Дослідження масштабованості систем біометричної автентифікації на основі ембеддінгів голосу

Study of the Scalability of Biometric Authentication Systems Based on Voice Embeddings

Христина Руда
аспірант, асистент кафедри захисту інформації, e-mail: khrystyna.s.ruda@lpnu.ua, ORCID: 0000-0001-8644-411X

Khrystyna Ruda
Postgraduate Student, Assistant Lecturer, e-mail: khrystyna.s.ruda@lpnu.ua, ORCID: 0000-0001-8644-411X

Національний університет "Львівська політехніка", м. Львів Україна

Lviv Polytechnic National University, Lviv, Ukraine

Received: February 21, 2025 | Revised: February 26, 2025 | Accepted: February 28, 2025

DOI: 10.33445/sds.2025.15.1.15

Мета роботи: порівнянні точності автентифікації за різної кількості користувачів для виявлення залежності між розміром бази даних та ефективністю системи.

Метод: кількісні та експериментальні методи, зокрема застосування методів машинного навчання для генерації ембеддінгів (модель TitaNet) і статистичного аналізу для оцінки масштабованості системи автентифікації.

Результати дослідження: проаналізовано вплив кількості користувачів на ефективність системи біометричної автентифікації на основі ембеддінгів голосу. Дослідження показало, що при невеликій кількості користувачів система демонструє високу стійкість і стабільність роботи. Однак зі збільшенням користувацької бази спостерігається поступове зниження точності, що свідчить про наявність обмежень масштабованості. Ця тенденція вказує на необхідність впровадження додаткових заходів для підтримання ефективності системи в умовах її розширення.

Теоретична цінність дослідження: дослідження поглиблює розуміння впливу розміру бази користувачів на ефективність систем біометричної автентифікації, а також демонструє потенціал ембеддінг-моделі **TitaNet** у цьому контексті. Це може стати основою для подальших розробок у галузі кібербезпеки стосовно методів голосової автентифікації.

Практична цінність дослідження: дослідження пропонує рекомендації для розробників, науковців і організацій, які впроваджують системи біометричної автентифікації на основі голосових ембеддінгів. Застосування оптимальних налаштувань порогових значень, тестування альтернативних моделей і збільшення обсягу вхідних даних під час реєстрації користувачів можуть суттєво підвищити ефективність і стійкість таких систем у масштабованих застосуваннях.

Цінність дослідження: використано інноваційний підхід до оцінки масштабованості системи біометричної автентифікації шляхом порівняння точності при різній кількості користувачів із застосуванням ембеддінг-моделі TitaNet та косинусної відстані. Це дозволяє виявити потенційні обмеження у продуктивності системи та сформувати рекомендації для покращення її ефективності у масштабованих застосунках.

Майбутні дослідження: розширення спектра моделей для генерації ембеддінгів та дослідження альтернативних метрик відстані, спрямованих на підвищення точності та масштабованості системи біометричної автентифікації. Це сприятиме формуванню більш комплексного розуміння впливу вибору моделі та метрики на ефективність автентифікації та вдосконаленню методів забезпечення безпеки.

Тип статті: Емпіричне дослідження.

Ключові слова: біометричні технології, Titanet, голосова автентифікація, кібербезпека, масштабованість систем.

Purpose: is to compare the authentication accuracy with varying numbers of users to identify the relationship between the database size and the system's efficiency.

Method: Quantitative and experimental methods were used, including the application of machine learning techniques for generating embeddings (using the TitaNet model) and statistical analysis to assess the scalability of the authentication system.

Findings: The study analyzed the impact of the number of users on the effectiveness of the biometric authentication system based on voice embeddings. The results demonstrated that the system maintains high stability and consistent performance with a small user base. However, as the number of users increases, a gradual decline in accuracy is observed, indicating scalability limitations. This trend highlights the need for additional measures to maintain system effectiveness under expanded usage conditions.

Theoretical implications: The study deepens the understanding of the impact of the user database size on the effectiveness of biometric authentication systems and demonstrates the potential of the TitaNet embedding model in this context. These findings may serve as a foundation for further developments in the field of cybersecurity related to voice authentication methods.

Practical implications: The study provides recommendations for developers, researchers, and organizations implementing biometric authentication systems based on voice embeddings. Applying optimal threshold settings, testing alternative models, and increasing the amount of input data during user registration can significantly enhance the effectiveness and robustness of such systems in scalable applications.

