

Scalability analysis of neural network architectures for voice biometric authentication^{*}

Khrystyna Ruda^{1,*}, Dmytro Sabodashko^{1,†}, Ihor Kos^{1,†}, Ivan Opirskyy^{1,†},
and Alina Akhmedova^{1,†}

¹ Lviv Polytechnic National University, 12 Stepan Bandera str., 79000 Lviv, Ukraine

Abstract

The rapid implementation of digital services in financial, governmental, and commercial domains is accompanied by a growing demand for reliable user identification tools. Voice-based biometric authentication systems represent one of the most promising directions, as they combine ease of use with the potential for integration across a wide range of services—from banking operations and call centers to voice assistants. At the same time, such systems face a number of challenges: the need to scale to tens of thousands of users, to ensure resilience against attacks employing synthetic speech, and to maintain high accuracy in real-time operation. Neural network architectures capable of generating robust embeddings and preserving stability as data volumes increase are of particular importance. This article examines the scalability challenges of modern voice biometrics models and substantiates approaches to enhancing their efficiency, including architectural optimization, indexed search, and the use of representative speech corpora. The study emphasizes the necessity of a comprehensive approach to the development of voice authentication systems, in which technical performance is combined with security requirements and protection against cyber threats.

Keywords

Voice biometrics, scalability, speaker verification, embeddings, authentication, ECAPA-TDNN, Pyannote, WavLM

1. Introduction

The scalability of a biometric system is defined by its ability to maintain efficiency as the number of users and the volume of data increase. Biometric technologies—such as fingerprint, face, voice, and iris recognition—must ensure high accuracy and rapid response even when operating with millions of enrolled templates [1]. This requirement is particularly relevant for national identification systems and voice assistants, which serve massive user bases daily, where performance must not degrade under growing loads. In this context, scalability implies the ability to continuously add new users without system reconstruction, while maintaining stable response times [2–4]. Such requirements can be met through optimized search and indexing algorithms, distributed computing methods, and cloud-based infrastructures [5, 6]. However, scaling biometrics to large datasets introduces a number of challenges. First, computational costs rise sharply as the user base expands, since 1:N identification requires millions of comparisons; without optimization, this leads to unacceptable response times. Equally critical are latency and throughput issues, as authentication must be performed in real time, even in large-scale scenarios such as call centers or border control. Furthermore, data storage and transmission requirements increase, demanding efficient handling of large biometric templates (fingerprints, images, embeddings) with compact representation and fast access. Finally, accuracy must be preserved despite a growing number of users, the presence of similar biometric patterns, and intra-user variability caused by recording conditions, background noise, or age-related factors [7–9].

^{*} CPITS-II 2025: Workshop on Cybersecurity Providing in Information and Telecommunication Systems, October 26, 2025, Kyiv, Ukraine

^{*} Corresponding author.

[†] These authors contributed equally.

✉ khrystyna.s.ruda@lpnu.ua (K. Ruda); dmytro.v.sabodashko@lpnu.ua (D. Sabodashko); ihor.kos.mkb.2025@lpnu.ua (I. Kos); ivan.r.opirskyy@lpnu.ua (I. Opirskyy); alina.akhmedova.mkb.2025@lpnu.ua (A. Akhmedova)

ORCID 0000-0001-8644-411X (K. Ruda); 0000-0003-1675-0976 (D. Sabodashko); 0009-0001-4558-5036 (I. Kos); 0000-0002-8461-8996 (I. Opirskyy); 0009-0005-4743-7140 (A. Akhmedova)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Therefore, scalability in biometric systems is a multifaceted characteristic that integrates performance, accuracy, and security. It requires the application of robust models capable of operating in real time while meeting stringent data protection requirements [10].

2. Related Work and Literature Review

The analysis of recent scientific sources demonstrates a significant growth of interest in scalable solutions in the field of voice biometrics. Current research focuses on the development of embedding models, architectural optimization, and the applied use of technologies in banking services, call centers, and voice assistants.

