

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського» Міністерства
освіти і науки України

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського» Міністерства
освіти і науки України

Кваліфікаційна наукова
праця на правах рукопису

ПАНАСКІН ДЕНИС ВАЛЕНТИНОВИЧ

УДК:004.93:616-073

ДИСЕРТАЦІЯ

**МЕТОД ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ ДІАГНОСТИЧНИХ
СИСТЕМ В АНАЛІЗІ ЗВУКУ НА ОСНОВІ НЕЙРОННИХ МЕРЕЖ**

121 Інженерія програмного забезпечення

12 Інформаційні технології

Подається на здобуття наукового ступеня доктора філософії

Дисертація містить результати власних досліджень. Використання ідей,
результатів і текстів інших авторів мають посилання на відповідне джерело

_____ Д.В. Панаскін

Науковий керівник: Стіренко Сергій Григорович, доктор технічних наук,
професор

Київ 2025

Анотація

Панаскін Д.В. Метод підвищення ефективності діагностичних систем в аналізі звуку на основі нейронних мереж. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 121 – Інженерія програмного забезпечення та 12 – Інформаційні технології. – Національний Технічний Університет України «Київський Політехнічний Інститут імені Ігоря Сікорського», Київ, 2025.

Метою дослідження було оцінити потенціал глибинного навчання для виявлення респіраторних захворювань у записах легневих звуків (аускультация), а також вивчити можливість збільшення обсягу тренувальних даних шляхом генерації синтетичних спектральних репрезентацій. У роботі запропоновано NASCA-метод (Neural-Augmented Sound Classification for Auscultation), що поєднує обробку аудіозаписів, згорткові варіаційні автокодувальники (CVAE) для аугментації та класичну CNN-модель.

Дослідження було проведено на основі набору даних ICBHI (Міжнародна конференція з біомедичної та медичної інформатики), який складався з 126 суб'єктів і загалом 920 звуків аускультатії легень, серед яких 810 із хронічними захворюваннями, 75 з не хронічними захворюваннями та 35 без будь-яких патологій. На етапі попередньої обробки частота була знижена до 4 кГц і було відфільтровано неінформативні частотні діапазони. Кожен зразок був перетворений у частотний спектр, і були згенеровані мел-спектрограми.

Для вирішення проблеми дисбалансу класів було згенеровано синтетичні спектрограми з використанням згорткових варіаційних автокодувальників.

Для побудови моделі була використана класична згорткова нейронна мережа. Ефективність алгоритму оцінювали за допомогою перехресної

валідації з 10-кратним поділом, а також через експерименти з розділенням аудіозаписів на навчальні та тестові набори на основі групування пацієнтів. Результати були оцінені за такими метриками, як чутливість, специфічність, F1-міра та коефіцієнт Каппа Коена, і досягли F1-міри в 98,45% для задачі класифікації на 5 класів. Дослідження підкреслює потенціал запропонованого методу для синтезу та збільшення чутливих медичних даних і визначає області для подальших досліджень у виявленні клінічних респіраторних захворювань.

Запропонований підхід до синтезу даних має особливе значення в медичних дослідженнях, де часто бракує якісних і збалансованих записів. Застосування згорткових варіаційних автокодувальників для генерування синтетичних зразків істотно покращує якість навчання моделей, що, своєю чергою, позитивно впливає на точність класифікації та ранню діагностику респіраторних захворювань.

Використання глибоких нейронних мереж для аналізу медичних аудіоданих відкриває нові перспективи автоматизації діагностичних процесів. Такий підхід може істотно скоротити час для первинної оцінки стану пацієнта, полегшити навантаження на медичних фахівців і підвищити ефективність лікування, даючи змогу зосередитися на критичних випадках.

Подальші дослідження варто зосередити на кількох взаємопов'язаних напрямках. Насамперед, планується розширити NASCA-метод шляхом інтеграції ембеддингів легеневих звуків із великими мовними моделями (LLM) для мультимодального аналізу (текст + аудіо) з використанням GPT, BERT тощо. Актуальним є також включення до аналізу інших біомедичних сигналів, наприклад фонокардіограм, і їх комбіноване використання з легеневими шумами задля підвищення діагностичної специфічності. Крім того, суттєві перспективи відкриває розробка та валідація федеративного навчання (FL) для обробки розподілених клінічних баз з урахуванням

вимог конфіденційності та захисту персональних даних. Поряд із цим, оптимізація архітектур глибоких мереж і вдосконалення методів генерації синтетичних даних залишаються важливими завданнями для ефективного подолання дисбалансу класів і забезпечення високої точності діагностики.

Ключові слова: Глибоке навчання, діагностика захворювань легень, аускультация, нейронні мережі, синтез даних, медичні дані, класифікація звуків, мел-спектрограма, аудіоаналіз, валідація моделі.

Abstract

Panaskin D.V. A method of increasing the efficiency of diagnostic systems in sound analysis based on neural networks.

Qualified scientific work on the rights of the manuscript. Dissertation for the degree of Doctor of Philosophy in the specialty 121 - Software Engineering and 12 - Information Technology. - National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute Kyiv, 2025.

This study aimed to assess the potential of deep learning for detecting respiratory diseases in lung auscultation recordings, as well as to explore the possibility of expanding the training dataset by generating synthetic spectral representations. The NASCA method (Neural-Augmented Sound Classification for Auscultation) is proposed, combining audio preprocessing, convolutional variational autoencoders (CVAE) for augmentation, and a classical CNN model.

The research utilized the ICBHI (International Conference on Biomedical and Health Informatics) dataset, which consisted of 126 subjects and a total of 920 lung auscultation sounds, including 810 with chronic diseases, 75 with non-chronic diseases, and 35 without any pathology. The preprocessing stage involved reducing the frequency to 4kHz and filtering out non-informative frequency bands. Each sample was then transformed into a frequency spectrum and Melspectrograms were generated.

To address the issue of class imbalance, synthetic spectrograms were generated using convolutional variational autoencoders. A classical convolutional neural network was employed to build the model.

The performance of the algorithm was evaluated through 10-fold cross-validation, as well as through experiments that split the audio recordings into training and test sets based on patient grouping. The results were evaluated using metrics such as sensitivity, specificity, F1-score, and Cohen's kappa, and achieved an F1-score of 98.45% for the 5-class classification problem. The study highlights the potential of the proposed method for synthesizing and

augmenting sensitive medical data and identifies areas for further research in the detection of clinical respiratory diseases

The proposed data synthesis approach is particularly significant in medical research settings, where access to high-quality and balanced datasets is often limited. The use of convolutional variational autoencoders to generate synthetic samples substantially improves model training quality, which in turn has a beneficial impact on classification accuracy and early detection of respiratory diseases.

Employing deep neural networks for the analysis of medical audio data opens new prospects for automating diagnostic procedures. This approach can considerably reduce the time needed for initial patient assessment, lessen the workload on medical professionals, and enhance treatment efficiency by allowing specialists to focus more on critical cases.

Future research should focus on several interrelated directions. First, it is planned to expand the NASCA method by integrating lung sound embeddings with large language models (LLMs) for multimodal analysis (text + audio) using GPT, BERT, etc. Additionally, the analysis of other biomedical signals, for instance phonocardiograms, and their combined use with lung sounds could enhance diagnostic specificity. Furthermore, there are substantial prospects in developing and validating federated learning (FL) for distributed clinical databases, taking confidentiality and data protection requirements into account. Alongside these efforts, optimizing deep neural network architectures and improving synthetic data generation methods remain crucial tasks for effectively addressing class imbalance and ensuring high diagnostic accuracy.

Keywords: Deep learning, respiratory disease diagnosis, auscultation, neural networks, data synthesis, medical data, sound classification, mel-spectrogram, audio analysis, model validation.

Список публікацій здобувача

Публікації у фахових виданнях, включених до переліку наукових фахових видань України з присвоєнням категорії «Б»:

1. Панаскін Д.В., Застосування методів аугментації CVAE для покращення навчання нейронних мереж в аналізі легеневих звуків; Вчені записки ТНУ імені В.І. Вернадського. Серія: Технічні науки, 2024, Том 35 (74) № 5, с. 240 - 245, DOI: 10.32782/2663-5941/2024.5.1/35.
2. Панаскін Д.В., Оцінка стійкості нейромережевих моделей до шуму та артефактів у складних умовах експлуатації; Науковий журнал "Комп'ютерно-інтегровані технології: освіта, наука, виробництво"; 2024 рік, №56, с 226 - 235; DOI:10.36910/6775-2524-0560-2024-56-29.

Публікації у наукових виданнях, включених до наукометричної бази Scopus:

3. D.V. Panaskin , Y. O. Bilokon , O.O. Pididiachyi; Methods for Improving the Efficiency of Diagnostic Systems in the Neural Networkbased Sound Analysis; International Journal of Biology and Biomedical Engineering; 2021; Vol 15; pp. 325-333; DOI : 10.46300/91011.2021.15.39.
4. D. Panaskin, D. Arhypenko, D.Babko; Automated differential diagnostics of respiratory diseases using an electronic stethoscope; Polish Journal of Medical Physics and Engineering; 2023; Vol 29, Issue 4; pp. 208-219; DOI : 10.2478/pjmpe-2023-0022.
5. D. Panaskin, S. Stirenko, D.Babko; Respiratory Disease Detection in Lung Auscultation with Convolutional Neural Networks and CVAE Augmentation; Медичні перспективи, 2024. Т. 29, №3. С. 96-107; DOI:10.26641/2307-0404.2024.3.313569.

Зміст

| | |
|--|-----------|
| Вступ..... | 13 |
| 1. Огляд літератури..... | 18 |
| 1.1. Загальні відомості: від традиційного стетоскопа до автоматизованих систем аускультатії..... | 18 |
| 1.1.1 Обмеження традиційної аускультатії..... | 18 |
| 1.1.2. Запис дихальних звуків: електронні стетоскопи..... | 23 |
| 1.1.3. Загальна архітектура автоматизованих систем аускультатії дихальних звуків..... | 24 |
| 1.2. Основні етапи розвитку автоматизованої системи для аускультатії дихальних звуків..... | 26 |
| 1.2.1. Анотування аудіоданих..... | 26 |
| 1.2.2. Попереднє опрацювання аудіозаписів дихальних звуків..... | 28 |
| 1.2.2.1. Усунення зайвого шуму..... | 28 |
| 1.2.2.2. Усунення серцевих тональностей..... | 33 |
| 1.2.2.3. Додаткові методи передпрацювання сигналу..... | 38 |
| 1.2.3. Візуалізація аудіоданих..... | 39 |
| 1.2.4. Доповнення даних..... | 42 |
| 1.3. Класифікація звуків аускультатії..... | 47 |
| 1.3.1. Ідентифікація ознак..... | 47 |
| 1.3.2. Машинне навчання..... | 52 |
| 1.3.3. Глибоке навчання..... | 54 |
| 1.3.4. Ансамблеве навчання..... | 60 |
| 1.3.5. Метрики оцінки..... | 64 |
| 2. Матеріали та методи..... | 72 |
| 2.1. Набір даних..... | 72 |
| 2.2. Попередня обробка..... | 76 |
| 2.3. Розширення даних..... | 78 |
| 2.4. Тренування і валідація..... | 82 |
| 2.5. Метод підвищення ефективності діагностичних систем в аналізі звуку на основі нейронних мереж..... | 86 |
| 2.5.1. Математичне формулювання методу..... | 87 |
| 2.5.2. Кроки реалізації методу..... | 89 |
| 2.6. Загальний план експерименту та обладнання..... | 91 |
| 3. Результати та обговорення..... | 92 |
| 3.1. Реалізація експерименту..... | 92 |

| | |
|---|------------|
| 3.1.1. Розробка основного набору даних..... | 92 |
| 3.1.2. Попереднє фільтрування аудіозаписів..... | 94 |
| 3.1.3. Створення та обробка мел-спектрограм..... | 96 |
| 3.1.4. Аугментація даних за допомогою змінювального згорткового автоенкодера..... | 99 |
| 3.1.5. Розподіл даних та класифікація..... | 106 |
| 3.2. Експериментальні результати..... | 114 |
| 3.2.1. Оцінка продуктивності генеративної моделі..... | 114 |
| 3.2.1.1. Відстань звукового відліку Frechette..... | 115 |
| 3.2.1.2. Кореляція між оригінальним та синтезованим сигналом..... | 116 |
| 3.2.1.3. Аналіз головних компонентів..... | 117 |
| 3.2.3. Оцінка ефективності моделі класифікації..... | 120 |
| 3.3. Обговорення результатів..... | 122 |
| 3.3.1. Місце запропонованої моделі в класифікації автоматизованих систем для аналізу звуків дихання..... | 122 |
| 3.3.2. Порівняння результатів поточної роботи з відповідними дослідженнями..... | 127 |
| 3.3.3. Особливості реалізації та обмеження систем автоматичної класифікації..... | 143 |
| 3.3.4. Проблема витоку даних..... | 145 |
| 3.3.4.1. Причини витоку даних..... | 145 |
| 3.3.4.2. Тестування гіпотези: другий етап експерименту..... | 147 |
| 3.3.4.3. Проблема витоку даних у подібних дослідженнях..... | 150 |
| 3.4. Проблеми та виклики у розробці автоматизованих систем класифікації звуків дихання та діагностики респіраторних захворювань.. | 158 |
| 3.4.1. Збір даних та їх конфіденційність..... | 158 |
| 3.4.2. Складний характер дихальних шумів та респіраторних захворювань..... | 164 |
| 3.4.3. Методи підвищення продуктивності..... | 167 |
| 3.4.4. Інтерпретація результатів машинного навчання..... | 175 |
| 3.4.5. Впровадження системи автоматичної класифікації дихальних звуків та респіраторних захворювань у клінічну практику..... | 179 |
| 4. Висновки..... | 181 |
| Список використаної літератури..... | 185 |

Перелік умовних позначень

COVID-19 – коронавірусна хвороба;

MARS (Marburg Respiratory Sounds database) – база даних звуків дихання Марбурга;

RALE (Respiration Acoustics Laboratory Environment repository) – репозиторій лабораторного середовища акустики дихання;

ICBHI (International Conference on Biomedical and Health Informatics 2017) – набір даних з Міжнародної конференції з біомедичної та медичної інформатики 2017 року;

BRON – бронхоектатична хвороба та бронхоліт;

SNR (signal-to-noise ratio) – співвідношення сигнал/шум;

PERCH () – велике дослідження етіології пневмонії;

FIR (finite impulse response) – фільтр кінцевої імпульсної характеристики;

STFT (short-time Fourier transform) – короткочасне перетворення Фур'є

ANOVA (analysis of variance) – дисперсійний аналіз;

GAN (generative adversarial network) – генеративно-змагальна мережа;

SMOTE (synthetic minority over-sampling technique) – техніка синтетичного оверсемплінгу меншини;

ADASYN (adaptive synthetic algorithm) – адаптивний синтетичний алгоритм;

VAE (variational autoencoder) – варіаційний автокодер;

CVAE (convolutional variational autoencoder) – згортковий варіаційний автокодер;

FAD (Frechet audio distance) – відстань Фреше для аудіо;

CNN (convolutional neural network) – згорткова нейронна мережа;

MFCC (mel-frequency cepstral coefficients) – мел-частотні кепстральні коефіцієнти. mel-frequency cepstral coefficients;

MRMR (minimum redundancy feature selection) – метод мінімальної надлишковості для вибору ознак;

MLP (multi-level perceptron) – багатошаровий перцептрон. multi-level perceptron;

GMM (gaussian mixture model) – гауссівська змішана модель;

SF (spectral flow) – спектральний потік;

DT (decision tree) – дерево рішень;

RF (random forest) – випадковий ліс;

KNN (k-nearest neighbours) – класифікатор k-найближчих сусідів;

SVM (support vector machines) – машина опорних векторів;

RNN (recurrent neural network) – рекурентні нейронні мережі;

ResNets (residual neural networks) – глибокі залишкові мережі;

LSTM (long short-term memory) – довготривала короткочасна пам'ять;

BiLSTM (bidirectional long short-term memory network) – двонаправлена нейронна мережа з довготривалою короткочасною пам'яттю;

GRU (gated recurrent units) – гейтований рекурентний блок;

BiGRU (bidirectional gated recurrent units.) – двонаправлені гейтовані рекурентні блоки;

KAUH (King Abdullah University Hospital) – Університетська лікарня Короля Абдулли;

MCC (Matthew correlation coefficient) – коефіцієнт кореляції Метьюза;

ROC (receiver operating characteristic) – робоча характеристика приймача;

AUC (area under ROC curve) – площа під кривою ROC;

PCA (principal component analysis) – аналіз головних компонент;

SR (sample rate) – частота дискретизації;

EL (ensemble learning) – ансамблеве навчання;

TL (transfer learning) – перенесення навчання;

FL (federated learning) – федеративне навчання;

DP (differential privacy) – диференційна конфіденційність;

AI HLEG (High-Level Expert Group on Artificial Intelligence) – група високого рівня експертів з питань штучного інтелекту;

ХОЗЛ– хронічне обструктивне захворювання легень;

ІВДШ – інфекція верхніх дихальних шляхів.

ІНДШ – інфекція нижніх дихальних шляхів.

ШІ – штучний інтелект;

МН – машинне навчання;

ГН– глибинне навчання.

Вступ

Актуальність теми. Рання та точна діагностика захворювань дихальної системи є одним із пріоритетних напрямів сучасної медицини, оскільки патології легень призводять до значних соціальних та економічних наслідків. За даними Всесвітньої організації охорони здоров'я, респіраторні хвороби, зокрема хронічне обструктивне захворювання легень та інші патології, що супроводжуються специфічними аускультативними ознаками, стабільно входять до переліку найбільш поширених причин смертності. Традиційний підхід до аускультативної діагностики легень (з використанням акустичного стетоскопа) залежить від суб'єктивних чинників – слуху лікаря, рівня шуму та резонансів у стетоскопі, досвіду медичного персоналу. Це зумовлює підвищений ризик помилкових діагнозів, особливо на ранніх стадіях захворювання.

Сучасні дослідження у сфері штучного інтелекту та глибокого навчання (Deep Learning) доводять, що нейронні мережі, зокрема згорткові (CNN) та генеративні автоенкодери (VAE, CVAE), здатні досягати високих показників чутливості й специфічності під час класифікації аудіосигналів. У розв'язанні завдань аналізу звуків легень (аускультативної діагностики) такі моделі покращують виявлення та диференціацію патологічних шумів (хрипів, свистів, трісків). Проте реалізація ефективних діагностичних систем у реальних умовах стикається з низкою викликів:

- недостатньою кількістю та незбалансованістю клінічних аудіозаписів (через рідкісність окремих патологій);
- високою варіативністю легневих шумів, що можуть накладатися на серцеві тони та фоновий шум;
- потребою у методах збільшення (аугментації) даних, які зберігають релевантність спектральних характеристик;
- ризиком «витоку даних» та переоцінки точності у випадку неправильного формування навчальних і тестових підвибірок.

Постає завдання розробити та дослідити метод підвищення ефективності діагностичних систем в аналізі звуків легень із застосуванням глибинних нейронних мереж, що враховуватиме проблему дисбалансу даних та необхідність синтетичного розширення вибірок, а також забезпечуватиме стійкість та високу точність класифікації у клінічно шумних умовах. Актуальність теми зумовлена як нагальною потребою у масштабованих, доступних та точних методах раннього скринінгу респіраторних захворювань, так і стрімким прогресом deep learning-підходів, що сьогодні формують state-of-the-art у різних завданнях розпізнавання та класифікації аудіосигналів.

Зв'язок роботи з науковими програмами, планами, темами. Тема дисертаційної роботи входить в план наукової роботи затвердженому на кафедрі обчислювальної техніки КПІ ім. Ігоря Сікорського, що враховує розпорядження Кабінету Міністрів України від 2 грудня 2020 р. No 1556-р про схвалення Концепції розвитку штучного інтелекту в Україні. Запропоновані у дисертації методи використані у науково-дослідній роботі НФДУ «Наука для безпеки людини та суспільства» – проєкт «Платформа штучного інтелекту для дистанційного автоматизованого виявлення та діагностики захворювань людини» реєстраційний номер проєкту: 2020.01/0490.

Мета і завдання дослідження. Метою дисертаційної роботи є підвищення ефективності діагностичних систем аналізу легеневих звуків шляхом розробки та застосування глибинних нейронних мереж, що здатні розпізнавати і класифікувати патологічні шумові патерни (хрипи, свисти тощо) за умов недостатніх та незбалансованих вибірок.

Для досягнення вказаної мети потрібно розв'язати такі завдання:

1. Дослідити сучасні підходи до обробки та виділення специфічних ознак із легеневих звуків (зокрема мел-спектрограм), а також оцінити стан та складність баз даних (ІСВНІ тощо).

2. Розробити метод збагачення (аугментації) наявних тренувальних вибірок шляхом генеративних нейронних мереж (CVAE) та перевірити його вплив на чутливість і специфічність класифікації.
3. Створити модель згорткової нейронної мережі (CNN) або ансамблевий підхід (CNN+LSTM), здатний до мультикласової класифікації респіраторних захворювань за аудіозаписами, та оптимізувати параметри навчання.
4. Проаналізувати особливості шуму та «витоку даних», розробити рекомендації для коректного розподілу на тренувальні й тестові набори з урахуванням груп пацієнтів.
5. Оцінити ефективність запропонованого методу в діагностиці захворювань дихальної системи (ХОЗЛ, пневмонія тощо) та визначити перспективи його практичного впровадження у клінічні системи.

Об’єкт дослідження. Процеси автоматизованого аналізу легеневих звуків (аускультатії) з метою виявлення та класифікації респіраторних патологій.

Предмет дослідження. Методи та алгоритми глибинного навчання (CNN, автоенкодери) для поліпшення показників чутливості й специфічності діагностики легень за аудіозаписами.

Методи дослідження. Під час розв’язання поставлених завдань застосовано системний аналіз підходів до обробки аудіосигналів, методи спектрального аналізу (мел-спектрограми), генеративні нейронні мережі (зокрема згорткові варіаційні автоенкодери – CVAE), глибинні згорткові моделі (CNN), а також статистичні метрики оцінки продуктивності (чутливість, специфічність, F1-міра, коефіцієнт Каппа Коена). Для написання та тестування моделей використовувалися програмні фреймворки PyTorch, а

також методи перехресної валідації (k-fold) для підвищення надійності результатів.

Наукова новизна отриманих результатів.

- Вперше розроблено комплексний метод до синтетичного збагачення баз даних легеневих звуків із використанням згорткового варіаційного автоенкодера (CVAE). Він забезпечує зменшення дисбалансу між різними класами захворювань та підвищує F1-показник класифікації на 14% без суттєвого збільшення ризику «надмірного узгодження» (overfitting), за рахунок оптимального поєднання CVAE-генерації з реальними вибірками, керованого налаштування внутрішніх репрезентацій та додаткової перевірки синтетичних зразків для збереження релевантних патернів дихальних шумів.
- Набуло подальшого розвитку вирішення проблеми «витоку даних» у дослідженнях з аускультатії легень: доведено, що при некоректному розподілі записів від одного пацієнта у тренувальному та тестовому множинах показники ефективності можуть бути завищені. Запропоновано схему розподілу за пацієнтами, що мінімізує ризик «витоку даних», за рахунок створення спеціальної процедури відбору, яка гарантує відсутність аудіозаписів того самого пацієнта одночасно в навчальній і тестовій вибірках.
- Удосконалено методику глибинної класифікації: поєднання CNN та технік фільтрування низькочастотного шуму для аналізу звуків легень. На відміну від існуючих робіт, вдалось підвищити стійкість до реальних шумів, у пропонованому підході надається стійкість до реальних шумів, що підвищує точність у клінічних умовах, за рахунок інтеграції додаткових процедур аугментації та адаптивного виділення корисних частотних ділянок, які дозволяють ефективніше

усувати низькочастотні артефакти і серцеві тони без втрати релевантної інформації.

- Подальшого розвитку набули підходи до мультикласової класифікації складних респіраторних захворювань із подібними аускультативними патернами шляхом вибору функцій втрат з ваговою корекцією меншинних класів, стратегій регуляризації (dropout, batch normalization) та застосування методів аугментації (CVAE, mixup) за браку даних. Це дало змогу водночас збалансувати різні класи, знизити ризик їх змішування та підвищити точність диференційованої діагностики в шумних умовах.

Практичне значення отриманих результатів. Створений метод синтетичного доповнення даних та розроблені моделі CNN можуть бути інтегровані у програмне забезпечення для дистанційної та амбулаторної діагностики респіраторних захворювань, забезпечуючи високу ефективність первинного скринінгу. Результати дослідження сприятимуть удосконаленню електронних стетоскопів і мобільних застосунків (телемедицині рішень) для моніторингу стану легень у домашніх умовах. Запропонований підхід дозволяє виявляти складні патологічні патерни за низькошумних і реальних клінічних умов, знижуючи кількість неправильних діагнозів. Запропоновані рекомендації щодо формування мультикласової моделі та оптимальних параметрів навчання.

1. Огляд літератури

1.1. Загальні відомості: від традиційного стетоскопа до автоматизованих систем аускультатії

1.1.1 Обмеження традиційної аускультатії

Аускультатія, яка зазвичай проводиться за допомогою акустичного стетоскопа, є одним із наявних методів неінвазивної діагностики патологій дихальної системи в реальному часі, який є дуже інформативним у пульмонології. На основі характеру та локалізації аномальних дихальних звуків спеціаліст робить висновки про наявність певних патологічних станів, а також оцінює структуру та функцію легенів і бронхів, набагато швидше, ніж підготовка інших досліджень: лабораторних аналізів, рентгенівських знімків, результатів томографії тощо. Проте, традиційна аускультатія за допомогою стетоскопа має кілька значних недоліків, пов'язаних як з об'єктивними, так і з суб'єктивними факторами.

Якість аускультатії органів дихальної системи в значній мірі залежить від характеристик стетоскопа. Незважаючи на те, що цей пристрій значно покращився з часів його винаходу Рене Ланнеком у 1817 році, його все ще не можна вважати ідеальним інструментом, оскільки він не має рівномірного передавання звуку на всіх частотах. Передача сигналу під час акустичної аускультатії ускладнюється шумом, ефектами резонансу трубок та іншими спотвореннями. Крім того, більшість сучасних стетоскопів підсилюють звуки на частотах нижче 112 Гц і ослаблюють на вищих частотах. Такі особливості цінуються кардіологами, але можуть ускладнити аускультатію, коли мова йде про легені та бронхи [3].

Стетоскопи рідко піддаються тестуванню, і лікарі зазвичай вибирають їх на підставі зовнішнього вигляду, зручності або репутації бренду, не звертаючи уваги на технічні характеристики, оскільки варіації матеріалів,

трубок та компонентів призводять до унікальних особливостей для кожного стетоскопа, що впливає на передачу конкретних діапазонів частот [3].

Нормальні та аномальні дихальні звуки, що виникають внаслідок вібрації тканини легенів та повітряних шляхів під час дихання, є складними та важкими для характеристики та опису. Загалом, для встановлення діагнозу потрібна послідовна аускультация:

- тип дихання (оцінка тривалості стадій вдихання та видихання);
- наявність або відсутність додаткових дихальних звуків та їх характер;
- фаза дихального циклу (вдихання, видихання), під час якої чуються додаткові шуми.

Фізіологічні дихальні звуки зазвичай називаються везикулярними та бронхіальними дихальними звуками, які слухаються як у здорових, так і у хворих людей. Однак захворювання легенів змінюють характер нормальних дихальних звуків. Зокрема, якщо в легенях є рідина, везикулярне дихання послаблюється, а при астмі його нормальна структура змінюється, як показано на Рисунку 1.1.1, і видихання помітно подовжується.



Рис. 1.1.1 Схема нормального везикулярного циклу дихання.

Фізіологічні ділянки, де проводиться аускультация, також є важливими для діагностики захворювань дихальної системи. Наприклад, при пневмонії бронхіальне дихання, яке зазвичай чути лише над трахеєю або великими бронхами, може бути почуте над периферичними легенями.

Хоча більшість аномальних дихальних звуків мають чітко визначені акустичні характеристики, як показано в Таблиці 1.1.1, кожне захворювання може бути охарактеризоване комбінацією цих звуків, і той самий звук часто зустрічається при різних захворюваннях.

Таблиця 1.1.1.

Характеристики патологічних дихальних звуків

| Звук | Причина | Місце | Характеристика звуку | Акустичні параметри | Хвороба |
|--|--|--------|---|--|---|
| Вологі дрібноміхурцеві хрипи (хрип високої частоти). <i>Дрібні хрипи.</i> | <ul style="list-style-type: none"> ▪ Раптове вирівнювання газового тиску між двома ділянками легенів. ▪ Відкриття закритих малих дихальних шляхів. | Легені | <ul style="list-style-type: none"> ▪ Переривчастий ▪ Високий. ▪ Інспіраторний. | <ul style="list-style-type: none"> ▪ Швидке заглушення хвильового відгалуження. ▪ Частота: приблизно 650 Гц. ▪ Коротке (близько 5-10 мс). | <ul style="list-style-type: none"> ▪ Пневмонія. ▪ Пульмональний міжструнний фіброз. ▪ Серцева недостатність. ▪ Пульмональний набряк. |
| Вологі великоміхурцеві хрипи (низькочастотна крепітація). <i>Грубі хрипи.</i> | <ul style="list-style-type: none"> ▪ Інтермітентне очищення дихальних шляхів, спричинене рідиною або секретами. | Легені | <ul style="list-style-type: none"> ▪ Короткий. ▪ Переривчастий. ▪ Низький. ▪ Інспіраторний. | <ul style="list-style-type: none"> ▪ Швидке заглушення хвильового відгалуження. ▪ Приблизна частота 350 Гц. ▪ Довге (приблизно 20-30 мс). | <ul style="list-style-type: none"> ▪ Пневмонія. ▪ Бронхоектатична хвороба. ▪ Пульмональний міжструнний фіброз. ▪ Серцева недостатність. |
| Сухі свистячі хрипи. Хрипи | <ul style="list-style-type: none"> ▪ Стиснуті дихальні шляхи. | Бронхи | <ul style="list-style-type: none"> ▪ Постійний. ▪ Високий. ▪ Видихальний/вдихальний. | <ul style="list-style-type: none"> ▪ Частота: 100–5000 Гц. ▪ Тривалість більше 80 мс. | <ul style="list-style-type: none"> ▪ Бронхіальна астма. ▪ Хронічне обструктивне |

| Звук | Причина | Місце | Характеристика звуку | Акустичні параметри | Хвороба |
|-------------------------------------|--|-----------------------|--|---|--|
| | <ul style="list-style-type: none"> Обмеження потоку повітря. | | | | захворювання легень. <ul style="list-style-type: none"> Бронхіоліт. • бронхоектатична хвороба. • Пульмональна міжструнна фіброза. • Обструктивний бронхіт. • Пухлина. • Чужорідне тіло. |
| <i>Басовий свистячий хрип.</i> | <ul style="list-style-type: none"> Порушення рідинних плівок секретій. • Вібрація стінки дихальних шляхів. | Бронхи | <ul style="list-style-type: none"> Постійний. Високий. • Видиховий/ вдиховий. | <ul style="list-style-type: none"> Частота менше 200 Гц. Тривалість більше 80 мс. | <ul style="list-style-type: none"> Бронхіт. Пневмонія. |
| <i>Стридор.</i> | <ul style="list-style-type: none"> Стиснуті дихальні шляхи. | Гортань, трахея | <ul style="list-style-type: none"> Постійний. Високий. Вдиховий. | <ul style="list-style-type: none"> Частота більше 500 Гц. | <ul style="list-style-type: none"> Епіглотит. • Чужорідне тіло. |
| <i>Тертя плевральної порожнини.</i> | Тертя плеври легеневої та парієтальної. | Стінка грудної клітки | <ul style="list-style-type: none"> Постійний. Низький. • Видиховий/ вдиховий. | <ul style="list-style-type: none"> Ритмічна коротка послідовність звуків. Частота менше 350 Гц. Тривалість більше 15 мс. | <ul style="list-style-type: none"> Плеврит. Перикардит. • Плевральна пухлина. |

Дихальні звуки, характерні для певних захворювань, є складними і дуже різноманітними, оскільки вони можуть змінюватися залежно від їх стадії, а також відрізняються від однієї людини до іншої, залежно від віку пацієнта та супутніх захворювань [4]. Наприклад, на початковій стадії бронхіту вислуховуються басові хрипи на вдиху і видиху, в міру подальшого розвитку захворювання вони трансформуються у вологі, великопухирчасті або дрібнопухирчасті хрипи.

Деякі дихальні звуки, які чуються під час аускультатії, містять не один, а кілька патологічних звуків легенів і особливо складно діагностуються. Крім того, вони накладаються на звуки серця та внутрішніх органів і змішуються з оточуючим шумом.

Крім того, існує велика варіабельність у описах як саме дихальних звуків, так і їх патологічних маркерів. Коли група дослідників під керівництвом А. Совіарві у 2000 році зібрала визначення всіх термінів, що описують дихальні звуки, використовувані в комп'ютерному аналізі дихальних звуків (CORSА), було отримано список із 162 слів. Такий великий набір даних не дозволяє лікарям мати єдине визначення використовуваних термінів (наприклад, сухі хрипи іноді описуються як "свисти", а вологі хрипи як "звук смаження рису на сковороді"), і також значно ускладнює завдання створення автоматизованої системи ідентифікації дихальних звуків [5].

На основі зазначених характеристик точна аускультативна діагностика потребує вмінь, які набуваються та вдосконалюються в клінічній практиці [2,6]. Існують декілька досліджень, які підтверджують, що здатність визначати та відрізнити незвичайні дихальні звуки сильно залежить від знань, досвіду та спеціалізації лікаря.

Ще одним очевидним недоліком традиційної аускультатії є відсутність дистанційного моніторингу, що ускладнює діагностику сидячих пацієнтів та людей, що проживають у віддалених районах. Такі обмеження особливо

важливі в контексті зростаючого попиту на телемедицину як більш доступну та дешевшу альтернативу традиційним медичним обстеженням.

Отже, існуючі діагностичні алгоритми для захворювань дихальної системи, які включають аускультацию за допомогою традиційного стетоскопа, мають значні обмеження, які потенційно збільшують кількість помилок, призводять до невірних діагнозів, неправильного застосування антибіотиків, збільшення захворюваності та смертності тощо.

1.1.2. Запис дихальних звуків: електронні стетоскопи

Альтернативою традиційним акустичним стетоскопам став вигляд електронних стетоскопів, які дозволили подолати проблему індивідуальних аудитивних характеристик лікарів, читаючи повний частотний спектр дихальних звуків.

В наш час для запису звуків легенів загалом використовують два типи трансдюсерів. До переваг електретних мікрофонів відносять низьку вартість і компактний розмір, але чим вони більш чутливі до вищих частот дихальних звуків, тим більше вони захоплюють надмірний шум навколишнього середовища. Контактні акселерометри менше чутливі до зовнішніх перешкод, але зазвичай дорожчі, крихкі і можуть захоплювати внутрішні резонанси [1.3].

Стив Крейман та спів. [7], досліджуючи придатність різних мікрофонів для запису акустичних сигналів дихальної системи, прийшли до висновку, що найбільш пристосованим мікрофоном для запису звуків легенів є електретний мікрофон з конічним роз'ємом і діаметром 10-15 мм.

Основною перевагою електронних стетоскопів є дистанційний та безперервний моніторинг стану пацієнтів, за умови, що вони використовуються з мобільною технологією. Як зауважив Тонг Сі у 2022 році, під час пандемії COVID-19 було введено кілька дистанційних автоматизованих систем для запису та аналізу кашлю або дихальних звуків. Вони дозволили лікарям віддалено проводити скринінг на коронавірусну хворобу, що

розширило медичні заклади і не порушило ізоляцію пацієнтів, зменшивши ризик поширення хвороби [8].

Проте головною перевагою електронних стетоскопів стала можливість запису та зберігання звуків легенів у вигляді електронних сигналів, що дозволило вивчати їх з кращим тлумаченням, ділитися ними, анотувати, створювати спеціальні бази даних і т. д. На сьогодні вже є кілька доступних, зокрема, мультимедійна база даних MARS (Marburg Respiratory Sounds), репозиторій RALE (Respiratory Acoustic Laboratory Environment), база даних ICBHI (International Conference on Biomedical and Medical Informatics) і т. д. [9].

1.1.3. Загальна архітектура автоматизованих систем аускультатії дихальних звуків

Поява електронних стетоскопів дозволяє подолати проблему індивідуального аудитивного сприйняття дихальних шумів, але вона не вирішила проблему суб'єктивності аускультатії та різниці в умінні інтерпретувати аускультатію між лікарями.

Для підвищення діагностичної значимості аускультатії було необхідно впровадити об'єктивний збір даних, а також забезпечити стандартизований підхід до їх інтерпретації. Саме для цього були розроблені комп'ютеризовані системи аналізу звуків легенів на основі останніх досягнень в області акустичної інженерії, обробки сигналів та штучного інтелекту (ШІ).

Загалом, принцип створення автоматизованих систем для аускультатії дихальних звуків представлений на малюнку 1.1.3.1.

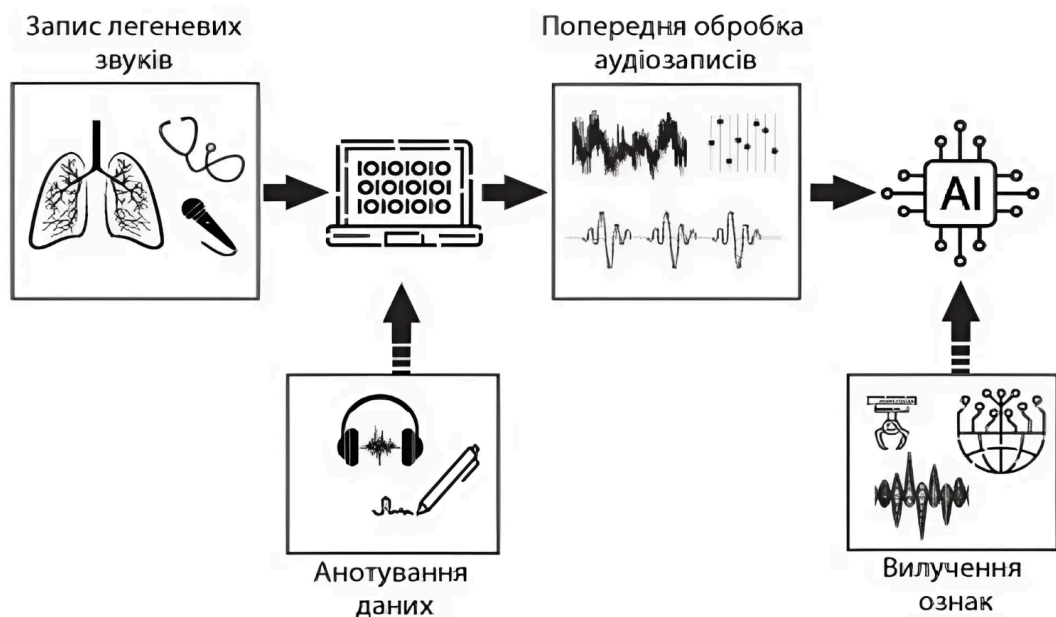


Рис. 1.1.3.1. Загальна архітектура автоматизованої системи аускультатії дихальних звуків.