Value: An innovative approach was applied to assess the scalability of the biometric authentication system by comparing accuracy across different numbers of users using the TitaNet embedding model and cosine distance. This approach enables the identification of potential performance limitations and the formulation of recommendations to improve the system's efficiency in large-scale applications.

Future research: Future studies are planned to expand the range of embedding models and explore alternative distance metrics aimed at enhancing the accuracy and scalability of the biometric authentication system. This will contribute to a more comprehensive understanding of how the choice of model and metric affects authentication effectiveness and improve methods for ensuring system security.

Paper type: Empirical study.

Key words: biometric technologies, Titanet, voice authentication, cybersecurity, scalability.

Вступ

У сучасних умовах стрімкого розвитку цифрових технологій питання безпечної та зручної ідентифікації користувачів стають дедалі актуальнішими. Серед різних методів біометричної автентифікації голосова автентифікація привертає особливу увагу завдяки своїй безконтактності, інтуїтивній зрозумілості та можливості застосування у широкому спектрі сфер — від мобільних додатків і фінансових сервісів до систем “розумного дому” та голосових асистентів [1]. Її популярність зумовлена не лише зручністю для кінцевого користувача, а й здатністю забезпечувати додатковий рівень безпеки у багатофакторних системах автентифікації. З огляду на зростання обсягів дистанційних послуг, активне використання голосових інтерфейсів і підвищення вимог до захисту персональних даних, дослідження у сфері голосової автентифікації є надзвичайно актуальними.

Суттєвий прогрес у розвитку голосових біометричних систем став можливим завдяки впровадженню ембеддінг-підходів, які дозволяють ефективно перетворювати голосові сигнали у векторні представлення. Ці методи забезпечують компактне та інформативне подання аудіоданих, що зберігає ключові характеристики мовця й дає змогу виконувати швидке й точне порівняння голосових зразків [2]. Ембеддінг-підхід вважається перспективним завдяки своїй здатності узагальнювати інформацію про голосові особливості користувача та стійкості до різних варіацій мовлення, фонового шуму чи зміни мовного контексту. Це відкриває широкі можливості для застосування таких моделей у реальних умовах, де якість даних може суттєво варіюватися [3].

Попри досягнення у точності й стабільності голосових біометричних систем, одним із ключових викликів залишається масштабованість — здатність системи ефективно функціонувати при збільшенні кількості користувачів. У практичних сценаріях, особливо у великих організаціях чи глобальних платформах, кількість зареєстрованих користувачів може сягати сотень тисяч або мільйонів, що суттєво ускладнює процес автентифікації. Недостатня масштабованість може призводити до зниження точності, збільшення часу обробки запитів і підвищення ймовірності помилкових спрацьовувань, що негативно позначається як на безпеці, так і на зручності користування системою.

Таким чином, оцінка масштабованості систем голосової автентифікації є критично важливим аспектом для їхнього ефективного впровадження у масштабних застосуваннях. Дослідження у цьому напрямі дозволяють не лише виявити можливі обмеження існуючих рішень, а й сприяють розробці стратегій оптимізації, що забезпечують стабільну роботу системи за умов збільшення кількості користувачів. Це має важливе значення для практичного використання голосових біометричних систем у сферах, де надійність і швидкодія автентифікації є визначальними факторами.

Теоретичні основи дослідження

Дослідження масштабованості систем голосової автентифікації ґрунтується на сучасних теоретичних засадах біометричної ідентифікації, методах представлення мовних сигналів у векторному просторі та алгоритмах порівняння ознак. Голосова автентифікація, як окремий напрям біометричної безпеки, базується на використанні індивідуальних фізіологічних та поведінкових особливостей голосового тракту, що дозволяє ідентифікувати або верифікувати особу за мовним сигналом [4]. У контексті цього дослідження розглядається задача верифікації — процесу встановлення належності тестового зразка конкретному користувачу, що має критичне значення для систем доступу з підвищеними вимогами до безпеки.

