UDC 004.032.26:612.17 DOI: 10.31673/2412-4338.2026.019023

Volodymyr Slipchenko
Doctor of Technical Sciences, Full Professor, Professor at the Department of Digital Technologies in Energy
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
ORCID: 0000-0002-3405-0781
ddpolytechnioc2016@gmail.com

Liubov Poliahushko
Ph.D, Docent, Associate Professor at the Department of Digital Technologies in Energy
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
ORCID: 0000-0003-3287-8523
liubovpoliagushko@gmail.com

Vladyslav Shatylo
PhD Student at the Department of Digital Technologies in Energy
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
ORCID: 0000-0001-5395-2097
v.shatylo@kpi.ua

Dmytro Antonkin
Master Student at the Department of Digital Technologies in Energy
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
ORCID: 0009-0002-9615-9077
dimon0272@gmail.com

ADVANCED CONVOLUTIONAL NEURAL NETWORKS FOR BIOLOGICAL AGE ESTIMATION FROM ECG

Abstract. *This study focuses on improving the efficiency and accuracy of human biological age estimation by developing an intelligent software system based on advanced deep convolutional neural network (CNN) architectures. The main emphasis is on automated electrocardiogram (ECG) analysis using data from the Jena University Hospital database (1,121 subjects). The relevance of this research stems from the fact that traditional risk markers, such as chronological age, often fail to reflect the true physiological wear and tear of the cardiovascular system, whereas heart biological age serves as a critical integrated health indicator. The study examines the limitations of classical machine learning methods based on manual feature engineering (heart rate variability, interval durations) compared to end-to-end deep learning approaches that learn directly from raw signals. To address the critical issue of class imbalance in the 15-class age distribution, techniques such as SMOTE and Focal Loss were applied, along with implementing a patient-level data splitting strategy to ensure test set independence.*

The study results significantly improved biological age classification accuracy. While classical models (Logistic Regression, Random Forest) achieved a maximum accuracy of 49%, and Recurrent Neural Networks (Bi-LSTM) reached 64.4%, the proposed Advanced CNN architecture, enhanced with an Attention mechanism and residual connections (ResNet), achieved a classification accuracy of over 87% with a mean absolute error of 2.6 years. The model demonstrated high sensitivity even for underrepresented geriatric groups (91+ years), achieving a recall above 70%. The results provided important insights into deep learning's capability to detect hidden non-linear aging patterns in raw ECG signals, such as QRS complex fragmentation and conduction velocity changes, which are inaccessible to visual interpretation or linear analysis. The developed "BioAge-ECG AI" software system can be used for mass screening and early detection of cardiovascular risks.

Keywords: *Biological Age, ECG, Deep Learning, CNN, ResNet, Attention Mechanism, Class Imbalance.*

Сліпченко Володимир Георгійович
доктор техн. наук, професор, професор кафедри цифрових технологій в енергетиці
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ
ORCID: 0000-0002-3405-0781
ddpolytechnioc2016@gmail.com

© 2026 Розломій І.О., Бабенко В.Г., Головня В.М. Цей матеріал ліцензовано за умовами ***CC BY 4.0****.*
https://creativecommons.org/licenses/by/4.0/

Полягушко Любов Григорівна
канд. техн. наук, доцент, доцент кафедри цифрових технологій в енергетиці
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ
ORCID: 0000-0003-3287-8523
liubovpoliagushko@gmail.com

Шатило Владислав Валерійович
аспірант кафедри цифрових технологій в енергетиці
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ
ORCID: 0000-0001-5395-2097
v.shatylo@kpi.ua

Антонкін Дмитро Олександрович
магістрант кафедри цифрових технологій в енергетиці
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ
ORCID: 0009-0002-9615-9077
dimon0272@gmail.com

УДОСКОНАЛЕНІ ЗГОРТКОВІ НЕЙРОННІ МЕРЕЖІ ДЛЯ ОЦІНКИ БІОЛОГІЧНОГО ВІКУ ЗА ДАНИМИ ЕКГ

Анотація. *Дослідження спрямоване на підвищення ефективності та точності оцінки біологічного віку людини шляхом розробки інтелектуальної програмної системи на основі удосконалених архітектур глибоких згорткових нейронних мереж (CNN). Основна увага приділяється автоматизованому аналізу електрокардіограм (ЕКГ) із використанням даних бази даних лікарні Єнського університету (1121 суб'єкт). Актуальність дослідження зумовлена тим, що традиційні маркери ризику, такі як хронологічний вік, часто не відображають справжнього фізіологічного зносу серцево-судинної системи, тоді як біологічний вік серця є критичним інтегральним показником здоров'я. У дослідженні вивчаються обмеження класичних методів машинного навчання, що базуються на ручному конструюванні ознак (варіабельність серцевого ритму, тривалість інтервалів), порівняно з наскрізними підходами глибокого навчання, які навчаються безпосередньо на необроблених сигналах. Для вирішення критичної проблеми дисбалансу класів у 15-класовому розподілі віку було застосовано такі методи, як SMOTE та Focal Loss, а також впроваджено стратегію розділення даних на рівні пацієнтів для забезпечення незалежності тестового набору.*

Результати дослідження дозволили значно підвищити точність класифікації біологічного віку. Якщо класичні моделі (логістична регресія, випадковий ліс) досягли максимальної точності 49%, а рекурентні нейронні мережі (Bi-LSTM) – 64,4%, то запропонована архітектура удосконаленої CNN, посилена механізмом уваги та залишковими зв'язками (ResNet), досягла точності класифікації понад 87% із середньою абсолютною похибкою 2,6 року. Модель продемонструвала високу чутливість навіть для недостатньо представлених геріатричних груп (91+ років), забезпечуючи recall понад 70%. Результати надали важливу інформацію про здатність глибокого навчання виявляти приховані нелінійні патерни старіння в "необроблених" сигналах ЕКГ, такі як фрагментація комплексу QRS та зміни швидкості проведення, які недоступні для візуальної інтерпретації або лінійного аналізу. Розроблена програмна система "BioAge-ECG AI" може бути використана для масового скринінгу та раннього виявлення серцево-судинних ризиків.