Автоматизована система аускультатії дихальних звуків передбачає отримання електронних аудіозаписів дихальних звуків, анотування цих аудіозаписів, їх попередню обробку з наступним вилученням ознак, і, нарешті, класифікацію цих записів за допомогою методів ШІ, зокрема, машинного навчання (МН) та глибокого навчання (ГН).

На сьогодні було продемонстровано багато прототипних систем для автоматизованого комп'ютеризованого аналізу дихальних звуків за допомогою глибокого аналізу на основі ШІ. Однак, переглядаючи прогрес, досягнутий в цій галузі, вчені вказують, що на сьогодні можна вважати завершеним лише розвиток електронних систем для читання та запису звуків дихання, тоді як всі інші кроки у розвитку автоматичних систем перебувають на стадіях активного розвитку. [11]

Незважаючи на ці досягнення, у галузі автоматизованого аналізу дихальних звуків залишаються кілька викликів і обмежень. Основною перешкодою є нестача маркованих даних, необхідних для навчання глибоких

моделей навчання при діагностиці захворювань. Ще однією складністю є значна змінність дихальних звуків, що створює виклики при створенні моделей, які були б стійкими в різних умовах і демографічних групах пацієнтів. Нарешті, потрібні подальші валідації запропонованих методів МН та ГН у реальних клінічних умовах, щоб продемонструвати їх відтворюваність та надійність у клінічній практиці.

Прогрес продовжується у сфері систем аналізу дихальних звуків, з постійними зусиллями по поліпшенню точності та надійності діагностики захворювань за допомогою новаторських методів. І хоча деякі з них вже показали перспективні результати у визначенні респіраторних захворювань, їх вдосконалення продовжується, а використання автоматизованих систем для аналізу дихальних звуків як стандартного інструменту клінічної практики для діагностики захворювань легенів залишається питанням найближчого майбутнього.

1.2. Основні етапи розвитку автоматизованої системи для аускультатії дихальних звуків

1.2.1. Анотування аудіоданих

Однією з основних перешкод у розробці автоматизованих систем аускультатії дихальних звуків є різноманітність і мінливість звуків легенів, що значно ускладнює розробку моделей, які можуть точно діагностувати хвороби. Анотування як нормальних, так і патологічних дихальних звуків допомагає вирішити цю проблему. Анотації є цінними на етапах перед опрацювання, зокрема для сегментації дихальних звуків, а також для створення тренувальних наборів даних, перевірки моделі розпізнавання звуку, розрахунку її ефективності тощо.

Тонг Сі у 2022 році описав характеристики більш ніж 10 баз даних з дихальними звуками, доступними для досліджень, вказуючи на їх різноманітність

за протоколом анотації записів. Деякі з них містять відомості самого пацієнта як мітки, в той час як інші містять експертні думки або незалежні результати тестування (ПЛР на COVID-19) [8].

Крім того, існує значна варіабельність в описах як самих дихальних сигналів, так і їх патологічних маркерів. Коли група дослідників під керівництвом А. Совіярві у 2000 році зібрала визначення всіх термінів, що описують дихальні звуки, використані в Комп'ютерному аналізі дихальних звуків (CORSA), був отриманий список з 162 словами. Очевидно, що такий великий набір даних не дозволяє лікарям мати єдине визначення використовуваних термінів (наприклад, свист - це іноді характеризується як "свист", а тріскотіння - як "звук плавлення рису на сковороді"), а також значно ускладнює завдання створення автоматизованої системи розпізнавання дихальних звуків [12].

Така різноманітність пов'язана не лише з відсутністю певних стандартизованих процедур для анотування записів дихальних звуків, а й зі складністю самої процедури. Дімітра Еммануїліду розповідає про процес створення анотацій аудіозаписів для комп'ютерної системи скринінгу дихальних звуків у дитячій популяції, у якому брали участь дев'ять експертів-рецензентів (педіатрів або лікарів з педіатричним досвідом). Два основні рецензенти анотували звуки, зібрані з восьми різних місць тіла кожного пацієнта, як нормальні, так і аномальні, з відповідною міткою впевненості для кожної анотації. Якщо на записі чітко було чути два або більше дихань і рецензент міг впевнено сказати, чи вони нормальні чи ні, анотацію позначали як "визначено". Якщо можна було чітко ідентифікувати лише одне дихання, анотацію позначали як "ймовірно". Якщо зв'язані з диханням шуми не були ідентифіковані, запис позначали як "неінтерпретоване". Якщо думки двох основних рецензентів відрізнялися з приводу конкретної анотації, інші фахівці прослуховували запис, щоб роз'яснити неоднозначність. На завершення, для кожної мітки рецензенти

вказували підінтервал запису, що включав у себе ділянку невизначеної довжини, що на найкращий спосіб показувала їх анотацію [13].

1.2.2. Попереднє опрацювання аудіозаписів дихальних звуків

Попереднє опрацювання аудіоданих, отриманих за допомогою електронних стетоскопів або інших пристроїв запису, є ключовим етапом у розробці автоматизованих систем аускультатії легеневих звуків, оскільки методи, які використовуються на цьому етапі, мають безпосередній вплив на ефективність моделі машинного навчання.

У більшості робіт, розглянутих у цьому огляді, для отримання більш гладких і менш шумних сигналів дихальні звуки проходять кілька етапів попереднього опрацювання, які включають видалення шуму, сегментацію, відсічку зайвої інформації, перевибіркування, нормалізацію та інше.

1.2.2.1. Усунення зайвого шуму

Звуки легенів, отримані за допомогою аускультатії, наповнюються різними шумами, включаючи ті, що походять від серця та інших внутрішніх органів, фонові розмови, оточуюче середовище, електричні перешкоди, відхилення стетоскопа та інше. Ці шуми забруднюють сигнал і іноді маскують важливі характерні сигнали, оскільки їхні часово-частотні шаблони в основному відповідають характерним шаблонам звуків легенів [14].

Ведучою характеристикою оцінки шуму аудіозапису є співвідношення сигнал-шум (SNR), тобто співвідношення потужності сигналу до потужності відповідного шуму під час запису:

$$SNR = \frac{P_s}{P_n} = \left(\frac{A_s}{A_n} \right)^2 \quad (1.2.1)$$

де A_n - середньоквадратична амплітуда шуму, A_s - середньоквадратична амплітуда сигналу, P_n - середня потужність шуму, P_s - середня потужність сигналу.

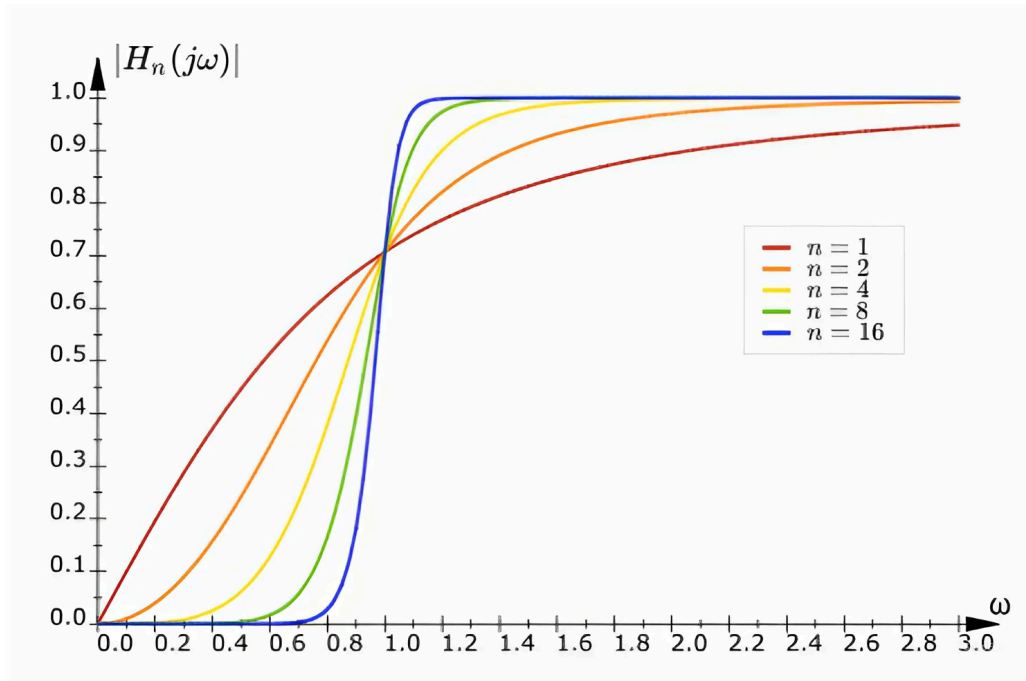


Рис. 1.2.1. Діаграма частотної характеристики фільтра високих частот Баттерворта

Частотна характеристика фільтра Баттерворта, яка представлена на рис. 1.2.1, обчислюється за формулою:

$$|H_{n,hp}(j\omega)| = \frac{\omega^n}{\sqrt{1+\omega^{2n}}} \quad (1.2.2),$$

де n - порядок фільтра, ω - частота зрізу.

Крім постійних перешкод, які присутні у кожному записі звуків легенів, також можуть з'являтися випадкові перешкоди. Наприклад, будь-яке механічне зміщення електронного стетоскопа призводить до низькочастотного хвилеподібного сигналу, що накладається на корисний звук. Іноді ці неправильні короточасні широкосмугові вибухи, також називають

артефактами відхилення, можуть маскувати аномальні звуки легенів, такі як шуми. Для усунення артефактів відхилень у дослідженні Фрейвана та інших у 2021 році була використана функція згладжування локальної поліноміальної регресії. У ній сигнал доповнюється зваженою регресійною функцією, зваження для точок збільшуються для точок ближче до прилягаючої лінії і зменшуються для точок далі від неї [16].

Для пригнічення складного шумового забруднення, викликаного розмовами, артефактами відхилення, випадковими розмовами, шумом на вулиці та іншими факторами, Дімітра Еммануїліду та інші розробили у 2015 році модифіковану схему спектрального віднімання [13].

Використовуючи два мікрофони, один з яких записує звуки легенів, забруднені фоновим шумом ($y(n)$), а інший записує лише фоновий шум ($d(n)$), вони шукали чистий звук легенів, який зазвичай виражається формулою:

$$y(n, \tau) = x(n, \tau) + d(n, \tau) \quad (1.2.3),$$

де y - вимірний сигнал довжиною N ; x - чисті звуки легенів, d - сигнал шуму, специфічний для пацієнта або кімнати.

Параметр τ використовується для представлення обробки протягом коротких вікон часу $w(n)$, тобто:

$$\begin{aligned} x(n, \tau) &= x(n)w(\tau - n), \\ y(n, \tau) &= y(n)w(\tau - n), \\ d(n, \tau) &= d(n)w(\tau - n) \end{aligned} \quad (1.2.4),$$

Алгоритм потім відображає короточасні вікна сигналів в спектральній області за допомогою короточасного перетворення Фур'є:

$$|Y(\omega, \tau)|e^{j\varphi_y(\omega, \tau)} = |X(\omega, \tau)|e^{j\varphi_x(\omega, \tau)} + |D(\omega, \tau)|e^{j\varphi_d(\omega, \tau)} \quad (1.2.5),$$

де $Y(\omega, \tau)$, $X(\omega, \tau)$, $D(\omega, \tau)$ відповідають короткочасному дискретному перетворенню Фур'є $y(\omega, \tau)$, $x(\omega, \tau)$, $d(\omega, \tau)$ відповідно.

Використовуючи зовнішній шумовий сигнал, записаний з зовнішнього мікрофона, автори визначили короткочасовий амплітудний спектр $D(\omega, \tau)$ як $D(\omega, \tau)$ і запропонували, що фазовий спектр шуму $\varphi_d(\omega, \tau)$ може бути замінений фазою шумового сигналу $\varphi_y(\omega, \tau)$. У результаті вони отримали формулу:

$$\hat{X}(\omega, \tau) = (|Y(\omega, \tau)| - |\hat{D}(\omega, \tau)|)e^{j\varphi_y(\omega, \tau)} \quad (1.2.6).$$

Припускаючи, що фоновий шум $d(n, \tau)$ є процесом з нульовим середнім значенням, некорельованим з сигналом легень $x(n, \tau)$, та враховуючи домен щільності спектральної потужності, та застосовуючи ту ж формулу, ми маємо:

$$|\hat{X}(\omega, \tau)|^2 = |Y(\omega, \tau)|^2 - |\hat{D}(\omega, \tau)|^2 \quad (1.2.7).$$

Для покращення ефективності цього загального підходу до звуків легень, які є складними за своєю природою, автори розширили його конструкцію з урахуванням різноманітних, непередбачуваних та гетерогенних характеристик шумових спотворень, а також необхідності збереження вмісту сигналу аускультатії як на низьких, так і на високих частотах. Зокрема, вони розширили схему віднімання на кілька частотних діапазонів ω_k , ввели відповідні правила вагового коефіцієнта (δk) , які ідентифікують спектральні діапазони, що несуть найбільше інформації в сигналах легень, а також забезпечили згладжування оцінок сигналу в межах сусідніх кадрів та спектральних діапазонів. Після внесення цих змін, спектральна щільність сигналу для кадру τ та частотного діапазону ω_k була визначена рівнянням:

$$\left| \hat{X}(\omega_k, \tau) \right|^2 = \left| \hat{Y}(\omega_k, \tau) \right|^2 - \alpha_{k,\tau} \delta_k \left| \hat{D}(\omega_k, \tau) \right|^2 \quad (1.2.8),$$

$\hat{Y}_{k,\tau}$ представляє вдосконалену апроксимацію $Y(\omega, \tau)$ через сусідні кадри; $\alpha_{k,\tau}$ позначає коефіцієнт зайвого віднімання, тоді як δ_k вказує на спектральний ваговий коефіцієнт, що відповідає за підкреслення нижніх частот, які загалом пов'язані з дихальними сигналами, та пом'якшення впливу вищих частот, які схильні до забруднення шумом.

Коефіцієнт зайвого віднімання α_k та спектральний ваговий коефіцієнт δ_k , що представляють тимчасові та спектральні параметри масштабування відповідно, динамічно налаштовуються для кожного інтервалу часу τ , який охоплює 50 мс, на основі переважаючого відношення сигнал/шум у цьому інтервалі. Додатково вносяться коригування для кожного частотного діапазону k для вирівнювання зі спектральними характеристиками сигналів. Оцінка чистого звуку легенів потім отримується за допомогою зворотного перетворення Фур'є.

Цей алгоритм був попередньо протестований на реальних даних, зібраних в умовах медичних установ в шумних середовищах у рамках Великомасштабного дослідження етіології пневмонії (PERCH) [18].

Вибір інших, більш специфічних засобів усунення непотрібного шуму в основному залежить від проблеми дослідження, характеру аудіозаписів та умов, у яких вони були створені. Наприклад, у вже згаданій роботі Дімітри Еммануїліду одним із важливих попередніх кроків було відфільтрування плачу, оскільки записи були зібрані з маленьких дітей, і плач був частим джерелом перешкод [13].

Гарретт Нельсон та спів.(2014) досліджували адаптивні методи скасування шуму для аускультатії в транспортному вертольоті Sikorsky UH-60 армії США в умовах, де фоновий шум перевищував 90 дБ [19].

1.2.2.2. Усунення серцевих тональностей

Інформативні дихальні шуми при аускультатії також завжди змішуються із тонами серця, діапазон яких становить від 5 Гц до 600 Гц, що збігається з низькочастотною частиною легневих звуків, тому усунути перешкоди серцевих тонів, не пошкодивши легневих звуків простим фільтром майже неможливо. Для цього застосовуються більш складні методики, наприклад, вейвлети, фільтри кінцевої імпульсної характеристики (*FIR*-фільтри), адаптивна фільтрація з рекурсивним алгоритмом найменших квадратів, фільтрація за часом/частотою, реконструкція, незалежний компонентний аналіз, метод на основі ентропії тощо [20].

Загалом методи видалення серцевих звуків з аудіозаписів легких можна розділити на дві категорії. Перша передбачає оцінку серцевих звуків та їх видалення за допомогою таких засобів, як хвильова фільтрація з адаптивним порогом або використання адаптивних фільтрів, а друга передбачає локацію та виділення сегментів серцевих звуків разом з дихальними звуками та заміну цих сегментів за допомогою лінійної або нелінійної передбачення.

Досліджуючи спектральні характеристики сигналів дихальних звуків за допомогою фільтрів хвильових перетворень, Ірина Хоссейн і Зара Мусаві (2003) спостерігали помітне зниження середньої потужності дихальних звуків по всьому частотному спектру. Однак нещодавній розробкою Абделкрім Семмад і Мохамед Бахура (2021) представили нову техніку на основі хвильових перетворень для приглушення дихальних звуків, яка ефективно знижує шум без пошкодження оригінальної якості звуку.

Для ефективного приглушення серцевих звуків з мінімізацією спотворень низькочастотних дихальних звуків використовуються адаптивні фільтри. У 1986 році Віджай Айєр запропонував адаптивний фільтр з використанням сигналу "розширеного електрокардіограми" як посилення. Цей підхід успішно зменшував серцеві шуми на 50-80%; проте вимога до

зовнішнього електрокардіографічного сигналу вводила практичні виклики в використанні фільтра [23].

Згодом, Мартін Компіс та спів. (1992) змінили цей алгоритм, щоб отримати адаптивний фільтр, який вимагав лише одного вхідного сигналу, але не змогли досягти значного покращення, їх метод зменшивши шум серця лише на 24–49% [24]. Для поліпшення цього показника Леонтіос Хаджілеонтіадіс та Ставрос Панас використали адаптивний фільтр на основі статистики четвертого порядку, який також не потребував посилення на додатковий сигнал. Завдяки складності алгоритму вони змогли знизити рівень шуму, що генерується серцевими тональностями, більш ніж на 90% [25].

Пізніше ті ж автори розробили новий адаптивний фільтр, який поєднував у собі хвильове перетворення, об'єднуючи метод адаптивного фільтру та композицію хвиль, що також ефективно зменшувало шум серця. Алгоритм, запропонований авторами, ґрунтувався на тому, що вибухові піки звуків серця в області часу мають помітний сигнал на багатьох масштабах хвиль, тоді як сигнали дихальних звуків швидко згасали зі збільшенням масштабу. Автори використовували представлення сигналів з різною роздільною здатністю i , за допомогою жорсткого порога в області хвильового перетворення, вони відокремлювали динамічний сегмент вхідного сигналу (звуки серця) від стабільного сегмента (дихальні звуки) [26].

Концепція шумопониження за допомогою хвильового перетворення, запропонована Хаджілеонтіадісом та Панасом, ґрунтується на припущенні, що хвильові коефіцієнти звуків легень значно нижчі, ніж хвильові коефіцієнти серцевих тональностей, тоді як забруднений сигнал кінцевої довжини N виражається наступним чином:

$$y_i = f_i + \sigma z_i, \dots i = 1, 2, \dots, N \quad (1.2.9),$$

де f_i - чистий звук легенів, z_i - гаусівський шум, що характеризується середнім значенням нуль і дисперсією σ^2 , який є незалежним від сигналу f_i .

Використовуючи дискретне хвильове перетворення з лівоінвертованою матрицею W , можна отримати наступний вираз:

$$Y_i = F_i + \sigma Z_i \dots i = 1, 2, \dots, N \quad (1.2.10),$$

Тут $Y = Wy$ представляє хвильові коефіцієнти шуму y_i , $F = Wf$ представляє хвильові коефіцієнти сигналу f_i , а $Z = Wz$ представляє хвильові коефіцієнти шуму z_i .

Основний принцип шумопониження за допомогою порогового методу хвильового перетворення полягає в тому, щоб елімінувати хвильові коефіцієнти, що відповідають шуму Z , зберігаючи при цьому ті, що відповідають корисному сигналу F , тим самим наближаючи відфільтрований сигнал до бажаного сигналу f_i . Для досягнення цього може бути визначений фіксований поріг:

$$\hat{Y} = T(Y, t) = \{Y \mid |Y| \geq t \mid Y| < t \quad (1.2.11),$$

або м'який поріг:

$$\hat{Y} = T(Y, t) = \{sgn(Y)(|Y| - t) \mid |Y| \geq t \mid 0 \mid Y| < t \quad (1.2.12),$$

де \hat{Y} - це оцінка хвильових коефіцієнтів Y .

Існують різні правила для вибору значення порогу. Найчастіше використовуються дві формули:

$$t = \sigma \sqrt{2 \lg N} \quad (1.2.13),$$

де N - кількість точок вибірки;

$$t = \sigma \sqrt{\frac{2 \lg N}{\lg(j+1)}} \quad (1.2.14),$$

де j - масштаб дискретного хвильового перетворення.

Серед більш сучасних розробок в цій галузі варто відзначити вдосконалений адаптивний алгоритм зменшення шуму, що ґрунтується на нормалізованому останньому середньоквадратичному алгоритмі, запропонованому Номаном Аль-Наггаром та інші у 2018 році.

Параметрами цього алгоритму є довжина фільтру (L_j), яка визначається в межах послідовності 2^n , починаючи з 2 і подвоюючись до досягнення значення 2048, та розмір кроку (μn), який автоматично визначається за допомогою випадкової оптимізації. Під час експериментів автори встановили, що оптимальний діапазон для μn знаходиться між 0 і 0.1. В цьому діапазоні адаптивний фільтр проявляв стабільну ефективність, особливо з нестационарними біосигналами, такими як звуки легенів. Оцінки продуктивності алгоритму в розділенні серцевих звуків від забруднених пульмональних звуків базувалися на співвідношенні сигнал-шум, що показали значні покращення зі зростанням довжини фільтру L [27].

Друга категорія методів, яка включає виявлення та видалення сегментів серцевих звуків разом зі звуками легень та їх заміну за допомогою лінійної або нелінійної передбачуваності, може включати такі методи, як виявлення серцевих звуків за ентропією звуків легень, використання мультимасштабних хвильових продуктів, часово-частотне фільтрування з наступною відновленням сигналу та інше [28].

Один з таких підходів, зокрема, був реалізований Лукуї Ші та спів. (2019). Для видалення серцевих звуків вони застосували хвильовий поріг, який відокремлював хвильові коефіцієнти на рівні розкладання хвиль. Хвильові коефіцієнти нижче порогу відкидалися, тоді як ті, що вище нього, зберігалися. Після цієї операції сигнали чистих дихальних звуків були відновлені за допомогою відновлення хвильових коефіцієнтів [15].

Ще один приклад методів приглушення серцевих звуків під час запису звуків легень відноситься до другої категорії, як це показано в алгоритмі, запропонованому Даніелем Флорес-Тапіа та спів. (2007). Цей метод використовує мульти масштабний добуток хвильових коефіцієнтів, отриманих зі звичайного сигналу, для виявлення сегментів, що містять серцеві тони. Ці сегменти в подальшому видаляються, а отримані прогалини заповнюються за допомогою серії фільтрів лінійного передбачення. Цей підхід показав перспективні результати, особливо для стаціонарних сегментів, позбавлених аудіо артефактів, таких як кліки. Димітра Еммануеліду (2018) також практично реалізувала цей метод у своїх дослідженнях [13].

Останнім часом спостерігається тенденція до використання інтегрованих алгоритмів для вилучення серцевих шумів зі звукових записів легень, які одночасно поєднують кілька методів. Наприклад, Фенг Менг та спів. (2020) запропонували послідовний підхід, який включав кілька технік. Спочатку вони застосували фільтр типу Чебишева I порядку з полосою пропускання від 20 Гц до 1600 Гц для зменшення шуму низької і високої частот. Потім вони використали модифікований хвильовий фільтр для приглушення навколишнього шуму, а також для зниження серцевих тонів і голосів людини. Нарешті, був використаний адаптивний фільтр з найменшим квадратом помилок з модифікованим довідковим сигналом. Практична реалізація цієї комплексної схеми показала ефективне приглушення шуму і відмінні акустичні характеристики [30].

Хай Чен та спів. (2019) запропонували альтернативу традиційному підходу до вилучення серцевих шумів з дихальних звуків. Їх метод передбачає трійну класифікацію дихальних звуків за допомогою глибоких нейронних мереж (ResNet) на основі оптимізованого S-перетворення. S-перетворення, застосоване до карт ознак дихального сегменту, виокремлює характеристики, пов'язані з частотою, за допомогою унікального вікна, залежного від частоти. Це вікно розширюється при нижніх частотах і стискається при вищих частотах, ефективно захоплюючи відповідні ознаки. Важливо відзначити, що цей підхід усуває потребу в передобробці, такі як вилучення серцевих тонів і змішаних артефактів, оскільки шум не перешкоджає навчанню ознак за допомогою глибоких мереж і продуктивність класифікаторів залишається незмінною [31].

1.2.2.3. Додаткові методи передпрацювання сигналу

Оскільки більшість нейронних мереж працюють з вхідними даними фіксованої довжини в аудіозаписах звуків легень, респіраторні цикли (періоди вдихання та видиху) визначаються в часовому домені, після чого виконується сегментація, обрізаються порожні фрагменти та заповнюються прогалини за допомогою методів, таких як нульове доповнення, копіювання респіраторних циклів, інтерполяція кубічними сплайнами та інші методи.

Вважається, що нормальний діапазон дихання у дорослих становить 12-18 вдихів за хвилину, тому один респіраторний цикл (вдихання та видихання) займає від 2 до 5 секунд, залежно від індивідуального темпу дихання. Таким чином, сегменти тривалістю від 2 до 7 секунд, як правило, забезпечують покриття обох частот дихання. Цей емпіричний підхід широко використовується в літературі [16, 32, 33].

Для сегментації записів бази даних звуків дихання ICBNI 2017 на окремі респіраторні цикли на основі експертних анотацій, Георгіос Петмезас (2022) встановив, що тривалість дихальних циклів коливається від 0,2 с до 16 с і

становила в середньому 2,7 с. Для ізоляції респіраторних циклів фіксованої довжини були використані наступні підходи: цикли, довші за 2,7 с, обрізаються, тоді як цикли, коротші за 2,7 с, навпаки, розширювалися за допомогою стратегії заповнення зразками [33].

Джинглонг Ніу та спів. (2019) ввели метод автоматичної ідентифікації респіраторних циклів, де визначення закінчення дихального циклу ґрунтується на ідентифікації вершини автокореляційної функції. У цьому підході автокореляція, що представляє кореляцію звукового сигналу з його відкладеною копією, обчислюється для набору L звукових зразків x_i , де i змінюється від 1 до L [34].

У деяких моделях передпрацювання також включає зміну частоти, на якій здійснюється вибірка аудіосигналів. Зменшення частоти дискретизації аудіосигналів зменшує обчислювальну складність і, внаслідок цього, прискорює процес класифікації без значного погіршення продуктивності та якості. Наприклад, база даних звуків дихання ICBHI містить зразки, записані за різними частотами дискретизації від 4000 Гц до 44 100 Гц. Зазначаючи, що корисні сигнали знаходяться на частотах нижче 2000 Гц, Георгіос Петмезас (2022) перевибирав усі записи з частотою 4000 Гц [33].

Мати чіткий сигнал без зміни тенденції з роками забезпечує кращу продуктивність алгоритмів глибокого навчання, тому нормалізація даних є важливим процесом перед впровадженням будь-якої стратегії машинного навчання. Існують кілька альтернатив для нормалізації даних. Наприклад, М.Т. Garcia-Ordas та спів. (2020) [35] і Бірук Абер Тессема (2022) [36] використовують алгоритм нормалізації MinMax, Фрейван (2022) [16] нормалізує сигнали за допомогою індексу z .

1.2.3. Візуалізація аудіоданих

Часто інформація, яка не є видимою у часовому представленні сигналу, розкривається у його частотному представленні. Двовимірні

аудіопредставлення, або часо-частотні зображення аудіосигналів, допомагають виділити ознаки та використовуються у глибокому аналізі. На таких зображеннях додаткові дихальні звуки, такі як шипи та скрипи, відображають впізнавані патерни, що можуть допомогти в їх ідентифікації. Одночасно, використовуючи візуалізацію, легко виділити частотні діапазони, де аудіо сигнали мають нульову енергію та вирізати їх, дозволяючи нейронним мережам акцентувати увагу на цікавих областях, що призводить до значних покращень продуктивності.

Існує кілька методів візуалізації аудіозаписів звуків дихання, зокрема, найчастіше використовуються три різні представлення: спектрограма, сколограма та мел-спектрограма [37].

Спектрограма - це візуальне представлення, де час відображається на горизонтальній вісі, частота - на вертикальній вісі, а інтенсивність сигналу зображується за допомогою палітри кольорів. Це часо-частотне зображення генерується за допомогою короткочасного перетворення Фур'є (STFT), широко використовуваного інструменту в аудіо аналізі та обробці. STFT дозволяє візуалізувати, як змінюються частотні компоненти з часом, надаючи цінні уявлення про тимчасову динаміку аудіосигналу. [38]

Зображення короткочасного перетворення Фур'є від заданого дискретного сигналу подається за формулою:

$$\Phi(n, w) = \sum_{i=-\infty}^{\infty} i w(n - i) e^{-j w i}, \quad (1.2.17)$$

де $w(i)$ - функція вікна, центрована у часі n .

Останнім часом, разом із спектрограмами, для візуалізації аудіоданих також використовуються інші методи - сколограми, тобто масштабоване представлення безперервного хвильового перетворення (англ. continuous wavelet transform, CWT) сигналу. Іншими словами, сколограми представляють неперервне вейвлет-перетворення у тривимірному тензорі, організованому з

фіксованою кількістю шарів, де кожен шар вказує на згортки даних сигналу з різними масштабами базового вейвлета [39,40].

Як правило, і спектрограми, і сколограми перетворюють аудіосигнал у складну двовимірну карту, яка може бути корисною для виявлення змінних особливостей у даних. Узагальнено, у обох методах сигнал $s(t)$ аналізується за допомогою варіацій однієї шаблонної функції $p(t)$ з двома параметрами p та q :

$$R_s(p, q) = \int s(t) p_{p,q}(t) dt, \quad (1.2.18)$$

де d - диференціальний приріст.

У випадку спектрограми, p представляє фіксоване (ненульове) вікно, а p, q вказує на те ж саме вікно p , зсунуте і модульоване на частоту, пропорційну q . З іншого боку, у контексті сколограми, p позначає фіксовану нульову функцію середнього значення, як правило, вейвлет, а p, q представляє той самий вейвлет p , зсунутий і розширений на фактор, пропорційний q [40].

Мел-спектрограма показує звуковий спектр на основі мел-шкали, яка є сприйняттєвою шкалою з рівномірно розташованими тонами, відповідно до сприйняття звуку людським вухом.

Зазвичай перехід між частотою (Гц) та висотою тона в мелах розраховується за формулою

$$m = 2595 \left(1 + \frac{f}{700} \right) = 1127 \ln \ln \left(1 + \frac{f}{700} \right) \quad (1.2.19)$$

Обернене рівняння перетворення виглядає так:

$$f = 700 \left(10^{\frac{m}{2595}} - 1 \right) = 700 (e^{\frac{m}{1127}} - 1) \quad (1.2.20)$$

1.2.4 Доповнення даних

Враховуючи особливості наборів даних, зібраних в природних умовах, що стосуються будь-якої медичної інформації і, зокрема, результатів аускультатії, розв'язання цієї задачі часто вимагає попереднього етапу - додавання або розширення даних, оскільки більшість доступних наборів аудіо даних звуків дихання є відносно невеликими і недостатньо збалансованими, що може бути найслабшим зв'язком автоматизованої категоризації звуків дихання на основі машинного навчання.

Одним із потенційних рішень цієї проблеми може бути уніфікація таких різних репозиторіїв, але враховуючи велику різноманітність даних, збережених у них, важко практично реалізувати такий підхід.

Тонг Сі (2022) зауважує, що деякі дослідники намагаються обійти цю проблему, застосовуючи техніки тренування на невеликих даних, включаючи напів наглядне навчання, самонавчання та трансфертне навчання, або використовуючи не дихальні аудіо дані або неанотовані записи дихальних звуків для тренування моделі [8].

Проте, більшість робіт використовують різні методи додавання даних. Це особливо справедливо при застосуванні моделей нейронних мереж для класифікації звуків, оскільки ці типові перенавчені моделі вимагають великих наборів даних для зменшення ймовірності перенавчання, підвищення точності та усунення чутливості моделі до фонового шуму або надмірної інформації [41].

Додавання даних спрямоване на вирішення двох основних завдань:

- Збільшення наборів тренувальних даних.
- Нормалізація даних.

Для вирішення першого завдання широко використовуються методи доповнення аудіоданих, такі як розтягування за часом, зміна висоти звуку, випадкові коливання, додавання шуму до сирого сигналу і т. д. Варіації

аудіофайлів, створені цими методами, можуть покращити загальні можливості узагальнення відповідних моделей.

Наприклад, для збільшення тренувального набору даних, Алі Мохаммед Алькоди та спів. (2022) використовують довільне уповільнення або прискорення звуку, зсув звуку вліво або вправо на випадкову величину, додавання шуму, а також довільне налаштування гучності звуку [42]. Ще один клас технік доповнення даних - обробка спектрограм, така як горизонтальне перевертання (кожна спектрограма випадковим чином перевертається) або обрізка спектрограми (кожна спектрограма випадковим чином обрізається певну кількість разів) [43].

Потреба у нормалізації даних виникає через те, що існуючі бази даних аудіо зразків аускультатії є несбалансованими, тобто кількість елементів одного класу набагато більша. Як і з будь-яким іншим медичними даними, це переважно пов'язано з рідкою появою аномальних зразків. Наприклад, у базах даних COVID-19 менше позитивних зразків на COVID-19, ніж негативних. Така нерівновага даних утруднює побудову надійної моделі класифікації [8].

Для вирішення цієї проблеми використовують так звані генеративні методи або методи на основі генеративних нейронних мереж, які називають генеративними протиріччями (Generative adversarial network, GAN). GAN - це клас алгоритмів штучного інтелекту, які є парою нейронних мереж у грі з нульовим сумісним результатом, при цьому одна мережа генерує кандидатів (генератор), а інша їх оцінює (дискримінатор).

Прикладами методів, використовуваних для нормалізації даних, є метод штучного перебору меншини (Synthetic Minority Oversampling Technique, SMOTE) і метод адаптивного штучного перебору (Adaptive Synthetic Sampling, ADASYN), які дозволяють контролювати розміри вибірок, штучно збільшуючи обсяг тренувальних даних меншого класу та зменшуючи обсяг тренувальних даних більшого класу, а також різні автоенкодері (VAE).

Дослідження, проведене Сіддхарттом Гайролою (2020) мало на меті розробити фреймворк на основі глибоких нейронних мереж для ідентифікації аномальних звуків легень з обмеженими даними. Незважаючи на великий обсяг набору даних ІСВНІ, який складає 6898 зразків, він стикається з проблемою дисбалансу: нормальний клас представлений 53% зразків. Щоб впоратися з цим викликом, автори широко проаналізували набір даних і експериментували з різними стандартними методами доповнення даних, такими як додавання шуму, зміна швидкості, випадковий перемикання, зміна висоти звуку, зсув і т. д.. Вони використовували зважений випадковий вибірковий пристрій для рівномірного вибору міні-паketу з кожного класу. Однак ці традиційні методи виявилися неефективними і погіршили узагальнення недооцінених класів (свист, тріскання або обидва). Для вирішення цього автори запропонували новий метод доповнення на основі конкатенації. Цей метод полягав у генерації нового зразка шляхом випадкового вибору та поєднання двох зразків з того самого класу. Цей підхід привів до значного покращення точності класифікації аномалій. Запропонований метод досяг правильності 95%, чутливості 92% і специфічності 97%.

Джейн Салдана та спів. (2022) проаналізували можливості кількох варіаційних автоенкодерів (багатошаровий, згортковий, умовний) для синтезу звуків дихання. Автори оцінювали реалістичність синтетичних звуків дихання за допомогою таких показників, як відстань Фреше для аудіо (Frechet audio distance, FAD), взаємкореляція та спотворення мель-кепстрального спектра. Виявилося, що моделі класифікації звуків легень на основі глибокого навчання можуть досягати значного покращення продуктивності після розширення неузгоджених наборів даних за допомогою автоенкодерів зміни [45].

Для вирішення проблеми неузгоджених наборів даних в галузі визначення респіраторної патології, де кількість зразків для певних класів

значно менша, ніж для інших, Марія Тереса Гарсія-Ордас та спів. (2020), використовуючи набір даних із респіраторних зразків ІСВНІ, випробували ряд методів генерації. Інтеграція нейронної мережі та варіаційного автоенкодера (VAE) продемонструвала перспективні результати виявлення респіраторної патології [35].

Автори провели експерименти з варіаційним згортковим автоенкодером (CVAE), ймовірнісною моделлю, призначеною для розуміння розподілу даних та генерації нових зразків, які відповідають цьому розподілу. Функція втрат, використана для навчання CVAE, складалася з двох компонентів: "відновлення", спрямованого на забезпечення точної роботи кодера-декодера, та "регуляризації", яка наближала розподіли, сформовані прихованим шаром кодера, до стандартного нормального розподілу, за допомогою дивергенції Кульбака-Лейблера:

$$VAE_{LOSS} = \|x - \tilde{x}\|^2 + KL[N(\mu_x, \sigma_x), N(0, 1)], \quad (1.3.1)$$

де x - відтворення \tilde{x} , $N(\mu_x, \sigma_x)$ - нормальний розподіл з середнім μ_x та стандартним відхиленням σ_x , $KL[p, q]$ - дивергенція Кульбака-Лейблера, обчислена за формулою:

$$KL[p, q] = - \int p(x) \log q(x) dx + \int p(x) \log p(x) dx \quad (1.3.2)$$

Дослідження використовувало набір даних, що включав звуки дихання, отримані від пацієнтів з ХОЗЛ та астмою, а також здорових осіб. Цей набір даних був розділений на навчальний, перевірочний та тестовий піднабори. Щоб вирішити проблему нерівності класів, штучні зразки недостатньо представлених класів були згенеровані за допомогою автоенкодера з варіаційним підходом (VAE). Наступно, на розширеному наборі даних навчали Згорткові Нейронні Мережі (CNN), а результати оцінювалися за

допомогою метрик, включаючи точність, точність, повернення та F1-показник.