Фундаментальною складовою сучасних систем голосової автентифікації є методи екстракції ознак, які дозволяють перетворити часово-частотні характеристики мовного сигналу у компактні векторні представлення — ембеддінги. Ембеддінг моделює унікальні

голосові параметри користувача, водночас забезпечуючи стійкість до змін мовного контексту, інтонаційних варіацій, темпу мовлення та впливу фонових шумів. Використання таких векторних репрезентацій знижує розмірність вхідних даних, спрощує обчислювальні процедури та підвищує ефективність системи у задачах порівняння голосових зразків, особливо у масштабних базах користувачів [5].

Процес формування ембеддінгів включає кілька етапів: первинну обробку аудіосигналу, вилучення спектральних ознак (наприклад, Mel-частотних кепстральних коефіцієнтів або спектрограм) та їх подання на вхід нейронної мережі, яка виконує проекцію вхідних характеристик у нижчий за розмірністю простір ембеддінгів. Такі представлення мають властивість зберігати релевантну інформацію про мовця, при цьому ігноруючи сторонні фактори, що не впливають на ідентифікацію особи. У системах верифікації ембеддінги, отримані на етапі реєстрації, використовуються для побудови еталонних профілів користувачів шляхом усереднення кількох ембеддінгів, що дозволяє підвищити їх стійкість та зменшити вплив випадкових флуктуацій у голосі [6].

TitaNet — це нейронна мережа, розроблена для вилучення представлень мовців (ембеддингів) із мовних сигналів із метою покращення задач ідентифікації та верифікації мовців. Архітектура моделі базується на використанні одно-вимірних глибинних сепарабельних згорток (1D depth-wise separable convolutions), що суттєво зменшує кількість параметрів і обчислювальну складність порівняно з класичними згортковими мережами. Крім того, TitaNet інтегрує механізми глобальної контекстної обробки через Squeeze-and-Excitation (SE) модулі, що дозволяють моделі акцентувати увагу на найбільш інформативних частинах мовного сигналу.

Однією з ключових особливостей TitaNet є її масштабованість: архітектура підтримує різні конфігурації (наприклад, TitaNet-S та TitaNet-M), що дозволяє обирати оптимальний баланс між точністю й обчислювальними ресурсами залежно від конкретного завдання. Модель перетворює мовні фрагменти змінної довжини у фіксовані вектори (t-вектори), які можна ефективно використовувати для класифікації мовців або порівняння схожості між записами.

TitaNet демонструє високу точність навіть при використанні меншої кількості параметрів порівняно з конкурентними моделями (наприклад, ECAPA-TDNN), що робить її придатною як для серверних систем, так і для пристроїв із обмеженими обчислювальними можливостями. Завдяки поєднанню ефективних згорткових шарів і контекстних механізмів модель забезпечує стійкість до фонових шумів і варіацій мовлення, що особливо важливо для практичного застосування в реальних умовах.

Оцінка схожості між тестовим та еталонним ембеддінгами є ключовим етапом верифікації. У цьому дослідженні для порівняння використовувалася косинусна відстань, яка є метрикою, що визначає кутову відмінність між двома векторами у багатовимірному просторі ознак. Перевагою косинусної відстані є її нечутливість до різниці в масштабі векторів, що дозволяє фокусуватися на формі спектральних характеристик голосу.[8] Прийняття рішення про належність двох зразків одному користувачу базується на порівнянні обчисленого значення косинусної відстані з наперед визначеним пороговим значенням. Визначення оптимального порогу є важливим для мінімізації помилок першого роду (False Acceptance Rate, FAR) — випадків несанкціонованого доступу, та помилок другого роду (False Rejection Rate, FRR) — необґрунтованих відмов у доступі для зареєстрованих користувачів [9].

Одним із основних викликів у побудові ефективних систем голосової автентифікації є забезпечення їх масштабованості — здатності системи зберігати високу ефективність при збільшенні кількості зареєстрованих користувачів. У великих системах масштабованість стає критично важливою характеристикою, оскільки розширення користувацької бази ускладнює підтримання стабільних показників точності, оскільки зі збільшенням кількості профілів зростає ймовірність збігу ознак між різними користувачами, що підвищує ризик помилкових верифікацій.

Постановка проблеми

Основною проблемою, розглянутою в цій статті, є недостатнє розуміння впливу кількості користувачів на ефективність систем голосової автентифікації, які знаходять широке застосування у різних сферах, зокрема у фінансових послугах, телекомунікаціях та інтелектуальних пристроях. Хоча такі системи демонструють високу точність у невеликих користувацьких базах, їх продуктивність може суттєво погіршуватися зі збільшенням кількості користувачів, що призводить до зниження точності автентифікації, підвищення ймовірності помилкових допусків та відмов, а також збільшення часу обробки запитів.

