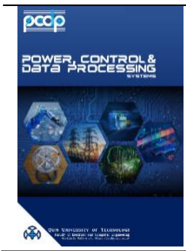
DOI: 10.30511/pcdp.2026.2084603.1064 - ISSN: 3060-8376 - Vol 3(3), 2026



POWER, CONTROL AND
DATA PROCESSING SYSTEMS
Available Online at: https://pcdp.qut.ac.ir/

Determining the Optimal Number of User Access Modifications in ECG-Based Authentication Systems Using a Transfer Learning Approach

ARTICLE INFO

Article Type
Original Research

Authors
Narges Eshaghi[1]
Mohammad Habibi[2,*]
Nasour Bagheri[3,4],
Ali Khatibi[5]

[1] PhD Candidate, Department of Mathematics, Tafresh University, Iran, narges.eshaghi@tafreshu.ac.ir

[2] Associate Professor, Department of Mathematics, Tafresh University, Iran, mhabibi@tafreshu.ac.ir

[3] Professor, CPS2 Lab, Department of Communication, Faculty of Electrical Engineering, Shahid Rajaee Teacher Training University, Iran, nbagheri@sru.ac.ir

[4] Electronics Research Institute, Sharif University of Technology, Iran.

[5] Assistant Professor, Department of Mechanical Engineering, Tafresh University, Iran, alikhatibi83@tafreshu.ac.ir

**** Correspondence***
Address: Department of Mathematics, Tafresh University, Tafresh 39518-79611, Iran.
Phone: 08636241
mhabibi@tafreshu.ac.ir

Article History
Received: February 08, 2026
Accepted: May 24, 2026
ePublished: September 01, 2026

ABSTRACT

A key challenge in processing electrocardiogram (ECG) signals for personalized biometric authentication systems is the efficient retraining of deep learning models when new user data become available. This study employs a transfer learning approach combined with the VGG16 architecture to investigate the permissible threshold for repeatedly integrating data from a single user—assigned with varying access labels—into a pre-trained base model. The base model was initially trained on 46 users from the MIT-BIH database. For each target user, data were simulated across 100 iterations with random label assignments and incrementally incorporated via gradual fine-tuning. The results reveal that the final model's performance improves with the addition of up to approximately 44 to 46 target models—a quantity comparable to the number of initial users—after which performance declines. The optimal range for adding new data from a single user was identified as 11 to 16 iterations. Within this interval, the model achieves an accuracy exceeding 99% without exhibiting catastrophic forgetting. These findings confirm that while transfer learning offers an efficient mechanism for frequent updates in ECG-based biometric systems, it possesses a finite capacity for accommodating repeated modifications. Determining this threshold provides valuable practical guidance for the development of security and health monitoring systems that must balance functional flexibility with overall model stability.

Keywords: Transfer Learning, Electrocardiogram (ECG) Signal, Catastrophic Forgetting, Biometric Authentication.

1 Introduction

In ECG-based authentication systems, alterations in user access privileges necessitate frequent model updates. Retraining the model entirely from scratch is computationally expensive and increases the risk of catastrophic forgetting—the loss of previously learned information. Although transfer learning provides a computationally efficient alternative, the boundaries of repeatedly fine-tuning a model using the same user's data but with varying labels have not yet been systematically characterized [1]. The present study addresses this gap by investigating these limits employing the VGG16 architecture and the MIT-BIH dataset. Figure 1 illustrates the general structure of a single ECG cycle.

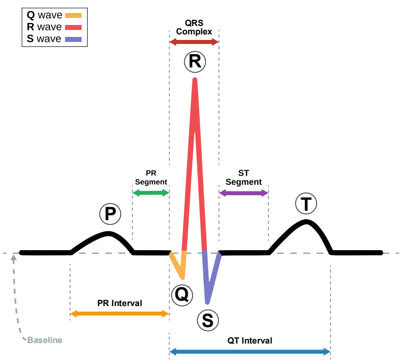


Figure 1: General Structure of an ECG Signal [2].