Результати свідчили про те, що використання VAE для розширення даних підвищувало ефективність CNN у виявленні респіраторних патологій, особливо для класів, які були недостатньо представлені. Найбільш вигідні результати були досягнуті завдяки поєднанню CNN з VAE, що призвело до точності на рівні 96% та F1-показника 0.92. Це підкреслює ефективність використання штучно створених синтетичних зразків, згенерованих VAE, для покращення здатності CNN виявляти респіраторні аномалії.

Для вирішення проблеми нерівності даних у базі даних ICBHI 2017, команда дослідників з Національного університету Сінгапуру під керівництвом Ї Ма (2020) запропонувала метод аугментації даних MixUp. Ця техніка сприяє лінійному переходу рішучих меж між класами, що призводить до більш гладкої оцінки невизначеності. Для реалізації MixUp автори уважно вивчили характеристики кожного типу випадкового звуку легенів, присутніх у наборі даних.

Наприклад, шуми були визначені за їх короткотривалим та вибуховим характером, часто відчутним як відмітна подія на тлі звичайних звуків легенів. Відповідно, дослідники поєднали "звичайні" цикли з "трісками", щоб збільшити кількість циклів "трісків" у наборі даних. Аналогічно, вони збільшили кількості циклів "хрипіння" та "трісків разом з хрипінням". Процес MixUp полягає в злитті двох випадково вибраних зразків разом із відповідними мітками у навчальному наборі даних, що призводить до створення нового набору зразків та міток:

$$\tilde{x} = \lambda x_i + (1 - \lambda)x_j,$$

$$\tilde{y} = \lambda y_i + (1 - \lambda)y_j \quad (1.3.3),$$

де (x_i, x_j) - це два вектори ознак, а (y_i, y_j) - класові мітки двох ознак, кодовані одночасно. $\lambda \in [0, 1]$ - це випадкове число, згенероване згідно з бета-розподілом.

Систематичний огляд методів доповнення даних та глибокого навчання у класифікації звуків, проведений Олусолою Абаєомі-Аллі та спів. (2022), який включав 56 статей, показав, що найбільша кількість публікацій використовувала додавання шуму (39,2%) як метод доповнення даних, а 26,7% використовували зсув у часі. Моделі на основі GAN та зсуву висоти використовувалися у 21,4% досліджень кожна. Інші техніки йшли в порядку зменшення: розтягування в часі, MixUp і фоновий шум. Деякі публікації використовували такі інструменти, як обертання та обрізання спектрограм, модуляція швидкості, порушення довжини вокального тракту (Vocal Tract Length Perturbation, VTLP) і т. д.

1.3. Класифікація звуків аускультатії

Зазвичай етап класифікації звуків легенів включає наступні кроки: вибір ознак, побудову та навчання класифікаторів та оцінку продуктивності.

1.3.1. Ідентифікація ознак

Більшість моделей машинного навчання для автоматичної ідентифікації респіраторних захворювань та класифікації дихальних звуків потребує додаткового кроку екстракції ознак, що пояснюють характеристики звуків та використовуються як змінні для виявлення або класифікації.

Екстракція ознак полягає в перетворенні початкових векторів високої розмірності у менш вимірені представлення, яке можна представити рівнянням:

$$Ph \rightarrow Pd, \text{ де } d < h \quad (1.3.1).$$

Іншими словами, це може бути розглянуто як процес зменшення обсягу даних, зберігаючи при цьому основні характеристики аналізованого сигналу.

Часто вивчаються акустичні характеристики, які використовуються в обробці музики чи мови, такі як часові характеристики (початок, темп, період, швидкість перетину нуля, гучність ударів тощо). Однак, при обробці дихальних звуків вони зазвичай недостатньо для розрізнення та правильної класифікації, оскільки варіації в аудіосигналах, пов'язаних з різними респіраторними станами, можуть бути складними, тонкими та неявними, а самі дихальні звуки можуть бути втрачені на тлі зовнішнього шуму. Тому в системах автоматичної класифікації дихальних звуків, як правило, використовуються інші, більш глибокі характеристики, включаючи високорівневі масивні дескриптори [8].

Екстракція ознак часто виконується на основі спектральних характеристик (щільність спектра потужності, середня частота, вибрана частота з найвищою потужністю, квантильні частоти, потужність частотних смуг тощо), кепстральних коефіцієнтів мел-частоти, коефіцієнтів вейвлета, хроматичності, аналізу авторегресії, лінійного прогнозування характеристик ентропії тощо.

Систематичний огляд, проведений Olusola Abayomi-Alli та спів. (2022), висвітлив поширеність певних методів екстракції ознак порівняно з іншими в попередніх дослідженнях. Зокрема, огляд показав, що: 25,6% переглянутих публікацій використовували представлення на основі MFCC, 18,6% застосовували логарифмічні методи мел-спектрограм, 16,3% використовували інші ознаки, отримані з мел-спектрограм, 9,3% використовували традиційні спектрограми та швидке перетворення Фур'є (FFT), решту 30% публікацій віддали перевагу альтернативним методам представлення ознак, таким як мішок слів, константне-Q перетворення (CQT) та частота перетину нуля (ZCR) [41].

Багато робіт також використовують багатовимірні вектори ознак, які є поєднанням згаданих методів. Наприклад, після оцінки кількох методів екстракції ознак та класифікаторів, Morten Grønnesby та спів. (2017) запропонували п'ятизмірний вектор ознак, що складається з чотирьох функцій у часовому домені та однієї функції у спектральному домені [47].

У роботі "Ефективна класифікація легеневих звуків з використанням глибинно-розділених моделей CNN з Fused STFT і MFCC" Шин-Юн Чжун запропонував метод, який використовує комбінацію короткочасного перетворення Фур'є (STFT) та малих частотних кепстральних коефіцієнтів (MFCC), витягнутих з сирого аудіосигналу для класифікації звуків легенів. Ці витягнуті ознаки потім подавалися на вхід у згорнутий нейромережі з розділеним глибинним шаром для класифікації. Інновація полягає в застосуванні архітектури глибоко розподіленої CNN, що допомагає зменшити кількість параметрів у моделі. Це зменшення параметрів підвищує обчислювальну ефективність моделі, при цьому зберігає або навіть покращує її точність класифікації.

Ефективність запропонованої моделі була оцінена на наборі аудіозаписів легеневих звуків. Результати показали, що запропонована модель досягла точності 97,6%.

У 2018 році Фатіх Демір запропонував модель, що складалася з комбінації згорткових і рекурентних шарів для екстракції ознак з сирого аудіосигналу. Згорткові шари використовувалися для вивчення локальних шаблонів в аудіосигналах, тоді як рекурентні шари використовувалися для захоплення часових залежностей між ними. Отримані ознаки потім використовувалися для навчання багатошарового персептрона (MLP) для класифікації звуків легенів з точністю 96,2% [38].

Мохамед Багура у 2009 році провів оцінку та порівняння різних методів екстракції ознак, які включали перетворення Фур'є, лінійне передбачення кодування, вейвлет-перетворення та мел-частотні кепстральні коефіцієнти

(MFCC). Ці методи були поєднані з різними підходами до класифікації, такими як векторна квантизація, гауссівські змішані моделі (GMM) та штучні нейронні мережі. Його дослідження показало, що використання кепстральних коефіцієнтів MFCC у поєднанні з GMM виявилось найефективнішим методом для розрізнення між звичайними та свистячими звуками дихання. Це поєднання продемонструвало переважну ефективність у точному класифікуванні шаблонів звуків дихання, підкреслюючи ефективність ознак MFCC у захопленні відповідних характеристик для аналізу звуків дихання [49].

У своєму дослідженні Bruno Rocha et al. (2021) дослідили застосування алгоритму мінімальної зайвості з максимальною важливістю (MRMR) для відбору ознак у контексті характеристики звуку. Вони використовували спектрограму, згенеровану за допомогою швидкого перетворення Фур'є (FFT) з вікном Хеммінга, а також спектрограму мелу. Спочатку автори обчислювали перші чотири стандартизовані моменти спектральних розподілів зі спектрограми, включаючи центроїд, розподіл, асиметрію та ексцес. Крім того, вони виділяли інші широко використовувані ознаки для характеристики тембру звуку, такі як кількість переходів через нуль, ентропія, рівномірність, шорсткість та нерівномірність.

Крім того, автори представили три додаткові ознаки на основі спектрального потоку (SF), який вимірює евклідову відстань між величиною спектра послідовних кадрів. Ці ознаки включали SF inc, що представляє суму позитивних різниць між кадрами; SF half-wave, отримане шляхом напівхвильової виправлення значень SF; та медіанне SF, де до співвідношення було застосовано медіанний фільтр для видалення фіктивних піків. Більш того, у дослідженні було оцінено енергію високих частот за допомогою двох методів: яскравість, що вказує на енергію високих частот вище визначених порогових частот (100, 200, 400 та 800 Гц); та згасання, що визначає частоту, нижче якої міститься певний відсоток загальної спектральної енергії.

Усього автори дослідження вилучили 81 ознаку з кожного кадру Мел-спектрограми та спектрограми, обчисленої за допомогою швидкого перетворення Фур'є з вікном Хеммінга. Ці ознаки включали 25 спектральних ознак, 26 ознак MFCC та 30 мелодичних ознак. Для кожної з цих ознак були обчислені п'ять статистичних характеристик: середнє значення, стандартне відхилення, медіана, мінімальне значення та максимальне значення, що загалом складає 2430 ознаки. Для відбору ознак використовувався алгоритм мінімальної зайвості з максимальною важливістю (MRMR). Цей алгоритм пріоритизує ознаки, оцінюючи коефіцієнт взаємної інформації, враховуючи як важливість, так і зайвість кожної ознаки. Цікаво, що дві найважливіші ознаки, вибрані MRMR для набору даних фіксованої тривалості, були мелодичними ознаками, тоді як три найважливіші ознаки для набору даних змінної тривалості були спектральними ознаками. Однак сім ознак MFCC були присутні в обох наборах даних, що свідчить про їхню сталу важливість у різних типах даних та тривалості [50].

Ще одним прикладом автоматичного ранжування ознак є дослідження Бірук Тессем та інші співавтори (2022). За допомогою дискретного вейвлет-перетворення автори ідентифікували ряд ознак, зокрема стандартне відхилення, середнє квадратичне значення, коефіцієнт форми, ексцес, вершина, перехресний фактор, фактор зазору, загальне гармонічне спотворення, діапазон енергії, тощо, і вибір найважливіших довіряється алгоритму ANOVA, який ранжує або впорядковує отримані ознаки на основі значення функції оцінки, якою була одновимірний аналіз дисперсії, зі схемою нормалізації MinMax [36].

Справді, важливо враховувати, що вибір характеристик для відбору значно залежить від природи аудіоданих. Як показано у дослідженні Бруно Роше (2021), різні класифікатори можуть по-різному взаємодіяти з певними характеристиками. Тому ефективність варто оцінювати в комплексі, враховуючи як самі характеристики, так і специфіку роботи класифікаторів.

Однак інформативність усіх спроб порівняти ефективність певних характеристик [49] або зробити їх рейтинг [36, 50] обмежена, оскільки вибір характеристик для відбору в значній мірі залежить від природи аудіоданих, що було продемонстровано у дослідженні Бруно Роше (2021) [50]. Крім того, різні класифікатори працюють по-різному з певними характеристиками, тому їх ефективність слід розглядати комплексно.

1.3.2. Машинне навчання

У літературі численні публікації зосереджуються на використанні методів машинного навчання (МН) для автоматизованої ідентифікації респіраторних захворювань та класифікації звуків легень. Серед найбільш поширених алгоритмів є:

- Дерево рішень (DT): Класифікатор, який використовує умови, подібні до дерев, для відображення ознак у кілька категорій.
- Випадковий ліс (RF): Ансамблевий метод, що складається з кількох дерев рішень і забезпечує підвищену стійкість до помилок.
- Класифікатор k-найближчих сусідів (KNN): Класифікує об'єкт на основі відстані між новим об'єктом та раніше ідентифікованими об'єктами.
- Наївні класифікатори Байєса: Група алгоритмів класифікації, що базуються на теоремі Байєса та припускають незалежність між ознаками.
- Метод опорних векторів (SVM): Алгоритм, що використовує ядра для представлення складних даних у низькорозмірному дискримінативному просторі, де об'єкти класифікуються на основі їхнього розташування відносно видимого проміжку в цьому просторі.

Ці алгоритми часто застосовуються в аналізі та класифікації звуків легень через їхню ефективність у роботі з різноманітними наборами даних та завданнями класифікації.

Завдяки своєму гнучкому вибору ядра та стабільній продуктивності, SVM є одним з найкращих та найбільш широко використовуваних методів машинного навчання. У дослідженні Дімітра Еммануельду (2018) для розділення респіраторних звуків на типові та аномальні з використанням параметрів спектрально-темпорального аналізу (час, частота, тимчасова динаміка та спектральні модуляції) у якості ознак, нелінійний класифікатор методу опорних векторів з м'якими межами продемонстрував точність 86,67%, чутливість 86,8% та специфічність 86,55% [13].

Мортен Греннесбі та спів. (2017) провели порівняльний аналіз кількох стандартних класифікаторів і виявили, що метод опорних векторів (SVM) з ядром базисних радіальних функцій перевершив інші. Цей SVM досягнув точності на рівні 86% і повноти 84% у класифікації хрипінь в респіраторних звуках, перевершивши результати традиційного аускультування, виконаного медичним персоналом. Автори використовували 5-вимірний вектор ознак, що складався з 4 характеристик з області часу та 1 з області спектра. Низька розмірність вектора ознак сприяла високій ефективності класифікатора SVM: модель навчалася на звичайному комп'ютері всього за 1,44 секунди та класифікувала 319 звуків легенів за 1,08 секунди. Цей швидкий час обробки демонструє практичну доцільність використання SVM для класифікації респіраторних звуків у реальному часі [47].

Бірук Тесема та спів. (2022) порівняли точність класифікації 7 основних захворювань легень для 4 моделей машинного навчання, навчених за однаковими даними та ознаками. У дослідженні використовувалися записи, отримані за допомогою електронного стетоскопу від госпіталізованих пацієнтів з клінічно підтвердженими діагнозами астми, бронхієктазу, пневмонії, бронхіоліту, ХОЗЛ та інших. Для порівняння та класифікації цих захворювань використовувалися прості класифікатори Байеса, k-найближчих сусідів, ансамбль моделей, що поєднував два дерева рішень, та метод опорних векторів. Результати дослідження показали, що SVM з ядром Fine Gaussian

виявився найбільш адекватною моделлю, а після оптимізації за методом баєсівської класифікації його точність досягла 98,8% [36].

Однак, незважаючи на досягнення високого рівня продуктивності, традиційні методи машинного навчання мають ряд недоліків, зокрема, вони потребують додаткового початкового етапу екстракції ознак, який виконується "вручну", і, отже, може потенційно призвести до помилки, пов'язаної з недостатньою вибіркою ознак для конкретного алгоритму або набору даних.

1.3.3. Глибоке навчання

З розвитком глибокого навчання та, зокрема, нейронних мереж, з'явилося ще одне способу класифікації, оскільки алгоритми глибокого навчання можуть вирішувати проблему класифікації звуків легень без необхідності у будь-яких попередніх процедурах екстракції ознак. Вони можуть бути згенеровані самими мережами глибокого навчання, що зменшує ймовірність людської помилки та вибору неоптимальних ознак.

Останнім часом глибоке навчання та, зокрема, нейронні мережі витіснили традиційні методи машинного навчання у побудові автоматичних систем для класифікації звуків легень. Систематичний огляд Олусоли Абайомі-Аллі та спів. (2022) показав, що згорткові нейронні мережі використовувалися в 78,5% публікацій, тоді як варіанти рекурентних нейронних мереж (RNN) посіли друге місце з 16,1%. Для порівняння, традиційні методи машинного навчання були використані лише в 10,7% публікацій [42].

Серед різноманітного спектру штучних нейронних мереж широко використовуються згорткові нейронні мережі (CNN), що призначені для екстракції ознак та класифікації звуків легень. CNN - це спеціалізований тип нейронних мереж прямого зв'язку, здатний ефективно обробляти двовимірні вхідні дані. Він досягає цього за допомогою навчаного фільтра, який

автоматично витягує витончені ознаки з вхідних даних, сприяючи точному розрізненню між різними класами чи категоріями. Ця здатність робить CNN особливо підходящими для завдань, пов'язаних з обробкою зображень або спектрограм, як це бачимо в аналізі звуків легень.

Згорткові нейронні мережі отримали свою назву від математичної операції багатовимірної дискретної згортки, яка визначається формулою:

$$y(n_1, n_2) = (n_1, n_2) * w(n_1, n_2) = \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} x(n_1, n_2) w(n_1 - k_1, n_2 - k_2)$$

(1.3.2.)

де $x(n_1, n_2)$ - вхідне зображення, $w(n_1, n_2)$ - імпульсна відповідь фільтра, а $y(n_1, n_2)$ - вихідне зображення.

Зазвичай структура згорткової нейронної мережі (CNN) включає кілька прихованих шарів, включаючи згорткові, пулінгові, нормалізації пакетів, активації та повністю з'єднані (або щільні) шари. Згорткові шари є ключовими структурними компонентами CNN та відповідають за вилучення ознак з вхідних даних. Під час навчання кожен згортковий шар вивчає ваги фільтрів для створення карти ознак. Ці фільтри, також відомі як ядра, переміщуються по вхідним даним та виконують згортку для створення карт ознак.

Після згорткового шару зазвичай додається пулінговий шар. Пулінгові шари використовуються для зменшення розмірів карт ознак, зменшуючи їх просторові виміри та кількість параметрів в мережі. Це допомагає зменшити обчислювальну складність та контролювати перенавчання за рахунок захоплення найбільш важливих ознак. Крім того, зазвичай вставляють шари нормалізації пакетів для нормалізації активацій попереднього шару, що сприяє стабільності навчання та прискорює збіжність. Шари активації вводять нелінійність в мережу, дозволяючи їй вивчати складні взаємозв'язки між ознаками. Нарешті, повністю з'єднані шари використовуються для

з'єднання кожного нейрона в одному шарі з кожним нейроном в наступному шарі, що в кінцевому підсумку дає остаточний вихід мережі. Ці шари зазвичай виконують класифікаційні або регресійні завдання на основі вилучених ознак.

Шари нормалізації пакетів використовують пакети для нормалізації вхідних даних, що прискорює процес навчання. Шари виключення відповідають за запобігання перенавчання за рахунок вимкнення певних нейронів. У кінцевому підсумку, повністю з'єднані шари, що знаходяться в кінці мережі, утворюють прості прямі нейронні мережі. Вони характеризують накопичені активації попередніх шарів у розподілі ймовірностей класу, встановлюючи зважені зв'язки між ними.

У CNN до сигналу $x_n^0 = [x_1, x_2, \dots, x_n]$, застосовуються кілька скалярних добутків (згорток), де N - загальна кількість точок, обчислених за формулою:

$$c_n^{uj} = h_a \left(b_j + \sum_{m=1}^M w_m^j x_{n+m-1}^j \right) \quad (1.3.3)$$

де u - індекс шару, h_a - функція активації, b_j - зміщення j -го відображення ознак b , M - розмір ядра, w_m^j - вага відображення ознаки, a індекс фільтра - m [16].

У своєму дослідженні, Крістіна Джакме (2018) використовувала згорткові нейронні мережі (CNN) з метою покращення точності та надійності виявлення фаз дихального процесу, що є ключовим етапом у діагностиці респіраторних захворювань. Дослідження використовувало набір даних, що містить звуки дихання від осіб з різними респіраторними захворюваннями. Ефективність запропонованого методу на основі CNN було оцінено за допомогою кількох метрик, зокрема, точності, чутливості та специфічності. Результати показали, що метод на основі CNN перевершив традиційні

підходи з точності та стабільності, досягнувши точності на рівні 89%, чутливості - 87% та специфічності - 92%. Крім того, дослідження показало стійкість CNN до різних умов, таких як зміни в розподілі даних та наявність шуму. Ці висновки підкреслюють ефективність CNN у виявленні фаз дихального процесу у звуках легенів та підкреслюють їх потенціал для покращення продуктивності автоматизованих методів аналізу звуків легень [51].

Нейронна мережа зі залишковими зв'язками (ResNets) - це модель глибокого навчання, у якій зважені рівні навчаються залишковими функціями з посиленням на вхід рівня. Вона виправляє поширену проблему глибоких мереж, а саме те, що збільшення глибини мережі спочатку призводить до її поліпшення, а потім стрімко погіршується. Це дозволяє тренувати моделі глибокого навчання з десятками або сотнями шарів. Зазвичай ResNet складається з 152 рівнів і доповнюється залишковим навчанням, що скорочує час, необхідний для навчання мережі, хоча вона залишається дуже глибокою.

Рекурентні нейронні мережі (RNN) спеціально розроблені для обробки часових послідовностей даних та мають вбудовану залежність від часу. Це означає, що дані в один момент часу впливають на наступні дані. RNN включають в себе вхід з попередніх кроків часу та мають механізм "пам'яті", щоб зберігати минулу інформацію. Мережі з довгостроковою та короткостроковою пам'яттю (LSTM), також відомі як розширені мережі з короткостроковою пам'яттю, є спеціалізованою формою рекурентних нейронних мереж. Вони спрямовані на вирішення викликів, таких як зникнення градієнта та вибух помилок зворотного розповсюдження, що виникають при навчанні довгих послідовностей. Архітектура LSTM складається з "вхідних", "вихідних" та "забуткових" воріт, які контролюють потік інформації в пам'ять та з неї. Ці ворота дозволяють LSTM ефективно захоплювати та зберігати відповідну інформацію протягом довгих

послідовностей, що робить їх добре підходящими для завдань, що включають обробку послідовних даних.

Мережі LSTM можуть керувати інформацією як у прямому напрямку (однонапрямлені), так і у прямому та зворотньому напрямках (двонапрямлені). Остання конфігурація широко використовується в додатках аналізу сигналів, особливо там, де часові розгляди відіграють вирішальну роль у процесі навчання.

Огляд наявної літератури показує, що глибоке навчання випередило традиційні методи в ідентифікації хвороб та класифікації звуків легень. Численні публікації підтверджують перевагу ГН над МН.

У дослідженні, проведеному Бруно Рошею (2021), порівнювалася ефективність чотирьох алгоритмів для класифікації трьох типів подій (хрип/інше, скрип/інше, хрип або скрип/інше) фіксованої та змінної тривалості. Автор дослідив три класичні алгоритми машинного навчання, які часто використовуються для класифікації подій: лінійний дискримінантний аналіз (LDA), опорний векторний класифікатор з радіальною базовою функцією (SVMrbf), збалансоване випадкове дерево (RUSBoost) та згорткові нейронні мережі (CNN). У дослідженні розглядалися три різних моделі CNN:

1. Модель CNN з двома входами: Ця модель використовувала як вхідні дані спектрограму, так і мел-спектрограму.
2. Модель CNN з використанням лише спектрограми: Ця модель використовувала лише спектрограму як вхідні дані.
3. Модель CNN з використанням лише мел-спектрограми: Ця модель використовувала лише мел-спектрограму як вхідні дані.

Метою дослідження було оцінити ефективність цих алгоритмів у класифікації респіраторних подій з різною тривалістю та порівняти їхню ефективність у вирішенні завдання класифікації.

Серед традиційних класифікаторів, RUSBoost досяг найвищої точності 92.3% у завданні з подвійною класифікацією (хрип або скрип/інше). Проте

згорткові нейронні мережі перевершили традиційні класифікатори, з моделлю CNN з двома входами досягнувши точності 96.9%, модель на основі мел-спектрограми досягла 96.7%, а модель на основі звичайної спектрограми отримала точність 96.2%. Варто зазначити, що в твердовидільному наборі даних різниця в точності між методами була відносно невеликою, але вона значно збільшилася в наборі даних змінної тривалості. RUSBoost залишався найефективнішим традиційним алгоритмом з точністю 62.9%, але модель CNN з двома входами видалася з точністю 81.8% [50].

Висновки Бруно Роше (2021) збігаються з висновками Фенга Менга та спів. (2020), де штучні нейронні мережі продемонстрували найвищу точність класифікації на рівні 85.43% серед різних методів, включаючи опорні векторні машини (SVM) та класифікатор k-найближчих сусідів (KNN). У дослідженні Фенга Менга та спів. використовувалося 705 звуків легенів, з яких 240 були хрипами, 260 - свистами та 205 - нормальними диханнями, зібраними від 130 пацієнтів. Вектори ознак формувалися за допомогою відносної енергії вейвлету, ентропії вейвлету та вимірів схожості між оригінальним сигналом та вейвлет-підсигналами. Обидва дослідження підкреслюють ефективність штучних нейронних мереж, особливо в контексті класифікації респіраторних подій за даними про звуки легенів. Ці висновки підкреслюють потенціал підходів на основі нейронних мереж для точного та надійного аналізу респіраторних звуків, що сприяє поліпшенню діагностики та контролю респіраторних захворювань [30].

На відміну від традиційних класифікаторів машинного навчання (MN), нейронні мережі проявляють вищу продуктивність у завданнях мультикласової класифікації. Марія Гарсія-Ордас та спів. (2020) дослідили згорткові нейронні мережі (CNNs) у двох відмінних завданнях: перше полягало в виявленні хронічних, нехронічних та нормальних звуків дихання, тоді як друге мало на меті відрізнити ці патології. У початковому експерименті результати показали високу чутливість на рівні 0.991 та точність

на рівні 0.994. Подальше, в більш складному експерименті, що включав п'ять відмінних класів, CNNs продовжили доставляти перспективні результати, досягаючи чутливості на рівні 0.988 та специфічності на рівні 0.986 [35].

Хай Чен та спів. (2019) розробили техніку виявлення та класифікації респіраторних шумів з використанням оптимізованого S-перетворення разом з глибокими нейронними мережами залишкових (ResNets). Їх експериментальні результати показали, що ResNets відзначаються в мультикласифікації звуків дихання, досягаючи вражаючих показників точності, чутливості та специфічності до 98.79%, 96.27% та 100% відповідно [31].

1.3.4. Ансамблеве навчання

Останнім часом методи ансамблевого навчання все більше використовуються для розробки автоматизованих систем для багатокласової класифікації дихальних звуків. Ансамблі або гібридні методи поєднують два різних типи алгоритмів глибокого навчання в одній архітектурі. Ці системи здатні значно підвищити ефективність класифікації і можуть бути застосовані не лише для класифікації окремих звуків (свистіння, хрипи тощо), але й для діагностики широкого спектру респіраторних захворювань.

Фу-Шун Хсу та його колеги (2021) провели порівняльну оцінку восьми архітектур нейронних мереж, включаючи довгострокову пам'ять (LSTM), керовані рекурентні одиниці (GRU), двосторонню довгострокову пам'ять (BiLSTM), двосторонні рекурентні одиниці (BiGRU), а також комбінації згорткових нейронних мереж (CNN) з LSTM, GRU, BiLSTM та BiGRU моделями. Їхнім завданням було виявлення фаз дихання та виявлення непередбачуваних звуків легенів. Дослідження показало, що ці моделі проявляють задовільну ефективність у аналізі звуків легенів. Особливо GRU-базові моделі виявили перевагу порівняно з LSTM-базовими моделями у більшості визначених завдань, як показано за показниками F1 та площею

під кривою характеристики робочої характеристики (AUROC). Крім того, двосторонні моделі постійно перевершували свої односторонні аналоги.

Крім того, включення згорткових нейронних мереж до ансамблів помітно підвищує точність класифікації випадкових звуків легенів. Ці висновки підкреслюють ефективність архітектур нейронних мереж, особливо моделей на основі GRU та двосторонніх архітектур, у аналізі звуків легенів та визначенні фаз дихання [52].

Георгіос Петмазас та його колеги (2022) використовували гібридну мережу CNN-LSTM з функцією втрати фокусу для автоматичної класифікації звуків легенів на 4 класи. Вони обробляють дихальні цикли з набору даних ICBHI 2017 в зображення спектрограм та передавали їх через згорткову нейронну мережу (CNN), щоб видобути важливі ознаки. Потім ці ознаки вводилися в мережу довгострокової пам'яті (LSTM), яка захоплювала та зберігала взаємозв'язки на довгий термін між ними. Щоб вирішити проблему нерівноваги даних у наборі даних, автори використали функцію втрати фокусу. Ця функція втрати спрямована на зменшення впливу нерівності класів, покараючи помилки класифікації меншість класів важче, тим самим підвищуючи ефективність моделі класифікації.

Оцінка помилок передбачення слугувала для оцінки ефективності моделі. Використовуючи цей гібридний підхід, який поєднує мережі CNN та LSTM з функцією втрати фокусу, дослідження мало на меті покращити точність та надійність автоматичної класифікації звуків легенів, особливо в умовах нерівноваги класів у наборі даних [32].

Алі Мохаммад Алкуда та його колеги (2022) провели порівняльний аналіз гібридного ансамблю CNN-LSTM поряд з індивідуальними моделями згорткової нейронної мережі (CNN) та моделями довгострокової пам'яті (LSTM). Вони використовували доповнені та недоповнені дані з двох наборів даних, ICBHI 2017 та KAUI, для оцінки. Висновки дослідження показали, що гібридна модель CNN-LSTM постійно перевершувала моделі, що

базуються виключно на моделях CNN або LSTM, на всіх наборах даних. Зокрема, модель CNN-LSTM досягла точності 99,8% без доповнення та 100% з доповненням. Ці результати підкреслюють ефективність підходу на основі ансамблю CNN-LSTM у завданнях класифікації звуків легенів, демонструючи її здатність досягати високої точності та надійної ефективності, особливо при роботі з різноманітними наборами даних та можливими нерівновагами класів [42].

У своєму дослідженні з назвою "Автоматична ідентифікація респіраторних захворювань за допомогою сигналів звуку стетоскопа легенів з використанням ансамблевих класифікаторів" Луай Фраїван та його колеги (2020) вперше використали ансамбль класифікаторів. Цей ансамблевий класифікатор складався з кількох окремих класифікаторів, таких як машини опорних векторів, дерева рішень і випадкові ліси. Їхній підхід спрямований на автоматичну ідентифікацію респіраторних захворювань на основі записів звуків легень. Основна передумова цього підходу полягає в тому, що поєднання передбачень з декількох класифікаторів підвищує точність та надійність діагностики захворювань. Дослідження використовувало набір даних, який містив записи звуків легень від осіб із різними респіраторними захворюваннями, що дозволило провести комплексний аналіз та оцінку запропонованого методу ансамблевого класифікатора.

Результати дослідження показали, що запропонований ансамбль класифікаторів значно підвищив точність та надійність автоматичної ідентифікації респіраторних захворювань, пропонуючи перспективне рішення для цієї складної проблеми. Перевершуючи окремі класифікатори за декількома метриками, включаючи правильність, чутливість та специфічність, ансамбль класифікаторів досяг вражаючих показників відповідно 95,2%, 95,8% та 94,5%. Крім того, дослідження підкреслює підвищену стійкість ансамблю класифікаторів за різних умов, особливо в умовах зміни розподілу даних та наявності шуму, у порівнянні з окремими класифікаторами. Це

підкреслює стійкість та ефективність ансамблевого підходу у вирішенні проблем, що властиві ідентифікації респіраторних захворювань за допомогою записів звуків легень [53].

У 2022 році Мохаммад Фраїван та його колеги використали гібридний підхід, поєднуючи згорткову нейронну мережу (CNN) і двонапрямлену короткочасну пам'ять (BiLSTM) з оптимізацією стохастичним градієнтним спуском з моментумом. У цьому підході CNN використовувалася для виділення ознак, що стосуються загального просторового виміру сигналу, тоді як BiLSTM захоплювала характеристики, що відповідають варіаціям у часовому домені.

Інтеграція цих двох мереж зазвичай призводить до покращення результативності при тренуванні моделей для передбачення сигналів на основі їх просторових та часових характеристик. Використовуючи як просторову, так і часову інформацію, комбінована модель CNN-BiLSTM досягла видатної середньої точності у 99,62% при класифікації різних респіраторних захворювань. Ця помітна точність демонструє ефективність використання гібридних моделей CNN-BiLSTM для аналізу респіраторних сигналів та ідентифікації респіраторних захворювань. Захоплюючи як просторові, так і часові ознаки, цей підхід поліпшує здатність моделі розуміти складні шаблони у даних та робити точні прогнози [16].

Підводячи підсумок результатів усіх вищезгаданих досліджень, можна зробити висновок, що ансамблеві класифікатори мають значний потенціал для створення автоматизованих систем для класифікації звуків легень та ідентифікації респіраторних захворювань. Використання різних комбінацій класифікаторів, різних методів виділення ознак та різних наборів даних показує, що ансамблеві класифікатори можуть забезпечувати підвищену точність та надійність таких систем.

1.3.5. Метрики оцінки

Матриця невідповідностей - основний інструмент для оцінки ефективності алгоритмів класифікації (рис. 1.3.1), у якій отримані результати поділяються на чотири категорії:

- Істинно-позитивні (True positives) - правильно класифіковані події відповідного класу.
- Істинно-негативні (True negatives) - правильно класифіковані події інших класів.
- Хибно-позитивні (False positives) - помилка, в результаті якої негативний клас був визнаний як позитивний (помилка класифікації типу I).
- Хибно-негативні (False negatives) - помилка, в результаті якої позитивний клас був визнаний як негативний (помилка класифікації типу II).



Рис. 1.3.1. Матриця невідповідностей

На основі матриці плутанини обчислюються різноманітні показники продуктивності алгоритму:

Чутливість:

$$\text{Чутливість} = \frac{TN}{TN+FP} \quad (1.3.5)$$

Міра правильності (англійською - Accuracy, A), яка відображає відсоток помилок:

$$A = \frac{TP+TN}{TP+TN+FP+FN} \quad (1.3.6)$$

Міра точності (англійською: Precision, P) характеризує, скільки позитивних відповідей, отриманих від класифікатора, є правильними і визначається тією ж формулою, що й чутливість:

$$P = \frac{TP}{TP+FP} \quad (1.3.7)$$

Міра повноти (англійською - Recall, R) характеризує здатність класифікатора «вгадати» якнайбільшу кількість позитивних відповідей серед очікуваних, і хибно позитивні відповіді не впливають на цей показник в жодний спосіб.

$$R = \frac{TP}{TP+FN} \quad (1.3.8)$$

Набір параметрів P і R надає досить комплексну характеристику класифікатора. Вони взаємопов'язані, оскільки зі збільшенням значення R через збільшення кількості хибно-позитивних відповідей точність P зменшується, і зі збільшенням точності індикатор R знижується, оскільки певна кількість правильних відповідей «вибуває». Саме тому для характеристики класифікатора часто використовується метрика F1 (англійською - F1-score), яка поєднує точність і повноту, представляючи середнє гармонійне значення P і R:

$$F_1 = 2 * \frac{P*R}{P+R} \quad (1.3.9)$$

Коефіцієнт кореляції Метьюза (MCC) є більш стійким статистичним показником, який надає високий бал тільки в тому випадку, якщо прогностичний засіб працює добре у всіх чотирьох категоріях матриці плутанини, пропорційно як розміру позитивних елементів, так і розміру негативних елементів у наборі даних.

$$MCC = \frac{TP*TN-FP*FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (1.3.10)$$

Кілька досліджень також використовують аналіз ROC для оцінки якості бінарної класифікації.

Характеристична крива ROC (крива ROC) - це графік TPR проти FPR, де TPR та FPR є відповідно пропорціями істинно-позитивних та істинно-негативних класифікацій, які можуть бути записані у формулах:

$$TPT(t) = \int_t^{\infty} P_1(t)dt \quad (1.3.11),$$

$$FPT(t) = \int_t^{\infty} P_0(t)dt \quad (1.3.12),$$

де $P_1(t)$ та $P_0(t)$ - функції розподілу ймовірностей.

Під час побудови кривої ROC на осі y відображається TPR(t), а на осі x - FPR(t), отримані при різних значеннях параметра t. Зазначимо, що цей параметр можна використовувати лише для оцінки якості бінарної класифікації. Для застосування оцінки ROC для класифікаторів декількох класів вихідні дані зводяться до бінарних значень, для яких класифікатор, налаштований для кожного класу, визначає цей клас як позитивний, а всі інші як негативний. На рисунку 1.3.2 показано приклад кривих ROC.

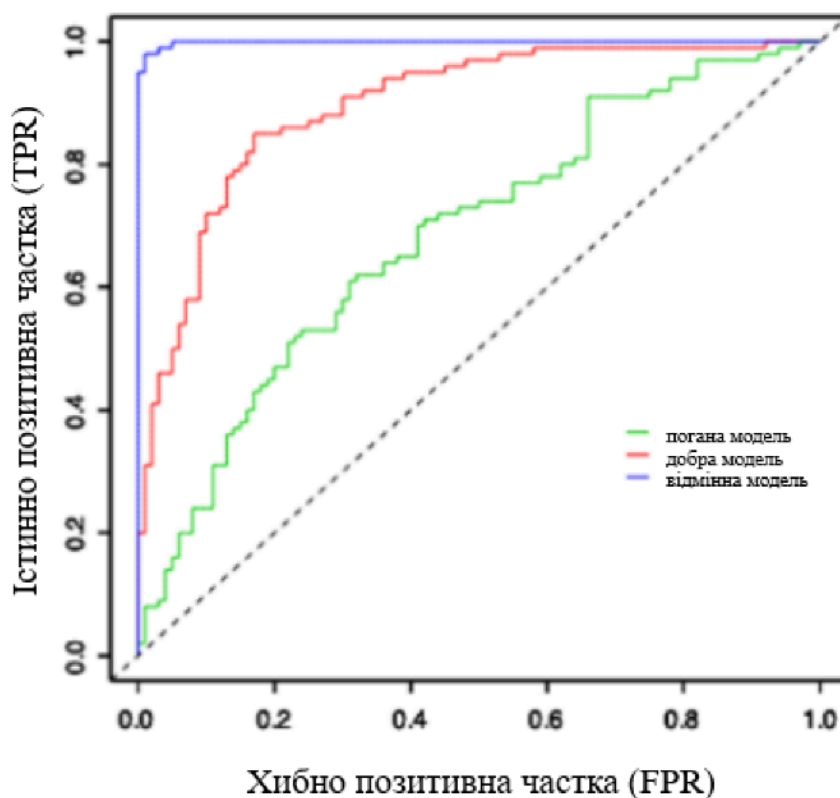


Рис. 1.3.2. Приклади кривих ROC

Кількісна інтерпретація кривої ROC полягає у її площі під кривою (AUC) - загальній характеристиці якості класифікації, яка не залежить від співвідношення цін на помилки, тому цей показник часто використовується для порівняння кількох моделей між собою. Якщо AUC дорівнює 0,5, класифікатор вважається непридатним, і чим вище AUC, тим краще.