Ця проблема стає особливо актуальною в умовах стрімкого зростання обсягів даних і масштабування біометричних систем до сотень тисяч або мільйонів користувачів. Ігнорування аспекту масштабованості може призвести до серйозних експлуатаційних труднощів, особливо у критично важливих застосуваннях, де помилки автентифікації можуть мати значні наслідки для безпеки й довіри користувачів. Попри важливість цього питання, у науковій літературі спостерігається недостатня увага до систематичного вивчення взаємозв'язку між розміром користувацької бази та ефективністю голосових біометричних систем.

Розв'язання цієї проблеми вимагає комплексного дослідження, що дозволить оцінити, як масштабування впливає на продуктивність автентифікації та сформулювати рекомендації для забезпечення стабільної роботи систем у реальних умовах експлуатації.

Методологія дослідження

Для оцінки впливу кількості користувачів на ефективність систем голосової автентифікації було побудовано п'ять окремих систем із різним розміром користувацьких баз. Кожна система проходила тестування за участі як зареєстрованих, так і незареєстрованих користувачів. На основі результатів спроб доступу здійснювалося оцінювання ефективності, що включало аналіз показників точності автентифікації та порівняння продуктивності між системами. Отримані результати дозволили виявити залежність між масштабом користувацької бази та ефективністю роботи системи.

Результати

Архітектура системи

Розроблена система голосової автентифікації має дворівневу архітектуру, що складається з двох основних модулів: модуля реєстрації користувачів та модуля верифікації. Така структурна організація забезпечує ефективне формування еталонних голосових зразків і подальшу перевірку належності тестових аудіозаписів певному користувачу.

Модуль реєстрації користувачів призначений для створення еталонних голосових профілів користувачів, які використовуються під час процесу верифікації. Архітектура цього модуля представлена на рисунку 1 та включає такі основні етапи:

1. Запис аудіозразків:

Користувач здійснює реєстрацію за допомогою мікрофона, через який записуються 10 голосових зразків. Кожен запис триває фіксований проміжок часу, що забезпечує узгодженість у подальшій обробці даних.

2. Екстракція ембеддінгів:

Оброблені аудіозразки надходять до нейронної мережі, що виконує функцію екстрактора ембеддінгів. На цьому етапі для кожного голосового фрагмента генерується ембеддінг — векторне представлення, яке містить ключові голосові характеристики мовця. Ембеддінги відображають унікальні особливості голосу, зберігаючи при цьому компактність і придатність для ефективного порівняння.

3. Формування еталонного ембеддінга:

Згенеровані ембеддінги для 10 записів користувача піддаються процесу усереднення, що дозволяє мінімізувати вплив можливих варіацій мовлення (інтонації, темпу, фонових шумів) і сформувати стабільний еталонний ембеддінг. Цей усереднений вектор представляє голосовий профіль користувача з високою узагальнюючою здатністю.

4. Збереження у базі даних:

Сформований еталонний ембеддінг зберігається у спеціалізованій базі даних, де він асоціюється з ідентифікатором користувача. База даних є оптимізованою для швидкого пошуку та забезпечує ефективний доступ під час верифікації.

Рисунок 1. – Схема модуля реєстрації системи біометричної автентифікації

Модуль верифікації користувачів призначений для визначення належності тестового голосового зразка до конкретного зареєстрованого користувача. Процес верифікації базується на порівнянні ембеддінга тестового зразка з еталонним ембеддінгом, збереженим у базі даних, із використанням косинусної відстані як метрики схожості. Архітектура модуля наведена на рисунку 2.

Етапи роботи модуля верифікації:

1. Отримання тестового зразка

Процес верифікації розпочинається з подання користувачем тестового голосового зразка, який записується за допомогою мікрофона. Записаний аудіофайл передається на подальшу обробку для видалення фонових шумів, нормалізації гучності та видалення нерелевантних ділянок сигналу.

2. Екстракція ембеддінга

Попередньо оброблений тестовий зразок надходить до нейронної мережі для екстракції ембеддінгів, яка генерує компактне векторне представлення голосових характеристик. Отриманий ембеддінг тестового зразка описує унікальні ознаки мовця, зберігаючи їх у зручній для обчислень формі.