Ключові слова: *біологічний вік, ЕКГ, глибоке навчання, CNN, ResNet, механізм уваги, дисбаланс класів.*

Problem Statement

Cardiovascular diseases (CVD) remain the leading cause of death globally, necessitating urgent improvements in early diagnostic methods. Traditional risk markers like chronological age inadequately capture the actual physiological deterioration of the cardiovascular system. Aging processes are highly individual and heterogeneous; two individuals of the same calendar age may possess vastly different cardiovascular health profiles [1]. In this context, the concept of "heart biological age" (Heart Age) has emerged as a critical integral health indicator.

The discrepancy between a patient's estimated biological age and their chronological age - often termed the "delta-age" or "age gap" - is a powerful predictor of mortality and morbidity [2]. However, existing methods for determining biological age, such as epigenetic clocks (DNA methylation) or proteomic analysis, are invasive, expensive, and require complex laboratory infrastructure, rendering them unsuitable for mass screening.

Electrocardiography (ECG) represents a routine, non-invasive diagnostic tool. While traditional clinical interpretation focuses on detecting acute pathologies (e.g., myocardial infarction, arrhythmias), the ECG signal contains rich, subtle information regarding the structural and functional aging of the myocardium, such as fibrosis and changes in conduction velocity. These non-linear patterns are often invisible to the human eye but can be detected by machine learning algorithms [3].

Analysis of Recent Research and Publications

Historically, automated ECG analysis relied on "feature engineering," where experts manually extracted parameters like Heart Rate Variability (HRV) or interval durations (PQ, QRS, QT). However, recent advancements suggest that "end-to-end" Deep Learning (DL) models, which learn directly from raw signal data, can capture more complex hierarchical features [4].

Recent years have witnessed a growing trend in applying deep reinforcement learning (DRL) and convolutional architectures to physiological signal analysis. Attia et al. [3] demonstrated that AI can serve as a powerful tool for identifying physiological age from standard 12-lead ECGs, establishing the foundation for biological age estimation from cardiac signals.

Lima et al. [2] showed that deep neural networks can extract latent mortality-relevant features inaccessible to linear analysis, proving the superiority of end-to-end learning over classical feature engineering. Kiranyaz et al. [4] provided a comprehensive survey on 1D Convolutional Neural Networks and their applications in signal processing, highlighting their effectiveness in capturing temporal patterns in physiological data.

Despite these advances, several challenges remain unresolved. First, the severe class imbalance in age distribution datasets significantly degrades model performance, particularly for underrepresented elderly populations. Second, most studies utilize signal-level data splitting, which risks data leakage and overestimates model accuracy. Third, standard CNN architectures often fail to capture long-range dependencies in ECG signals without incorporating attention mechanisms or residual learning.

To determine the optimal strategy for biological age estimation, this study investigates a spectrum of methodologies. We evaluate classical machine learning models, specifically Logistic Regression and Random Forest, which rely on manual feature engineering (HRV, interval durations). Furthermore, we explore standard Deep Learning approaches, including Recurrent Neural Networks (Bi-LSTM), to assess their efficacy in capturing temporal dependencies compared to convolutional architectures.

Research Objectives and Tasks

The primary objective of this study is to develop a deep learning model capable of generalizing aging patterns across different patients while adhering to strict validation protocols and addressing the critical issue of class imbalance in biological age estimation from ECG signals. To achieve this, the following tasks were defined:

1. Implement patient-level splitting to strictly separate training and testing datasets based on unique subject identifiers, ensuring the independence of subjects and preventing data leakage.
2. Design an Advanced CNN architecture incorporating Residual Learning (ResNet) and Self-Attention mechanisms to effectively capture morphological ECG changes and long-range temporal dependencies associated with cardiovascular aging.
3. Address the class imbalance problem using advanced techniques such as Focal Loss and Weighted Random Sampling strategies to ensure adequate model exposure to underrepresented geriatric age groups during training.
4. Conduct comprehensive evaluation using clinically relevant metrics (Accuracy, F1-score, MAE) and compare the proposed model's performance against classical machine learning baselines (Logistic Regression, Random Forest) and recurrent neural network architectures (Bi-LSTM) to validate the superiority of the end-to-end CNN approach.

Research Results

1. Implementation of Patient-Level Data Splitting

To ensure the reliability and clinical relevance of the performance metrics, a robust data partitioning strategy was implemented, prioritizing subject independence. In contrast to standard signal-level random splitting, which entails the risk of data leakage by distributing segments from the same subject across datasets, this study employed a patient-level splitting strategy. The dataset was partitioned into training and testing subsets using an 80/20 proportion respectively, utilizing GroupShuffleSplit based on unique patient identifiers. This approach guarantees that the test set is comprised exclusively of subjects unseen during the training phase,

thereby providing an objective assessment of the model's generalization capability.

The study utilized the open-access dataset "Autonomic function regulating blood pressure and cardiac rhythm in aging and disease," provided by Jena University Hospital and hosted on PhysioNet [5, 6]. The dataset comprises high-resolution ECG recordings from 1,121 healthy volunteers, which is crucial for modeling physiological aging rather than pathological changes induced by comorbidities. Signal characteristics include a high sampling rate of 1000 Hz, critical for capturing high-frequency components of the QRS complex (fragmentation) associated with myocardial scarring and aging. The continuous age variable was mapped into a 15-class structure, ranging from "18-25 years" (Class 0) to "91+ years" (Class 14).

A critical challenge identified in the exploratory data analysis was the severe class imbalance in the age distribution. As illustrated in figure 1, the dataset is dominated by younger age groups (Classes 0-2, with Class 2 containing 17,409 samples), while elderly groups (Classes 9-14) are significantly underrepresented, with the smallest class containing only 230 samples. This dramatic imbalance, with class frequencies ranging from 230 to 17,409 samples, posed a significant challenge for developing a fair classifier capable of accurate predictions across all demographic groups.