Ілюстрація використання матриці плутанини для оцінки ефективності системи автоматичної класифікації звуків легень демонструється у дослідженні Мохаммада Фраївана та спів. (2022). Дослідники оцінювали згорткову нейромережу, двонаправлену нейромережу з довгостроковою пам'яттю та їх гібрид для класифікації звуків легень з бази даних ICBNI. Порівняння цих класифікаторів передбачало використання матриці плутанини, зображеної на малюнку 1.3.2.

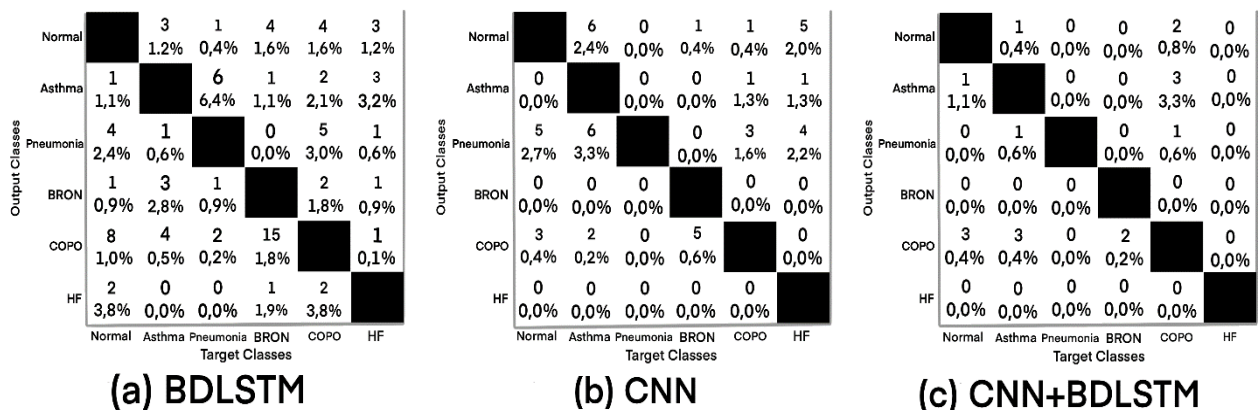


Рис. 1.3.2. Матриця плутанини для різних класифікаторів, складена М. Фраїваном та його колегами (2022)

Результати показали, що комбінація CNN і BDLSTM досягла найвищої точності в класифікації звуків легенів з показником $F1=0,9845$ [16].

Систематичний огляд Олусола Абаомі-Аллі та спів. (2022) визначив, що найбільш поширеними мірами оцінки продуктивності класифікаторів в автоматизованих системах класифікації звуків дихання є правильність, метрика F1, повнота та точність. Інші показники використовуються рідше [42].

Дослідження Аннамарії Мезарос (2016) буде спрямоване на оцінку систем виявлення поліфонічних звукових подій, що є важливим завданням у галузі аудіоаналізу, оскільки це засіб кількісної оцінки ефективності цих систем, що дозволяє порівняти їх ефективність.

У ньому автор запропонував декілька метрик для оцінки поліфонічної продуктивності системи виявлення звукових подій, включаючи правильність (1.3.7), повноту (1.3.8) та міру F1 (1.3.9). Вона також запропонувала використовувати такий показник, як швидкість помилок.

У сегменті k , кількість помилок $S(k)$, яка вказує на кількість реальних подій, для яких не було зроблено правильних передбачень, визначається шляхом поєднання хибних позитивів та хибних негативів. Залишкові помилки складаються з вставок та видалень: $D(k)$ представляє кількість посилян на події, які не були правильно визначені (де помилкові негативи після заміщень розглядаються), а $I(k)$ представляє кількість подій у вихідних даних системи, які є неправильними. Ці параметри можуть бути представлені наступними формулами:

$$\begin{aligned} S(k) &= (FN(k), FP(k)) \\ D(k) &= (0, FN(k) - FP(k)) \\ I(k) &= (0, FP(k) - FN(k)) \end{aligned} \quad (1.3.13)$$

Загальний коефіцієнт помилок обчислюється шляхом інтегрування підрахунків по сегментах з загальною кількістю сегментів K , де $A(k)$ представляє кількість аудіо подій, позначених як активні у сегменті k . Загальний коефіцієнт помилок ER можна обчислити за формулою:

$$ER = \frac{\sum_{k=1}^K S(k) + \sum_{k=1}^K D(k) + \sum_{k=1}^K I(k)}{\sum_{k=1}^K N(k)} \quad (1.3.14)$$

Порівнюючи результати всіх метрик та їх комбінацій за використання реальних наборів даних, дослідники зробили висновок, що метрика F1-виміру є найбільш придатною для оцінки продуктивності систем виявлення поліфонічних звукових подій, оскільки вона забезпечує баланс між точністю та повнотою та менше чутлива до змін у кількості виявлених звукових подій. Метрика F1-виміру коливається від 0 до 1, де 1 представляє ідеальну продуктивність, а 0 - найгіршу можливу продуктивність. Автори також зробили висновок, що вибір метрики оцінки має значний вплив на

продуктивність систем виявлення поліфонічних звукових подій та що важливо враховувати кілька метрик продуктивності під час оцінки.

В цілому, систематичний огляд, проведений Олусолою Абаїомі-Аллі та спів. (2022), визначив, що найбільш часто використовувані показники продуктивності класифікаторів у автоматизованих системах класифікації респіраторних звуків - це правильність, метрика F1, повнота та точність. Інші метрики використовуються менше частіше.

2. Матеріали та методи

2.1. Набір даних

Джерелом даних для нашого дослідження був відкритий набір даних Міжнародної конференції з біомедичної та медичної інформатики (International Conference on Biomedical and Health Informatics, ICBHI) 2017 [6]. На момент дослідження цей набір даних був одним із найбільших загальнодоступних джерел аудіозаписів аускультатії легень.

База даних ICBHI 2017 містить 920 аудіофайлів, зібраних двома незалежними дослідницькими групами в Португалії та Греції від 126 здорових і хворих осіб, демографічні характеристики яких наведені на Рисунку 2.1.1.

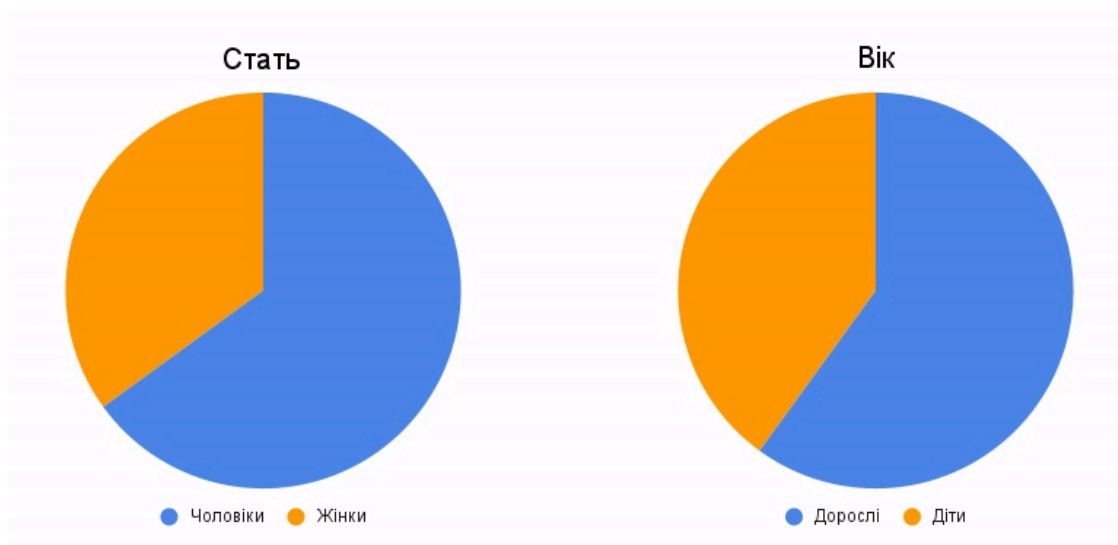


Рис. 2.1.1. Демографічні характеристики набору даних Міжнародної конференції з біомедичної та медичної інформатики

Зразки у базі даних ICBHI 2017 професійно анотовані двома способами:

1. Згідно з патологічним станом пацієнта: Ця анотація надає інформацію про загальний стан здоров'я або конкретні умови пацієнта, з якого зібрані звуки легенів.
2. Згідно з наявністю порушень дихання в кожному дихальному циклі: Ця анотація описує наявність або відсутність порушень дихання в кожному окремому дихальному циклі. Кожен дихальний цикл складається з кількох записів звуків легенів і характеризується випадковими мітками класифікації звуків легенів.

Кожен з 920 записів у базі даних представляє звуки легенів, зібрані від одного пацієнта. Ці записи містять кілька дихальних циклів, кожен зі своїми власними анотаціями на основі зазначених вище критеріїв.

У базі даних ІСВНІ 2017 є 3642 записи "нормальних циклів", 1864 цикли з "тріском", 886 циклів з "хрипами" та 506 циклів з "тріском та хрипами" (Таблиця 2.1.1.). Загальна кількість циклів становить 6898.

Таблиця 2.1.1.

Дихальні цикли, що містяться в базі даних ІСВНІ.

| Тип циклу | Кількість |
|---------------------------|-----------|
| Цикл із тріском | 1864 |
| Цикл із хрипами | 886 |
| Цикл із тріском і хрипами | 506 |
| Нормальний цикл | 3642 |
| Загальна кількість циклів | 6898 |

ІСВНІ містить зразки звуків легенів від пацієнтів із захворюваннями дихальних шляхів, такими як ХОЗЛ, пневмонія, інфекція верхніх

дихальних шляхів (ІВДШ), бронхоектатична хвороба, бронхіоліт, інфекція нижніх дихальних шляхів (ІНДШ), астма, а також від здорових осіб.

Крім 26 здорових осіб, розподіл пацієнтів із різними діагнозами дихальних захворювань показано на малюнку 2.1.2.

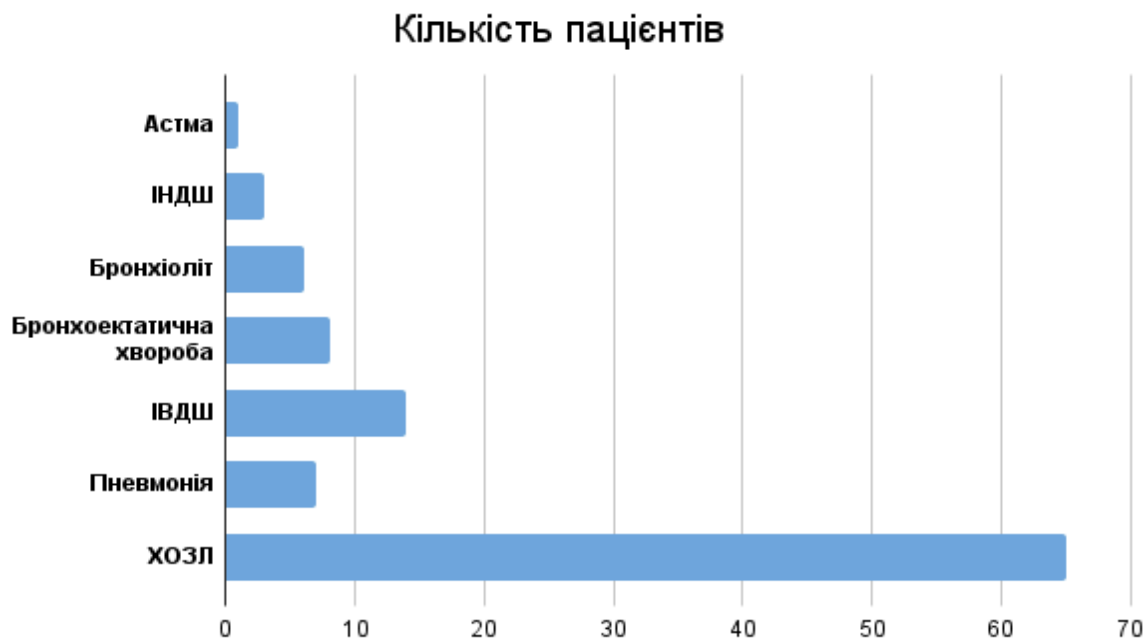


Рис. 2.1.2. Діагнози пацієнтів у наборі даних ІСВНІ 2017

Щодо окремих записів, що складають базу даних ІСВНІ 2017, їх розподіл за різними діагнозами показано на зображенні 2.1.3.

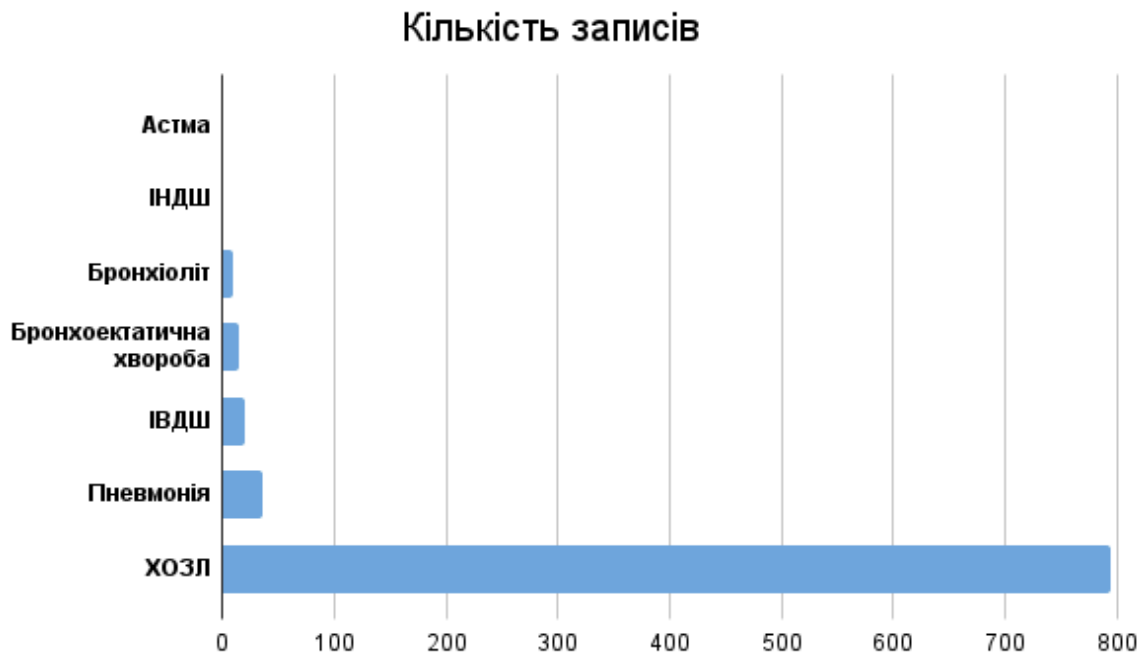


Рис. 2.1.3. Розподіл аудіозаписів звуків легенів за діагнозами пацієнтів у наборі даних ICBHI

Датасет ICBHI 2017 містить 5,5 годин аудіозаписів на різних частотах (4 кГц, 10 кГц і 44 кГц), тривалість яких коливається від 10 до 90 секунд. Дані аускультатії для кожного пацієнта були зібрані в різних місцях та за допомогою різних типів запису, зокрема, використовувалися мікрофон AKG C417L, стетоскоп 3M Littmann Classic II SE, електронний стетоскоп 3M Littmann 3200 і електронний стетоскоп Welch Allyn Meditron Master Elite.

Детальний аналіз записів датасету ICBHI, який буде описаний в наступному розділі, виключив аудіозаписи з астмою та ІНДШ як не характерні для цілей поточного дослідження, а також об'єднав два класи (бронхоектатичну хворобу і бронхіоліт) в один клас BRON через їх схожість. Таким чином, з бази даних ICBHI, яка спочатку включала вісім

класів і 920 аудіозаписів, для цього дослідження використовувався датасет, що складається з 917 записів і п'яти класів.

Датасет був розділений на дві частини: навчальний набір і тестовий набір. Для цього створено випадкову вибірку без повернення, яка включає 734 записи для навчального набору. Решта 20% записів утворюють тестовий набір.

2.2. Попередня обробка

Як і будь-які інші електронно записані природні звуки, записи аускультатії легенів містять різноманітні спотворення, додаткові шуми, звуки навколишнього середовища тощо. Крім того, оскільки набір даних ICBNI був зібраний з різних пацієнтів з декількох лікарень за різних умов запису та з використанням різного обладнання, частота дискретизації, рівень навколишнього шуму та кількість дихальних циклів у записах різняться. Аудіо також піддається впливу пристроїв, які використовувалися для запису, оскільки кожен має свою власну амплітудно-частотну відповідь. Все це безпосередньо впливає на якість інформаційного змісту зразка.

Набір даних ICBNI має різні частоти дискретизації від 4 кГц до 44,1 кГц. Зауваживши, що корисні сигнали знаходяться на частотах нижче 2000 Гц, всі записи були перевибрані до 4000 Гц.

Згідно з літературою, частотний діапазон, в якому можна локалізувати корисний шум легенів, становить 100-1500 Гц. Був застосований полосно-пропускний фільтр Баттерворта п'ятого порядку для виділення корисного діапазону, і додаткові частоти, які не несли корисної інформації, були відсічені.

Заснований на цілях і завданнях поточного дослідження, спектральні характеристики, отримані після попередньої стандартизації, були обрані для вибору ознак. Найбільш прийнятним виявилось використання

Мел-спектрограм, тобто візуальних зображень звуку за шкалою Мел, яка є перцептивною шкалою висот, яку люди, відповідно до свого психофізіологічного сприйняття, оцінюють як однаково віддалені один від одного.

Одиницею виміру між цією шкалою та звичайним вимірюванням частот є визначення звукового сприйняття на 1000 мел для тону на 1000 Гц, що перевищує поріг слухача на 40 дБ. Далі шкала змінюється нелінійно, оскільки особливості звукового сприйняття призводять до того, що, на думку слухачів, на частотах вище 500 Гц все більші інтервали створюють однакові висотні інтервали.

Вибір чалкових спектрограм для візуальної презентації даних базувався на кількох принципах. По-перше, огляд літератури показав, що ознаки, витягнуті з чалкових спектрограм, найбільш підходять для навчання згорткових нейронних мереж. По-друге, чалкова спектрограма зберігає більше інформації на низьких частотах, тоді як значення з більш широкого діапазону усереднюються на високих частотах. І, нарешті, третя перевага мел-спектрограми була визнана через переважання неперервних звуків легенів з мелодійними характеристиками, а не інтермітентними звуками.

Для генерації мел-спектрограм у поточному дослідженні використовувалися 2048 кадрів із кроком 256. Для отримання мел-спектрограм аудіозаписи були перетворені з області часу в область частоти за допомогою швидкого перетворення Фур'є, після чого шкала Гц була переведена в мел-шкалу, а амплітуда виражена за допомогою виміру кольору. Перерахування частот, представлених в герцах, в мел виконувалось згідно з формулою, наведеною в попередньому розділі (1.2.19).

Оскільки база даних ІСВНІ 2017 мала різну тривалість (від 10 до 90 секунд), були отримані спектрограми різних розмірів. Не рекомендується

використовувати їх у такому вигляді, оскільки подальші етапи дослідження, зокрема додавання даних за допомогою CVAE (згортковий варіаційний автоенкодер) і використання згорткових нейронних мереж, вимагали використання вхідних даних однакового розміру. Для вирішення цієї проблеми отримані чалкові спектрограми лінійно інтерполювалися до однакового розміру на основі середнього розміру отриманої спектрограми.

Отримані спектрограми були нормалізовані в діапазоні від 0 до 1 за допомогою мінімально-максимальної нормалізації:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad 2.2.1$$

2.3. Розширення даних

Як і в більшості загальнодоступних медичних наборах даних, в базі даних звуків легенів ICBNI 2017 спостерігається помітний дисбаланс, що чітко ілюструється на рисунках 2.1.2 і 2.1.3. Завдяки різній спеціалізації медичних установ, в яких проводяться записи, відмінній поширеності захворювань і специфіці їх перебігу (наприклад, верхні інфекції дихальних шляхів рідко потребують госпіталізації або навіть візиту до лікаря), майже в кожному наборі даних аускультатії легенів формується таким чином, що для одного захворювання з розглянутого є значно більше аудіозаписів, ніж для інших. Зокрема, в базі даних ICBNI 2017 чітко переважає ХОЗЛ, кількість пацієнтів з цим захворюванням становить 51% від загального числа обстежених осіб, а кількість зразків класу ХОЗЛ досягає 86,5% від загальної кількості аудіозаписів. Існує багато рішень для проблеми дисбалансу класів, які були детально обговорені в огляді літератури. Обмеження медичних даних та потреба в їх раціональному використанні не дозволяє використовувати методи перевибірki та недовибірki. З урахуванням розміру доступного набору даних, використання зважених цільових змінних також може бути неефективним. Тому найочевиднішим

варіантом для балансування набору даних є збагачення даними за допомогою методів генерації. Збагачення даними в часовому і частотному доменах може бути використано для задачі класифікації респіраторних захворювань. У статті пропонується метод, який використовує частотне представлення аускультатії легенів. В межах його обмежень згенеровані нормалізовані мел-спектрограми можуть розглядатися як одноканальні зображення фіксованого розміру, де кожен піксель представляє потужність спектра частот у певний момент часу. Чалкові спектрограми кожного класу мають спільні характеристики, які можуть бути використані для генерації синтетичних даних недорогих класів. Цей підхід є більш стійким до перевищення.

Для генерації синтетичних даних робота пропонує використовувати архітектуру автоенкодерів, зокрема згорткові варіаційні автоенкодери (CVAE). Автоенкодер складається з двох з'єднаних мереж: енкодера і декодера. Енкодер приймає вхідні дані і кодує їх характеристики в кодуючі вектори, тоді як декодер отримує вихідні характеристики знову з кодуючого вектора. Енкодер розроблений таким чином, що він продукує приховану змінну на вихід, і суттєво стискає форму вхідного сигналу. Між тим, декодер розроблений для розпакування прихованих змінних, щоб виконати зворотне перетворення вихідного сигналу.

Основна відмінність між VAE та автоенкодером полягає в способі отримання представлення прихованих атрибутів. Автоенкодер представляє приховані атрибути як функцію, тоді як VAE робить це, обчислюючи ймовірнісний розподіл. Таким чином, варіаційні автоенкодери можуть бути позиціоновані як ймовірнісна модель, яка має на меті симуляцію розподілу даних для подальшої генерації синтетичних даних, які можуть належати до початкового розподілу. Отже, на свою сутність, VAE має генеративну властивість.

Ще одна фундаментально унікальна властивість VAE, яка відрізняє їх від інших, зокрема звичайних автоенкодерів, і робить їх корисними для генеративного моделювання, полягає в тому, що їхні приховані простори є неперервними, що дозволяє легке випадкове вибіркве відбір і інтерполяцію.

Зворотно, згорткові варіаційні автоенкодери відрізняються тим, що згорткові шари використовуються як для виходу, так і для виходу для формування та наступної кодування в прихований (латентний) простір просторових характеристик та наступного декодування просторового тензору. Основна схема моделі CVA:

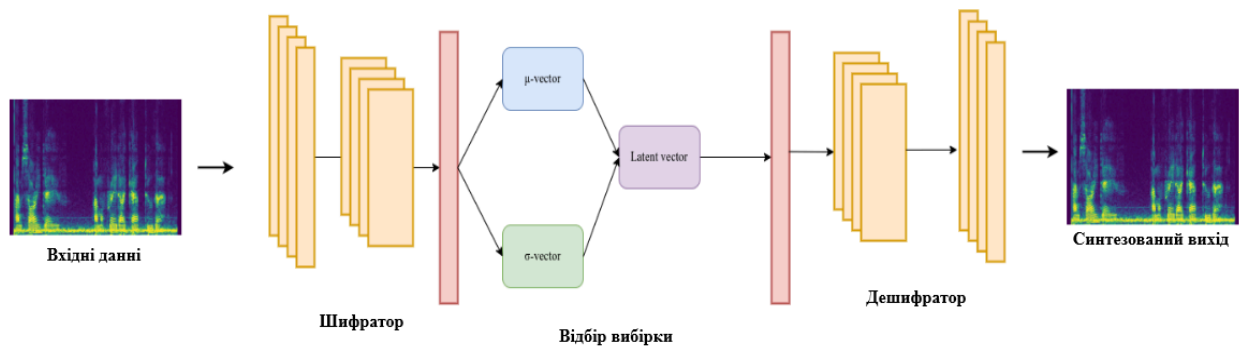


Рис. 2.3.1. Структура та схема роботи згорткового варіаційного автоенкодера

Під час тренування вхідне зображення X (Input) представлено двома латентними змінними Z_μ та Z_σ (вектор μ та вектор σ), з яких вибирається латентний вектор і подається на вхід декодера для виведення зображення. Однак ця стохастична операція вибору робить латентний простір випадковим вузлом, оскільки градієнти не можуть проходити через шар вибірки, і параметри Z_μ та Z_σ не можуть бути вивчені. Для того щоб ітеративно перетинати градієнти та застосовувати правило ланцюжка у зворотному розповсюдженні, вузол повинен бути детермінованим. Для цього у варіаційному автоенкодері використовується трюк з

перепараметризації, який перетворює випадковий вузол у детермінований, залишаючи вектор μ та вектор σ як параметри мережі, які можна вивчити, при цьому зберігаючи стохастичність всієї системи через функцію ϵ .

Замість вектора вибірки з нормального розподілу з параметрами μ та σ , який виражений як $Z \sim N(Z_\mu + Z_\sigma^2)$ та не дозволяє обчислення градієнтів, вектор виражений як сума фіксованого середнього вектора μ та фіксованого вектора стандартного відхилення масштабується випадковими константами, витягнутими з одиничного гауссів розподілу. Ця операція вибірки може бути записана наступним чином:

$$Z = Z_\mu + Z_\sigma^2 \epsilon, \quad 2.3.1$$

де $\epsilon \sim N(0,1)$, а \odot - це поелементне множення. Принцип трюку з параметризації показаний детально на Рисунку 2.3.2.

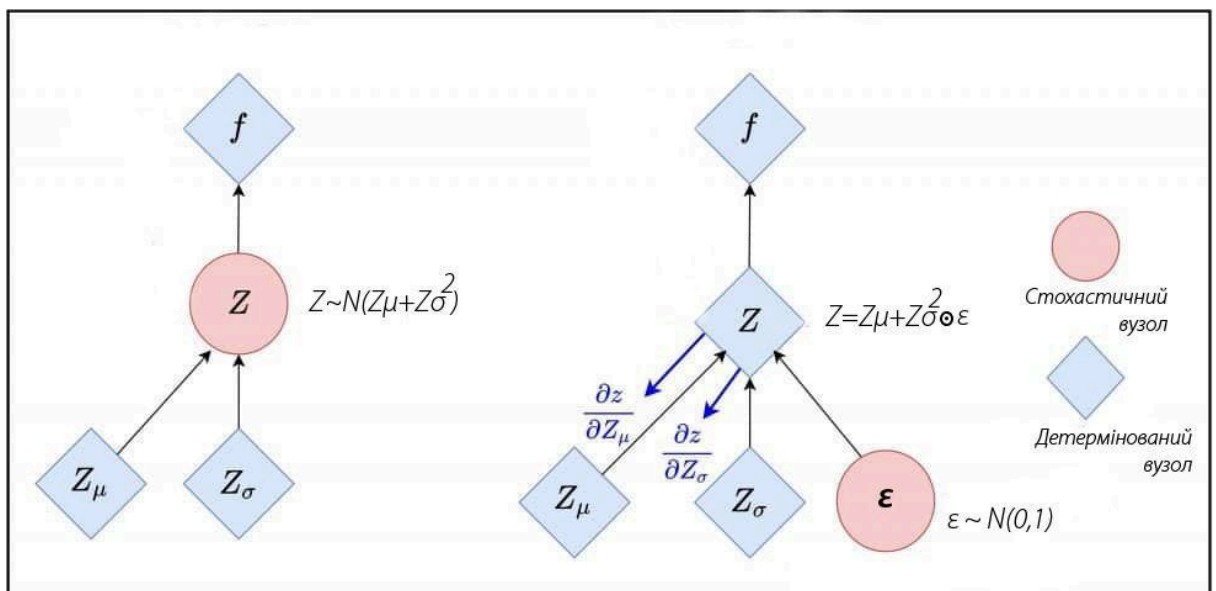


Рис. 2.3.2. Трюк параметризації згорткового варіаційного автоенкодера

Функція втрат варіаційного автоенкодера (VAE) складається з двох зважених частин: реконструкція, яка відповідає за умовну схожість

вхідного тензора із вихідним тензором; регуляризація, яка зменшує розподіл латентного простору, створеного кодерами, до нормального розподілу.

Для визначення реконструкції використовувалась середньоквадратична помилка (MSE), а для регуляризації - дивергенція Кульбака-Лейблера. Функція втрат VAE була визначена за формулами 1.3.1 та 1.3.2.

Таким чином, у поточному дослідженні згортковий варіаційний автоенкодер синтезував дані з кожного недостатньо представленого класу, кодуючи нормалізовані мел-спектрограми в нормально розподілений латентний простір, а потім декодуючи вектори у зображення, які можуть належати розподілу зображень цього класу.

2.4. Тренування і валідація

Згорткові нейронні мережі є поширеним методом аналізу просторових характеристик зображення, які у поточній роботі представлені мел-спектрограмами звуків легень і узагальненням даних, отриманих з них. Застосування CNN не потребує попереднього етапу вручного вилучення ознак, оскільки цей алгоритм глибокого навчання автоматично видобуває і обробляє складні залежності даних з двовимірних представлень за допомогою наборів ядер.

У цифрових зображеннях значення пікселів зберігаються у двовимірній сітці, тобто у масиві чисел. У згортковій нейронній мережі (CNN) невелика сітка параметрів, відома як ядро, служить оптимізованим вилучачем ознак. Це ядро застосовується до кожної позиції на вхідному зображенні для вилучення ознак, які є відповідними для даної задачі. При проходженні вхідного зображення через послідовні шари мережі вилучені ознаки стають все складнішими. Під час тренування оптимізаційний процес цих ядер відбувається для мінімізації різниці між виходами мережі

та істинними мітками. Цей процес, часто називається тренуванням, включає оновлення параметрів мережі, включаючи ядра, за допомогою алгоритму оптимізації, такого як зворотне поширення або градієнтний спуск. Мета полягає у ітеративному налаштуванні параметрів для покращення здатності мережі точно класифікувати або прогнозувати на основі вхідних даних. Через цей ітеративний оптимізаційний процес CNN навчається видобувати значущі ознаки з вхідних даних і робити точні прогнози або класифікації.

У цьому дослідженні використовується стандартна архітектура згорткової нейронної мережі, показана на Рисунку 2.4.1.

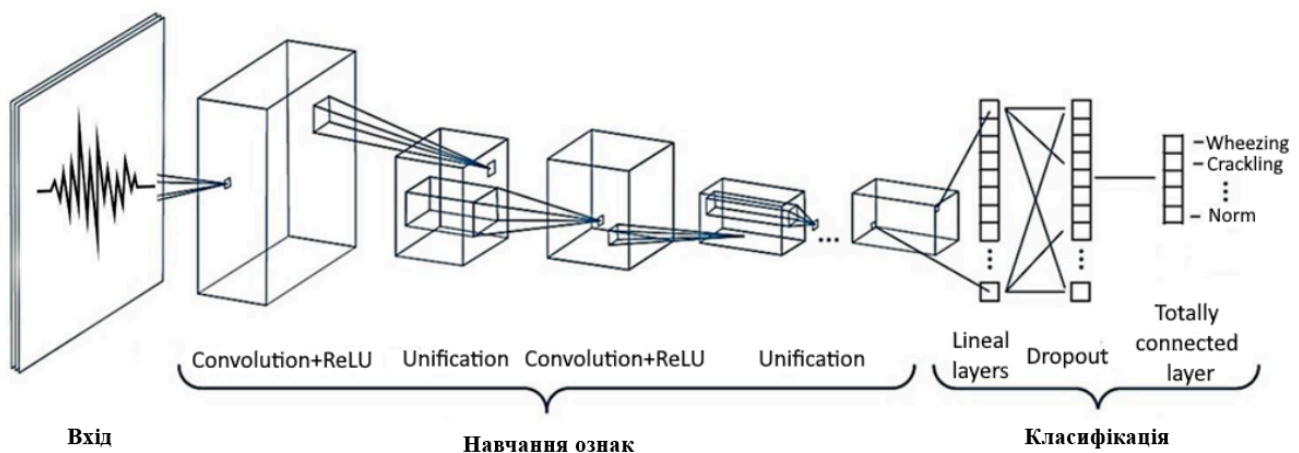


Рис. 2.4.1. Архітектура згорткової нейронної мережі

Ця архітектура включає використання звичайних згорткових шарів, що межують із шарами пулінгу.

Для видобуття ознак згорткові нейронні мережі (CNN) використовують спеціалізований тип лінійної операції, відомий як згортка. У цій операції ядро, представлене малим масивом чисел, застосовується до вхідних даних, які також представлені як масив чисел (тензор). На кожній позиції всередині вхідного тензора обчислюється добуток кожного елемента у ядрі і відповідного елемента у вхідному тензорі. Це обчислення

виконується на всіх позиціях у тензорі, що призводить до отримання карти ознак. Фактично згорткова операція систематично сканує вхідний тензор за допомогою ядра, виконуючи локалізовані множення і агрегації для видобуття ознак, важливих для поточної задачі. Цей процес допомагає CNN захоплювати просторові ієрархії і шаблони, присутні во вхідних даних, що сприяє завданням, таким як визначення зображень, виявлення об'єктів та іншим.

Згорткові шари узагальнюють просторові характеристики спектрограми і виявляють складні залежності в даних, в той час як шари злиття забезпечують комбінацію знайдених ознак і перехід до вищого рівня аналізу. Шари пулінгу виконують стандартну операцію зменшення вимірності всередині площини карти ознак. Це вводить інваріантність перекладу до невеликих зміщень і спотворень та допомагає зменшити кількість параметрів, які доступні для навчання.

Для забезпечення стабільності оптимізаційного алгоритму нейронної мережі кожен згортковий шар зазвичай слідує за шаром нормалізації партій (batch normalization). Нормалізація партій допомагає стабілізувати поширення градієнтів, забезпечуючи, що ваги в кожному шарі мережі під час тренування підтримують нормальний розподіл. Додатково, нелінійна функція ReLU (прямої лінійної одиниці) часто використовується для розширення простору гіпотез. Після лінійної згорткової операції вихідні дані проходять через активаційну функцію ReLU. Ця функція 2.4.1, позначена як $\text{ReLU}(x)$, визначається наступним чином:

$$f(x) = \max(0, x) \quad 2.4.1$$

Після досягнення необхідного рівня узагальнення характеристик вихідні дані згорткової нейронної мережі перетворюються у лінійний простір і обробляються лінійними шарами. Подальше простір лінійних характеристик скорочується до повністю зв'язаного шару, кількість

нейронів якого відповідає кількості класів, конкретно 5 у нашому випадку. Цей шар обчислює ймовірність того, що тензор вхідних даних належить кожному класу.

Під час експерименту також були додані Dropout шари з параметром $p = 0.3$ між лінійними шарами для запобігання перенавчанню. Dropout є поширеним методом регуляризації штучних нейронних мереж, призначеним для запобігання перенавчанню і підвищення швидкості навчання. Його суть полягає в тому, що під час навчання загальної мережі випадковим чином вибирається підмережа, для якої виконується навчання. Після навчання обраної підмережі випадковим чином вибирається нова підмережа, і навчання продовжується. Вибір нейронів для підмережі відбувається випадковим чином з ймовірністю, яку називають коефіцієнтом Dropout, причому навченіші нейрони отримують більший ваговий коефіцієнт у мережі.

Для задачі багатокласової класифікації була обрана функція втрат кросс-ентропії, яка працює з початковими ймовірностями моделі.

$$L = - \sum_{c=1}^M y_{o,c} \log \log (p_{o,c}) \quad 2.4.2$$

У цьому випадку M - це кількість класів, y - бінарний індикатор, що дорівнює 0 або 1, якщо мітка класу c є правильним класифікаційним значенням для спостереження o , p - передбачена ймовірність того, що спостереження o належить класу c .

Для оптимізації ваг використовувався алгоритм оптимізації Adam з коефіцієнтом навчання $l = 0.0001$. Adam є одним з найефективніших алгоритмів оптимізації для тренування нейронних мереж. [56]

Оптимізатор Adam є розширенням алгоритму стохастичного градієнтного спуску (SGD), що пропонує адаптивне оновлення швидкості навчання. Він використовує середні значення перших і других моментів

градієнтів для оновлення параметрів моделі. Зокрема, алгоритм обчислює експоненційно згладжене середнє значення градієнту (середнє значення градієнтів) і квадратичний градієнт (дисперсія градієнтів), контролюючи швидкість згладжування цих експоненційних середніх. У порівнянні з іншими існуючими оптимізаторами, такими як RMSprop, SGD, ADAGrad і NAdam, Adam забезпечує швидше збіжність разом із оптимізованими параметрами продуктивності. Його адаптивний механізм швидкості навчання дозволяє ефективно оптимізувати шляхом динамічного адаптування швидкості навчання в залежності від характеристик градієнтів, що призводить до покращення швидкості збіжності і загальної продуктивності.

Додатково, для кращої збіжності мережі використовувався планувальник швидкості навчання з параметром $\gamma = 0.9$. Навчання проводилося після 10-кратної перехресної перевірки для забезпечення охоплення всіх можливих комбінацій у наборі даних. Як метрики оцінки використовувалась матриця невідповідностей, а також значення чутливості (формула 1.3.4), специфічності (формула 1.3.5) розробленого методу, F1-оцінка (формула 1.3.9) і коефіцієнт Коена, обчислений за формулою:

$$\kappa = \frac{P_o - P_c}{1 - P_c} \quad 2.4.3$$

де P_o — спостережена згода і P_c — очікувана випадкова згода.