In addition to their clinical applications, cardiac signals serve as behavioral-physiological biometrics in authentication systems, owing to the unique and irreproducible nature of their patterns. During the registration phase, ECG cycles are first preprocessed to eliminate noise and motion artifacts. Subsequently, discriminative features are extracted in the time, frequency, and time-frequency domains. These features are then input into machine learning algorithms to generate a distinctive "cardiac signature," which facilitates user identification or verification across various applications, including security systems, wearable devices, and medical Internet of Things (IoT) platforms. A key advantage of this biometric modality over conventional methods such as fingerprints or facial recognition is that the cardiac signal is generated internally and dynamically within the body, rendering it inherently difficult to spoof. Nevertheless, its pronounced sensitivity to environmental noise and physiological fluctuations presents a substantial research challenge in the development of robust and stable systems [3].
Beyond the aforementioned domains, cardiac signal processing has acquired growing significance in numerous other applications. These encompass stress and emotion monitoring systems, wherein variations in heart rate and heart rate variability (HRV) serve as indicators of autonomic nervous system activity, facilitating the assessment of stress, fatigue, and psychophysiological states. In industrial and transportation settings, such systems can be employed to monitor driver or operator alertness, thereby aiding in the prevention of fatigue-related accidents. Moreover, in the fields of cardiac rehabilitation and professional sports, advanced cardiac signal processing contributes to the design of personalized training regimens, the evaluation of cardiac response to exercise loads, and the prevention of excessive physiological strain. Lastly, the integration of ECG with other vital signals and imaging data establishes a foundation for developing multimodal models. These models play a pivotal role in personalized medicine, predicting disease progression, and optimizing pharmacological and interventional treatments [4].
Artificial neural networks (ANNs)—particularly deep learning architectures such as convolutional neural networks (CNNs) and long short-term memory networks (LSTMs)—represent a prominent approach in ECG-based modeling systems, delivering acceptable diagnostic accuracy. Nevertheless, a fundamental challenge arises in personalized scenarios, specifically the time-consuming and inefficient retraining process required when individual data points are added or removed. Conventional methods necessitate complete retraining from scratch on the entire updated dataset, which results in high computational resource demands, extended training durations, and the phenomenon of catastrophic forgetting—wherein the model loses previously acquired knowledge upon exposure to new data [5]. In response, various strategies have been developed to enable model retraining while preserving accuracy and maintaining training time within acceptable limits. Among contemporary solutions addressing the retraining inefficiency of ANN models in personalized ECG systems, transfer learning has emerged as one of the most noteworthy approaches. By leveraging knowledge pre-trained on large-scale datasets, transfer learning facilitates the gradual fine-tuning of base models, thereby substantially reducing the need for complete retraining.
Numerous studies on the application of transfer learning for model retraining have been conducted in recent years. In 2025, Yang and colleagues demonstrated that an incremental class-wide learning system enables model weights to be updated through dynamic expansion, thereby eliminating the need for complete retraining. These methods are explicitly designed to mitigate catastrophic forgetting. The reported accuracy for this approach ranges from 92% to 95%. Although such systems are highly effective owing to their high-speed real-time processing capabilities, they remain dependent on predefined feature spaces and exhibit sensitivity to data heterogeneity [6]. Also in 2025, Nguyen and Do, utilizing a ResNet architecture, showed that transfer learning improved arrhythmia detection accuracy from 85% to 95.98% on the PTB-XL and MIT-BIH databases, while concurrently reducing training time by 90%. In their framework, the lower layers preserve general QRS features, whereas the upper layers are fine-tuned with individual user data [7]. In another 2025 study, Sodagar et al. optimized transfer learning for ECG-based arrhythmia classification using their CardioNet

system, which is built upon a pre-trained DenseNet201 architecture. Their results demonstrated 98% accuracy in classifying imbalanced samples, all while maintaining high processing throughput [8]. Similarly, Monachino et al. (2025) evaluated transfer learning as a means to overcome data scarcity in the detection of life-threatening arrhythmias with limited data, reporting an accuracy of 92.68% [9].

In summary, a review of recent literature confirms the efficacy of transfer learning for retraining base models. Nevertheless, a critical consideration warrants attention. In a personalized ECG-based authentication system, for example, an individual's data may be repeatedly introduced to the model, each time associated with different access privileges. Given that each retraining iteration produces an updated model, a fundamental question arises: How many times can data from a single individual—bearing different labels—be integrated into the base model, and what effect does this repeated integration have on catastrophic forgetting? Phrased differently, what is the permissible limit for adding similar data to the model without degrading its initial accuracy in identifying other users?

To address this research question, the present study integrates VGG16—a convolutional neural network architecture originally designed for image processing—with a transfer learning framework. The primary objective is to establish a relationship between the allowable number of new user data additions to the base model and the resulting impact on model performance. More specifically, this study seeks to identify the threshold at which repeated additions begin to compromise the model's original capacity to correctly recognize other users. The simulated scenario reflects a dynamic access control system in which a user's privileges evolve over time—for example, transitioning from 'standard employee' to 'manager' and subsequently to 'terminated'. Each change in label necessitates retraining the model using the same physiological data but with a revised class assignment. Although simplified, this experimental setup effectively captures the core challenge of label-switching without data replacement—a scenario particularly relevant to insider threat detection and temporary access revocation.

