

Вдосконалення методу виявлення deepfake аудіо на основі аналізу аномалій спектральних характеристик сигналу

Вінницький національний технічний університет

Анотація. Робота присвячена вдосконаленню методу виявлення синтетично згенерованих аудіозаписів (deepfake) шляхом аналізу аномалій у спектральних характеристиках сигналу. У роботі проведено огляд сучасних технологій синтезу мовлення, проаналізовано переваги та недоліки існуючих методів детекції, а також розроблено новий алгоритм, що поєднує спектральний аналіз із методами машинного навчання. Практичною реалізацією є програмний додаток на Python, який дозволяє автоматично аналізувати аудіофайли, візуалізувати їх спектрограми та формувати звіт про можливу наявність ознак deepfake. Результати тестування підтверджують високу ефективність алгоритму. Розробка має практичне значення для систем кібербезпеки, аудіофоре́нзики та перевірки автентичності звукових даних.

Ключові слова: deepfake, аудіо, спектральний аналіз, виявлення аномалій, машинне навчання, MFCC, кібербезпека.

Abstract. Thesis is devoted to improving the method of detecting synthetically generated audio recordings (deepfake) by analyzing anomalies in the spectral characteristics of the signal. The work includes a review of modern speech synthesis technologies, an analysis of the advantages and disadvantages of existing detection methods, and the development of a new algorithm that combines spectral analysis with machine learning methods. The practical implementation is a Python-based software application that allows automatic analysis of audio files, visualization of their spectrograms, and generation of reports on possible signs of deepfake. Testing results confirm the high efficiency of the algorithm. The development has practical value for cybersecurity systems, audio forensics, and authenticity verification of audio data.

Keywords: deepfake, audio, spectral analysis, anomaly detection, machine learning, MFCC, cybersecurity.

Вступ

Сучасний розвиток технологій штучного інтелекту та глибинного навчання призвів до появи нових інструментів генерації мультимедійного контенту, серед яких особливе місце займають deepfake-технології. Спочатку ці технології позиціонувалися як засіб для створення реалістичних аудіо- та відеоефектів у розважальній сфері, проте згодом вони набули широкого поширення у сфері інформаційних атак, шахрайства та маніпуляцій громадською думкою. Зокрема, deepfake аудіо може використовуватися для імітації голосу конкретної особи, що створює серйозні ризики для безпеки та довіри у цифровому середовищі.

Зростання кількості підроблених аудіозаписів актуалізує проблему їх виявлення та протидії. Традиційні методи детекції, що базуються на статистичних характеристиках сигналу, поступово втрачають ефективність у зв'язку з удосконаленням алгоритмів синтезу мовлення. У цьому контексті особливого значення набувають підходи, що аналізують спектральні характеристики мовних сигналів, оскільки саме в спектральному просторі зберігаються відмінності між натуральними та синтетичними голосами, які не завжди помітні для слухача.

Для досягнення поставленої мети у дипломній роботі передбачено виконання таких завдань:

- провести огляд існуючих методів виявлення deepfake аудіо та визначити їхні переваги й недоліки;
- дослідити спектральні характеристики реальних та синтетичних мовних сигналів;
- розробити алгоритм аналізу аномалій у спектральному просторі;
- реалізувати програмний прототип системи виявлення deepfake аудіо;

– провести експериментальні дослідження ефективності запропонованого підходу та порівняти його з відомими методами.

Наукова новизна роботи полягає у вдосконаленні методу виявлення deepfake аудіо за рахунок використання спектрального аналізу сигналів та алгоритмів машинного навчання, що дозволяє підвищити точність визначення підробленого мовлення в умовах наявності шумів і спотворень.

Результати дослідження

Розробка алгоритму виявлення deepfake-аудіо на основі спектрального аналізу передбачає створення загальної моделі, яка відображатиме послідовність обробки сигналу та логіку прийняття рішення. Така модель дозволяє формалізувати процес виявлення аномалій, визначити ключові етапи аналізу та встановити взаємозв'язки між ними.

Загальна модель алгоритму виявлення deepfake-аудіо складається з кількох послідовних етапів обробки. На першому етапі здійснюється збір вхідних даних, що включає аудіозаписи природного мовлення та потенційно синтетично згенеровані записи. На цьому кроці особлива увага приділяється забезпеченню сумісності формату аудіо та підтримці різних джерел запису, що дозволяє алгоритму працювати з різноманітними умовами отримання сигналу.

