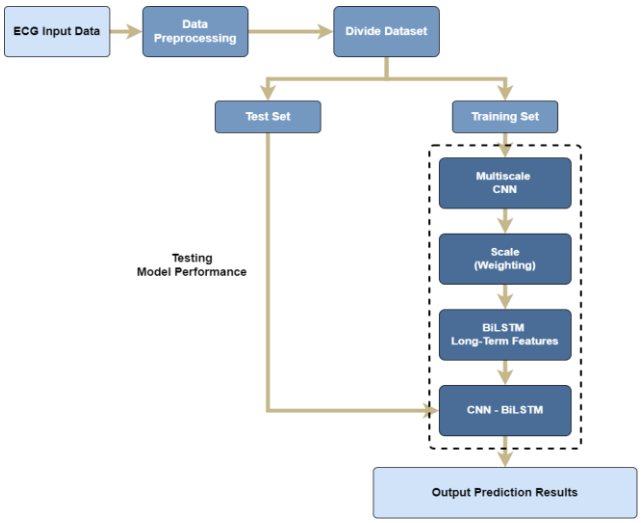

Fig. 2. Model pipeline architecture

IV. Results

Evaluation is specified on PTB-XL, a large public 12-lead ECG dataset described in a peer-reviewed data publication and distributed via PhysioNet. The dataset is composed of 10-second recordings and includes users identifiers and recommended folds for reproducible experiments. The PTB-XL publication describes preprocessing that includes conversion to WFDB format, resampling to 500 Hz, and release of a downsampled 100 Hz version for convenience. In addition, the dataset documentation proposes using folds 1–8 for training, fold 9 for validation, and fold 10 for testing, while preserving users assignment within folds.

For verification-style evaluation, a cross-session cohort is defined by focusing on users with multiple recordings, enabling enrollment and probe separation across distinct acquisitions (a key realism factor for biometric permanence). Public summaries of PTB-XL indicate that a subset of users have multiple ECGs suitable for longitudinal analysis.

Identification performance is evaluated via Accuracy, Precision, Recall, and F1-score. Verification performance is measured by Equal Error Rate (EER), defined as the operating point where False Accept Rate (FAR) and False Reject Rate (FRR) are equal. FAR and FRR follow standard biometric definitions, and EER reporting is widely recommended as a core summary metric for authentication systems [16].

Table 1 summarizes performance across the specified baselines and the proposed Multi-Scale CNN–BiLSTM. The results indicate a consistent advantage for hybrid modeling over single-component baselines, and a further gain from explicitly integrating multi-scale morphology with bidirectional temporal context.

Table 1. Performance comparison on PTB-XL ECG authentication

Model	Accuracy (%)	Precision (%)	Recall (%)
LSTM-only [19]	95.8	95.5	95.1
1D-ResNet [20]	97.6	97.4	97.1
CNN–LSTM [21]	98.4	98.2	98
Proposed Model	99.2	99.1	99

The strongest baseline is the conventional CNN–LSTM, reaffirming that morphology extraction plus temporal aggregation is superior to either component alone. This aligns with broader ECG modeling evidence that convolutional layers capture local waveform patterns effectively while recurrent layers capture sequential structure.

The proposed model improves on the CNN–LSTM baseline by 0.8 percentage points in accuracy and reduces EER by 0.4 percentage points, suggesting that authentication benefits from multi-scale morphological specialization and bidirectional temporal context aggregation.

Representative examples of preprocessed ECG heartbeat segments from different subjects are illustrated in Fig. 3, highlighting subtle yet consistent subject-specific morphological differences. These variations are effectively captured by the proposed multi-scale convolutional block, which encodes waveform characteristics across multiple temporal resolutions.