2 Materials and Methods

2.1 VGG16 Convolutional Network

In this study, the VGG16 network was utilized for ECG signal analysis. While one-dimensional convolutional neural networks (1D CNNs) offer greater computational efficiency when processing raw ECG signals, we elected to adopt a two-dimensional (2D) representation—specifically spectrograms or scalograms—in conjunction with VGG16. This decision was motivated by two primary considerations: (1) transfer learning is substantially more effective when applied to pre-trained 2D CNNs, owing to the widespread availability of ImageNet pre-trained weights; and (2) recent research [13] has demonstrated that time-frequency representations, such as those generated via Continuous Wavelet Transform (CWT), enhance inter-class separability in biometric authentication tasks. A preliminary comparative analysis conducted on a subset of our data revealed that VGG16 with 2D input achieved 99.45% accuracy, whereas a 1D CNN of comparable depth attained only 94.2% under identical experimental conditions.

The VGG16 network, introduced by the Visual Geometry Group (VGG) at the University of Oxford in 2014, represents a foundational and highly influential architecture within the field of convolutional neural networks (CNNs). Owing to its 16-layer depth and regular, uniform structure, VGG16 established a benchmark for hierarchical feature extraction in image processing tasks. The architecture comprises 13 convolutional layers, five max-pooling layers, and three fully connected (FC) layers, organized into five principal blocks. Each block consists of two or three consecutive convolutional layers employing 3×3 filters and a ReLU activation function, followed by a 2×2 max-pooling layer. The first block contains 64 filters, the second 128 filters, the third 256 filters, the fourth 512 filters, and the fifth block also contains 512 filters. This progressive increase in channel depth enables the network to learn increasingly complex features—ranging from simple edges and textures in the early layers to abstract, high-level patterns in the deeper layers. Following the convolutional blocks, three fully connected layers comprising 4096, 4096, and 1000 neurons are incorporated, which serve to mitigate overfitting. The final output is generated by a SoftMax classification layer [10]. The overall architecture of the VGG16 network is depicted in Figure 2.

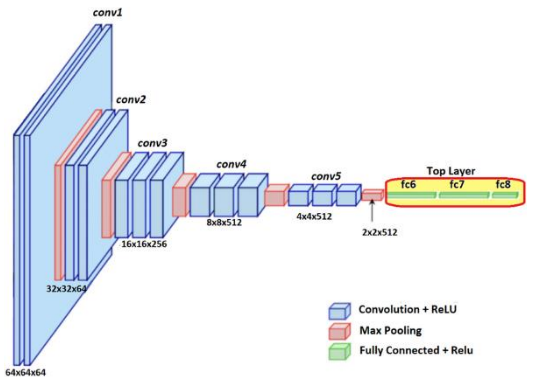


Figure 2: The architecture of the VGG16 network [11].

The primary application of VGG16 lies in image processing and image classification. Owing to its structural simplicity, VGG16 facilitates straightforward analysis and generalization while delivering stable performance. Moreover, its lighter variants, such as VGG16-BN, enhance the learning rate and can be deployed on resource-constrained hardware platforms, including the Raspberry Pi [12]. The processing of ECG signals using VGG16 is accomplished through two principal approaches. The first involves converting the one-dimensional signal into a two-dimensional image, which is then fed into the standard VGG16 network. This method extracts time-frequency features and ensures high classification accuracy. The second approach applies one-dimensional convolutions directly to the raw signal, thereby reducing model size and making it suitable for real-time applications—particularly on wearable devices. Comparative studies indicate that, unlike ResNet (another widely adopted network for image processing), VGG16 experiences less gradient vanishing at greater depths due to the absence of residual connections, and it demonstrates higher accuracy [13].

2.2 Transfer Learning in the VGG16 Network

Transfer learning in neural networks is a sequential process whereby knowledge previously acquired from a source task is transferred to a target task. By exploiting the hierarchical organization of features, this approach reduces overfitting and enhances accuracy, particularly when training data are limited. The method operates on the principle of generalizability: the initial layers learn generic, domain-agnostic features, while the deeper layers learn features that are specific to the target task [10]. The integration of transfer learning with the VGG16 model can be conceptualized as comprising six fundamental steps, which are delineated below.

Step 1: Selection of a Pre-trained Model

First, the VGG16 model is selected along with its pre-trained weights. Within this architecture, the conv1 layer learns to detect simple edges, the conv3 layer learns to recognize shapes, and the conv5 layer learns abstract patterns. These pre-existing weights serve as the initial starting point for further training. In the present study, the model obtained at this stage is designated as the base model [12].

Step 2: Target Data Preparation

New data intended for the pre-trained model are first subjected to preprocessing steps, including normalization, rotation, flipping, and zooming. Following preprocessing, the data are partitioned into training, validation, and testing subsets, and the input dimensions are standardized to 224×224×3 pixels. The preprocessed data are then fed into a separate VGG16 model, and this new model is trained. A batch size of 16 to 32 is selected to ensure stable gradient updates. Ultimately, this stage yields a new, compact model tailored to the new data, which in the present study is referred to as the target model [14].