3. Обчислення косинусної відстані

Згенерований ембеддінг порівнюється з еталонним ембеддінгом користувача, збереженим у базі даних. На цьому етапі обчислюється косинусна відстань між двома векторами, що відображає ступінь їхньої схожості. Чим менше значення косинусної відстані, тим більш подібними є два ембеддінги, що свідчить про вищу ймовірність належності зразків одному користувачу.

4. Прийняття рішення

Обчислене значення косинусної відстані порівнюється із заздалегідь встановленим пороговим значенням:

Якщо косинусна відстань менша за порогове значення, система приймає рішення, що користувача верифіковано (тестовий зразок належить тому самому мовцю, що й еталонний зразок).

Якщо косинусна відстань перевищує порогове значення, система визначає, що користувача не верифіковано (зразки належать різним мовцям).

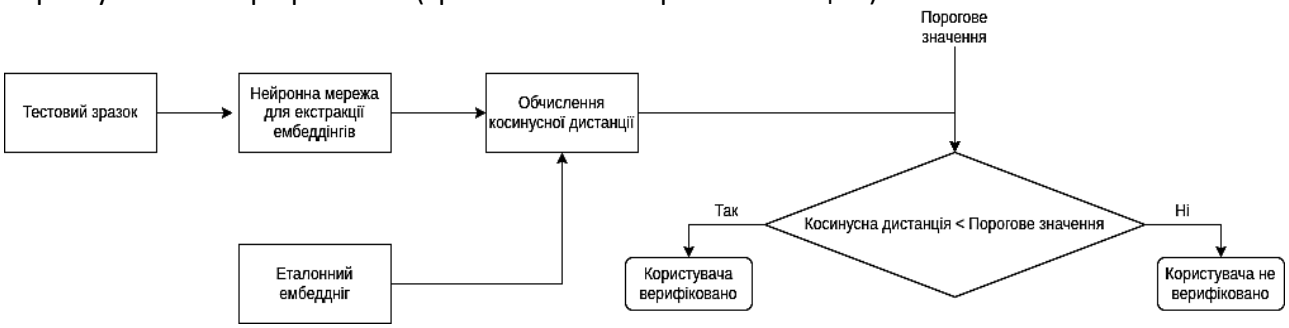

Рисунок 2. –Схема модуля верифікації системи біометричної автентифікації

Підготовка даних

Для проведення дослідження було використано датасет із загальною кількістю 70 мовців, сформований із двох джерел. Першу частину даних, що сформувала авторський датасет, було зібрано вручну з відкритих джерел і вона включала 20 англомовних спікерів, серед яких було 10 чоловіків і 10 жінок. Ця частина відзначалася різноманітністю акцентів і варіативною якістю аудіозаписів, що дозволило врахувати різні умови реального використання системи. Друга частина даних складалася з 50 спікерів із відкритого набору даних [11], що забезпечило ширше охоплення голосових варіацій.

Для кожного спікера було відібрано 100 секунд аудіоматеріалу для первинного навчання системи. Цей матеріал було розбито на 10-секундні сегменти для формування ембеддінгів. Окремо для тестування системи було підготовлено 20 незалежних 10-секундних фрагментів на кожного мовця, що дало змогу оцінити стабільність та узагальнюючу здатність моделей.

З метою дослідження впливу кількості користувачів на ефективність системи голосової автентифікації було сформовано п’ять підмножин (сабсетів) із різною кількістю мовців: 8, 16, 20, 50 та 70. Це дозволило провести поетапне оцінювання продуктивності системи під час збільшення розміру користувацької бази та проаналізувати її масштабованість. Структура сабсетів представлена в таблиці 1.