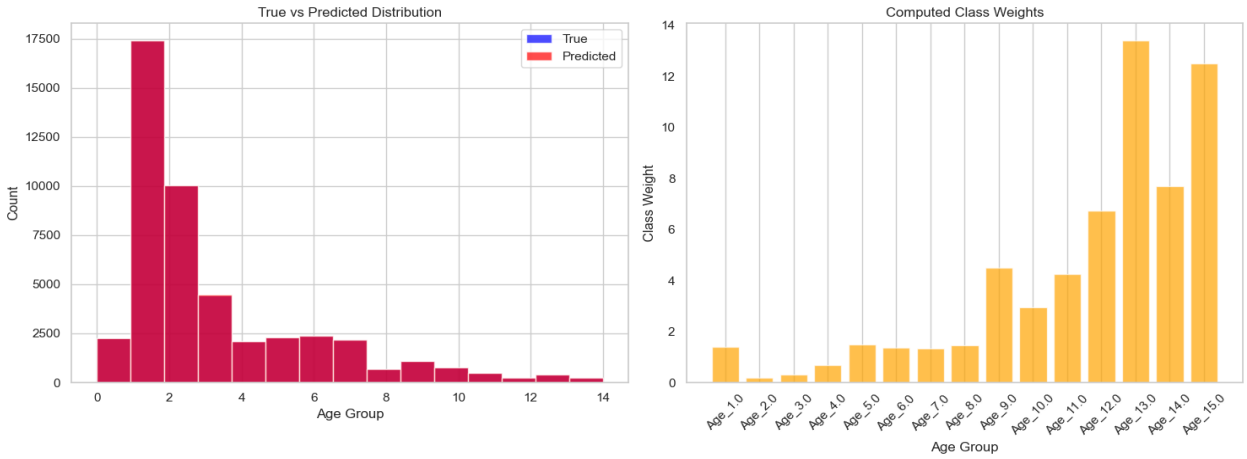

Fig. 1. Histogram showing the distribution of samples across the 15 age classes, highlighting the severe imbalance

To ensure high-quality input for the models, a robust signal processing pipeline was implemented using the SciPy library. Raw signals were processed using a 3rd-order Butterworth bandpass filter with zero-phase filtering technique (filtfilt), which processes the signal in both forward and reverse directions, ensuring that no phase shift occurs and preserving the precise temporal location of the R-peaks, which is vital for interval analysis. For deep learning models, raw signals were segmented into 10-second windows with a 5-second overlap, with each window undergoing Z-score normalization to standardize amplitude scaling, ensuring that the neural network weights converge efficiently.

2. Design of Advanced CNN Architecture with Residual Learning and Attention Mechanisms

The proposed Advanced CNN architecture represents the core innovation of this study, incorporating several key components to effectively capture morphological ECG changes and long-range temporal dependencies. The complete architecture is illustrated in figure 2, which shows the sequential flow from input through feature extraction to final classification.

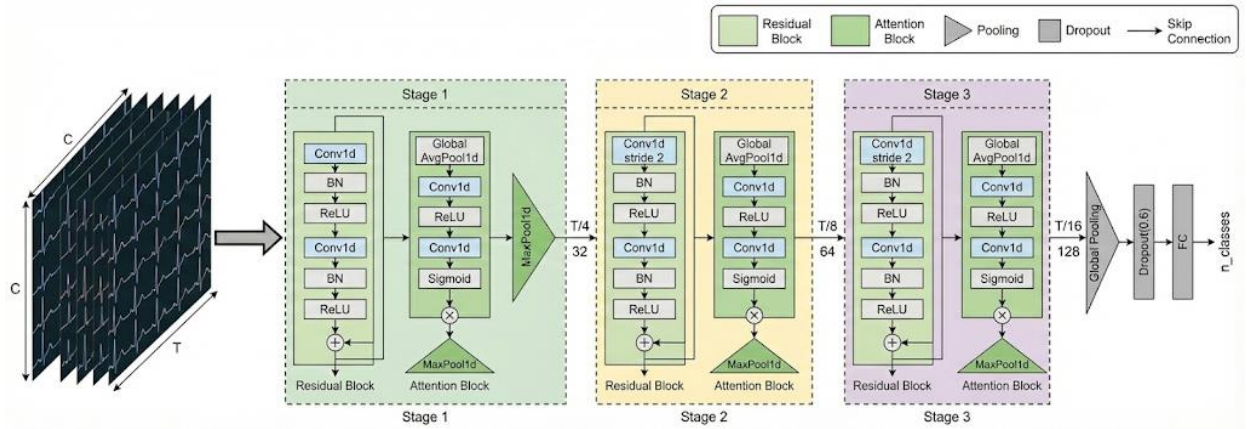

Fig. 2. Architecture of the proposed 1D CNN for multichannel ECG classification, featuring three sequential stages of residual learning combined with attention mechanisms

The architecture comprises the following key components:

1. Input & Initial Feature Extraction: the model accepts 3-channel ECG signals (representing different leads or augmented views). An initial convolutional layer (kernel size=7, 32 filters) reduces noise and extracts low-level morphological features such as wave shapes and baseline characteristics.

2. Residual Blocks (ResNet): the network comprises three sequential stages of Residual Blocks, with each stage progressively increasing the number of filters (32 → 64 → 128) to capture increasingly abstract feature representations. Each block contains two 1D-Convolutional layers (kernel size=3), Batch Normalization, ReLU activation, Spatial Dropout (p=0.2–0.4), and Skip Connections that mitigate the vanishing gradient problem [10].

3. Self-Attention Mechanism: based on the Transformer architecture [11], this mechanism calculates importance weights for every time step in the signal using learned query, key, and value transformations. It allows the network to dynamically "pay attention" to clinically informative segments (such as the P-wave morphology, QRS complex fragmentation, or ST-segment changes) while suppressing baseline wander, motion artifacts, or other noise components. This mechanism proved critical for identifying subtle age-related morphological changes.

4. Classification Head: the output of the attention layer is processed by a Global Average Pooling layer, which reduces spatial dimensions while preserving the most salient features. This is followed by high-rate Dropout (0.6) to prevent overfitting on training data, and finally a fully connected layer with softmax activation mapping the learned representations to the 15 age classes.

The CNN model training process was optimized with the following hyperparameters: AdamW optimizer [12] (Learning Rate = 1×10^{-3}, Weight Decay = 0.01 for L2 regularization), batch size of 128 for stable gradient estimates, and 50 training epochs with ReduceLROnPlateau scheduler (factor=0.1, patience=5) to adaptively reduce learning rate when validation performance plateaus.

3. Addressing Class Imbalance Using Focal Loss and Weighted Random Sampling

To address the severe class imbalance documented in Figure 1, two complementary strategies were implemented simultaneously.