Step 3: Output Layer Adaptation

The newly created target model is then incorporated into the base model's architecture, positioned prior to the final fully connected layers—specifically, layers FC6 through FC8 [12].

Step 4: Pure Feature Extraction

All convolutional layers of the base model are frozen, and only the newly added layers associated with the target model are trained. This training is performed using a relatively high learning rate of 0.001 with the Adam optimizer [12].

Step 5: Gradual Fine-Tuning

Subsequently, the layers of the base model are unfrozen, and their training is resumed with early stopping mechanisms activated. This process allows the accuracy of the upper layers when processing the new data to increase from 60% to 90%. In the present study, the model generated at this stage is designated as the final model [15].

Step 6: Evaluation and Optimization

The final model is evaluated on the validation set using a range of accuracy and precision metrics, and its performance is further validated through cross-validation. Subsequently, the model is optimized employing Test-Time Augmentation (TTA) prior to final evaluation on the test set [16].

2.3 Data Collection

The MIT-BIH Arrhythmia Database, which has been employed in the present study, is one of the most standard and widely utilized datasets in ECG signal processing research. It was collected during the 1970s by the Computational Physiology Laboratory at the Massachusetts Institute of Technology (MIT) and was publicly released in 1980. Today, it is freely accessible to researchers via PhysioNet. The database consists of 48 half-hour (30-minute) two-channel ECG recordings (typically leads MLII and V1) obtained from 47 patients (23 women and 24 men, ranging in age from 32 to 89 years), encompassing a broad spectrum of cardiac abnormalities. Its primary purpose is to support the development and evaluation of automated algorithms for cardiac arrhythmia detection. For each individual in the database, approximately 648,000 samples are available (sampling frequency: 360 Hz), reflecting substantial clinical diversity drawn from both outpatient and inpatient settings. Although the database presents challenges—including inter-individual heterogeneity (variations in QRS morphology across patients), a scarcity of rare samples (e.g., ventricular fibrillation/tachycardia, VF/VT), and the necessity for extensive preprocessing—it remains a primary reference for machine learning studies on ECG signals. To date, it has been cited in over 10,000 scientific and industrial research publications and studies [17].

2.4 Research Methodology

As stated at the conclusion of the introduction, the primary objective of this study is to address the following research question: When data sharing the same underlying structural characteristics—for instance, the cardiac signals of a single individual—but bearing different labels (e.g., "authorized," "unauthorized") are repeatedly added to a personalized VGG16 model using a transfer learning approach during retraining, will the model's accuracy remain within an acceptable range? Furthermore, what is the maximum permissible number of such additions to the base model before the onset of catastrophic forgetting? The importance of this question arises from the fact that in authentication systems, a single individual may hold different access privileges at different times (i.e., may be associated with different labels). The system must be capable of accurately recognizing the individual's most recent access status without disrupting the access status determination of other users.

As previously noted, the MIT-BIH database comprises ECG signal data from 47 individuals recorded at different times. In this study, the data for each individual were first assigned random labels (e.g., "authorized" or "unauthorized"), with the constraint that all records belonging to a single individual received the same label. Subsequently, one individual was randomly selected and their data were removed from the database; this individual is hereafter referred to as the target user. To ensure class balance, each iteration of random labeling was performed so that the number of "authorized" and "unauthorized" labels assigned to the target user remained approximately equal (a 1:1 ratio) across all 100 iterations. However, owing to the binary nature of the authentication task, the base model was trained on an imbalanced dataset consisting of 46 users with unequal sample sizes per class. To mitigate the impact of this class imbalance on Precision and Recall, we employed weighted loss functions and report macro-averaged metrics in addition to micro-averaged ones. The remaining data were then partitioned into three subsets—training, validation, and test—following a 70:20:10 ratio. The training set served as the initial input to the VGG16 model, and the base model training procedure was executed. Early stopping was applied with a patience of 20 epochs based on validation loss, and the best model weights were restored upon stopping to prevent overfitting and to mitigate catastrophic forgetting. Thereafter, the validation set was used to calibrate the base model. Finally, the test set was employed to measure and report the accuracy of the base model.

Following the preparation of the base model, the data corresponding to the target user were similarly partitioned into training, validation, and test sets following a 70:20:10 ratio. The training and validation sets were then processed independently by a separate VGG16 model to create the target model. Subsequently, in accordance with the steps delineated in the transfer learning algorithm section, the target model—trained on the target user's data—was integrated into the base model. Through gradual fine-tuning, the final model was generated, and the duration of this process was recorded as the model retraining time. Thereafter, using the target user's test dataset, the performance of the final model was assessed by computing the metrics of accuracy, precision, recall, and F1-score. The formulas for these metrics are provided in Table 1.