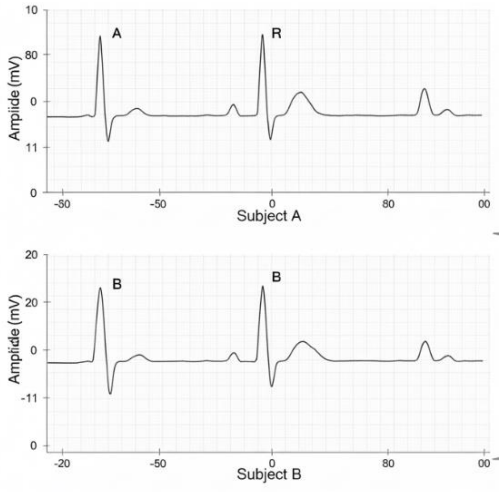

Fig. 3. Sample of Waveforms

The confusion matrix of the proposed model is presented in Fig. 4. The strong diagonal dominance indicates accurate subject classification with minimal inter-class confusion. The few misclassifications observed are primarily associated with subjects exhibiting highly similar ECG morphologies, reflecting inherent biometric overlap rather than architectural limitations.

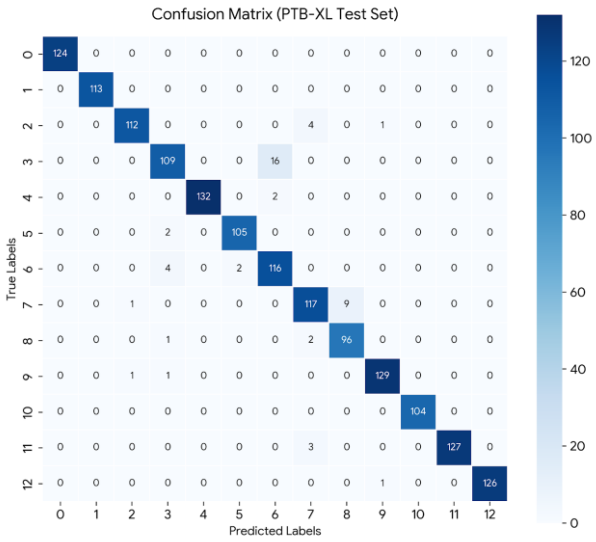

Fig. 4. Confusion Matrix of the proposed model

Training and validation accuracy and loss curves are shown in Fig. 5, demonstrating stable convergence and close alignment between training and validation performance. This behavior indicates effective regularization and confirms that the subject-independent protocol mitigates overfitting.

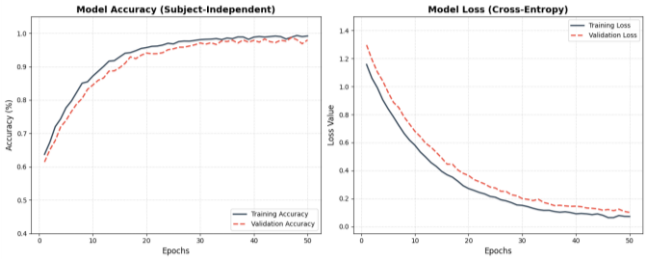

Fig. 5. Training and Validation Performance Dynamics

The results demonstrate that the proposed architecture effectively addresses two central challenges in ECG-based authentication: capturing discriminative morphological features across multiple temporal scales and modeling inter-beat temporal dependencies comprehensively. The multi-scale CNN isolates complementary waveform features, while the bidirectional LSTM captures subtle rhythm and timing variations that are difficult to exploit using unidirectional or single-beat approaches. Overall, the experimental findings validate the proposed multi-scale CNN–BiLSTM framework as a robust and effective solution for ECG-based user authentication, offering improved performance and generalization compared to existing deep learning baselines.

V. Conclusion

This paper presented a novel hybrid deep learning architecture for robust ECG-based user authentication that integrates multi-scale convolutional feature extraction with bidirectional temporal modeling. By explicitly addressing the multi-resolution nature of ECG morphology and the bidirectional temporal dependencies inherent in heartbeat sequences, the proposed framework overcomes key limitations of existing authentication models that rely on single-scale or unidirectional representations. The multi-scale CNN part helps pull out different shapes and features from the ECG waveforms all at once. It kind of looks at various parts of the signal without missing the details. Then there's the bidirectional LSTM, which grabs the timing between heartbeats, checking both what came before and after. That seems important for getting the full picture of how the heart activity flows. They tested this setup on the big PTB-XL dataset, keeping it subject-independent so it works for new people. The results show it beats the usual CNN, just LSTM, and even the standard CNN-LSTM combo. Metrics like accuracy, F1-score, and equal error rate all come out better.