In [11], the scalability of a voice authentication system based on the TitaNet architecture was investigated. The author showed that the model maintained high efficiency on smaller datasets but gradually degraded as the number of users increased. This highlights the need for adaptive threshold calibration and additional methods to improve accuracy. Thus, even state-of-the-art architectures reveal scalability limitations, underscoring the necessity of further optimization.

A more comprehensive comparative study was presented by Brydinskyi et al. (2024) [12], who analyzed several modern speaker verification models, including ECAPA, TitaNet, WavLM, and PyAnnote. The authors concluded that speaker embedding technology exhibits strong scalability, as new users can be added without retraining the model. Particular attention was given to the ECAPA architecture, which demonstrated a balanced trade-off between accuracy (EER ~1.7%) and inference speed (~69 ms), making it a promising candidate for large-scale authentication systems.

Continuing the topic of optimization, Thienpondt and Demuynck (2023) [13] proposed the hybrid ECAPA2 architecture, which enhances the robustness of speaker embeddings to noise and short utterances. The model achieved state-of-the-art results on the VoxCeleb1 dataset with fewer parameters compared to earlier approaches. This indicates the effectiveness of reducing computational complexity without sacrificing performance—an important factor in system scalability.

Another relevant challenge is speaker identification from short speech fragments. Deng et al. (2025) [14] introduced the Dense-Fusion2Net architecture with a Time-Frequency Channel Attention (TFCA) module, enabling efficient processing of signals of limited duration. The model demonstrated strong performance on VoxCeleb datasets, preserving both high accuracy and low computational cost. This makes it highly suitable for banking and service applications, where user interaction lasts only a few seconds.

A broader historical perspective on the evolution of technologies was provided by Sharma et al. (2024) [15]. The authors traced the transition from classical i-vector approaches to modern neural embedding models, emphasizing that the emergence of x-vector and subsequent deep architectures enabled the move toward scalable systems. Their review highlighted that contemporary methods allow operation in open-set scenarios, where user databases expand continuously, aligning with real-world application needs.

The issue of security in scalable systems was addressed by Chen et al. (2023) [16], who studied attempts to deceive speaker recognition systems using specially crafted audio adversarial examples. The authors demonstrated that combining audio signal transformations with adversarial training improved accuracy by ~13.6% and significantly increased system resistance to such attacks. This forms the basis for developing more secure and scalable biometric systems.

Another important security aspect is privacy and heterogeneous training conditions. Chen & Xu (2023) [17] proposed a personalized federated learning (PFL) approach for speaker verification and identification tasks. Their framework preserved user data privacy, enabled adaptation to diverse acoustic domains (rooms, noise conditions, languages), and avoided catastrophic forgetting in continuous learning (C-PFL). Across 12 simulated scenarios, this approach achieved lower average EER and better convergence compared to centralized training, highlighting its promise for scalable real-world systems.

The practical dimension of scalability is illustrated by industrial studies. For example, in the RudderAnalytics project (2024) [18], a system based on SpeechBrain with the VoxCeleb2 dataset was deployed. The developers showed that the system maintained high accuracy under scaling, successfully handling a 115% increase in registered users without performance degradation.

Meanwhile, even classical architectures remain subject to optimization. Sharif-Noughabi et al. (2025) [19] demonstrated that a modified VGG-CNN, combined with data augmentation techniques, significantly improved identification accuracy on VoxCeleb1 (from ~84% to over 91%). This indicates the effectiveness of integrating classical approaches with modern training techniques to ensure performance on large user databases.

In summary, the analysis of recent publications indicates that the scalability of voice biometric systems relies on compact and robust embeddings, lightweight architectures, quantization methods, and solutions for short speech recordings, complemented by security and privacy mechanisms. Modern neural models are moving toward unifying high accuracy, optimal inference speed, and resilience to attacks, thereby opening the way for large-scale deployment in banking services, call centers, and voice assistants.