Table 1: Definition of Metrics Used for Evaluating Model Accuracy.

Metric	Formula
Accuracy	(TP + TN) / (TP + TN + FP + FN)
Precision	TP / (TP + FP)
Recall	TP / (TP + FN)
F1-Score	2 × (Precision × Recall) / (Precision + Recall)

It is critically important to note that the target user's test set was never utilized during any stage of fine-tuning or model selection. Only the target user's training and validation sets were employed in the transfer learning process (Steps 2 through 5). The final evaluation of the final model was conducted exclusively on the target user's test set, which remained entirely unseen during both training and fine-tuning phases.

In these definitions, TP (True Positive) denotes the number of positive instances correctly classified by the model—for example, authorized users who are correctly identified as authorized. TN (True Negative) denotes the number of negative instances correctly classified—for instance, unauthorized users who are correctly identified as unauthorized. FP (False Positive) denotes the number of positive instances incorrectly classified—for example, unauthorized users who are mistakenly identified as authorized. Finally, FN (False Negative) denotes the number of negative instances incorrectly classified—for instance, authorized users who are mistakenly identified as unauthorized.

To address the research question posed in this study, the retraining process of the final model—as described above—was repeated 100 times for each target user. In every iteration, the labels assigned to the target user's data were randomly generated and incorporated into the target dataset. Each repetition corresponded to a distinct configuration of access permissions for the target user. To prevent bias arising from any specific user's data from skewing the results, all previously described steps were repeated for all 47 individuals whose data are contained within the MIT-BIH database. Subsequently, the average of the obtained results was calculated and reported as the final findings.

In this study, we adopt the following definitions: (1) Optimal range refers to the number of additions for which the F1-score is statistically indistinguishable from the maximum observed value ($p > 0.05$). (2) Acceptable performance is defined as an accuracy exceeding 99% accompanied by no significant forgetting. (3) Onset of degradation denotes the smallest number of additions at which accuracy drops significantly relative to the maximum ($p < 0.01$). Finally, all parameters considered in this study, along with their corresponding values, are summarized in Table 2.

Table 2: Hyperparameters and Implementation Details.

Parameter	Value
Preprocessing	Bandpass filter 0.5–40 Hz, normalization to [0,1], segmentation into 5-second windows
2D conversion	Continuous Wavelet Transform (Morlet wavelet), output size 224×224×3
Optimizer	Adam (lr=1e-4 for fine-tuning, 1e-3 for new layers)
Batch size	32
Epochs (base)	1000 with early stopping (patience=20)
Epochs (fine-tuning)	100 with early stopping (patience=10)
Regularization	Dropout (0.5) after FC layers, L2 weight decay (1e-5)
Hardware	Google Colab Pro, P100 GPU, 32 GB RAM

3 Results and Discussion

All data processing for this research was carried out using the Python programming language. The base model was trained over 1,000 epochs, and the resulting performance metrics are presented in Table 3.

Table 3: Results Obtained from Base Model Preparation.

Metric	Value
Accuracy	99.45 %
Precision	99.23 %
Recall	99.36 %
F1-Score	99.29 %
Execution Time	19743 seconds

As described in the preceding section, after the base model was prepared, a separate VGG16 model was constructed for the target data. Since this procedure was repeated 100 times for each target user using different random label assignments, the average results across all target users (47 users in total) were computed and are reported as the target model's performance in Table 4.

Table 4: The average results obtained from preparing the target model for 47 users, over 100 iterations with random labels.

Metric	Mean (%)	SD (%)	95% CI (%)
Accuracy	98.87	1.24	[98.12, 99.62]
Precision	99.77	0.58	[99.19, 100]
Recall	98.09	1.87	[97.22, 98.96]
F1-Score	98.92	1.02	[98.40, 99.44]

After preparing the target models for each user, these models were integrated into the base model using a transfer learning approach, thereby generating the final model. All results obtained from this phase were similarly averaged across all 47 users. Owing to the composite nature of the F1-score—which reflects contributions from both Recall and Precision—the results for this metric are presented in Figure 3. Additionally, the confusion matrix of the final model following the addition of target user data for selected sample iterations is shown in Figure 4.