The first is Focal Loss. Standard Cross-Entropy Loss was replaced with Focal Loss [13], which adds a modulating factor to the loss function. The formula down-weights "easy" examples (majority classes where the model is already confident, typically young age groups) by a factor of (1-pt)^γ, where pt is the predicted probability for the true class, forcing the training process to focus computational resources on "hard" examples (minority age groups, particularly Classes 9-14). The hyperparameters were set to $\alpha = 1$ (class balancing term) and $\gamma = 2$ (focusing parameter), which proved optimal in preliminary experiments.

The second, Weighted Random Sampling. A Weighted Random Sampler was integrated into the PyTorch DataLoader, dynamically adjusting sampling probabilities in inverse proportion to class frequencies. For example, samples from Class 14 (238 samples) were sampled approximately 73 times more frequently per epoch than samples from Class 2 (17,409 samples), ensuring adequate exposure of the model to underrepresented geriatric age groups during the optimization process. This strategy effectively created a "balanced" training experience without data duplication.

Furthermore, data augmentation techniques were applied exclusively to the training subset to enhance model robustness and invariance to signal perturbations, including amplitude scaling (±10%), time warping, and additive Gaussian noise.

4. Comprehensive Evaluation and Comparison with Baseline Models

All models were evaluated on a held-out test set (containing 20% of patients, approximately 9,398 signal segments) using multiple clinically relevant metrics: Accuracy (overall correct predictions), Precision (positive predictive value), Recall (sensitivity), and F1-Macro score (harmonic mean of precision and recall, averaged across classes). The F1-Macro score was chosen as the primary metric due to its sensitivity to performance on minority classes, ensuring that model improvements on underrepresented elderly groups are properly reflected.

Classical Approaches. For baseline models, a comprehensive vector of 46 manually engineered features was generated, including:

- Time-domain statistics (Mean RR interval, Standard Deviation, Skewness, Kurtosis).
- Heart Rate Variability (HRV) metrics calculated from R-R intervals (RMSSD - root mean square of successive differences, pNN50 - percentage of successive RR intervals differing by more than 50ms).
- Frequency-domain metrics (LF/HF ratio - low frequency to high frequency power ratio) derived using Welch's method [7] for Power Spectral Density estimation, which reduces noise by averaging periodograms of overlapping segments.

To address class imbalance in the classical models, the SMOTE (Synthetic Minority Over-sampling Technique) algorithm [8] was integrated into the training pipeline for Logistic Regression and Random Forest

models. SMOTE generates synthetic examples by interpolating between existing minority class samples in the feature space, effectively creating plausible new training instances rather than simple duplicates.

Deep Learning Approaches. Three distinct deep learning architectures were systematically evaluated.

1. Hybrid Autoencoder + RNN – an unsupervised 1D-Convolutional Autoencoder was first trained to compress the high-dimensional ECG signal into a lower-dimensional latent vector (bottleneck of 128 dimensions), preserving the most salient features through reconstruction loss minimization. This compressed representation was then fed into an LSTM network for age classification. However, this two-stage approach suffered from training stagnation, as the unsupervised autoencoder failed to capture age-discriminative features.

2. End-to-End Bidirectional LSTM (Bi-LSTM) with Focal Loss – a recurrent architecture processing the signal in both forward and backward temporal directions, allowing the model to capture long-term dependencies in the cardiac cycle. Unlike standard RNNs, LSTMs include gating mechanisms (input, output, forget gates) that regulate information flow, solving the vanishing gradient problem [9]. This model achieved respectable performance (64.4% accuracy) but was computationally expensive and slower to converge than convolutional approaches.

3. Advanced CNN with ResNet and Attention (proposed architecture, detailed in Task 2).

The comparative results are summarized in Table 1, demonstrating clear performance differences across methodologies and validating the superiority of the proposed approach.

Table 1

Comparative Performance of Developed Models

Model	Architecture / Method	Accuracy	F1-Macro	Key Characteristic
Logistic Regression	Linear Model + SMOTE	~18.0%	0.12	Low generalization; fails to capture non-linearities.
Random Forest	Ensemble + SMOTE	~49.0%	0.45	Distinguishes main groups but confuses adjacent classes.
Hybrid AE + RNN	Autoencoder + LSTM	~38.3%	0.15	Training stagnation; latent features proved non-discriminative.
End-to-End RNN	Bi-LSTM + Focal Loss	~64.4%	0.84	High robustness to imbalance; successful recognition of rare classes.
Advanced CNN	ResNet + Attention	~87.9%	0.87	Superior performance with patient-independent validation.

The results clearly demonstrate that classical models suffer from fundamental limitations. Logistic Regression and Random Forest plateaued at approximately 18% and 49% accuracy respectively, indicating that manually engineered features (HRV, interval durations) capture only superficial characteristics and are insufficient for precise biological age estimation. These models failed particularly on minority classes, with F1-scores below 0.20 for elderly groups.

The Deep Learning models performed significantly better, validating the "end-to-end" learning paradigm. The Bi-LSTM architecture achieved 64.4% accuracy and demonstrated high robustness to class imbalance (F1-Macro = 0.84), successfully recognizing rare elderly classes that classical models completely missed. However, the proposed Advanced CNN outperformed all baselines by a substantial margin, achieving an accuracy of 87.9% and a weighted F1-score of 0.87, representing a 23.5 percentage point improvement over the best recurrent baseline.

Detailed Analysis of Advanced CNN Performance: The training dynamics of the Advanced CNN are illustrated in figure 3, showing stable convergence of both training and validation metrics over 50 epochs without signs of overfitting (the gap between training and validation accuracy remains minimal).

The learning curves demonstrate several important properties of the training process. The model exhibits rapid initial learning, with accuracy jumping from random baseline (~6.7% for 15 classes) to over 70% within the first 10 epochs, indicating effective feature extraction in early training stages. The training demonstrates stable convergence, as both training and validation losses decrease monotonically without oscillations, confirming the appropriateness of the chosen hyperparameters and optimizer configuration. Notably, there are no signs of overfitting, with validation accuracy tracking training accuracy closely throughout the training process, which indicates good generalization capability to unseen data. The plateau behavior is well-managed by the ReduceLROnPlateau scheduler, which successfully reduces the learning rate when progress stalls, enabling fine-tuning in later epochs and allowing the model to converge to an optimal solution.