3. Research Methodology

Highly accurate neural networks are often very large, which complicates their deployment on devices with limited resources or under large-scale cloud access. To address the computational challenges posed by such models, optimization methods are applied, in particular quantization—reducing the dimensionality of neural networks. For instance, Amazon reduced the Alexa voice assistant model to less than 1% of its original size without significant loss of accuracy [20]. This was achieved through parameter quantization and knowledge distillation, where a smaller student model learns from the outputs of a larger teacher model. Optimized models of this kind can be deployed on smartphones, in-car systems, and even microcontrollers. At the same time, for speaker identification, fast search in embedding databases is critical. Instead of exhaustive matching, index-based methods are employed, enabling scalability to tens of millions of templates [21].

Another critical component of scalable biometric systems is training data. The quality, size, and representativeness of training corpora directly influence model performance and adaptability. To ensure reliable operation across diverse conditions, models must be trained on large and heterogeneous datasets. Representativeness is especially important: datasets should cover a wide variety of accents, languages, and user groups to avoid bias and accuracy drops for underrepresented populations. For example, voice services trained predominantly on a single demographic may perform poorly for other groups with distinct accents, timbres, or vocabulary. Thus, investment in the collection and annotation of balanced, diverse training data is fundamental for scalable and universal AI systems.

Scalability is also inseparable from reliability. In high-availability services, such as banking voice platforms or consumer voice assistants, models must operate continuously (24/7). This requires deployment in clustered environments with effective load balancing across nodes. Such architectures prevent bottlenecks, ensure stable performance under peak loads, and support automatic failover: if one node fails, others seamlessly take over the workload. This approach guarantees service continuity, fault tolerance, and minimizes downtime risks—critical for online services, financial platforms, healthcare applications, and other domains where even brief disruptions can result in financial losses or reduced user trust.

Accordingly, system scalability encompasses not only the ability to process increasing data volumes and user requests, but also the assurance of robustness, fault tolerance, and uninterrupted service availability under real-world conditions.

4. Dataset Description

For the experimental study of neural model scalability in voice biometric authentication systems, a multilingual audio dataset was constructed, including both Ukrainian and English speakers [22]. The Ukrainian portion of the dataset consisted of recordings from 50 public figures (primarily politicians, government officials, and communicators), collected from open sources. As the main source, we used a public dataset hosted on the Hugging Face platform [23], which provided structured Ukrainian audio files of sufficient quality for the research. Each speaker was represented by 10 recordings with an average duration of approximately 10 seconds, ensuring an adequate amount of data for generating reference embeddings.

To evaluate model scalability under conditions of linguistic variability, the dataset was supplemented with recordings of 20 English-speaking individuals, including well-known actors, journalists, and public figures. These audio materials were manually collected from open sources, primarily video interviews published on YouTube.

To prevent overlap between training and testing examples, an extended dataset was prepared. Specifically, for each of the 70 speakers, an additional 20 unique audio recordings were created, which did not duplicate the samples used for constructing reference embeddings. This ensured the correctness of verification accuracy evaluation and the validity of experimental results.

5. System Architecture and Verification Methodology

In the first stage of the study, all audio files belonging to a single speaker were processed by each of the selected neural networks: Pyannote, WavLM base-sv, WavLM base-plus-sv, and ECAPA. As a result, 10 embeddings were generated and then averaged into a single vector—the speaker’s reference embedding. Averaging reduces the impact of natural voice variations (intonation, tempo, background noise). The resulting reference embeddings were stored in a database together with a unique speaker identifier.

In the next phase of the study, it was necessary to determine the optimal verification threshold. To this end, a full authentication process was simulated to assess system behavior under real-world conditions. The objective was to achieve a balanced trade-off between false negatives and false positives. Within this simulation, pairs of recordings were formed to cover two scenarios: verification of valid users, where test samples were compared with the reference embeddings of the same individual; and verification of invalid users, where test embeddings were compared with the reference vectors of other speakers [24–30].

Once the reference database was created, the system proceeded to the verification stage. Using the same neural model employed for reference construction, an extended set of voice samples from each speaker was processed to extract test embeddings. Each test vector was then compared with the corresponding reference embedding using cosine similarity. If the cosine distance value was below a predefined threshold, verification was considered successful; otherwise, the sample was not recognized as belonging to the claimed user (Figure 1).