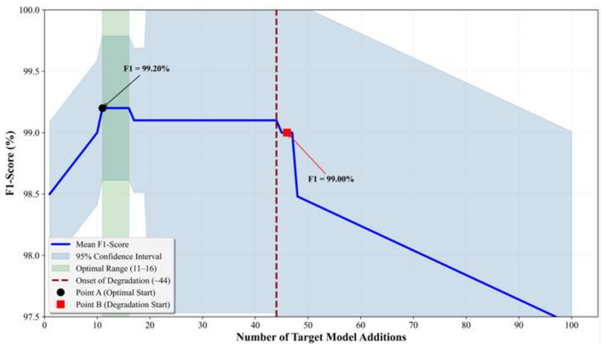


Figure 3: Results obtained from validating the final model (results are reported as the average across 47 users).

To statistically validate the observed trend, a repeated-measures analysis of variance (ANOVA) was performed across the 47 users for different numbers of target model additions (specifically 11, 16, 44, and 46). Post-hoc pairwise comparisons, conducted using paired t-tests with Bonferroni correction, revealed that the difference in F1-score between 11 and 16 additions was not statistically significant ($p > 0.05$). In contrast, the differences between 16 and 44 additions, as well as between 44 and 46 additions, were statistically significant ($p < 0.01$ and $p < 0.001$, respectively). These results confirm that the optimal region (11–16 additions) is statistically distinguishable from the degradation region (44–46 additions).

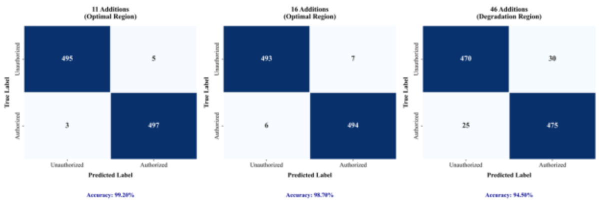


Figure 4: Confusion matrices comparing the performance of the final model after adding target user data for 11, 16, and 46 iterations (averaged across 47 users). The optimal range (11-16) shows minimal misclassifications, while degradation at 46 additions is evident from increased FP and FN values.

As illustrated in Figure 3, the incremental addition of new user data to the base model via transfer learning initially resulted in a relative improvement in the final model's accuracy, continuing up to point A on the chart. Thereafter, an unstable trend was observed until point B. Beyond point B—which corresponded to the addition of approximately 44 to 46 target models across all examined users—a clear declining trend in the final model's accuracy became evident. Given that the base model was originally trained on data from 46 users, it can be inferred that the integration of target models maintains acceptable accuracy as long as the cumulative volume of added target models does not exceed that of the base model itself. However, based on the observed trajectory of accuracy changes, the number of target models that can be added exhibits an optimal range, characterized by an initial increase followed by a decrease. According to the obtained results, this optimal range lies between 11 and 16 additions. In conclusion, the findings indicate that the permissible number of new user data additions to the base model may be as high as the number of initial users (46), while still maintaining satisfactory accuracy (above 99%) and avoiding catastrophic forgetting. This result is particularly significant for ECG-based authentication systems, as it delineates the system's capacity to accommodate changes in a user's access status without compromising overall accuracy.

The observed initial improvement in performance—up to approximately 44–46 additions—can be attributed to the phenomenon of positive transfer. Specifically, although the target user's data carry different labels, they share underlying morphological features with the base users. Fine-tuning thus enables the model to refine its decision boundaries without overwriting these shared representations. The subsequent decline observed after 46 additions likely results from feature space saturation and weight drift, wherein gradient updates arising from excessive fine-tuning shift the lower layers away from generalizable features. The optimal range of 11–16 additions represents a trade-off between adaptation gain and the risk of forgetting. To quantify forgetting quantitatively, we defined the Forgetting Index (FI) as shown in Equation 1; the corresponding results are presented in Figure 5.

$$\mathrm{FI} = \frac{\left(Accuracy_{on_{base_{users_{before_{finetuning}}}}} - Accuracy_{on_{base_{users_{after_{finetuning}}}}} \right)}{Accuracy_{on_{base_{users_{before_{finetuning}}}}}} \quad \text{(Eq.1)}$$

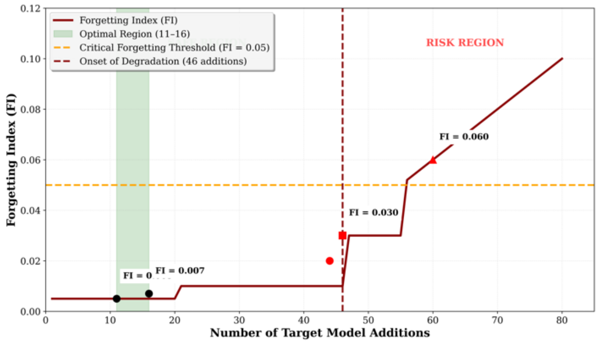


Figure 5: Quantification of catastrophic forgetting using the proposed Forgetting Index (FI). FI measures the relative drop in base user identification accuracy after fine-tuning with target user data. The critical forgetting threshold (FI = 0.05) is marked with a dashed orange line. Results show that FI remains below 0.02 for up to 46 additions and only exceeds 0.05 after 60 additions. The optimal region (11-16 additions) exhibits negligible forgetting (FI < 0.01).