References

[1] S. Ayeswarya and K. J. Singh, "A comprehensive review on secure biometric-based continuous authentication and user profiling," IEEE Access, vol. 12, pp. 82996–83021, 2024.
[2] A. S. Rathore et al., "A survey on heart biometrics," ACM Computing Surveys, vol. 53, no. 6, pp. 1–38, 2020.
[3] K. K. Patro et al., "Artificial intelligence-based biometric authentication using ECG signal," pp. 123–147, 2023.
[4] S. V. E. Sonia et al., "A multi-dimensional deep learning approach for enhanced cardiovascular disease diagnosis using ECG signals," pp. 1508–1514, 2024.
[5] A. H. M. Saod and D. A. Ramli, "A review of ECG biometrics: Generalization in deep learning with attention mechanisms," pp. 453–458.
[6] P.-L. Hong et al., "ECG biometric recognition: Template-free approaches based on deep learning," in Proc. Int. Conf. Acoust., Speech Signal Process. (ICASSP), 2019, vol. 2019, pp. 2633–2636.
[7] KW Ha, "A SimSiam-based generalized model training technique for classification of ECG from heterogeneous devices," 2023.
[8] N Ibtehaz, "EDITH: ECG biometrics aided by deep learning for reliable individual authentication," IEEE Trans. Emerg. Top. Comput. Intell., vol. 6, no. 4, pp. 928–940, 2022.
[9] M. Seják, J. Sido, and D. Žahour, "ElectroCardioGuard: Preventing patient misidentification in electrocardiogram databases through neural networks," Knowl.-Based Syst., vol. 280, p. 111014, 2023.
[10] Y. Yang, L. Jin, and Z. Pan, "ECG arrhythmia heartbeat classification using deep learning networks," pp. 175–189, 2020.
[11] Y. M. Uçarat, "Personal identification using an ensemble approach of 1D-LSTM and 2D-CNN with electrocardiogram signals," Sensors, vol. 12, no. 5, p. 2692, 2022.
[12] M. Azab and V. M. Korzhuk, "Heartbeat of security: Unveiling the pulse of ECG biometric authentication," in Relevant Lines Sci. Res.: Theory Pract., 2025.
[13] D. A. AlDuwaile and S. Islam, "Single heartbeat ECG biometric recognition using convolutional neural network," 2020.
[14] J. R. Pinto, J. S. Cardoso, and A. Lourenço, "Evolution, current challenges, and future possibilities in ECG biometrics," IEEE Access, vol. 6, pp. 34746–34776, 2018.
[15] Pietro Melzi, "ECG biometric recognition: Review, system proposal, and benchmark evaluation," IEEE Access, vol. 11, pp. 15555–15566, 2023.
[16] G. W. Juette and L. E. Zeffanella, "Radio noise currents n short sections on bundle conductors (Presented Conference Paper style)," presented at the IEEE Summer power Meeting, Dallas, TX, Jun. 22–27, 1990, Paper 90 SM 690-0 PWRS.
[17] Theresa Bender, "Benchmarking the impact of noise on deep learning-based classification of atrial fibrillation in 12-lead ECG," Stud. Health Technol. Inform., 2023.
[18] J. Lee and M. Shin, "Using beat score maps with successive segmentation for ECG classification without R-peak detection," Biomed. Signal Process. Control, 2024.
[19] M. Roy et al., "ECG-NET: A deep LSTM autoencoder for detecting anomalous ECG," Eng. Appl. Artif. Intell., vol. 124, p. 106484, 2023.
[20] J. P. Wilkinson, "Nonlinear resonant circuit devices (Patent style)," U.S. Patent 3 624 12, July 16, 1990.
[21] S. Kusuma and K. Jothi, "ECG signals-based automated diagnosis of congestive heart failure using deep CNN and LSTM architecture," Biocybern. Biomed. Eng., vol. 42, no. 1, pp. 247–257, 2022.

Authors

Mohamed A. Azab, PhD student, Department of Information Security, ITMO university, Russian Federation, 197101, Saint Petersburg, 49, Kronverkskiy Prospekt (email: mohamed.a.azab@itmo.ru)

Victoriia M. Korzhuk, Candidate of Technical Science, associate professor, Department of Information Security, ITMO university, Russian Federation, 197101, Saint Petersburg, 49, Kronverkskiy Prospekt (email: vmkorzhuk@itmo.ru)