To assess the results of the experimental study, a set of commonly accepted metrics traditionally used in biometric verification was applied. The most frequently considered indicators for quantitative analysis of system performance are Accuracy, False Acceptance Rate (FAR), False Rejection Rate (FRR), and the integral criterion Equal Error Rate (EER).

Accuracy is a basic metric that reflects the overall proportion of correctly classified cases in the binary speaker verification task. It is defined as the ratio of correct decisions (successful authentications and correct rejections) to the total number of verification attempts. Despite its intuitive clarity, this metric is sensitive to class imbalance and is therefore not considered decisive in verification studies.

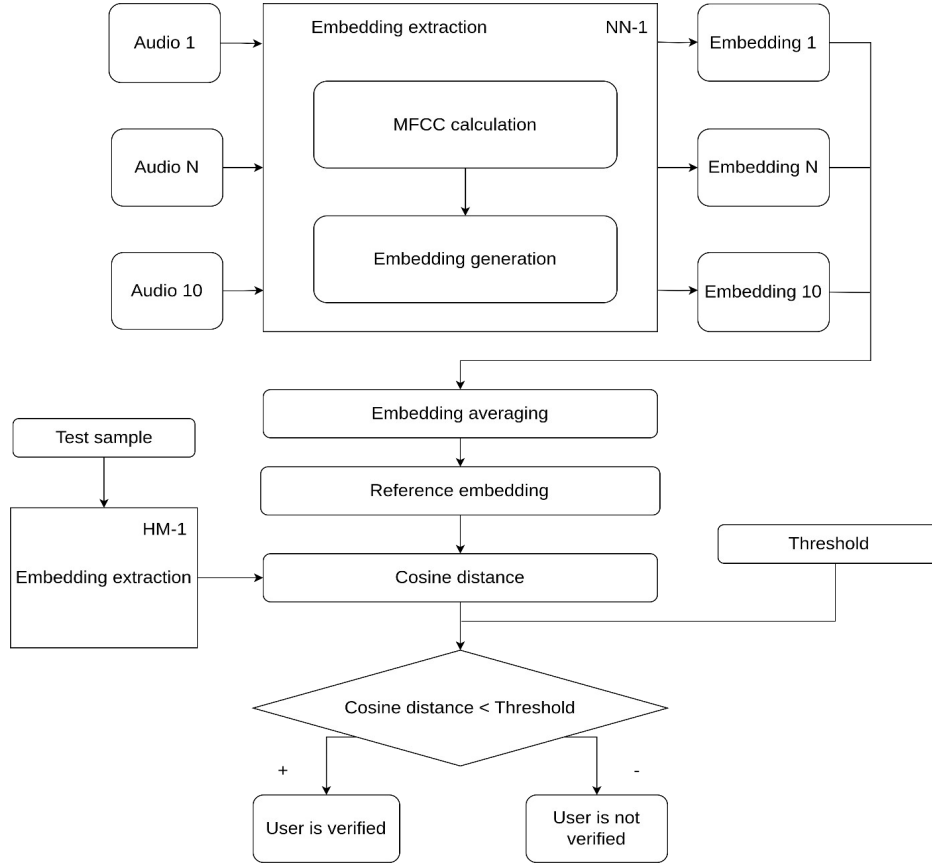


Figure 1: Workflow of the Voice Biometric Authentication System

False Acceptance Rate (FAR) characterizes the security of the system and represents the proportion of cases in which an unauthorized user is incorrectly accepted as legitimate. Formally, FAR is calculated as the ratio of false acceptances to the total number of access attempts made by impostors. Minimizing FAR is critically important for improving system resilience to attacks and ensuring reliability.

False Rejection Rate (FRR), on the contrary, reflects system usability and measures the proportion of cases in which a registered user is incorrectly rejected as unauthorized. FRR is calculated as the ratio of false rejections to the total number of access attempts made by genuine users. A low FRR ensures better user experience and higher system availability. Obviously, it is impossible to minimize both FAR and FRR simultaneously. The optimal balance between them is achieved by selecting an appropriate similarity threshold for embeddings, which determines the priority: minimizing false acceptances or false rejections.