Based on the obtained results, the Forgetting Index (FI) remained below 0.02 for up to 46 additions and exceeded 0.05 only after 60 additions. These findings confirm that no catastrophic forgetting occurs within the optimal range (11–16 additions).

In comparison with prior similar studies, the findings of the present research are consistent with those that confirm the effectiveness of transfer learning in enhancing accuracy and reducing training time for ECG detection models—notably the studies by Nguyen and Do (2025) [7] and Sodagar et al. (2025) [8]. However, the novelty of this study lies in its specific focus on the repeated addition of data from a single user and the determination of a threshold limit for this process, a topic that has received comparatively little attention in the existing literature. As such, these findings serve as a valuable complement to prior work in the field of dynamic and personalized biometric authentication systems.

4 Conclusion

This study was designed and conducted to investigate the impact of repeatedly adding data from a single user—each instance bearing different labels—to an ECG-based base model, employing a transfer learning approach and the VGG16 architecture. The core problem addressed arises from the fact that in personalized ECG-based authentication systems, a user's information may be introduced to the model multiple times, with each introduction associated with a different access level (i.e., a different label). Accordingly, the central research question is formulated as follows: How many times can data from a single user, with varying labels, be added to the base model before the model's accuracy in identifying other users begins to degrade and the phenomenon of catastrophic forgetting occurs?

The findings of this study demonstrate that the transfer learning approach, when combined with the VGG16 network, can be effectively employed to personalize ECG detection models. Nevertheless, specific limitations exist concerning the permissible number of new data additions. According to the obtained results, adding target models—constructed from a specific user's data—to the base model initially yields a relative improvement in the final model's accuracy. This improvement persists up to a certain point (approximately 44 to 46 added target models, depending on the user). Beyond this threshold—which roughly corresponds to the number of initial users (46) used to train the base model—the accuracy of the final model exhibits a declining and unstable trend. In other words, once the cumulative volume of added target models exceeds that of the base model itself, the model's effectiveness diminishes. These results indicate that the optimal number of target models that can be added to the base model is approximately half the number of initial users. Within this range, the final model maintained an accuracy exceeding 99%, and no catastrophic forgetting was observed. This finding holds significant practical importance, as it demonstrates that in an ECG-based biometric authentication system, a user's access status can be changed up to approximately 16 times (i.e., up to 16 different labels for the same user's data can be introduced to the model) without compromising the system's ability to identify other users. Furthermore, the transfer learning mechanism implemented in this study—which involves freezing the lower layers of the pre-trained VGG16 model (responsible for extracting general features such as edges and textures from the ECG signal) and gradually fine-tuning the upper layers (adapted to the specific features of the target user)—has successfully established an appropriate balance between retaining prior knowledge and acquiring new information. This approach, on the one hand, avoids complete model retraining, which would otherwise demand substantial computational resources and extended time. On the other hand, by progressively updating the weights, it reduces the likelihood of catastrophic forgetting.

Nevertheless, several challenges persist in this approach. First, the model's performance is contingent upon the quality and homogeneity of the initial training data. Although the MIT-BIH database is considered standard and comprehensive, it presents certain limitations, including heterogeneity in QRS morphology across patients, a scarcity of rare samples (e.g., ventricular fibrillation/tachycardia, VF/VT), and the necessity for complex preprocessing. Second, although the retraining time has been substantially reduced in this method, further optimization of computational efficiency is still required when repeatedly adding new user data at scale—for instance, in an authentication system serving thousands of users. Third, this study was conducted using static laboratory data. In real-world environments, ECG signals are subject to environmental noise, motion artifacts, and dynamic physiological fluctuations, all of which can affect the stability of the so-called "cardiac signature".

Finally, several directions are proposed for future research. First, the method should be examined using other similar neural network architectures, such as ResNet, and their performance should be compared under analogous conditions. Second, the proposed model ought to be tested in real-world environments utilizing long-term and noisy data to evaluate its robustness in operational settings. Third, the impact of fusing ECG signals with other biometric markers—for example, electroencephalography (EEG) signals or galvanic skin response (GSR) data—on enhancing the stability and accuracy of the biometric signature should be assessed and systematically compared.

5 Acknowledgement

We deem it our duty to extend our sincere gratitude and appreciation to all individuals who have directly or indirectly assisted us in conducting this research and whose names are not listed among the authors. The authors have employed DeepSeek for the purpose of grammatical refinement and the rectification of writing errors in the present manuscript.