2.5. Метод підвищення ефективності діагностичних систем в аналізі звуку на основі нейронних мереж

У цьому пункті пропонується метод підвищення ефективності автоматизованого аналізу легеневих звуків, що поєднує підготовку та попередню обробку аудіоданих, синтетичне збагачення набору записів на основі згорткового варіаційного автоенкодера (CVAE) та глибинну

класифікацію (CNN) з контролем ризику «витоку даних» і крос-валідацією. У подальшому цей підхід іменується NASCA-метод (Neural-Augmented Sound Classification for Auscultation). Схему NASCA-методу ілюструє рисунок 2.5.1.

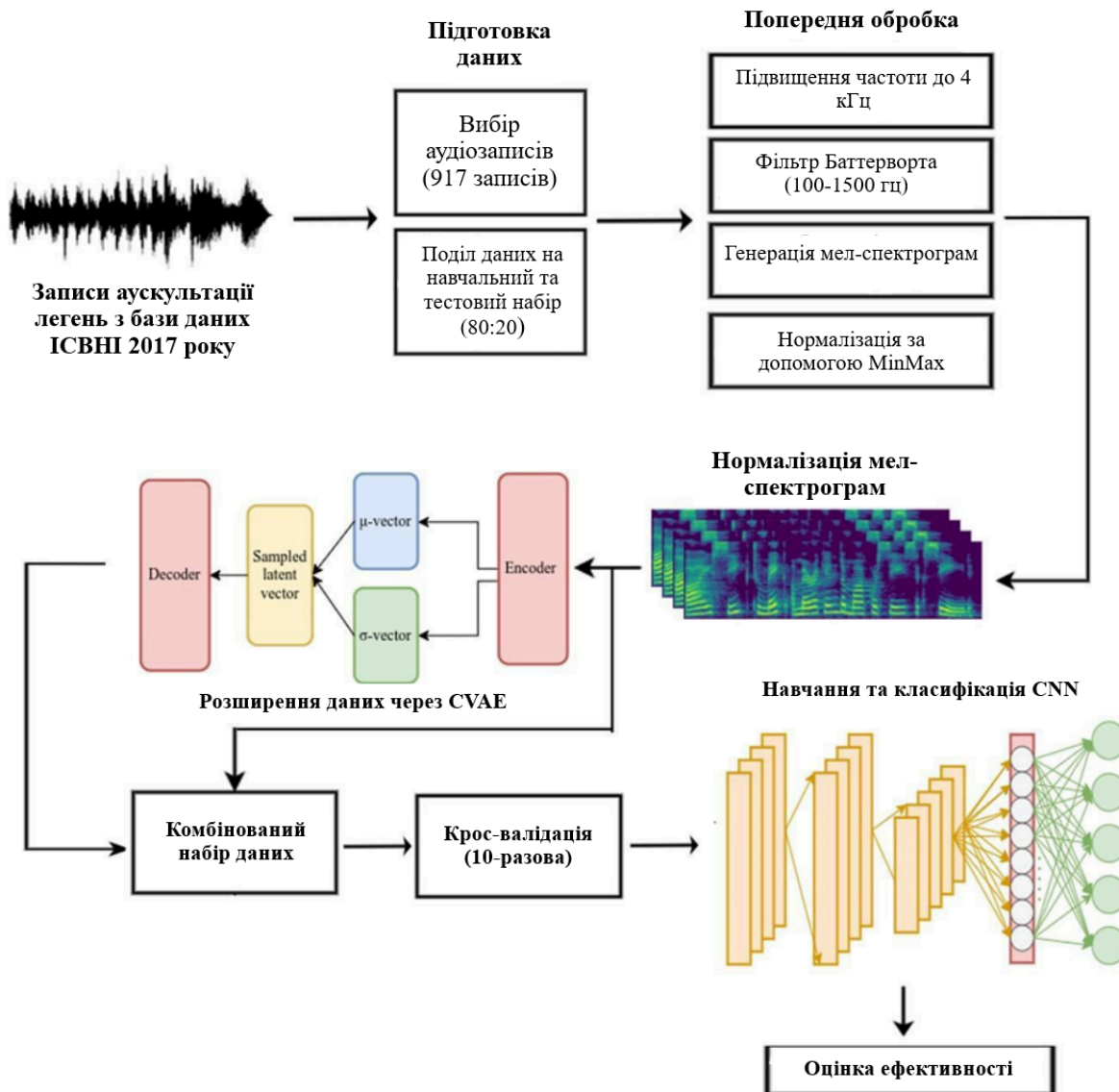


Рис. 2.5.1. Схематичне відображення кроків NASCA-методу для автоматизованої класифікації легеневих звуків.

2.5.1. Математичне формулювання методу

Нехай D - вихідний набір аудіозаписів аускультатії легень:

$$D = \{x_i | i = 1, \dots, N\}, \quad 2.5.1.1$$

де N - загальна кількість записів у дослідженні. Кожен запис x_i має мітку $y_i \in C$, де C - множина класів респіраторних захворювань або «норма».

Залежно від умов експерименту, у просторі D визначимо дві неперетинні підмножини:

$$D_{train} \subset D, D_{test} \subset D, D_{train} \cap D_{test} = \emptyset, D_{train} \cup D_{test} = D. \quad 2.5.1.2$$

Задля зменшення ризику «витоку даних», розподіл D здійснюється “пацієнт-орієнтованим” способом: записи одного пацієнта включаються лише в одну з підмножин.

Над елементами D_{train} виконується низка перетворень (попереднє фільтрування, побудова мел-спектрограм тощо). Означимо оператор попередньої обробки Φ , що діє як:

$$\Phi: x_i \mapsto \bar{x}_i, \quad 2.5.1.3$$

де \bar{x}_i - зображення нормалізованої мел-спектрограми (рис. 2.5.1, блок «Попередня обробка»). Тоді перетворена підмножина:

$$\bar{D}_{train} = \Phi(D_{train}) = \{\bar{x}_i | \bar{x}_i = \Phi(x_i), x_i \in D_{train}\}. \quad 2.5.1.4$$

Далі для розв’язання проблеми дисбалансу класів і браку аномальних записів застосовується згортковий варіаційний автоенкодер CVAE, яку розглядаємо як стохастичний оператор G :

$$G: \bar{x}_i \mapsto \hat{x}_j, \quad 2.5.1.5$$

де \hat{x}_j — синтезовані спектрограми. Сукупність усіх синтетичних зразків утворює множину:

$$\hat{D} = \{\hat{x}_j\}_{j=1}^M \quad 2.5.1.6$$

де $M = |\widehat{D}|$ – кількість синтетичних прикладів, необхідних для збалансованого навчання.

Після аугментації формується розширений набір:

$$D_{train}^* = \overline{D}_{train} \cup \widehat{D}. \quad 2.5.1.7$$

Кожен його елемент має мітку $y_i \in C$, (клас) захворювання або «норма».

Кінцевим результатом є глибинна модель f_θ , побудована за допомогою згорткової нейронної мережі (CNN) із параметрами θ . Для пошуку її параметрів θ використовується певна функція втрат L , з урахуванням ваг для різних класів і регуляризації:

$$\theta^* = \arg \min_{\theta} \left[\sum_{(x^*, y) \in D_{train}^*} L(f_\theta(x^*, y)) \right] \quad 2.5.1.8$$

Після визначення θ^* отримуємо кінцеву модель f_{θ^*} . Таким чином, фінальна класифікація для нового запису x виконується як:

$$\widehat{y} = f_{\theta^*}(\Phi(x)). \quad 2.5.1.9$$

Ця сукупність операцій і складає NASCA-метод (Neural-Augmented Sound Classification for Auscultation), спрямований на підвищення ефективності діагностичних систем для аналізу легеневих звуків.

2.5.2. Кроки реалізації методу

1. Формування датасет і розподіл на множини.

- Зі всього набору D виділяються записи від певних пацієнтів у

D_{train} і D_{test} в пропорції 80:20.

- Пацієнт-орієнтований принцип забезпечує уникнення «витоку даних».

2. Первинне фільтрування та підготовка аудіо.

- Пониження частоти дискретизації до 4 кГц (oversampling).
- Застосування фільтра Баттерворта (100–1500 Гц).
- Формування мел-спектрограм і MinMax-нормалізація.

3. Аугментація даних за допомогою CVAE.

- Для частини вибірки, де суттєво бракує прикладів, здійснюється синтетична генерація спектрограм \hat{D} .
- Перетворення G виконує стохастичне кодування та декодування у латентному просторі, створюючи нові інформативні зразки.
- Об'єднуємо оригінальний набір D_{train} із синтетичним \hat{D} .

4. Перехресна валідація та збалансований відбір.

- Використовується 10-кратна крос-валідація: D_{train} розбивається на 10 блоків. Кожен блок по черзі стає валідаційним, а решта блоків - навчальним.
- Враховується вагова корекція класів у випадку, коли деякі діагнози мають меншу кількість записів.

5. Навчання CNN і класифікація.

- Результатом є набір моделей, усереднення показників яких (середнє F1, точність, чутливість, специфічність тощо) відображає загальну ефективність.

6. Остаточне оцінювання.

- Здійснюється на відкладеному наборі D_{test} .
- Аналізуються матриця невідповідностей (confusion matrix), значення F1-міри, Cohen's Kappa, MCC тощо.

- За результатами робиться висновок щодо покращення ефективності системи порівняно з базовими методами (без аугментації чи без фільтрації).

2.6. Загальний план експерименту та обладнання

У межах експерименту було використано 917 аудіозаписів аускультації легень із загальнодоступної бази даних. Їх розподілили на навчальний та тестовий набори у співвідношенні 80:20, причому всі записи одного пацієнта включалися лише в одну з підвибірок, щоб уникнути «витоку даних». Для зниження частоти дискретизації застосували 4 кГц, а видалення сторонніх шумів виконали фільтром Баттерворта.

Кожен з отриманих аудіозразків представили у часово-частотній формі, утворивши спектрограми (мел-спектрограмування). Так стало можливим відображати та аналізувати реальні й згенеровані спектрограми як зображення. Проблему дисбалансу даних у наборі було розв'язано завдяки згортковому варіаційному автокодувальнику (CVAE), що генерував синтетичні зразки для рідкісних класів.

На наступному етапі згорткова нейронна мережа (CNN) проходила навчання з використанням 10-кратної перехресної валідації, без додаткової ручної екстракції ознак. Показники чутливості, специфічності та F1-міра та інші метрики слугували для оцінювання ефективності класифікації.

Експеримент проводився на машині з наступною конфігурацією:

- CPU: Intel(R) Xeon(R) @ 2.20GHz;
- GPU: Nvidia A100 (80 GB HBM2);
- ОЗП: 25 GB RAM;
- Твердотільний накопичувач з динамічною хмарною пам'яттю.

Середній час навчання моделі становив близько 7 хвилин.

3. Результати та обговорення

3.1. Реалізація експерименту

3.1.1. Розробка основного набору даних

Ця частина дослідження мала завдання розробити автоматизовану систему для класифікації респіраторних захворювань на основі звуків легень за допомогою машинного навчання, порівняти її з подібними системами, вивчити їх переваги та недоліки, а також розглянути можливі шляхи підвищення їх ефективності.

Область поточної роботи - використання методів глибинного навчання для виявлення деяких типів респіраторних патологій, таких як ХОЗЛ, ІВДШ, пневмонія тощо - вважається більш складною проблемою, ніж класифікація окремих звуків легень, таких як сухі та вологі хрипи. Оскільки ефективність будь-яких алгоритмів МН та ГН безпосередньо залежить від кількості навчальних даних, для моделювання системи було обрано набір даних ІСВНІ 2017, оскільки він був найбільшим загальнодоступним набором записів аускультатії на момент проведення дослідження. Однак, щоб протестувати модель, яка могла б класифікувати конкретні захворювання за їхніми респіраторними звуками, було вирішено дещо модифікувати оригінальний набір даних. На першому етапі були виключені два класи даних: астма та інфекції нижніх дихальних шляхів, оскільки вони не були репрезентативними для поточного дослідження, оскільки містили лише 2 та 1 зразок відповідно. Таким чином, у дослідження було включено 917 записів з набору даних ІСВНІ 2017, які належали до 6 класів: ХОЗЛ, пневмонія, інфекції верхніх дихальних шляхів, бронхоектатична хвороба, бронхіоліт та здорові індивіди. Усі вони містили сухі хрипи, вологі хрипи, як сухі, так і вологі хрипи або відсутність патологічних звуків.

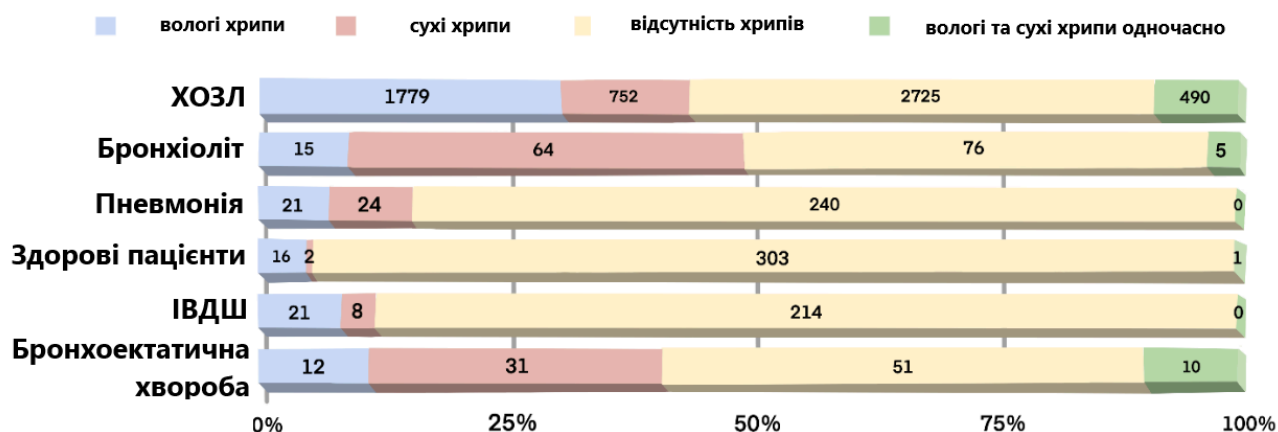


Рис. 3.1.1.1. Розподіл сухих і вологих хрипів у дихальних циклах при різних захворюваннях.

Рисунок 3.1.1.1 показує розподіл аномальних і нормальних звуків дихання у дихальних циклах пацієнтів з різними захворюваннями та здорових осіб, зібраних у наборі даних ICBNI 2017. Ця ілюстрація на краще демонструє одну з найбільш актуальних проблем у створенні автоматизованих систем класифікації захворювань дихальних шляхів - складність природи звуків легенів. Як видно з малюнка, різні типи патологічних звуків дихання можуть присутні у пацієнтів з одним і тим же захворюванням. Зокрема, аудіозаписи пацієнтів з ХОЗЛ містять сухі та вологі хрипи, а також одночасно сухі та вологі хрипи. З іншого боку, такі самі патологічні звуки можуть бути характерні для різних захворювань, наприклад, цикли дихання з вологими хрипами зустрічаються у всіх діагнозах без винятку.

Крім того, на малюнку 3.1.1.1 чітко демонструється схожість у розподілі патологічних звуків дихання при бронхоектатичній хворобі та бронхіоліті. Згідно з метою дослідження, через цю схожість обидва цих класи були об'єднані в один клас BRON. Таким чином, кінцевий набір даних містив 793 записи про ХОЗЛ, 29 записів про BRON, 37 записів про пневмонію, 23 записи про ІВДШ і 35 записів про нормальне дихання, як це

представлено на малюнку 3.1.1.2. Ці записи належали 64 пацієнтам із ХОЗЛ, 26 здоровим особам, 14 пацієнтам з ІВДШ, 13 пацієнтам з бронхоектатичною хворобою або бронхіолітом і 6 пацієнтам з пневмонією.

Малюнок показує, що розподіл класів, отриманий після вищезазначеної обробки бази даних ІСВНІ 2017, залишався надзвичайно незбалансованим. Лише клас ХОЗЛ мав достатню кількість вибірок, тоді як інші класи були меншими і мали недостатньо даних для класифікації за допомогою згорткових нейронних мереж.

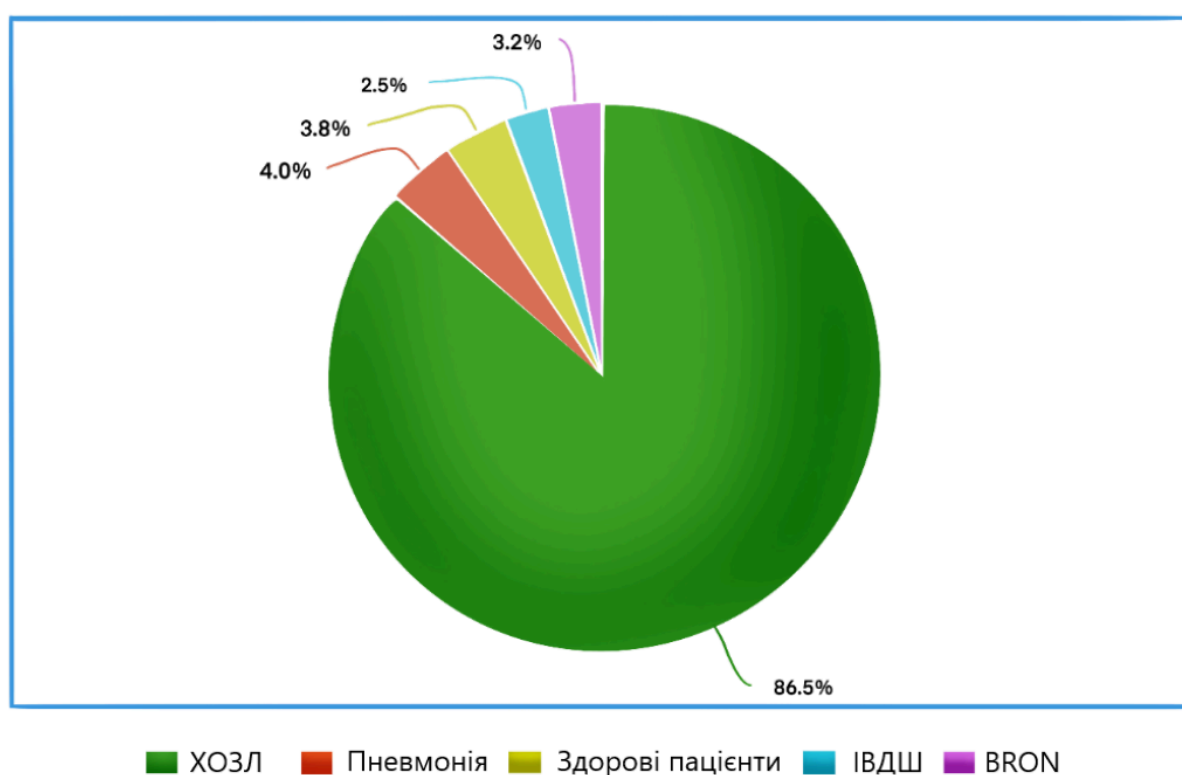


Рис. 3.1.1.2. Розподіл класів для поточного дослідження до отримання додаткових даних.

3.1.2. Попереднє фільтрування аудіозаписів

Усі 917 записів з п'яти класів, які увійшли до кінцевої вибірки, пройшли передпроцесінг, який описаний в розділі 2.

Частоти, які не несуть корисної інформації (нижче 100 Гц і вище 1500 Гц), були відфільтровані за допомогою банд-пас фільтру Баттерворта п'ятого порядку з частотою перевибірки 4000 Гц (див. Рисунок 3.1.2.1).

```
from scipy.signal import butter, lfilter

def butter_bandpass(lowcut, highcut,fs,order=5):
    return butter (order,[lowcut, highcut], fs=fs, btype='band')

def butter_bandpass_filter(data,lowcut,highcut,fs,order=5):
    b, a = butter_bandpass(lowcut, highcut,fs,order=order)
    y = lfilter (b,a,data)
    return y

fs = 4000.0
lowcut = 100.0
highcut = 1500.0
```

Рис. 3.1.2.1. Перелік реалізації 5-го порядку фільтра Баттерворта

У цьому дослідженні звуки легенів не були відокремлені від звуків серця. З одного боку, такий підхід трохи ускладнив завдання класифікації, оскільки (як детально описано у розділі 1.2.2.2) інформативні дихальні звуки завжди змішані з серцевими звуками під час аускультції. З іншого боку, це спростило загальну архітектуру системи, що додало до її переваг у відношенні потенційного практичного застосування і швидкості.

Можливість навчання на шумних даних є фундаментальною умовою для правильного функціонування автоматизованої системи класифікації звуків дихання в клінічних умовах, оскільки навчання з використанням

неоптимальних даних збільшує адаптивність алгоритмів штучного інтелекту до реальних сценаріїв. Крім того, вибраний підхід базувався на деяких інших дослідженнях [31, 32, 44, 45], які показали, що шум серця є основною перешкодою для ручного вилучення ознак, але мало перешкоджує тренуванню глибоких мереж.

3.1.3. Створення та обробка мел-спектрограм

Оскільки сирі аудіозаписи в базі даних ICBN1 2017 збиралися в реальному середовищі, для глибокого навчання використовувалися мел-спектрограми порівняно з іншими загальноновживаними методами. Зокрема, у випадку представлення даних в частотній області за допомогою перетворення Фур'є, тимчасова інформація не відображається належним чином в ознаках, і логарифмічна мел-спектрограма мала б занадто високу розмірність у контексті поточного дослідження.

Таким чином, на наступному етапі експерименту оброблені записи візуалізувалися у вигляді мел-спектрограм за допомогою алгоритму, зображеного на рисунку 3.1.3.1.

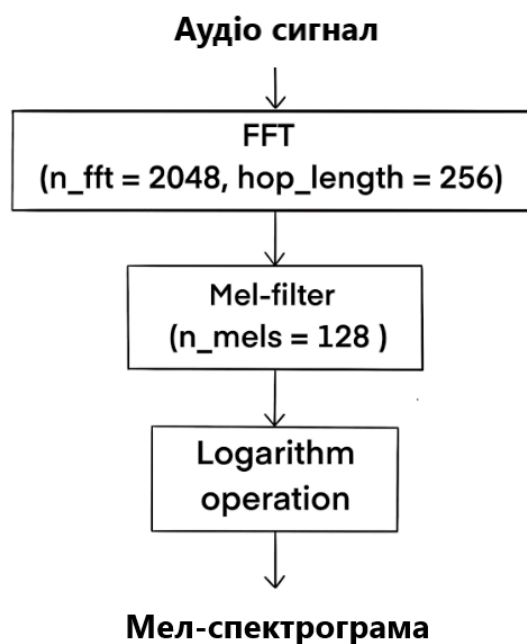


Рис. 3.1.3.1. Алгоритм для генерації мел-спектрограм

На початку отримання мел-спектрограм дані відбиралися з використанням розміру вікна $nfft=2048$ і кроку $hop_length=256$ для наступного вікна. Для кожного вікна виконувалася швидке перетворення Фур'є (FFT), яке перетворювало сигнал з області часу в область частот. Після цього весь частотний спектр був розділений на $n_{mels} = 128$ бінів або рівномірно розподілені частоти. Цей процес здійснюється для відображення лінійної частотної шкали на мел-шкалу, яка наближає сприйняття звуку аудіторною системою людини.

Для кожного вікна величина сигналу розкладалася на компоненти, що відповідають частотам на мел-шкалі. Це розкладання досягалося за допомогою формули (1.2.19). Повний перелік параметрів мел-спектрограм подано в Таблиці 3.1.3.1. Ці параметри включають розмір вікна, крок, кількість бінів мел та інші важливі деталі для обчислення мел-спектрограм.

Таблиця 3.1.3.1

Параметри для створення мел-спектрограм

| Параметр | Опис | Значення |
|---------------|--|----------|
| SR | Частота дискретизації аудіосигналу (SR), (Гц) | 4000 |
| n_fft | Довжина вікна FFT | 2048 |
| hop_length | Крок, тобто кількість вибірок між послідовними кадрами | 256 |
| n_mels | Кількість mel-смуг для створення спектрограми | 128 |
| win_length | Довжина вікна | 2048 |
| f_{min} | Найнижча частота (Гц) | 0 |
| f_{max} | Найвища частота (Гц), $f_{max}=sr/2$ | 2000 |

Як зазначено в Таблиці 3.1.3.1, найвища частота становила 2000 Гц. Цей показник було розраховано за формулою:

$$f_{max} = \frac{SR}{2} \quad (3.1.1),$$

де SR - частота дискретизації аудіосигналу.

На наступному етапі мел-спектрограми було стандартизовано до діапазону від 0 до 1 за допомогою мінімакс стандартизації (формула 2.2.1). Мел-спектрограми, отримані на цьому етапі, мали різну довжину, оскільки база даних ICBN1 2017 містить аудіозаписи звуків легень тривалістю від 10 до 90 секунд. Тому отримані часткові спектрограми лінійно інтерполювалися до однакового розміру на основі середнього розміру отриманої спектрограми. Враховуючи, що кожен стовпчик у мел-спектрограмі представляє одиницю часу, розмір усіх зображень було змінено так, щоб вони містили однакову кількість стовпців, рівну середньому числу стовпців у всіх спектрограмах. Розрахунок виконувався за формулою:

$$Columns_{mean} = \sum_{i=1}^N \frac{cols(dataset(i))}{N} \quad (3.1.2)$$

де N - кількість спектрограм у експерименті, а cols(x) - кількість стовпців зображення x.

В результаті цього кроку отримується набір зображень мел-спектрограм розміром 128x336 пікселів зі стандартизованими значеннями. Приклади мел-спектрограм, отриманих за допомогою вищезгаданого процесу для різних класів, наведено на Рисунку 3.1.3.2.

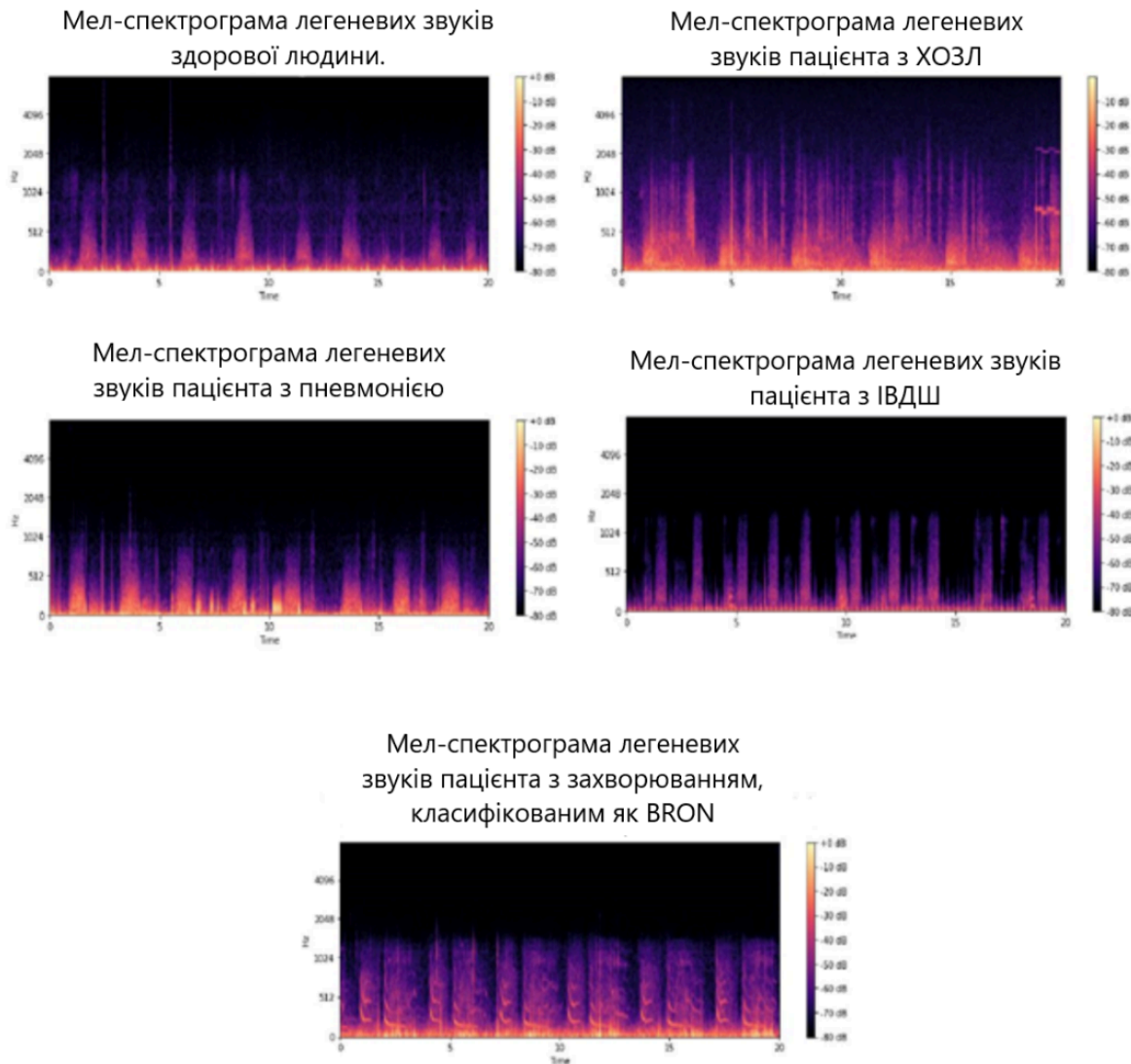


Рис. 3.1.3.2. Приклади стандартизованих мел-спектрограм для різних класів: здорової людини, ХОЗЛ, пневмонії, ІВДШ, BRON

У поточній роботі використовувалась структура класифікації «від початку до кінця», мел-спектрограми були взяті як вхід до мережі. Це дозволило уникнути етапу додаткового «ручного» вилучення ознак.

3.1.4. Аугментація даних за допомогою змінювального згорткового автоенкодера.

Оскільки, як вже було показано вище, база даних ICBHI 2017 має значну незбалансованість, наступним кроком було доповнення даних.

Згідно з експериментальним планом, описаним детально в Розділі 3, очевидну незбалансованість класів (див. Рисунок 3.1.1.2) було усунуто за допомогою згорткового автоенкодера CVAE згідно з алгоритмом, представленим на Рисунку 3.1.4.1.

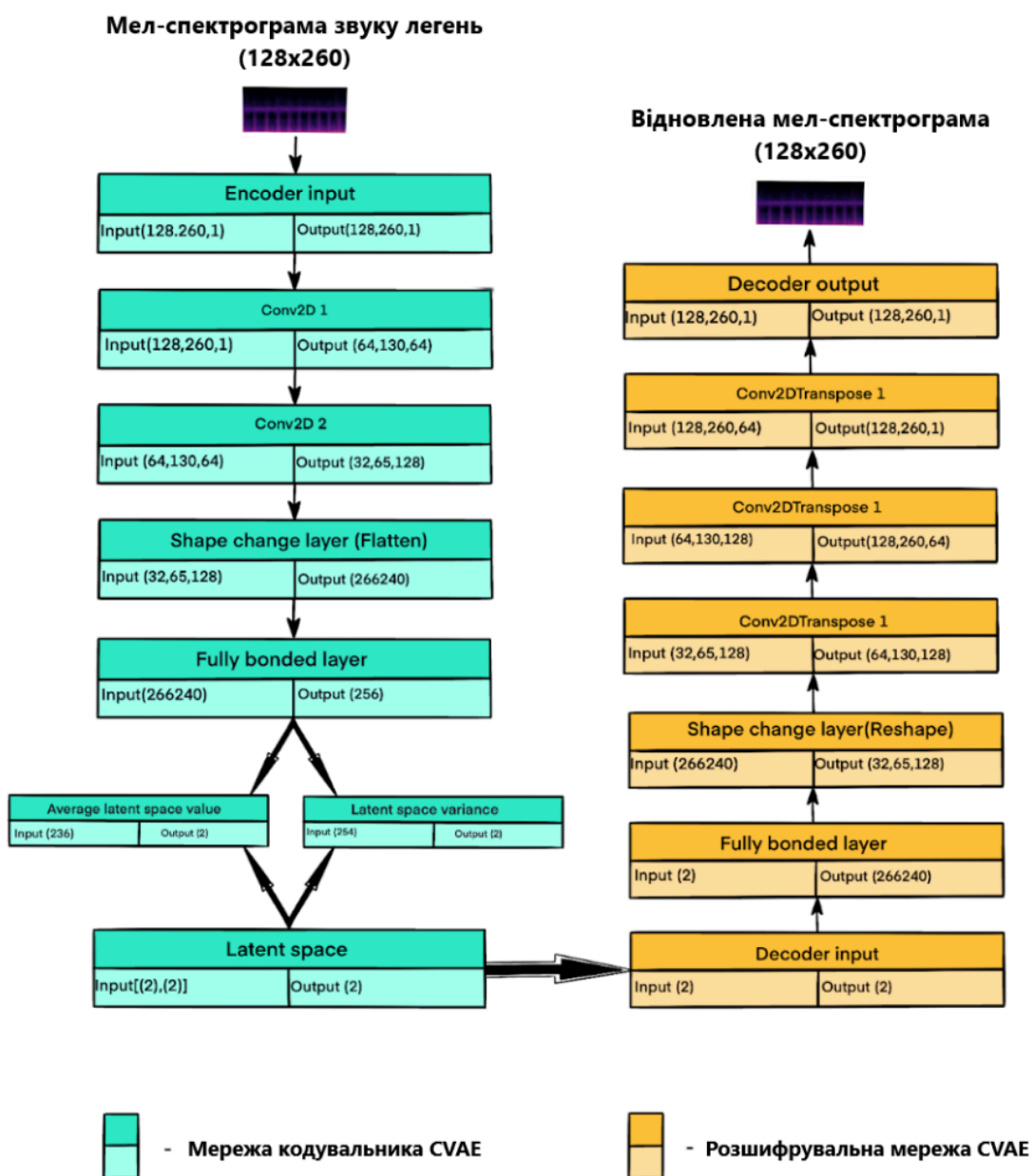


Рис. 3.1.4.1. Алгоритм навчання зі згортковим варіаційним автоенкодером

На вищезгаданій діаграмі показано, що вхідний сигнал проходить через два 2D згорткові шари енкодера Conv2D (розмір ядра = 3, кількість кроків = 2), при цьому кожен з цих шарів має удвічі більше згорткових фільтрів, ніж попередній (16 і 32 відповідно), що дозволяє моделі досліджувати складніші характеристики потоку. Далі інформація проходить до шару плоскості (flatten layer), який зменшує розмірність, тобто перетворює двовимірну форму в одновимірну для подальшої обробки, і потім до щільного або повністю з'єднаного шару, в якому кожен нейрон з'єднаний з усіма нейронами попереднього шару. Він використовується для комбінації всіх карт ознак, отриманих з останнього шару енкодера.

Щільний шар з'єднаний з варіаційним шаром, який розраховує параметри розподілу вихідних даних. Середнє (μ) і дисперсія (σ), параметри, отримані під час тренування моделі, визначають ймовірнісний розподіл за допомогою трюку з перепараметризації, що детально описано в Розділі 2.3, і латентний вектор Z вибирається з розподілу з середнім $Z\mu$ і дисперсією $Z\sigma$, після чого передається в декодер для отримання передбачуваних виходів.

Все декодуюча мережа відображає цю архітектуру, включаючи повністю зв'язаний шар, шар Reshape і три шари Conv2DTranspose, відповідальні за підвищення розміру та відновлення зображення до його початкового розміру. Транспоновані шари згортки виконують перетворення, яке працює в протилежному напрямку звичайної згортки, при цьому зберігаючи зв'язність, сумісну з конволюційною мережею, при цьому кожен шар містить удвічі менше кількість згорткових фільтрів, ніж попередній (відповідно 64, 32 і 16).

В таблиці 3.1.4.1 представлено список гіперпараметрів для згорткового автоенкодера.

Таблиця 3.1.4.1.

Налаштування гіперпараметрів CVAE

| Шар | Гіперпараметр | Опис | Значення |
|-------------------|--------------------|---|----------|
| <i>Conv2D 1</i> | <i>Фільтри</i> | Розмір вихідного простору, що вказує на кількість вихідних фільтрів у згортці. | 16 |
| <i>Conv2D 2</i> | <i>Фільтри</i> | Розмір вихідного простору, що вказує на кількість вихідних фільтрів у згортці. | 32 |
| <i>Conv2D 1,2</i> | <i>Розмір ядра</i> | Розмір ядра (висота і ширина вікна першого шару двовимірної згортки). | [3,3] |
| | <i>Кроки</i> | Кроки для переміщення вікна об'єднання. | [2,2] |
| <i>Dense 1</i> | <i>Одиниці</i> | Розмір вихідного простору, що вказує на кількість вихідних фільтрів у згортці.. | 256 |

| Шар | Гіперпараметр | Опис | Значення |
|------------------------------|--------------------|--|----------------|
| <i>Dense 2</i> | <i>Одиниці</i> | Розмір вихідного простору, що вказує на кількість вихідних фільтрів у згортці. | 266240 |
| <i>Dense 1,2</i> | <i>Активація</i> | Функція активації повністю з'єданого шару. | <i>Softmax</i> |
| <i>Conv2DTranspose 1,2,3</i> | <i>Розмір ядра</i> | Розмір ядра (висота та ширина вікна першого шару двовимірної згортки) | [3,3] |
| <i>Conv2DTranspose 1</i> | <i>Фільтри</i> | Розмір вихідного простору, що вказує кількість вихідних фільтрів у згортці. | 64 |
| <i>Conv2DTranspose 1</i> | <i>Фільтри</i> | Розмір вихідного простору, що вказує кількість вихідних фільтрів у згортці. | 32 |
| <i>Conv2DTranspose 3</i> | <i>Фільтри</i> | Розмір вихідного простору, що вказує кількість вихідних фільтрів у згортці. | 16 |

У поточному експерименті CVAE було навчено протягом 100 епох, тобто повних проходів через весь навчальний набір даних туди і назад. Наприкінці кожної епохи параметри моделі оновлювалися на основі двох функцій втрат: середньоквадратичної похибки (Рівняння 1.3.1) для реконструкції та дивергенції Кульбака-Лейблера (Рівняння 1.3.2) для регуляризації латентного простору. На кожному кроці протягом усього процесу навчання розраховувалися та мінімізувалися втрати реконструкції та дивергенції.