Equal Error Rate (EER) serves as an integral indicator of biometric system quality, defined at the point where FAR and FRR intersect. The lower the EER, the more effective the system is considered. Due to its independence from a specific threshold value, this metric is widely used as a universal criterion for comparing different biometric authentication algorithms [31].

6. Research Results

A series of experiments was conducted to evaluate the scalability of the voice biometric authentication system at five levels: 10, 20, 35, 50, and 70 registered users. For each level, comparative testing was performed on four neural models—**Pyannote**, **WavLM base-sv**, **WavLM base-plus-sv**, and **ECAPA**—followed by analysis of accuracy, false acceptance rate (FAR), false rejection rate (FRR), and the equal error rate (EER).

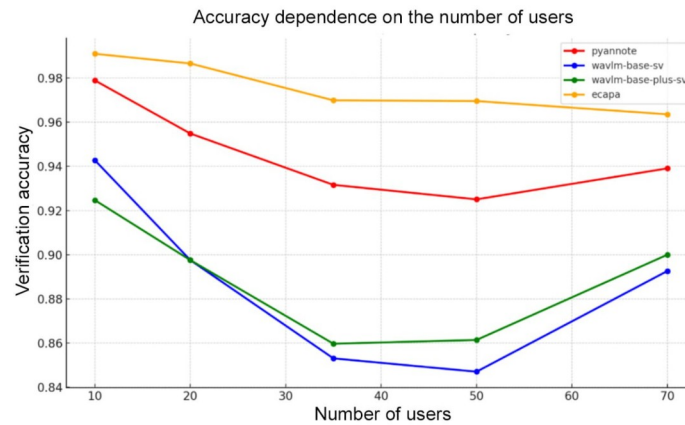
Table 1 presents the EER results obtained from the scalability experiments for the selected models.

Table 1

EER metrics for each model across all stages of the experiment.

Model	10, %	20, %	35, %	50, %	70, %
Pyannote	2.73	4.46	5.40	5.30	4.55
WavLM base-sv	7.42	10.13	11.72	10.90	8.09
WavLM base-plus-sv	9.77	10.12	11.27	9.92	7.47
ECAPA-TDNN	1.17	1.33	2.38	2.14	03.04

Regardless of scale, ECAPA-TDNN demonstrates the best performance: the EER remains within approximately 1.2–3.0% and consistently achieves the lowest value at each stage. Pyannote ranks second (about 2.7–5.4%), showing a moderate increase in error with the growing number of users and a slight improvement at 70 speakers. Both versions of WavLM exhibit the highest EER values (approximately 7–11%). A general trend is observed (Figure 2): as the number of speakers increases, the EER rises; however, partial stabilization occurs at the largest dataset size.

**Figure 2:** Accuracy as a function of the number of users

Based on the constructed graph of accuracy versus the number of users, it was established that the ECAPA-TDNN model demonstrates the highest stability and accuracy across all configurations. Its accuracy remains in the range of 95–98%, and even under maximum scaling (70 users) it maintains a high level of performance. Pyannote also showed stable accuracy, only slightly behind ECAPA-TDNN, with a mostly uniform degradation as the number of users increased.

The WavLM base-sv and WavLM base-plus-sv models demonstrated somewhat lower accuracy. A particularly noticeable decline was observed after exceeding the threshold of 35 users. This may indicate that, for these models, the increasing number of speakers complicates the discrimination of embeddings, leading to a higher number of system errors.

However, an interesting phenomenon was observed in the final iteration (70 users): the accuracy of Pyannote, WavLM base-sv, and WavLM base-plus-sv increased compared to the previous stage, breaking the general trend of degradation. This can be attributed to the larger pool of speakers, which allowed the models to establish better threshold values and reduce both false acceptances and false rejections.