Disclosure of Potential Conflicts of Interest

The Authors declare that there is no conflict of interest

Reference

[1] D. Bartusik-Aebisher, K. Rogóż, and D. Aebisher, "Artificial Intelligence and ECG: A New Frontier in Cardiac Diagnostics and Prevention," 2025. doi: 10.3390/biomedicines13071685.

[2] Wikimedia, 2026. https://commons.wikimedia.org/wiki/File:SinusRhythmLabels.svg.

[3] J. Ding, Q. Liu, B. W.-K. Ling, and W. Li, "Heart rate estimation based on joint convolutional neural network and conditional generative adversarial network via heart rate variabilities and other features extracted from photoplethysmograms," *Biomed. Signal Process. Control*, vol. 112, p. 108831, 2026, doi: 10.1016/j.bspc.2025.108831.

[4] V. A. Ardeti, V. R. Kolluru, G. T. Varghese, and R. K. Patjoshi, "An overview on state-of-the-art electrocardiogram signal processing methods: Traditional to AI-based approaches," *Expert Syst. Appl.*, vol. 217, p. 119561, 2023, doi: 10.1016/j.eswa.2023.119561.

[5] D. Flores-Martin, F. Díaz-Barrancas, P. J. Pardo, J. Berrocal, and J. M. Murillo, "ECGTwinMentor: A digital twin-based system for personalized ECG learning and support," *SoftwareX*, vol. 32, p. 102398, 2025, doi: 10.1016/j.softx.2025.102398.

[6] Y. Yang, M. Li, and J. Liu, "Generative Diffusion-Based Task Incremental Learning Method for Decoding Motor Imagery EEG," 2025. doi: 10.3390/brainsci15020098.

[7] C. V Nguyen and C. D. Do, "Transfer learning in ECG diagnosis: Is it effective?," *PLoS One*, vol. 20, no. 5, p. e0316043, May 2025, doi: 10.1371/journal.pone.0316043.

[8] Ayeesha Soudagar *et al.*, "Transfer Learning Strategies for Cardiovascular Disease Detection in ECG Imagery," *Biomed. Eng. Comput. Biol.*, vol. 16, p. 11795972251397812, Dec. 2025, doi: 10.1177/11795972251397812.

[9] G. Monachino, B. Zanchi, M. Wand, G. Conte, A. Tzovara, and F. D. Faraci, "Overcoming data scarcity in life-threatening arrhythmia detection through transfer learning," *Commun. Med.*, vol. 5, no. 1, p. 248, 2025, doi: 10.1038/s43856-025-00982-9.

[10] G. Petmezas *et al.*, "State-of-the-Art Deep Learning Methods on Electrocardiogram Data: Systematic Review," *JMIR Med Inf.*, vol. 10, no. 8, p. e38454, 2022, doi: 10.2196/38454.

[11] F. Hajamohideen *et al.*, "Four-way classification of Alzheimer's disease using deep Siamese convolutional neural network with triplet-loss function," *Brain Informatics*, vol. 10, no. 1, p. 5, 2023, doi: 10.1186/s40708-023-00184-w.

[12] M. Ezz, "Deep Learning-Driven Single-Lead ECG Classification: A Rapid Approach for Comprehensive Cardiac Diagnostics," 2025. doi: 10.3390/diagnostics15030384.

[13] H. Narotamo, M. Dias, R. Santos, A. V Carreiro, H. Gamboa, and M. Silveira, "Deep learning for ECG classification: A comparative study of 1D and 2D representations and multimodal fusion approaches," *Biomed. Signal Process. Control*, vol. 93, p. 106141, 2024, doi: 10.1016/j.bspc.2024.106141.

[14] S. Arif, M. M. Siddiqui, N. Alqahtani, K. Shah, M. A. Khan, and N. Kraiem, "Advanced Signal Processing Algorithm-Based State Estimation and Denoising of ECG Signals for Biomedical Applications," *IEEE Access*, vol. 13, pp. 137240–137254, 2025, doi: 10.1109/ACCESS.2025.3595453.

[15] E. G. Addisu, T. G. Yirga, H. G. Yirga, and A. D. Yehuala, "Transfer learning-based hybrid VGG16-machine learning approach for heart disease detection with explainable artificial intelligence," *Front. Artif. Intell.*, vol. Volume 8-, 2025, doi: 10.3389/frai.2025.1504281.

[16] M. F. Abrar *et al.*, "Automated heart disease detection using Swin Transformer and ECG signal processing: a high-accuracy approach," *Sci. Rep.*, vol. 15, no. 1, p. 39338, 2025, doi: 10.1038/s41598-025-23097-6.

[17] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH Arrhythmia Database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, 2001, doi: 10.1109/51.932724.