Після отримання правильного екземпляра моделі аугментації були згенеровані синтетичні мел-спектрограми аускультатії легень. Приклади даних для цільових класів, отриманих за допомогою CVAE, що працює в режимі без нагляду, представлені на Рисунку 3.1.4.2.

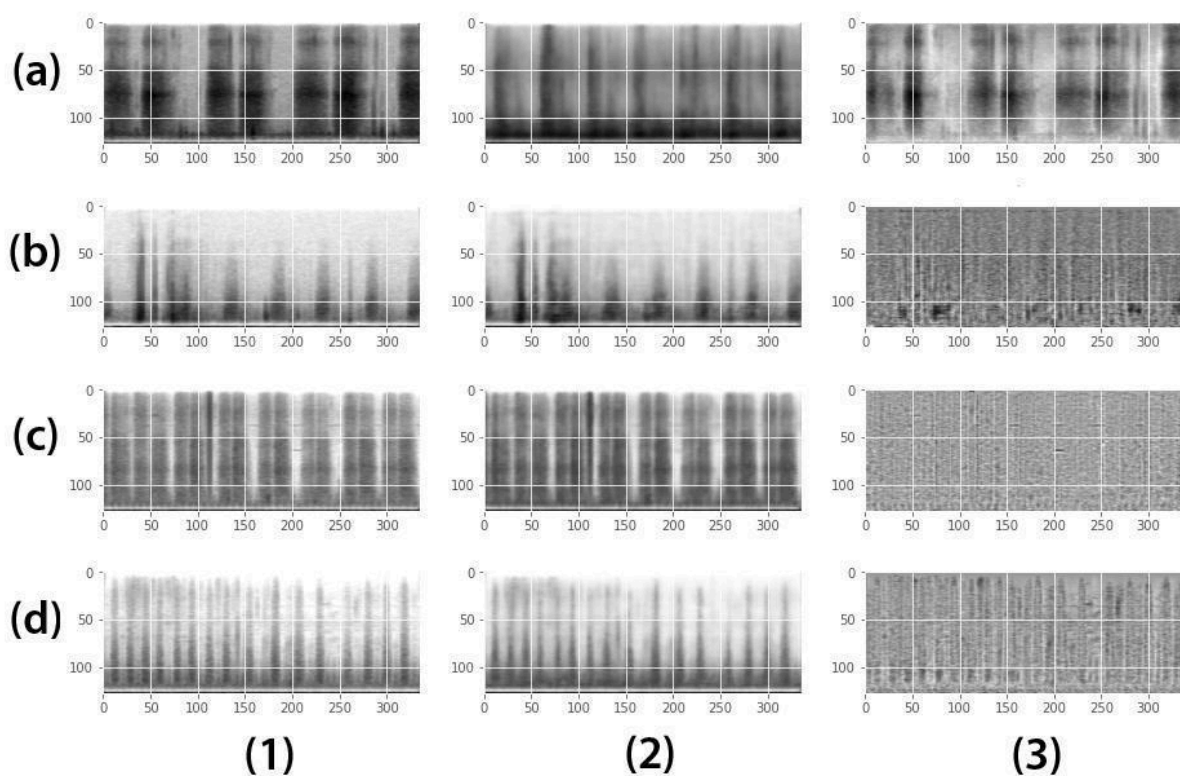


Рис. 3.1.4.2. Синтезовані навчальні зразки, отримані під час збільшення даних за допомогою CVAE.

Умовні позначки на Рисунку 3.1.4.2. (1) - оригінальний зразок, (2) - синтезований зразок, (3) - крос-кореляція. Зразки наведені для класів (a) - BRON, (b) - пневмонія, (c) - стандарт, (d) - ІВДШ.

Число було обрано для врахування відсотка кожного класу меншини в незбалансованому наборі даних. Загальна кількість зразків, створених CVAE, представлена в Таблиці 3.1.4.2.

Таблиця 3.1.4.2.

Кількість зразків, створених CVAE, та загальна кількість реальних і синтезованих зразків у класах меншини.

| | BRON | Пневмонія | Здоровий пацієнт | ІВДШ |
|---------------------------|-------------|------------------|-------------------------|-------------|
| CVAE | 760 | 764 | 765 | 758 |
| Загальна кількість | 789 | 801 | 800 | 781 |

Нарешті, було створено спільний набір реальних та синтетичних даних, у якому кількість зразків кожного класу була приблизно однаковою (див. Рисунок 3.1.4.3).

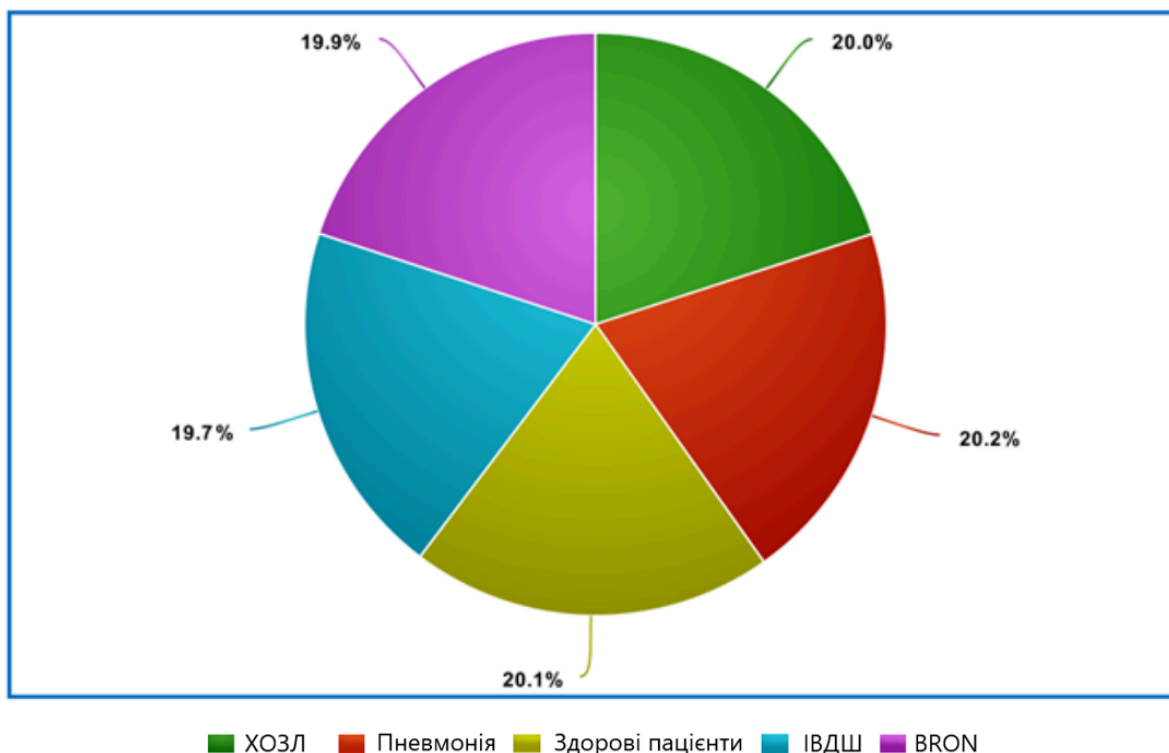


Рис. 3.1.4.3. Розподіл класів для поточного дослідження після збільшення даних

Отриманий набір даних був придатний для подальшого моделювання.

3.1.5. Розподіл даних та класифікація

Після визначення та розробки робочого набору даних його було розділено на навчальні та тестові набори. Як відомо, існують різні методи розподілу даних для моделей МН та ГН, вибір яких значною мірою визначається якістю та кількістю даних. Найпоширенішим варіантом є випадкова вибірка, але в поточній роботі було обрано метод стратифікованого розподілу, який зазвичай використовується для незбалансованих наборів даних, де деякі класи містять значно менше даних, ніж інші. Він забезпечує, що навчальні та тестові набори даних адекватно представляють розподіл класів, щоб уникнути упередженості в кінцевій моделі.

При стратифікованому розподілі набір даних ділиться з урахуванням відносних пропорцій класу, тобто навчальні та валідаційні набори містять репрезентативну підмножину кожного класу, зберігаючи такий самий відсоток зразків з кожного класу, як і оригінальні дані. Таким чином, модель може навчитися розпізнавати шаблони та робити прогнози для всіх класів, що значно підвищує надійність алгоритму МН.

Стратифікований розподіл було виконано за допомогою функції `train_test_split` з параметром `stratify` (Рис. 3.1.5.1).

```
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.8, stratify=y)
```

Рис. 3.1.5.1. Фрагмент коду для стратифікованого поділу даних на навчальні та тестові набори

Використовуючи стратифікований розподіл, набір даних було розділено на 80% навчальний набір та 20% тестовий набір. Цю пропорцію було обрано на основі розрахунків Афшина Голамі та спів. (2018), які довели, що такий розподіл забезпечує оптимальний баланс між достатнім навчанням та перенавчанням системи [57].

Для уникнення перенавчання було використано метод 10-кратної крос-валідації, алгоритм якого представлений на Рисунку 3.1.5.2.

Десятикратна крос-валідація (або більш загально, k -кратна валідація з $k=10$) є широко прийнятою практикою в машинному навчанні для зменшення перенавчання шляхом використання всього доступного набору даних для навчання та валідації. Основна концепція полягає у поділі навчального набору даних на k менших наборів, які часто називаються "фолдами". Модель навчається, використовуючи $(k-1)$ фолди як навчальні

дані, тоді як її точність оцінюється на решті даних. Цей процес повторюється k разів. У випадку $k = 10$, кожен зразок у наборі даних використовується один раз для тестування та дев'ять разів для навчання. Оцінки, отримані в кожній з 10 ітерацій, потім усереднюються для отримання загального результату.

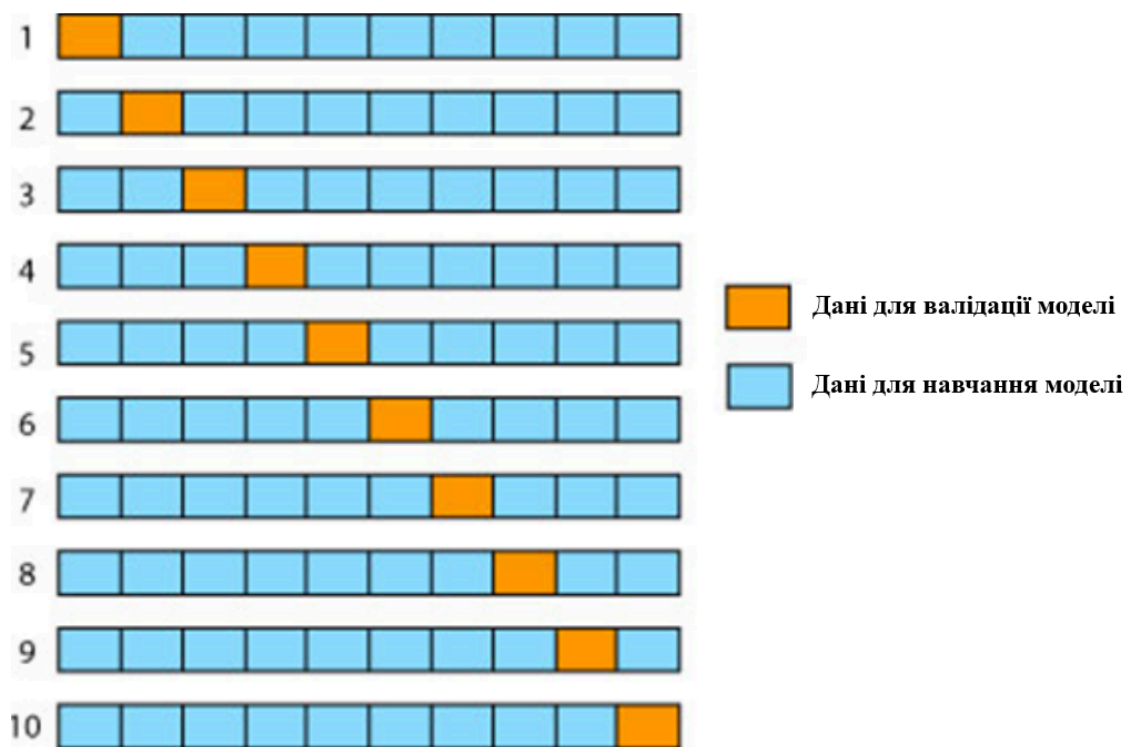


Рис. 3.1.5.2. Алгоритм десятикратної крос-валідації

У даному дослідженні використовувався стратифікований підхід k -кратної валідації, при якому дані поділяються на k згорток, що містять однакову пропорцію мічених класів. Для класифікації записів аускультації легень була використана модель згорткової нейронної мережі, яка детально описана в розділі 2.4. Налаштування параметрів цієї моделі наведені в таблиці 3.1.5.1.

Таблиця 3.1.5.1.

Налаштування параметрів згорткової нейронної мережі (CNN)

| Шар | Параметр | Опис | Значення |
|------------------------|--------------------------|---|----------------|
| <i>Conv2D 1,2</i> | <i>Фільтри</i> | Розмір вихідного простору, що вказує кількість вихідних фільтрів у згортці. | 32 |
| | <i>Кроки</i> | Кроки згортки у висоту та ширину | [1,1] |
| | <i>Активація</i> | Функція активації | <i>ReLu</i> |
| <i>Conv2D layer 1</i> | <i>Розмір ядра</i> | Висота та ширина вікна першого шару двовимірної згортки. | [3,3] |
| <i>Conv2D layer 2</i> | | | [2,2] |
| <i>MaxPooling2D</i> | <i>Розмір об'єднання</i> | Розмір вікна об'єднання | [2,2] |
| | <i>Кроки</i> | Кроки переміщення вікна об'єднання. | [2,2] |
| | <i>Активація</i> | Функція активації | <i>ReLu</i> |
| <i>Dropout</i> | <i>Частота</i> | Відсоток введених одиниць, які будуть викинуті. | 0.3 |
| <i>Dense layer 1,4</i> | <i>Одиниці</i> | Розмір початкового простору у повністю з'єднаних шарах (1-4). | 64 |
| <i>Dense layer 2,3</i> | | | 128 |
| <i>Dense layer</i> | <i>Активація</i> | Fully bonded layer activation function | <i>Softmax</i> |
| | <i>Оптимізатор</i> | Fully connected layer optimiser | <i>Adam</i> |

У рекурентній нейронній мережі Conv2D та MaxPooling2D використовується активаційна функція ReLU, тоді як повністю зв'язаний шар використовує Softmax. Softmax зазвичай використовується для правильної багатокласової класифікації вхідних даних.

Останній шар згорткової нейронної мережі складається з повністю зв'язаного шару з п'ятьма нейронами. Його завдання - надавати ймовірності кожного з п'яти респіраторних класів (стандартний, BRON, ХОЗЛ, ІВДШ, пневмонія), що розглядаються в дослідженні.

Кожен нейрон повинен визначати ймовірність належності об'єкта до класу, тобто значення від нуля до одиниці. Всі п'ять нейронів разом мають утворювати загальну ймовірність, рівну одиниці. Для цього використовується функція Softmax, формула якої подана нижче.

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{k=1}^K e^{z_k}} \quad (3.1.3)$$

де K - кількість класів (у поточній роботі 5).

Отже, функція Softmax приймає на вхід набір чисел для кожної категорії, званих логітами, і повертає ймовірнісний розподіл того, що об'єкт належить кожному з визначених класів, виконуючи дві основні умови: ймовірність кожного класу не може бути менше нуля, і сума всіх ймовірностей дорівнює одиниці.

Модель класифікації була побудована з використанням функції втрат категоріальної перехресної ентропії (формула 2.4.2), яка працює зі значеннями виходу моделі, які представляють ймовірності належності вхідних даних до різних класів.

Для оптимізації використовувався оптимізатор Adam, алгоритм якого наведений на рисунку 3.1.5.4.

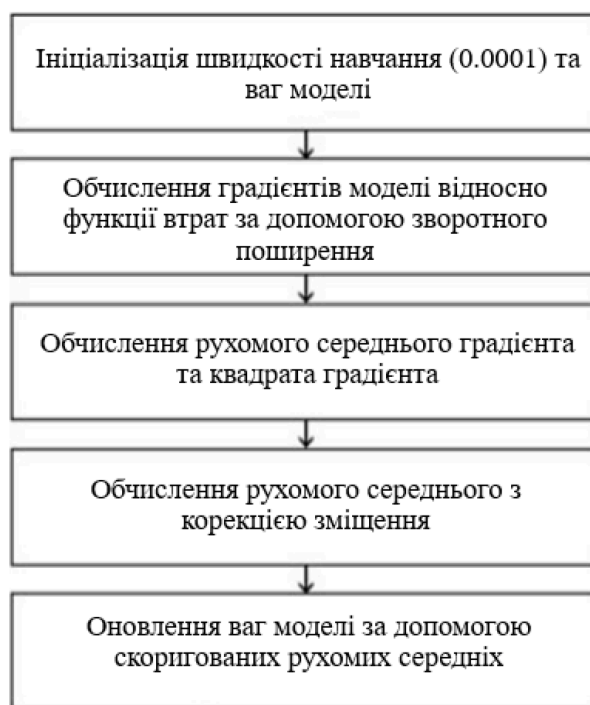


Рис. 3.1.5.4. Алгоритм оптимізатора Adam

У математичних термінах цей алгоритм, з вхідними параметрами $\gamma(lr)$, β_1 , β_2 , θ , $f(\theta)$ and λ (вага втрат), та початковими умовами $m_0 \leftarrow 0$ (перший момент), $v_0 \leftarrow 0$ (другий момент) і $v_0^{max} \leftarrow 0$, може бути описаний наступним чином:

for $t=1$ *to* ... *do*
 if *maximise*:

$$g_t \leftarrow -\nabla_{\theta} f_t(\theta_{t-1})$$

else

$$g_t \leftarrow \nabla_{\theta} f_t(\theta_{t-1})$$

if $\lambda \neq 0$

$$g_t \leftarrow g_t + \lambda \theta_{t-1}$$

$$m_t \leftarrow \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad 3.1.4$$

$$v_t \leftarrow \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$$

$$\hat{m}_t \leftarrow \frac{m_t}{1 - \beta_1^t}$$

$$\hat{v}_t \leftarrow \frac{v_t}{1 - \beta_2^t}$$

if amsgrad

$$\hat{v}_t^{max} \leftarrow \max(\hat{v}_t^{max}, \hat{v}_t)$$

$$\theta_t \leftarrow \theta_{t-1} - \frac{\gamma \hat{m}_t}{(\sqrt{\hat{v}_t^{max}} + \epsilon)}$$

else

$$\theta_t \leftarrow \theta_{t-1} - \frac{\gamma \hat{m}_t}{(\sqrt{\hat{v}_t} + \epsilon)}$$

return θ_t

Метод Adam характеризується швидкістю навчання та має два гіперпараметри, по одному для кожного з рухомих середніх. Застосовані параметри оптимізатора наведені в Таблиці 3.1.5.2.

Таблиця 3.1.5.2

Встановлення параметрів оптимізатора Adam

| Аргумент | Значення |
|--------------------|----------|
| β_1 | 0.9 |
| β_2 | 0.999 |
| Швидкість навчання | 0.001 |

За замовчуванням оптимізатор *Adam* має постійну швидкість навчання. Однак для мінімізації втрат моделі необхідно масштабувати величину оновлень ваг моделі, змінюючи швидкість навчання, яка є одним з ключових гіперпараметрів для градієнтного спуску. Для вирішення цієї

проблеми був обраний експоненційний план зменшення швидкості навчання з коефіцієнтом 0.9, реалізований функцією *ExponentialLR*, який зменшує швидкість навчання кожної групи параметрів на гамму кожної епохи.

Таблиця 3.1.5.3.

Налаштування для плану зменшення експоненційної швидкості навчання

| Параметр | Опис | Значення |
|-------------------|---|--------------|
| <i>optimiser</i> | Оптимізатор | <i>Adam</i> |
| <i>gamma</i> | Мультиплікативний коефіцієнт зниження швидкості навчання (γ) | 0.9 |
| <i>last_epoch</i> | Індекс останньої епохи. | -1 |
| <i>verbose</i> | Додаткові дані | <i>False</i> |

Гіперпараметри планувальника наведено у Таблиці 3.1.5.3, а уривок коду - на Рисунку 3.1.5.5.

```

class ExponentialLR(LRScheduler):

    def __init__(self, optimizer, gamma, last_epoch = -1, verbose = False):
        self.gamma = gamma
        super().__init__(optimizer, last_epoch, verbose)

    def get_lr(self):
        if not self._get_lr_called_within_step:
            warnings.warn

        if self.last_epoch == 0:
            return [ group['lr'] for group in self.optimizer.param_groups]
        return [group['lr']* self.gamma
                for group in self.optimizer.param_groups]

    def _get_closed_form_lr(self):
        return [base_lr * self.gamma ** self.last_epoch
                for base_lr in self.base_lrs]

```

Рис. 3.1.5.5. Планувальник ExponentialLR, який зменшує швидкість навчання кожної групи параметрів на γ кожної епохи.

Коефіцієнт був обраний експериментально шляхом проведення кількох міні-серій навчання, після кожної з яких швидкість навчання поступово зростала експоненційно, а втрати на кожному збільшенні були записані.

3.2. Експериментальні результати

3.2.1. Оцінка продуктивності генеративної моделі

Для оцінки якості звуків дихання, синтезованих генеративною моделлю, використовувались Frechette audio distance (FAD), кореляція та порівняння характеристик синтезованого та оригінального звуку з використанням аналізу головних компонент (PCA).

3.2.1.1. Відстань звукового відліку Frechette

Frechette Audio Distance (FAD) - метрика, яка може бути використана для оцінки синтезованого аудіо навіть відсутності посилання на зразок [58]. Замість оцінки окремих синтезованих аудіозаписів ця метрика порівнює статистику вбудовання, згенеровану на повному наборі даних для оцінки зі статистикою вбудовання, згенерованою на навчальному наборі. FAD для оригінальних і синтетичних наборів даних представлений наступною формулою:

$$FAD = \|\mu_r - \mu_g\|^2 + trace(\Sigma_r + \mu_g - 2\sqrt{\Sigma_r \Sigma_g}) \quad 3.2.1$$

де μ_r - середнє значення розподілу реальних даних, μ_g - середнє значення розподілу синтезованих даних, Σ_r - коваріація розподілу реальних даних, Σ_g - коваріація розподілу синтезованих даних.

Оцінки FAD синтетичного аудіо, встановлені з використанням CVAE, наведені в Таблиці 3.2.1.1, показують стабільні значення в діапазоні від 10 до 13, що свідчить про стабільно задовільну якість синтезованих записів у всіх категоріях.

Таблиця 3.2.1.1.

Значення FAD для синтезованого аудіо за класами.

| Клас | Значення <i>FAD</i> |
|-----------|---------------------|
| BRON | 12.13 |
| Пневмонія | 11.32 |
| ІВДШ | 10.12 |
| Стандарт | 11.07 |

3.2.1.2. Кореляція між оригінальним та синтезованим сигналом

Ще одна оцінка якості синтезованих записів CVAE — кореляція, яка корисна для вимірювання схожості між двома сигналами, оскільки це міра того, наскільки два сигнали схожі один на одного.

Для одновимірних часових рядів x та y взаємна кореляція з лагом k визначається рівнянням:

$$\Phi_{xy}[k] = \sum_{n=-\infty}^{\infty} x[k+n]y[n] \quad 3.2.2$$

Іншими словами, кореляція є добутком сигналу $y[n]$ та зміщеної версії сигналу $x[n]$, яка отримується шляхом введення затримки k . Щоб визначити взаємну кореляцію двох двовимірних зображень - мел-спектрограм реального та синтетичного сигналу - використовувалася функція `scipy.signal.correlate2d` з бібліотеки SciPy. Приклади цієї кореляції представлені на рисунку 3.1.4.2. Для оцінки кореляції між реальними та синтезованими сигналами було взято вибірку з 50 зразків з кожного класу та доповнено CVAE. Для кожного з обраних синтезованих зразків була виміряна кореляція стосовно реальних сигналів, і була обрана максимальна кореляція. Середнє значення та стандартне відхилення максимально досяжної кореляції представлені в таблиці 3.2.1.2.

Таблиця 3.2.1.2.

Середнє значення та стандартне відхилення максимально досяжної кореляції між синтетичними та реальними звуками

| Клас | Кореляція |
|-----------|-----------|
| BRON | 0,49±0,16 |
| Стандарт | 0,54±0,15 |
| Пневмонія | 0,55±0,14 |

| Клас | Кореляція |
|------|-----------|
| ІВДШ | 0,59±0,13 |

Як можна побачити з таблиці 3.2.1.2, найвищу кореляцію було зафіксовано для класу URTI, а найнижчу – для класу BRON, що відповідає індикаторам аудіо дистанції Фреше (див. таблицю 3.2.1.1), які відповідно були найвищими для BRON і найнижчими для ІВДШ.

3.2.1.3. Аналіз головних компонентів

Для аналізу якості та продуктивності генеративної моделі, запропонованої в даному дослідженні, була оцінена схожість між аудіозаписами, синтезованими за допомогою Conditional Variational Autoencoder (CVAE), та реальними записами, отриманими під час аускультатії легень. Ця оцінка була проведена за допомогою аналізу головних компонент (PCA) та візуалізована у двовимірному просторі.

Аналіз головних компонент є широко використовуваною технікою у статистиці для факторного аналізу. Він передбачає застосування ортогонального перетворення до набору спостережень з потенційно корельованими змінними. Ці змінні представляють сутності, кожна з яких має окремі числові значення. Метою PCA є перетворення цього набору спостережень у новий набір змінних, які позбавлені лінійної кореляції. Ці нові змінні називаються головними компонентами.

У контексті оцінки схожості між синтезованими та реальними аудіозаписами PCA може бути використаний для зменшення розмірності даних та виявлення патернів або схожостей між ними. Візуалізація даних у нижчому вимірі, наприклад, у двовимірному просторі, дозволяє дослідникам спостерігати будь-яке кластерування або розділення між

синтезованими та реальними записами. Ця візуалізація допомагає зрозуміти ступінь схожості або розбіжності між двома наборами записів.

Для проведення PCA для кожного меншого класу звуків легень, який був розширений за допомогою Conditional Variational Autoencoder (CVAE), було відібрано 200 синтетичних аудіозаписів і порівняно з реальними аудіозаписами у кожній категорії. Процес вилучення ознак, який включав обчислення статистик, таких як середнє значення, стандартне відхилення та похідна першого порядку коефіцієнтів тонкої частоти кепстра, призвів до отримання вектору ознак розміром 39 вимірів для кожного аудіозапису.

Після цього був застосований PCA для зменшення розмірності векторів ознак з 39 вимірів до двох головних компонент. Ці головні компоненти захоплюють найбільш значущу варіабельність у даних, що дозволяє представити аудіозаписи у двовимірному просторі. Отримана діаграма розсіювання, яка зображає синтетичні та оригінальні аудіозаписи у кожній розширеній категорії, показана на рисунку 3.2.1.1. Ця візуалізація дозволяє оцінити схожість між синтетичними та реальними аудіозаписами у кожній категорії та полегшує оцінку продуктивності методу розширення на основі CVAE.

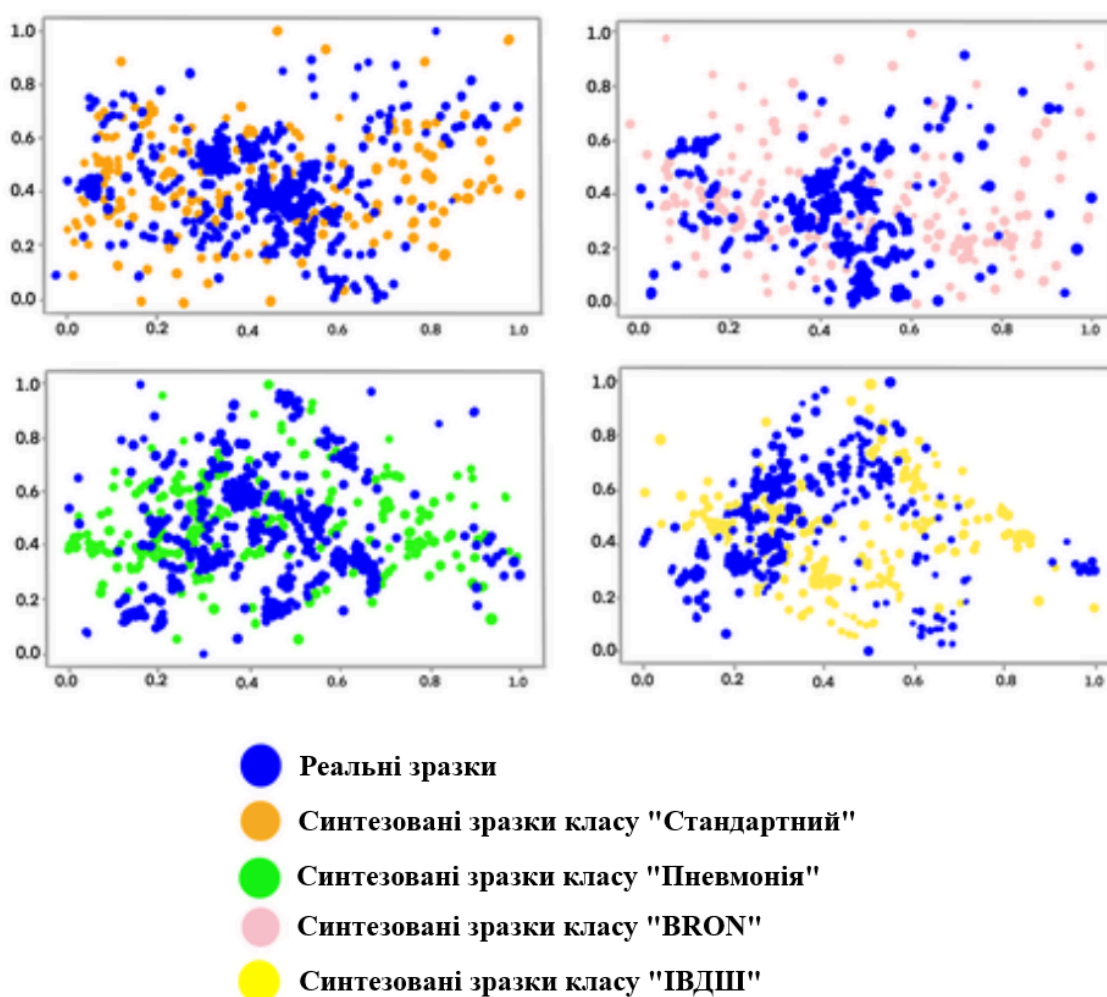


Рис. 3.2.1.1. Діаграма розсіювання синтезованих і оригінальних аудіозаписів у розширених категоріях

Рисунок 3.2.1.1 показує, що розподіл ознак аудіозаписів, згенерованих CVAE, відповідає характеристикам оригінальних звуків, тому синтезовані аудіозаписи схожі на оригінальні і можуть бути використані для подальшої класифікації.

3.2.3. Оцінка ефективності моделі класифікації

Щоб перевірити якість моделі та проаналізувати матрицю змішування, використовувалися декілька ключових показників, включаючи чутливість (формула 1.3.4), специфічність (формула 1.3.5), міру коректності (формула 1.3.6) та метрику F_1 (формула 1.3.9). Крім того, було обчислено параметр капа Кохена (формула 2.4.3), який служить надійним заходом ступеня згоди між оригінальними та передбаченими мітками.

Матриця змішування систематично генерувалася після кожної згортки, і всі метрики обчислювалися на основі кумулятивної матриці змішування, отриманої за допомогою десятикратної крос-валідації. Експеримент проводився для змішаного набору реальних та синтезованих аудіозаписів з стратифікованим поділом на навчальні та тестові набори, які мали перетин за ідентифікаторами пацієнтів, тобто різні звуки дихання, зібрані від одного досліджуваного учасника, могли одночасно включатися як до навчальних, так і до тестових даних. Використовуючи зазначені показники, ефективність класифікацій запропонованої моделі для всіх п'яти класів (стандарт плюс 4 захворювання) представлена в Таблиці 3.2.3.1.

Таблиця 3.2.3.1.

Оцінка ефективності (%) запропонованої моделі без групування ідентифікаторів пацієнтів на основі десятикратної крос-валідації.

| Діагноз | Чутливість, % | Специфічність, % | Точність, % | Параметр Капа Кохена, % | Метрика F1 (F1-score), % |
|-----------------------------|------------------|---------------------|----------------|----------------------------------|-----------------------------------|
| Здоровий пацієнт (стандарт) | 98.46 | 99.61 | | | 98.46 |
| ХОЗЛ | 100.00 | 98.45 | | | 96.96 |
| BRON | 95.31 | 100.00 | | | 96.06 |
| Пневмонія | 98.46 | 100.00 | | | 99.22 |
| ІВДШ | 100.00 | 100.00 | | | 100.00 |
| Загалом (середнє) | 98.44 | 99.61 | 98.44 | 98.05 | 98.45 |

Високі показники специфічності в експерименті свідчать про достатню якість імітації паттернів патології при збільшенні даних за допомогою варіаційного згорткового автокодувальника. Подібний підхід до збільшення даних працював з десятикратною крос-валідацією і дозволяє узагальнити характеристики для малих класів.

Найвищі значення специфічності були отримані для класів ІВДШ, пневмонія та бронхіт, що свідчить про найбільшу варіативність цих класів за представленими характеристиками. Найвищі значення F1 та чутливості для класу ІВДШ в першому експерименті вказують на найбільше розмаїття

зразків серед інших класів, що забезпечує узагальнення характеристик всередині класу.

Середнє значення параметра Капа Кохена, який вважається більш точним, ніж звичайний відсоток відповідності, оскільки враховує можливість випадкової відповідності, становило 98.44%, що є показником досить високого ступеня згоди між передбаченими та оригінальними мітками. Капа досягає свого теоретичного максимального значення 1 лише тоді, коли ця згода є досконалою, але зазвичай значення більше 0.8 вважається чудовим результатом [59].

3.3. Обговорення результатів

3.3.1. Місце запропонованої моделі в класифікації автоматизованих систем для аналізу звуків дихання

Для аналізу результатів поточної роботи в контексті загального стану розвитку автоматизованих систем аналізу звуків дихання необхідно виявити відповідні проекти, для чого потрібно повернутися до їх класифікації згідно з основними завданнями, які вони повинні вирішувати.

Останніми роками було розроблено та оцінено багато різних моделей для автоматичного виявлення та класифікації респіраторних патологій. Важливо зазначити, що кінцевою метою розробки таких систем є покращення діагностики захворювань верхніх та нижніх дихальних шляхів, багато з яких, таких як ХОЗЛ або астма, можуть призвести до інвалідності та передчасної смерті. Однак, існуючі автоматизовані системи, описані в літературі, підходять до цього завдання з різних кутів. Вони можуть бути поділені на дві широкі категорії згідно з їх призначенням:

1. Автоматизовані системи, які розпізнають і класифікують звуки легень (вдихи та видихи, вологі та сухі хрипи тощо) .

2. Автоматизовані системи, які розпізнають специфічні патерни звуків легень та класифікують діагнози або захворювання, що відповідають цим патернам.

Кожна з вищезгаданих категорій, однак, може бути поділена на конкретні області згідно зі ступенем складності завдання, яке повинна вирішити автоматизована система класифікації. Повна класифікація існуючих автоматизованих систем класифікації респіраторних звуків з прикладами та посиланнями на їх реалізацію представлена в Таблиці 3.3.1.1.

Таблиця 3.3.1.1.

Класифікація автоматизованих систем розпізнавання та класифікації респіраторних звуків

| Тип системи | Мета системи | Приклади реалізації |
|---|--|--|
| Системи, що розпізнають та класифікують звуки легень. | Пошук конкретних дихальних звуків | <ul style="list-style-type: none"> • Виявлення дихальних звуків з мокротинням [34]. • Виявлення вологого свисту [47]. • Виявлення безперервних випадкових звуків легень [52]. • Виявлення фази дихання [52]. |
| | Двохфакторна класифікація звуків легень. | <ul style="list-style-type: none"> • Класифікація аномальних звуків/стандарт"[44]. • Класифікація “сухий свист/стандарт” [49, 50]. • Класифікація “вологий свист/стандарт” [50]. |

| Тип системи | Мета системи | Приклади реалізації |
|--|---|--|
| | Багатофакторна класифікація. | <ul style="list-style-type: none"> • Трифакторна класифікація "вологий свист/хрип/стандарт" [30]. • Трифакторна класифікація "сухий свист/вологий свист/стандарт" [30]. • Чотирифакторна класифікація "сухі хрипи/вологі хрипи/сухі хрипи + вологі хрипи/стандарт" [33, 38, 46, 50]. • Чотирифакторна класифікація "стандартні звуки/безперервні звуки/переривчасті звуки/невідомі звуки" [48] • Семифакторна класифікація "вологі дрібноміхурцеві/вологі великі міхурцеві/монофонічні сухі хрипи/поліфонічні сухі хрипи/стидор/свист/стандарт" [43]. |
| Системи, які розпізнають конкретні патерни звуків легень та класифікують | Пошук конкретних захворювань (двохфакторна класифікація | <ul style="list-style-type: none"> • Двофакторна класифікація "стандарт/астма", "стандарт/пневмонія", "стандарт/ХОЗЛ", |

| Тип системи | Мета системи | Приклади реалізації |
|--|---|--|
| діагнози або захворювання, що їм відповідають. | “захворювання-стан дарт”). | “пневмонія/астма”, “астма/ХОЗЛ” [58] |
| | Діагностика захворювань (багатофакторна класифікація респіраторних захворювань, розпізнавання конкретних патернів дихання, які містять повний набір звуків легень, характерних для певного захворювання). | <ul style="list-style-type: none"> ● Стандарт / астма / пневмонія [15] ● Стандарт / астма / пневмонія / BRON/ ХОЗЛ / серцева недостатність [16], [53] ● Стандарт / ХОЗЛ / ІВДШ/ пневмонія / бронхектази / бронхіоліти [35] ● Стандарт / ХОЗЛ / ІВДШ/ ІНДШ/ пневмонія / бронхектази / бронхіоліти / астма [36] ● Стандарт / ХОЗЛ / пневмонія / бронхектази / бронхіоліти / ІВДШ/ ІНДШ/ астма [45] ● Стандарт / астма / BRON/ ХОЗЛ, серцева недостатність / легенева фіброз / плевральний випіт / пневмонія [42] ● Стандарт / астма / бронхектази / бронхіоліти / ХОЗЛ / ІНДШ , пневмонія / SARS [42] ● Стандарт / астма / легенева фіброз / бронхіт / |

| Тип системи | Мета системи | Приклади реалізації |
|-------------|--------------|---|
| | | <p>ХОЗЛ / серцева недостатність / серцева недостатність + ХОЗЛ / серцева недостатність + легеневий фіброз / легеневий фіброз / плевральний випіт / пневмонія [42]</p> <p>● Стандарт / астма / пневмонія / серцева недостатність, бронхоектази / бронхіт / ХОЗЛ [53]</p> |

З практичної точки зору системи, які виконують будь-які з вищезазначених завдань, можуть бути використані в клінічній практиці. Наприклад, класифікація дихальних звуків як "патологія/стандарт" може бути корисною для початкового огляду широкого спектру пацієнтів. Результати такого огляду можуть бути причиною подальшої діагностики пацієнтів, чиї дихальні звуки містять патологічний шум. Класифікація конкретних аномальних звуків легень і багатофакторна класифікація повинні допомагати медичним працівникам у початковій діагностиці пацієнтів, сприяти у прийнятті рішень і вдосконалювати точність аускультатії клініків. Сценарії, спрямовані на пошук конкретних захворювань, можуть потенційно використовуватися для скринінгу. Скринінг можна охарактеризувати як завдання класифікації, де вхідними даними є дихальні звуки, а вихідним – прогноз (так/ні, високий ризик/низький ризик тощо) для конкретних респіраторних станів.