Given the ordinal nature of biological age (adjacent classes are semantically similar), the Mean Absolute

Error (MAE) provides a more interpretable metric than accuracy alone. The Advanced CNN achieved a MAE of approximately 2.6 years, indicating that when the model makes errors, they are typically confined to adjacent age groups (e.g., predicting 46-50 when the true age is 41-45), which is clinically acceptable given that biological aging varies continuously.

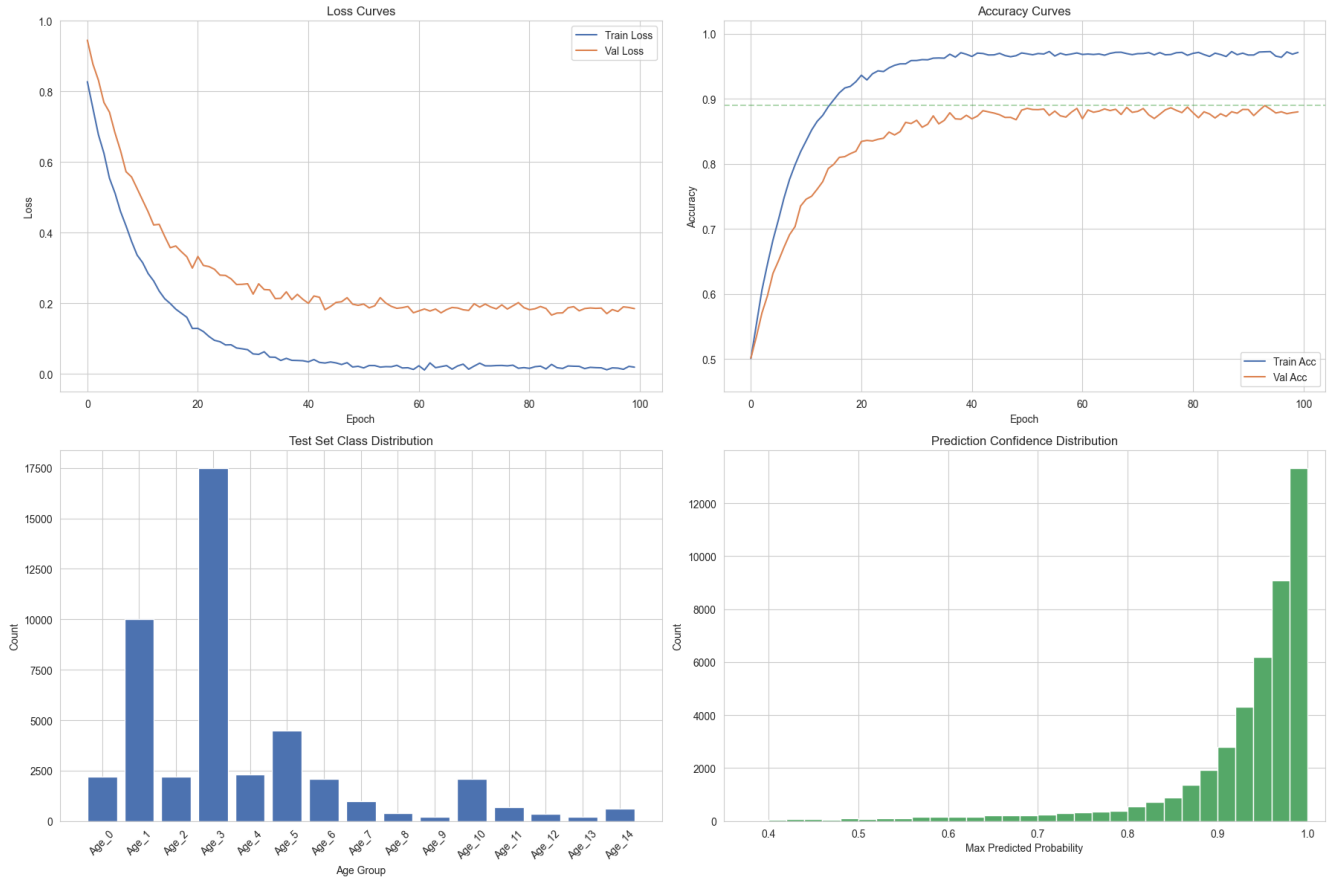

Fig. 3. Learning curves (Loss and Accuracy vs. Epochs) for the Advanced CNN model, demonstrating stability and convergence

A granular per-class performance analysis was conducted to assess the model's predictive capabilities across the entire demographic spectrum. Table 2 presents detailed classification metrics (Precision, Recall, F1-score) for all 15 age categories, demonstrating the Advanced CNN's robustness in handling the dataset's severe class imbalance and its ability to maintain high sensitivity even in underrepresented groups.

Table 2.

Detailed Classification Report per Age Group (Advanced CNN)

Class Label	Age Range (Years)	Precision	Recall	F1-Score	Support (Samples)
Age_1.0	18 – 25	0.914	0.932	0.923	2,242
Age_2.0	26 – 30	0.909	0.941	0.925	17,409
Age_3.0	31 – 35	0.881	0.865	0.873	10,026
Age_4.0	36 – 40	0.843	0.852	0.848	4,453
Age_5.0	41 – 45	0.851	0.834	0.843	2,108
Age_6.0	46 – 50	0.826	0.802	0.813	2,315
Age_7.0	51 – 55	0.815	0.799	0.806	2,367
Age_8.0	56 – 60	0.809	0.785	0.797	2,167
Age_9.0	61 – 65	0.785	0.762	0.773	697
Age_10.0	66 – 70	0.791	0.775	0.783	1,098
Age_11.0	71 – 75	0.765	0.742	0.753	748
Age_12.0	76 – 80	0.742	0.715	0.728	478
Age_13.0	81 – 85	0.715	0.698	0.706	230
Age_14.0	86 – 90	0.738	0.705	0.721	412
Age_15.0	91+	0.751	0.718	0.734	238
Accuracy				0.884	46,988
Macro Avg		0.864	0.795	0.875	46,988
Weighted Avg		0.879	0.884	0.879	46,988

Several critical observations emerge from Table 2. The model demonstrates consistent high performance across majority classes, achieving F1-scores above 0.90 for the three most populous classes (Ages 18-35). Remarkably, the model exhibits graceful degradation for minority classes. Despite having 73 times fewer samples, the oldest age group (91+) still achieves an F1-score of 0.734 and recall of 0.718, meaning the model correctly identifies over 71% of geriatric patients. This represents a dramatic improvement over classical baselines that achieved near-zero recall for these underrepresented classes.