Overall, the observed trend confirms the classical principle: as the number of users in the system increases, the task of distinguishing between voice embeddings becomes more complex, which can lead to higher error rates and reduced accuracy.

A histogram was constructed based on three key metrics: False Acceptance Rate (FAR) (red), False Rejection Rate (FRR) (blue), and Equal Error Rate (EER) (green). This visualization enabled a deeper analysis of model behavior under scalability conditions. The results provide a comprehensive assessment of the scalability and robustness of different architectures in the speaker verification task.

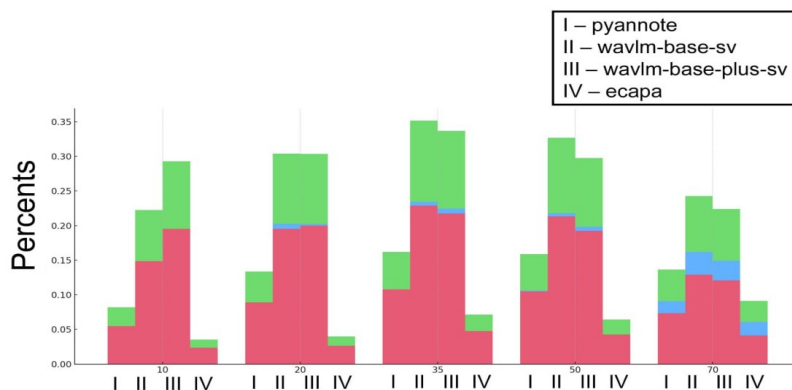


Figure 3: Histogram of FAR, FRR, and EER across scalability levels for the tested models

ECAPA-TDNN confirmed its high effectiveness: it maintained minimal EER values across all scalability levels and demonstrated a balanced trade-off between FAR and FRR. This indicates optimal threshold calibration and the model’s ability to generate discriminative embeddings that remain robust as the user base grows. Such characteristics make ECAPA-TDNN the most suitable choice for real-world deployment in systems with strict requirements for performance and security.

Pyannote also showed relatively low error rates; however, at higher scales (50–70 users), a gradual increase in FRR was observed. This reflects a reduced ability of the model to correctly recognize legitimate users as the database expands. Despite this, the overall error levels remain acceptable, making Pyannote suitable for medium-scale applications.

In contrast, WavLM base-sv revealed significant limitations: elevated FAR values indicate a higher probability of false acceptances. Such behavior is critically dangerous in security-sensitive scenarios, such as banking services or government registries, where even isolated false acceptances can have severe consequences.

The WavLM base-plus-sv variant slightly reduced EER compared to the base version but did not demonstrate stability comparable to ECAPA or Pyannote. As the number of users increased, FRR rose substantially, indicating the model’s declining ability to reliably recognize legitimate speakers. This significantly reduces usability and can negatively affect user experience.

The overall trend shows that scaling the user base is not a neutral factor across all models: the most robust architectures (such as ECAPA) maintain stable or even improved EER values due to better alignment of embeddings, while weaker models exhibit simultaneous growth of both FAR and FRR.

Thus, the results confirm that the scalability of biometric systems directly depends on the choice of model architecture. Powerful solutions such as ECAPA provide high tolerance to user base expansion, while other architectures require additional measures—embedding optimization, adaptive threshold selection, or even retraining on representative data. This underscores the need not only for careful model selection but also for the application of comprehensive methods to maintain scalability in the development of real-world voice biometric systems.

7. Conclusions

This study presented an experimental comparison of four neural models for voice authentication in order to evaluate their scalability. It was established that ECAPA-TDNN demonstrates the highest stability and the lowest EER values across all user levels, confirming its suitability for large-scale real-time systems. The Pyannote model also maintains acceptable accuracy, although it is characterized by a gradual increase in FRR as the database expands. In contrast, the WavLM

architectures revealed scalability limitations: the base-sv variant exhibited elevated FAR, while the base-plus-sv version showed instability in FRR.