Нарешті, системи, які виконують останнє, найскладніше завдання, можуть бути використані в телемедицині, автоматизації щоденної клінічної практики, діагностиці людей, що проживають в віддалених і ізольованих

районах, а також самодіагностиці за допомогою носимих пристроїв – найпопулярніших і найбільш перспективних напрямків медичної допомоги. Модель автоматизованої системи для аналізу дихальних звуків, запропонована в поточній роботі, належить до цих напрямів.

3.3.2. Порівняння результатів поточної роботи з відповідними дослідженнями

На основі вищезазначеної класифікації очевидно, що для відповідного порівняння результатів поточної роботи варто розглядати лише автоматизовані системи для класифікації певних аномалій дихальних шляхів (ХОЗЛ, пневмонія, бронхіт, астма і т.д.) на основі розпізнавання їх вроджених патологічних звуків легень, що можуть бути визначені як системи для автоматизованої діагностики легеневих захворювань. Дуже мало досліджень у літературі пропонують моделі для класифікації багатьох респіраторних захворювань, що передусім пов'язано з складністю завдання та відсутністю анотованих даних. Більшість цих досліджень використовують базу даних ICBNI 2017 або свої набори даних. Список та загальні характеристики моделей автоматизованої діагностики легеневих захворювань, знайдених у літературі, наведено у Таблиці 3.3.2.1.

Таблиця 3.3.2.1.

Характеристики моделей для багатофакторної класифікації
респіраторних захворювань

| Дослідження | Методологія | База даних | Класи | Метрики оцінювання |
|------------------|-------------------------------------|---------------------|------------------------------|---------------------------------|
| Лікві Ши та інші | Попередньо навчену мережу VGGish та | Власний набір даних | Стандарт / астма / пневмонія | Точність, повнота, оцінка F_1 |

| Дослідження | Методологія | База даних | Класи | Метрики оцінювання |
|---|---|--|---|--|
| співавтори (2019) [15]. | бінаправлену мережу керованих рекурентних одиниць (BiGRU) | | | |
| Бірюк Абер Тесема та інші співавтори (2022) [36]. | Моделі машинного навчання: <ul style="list-style-type: none"> ▪ Наївний байєсівський класифікатор ▪ <i>KNN</i>, ▪ Ансамблевий алгоритм (RUS Boosted та Bagged trees) ▪ <i>SVM</i> | Власний набір даних та вибірка з бази даних ICBHI 2017 | Стандарт / астма / пневмонія / ІВДШ/ ІНДШ/ ХОЗЛ / бронхектази / бронхіоліти | Точність, чутливість, специфічність |
| Джейн Салданья та інші співавтори (2022) [45] | 5 архітектур глибокого навчання: <ul style="list-style-type: none"> ▪ Багатошарний перцептрон, ▪ <i>CNN</i> *, ▪ <i>LSTM</i>, ▪ Попередньо навчена модель <i>Resnet-50</i>, | Доповнюючий набір даних ICBHI 2017 | Стандарт / ХОЗЛ / пневмонія / бронхектази / бронхіоліти / ІВДШ/ ІНДШ/ | Специфічність, чутливість, міра точності, оцінка F_1 |

| Дослідження | Методологія | База даних | Класи | Метрики оцінювання |
|---|--|--|--|--|
| | ▪ <i>Архітектура Efficient Net B0</i> | | | |
| Марія Тереза Гарсія-Ордас та інші співавтори (2020) [35]. | <i>CNN</i> * | Доповнюючий набір даних ICBHI 2017 | Два набори даних: • Стандарт / хронічні стани / не хронічні стани • Стандарт / ХОЗЛ / пневмонія / бронхектази / бронхіоліти / ІВДШ** | Специфічність, чутливість, міра точності, міра повноти, оцінка F_1 |
| Мохаммад Фрейвана та інші співавтори (2022) [16]. | 3 моделі глибокого навчання: ▪ <i>CNN</i> ,* ▪ бінаправлене LSTM ▪ <i>CNN</i> ensemble + bi-directional LSTM. | Власний набір даних та вибірка з бази даних ICBHI 2017 | Стандарт / астма / пневмонія / BRON/ ХОЗЛ / серцева недостатність | Специфічність, чутливість, міра точності, оцінка F_1 |
| Луай Фрайван та інші співавтори (2020) [53]. | Ансамблевий класифікатор на основі дерева рішень | Власний набір даних та вибірка з бази даних ICBHI 2017 | Стандарт / астма / пневмонія / BRON/ ХОЗЛ / серцева недостатність | Специфічність, чутливість, міра точності, оцінка F_1 |

| Дослідження | Методологія | База даних | Класи | Метрики оцінювання |
|--|--|---|--|---|
| Алі Мохаммад Алькода та інші співавтори (2022) [42]. | 3 моделі глибокого навчання: ▪ <i>CNN</i> , * ▪ <i>LSTM</i> , ▪ <i>CNN ensemble</i> + <i>LSTM</i> . | Бази даних ICBHI 2017, база даних <i>KAUH</i> . | Чотири набори даних: ● Стандарт / хронічні захворювання / нехронічні захворювання ● Стандарт / астма / бронхектази / бронхіоліти / ХОЗЛ / ІВДШ/ пневмонія / SARS.** ● Стандарт / астма / бронхіт / ХОЗЛ / серцева недостатність / фіброз легенів / плевральний ефузія / пневмонія ● Стандарт / астма / фіброз легенів / бронхіт / ХОЗЛ / серцева недостатність / серцева недостатність + | Чутливість, специфічність, правильність, точність |

| Дослідження | Методологія | База даних | Класи | Метрики оцінювання |
|-------------|-------------|------------|---|--------------------|
| | | | ХОЗЛ / серцева недостатність + фіброз легенів / фіброз легенів / плевральний ефузія / пневмонія | |

*Примітки: * – найбільш відповідна модель глибокого навчання, вибрана для порівняння з результатами поточного дослідження; ** – найбільш відповідні набори даних, вибрані для порівняння з результатами поточного дослідження.*

У своєму дослідженні Лікві Ши та спів. (2019) здійснили потрібну класифікацію захворювань легень (стандарт/астма/пневмонія) за допомогою алгоритму на основі перенесеного навчання. Цей алгоритм поєднував попередньо навчену мережу VGGish з двонаправленою рекурентною нейронною мережею з блоками GRU (BiGRU). Для експерименту дослідники використовували власний набір даних, зібраний від здорових добровольців та пацієнтів з пневмонією та астмою. Збір даних проводився професійними лікарями за допомогою електронного стетоскопа 3M Littmann 3200. Під час збору даних отримувалася 30-секундна аудіозапис дихальних звуків кожної особи.

Через різноманітність ступенів тяжкості захворювань та умов запису якість деяких зібраних зразків була недостатньою. Щоб вирішити цю проблему, дослідники вибрали 384 з найкращих зразків типових звуків легень. Ці зразки включали 120 нормальних звуків легень, 156 звуків пневмонії та 108 звуків астми. В експерименті кожен з цих записів був додатково розділений на три 10-секундні сегменти, що призвело до отримання остаточного набору даних, що складається з 360 нормальних

звуків легень, 468 зразків дихання від пацієнтів з пневмонією та 324 від пацієнтів з астмою. У загальній складності набір даних містив 1152 10-секундні аудіозаписи.

Вхідні дані для моделі складали мел-спектрограми, отримані зібраними звуками легень. Експеримент охоплював шість компонентів: оцінка загальної продуктивності моделі, дослідження впливу серцевих тонів, вивчення різних методів частотно-часового аналізу, проведення перехідного навчання, оцінка різних рівнів повторного навчання та порівняння різних методологій. Для оцінки здатності моделі до узагальнення дослідники використовували техніку шестикратної крос-перевірки.

Результати показали, що модель, запропонована дослідниками, яка використовувала поєднання попередньо навченої мережі VGGish та двонаправленої рекурентної нейронної мережі з блоками GRU (BiGRU) через перенесене навчання, перевершила як традиційні моделі машинного навчання, такі як SVM, KNN і RF, так і певні моделі глибокого навчання, такі як CNN, у класифікації захворювань дихальної системи.

Бірук Аберра Тессема та його колеги (2022) використовували власний набір даних, отриманий за допомогою спеціально розробленого електронного стетоскопа у пацієнтів, що поступили до відділення пульмонології Медичного центру Джімма в Ефіопії. Їхній набір даних складався з записів від 47 осіб, які охоплювали різні респіраторні стани. Зокрема, він включав записи від 9 пацієнтів з астмою, 6 з бронхоектазією, 4 з пневмонією, 3 з ГРЗ, 6 з бронхіолітом, 7 з інфекціями нижніх дихальних шляхів, 6 з хронічним обструктивним захворюванням легень та 6 здорових учасників.

У загальній складності було зібрано 272 аудіозаписи з розподілом наступним чином: 60 для астми, 49 для бронхоектазії, 23 для пневмонії, 15 для ГРЗ, 47 для бронхіоліту, 48 для ІНДШ, 30 для стандартних звуків

дихання та 15 для ХОЗЛ. Крім того, дослідження включало записи звуків дихання з набору даних ІСВНІ 2017. Ці записи включали 10 для астми, 16 для бронхоектазії, 37 для пневмонії, 45 для інфекцій верхніх дихальних шляхів, 13 для бронхіоліту, 2 для ІНДШ, 45 для ХОЗЛ та 45 стандартних звуків дихання. Щоб забезпечити узгодженість даних, з бази даних ІСВНІ було відібрано лише звуки легень, отримані за допомогою одноканального стетоскопа. Цей ретельний процес відбору був спрямований на забезпечення однорідності та надійності набору даних, використаного для аналізу.

Для навчання та перевірки моделі машинного навчання в цьому дослідженні використовувалися загалом 500 звуків легень (287 з Медичного центру Джімма та 213 з онлайн-даних), включаючи 75 записів від здорових суб'єктів, 70 - від пацієнтів з астмою, 65 - від пацієнтів з бронхоектазією, 60 - від пацієнтів з пневмонією, ІВДШ і ХОЗЛ та 50 - від пацієнтів з ІНДШ.

Дослідники використовували традиційні алгоритми МН у поєднанні з ручним виявленням ознак, яке використовувало досить складний алгоритм на основі DWT. Перед навчанням моделей МН функції відбиралися шляхом ранжування результатів однофакторного дисперсійного аналізу. Було відібрано 13 функцій, включаючи формфактор, стандартне відхилення, середньоквадратичне відхилення, перехресний фактор, фактор зазору, імпульсний фактор, співвідношення сигнал/шум, співвідношення сигнал/шум та спотворення, смугу пропускання, пікову амплітуду, середнє значення, ексцес, пікову частоту.

Дані були розділені на навчальні та тестові набори за допомогою методу утримання. Після цього було реалізовано восьмикратну крос-перевірку, що дозволило кожній точці даних пройти перевірку у восьми окремих ітераціях. Зокрема, 80% даних (400 зразків) були призначені для навчання, а решта 20% (100 зразків) - для тестування.

Різні моделі, включаючи просту баєсову, модель найближчих сусідів (KNN), ансамблеву модель та моделі машин підтримки векторів (SVM), були навчені та порівняні для класифікації захворювань легень у кілька класів: нормальні, астма, пневмонія, ILD, ILDP, ХОЗЛ, бронхоектазія та бронхіоліт. Серед цих моделей SVM з тонким гаусовим ядром продемонструвала найкращі результати, досягнувши точності класифікації 97,8%. Відповідно, вона була обрана для подальшого покращення точності за допомогою методу байєсової оптимізації. Цей процес оптимізації призвів до кінцевої точності системи 98,8% (див. Таблицю 3.3.2.2).

Джейн Салданья та спів. (2022) [45] використовували набір даних ICBHI 2017, доповнений трьома різними варіаційними автокодерами: багат шаровим персептроном VAE (MLP-VAE), згортковим VAE (CVAE) та умовним VAE, які були навчені на мел-спектрограмах малих класів. У цій роботі було додано етап вилучення необхідних характеристик з аудіо за допомогою коефіцієнтів частотно-часового кепстра.

Для класифікації всіх шести класів захворювань, присутніх у наборі даних ICBHI 2017, включаючи ХОЗЛ / пневмонію / бронхоектазію / бронхіоліт / ІВДШ/ ІНДШ / астму / стандартні, було запропоновано п'ять архітектур глибокого навчання: MLP, CNN, LSTM, попередньо навчену модель Resnet-50 та архітектуру Efficient Net B0.

Усі моделі були навчені з використанням функції втрат категоричної крос-ентропії, оптимізатора Adam та фіксованої швидкості навчання 0,0001. Процес навчання виконувався як на незбалансованих, так і на доповнених наборах даних. Метрики оцінювання, включаючи специфічність, чутливість, точність та F1-оцінку, використовувалися для оцінки продуктивності моделей. Важливо зазначити, що кожен класифікатор пройшов навчання на чотирьох різних наборах: незбалансованому наборі та трьох доповнених наборах. Цей процес повторювався тричі, охоплюючи 100 епох. Приблизно 70% аудіосегментів

для кожного класу використовувалися для навчання як варіаційного автокодера (VAE), так і моделей класифікації. Решта сегментів були зарезервовані для тестування та оцінки продуктивності класифікаторів. Такий ретельний підхід до навчання та оцінки був прийнятий для забезпечення надійності та ефективності моделей у виявленні респіраторних захворювань.

Результати експерименту довели, що доповнення даних за допомогою варіаційних автокодерів значно покращує продуктивність класифікаторів. Автори роблять висновок, що MLP-VAE і CVAE перевершують умовний варіаційний автокодер за якістю синтезованих даних для всіх класифікаторів.

Найкращі результати, отримані в цьому дослідженні за допомогою попередньо навченої моделі Resnet-50 з набором даних, доповненим багат шаровим персептроном VAE, наведено в таблиці 3.3.2.2, а оцінки продуктивності, отримані для згорткової нейронної мережі, навченої на наборі даних, доповненому згортковим варіаційним автокодером, включені в таблицю 3.3.2.3 для надання найбільш релевантного порівняння з поточним дослідженням.

Крім того, Марія Тереза Гарсія-Ордас та спів. (2020) [35] використовували набір даних ICBHI 2017, доповнений трьома генеративними методами (методом синтетичного меншості ресемплінгу (SMOTE), адаптивним методом синтетичного ресемплінгу (ADASYN) та варіаційним автокодером (VAE)).

Для класифікації вони використовували згорткову нейронну мережу, а сам експеримент проводився у два етапи. На першому етапі проводилася потрібна класифікація хронічних станів / не хронічних станів / стандартних. На другому етапі класифікація здійснювалася за шістьма класами: ХОЗЛ / пневмонія / бронхоектазія / бронхіоліт / ІВДШ/ стандарт.

Перед навчанням дані розділялися (80/20), і проводилася 10-кратна крос-перевірка для уникнення випадкових ефектів. Метрики оцінювання включали чутливість, специфічність, точність та повноту, а також F1-оцінку.

Для порівняння з поточним дослідженням таблиці 3.3.2.2 та 3.3.2.3 містять результати дослідження Марії Терези Гарсії-Ордас та спів. (2020), отримані авторами на другому, більш складному етапі експерименту з шістьма різними класами, де модель CNN показала чутливість 98,8% та специфічність 98,6%, і досягла точності 99,4% [35].

У дослідженні, проведеному М. Фріваном та спів. (2022), набір даних складався з записів звуків легень, зібраних у 103 пацієнтів у лікарні короля Абдалли (Йорданія) за допомогою електронного стетоскопа. Крім того, дані від 110 пацієнтів були отримані з набору даних ICBHI 2017. У загальній складності в дослідженні брали участь сегментовані записи тривалістю 1/5 секунди від 213 пацієнтів.

Захворювання були розділені на шість класів: стандарт, астма, пневмонія, BRON, ХОЗЛ і серцева недостатність. Архітектура мережі глибокого навчання складалася з двох блоків: згорткової нейронної мережі (CNN) та двонаправленої довготривалої пам'яті (BDLSTM). Ці блоки оцінювалися як окремо, так і в комбінації. Відмінністю від інших досліджень було використання лише одновимірних (1D) сигналів, що складали невеликий фрагмент 5-секундних записів звуків легень, як вхідних даних для моделей глибокого навчання. У цьому дослідженні не застосовувалася жодна методика доповнення даних.

Щоб вирішити проблему дисбалансу класів, було додано модифікований шар класифікації до кінця мереж CNN, BDLSTM та CNN+BDLSTM. Цей шар обробляв втрату суми квадратів помилок, коли мітки даних були нерівномірно розподілені. Вага кожного класу (w_d) визначалася за допомогою наступного рівняння:

$$w_d = 1 - \frac{N_d}{N_t} \quad (3.3.2)$$

де N_d та N_t — кількість зразків у класі та загальна кількість зразків відповідно.

Оцінка навчання моделі проводилась з використанням 10-кратної крос-валідації, і використовувались різні метрики продуктивності. Ці метрики включали точність, чутливість, специфічність, точність і F1-оцінку. Серед класифікаторів, що оцінювалися, ансамблевий класифікатор (CNN + двонаправлений LSTM) показав найвищу класифікаційну продуктивність. Зокрема, ансамблевий класифікатор досяг середньої точності 99,62%, чутливості 98,43%, специфічності 99,69%, точності 98,7% і F1-оцінки 98,56% (див. таблицю 3.3.2.2). Для надання релевантного порівняння з поточним дослідженням у таблиці 3.3.2.3 представлені середні значення для моделі CNN.

У дослідженні, проведеному Луайем Фріваном та спів. (2020), для класифікації респіраторних захворювань використовувався ансамблевий класифікатор на основі дерев рішень. Функції, використані в цьому дослідженні, включали ентропію Шеннона, логарифмічну енергійну ентропію та спектральну ентропію, отримані зі спектрограми. Оптимальну конфігурацію для ансамблевих моделей було визначено за допомогою байєсівської оптимізації гіперпараметрів. Вибірка дослідження складалася з 215 суб'єктів, з загальною кількістю 308 клінічних записів звуків легень, отриманих за допомогою електронного стетоскопа. Додатково було включено 1176 записів з бази даних ICBHI 2017, які були класифіковані на такі класи: стандарт, астма, пневмонія, серцева недостатність, бронхоектазія (BRON) і ХОЗЛ.

Порівняння між ансамблевим класифікатором та звичайними моделями машинного навчання показало перевагу ансамблевого підходу. Ансамблевий класифікатор досяг середньої точності, чутливості,

специфічності та F1-оцінки 98,27%, 95,28%, 98,9% і 93,61% відповідно. Ці результати підкреслюють ефективність ансамблевого класифікатора у точній класифікації респіраторних захворювань (див. таблицю 3.3.2.2).

Алі Мохаммад Алкуда та колеги (2022) досліджували продуктивність трьох різних моделей глибокого навчання: CNN, LSTM і CNN + LSTM на не підвищених і підвищених наборах даних. Сигнали, використані в дослідженні, були отримані шляхом об'єднання двох загальнодоступних наборів даних: набору даних ICBHI 2017 та набору даних звуків дихання з лікарні короля Абдалли (KAUH). Ці записи представляли сигнали від пацієнтів з різними респіраторними захворюваннями, такими як астма, легеневий фіброз, бронхоектазія (BRON), хронічне обструктивне захворювання легень (ХОЗЛ), серцева недостатність, серцева недостатність + ХОЗЛ, серцева недостатність + легеневий фіброз, плевральний випіт, пневмонія, а також записи від здорових людей.

Для експериментів були створені чотири набори даних. Перший набір даних був позначений для потрійної класифікації, що складалася з класів хронічне захворювання, не хронічне захворювання та стандарт. Другий набір даних включав позначення для восьми класів: стандарт, астма, бронхоектазія, бронхіоліт, ХОЗЛ, ІВДШ, пневмонія та тяжкий гострий респіраторний синдром (SARS). Третій набір даних складався з восьми класів: стандарт, астма, бронхіоліт, ХОЗЛ, серцева недостатність, легеневий фіброз, плевральний випіт і пневмонія. Нарешті, четвертий набір даних містив позначення для 11 класів: стандарт, астма, легеневий фіброз, бронхоектазія, ХОЗЛ, серцева недостатність, серцева недостатність + ХОЗЛ, серцева недостатність + легеневий фіброз, плевральний випіт і пневмонія.

Для доповнення даних і нормалізації набору даних використовувалися різні техніки, включаючи уповільнення або прискорення звуку, зсув звуку вліво або вправо, додавання випадкового шуму та регулювання гучності

звучу. Однак, незважаючи на ці методи доповнення, експериментальні результати не показали значного покращення продуктивності моделі. Коли застосовувалось доповнення, точність трохи покращилась до 100%, 99,8%, 98,0% і 99,5% для наборів даних 1, 2, 3 і 4 відповідно. Хоча деякі покращення були, вони не були суттєвими у всіх наборах даних, що вказує на те, що методи доповнення не значно покращили продуктивність моделі.

Досягнута правильність / чутливість / специфічність / точність становили 100%, 100%, 100% і 100% при використанні гібридної моделі CNN-LSTM для трифакторної класифікації на доповненому наборі даних, але її продуктивність трохи зменшувалась при виконанні більш складних завдань восьмифакторної та одинадцятифакторної класифікації. Точність / чутливість / специфічність / точність для другого набору даних становили 99,81%, 99,97%, 99,97% і 99,16%, а для третього набору даних 98,03%, 96,47%, 99,65% і 99,21%. У четвертому і найскладнішому завданні ці показники становили 99,53%, 98,8%, 99,92% і 99,47% відповідно.

Як релевантні дані для порівняння результатів цього дослідження з результатами поточної роботи (див. таблицю 3.3.2.3), були обрані оцінки продуктивності моделі CNN для другого доповненого набору даних, що містив позначення восьми класів (стандарт, астма, бронхоектазія, бронхіоліт, ХОЗЛ, ІНДШ, пневмонія та SARS).

Таблиця 3.3.2.2.

Найкращі результати, отримані автоматизованими системами
діагностики дихальних шляхів

| Дослідження | Класифікатор, який показав найкращі результати | Кількість класів | Міра правильності | <i>F1</i> |
|--|---|------------------|-------------------|-----------|
| Лікві Ши та інші співавтори (2019) [15] | Попередньо навчена мережа VGGish та бінаправлена мережа керованих рекурентних одиниць (BiGRU) | 3 | НД | 87,41% |
| Бірюк Абер Тесема та інші співавтори (2022) [36] | <i>Тонка Гауссова SVM з байєсівською оптимізацією</i> | 8 | 98,8% | НД |
| Джейн Салданья та інші співавтори (2022) [45] | Попередньо навчена модель Resnet-50 з доповненим набором даних з MLP-VAE | 7 | НД | 74% |
| Марія Тереза Гарсія-Ордас та інші співавтори (2020) [35] | <i>CNN</i> | 6 | 99,4% | 90% |

| Дослідження | Класифікатор, який показав найкращі результати | Кількість класів | Міра правильно сті | <i>F1</i> |
|---|--|------------------|--------------------|-----------|
| Мохаммад Фрейвана та інші співавтори (2022) [16] | Ансамблевий класифікатор CNN + бінаправлене LSTM | 6 | 99,62% | 98,56% |
| Луай Фрайван та інші співавтори (2020) [53]. | Ансамблевий класифікатор на основі дерева рішень | 6 | 98,27% | 93,61% |
| Алі Мохаммад Алькода та інші співавтори (2022) [42] | Ансамблевий класифікатор CNN + LSTM | 8 | 99,81%, | НД |
| Поточні дослідження | CNN | 5 | 98,44% | 98,45% |

Примітка: НД - немає даних

Оскільки поточне дослідження класифікувало респіраторні захворювання за допомогою конволюційної нейронної мережі на п'ять класів на наборі даних ICBHI 2017, доповненому та нормалізованому за допомогою синтетичних даних, згенерованих конволюційним варіаційним автоенкодером, найбільш релевантні результати відповідних досліджень, тобто метрики для оцінювання продуктивності моделі CNN, отримані в подібних дослідженнях, були вибрані та наведені в Таблиці 3.3.2.3.

Таблиця 3.3.2.3

Порівняння метрик оцінювання продуктивності моделей CNN для багатofакторної класифікації станів, що впливають на респіраторну систему, запропонованих у поточній роботі та відповідних дослідженнях

| | Джейн Салдань я та інші співавтори (2022) [45] | Марія Тереза Гарсія-Ор дас та інші співавтори (2020) [35] | Мохаммад Фрейвана та інші співавтори (2022) [16] | Алі Мохаммад Алькада та інші співавтори (2022) [42] | Поточна робота |
|-------------------------------|---|--|---|--|---------------------------|
| Кількість класів | 7 | 6 | 6 | 8 | 5 |
| Чутливість | 96,0% | 98,8% | 93,02% | 99,97% | 98,44% |
| Специфічність | 62,0% | 98,6% | 99,33% | 99,97% | 99,61% |
| Точність | НД | НД | 99,04% | 99,81% | 98,44% |
| Прецизійність | 76,0% | 99,4% | 96,88% | 99,66% | НД |
| Повнота | НД | 98,6% | НД | НД | НД |
| F_1 score | 65,0% | 90,0% | 94,68% | НД | 98,45% |

Примітка: НД - немає даних

Результати, представлені в Таблиці 3.3.2.3, показують, що поточне дослідження досягло високих результатів у точності класифікації,

перевершило середні показники чутливості та специфічності моделі, а також покращило оцінку F1 серед усіх порівнюваних досліджень

3.3.3. Особливості реалізації та обмеження систем автоматичної класифікації

Залежно від завдань, покладених на конкретну систему автоматичної класифікації легеневих звуків, кожна з них має свої обмеження та особливості реалізації. Наприклад, завдання пошуку конкретних звуків дихання та їх двофакторної класифікації досить ефективно виконуються системами на основі традиційних методів машинного навчання. Однак для трифакторної або багатофакторної класифікації ці моделі зазвичай поступаються за продуктивністю системам на основі глибокого навчання (ГН), зокрема нейронних мереж. Перш за все, це пов'язано з тим, що використання моделей машинного навчання (МН) потребує ручного вибору функцій, що залежить від знань дослідників у певній галузі та значною мірою визначає ефективність моделі.

Деякі дослідники, такі як Захангір Алом та спів. (2019) [60], вважають, що продуктивність традиційних моделей машинного навчання вже досягла певного плато, і нові дослідження в цій галузі не можуть досягти значного покращення, тоді як інші, такі як Луай Фрайван та спів. (2020) [53] або Бірюк Абер Тесема та спів. (2022) [36], намагаються як покращити точність таких моделей шляхом об'єднання кількох класифікаторів в єдиний ансамбль та застосування різних методів оптимізації, так і застосувати їх до набагато складнішого завдання - класифікації респіраторних захворювань. Однак їхні досить високі результати є сумнівними з причин, які будуть розглянуті в наступному розділі роботи. Більшість досліджень демонструють, що ручне виділення ознак та єдиний класифікатор не підходять для складних завдань класифікації.

Для цього завдання частіше використовуються глибокі нейронні мережі, оскільки вони працюють з великими обсягами даних та мають високу масштабованість для вирішення завдань різної складності. Однак вони втрачають ефективність, коли завдання класифікації стають складнішими, що підтверджується кількома дослідженнями, зокрема, Бруно Роше та спів. (2021). У сценарії двофакторної класифікації ("вологий хрип проти інших") конволюційна нейронна мережа з мел-спектрограмами як вхідними даними показала точність 99.4%, тоді як у завданні трифакторної класифікації її точність становила 96.7%. При цьому зниження ефективності традиційних класифікаторів машинного навчання було більш помітним: зокрема, точність методу опорних векторів (SVM) у першому завданні становила 93.5%, а у другому - 89.7%. Таким чином, при переході від двофакторної до трифакторної класифікації, CNN втратила 2.5% точності, а SVM - 4%. [50]

Зниження ефективності класифікації звуків дихання зі збільшенням складності завдань ілюструється в роботі Мурата Айканата та спів.(2017). Використовуючи досить велику базу даних, що містить 17930 записів звуків легенів, зібраних у 1630 осіб за допомогою саморозробленого електронного стетоскопа, автори порівняли ефективність тих самих класифікаторів для вирішення завдань різної складності:

- двофакторна класифікація (здоровий звук проти патологічного звуку);
- трифакторна класифікація (вологий хрип, сухий хрип та стандартний звук);
- багатофакторна класифікація (визначення типу звуків дихання);
- класифікація типів аудіо з усіма типами звуків.

Використовуючи конволюційну нейронну мережу, яка використовувала 28x28 мел-спектрограми, перетворені у відтінки сірого, як вхідні дані, вони отримали наступні результати: для першого завдання

(двофакторна класифікація) точність CNN склала 86%, для другого - 76%, а для третього - 80%. Але при вирішенні четвертого, найскладнішого завдання, цей показник знизився до 62% [61].

На основі результатів цих досліджень слід припустити, що при виконанні найскладнішого завдання, пов'язаного з діагностикою респіраторних станів шляхом аналізу звуків аускультації легенів, ефективність алгоритмів ГН, і ще більше алгоритмів МН, не може бути надто високою. Однак у дослідженнях, представлених у Таблицях 3.3.2.2 та 3.3.2.3, показники продуктивності дуже високі. Крім того, подібні результати були отримані в експерименті, описаному в цій роботі.

Отримання дуже високих оцінок для моделей, які вирішують таке складне і складне завдання, як діагностика респіраторних захворювань, може саме по собі служити "червоним прапором" для методологічної помилки, що призводить до витоку даних.

3.3.4. Проблема витоку даних

3.3.4.1. Причини витоку даних

Загалом техніка розділення навчання та тестування (відомий як контейнмент), яка передбачає поділ наявних даних на дві частини: одну для навчання та іншу для оцінки, вважається надійним запобіжником проти витоку даних. Щоб оптимізувати продуктивність моделі, розмір навчальної вибірки повинен задовольняти дві умови:

- представляти набір даних в цілому,
- бути досить великим для отримання значущих результатів, але не надто великим, щоб уникнути перенавчання моделі.

У поточному дослідженні перша умова виконувалась за допомогою стратифікованого поділу, що дозволяє зберегти відносні пропорції кожного

класу у навчальних та валідаційних наборах. Як результат, модель навчається на всіх класах, що підвищує продуктивність та надійність прогнозів алгоритму машинного навчання та уникає упередженості у фінальній моделі. Щоб дотриматися другої умови, було обрано розподіл 80/20, де 80% даних належить навчальній вибірці і 20% - тестовій. Варто зазначити, що застереження щодо розділення даних стосується техніки крос-валідації. Набори валідації повинні бути приховані, щоб залишатися невідомими для моделі. Таким чином, у поточній роботі та у відповідних дослідженнях, представлених у Таблиці 3.3.2.1, всі формальні вимоги щодо запобігання витоку даних були дотримані. Однак природа медичних даних, використаних у цих дослідженнях, зокрема записи аускультатії легенів, зібрані у базі даних ICBHI 2017 та інших наборах звуків дихання, натякає на наявність іншого специфічного механізму витоку даних з навчальної до тестової вибірки. Можна припустити, що головна проблема полягає в тому, що дані, зібрані від одного пацієнта, зазвичай записуються протягом короткого проміжку часу, в однакових умовах (з однаковим рівнем та характером фонових шумів) і з використанням одного й того самого обладнання. Це робить їх дуже схожими один на одного, і саме ця схожість сприяє потенційній причині витоку даних. Зокрема, база даних ICBHI 2017 містить 920 аудіофайлів аускультатії легенів та 6898 записів циклів аускультатії легенів, зібраних у 126 пацієнтів. Таким чином, в середньому, до бази даних включено 7.8 записів та 54.7 циклів від кожного пацієнта. Коли цикли, що походять від одного пацієнта, одночасно включені до навчальної та тестової вибірок, може відбутися витік даних через їх схожість, що призводить до того, що модель демонструє надто оптимістичні результати.

3.3.4.2. Тестування гіпотези: другий етап експерименту

Щоб перевірити цю гіпотезу та детальніше дослідити вплив розподілу даних на продуктивність моделі автоматичної класифікації звуків дихання, експеримент, описаний у цьому дослідженні, був повторений з іншим розподілом даних. Розподіл даних був групуваний за суб'єктами таким чином, що аудіозаписи суб'єктів з тестової вибірки не перекривалися з аудіозаписами суб'єктів з навчальної вибірки. У той час як у головному експерименті різні аудіозаписи, отримані від одного пацієнта, могли бути включені як до навчальної, так і до тестової вибірок, у другому експерименті вони були включені лише в одну з цих груп. Таким чином, аудіозаписи звуків дихання, отримані від однієї людини, були включені або до навчального, або до тестового набору даних. Експеримент був повторений з новим розподілом даних, як описано в розділі 3.1.5, і результати були оцінені, як описано в розділі 3.2.3, за допомогою десятикратної крос-валідації. Результати другого експерименту представлені в Таблиці 3.3.4.1

Таблиця 3.3.4.1.

Оцінка ефективності (%) запропонованої моделі з групуванням ідентифікаторів пацієнтів на основі десятиразової крос-валідації

| Діагноз | Чутливість, % | Специ- фічність, % | Точність, % | Параметр капа Кохена,% | F1-score, % |
|---------------------|------------------|--------------------------|----------------|------------------------------|----------------|
| Здоровий пацієнт | 28.57 | 90.34 | | | 26.67 |
| ХОЗЛ | 93.08 | 40.17 | | | 91.93 |
| BRON | 16.67 | 100.00 | | | 28.57 |

| | | | | | |
|-----------------------------|--------------|--------------|--------------|--------------|--------------|
| Пневмонія | 22.22 | 95.95 | | | 26.67 |
| ІВДШ | 0.00 | 88.65 | | | 0.00 |
| Усього (середнє) | 32.11 | 83.02 | 82.70 | 27.48 | 34.76 |

Результати другого варіанту експерименту, де тренувальні та тестові набори були розділені таким чином, що аудіозаписи респондентів з тестового набору не перекривалися з аудіозаписами респондентів з тренувального набору (Таблиця 3.3.4.1), досить середні і суттєво відрізняються від результатів першого етапу. Ця різниця свідчить про потенційні методологічні недоліки і вказує на те, що в першому експерименті відбулося деяке витікання даних з тренувального набору в тестовий, що не було виправлено крос-валідацією.

Виявлена проблема витікання даних, ймовірно, є специфічною для медичних даних, а саме для аудіозаписів сигналів респіраторної аускультатії, оскільки різні записи та їх сегменти, зібрані у пацієнтів за короткий період у тих же умовах, є дуже схожими.

Враховуючи відносно високий F_1 -рахунок і чутливість для класу ХОЗЛ на другому етапі експерименту на фоні значно гіршої продуктивності для інших класів, можна припустити, що зазначена проблема витікання даних впливає на збільшення даних за допомогою згорткового автоенкодера. Збільшення для інших класів з відносно вузькими розподілами було недостатнім для узагальнення їх характеристик та досягнення ідеальної класифікації. Відносна недостатність узагальнення, про що свідчать низькі оцінки продуктивності для менш значних класів, може бути пояснена необхідністю синтезу великої кількості даних з доступних досить малих масивів схожих даних.

Таблиця 3.3.4.2

Співвідношення кількості пацієнтів, оригінальних та синтезованих записів для кожного класу

| | Кількість пацієнтів | Кількість оригінальних записів | Кількість синтезованих записів |
|-------------------------|--------------------------------|---|---|
| ІВДШ | 14 | 23 | 758 |
| BRON | 13 | 29 | 760 |
| Пневмонія | 6 | 37 | 764 |
| Здоровий пацієнт | 26 | 35 | 765 |
| ХОЗЛ | 64 | 793 | 0 |

Співвідношення кількості пацієнтів, кількості оригінальних записів, отриманих від них, та кількості синтезованих записів, використаних у поточному дослідженні, представлене в Таблиці 3.3.4.2 і зображене на Рисунку 3.3.4.1. Очевидно, що кількість пацієнтів (окрім людей з ХОБЛ), залучених до створення набору даних ІСВНІ 2017, була недостатньою для якісного узагальнення, що призвело до витікання даних.