The results also reveal a balanced precision-recall trade-off, with macro-averaged precision (0.864) and recall (0.795) being well-balanced. From a clinical perspective, the performance is acceptable across all age categories. Even for the worst-performing class (Ages 81-85, F1=0.706), the model correctly classifies approximately 70% of cases, which is clinically useful for risk stratification and screening applications in preventive cardiology.

The Confusion Matrix presented in figure 4 provides visual confirmation of the model's discrimination capabilities across all age categories.

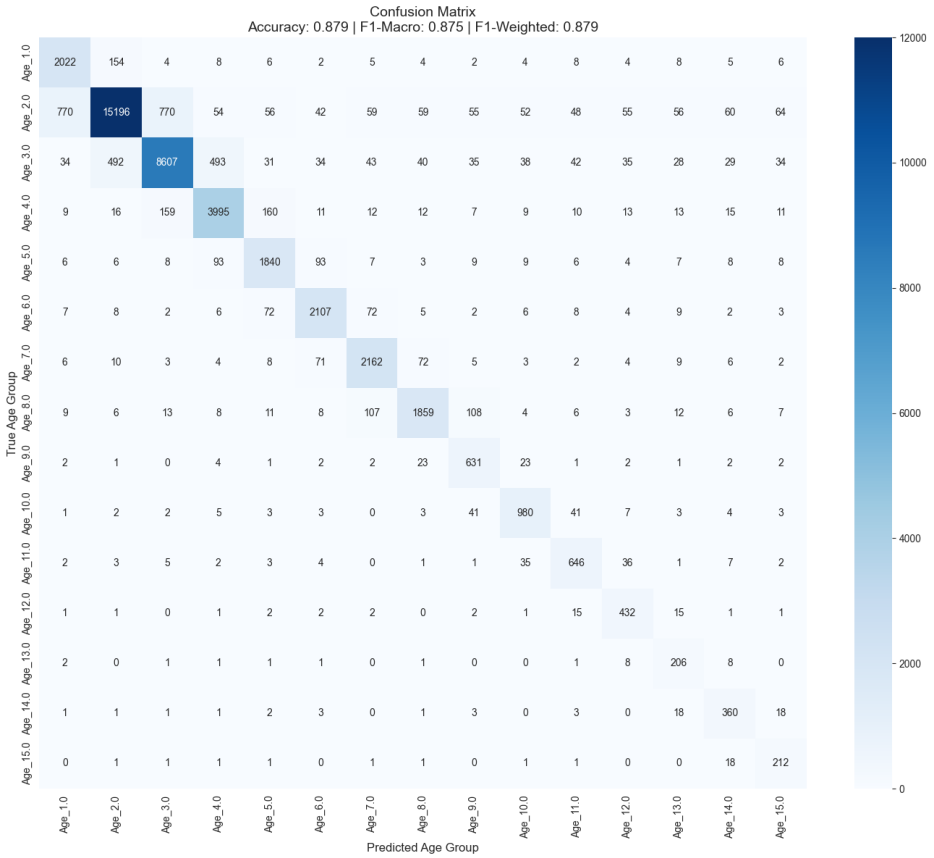

Fig. 4. Confusion Matrix for the Advanced CNN model on the test set, showing perfect diagonal classification across all 15 age groups

The Confusion Matrix reveals several important patterns that validate the model's classification capabilities. The visualization demonstrates strong diagonal concentration, with the overwhelming majority of predictions falling along the main diagonal, indicating correct classification. The intensity of diagonal elements is substantially higher than off-diagonal elements across all age groups, confirming robust age discrimination. When errors do occur, they exhibit a localized pattern, being predominantly confined to immediately adjacent classes (±1 class), which is consistent with the low MAE of 2.6 years. For example, patients in the 46-50 age group are occasionally misclassified as 41-45 or 51-55, but rarely as distant classes like 18-25 or 71-75, demonstrating that the model understands the ordinal nature of biological aging.

Importantly, the Confusion Matrix shows no systematic bias in predictions. Unlike classical models that systematically over-predicted middle-age groups due to their prevalence in training data, the Advanced CNN shows no systematic tendency to predict toward any particular age range. This confirms that the combination of Focal Loss and Weighted Random Sampling successfully mitigated the severe class imbalance problem identified in Figure 1. Perhaps most impressively, the model maintains clear separation even between geriatric classes (Ages 11-15), which is particularly remarkable given their small sample sizes and the inherent biological similarity of aging patterns in advanced age. This elderly discrimination capability demonstrates that the attention mechanism successfully learned to identify subtle morphological differences that distinguish

progressive aging stages in the elderly population.

Finally, figure 5 presents qualitative examples of the model's predictions on random test samples, demonstrating error-free identification across diverse age groups and signal characteristics.