Таблиця 1 – Структура сабсетів

Кількість класів	Розподіл мовців	Навчальні дані (сек)	Тестові дані (сек)
8	8 з авторського датасету	800	1600
16	16 з авторського датасету	1600	3200
20	20 з авторського датасету	2000	4000
50	20 з авторського датасету і 30 з відкритого датасету	5000	10000
70	20 з авторського датасету і 50 з відкритого датасету	7000	14000

Побудова і калібрація систем

Для дослідження масштабованості було побудовано п’ять систем біометричної автентифікації, кожна з яких відповідала одному з датасетів із різною кількістю зареєстрованих користувачів: 8, 16, 20, 50 та 70. Це дозволило оцінити, як збільшення розміру користувацької

бази впливає на ефективність роботи системи. Процес побудови кожної системи розпочинався з формування еталонних ембеддінгів для кожного користувача. Для цього використовувалися тренувальні датасети, що складалися з десяти аудіозаписів тривалістю по десять секунд. Кожен аудіозразок пропускався через нейронну мережу TitaNet, яка виконувала екстракцію ембеддінгів — векторних представлень, що містять ключові голосові характеристики користувача. Для кожного мовця було отримано десять ембеддінгів, які згодом усереднювалися, щоб зменшити вплив можливої варіативності у вимові, шумових перешкод і коливань інтонації. Усереднений вектор формував еталонний ембеддінг користувача, який зберігався у базі даних і використовувався на етапі верифікації.

Після завершення формування еталонних ембеддінгів проводилося калібрування систем, спрямоване на визначення оптимального порогового значення косинусної відстані, що використовується для прийняття рішень про належність тестового зразка певному користувачу. Коректне встановлення порогу є критичним для досягнення балансу між двома видами помилок: помилками першого роду (верифікація невалідного користувача) та помилками другого роду (невірна відмова у верифікації валідного користувача). Занижене порогове значення може призвести до надмірної кількості відмов для справжніх користувачів, тоді як завищене збільшує ризик несанкціонованого доступу.

Для пошуку оптимального порогу було змодельовано процес автентифікації, який охоплював два сценарії: перевірку валідних користувачів, де тестові зразки порівнювалися з еталонами тих самих осіб, та перевірку невалідних користувачів, у якій тестові ембеддінги зіставлялися з еталонними векторами інших користувачів. Аналіз результатів дозволив визначити таке порогове значення, при якому досягалося найкраще співвідношення помилок першого і другого роду. Зі збільшенням кількості зареєстрованих користувачів спостерігалася тенденція до зниження порогу, що пояснюється необхідністю підвищення суворості критеріїв верифікації у масштабованих системах, аби уникнути зростання кількості хибних спрацьовувань.

Отримані результати калібрування стали основою для подальшого порівняльного аналізу ефективності п’яти побудованих систем і дозволили зробити висновки щодо впливу масштабування на їхню продуктивність та надійність.

Тестування і оцінювання результатів

У результаті проведеного дослідження було отримано показники ефективності системи голосової автентифікації для п’яти різних конфігурацій із кількістю користувачів: 8, 16, 20, 50 та 70. Оцінювання ефективності базувалося на співвідношенні помилок першого та другого роду, а також на загальній точності верифікації при оптимально налаштованому пороговому значенні. Отримані результати наведені в таблиці 2

Таблиця 2 – Результати експериментів із масштабування

Кількість користувачів	FAR, %	FRR, %	EER, %	Точність верифікації, %
8	2.86	15.42	9.14	88.36
16	3.73	8.02	5.88	93.92
20	4.85	6.13	5.49	94.50
50	14.50	3.90	9.20	88.56
70	11.01	2.85	6.93	90.83

Аналіз отриманих результатів свідчить про наявність залежності точності верифікації від кількості користувачів. Максимальне значення точності (94.50%) спостерігається при 20 користувачах, що вказує на оптимальне співвідношення між чисельністю вибірки та

можливостями системи коректно розпізнавати користувачів. Схожий рівень точності (93.92%) зафіксовано для 16 користувачів, що підтверджує стабільність системи у цьому діапазоні.

Зі збільшенням кількості користувачів до 50 і 70 спостерігається зниження точності до 88.56% та 90.83% відповідно. Така динаміка може бути зумовлена підвищенням складності розмежування ознак серед більшої кількості користувачів, що призводить до зростання кількості хибних рішень системи. Водночас низька точність при 8 користувачах (88.36%) може свідчити про нестачу навчальних даних, що обмежує здатність системи до якісної генералізації.

Таким чином, результати демонструють, що найефективніше функціонування системи верифікації досягається при кількості користувачів у межах від 16 до 20, де забезпечується високий рівень точності за мінімального ризику помилкової класифікації. Хоча рівень точності залишається порівняно високим, тенденція до погіршення ефективності при масштабуванні сигналізує про потребу у додаткових заходах для підтримання стабільної роботи системи у великих користувацьких базах.