The findings indicate that maintaining high system performance and reliability requires the use of compact embeddings, indexed search, and model optimization methods such as quantization and knowledge distillation. The practical significance of the results lies in their applicability for the development of scalable biometric services in domains such as banking, call centers, and voice assistants.

Declaration on Generative AI

While preparing this work, the authors used the AI programs Grammarly Pro to correct text grammar and Strike Plagiarism to search for possible plagiarism. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the publication's content.

References

- [1] B. Desplanques, J. Thienpondt, K. Demuynck, ECAPA-TDNN: Emphasized Channel Attention, Propagation and Aggregation in TDNN-based Speaker Verification, in: *Interspeech 2020*, 3830–3834. doi:10.21437/Interspeech.2020-2650
- [2] A. Nagrani, J. S. Chung, A. Zisserman, VoxCeleb: Large-Scale Speaker Verification in the wild, *Comput. Speech Lang.*, 60 (2020) 101027.
- [3] K. Okabe, T. Koshinaka, K. Shinoda, Attentive Statistics Pooling for Deep Speaker Embedding, in: *Interspeech 2018*, 2252–2256. doi:10.21437/Interspeech.2018-993
- [4] Z. Fan, M. Li, S. Zhou, B. Xu, Exploring wav2vec 2.0 on Speaker Verification and Language Identification, in: *Interspeech 2021*, 1509–1513. doi:10.21437/Interspeech.2021-1280
- [5] V. Lakhno, et al., Management of Information Protection based on the Integrated Implementation of Decision Support Systems, *East.-Eur. J. Enterp. Technol.*, 5(9)(89) (2017) 36–41. doi:10.15587/1729-4061.2017.111081
- [6] O. Vakhula, I. Opirskyy, O. Mykhaylova, Research on Security Challenges in Cloud Environments and Solutions based on the “Security-as-Code” Approach, in: *Cybersecurity Providing in Information and Telecommunication Systems II*, vol. 3550, 2023, 55–69.
- [7] V. Susukailo, I. Opirsky, O. Yaremko, Methodology of ISMS Establishment against Modern Cybersecurity Threats, in: *Lect. Notes Electr. Eng.*, Springer, Cham, 2021, 257–271. doi:10.1007/978-3-030-92435-5_15
- [8] I. Opirskyy, et al., Modern Methods of Ensuring Information Protection in Cybersecurity Systems using Artificial Intelligence and Blockchain Technology, O. Harasymchuk (Ed.), Technology Center PC, Kharkiv, 2025. doi:10.15587/978-617-8360-12-2
- [9] X. Wang, et al., ASVspoof 2019: A Large-Scale Public Database of Spoofed Speech for Speaker Verification, *Comput. Speech Lang.*, 60 (2020) 101027.
- [10] Implementing Biometrics for Large-Scale Applications: Overcoming 6 Challenges, Biostatistics.io (2023). <https://biostatistics.io/qa/implementing-biometrics-for-large-scale-applications-overcoming-6-challenges>
- [11] K. Ruda, Study of the Scalability of Biometric Authentication Systems based on Voice Embeddings, *Soc. Dev. Secur.*, 15(1) (2025) 161–170. doi:10.33445/sds.2025.15.1.15
- [12] V. Brydinskyi, et al., Comparison of Modern Deep Learning Models for Speaker Verification, *Appl. Sci.*, 14(4) (2024) 1329. doi:10.3390/app14041329
- [13] J. Thienpondt, K. Demuynck, ECAPA2: A Hybrid Neural Network Architecture and Training Strategy for Robust Speaker Embeddings, in: *IEEE Autom. Speech Recognit. Underst. Workshop (ASRU 2023)*, Taipei, Taiwan, 2023, 1–8. doi:10.1109/ASRU57964.2023.10389750
- [14] F. Deng, R. Huang, P. Jiang, L. Deng, Dense-Fusion2Net: A More Efficient and Lightweight Short Speech Speaker Recognition System with Time-Frequency Channel Attention, *Sci. Rep.*, 15 (2025) 9601. doi:10.1038/s41598-025-93873-x