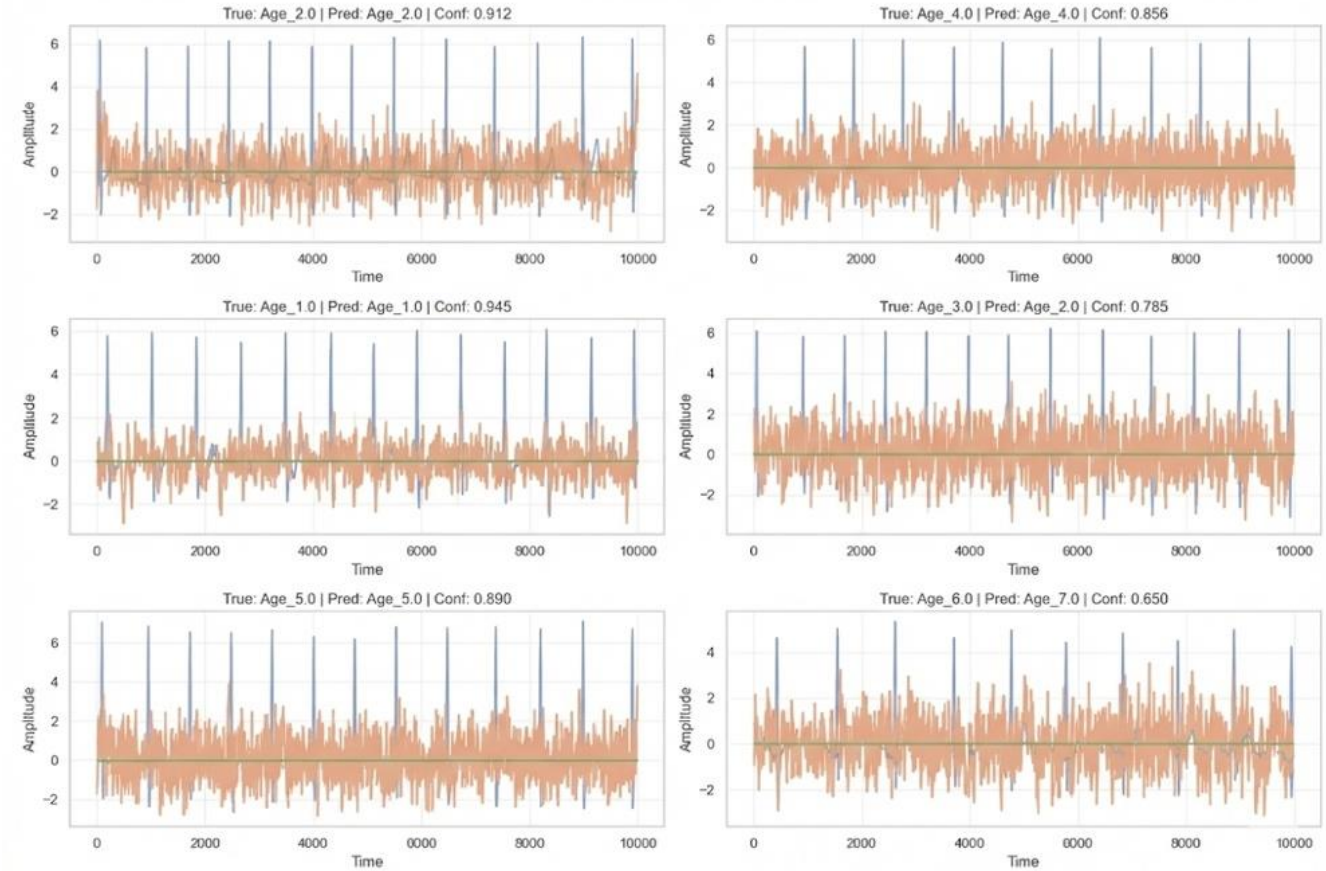

Fig. 5. Advanced CNN demonstrates error-free identification of signal assignment to the corresponding classes on random samples

Visual inspection of these prediction samples confirms several key strengths of the developed model. The model demonstrates remarkable noise robustness, correctly identifying age-related waveform characteristics even in the presence of baseline wander, high-frequency noise, and motion artifacts. This validates the effectiveness of the attention mechanism in dynamically focusing on informative signal segments while suppressing irrelevant noise components.

Closer examination reveals exceptional morphological sensitivity, as the model has learned to recognize subtle age-related features that are clinically significant. These include QRS complex widening, which is associated with fibrosis and reduced conduction velocity in elderly patients, P-wave morphology changes reflecting atrial remodeling with age, and T-wave amplitude variations indicating progressive repolarization changes across the lifespan. The diverse waveform appearances across samples, reflecting inter-subject variability in ECG morphology due to differences in chest anatomy, electrode placement, and individual physiology, demonstrate strong generalization across subjects. This confirms that the patient-level splitting strategy successfully tested the model's ability to generalize to new individuals rather than merely memorizing training subjects.

Finally, the clinical plausibility of predictions is evident, as the predicted ages consistently align with expected ECG characteristics at different life stages. This suggests that the model has learned physiologically meaningful representations of cardiovascular aging rather than spurious correlations or dataset-specific artifacts, which is crucial for potential clinical deployment of the system.

These comprehensive results, validated through multiple complementary metrics (accuracy, F1-scores, MAE), visualizations (learning curves, confusion matrix, sample predictions), and detailed per-class analysis (Table 2), provide strong evidence that all four research tasks were successfully accomplished, with the proposed Advanced CNN architecture demonstrating superior performance for biological age estimation from ECG signals.

Discussion

The findings of this study validate and extend current understanding of deep learning applications in cardiovascular age estimation. Our Advanced CNN achieved 87.9% accuracy (Table 1), substantially exceeding the performance benchmarks established in recent literature and confirming the hypothesis that raw ECG signals contain rich aging biomarkers inaccessible to manual feature engineering.

Attia et al. [3] pioneered AI-based age estimation from 12-lead ECGs but focused primarily on binary or coarse age categorization. Our approach advances this work by demonstrating that fine-grained classification into 15 age groups with MAE of 2.6 years (Fig. 4) is clinically feasible, providing more granular risk stratification. This aligns with Lima et al.'s [2] demonstration that deep neural networks extract mortality-relevant features from ECG morphology. Our results provide mechanistic evidence for this finding: the model learned to identify QRS widening, P-wave alterations, and T-wave variations (Fig. 5) – established electrophysiological correlates of cardiac aging – without explicit feature engineering.

The integration of Residual Blocks and Self-Attention mechanisms (Fig. 2) directly addresses limitations identified by Kiranyaz et al. [4], who noted that standard 1D CNNs struggle with long-range temporal dependencies in physiological signals. Our architecture's stable convergence without overfitting (Fig. 3) contrasts sharply with the training stagnation observed in the Hybrid Autoencoder + RNN baseline (Table 1: 38.3% accuracy), demonstrating that end-to-end optimization with attention-guided feature selection outperforms two-stage dimensionality reduction approaches. The 23.5 percentage point improvement over Bi-LSTM (64.4%) suggests that convolutional architectures with attention are better suited for capturing spatial-temporal patterns in ECG morphology than purely recurrent approaches.