На наступному етапі проводиться попередня обробка сигналу, яка передбачає нормалізацію амплітуди, усунення фонових шумів та приведення аудіофайлів до єдиного формату. Для підвищення ефективності спектрального аналізу сигнал поділяється на короткі часові вікна, що дозволяє враховувати нестационарну природу мовного сигналу та підвищує точність виділення часово-частотних характеристик.

Цей підхід забезпечує підготовку даних для подальшого використання методів часово-частотного аналізу, виділення інформативних ознак та застосування алгоритмів машинного навчання для класифікації сигналів як справжніх або синтетичних.

Наступним етапом обробки є перетворення аудіосигналу у спектральну область, що дозволяє отримати часово-частотне представлення сигналу. Найбільш поширеним підходом є короткочасне перетворення Фур'є (STFT), яке забезпечує поділ сигналу на короткі сегменти і визначає спектр кожного з них. Це дозволяє відслідковувати зміну частотних компонентів у часі та виявляти аномалії, характерні для синтетичних записів.

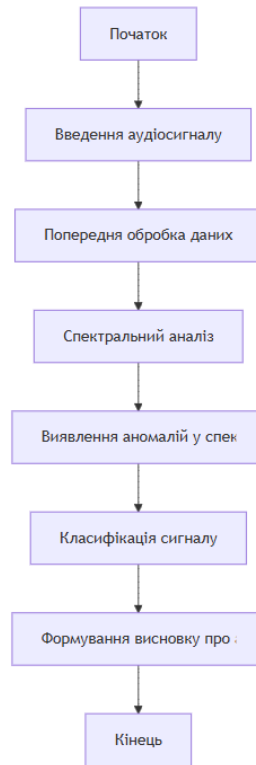
Для більш детального відображення особливостей людського слуху застосовується мел-шкала, на основі якої обчислюються мел-частотні мел-частотні коефіцієнти (MFCC). Ці ознаки широко використовуються у задачах розпізнавання мовлення та класифікації голосу, оскільки вони відображають акустичні характеристики, що найбільш сприйнятні людським вухом, і дозволяють ефективно відрізнити природний голос від синтетичного.

Додатково для аналізу локальних особливостей сигналу можна застосовувати вейвлет-перетворення, яке забезпечує змінну часово-частотну роздільну здатність та дозволяє виявляти артефакти, що можуть залишатися непоміченими при класичному спектральному аналізі. Завдяки такому підходу формується комплексний набір спектральних ознак, які надалі використовуються для класифікації сигналів як реальних або синтетичних.

На наступному етапі здійснюється аналіз аномалій у спектральних характеристиках аудіосигналу. Природні голосові записи характеризуються властивою людині спектральною нерівномірністю, мікроеваріаціями тембру, а також варіативністю формантів, що обумовлено фізіологічними особливостями голосового апарату. У свою чергу, синтетичне deepfake-аудіо часто демонструє надмірну регулярність, штучну гладкість спектра або неприродні відхилення у високочастотних ділянках.

Завдяки виявленим структурним і статистичним відмінностям стає можливим класифікувати сигнал, оцінюючи наявність аномалій, що характерні для генеративних моделей. Такий підхід дозволяє алгоритму ефективно відрізнити реальні голоси від синтетичних, навіть у випадках, коли зовнішні ознаки сигналу (тембр, гучність, інтонація) залишаються подібними до природних.

Заключним етапом є класифікація сигналу, яка може здійснюватися за допомогою машинного навчання або евристичних правил. У випадку використання моделей штучного інтелекту (наприклад, нейронних мереж), класифікатор навчається на попередньо зібраних даних, відрізняючи справжні записи від підробок. Результатом роботи алгоритму є рішення про достовірність вхідного аудіофрагмента.



Загальна модель алгоритму виявлення deepfake-аудіо

Узагальнена блок-схема алгоритму включає такі основні блоки:

1. Введення аудіосигналу.
2. Попередня обробка даних.
3. Спектральний аналіз (отримання спектрограми).
4. Виявлення аномалій у спектральних характеристиках.
5. Класифікація сигналу.
6. Формування висновку про автентичність.

Виходячи з вищеперерахованого, побудована модель дозволяє системно описати процес виявлення deepfake-аудіо та створює основу для подальшої програмної реалізації алгоритму.