Рекомендації для підвищення стійкості системи:

1. Оптимізація порогового значення. Для зменшення кількості хибних допусків при великій кількості користувачів доцільно налаштувати порогове значення косинусної відстані з акцентом на підвищення показника відсікання нелегальних користувачів. Хоча це може незначно збільшити кількість відмов для валідних користувачів, такий підхід є виправданим у системах із підвищеними вимогами до безпеки.

2. Випробування альтернативних моделей. Рекомендується протестувати інші нейронні мережі для генерації ембеддінгів. Деякі моделі можуть забезпечити кращу розмежувальну здатність між голосовими профілями, що дозволить зменшити ймовірність помилкових збігів при збільшенні користувацької бази.

3.Збільшення обсягу вхідного аудіо під час реєстрації. Розширення тривалості або кількості аудіозаписів на етапі формування еталонних ембеддінгів дозволить підвищити їх репрезентативність. Це допоможе краще врахувати індивідуальні варіації голосу та зменшити ризик помилок при верифікації, особливо у великих системах.

4. Інтеграція багаторівневої верифікації. Запровадження багаторівневого підходу з додатковими перевірками для випадків, коли косинусна відстань близька до порогового значення, може сприяти зниженню кількості помилкових рішень без суттєвого погіршення загальної продуктивності системи.

Підсумовуючи, результати дослідження показали, що система голосової автентифікації досягає найвищої ефективності при кількості користувачів від 16 до 20, проте масштабування до 50 користувачів спричиняє спад точності. Хоча зниження не є критичним, воно вказує на потенційні обмеження масштабованості та потребу у впровадженні заходів для збереження високого рівня ефективності в умовах зростання кількості користувачів. Реалізація запропонованих рекомендацій дозволить підвищити стійкість системи та забезпечити її надійну роботу у масштабних застосуваннях.

Висновки

У результаті проведеного дослідження було розроблено та проаналізовано систему голосової автентифікації з дворівневою архітектурою, що включає модулі реєстрації та верифікації користувачів. Реалізована система забезпечує ефективне формування еталонних голосових профілів та їх подальше порівняння із тестовими зразками на основі косинусної відстані.

Дослідження масштабованості системи, проведене на п’яти конфігураціях із кількістю користувачів від 8 до 70, показало, що оптимальна ефективність досягається при кількості користувачів у діапазоні від 16 до 20, де точність верифікації становить понад 94 %. Це зумовлено достатнім обсягом навчальних даних для формування стійких еталонних

ембеддінгів. Проте подальше збільшення кількості користувачів до 50 і 70 спричиняє спад точності до 88,6 %, що вказує на ускладнення у розмежуванні голосових профілів та необхідність додаткових заходів для підтримання стабільності системи.

Запропоновані рекомендації для покращення масштабованості системи включають оптимізацію порогового значення косинусної відстані, випробування альтернативних моделей нейронних мереж, збільшення обсягу вхідних аудіоданих під час реєстрації та впровадження багаторівневої верифікації. Реалізація цих заходів дозволить покращити стійкість системи до зростання користувацької бази та забезпечити її надійну роботу у великих масштабах застосування.

Отримані результати демонструють перспективність розробленої системи для практичного використання у сфері біометричної безпеки та свідчать про доцільність подальшого удосконалення з метою підвищення її ефективності та масштабованості.

Фінансування

Це дослідження не отримало конкретної фінансової підтримки.

Конкуруючі інтереси

Автори заявляють, що у них немає конкуруючих інтересів.