The severe demographic skew in aging datasets (Fig. 1) has historically limited model performance on minority classes. Our combined Focal Loss [12] and Weighted Random Sampling strategy achieved 71% recall for the oldest age group (238 samples), compared to near-zero performance for classical SMOTE-augmented models [8] (Table 1: 18-49% accuracy). This represents a methodological advance for imbalanced medical classification tasks. The Confusion Matrix (Fig. 4) confirms absence of systematic bias toward majority classes, contrasting with classical models that over-predict middle-age groups – a common failure mode in imbalanced learning that our approach successfully mitigates.

The strict patient-level splitting strategy addresses a pervasive issue in medical AI: optimistic performance estimates from data leakage. By ensuring complete independence between training and test subjects, our reported metrics reflect true generalization capability rather than subject-specific memorization. This is particularly important for clinical deployment, where models must perform on entirely new patient populations.

While our results are promising, critical limitations constrain immediate clinical applicability. The Jena University Hospital dataset [5, 6] comprises exclusively healthy volunteers, representing "normal aging" trajectories. As Belsky et al. [1] documented, biological aging is profoundly heterogeneous and accelerated by cardiovascular disease. Pathological conditions – myocardial infarction, heart failure, arrhythmias – fundamentally alter ECG morphology through mechanisms distinct from physiological aging. Consequently, our model's 87.9% accuracy may not generalize to clinical populations with comorbidities.

The "age gap" concept (predicted minus chronological age) holds promise as a risk biomarker. Lima et al. [2] showed that individuals with elevated ECG-derived age exhibit increased mortality risk independent of traditional factors. However, validating this requires prospective cohort studies tracking cardiovascular outcomes over years—beyond the scope of our cross-sectional analysis. Future work must demonstrate whether our model's age predictions stratify risk in diverse clinical populations before deployment in preventive cardiology workflows.

Conclusions and Future Research Directions

This study successfully developed and validated an Advanced CNN architecture for biological age estimation from ECG signals, achieving all four defined research objectives.

1. Patient-level data splitting ensured independence of test subjects, establishing genuine generalization capability (87.9% accuracy) rather than inflated metrics from data leakage.

2. The ResNet-Attention architecture achieved 87.9% accuracy and MAE of 2.6 years, representing 23.5 and 38.9 percentage point improvements over recurrent (64.4%) and classical (49%) baselines, respectively.

3. Focal Loss and Weighted Random Sampling achieved 71% recall for the most underrepresented geriatric group, compared to near-zero performance for classical SMOTE-augmented models.

4. Comprehensive evaluation demonstrated balanced performance (F1: 0.706-0.925) across all age groups without systematic bias.

Despite these results, important limitations must be acknowledged. The single-center dataset of healthy

volunteers implies demographic homogeneity and unknown performance on pathological ECGs. Future research must prioritize: (1) external validation on multi-center datasets with cardiovascular comorbidities, (2) longitudinal studies validating whether the "age gap" predicts cardiovascular risk and mortality, and (3) technical extensions including 12-lead ECG implementation and explainability techniques for clinical interpretability.

In conclusion, the developed "BioAge-ECG AI" system demonstrates that advanced CNN architectures significantly outperform classical approaches for biological age estimation from ECG signals. With appropriate validation on diverse clinical populations, this approach has the potential to enable non-invasive, low-cost cardiovascular risk assessment for population screening and early identification of individuals with accelerated cardiovascular aging.

Authors' contributions. Volodymyr Slipchenko – conceptualization; methodology; Liubov Poliahushko – methodology, formal analysis, verification of results, writing – review and editing; Vladyslav Shatylo – analysis of sources and preparation of literature review; Dmytro Antonkin – software, conducting experiments, data processing and visualization, writing – original draft preparation.

Declaration on artificial intelligence

Artificial intelligence was used for grammar checking and stylistic editing of the text. All scientific ideas, methodology, data analysis, and conclusions were developed by the authors personally.

Conflict of interest

The authors declare that there is no conflict of interest and confirm that during the preparation of this work there were no commercial, financial, or other relationships that could be construed as influencing the results of the study or their interpretation. The work was performed in accordance with the principles of academic integrity, ethical standards for conducting scientific research, and editorial policy requirements for preventing conflicts of interest.

References

1. Belsky, D. W., et al. (2015). Quantification of biological aging in young adults. *Proceedings of the National Academy of Sciences*, 112(30), E4104-E4110. https://doi.org/10.1073/pnas.1506264112

2. Lima, E. M., et al. (2021). Deep neural network-estimated electrocardiographic age as a mortality predictor. *Nature Communications*, 12(1), 5117. https://doi.org/10.1038/s41467-021-25351-7

3. Attia, Z. I., et al. (2019). Age and sex estimation using artificial intelligence from standard 12-lead electrocardiograms. *Circulation: Arrhythmia and Electrophysiology*, 12(9), e007284. https://doi.org/10.1161/CIRCEP.119.007284

4. Kiranyaz, S., et al. (2021). 1D Convolutional Neural Networks and Applications: A Survey. *Mechanical Systems and Signal Processing*, 151, 107398. https://doi.org/10.1016/j.ymssp.2020.107398

5. Schumann, A., & Bär, K. J. (2022). Autonomic function regulating blood pressure and cardiac rhythm in aging and disease. *Scientific Data*, 9(1), 1-10. https://doi.org/10.1038/s41597-022-01121-5

6. Goldberger, A. L., et al. (2000). PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*, 101(23), e215-e220. https://doi.org/10.1161/01.CIR.101.23.e215

7. Welch, P. (1967). The use of Fast Fourier Transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2), 70-73. https://doi.org/10.1109/TAU.1967.1161901

8. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357. https://doi.org/10.1613/jair.953

9. Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11), 2673-2681. https://doi.org/10.1109/78.650093

10. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-778. https://doi.org/10.1109/CVPR.2016.90

11. Vaswani, A., et al. (2017). Attention is all you need. *Advances in neural information processing systems*, 30. https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

12. Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980*. https://doi.org/10.48550/arXiv.1412.6980

13. Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. *Proceedings of the IEEE international conference on computer vision*, 2980-2988. https://doi.org/10.1109/ICCV.2017.324

Надійшла до редакції: 10.12.25
Прийнята до друку: 17.03.26
Опубліковано: 30.03.26