Висновки

Проведене дослідження було спрямоване на вдосконалення методів виявлення підроблених аудіозаписів (deepfake) через комплексний і системний аналіз спектральних характеристик аудіосигналу. В процесі роботи було детально вивчено сучасні технології генерації синтетичного голосу, включаючи алгоритми текст-у-мову (TTS), нейромережеві генератори та методи стилізації голосу, які використовуються для створення високоякісних deepfake-записів. Крім того, було проаналізовано різноманітні механізми формування підробок, що дозволяє зрозуміти, які артефакти та аномалії виникають у спектральному та часово-частотному представленні сигналу. Значна увага приділялася огляду існуючих методів виявлення синтетичних голосових записів, як класичних статистичних підходів, так і сучасних алгоритмів машинного та глибинного навчання. Особливу увагу було зосереджено на оцінці ефективності різних методів, визначенні найбільш інформативних спектральних і статистичних ознак та розробці критеріїв, що дозволяють достовірно розрізняти природні та синтетичні голосові записи, забезпечуючи при цьому високу точність і стабільність роботи алгоритму в різних умовах аудіозапису.

Аналіз отриманих результатів показав, що спектральний аналіз аудіосигналів є ефективним та надійним інструментом для виявлення прихованих аномалій, які характерні для синтетично згенерованого голосу. Застосування короткочасного перетворення Фур'є дозволяє оцінити зміну частотних компонентів у часі, а використання мел-спектрів враховує особливості сприйняття звуку людським слухом, що робить виділені ознаки більш релевантними для аналізу мовлення. Додатково залучення вейвлет-перетворення забезпечує змінну часово-частотну роздільну здатність, дозволяючи виявляти локальні артефакти та аномалії, які можуть залишатися непомітними при класичному спектральному аналізі. У сукупності ці методи дають змогу формувати інформативний набір ознак, що суттєво підвищує точність та надійність класифікації справжніх і синтетичних аудіозаписів,

забезпечуючи стабільну роботу алгоритму навіть за наявності фонового шуму та різних характеристик голосу.

Отримані результати свідчать про високу ефективність інтеграції спектрального аналізу з сучасними методами машинного навчання для виявлення deepfake-аудіо. Запропонований підхід забезпечує не лише точну класифікацію сигналів, а й стабільну роботу алгоритму за різноманітних умов, включаючи варіації тембру, швидкості мовлення та наявності фонового шуму. Завдяки цьому система є придатною для практичного використання в галузях інформаційної безпеки, аудіофоренстики та інших сферах, де критично важливо гарантувати достовірність і автентичність звукових даних.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Yi, J., Wang, C., Tao, J., Zhang, X., Zhang, C. Y., & Zhao, Y. (2023). Audio Deepfake Detection: A Survey. *Journal of LaTeX Class Files*, 14(8), 1–10.
2. Zhang, B. (2025). Audio Deepfake Detection: What Has Been Achieved and What Remains to Be Done. *Sensors*, 25(7), 1989.
3. Tahaoglu, G., & Yildirim, S. (2025). Deepfake Audio Detection with Spectral Features and Machine Learning. *Expert Systems with Applications*, 202, 117–130.
4. Zhang, B., & Zhang, Y. (2025). Deepfake Audio Detection Using Spectrogram-based Deep Learning Models. *arXiv preprint arXiv:2407.01777*.
5. Almutairi, Z., & Al-Dosari, M. (2022). A Review of Modern Audio Deepfake Detection Methods. *Electronics*, 15(5), 155.
6. Singh, S., & Gupta, P. (2025). Unmasking Digital Deceptions: An Integrative Review of Deepfake Generation and Detection. *Journal of Digital Forensics*, 12(2), 45–60.

Молчан Єгор Віталійович – студент групи КІТС-24м, факультет менеджменту і інформаційної безпеки, Вінницький національний технічний університет, Вінниця, e-mail: egor2014.9595@gmail.com

Науковий керівник: Грицак Анатолій Васильович – доцент каф. Менеджменту та безпеки інформаційних систем, Вінницький національний технічний університет, м. Вінниця, e-mail: grytsak.a.v@gmail.com

Molchan Yehor V. – student of the KITS-24m group, Faculty of Management and Information Security, Vinnytsia National Technical University, Vinnytsia, e-mail: egor2014.9595@gmail.com

Supervisor: Hrytsak Anatolii Vasyliovych – docent of the Department of Management and Security of Information Systems, Vinnytsia National Technical University, Vinnytsia, e-mail: grytsak.a.v@gmail.com