- [15] R. Sharma, et al., Milestones in Speaker Recognition, *Artif. Intell. Rev.*, 57 (2024) 58. doi:10.1007/s10462-023-10688-w
- [16] G. Chen, et al., Towards Understanding and Mitigating Audio Adversarial Examples for Speaker Recognition, *IEEE Trans. Dependable Secure Comput.*, 20(5) (2023) 3970–3987. doi:10.1109/TDSC.2022.3220673
- [17] Z. Chen, S. Xu, Learning Domain-Heterogeneous Speaker Recognition Systems with Personalized Continual Federated Learning, *EURASIP J. Audio Speech Music Process.*, 33 (2023). doi:10.1186/s13636-023-00299-2
- [18] RudderAnalytics, Building a Robust Speaker Verification System for Secure Voice Authentication, *Medium* (2023). <https://medium.com/@rudderanalytics/voice-based-security-implementing-a-robust-speaker-verification-system-12c5fd98f1c1>
- [19] M. Sharif-Noughabi, S. M. Razavi, S. Mohamadzadeh, Improving the Performance of Speaker Recognition System using Optimized VGG Convolutional Neural Network and Data Augmentation, *Int. J. Eng.*, 38(10) (2025) 2414–2425. doi:10.5829/ije.2025.38.10a.17
- [20] On-Device Speech Processing Makes Alexa Faster, Lower Bandwidth, *Amazon Sci. Blog* (2023). <https://www.amazon.science/blog/on-device-speech-processing-makes-alexa-faster-lower-bandwidth>
- [21] An Overview of Speech Recognition Techniques, *Google Res.* (2023). <https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/42535.pdf>
- [22] P. Petriv, I. Oprisky, N. Mazur, Modern Technologies of Decentralized Databases, Authentication, and Authorization Methods, in: *Cybersecurity Providing in Information and Telecommunication Systems II*, vol. 3826, 2024, 60–71.
- [23] ua-polit-tiny, Hugging Face – the AI Community Building the Future (2023). <https://huggingface.co/datasets/vbrydik/ua-polit-tiny>
- [24] I. Iosifov, O. Iosifova, V. Sokolov, Sentence Segmentation from Unformatted Text using Language Modeling and Sequence Labeling Approaches, in: *7th International Scientific and Practical Conference Problems of Infocommunications. Science and Technology* (2020) 335–337. doi:10.1109/PICST51311.2020.9468084
- [25] O. Iosifova, et al., Analysis of Automatic Speech Recognition Methods, in: *Cybersecurity Providing in Information and Telecommunication Systems*, vol. 2923 (2021) 252–257.
- [26] I. Iosifov, et al., Transferability Evaluation of Speech Emotion Recognition between Different Languages, *Advances in Computer Science for Engineering and Education* 134 (2022) 413–426. doi:10.1007/978-3-031-04812-8_35
- [27] O. Romanovskiy, et al., Prototyping Methodology of End-to-End Speech Analytics Software, in: *4th Int. Workshop on Modern Machine Learning Technologies and Data Science*, vol. 3312 (2022) 76–86.
- [28] O. Romanovskiy, et al., Automated Pipeline for Training Dataset Creation from Unlabeled Audios for Automatic Speech Recognition, *Advances in Computer Science for Engineering and Education IV*, vol. 83 (2021) 25–36. doi:10.1007/978-3-030-80472-5_3
- [29] I. Iosifov, et al., Natural Language Technology to Ensure the Safety of Speech Information, in: *Cybersecurity Providing in Information and Telecommunication Systems*, vol. 3187, no. 1 (2022) 216–226.
- [30] O. Romanovskiy, et al., Accuracy Improvement of Spoken Language Identification System for Close-Related Languages, *Advances in Computer Science for Engineering and Education VII*, vol. 242 (2025) 35–52. doi:10.1007/978-3-031-84228-3_4
- [31] Defining the Core Accuracy Metrics of Biometric Systems, *Alice Biometrics* (2023). <https://alicebiometrics.com/en/defining-the-core-accuracy-metrics-of-biometric-systems>