Список використаних джерел

1. Fortune Business Insights. (2024). Voice biometric solutions market size, share & industry analysis, by component, application, end-user, and regional forecast, 2024–2032. Available from : https://www.fortunebusinessinsights.com/industry-reports/voice-biometric-solutions-market-100509?utm_source=chatgpt.com
2. De Prisco, R., Fusco, C., Malandrino, D., & Zaccagnino, R. (2023). Text-independent voice recognition based on Siamese networks and fusion embeddings. In Proceedings of ITASEC 2023: The Italian Conference on CyberSecurity (May 03–05, 2023, Bari, Italy). CEUR Workshop Proceedings, 3488.
3. Quang, C. T., Nguyen, Q. M., Phuong, P. N., & Do, Q. T. (2021). Improving speaker verification in noisy environment using DNN classifier. In 2021 RIVF International Conference on Computing and Communication Technologies (RIVF) (pp. 1–5). IEEE. https://doi.org/10.1109/RIVF51545.2021.9642074
4. Jain, A. K., Hong, L., & Pankanti, S. (2000). Biometric identification. Communications of the ACM, 43(2), 91–98. https://doi.org/10.1145/328236.328110
5. Guo, C., & Berkhahn, F. (2016). Entity Embeddings of Categorical Variables. ArXiv, abs/1604.06737.
6. Zaiets, I., Brydinskyi, V., Sabodashko, D., Khoma, Y., & Ruda, K. (2024). Integrated system for speaker diarization and intruder detection using speaker embeddings. In CEUR Workshop Proceedings, 3654: Cybersecurity providing in information and telecommunication systems 2024 (pp. 228–238). CEUR-WS.
7. Koluguri, N. R., Park, T., & Ginsburg, B. (2022). TitaNet: Neural model for speaker representation with 1D depth-wise separable convolutions and global context. In ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 8102–8106). IEEE. https://doi.org/10.1109/ICASSP43922.2022.9746806
8. Steck, H., Ekanadham, C., & Kallus, N. (2024). Is cosine-similarity of embeddings really about similarity? In Companion Proceedings of the ACM Web Conference 2024 (WWW '24) (pp. 887–890). Association for Computing Machinery. https://doi.org/10.1145/3589335.3651526

9. Biometric Update. (2024). How are biometric systems evaluated? Available from : https://www.biometricupdate.com/202405/how-are-biometric-systems-evaluated
10. Vjcalling. (2019). Speaker Recognition Audio Dataset [Data set]. Kaggle. Available from : https://www.kaggle.com/datasets/vjcalling/speaker-recognition-audio-dataset

References

1. Fortune Business Insights. (2024). Voice biometric solutions market size, share & industry analysis, by component, application, end-user, and regional forecast, 2024–2032. Available from : https://www.fortunebusinessinsights.com/industry-reports/voice-biometric-solutions-market-100509?utm_source=chatgpt.com
2. De Prisco, R., Fusco, C., Malandrino, D., & Zaccagnino, R. (2023). Text-independent voice recognition based on Siamese networks and fusion embeddings. In Proceedings of ITASEC 2023: The Italian Conference on CyberSecurity (May 03–05, 2023, Bari, Italy). CEUR Workshop Proceedings, 3488.
3. Quang, C. T., Nguyen, Q. M., Phuong, P. N., & Do, Q. T. (2021). Improving speaker verification in noisy environment using DNN classifier. In 2021 RIVF International Conference on Computing and Communication Technologies (RIVF) (pp. 1–5). IEEE. https://doi.org/10.1109/RIVF51545.2021.9642074
4. Jain, A. K., Hong, L., & Pankanti, S. (2000). Biometric identification. Communications of the ACM, 43(2), 91–98. https://doi.org/10.1145/328236.328110
5. Guo, C., & Berkhahn, F. (2016). Entity Embeddings of Categorical Variables. ArXiv, abs/1604.06737.
6. Zaiets, I., Brydinskyi, V., Sabodashko, D., Khoma, Y., & Ruda, K. (2024). Integrated system for speaker diarization and intruder detection using speaker embeddings. In CEUR Workshop Proceedings, 3654: Cybersecurity providing in information and telecommunication systems 2024 (pp. 228–238). CEUR-WS.
7. Koluguri, N. R., Park, T., & Ginsburg, B. (2022). TitaNet: Neural model for speaker representation with 1D depth-wise separable convolutions and global context. In ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 8102–8106). IEEE. https://doi.org/10.1109/ICASSP43922.2022.9746806
8. Steck, H., Ekanadham, C., & Kallus, N. (2024). Is cosine-similarity of embeddings really about similarity? In Companion Proceedings of the ACM Web Conference 2024 (WWW '24) (pp. 887–890). Association for Computing Machinery. https://doi.org/10.1145/3589335.3651526
9. Biometric Update. (2024). How are biometric systems evaluated? Available from : https://www.biometricupdate.com/202405/how-are-biometric-systems-evaluated
10. Vjcalling. (2019). Speaker Recognition Audio Dataset [Data set]. Kaggle. Available from : https://www.kaggle.com/datasets/vjcalling/speaker-recognition-audio-dataset