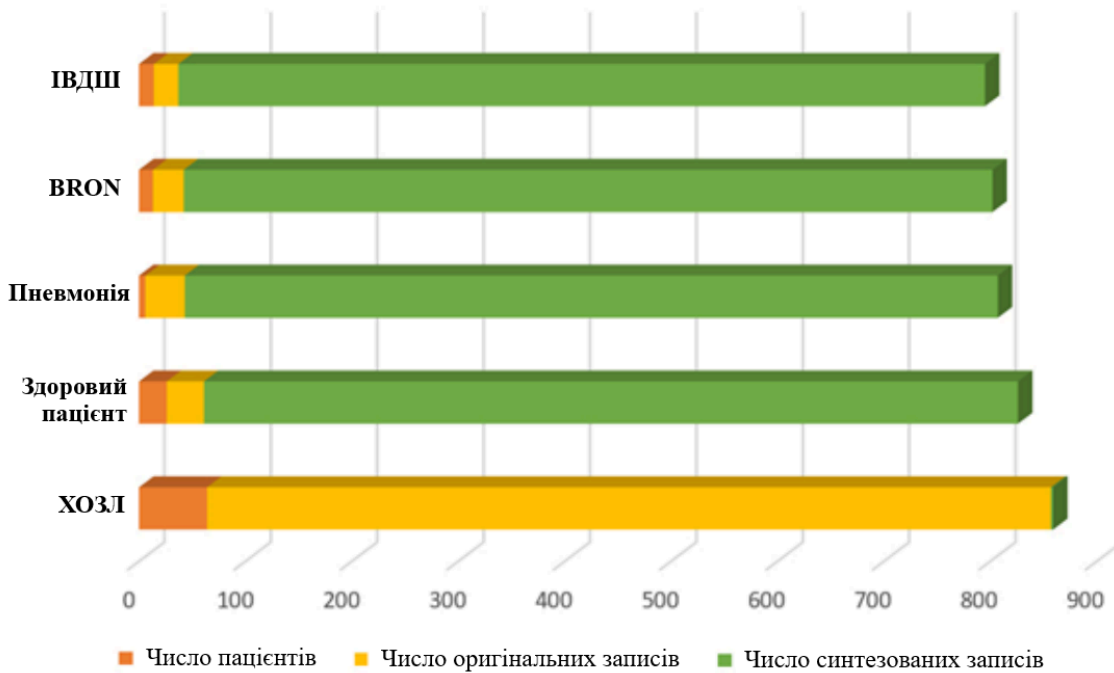


Рис. 3.3.4.1. Гістограма розподілу кількості пацієнтів, оригінальних і синтезованих записів для кожного класу

Синтезовані дані, отримані зі спектрограм менш представлених класів, є достатньо якісними для використання. Але, на жаль, обмеження набору даних і варіації технічних характеристик записуючих пристроїв у наявному наборі даних не дозволяють однозначно розподілити та узагальнити нові дані, що частково пояснює результати другого експерименту. Інакше кажучи, клас ХОЗЛ має позитивніші оцінки порівняно з іншими класами, ймовірно, саме тому, що він не містить синтезованих записів.

3.3.4.3. Проблема витоку даних у подібних дослідженнях

На жаль, проблема потенційного витоку даних, що виникає через схожість записів, отриманих від одного й того самого пацієнта, майже не висвітлюється в літературі та не розглядається під час розробки

автоматизованих систем класифікації дихальних звуків як один з факторів, що обмежують дослідження та необхідність коригування його методології.

Таблиця 3.3.4.3. Загальні параметри розробки набору даних та його розподіл, використаний у багатокласовій класифікації респіраторних патологій у подібних дослідженнях.

| | [36] | [45] | [35] | [16] | [53] | [42] |
|--|--|--|--|--|--|--|
| База даних | Власний набір даних та вибірка з набору даних ICBHI 2017 | Додатковий набір даних ICBHI за 2017 рік | Додатковий набір даних ICBHI за 2017 рік | Власний набір даних та вибірка з набору даних ICBHI 2017 | Власний набір даних та вибірка з набору даних ICBHI 2017 | <i>Набір даних ICBHI 2017, набір даних КАУН*</i> |
| Кількість пацієнтів | 125 | 126 | 124 | 213 | 215 | 196 |
| Кількість записів | 500 | 40614 | 4874 | 1182 | 1448 | 3977 |
| Коефіцієнт утримання | 80/20 | 70/30 | 80/20 | НД | НД | 86,5/13,5 |
| k (Кратність перехресної перевірки) | 8 | НД | 10 | 10 | 10 | НД |

Примітки: НД- немає даних; * - дані надані для другого розширеного набору даних з 8 категоріями маркування, який забезпечив найкращі результати для багатокласової класифікації респіраторних захворювань (Таблиця 3.3.2.2).

Як показано аналізом подібних досліджень, обговорених детально вище (див. Таблицю 3.3.2.3), більшість авторів не приділяли увагу

проблемі витоку даних, пов'язаній з цими особливостями записів аускультації легенів. Таблиця 3.3.4.3 демонструє, що в кожному з цих досліджень кількість пацієнтів була значно меншою за кількість записів або сегментів, використаних для навчання та тестування моделі. Іншими словами, велика кількість записів від одного пацієнта була потенційно дуже схожа між собою, що спричиняло витік даних і призводило до надто оптимістичних результатів класифікаторів.

Проблема витоку даних обговорювалася лише в одному з досліджень, представлених у Таблиці 3.3.4.2, а саме у дослідженні Джейн Салданья та спів. (2022) [45]. Однак, оскільки основною метою цієї роботи було тестування різних типів варіаційних автокодерів для розширення даних, автори підходили до цього з іншого боку. Їх підхід полягав у розподілі даних на навчальні та тестові набори до класифікації записів і до розширення набору даних за допомогою VAE.

Розподіл реальних аудіосегментів на навчальні та тестові набори, запропонований авторами, представлений на Рис. 3.3.4.2.

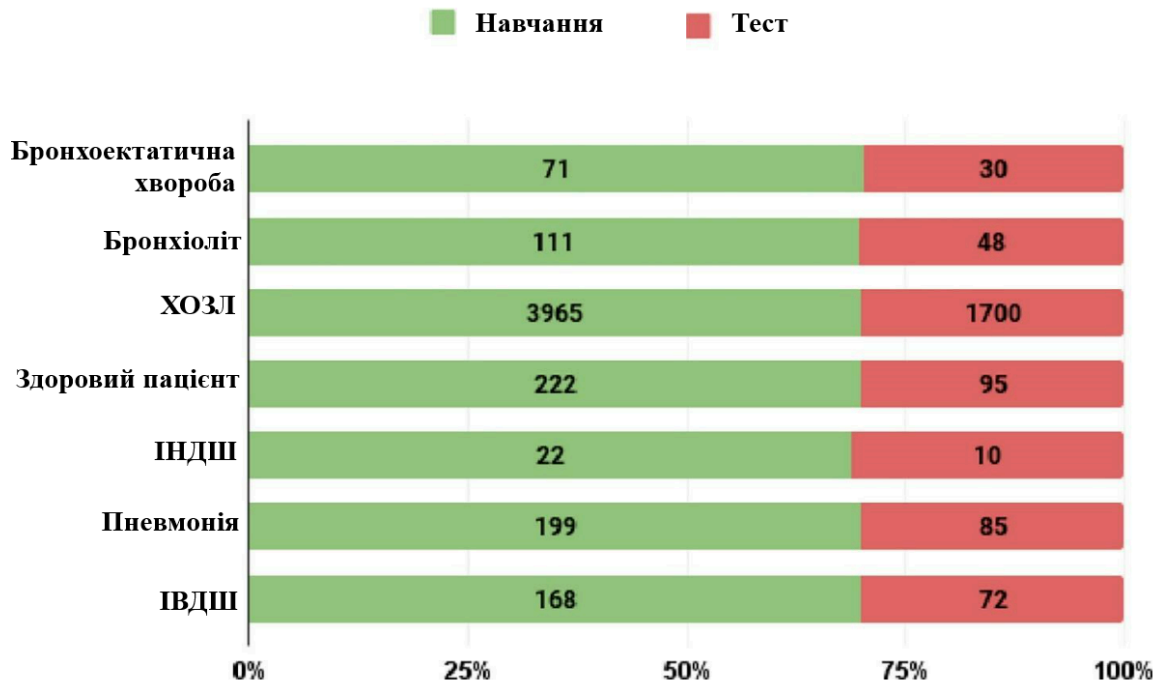


Рис. 3.3.4.2. Розподіл реальних аудіосегментів на навчальні та тестові набори в дослідженні Джейн Салданья та спів. (2022) [45]

На представлений гістограмі на Рис. 3.3.4.2 показано, що приблизно 70% аудіосегментів у кожному класі використовувалися для навчання як варіаційного автокодера (VAE), так і моделей класифікації.

Решту сегментів було відкладено для тестування та оцінки продуктивності класифікатора. Підхід до розподілу даних, запропонований авторами, дозволив точніше оцінювати якість записів, синтезованих різними автокодерами, оскільки в кожній категорії вони порівнювалися із збереженими наборами оригінальних даних. Крім того, це могло покращити точність оцінок продуктивності класифікатора, оскільки для навчання використовувалися як згенеровані, так і оригінальні записи, а тестувалися лише оригінальні записи. Таким чином, розподіл даних, представлений у цій роботі, вирішив проблему витoku даних через схожість синтезованих та оригінальних аудіозаписів, але не вирішив проблему схожості записів, отриманих від одного пацієнта.

Примітно, що навіть часткове вирішення проблеми витоку даних, запропоноване Джейн Салданья та спів. (2022) [45], значно вплинуло на продуктивність розробленої ними системи. Порівнюючи продуктивність моделей CNN для багатofакторної класифікації респіраторних захворювань у Таблиці 3.3.2.4 та загальну кількість записів, на яких навчалася система (Таблиця 3.3.4.3), можна помітити, що в цьому дослідженні були отримані найгірші результати, хоча обсяг даних був набагато більшим (Таблиця 3.3.4.4).

Таблиця 3.3.4.4.

Порівняння метрик продуктивності та кількості записів у системах автоматизованої класифікації респіраторних захворювань, реалізованих на основі згорткової нейронної мережі з використанням розширених даних

| | Джейн Салданья та інші співавтори (2022) [45] | Марія Тереза Гарсія-Ордас та інші співавтори (2020) [35] | Поточні дослідження без дезагрегації даних та безвідносно до групування суб'єктів | Поточні дослідження без дезагрегації даних на основі групування суб'єктів |
|----------------------|---|--|---|---|
| Кількість класів | 7 | 6 | 6 | 6 |
| Кількість записів | 40614 | 4874 | | |
| <i>Чутливість</i> | 96,0% | 98,8% | 98,44% | 32,11% |
| <i>Специфічність</i> | 62,0% | 98,6% | 99,61% | 83,02% |
| <i>Прецизійність</i> | 76,0% | 99,4% | НД | НД |

| | | | | |
|-------------|-------|-------|--------|--------|
| Точність | НД | НД | 98,44% | 82,70% |
| F_1 score | 65,0% | 90,0% | 98,45% | 82,70% |

Примітки: НД - немає даних;

Оскільки ефективність класифікації моделей глибинного навчання (ГН), зокрема згорткових нейронних мереж, безпосередньо залежить від кількості даних, очікувалося, що дослідження Джейн Салданья та спів. (2022) [45] матиме найвищу ефективність серед наведених досліджень. Проте протилежний результат, представлений у Таблиці 3.3.4.4, опосередковано підтверджує існування проблеми витоку даних, пов'язаної зі схожістю синтезованих та оригінальних записів, і свідчить про переоцінку ефективності інших досліджень.

Проблема витоку даних, пов'язана зі схожістю записів, отриманих від одного пацієнта, детально розглядається лише в одній роботі з автоматизованої класифікації звуків дихання. Дослідження Фу-Шун Хсу та спів. (2021) вивчало та порівнювало ефективність восьми типів нейронних мереж для виявлення фаз дихання та випадкових звуків у записах аускультатії легень [52]. Автори створили свою базу даних звуків легень, HF_Lung_V1, яка містить 9765 сегментів записів звуків легень (по 15 с кожен), зібраних у 279 пацієнтів. Записи були анотовані відповідними мітками: 34095 міток вдиху, 18349 міток видиху, 13883 міток безперервного випадкового звуку (включаючи 8457 міток сухих хрипів, 686 міток стридора та 4740 міток сухих хрипів), і 15606 міток переривчастих звуків (усі мокрі хрипи).

Вони помітили, що, оскільки кілька раундів аускультатії для кожного окремого пацієнта проводилися в одному місці протягом короткого часу (кілька годин), звукові патерни могли проявляти схожість. Щоб запобігти потенційному витоку даних, записи, зібрані в один день, включалися

виключно в один із наборів даних (навчальний, валідаційний або тестовий).

У цьому дослідженні співвідношення кількості записів у навчальному та тестовому наборах стабільно підтримувалося на рівні приблизно 4:1. Крім того, автори застосовували п'ятикратну перехресну перевірку всередині навчального набору даних для навчання та валідації моделей. Порівнювати результати цього дослідження з результатами поточної роботи складно, оскільки вони мали різні цілі, використовували різні бази даних та різні архітектури. Дослідження Фу-Шун Хсу та спів. (2021) було спрямоване на виявлення сегментів та подій (вдих, видих, переривчасті/безперервні випадкові звуки легень), тобто відповідно до класифікації, наведеної в Таблиці 3.3.1.1, виконувало найпростішу задачу (пошук конкретних звуків дихання), тоді як поточна робота, відповідно до тієї ж класифікації, виконувала найскладнішу задачу - діагностику респіраторних захворювань.

База даних HF_Lung_V1 перевершує базу даних ICBHI 2017 як за кількістю пацієнтів (279 проти 126), так і за кількістю аудіофайлів (9765 проти 6898).

Дослідження Фу-Шун Хсу та спів. (2021) не мало проблеми дисбалансу даних, оскільки класифікація проводилася лише за одним параметром. Фу-Шун Хсу та спів. (2021) тестували продуктивність кількох архітектур ГН, включаючи довгострокову короткострокову пам'ять (LSTM), спрямований рекурентний блок (GRU), двонаправлений LSTM (BiLSTM), двонаправлений GRU (BiLSTM) та ансамблеві класифікатори CNN-LSTM, CNN-GRU, CNN-BiLSTM і CNN-BiGRU.

У порівнянні з результатами інших досліджень, представлених у Таблиці 3.3.2.3, автори отримали досить скромні результати. Так, найвищу ефективність продемонстрував ансамблевий класифікатор CNN-BiGRU. При виявленні вдиху його F1-оцінка склала 86,3%, при виявленні видиху -

70,9%, при виявленні безперервних випадкових звуків - 51,6%, при виявленні переривчастих випадкових звуків - 70,8%. З огляду на численні спрощення та переваги (простішу класифікацію великої бази даних, відсутність дисбалансу тощо), результати дослідження Фу-Шун Хсу та спів. (2021) цілком порівнянні з результатами, отриманими в поточній роботі при проведенні експерименту з урахуванням групування даних за суб'єктами.

Таким чином, можна стверджувати, що зміни в розподілі даних, а саме усунення перетину записів, що належать одному пацієнту, у навчальних та тестових наборах, призводять до нижчих, але при цьому більш реалістичних показників продуктивності. Це підтверджує гіпотезу про те, що існує витік даних через схожість звуків легень, отриманих від однієї людини за короткий час. Таким чином, були виявлені два механізми витоку даних, притаманні системам автоматизованої класифікації звуків аускультатії легень.

Перший з них пов'язаний з розширенням даних за допомогою варіаційних автокодерів, техніки, що використовується для подолання проблеми дисбалансу класів даних у загальнодоступних наборах звуків легень, які використовуються для навчання та тестування таких систем. Варіаційні автокодери, і зокрема згортковий варіаційний автокодер, використаний у поточному дослідженні, генерують дані досить високої якості для використання у ГН, як було доведено в розділі 3.2.1. "Оцінка продуктивності генеративної моделі", але через схожість синтезованих записів з оригінальними відбувається витік даних, що завищує реальну продуктивність системи.

Другий механізм витоку даних пов'язаний зі схожістю записів, отриманих від одного пацієнта протягом короткого періоду за одних і тих же умов. На жаль, усі існуючі загальнодоступні набори даних звуків легень містять велику кількість записів, зібраних у відносно невеликій кількості

пацієнтів, що є основною причиною витоку даних та завищення продуктивності систем автоматизованої класифікації звуків або захворювань дихальної системи, які використовують ці набори даних для навчання та тестування.

3.4. Проблеми та виклики у розробці автоматизованих систем класифікації звуків дихання та діагностики респіраторних захворювань

3.4.1. Збір даних та їх конфіденційність

Аналіз результатів, отриманих у поточній роботі та кількох аналогічних дослідженнях, дозволяє зробити висновок, що основною проблемою для створення систем розпізнавання звуків дихання та діагностики респіраторних захворювань є недостатня кількість добре анотованих записів аускультатії легень, придатних для навчання моделей штучного інтелекту.

На момент написання роботи у публічному доступі можна знайти близько 10 баз даних звуків легень, які охоплюють різні типи звуків, включаючи звуки легень, дихання, кашлю та голосу, а також різні респіраторні захворювання, такі як астма, ХОЗЛ та COVID-19. Найбільша з них - ICBHI 2017, але вона містила записи лише 126 пацієнтів, що було недостатньо для належного навчання моделі для всіх класів респіраторних захворювань. Іншим недоліком був значний дисбаланс даних за класами, що вимагало доповнення даних для навчання моделей МН та ГН.

Теоретичним рішенням проблеми браку даних могло б стати об'єднання різних існуючих аудіобаз даних, проте це ускладнюється їхньою гетерогенністю за методологією збору даних. Крім того, деякі загальнодоступні бази даних збираються за допомогою краудсорсингу без клінічної валідації або містять самозвіти про стан здоров'я, що призводить

до великої кількості помилок при використанні для навчання моделей МН та ГН.

Більш реалістичним рішенням проблеми видається встановлення єдиного стандарту для запису звуків аускультатії, який включав би вимоги до збору даних, типу даних, тривалості, анотації тощо. Керуючись таким стандартом, можна було б створити єдину загальнодоступну базу даних респіраторних звуків, яку б створювали численні, але невеликі дослідницькі групи. У міру надходження нових даних про респіраторні звуки, їх можна було б використовувати як для створення нових моделей, так і для подальшого вдосконалення існуючих.

Подібний підхід, алгоритм якого представлений на Рисунку 3.4.1.1, називається data lake і вже використовується в деяких медичних базах даних, зокрема, IBM Merge Healthcare, French Health Data Hub, Health Data Research UK тощо.

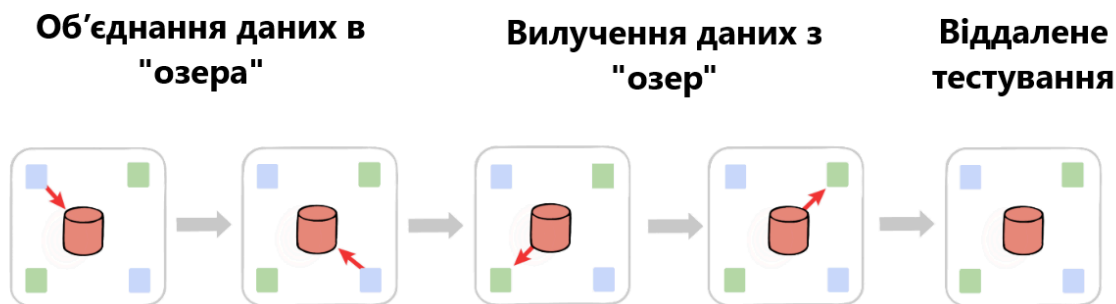


Рис. 3.4.1.1. Алгоритм створення та використання централізованих “озер даних”

Поки вчені змушені збирати медичну інформацію для кожного дослідження, практики постійно накопичують великі її обсяги. Тому теоретично використання “озер даних”, що об'єднують дані з різних медичних установ, може вирішити проблему великих даних для

застосування ІІІ в медицині. Однак використання цього підходу для збору та об'єднання даних обмежується трьома основними факторами.

По-перше, дослідникам важко координувати свої зусилля через відсутність єдиного протоколу збору даних або єдиного стандарту для електронних медичних записів.

По-друге, загальнодоступні медичні дані зазвичай асоціюються з конкретним завданням або захворюванням і часто публікуються з різними ступенями ліцензійних обмежень, що обмежує їх використання. Високоякісні набори даних часто мають значну комерційну цінність, тому окремі лікарі та медичні установи неохоче розкривають їх дослідникам. По-третє, використання будь-яких медичних даних обмежується суворими юридичними та етичними вимогами щодо захисту персональних даних та забезпечення їх конфіденційності. Більшість цих проблем можна вирішити за допомогою федеративного навчання (FL), яке базується на спільному навчанні алгоритмів без обміну самими даними. Існує дві основні моделі федеративного навчання, які представлені на Рисунках 3.4.1.2 і 3.4.1.3.

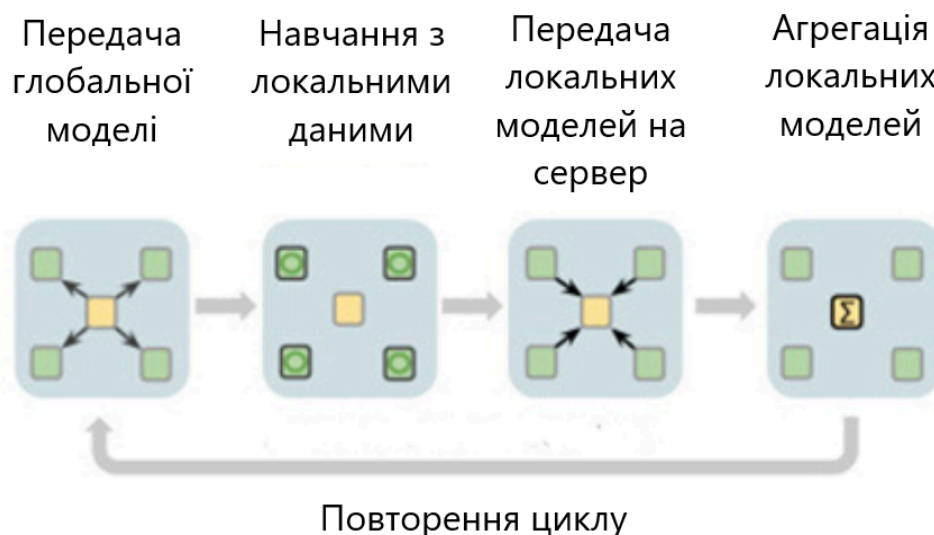


Рис.3.4.1.2. Модель федеративного навчання "Aggregation Server"

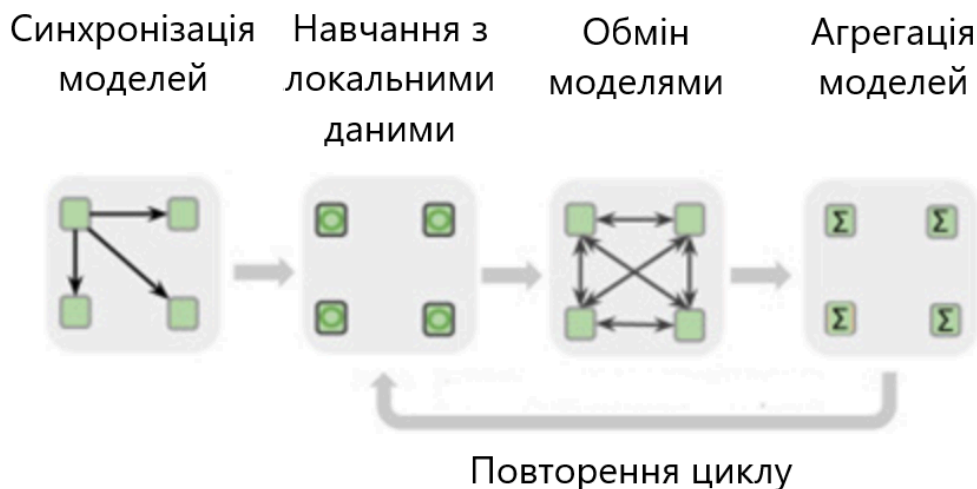


Рис.3.4.1.3. Модель федеративного навчання "Peer-to-peer"

У першій моделі вузли навчання пов'язані через єдиний сервер агрегації. Кожен вузол навчання надсилає свої частково навчені моделі на центральний сервер для агрегації і потім продовжує навчання на основі узгодженої моделі, наданої сервером. Друга модель, Peer-to-Peer, є архітектурою федеративного навчання, в якій всі навчальні вузли обмінюються своїми частково навченими моделями один з одним, одночасно виконуючи агрегацію даних.

Останні дослідження показують, що моделі федеративного навчання можуть досягати рівнів продуктивності, аналогічних моделям, навченим на централізованих наборах даних, і перевершувати моделі, які використовують дані з однієї установи. Таким чином, дозволяючи кільком сторонам навчатися разом без необхідності обміну або централізації наборів даних, федеративне навчання є перспективним підходом до створення потужних, точних, безпечних, надійних та неупереджених моделей.

Однак, хоча федеративне навчання вирішує проблеми управління даними, воно саме по собі не гарантує безпеку та конфіденційність медичних даних. Тому також необхідно розробити механізми конфіденційності, які можна застосувати до всього ланцюга глибокого навчання. Звичайні методи деідентифікації та захисту персональних даних, такі як анонімізація (видалення всіх відповідних метаданих, таких як ім'я пацієнта, стать, дата народження тощо) або псевдонімізація (заміна реальних записів синтетичними даними та створення таблиці відповідностей, яка зберігається окремо) часто є недостатніми для збереження конфіденційності, оскільки повторно ідентифіковані записи пацієнтів є вигідною ціллю для страхових компаній і компаній з видобутку та продажу даних.

Таким чином, якщо деідентифікація обмежена, а федеративне навчання не поєднується з додатковими методами захисту даних, такими як функціональне шифрування, квантування, гомоморфне шифрування, захищене багатостороннє обчислення тощо, воно саме по собі не гарантуватиме безпеку та конфіденційність, оскільки особисті дані пацієнтів можуть бути викрадені безпосередньо з вузлів або під час передачі даних у процесі комунікації [63].

Перспективною технологією захисту персональних даних у медичних записах вважається диференційна конфіденційність (Differential Privacy, DP), формальна структура для кількісної оцінки конфіденційності. Принцип цієї технології полягає в тому, що вона встановлює анонімізовані дані, вводячи шум у набори даних, що найчастіше досягається шляхом внесення довільних невеликих змін до індивідуальних даних, які не змінюють загальної статистики. Диференційну конфіденційність можна реалізувати як локально, вводячи шум до окремих даних перед централізацією в базі даних, так і глобально, додаючи шум до необроблених даних після їх збору від окремих осіб.

Основна ідея DP полягає в тому, що конфіденційність слід розглядати як ресурс, щось, що використовується при вилученні інформації з набору даних. З цієї точки зору, метою приватного видобутку даних є вилучення максимально корисної інформації при мінімальному споживанні конфіденційності. DP має перевагу бути кількісною. Диференційна конфіденційність розраховується шляхом введення в набір даних двох параметрів: втрати конфіденційності (δ) і параметра бюджету конфіденційності (ϵ). Ці параметри контролюють, скільки шуму або випадковості додається до необробленого набору даних. Це забезпечує числове представлення рівня конфіденційності, який можна очікувати: чим менші ϵ і δ , тим більше зберігається конфіденційність[63].

Отже, DP є математичним визначенням втрати конфіденційності даних окремих осіб, чиї особисті дані використовуються. Математично диференційну конфіденційність можна описати наступним чином. Нехай A - рандомізований алгоритм, який приймає входні дані D , конфіденційність яких має бути збережена, а P - функція ймовірності. Можна стверджувати, що алгоритм A забезпечує ϵ -диференційну конфіденційність, якщо для всіх наборів даних D і D' , які відрізняються одним елементом (тобто даними однієї особи) і всіх підмножин можливих результатів S алгоритму A , вираз є істинним:

$$P[A(D_1) \in S] \leq e^\epsilon P[A(D_2) \in S] + \delta \quad (3.4.1.1)$$

Якщо обидва параметри ϵ та δ є малими позитивними числами, результати алгоритму A майже не зміняться в розподілі, якщо змінити одну точку даних у базі даних. Іншими словами, додавання даних одного пацієнта до одного приватного дослідження не вплине на результати з високою ймовірністю. З практичної точки зору, алгоритм A можна вважати диференційно конфіденційним, якщо спостерігач, який бачить результат,

не може визначити, чи була інформація конкретної особи використана в обчисленні. Сьогодні DP широко використовується комерційними компаніями, соціальними мережами та цифровими сервісами, які зацікавлені в зборі та аналізі даних без розкриття особистої інформації людей. Для дослідників, які працюють з штучним інтелектом і машинним навчанням у медичній сфері, ця технологія надає доступ до великої кількості конфіденційних даних для досліджень без шкоди для пацієнтів [63].

Таким чином, створення великих наборів даних, що містять медичну інформацію, та збереження їх конфіденційності є наріжним каменем застосування ШІ у діагностиці та, зокрема, у розвитку автоматизованих систем діагностики захворювань дихальної системи. Однак ця проблема ще не вирішена повністю, насамперед через відсутність єдиних стандартів даних та розрив між захистом персональних даних і використанням цих даних для клінічної практики та досліджень.

3.4.2. Складний характер дихальних шумів та респіраторних захворювань

Штучний інтелект, а зокрема глибоке навчання, сьогодні використовується у багатьох галузях медицини. Зокрема, автоматизовані системи, засновані на нейронних мережах, застосовуються для аналізу результатів електрокардіограм, магнітно-резонансної томографії мозку тощо. Однак розробка автоматизованих систем для аналізу та класифікації результатів аускультатії має свої особливості, які пов'язані, насамперед, зі складністю та варіабельністю дихальних шумів у легенях.

Хоча більшість патологічних шумів у легенях мають досить чітко визначені акустичні характеристики, які представлені у Таблиці 1.1.1 у першому розділі цієї роботи, кожне захворювання може характеризуватися

комбінацією цих шумів, і один і той самий шум часто зустрічається при різних захворюваннях.

Особливо складно діагностувати дихальні шуми, які містять не один, а кілька патологічних легеневих шумів. Крім того, ансамблі дихальних шумів, характерні для певних захворювань, можуть змінюватися залежно від їх стадії. Наприклад, на початковій стадії бронхіту під час вдиху і видиху чути низькі хрипи, але з розвитком хвороби вони перетворюються на вологі велико- або дрібнопухирцеві хрипи.

Для точної діагностики захворювань дихальних шляхів важливо враховувати як патологічні шуми, так і фізіологічні дихальні шуми, до яких зазвичай належать везикулярне і бронхіальне дихання. Їх можна почути як у здорових, так і в хворих людей, але при захворюваннях легень їхній характер змінюється. Зокрема, при емфіземі або рідині в легенях везикулярне дихання послаблюється, а при астмі його стандартна структура змінюється (див. Рис. 1.1.1) і видих помітно подовжується.

Для діагностики захворювань легень слід враховувати фізіологічні області, в яких проводилася аускультация. Так, при пневмонії бронхіальне дихання, яке зазвичай чути лише над трахеєю або великими бронхами, починає чути над периферійними частинами легень.

Крім того, на результати можуть впливати інші фактори: вік, стать, спосіб життя, супутні захворювання, глибина дихання тощо. Навіть ліки, які приймають пацієнти, можуть сприяти розвитку або загостренню симптомів, а отже, змінювати дихальні шуми.

Більшість існуючих досліджень, присвячених розробці автоматизованих систем розпізнавання дихальних шумів, не досліджували вплив жодного з вищезазначених факторів.

Варто зазначити, що діагностика захворювань за записами аускультация легень є ще більш нетривіальним завданням для машинного навчання, оскільки вона не обмежується лише аналізом і класифікацією

аномальних дихальних шумів. У широкому сенсі, ідеальна автоматизована діагностична система повинна мати можливість виявляти:

- Фаза респіраторного циклу (вдих, видих), під час якої чутні додаткові шуми;
- Тип дихання (порівняння тривалості фаз вдиху і видиху, паузи між ними тощо);
- Зміни у стандартному диханні в різних зонах аускультації;
- Наявність або відсутність додаткових дихальних шумів;
- Тип і характер додаткових дихальних шумів;
- Змішування та комбінації дихальних шумів.

Очевидно, що додавання кожного з цих векторів, які є важливими для аналізу дихальних шумів та діагностики респіраторних захворювань, значно ускладнить будь-яку автоматизовану систему, що базується на алгоритмах машинного або глибокого навчання. Включення всіх цих векторів до умов зробить систему занадто складною і майже непридатною для практичного використання в клінічних умовах або на носимих пристроях через необхідність значних обчислювальних ресурсів.

Таким чином, враховуючи складність і багатовекторність діагностики респіраторних захворювань, доцільно розвивати та вдосконалювати автоматичні системи аналізу звуків іншим шляхом, а саме, використовувати не окремі фізіологічні маркери, а повні патерни певного захворювання. Іншими словами, вхідними навчальними даними для автоматизованих систем повинні бути не окремі ознаки певних дихальних шумів, а приклади дихання пацієнтів із конкретними захворюваннями. Лише алгоритми глибокого машинного навчання, які працюють з великими обсягами даних і не залежать від ручного вибору ознак, здатні виконувати це завдання, але навіть їхні можливості обмежені.

3.4.3. Методи підвищення продуктивності

Можливими способами подолання цих обмежень є ансамблеве навчання, перенесене навчання або глибоке вилучення ознак, підходи, які наразі вважаються найперспективнішими напрямками розвитку штучного інтелекту. Ансамблеве навчання (EL), загальний алгоритм якого представлено на Рисунку 3.4.3.1, — це техніка машинного навчання, яка використовує кілька навчальних алгоритмів для отримання найкращої попередньої продуктивності, яку можна було б отримати від кожного алгоритму окремо.

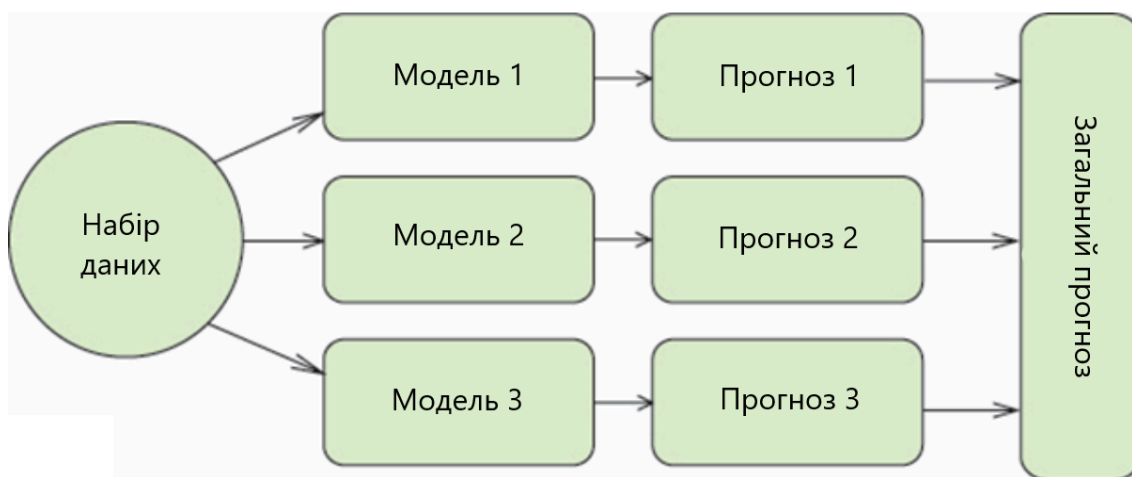


Рис. 3.4.3.1. Алгоритм ансамблевого навчання.

Нині ця техніка широко використовується в медичній сфері, наприклад, для виявлення нейрокогнітивних розладів на основі наборів даних МРТ або для класифікації результатів цитологічних тестів. Серед відповідних досліджень, розглянутих у цій роботі, деякі використовують ансамблі моделей МН та ГН для класифікації звуків дихання або діагностики захворювань легень [16, 36, 42, 53].

Примітно, що в усіх дослідженнях, де тестувалися індивідуальні моделі та їхні ансамблі, ансамблі досягали кращих результатів. Наприклад, Мохаммед Фріван та спів. (2022) змогли підвищити показник F1 майже на 3% порівняно з найкращим окремим класифікатором (з 95,76% до 98,57%), поєднавши класифікатори CNN та двонаправлений LSTM. У дослідженні Алі Мохаммада Алкоуда та спів. (2022) класифікатор ансамблю CNN + LSTM перевершив індивідуальні моделі за деякими показниками як на покращених, так і на непокращених наборах даних.

Перенесене навчання (TL) є ще більш перспективним підходом для покращення продуктивності автоматичних систем класифікації звуків дихання та респіраторних захворювань. Тоді як звичайні алгоритми машинного та глибокого навчання розроблені для ізольованої роботи та вирішення конкретних завдань, що передбачає, що кожного разу, коли змінюється розподіл у просторі функцій, модель потрібно навчати заново, перенесене навчання пропонує використовувати знання, отримані під час вирішення одного завдання, для вирішення іншого, але пов'язаного завдання.

Фундаментальне припущення в алгоритмах машинного навчання полягає в тому, що навчальні та тестові дані повинні існувати в одному просторі ознак і мати однаковий розподіл. Однак на практиці це припущення іноді не виконується, наприклад, розпізнавання цифр може допомогти в розпізнаванні літер. Іншими словами, у деяких випадках повторне використання або перенесення інформації з раніше виконаних завдань на нове навчання може покращити продуктивність класифікаторів, навіть якщо вони знаходяться в іншому просторі ознак або слідує іншому розподілу даних. Такий перенесення знань, якщо він виконаний успішно, також спрощує вимоги до розміру бази даних, анотації даних тощо.

Перенесене навчання спрямоване на покращення навчання цільової предиктивної функції f_t у цільовій області D_t і навчального завдання T_t , використовуючи знання з іншої джерельної області D_s і навчального завдання T_s . [65] Загальний алгоритм перенесеного навчання представлений на Рисунку 3.4.3.2.



Рис. 3.4.3.2. Порівняння традиційних (ізолюваних) та переносних алгоритмів машинного навчання.

Спочатку модель проходить навчання на великому наборі даних, а потім цей процес повторюється на меншому наборі даних для точного налаштування параметрів.

Одним із видів переносного навчання, що може бути перспективною технологією для підвищення ефективності класифікації дихальних звуків або респіраторних захворювань, є глибоке вилучення ознак. У цьому методі рівні попередньо навчених мереж комбінуються або переносяться в інші мережі для точного налаштування або вилучення ознак [66].

У медичній сфері вже є успішні приклади застосування переносного навчання, зокрема для виявлення підтипів раку [67].

Крім того, цей метод може бути перспективним для навчання автоматизованих систем розпізнавання дихальних звуків і діагностики респіраторних захворювань, де передача знань потенційно може здійснюватися від попередньо навчених і протестованих систем розпізнавання мови або аналізу серцевого ритму тощо.

Переносне навчання, яке забезпечує задовільне налаштування, є більш продуктивним, ніж навчання моделі з нуля. Це твердження, зокрема, підтверджується дослідженням Джейн Салданья та спів. (2022) [45], у якому найкращі результати класифікації були продемонстровані попередньо навченою моделлю Resnet-50.

У дослідженні Фатіха Деміра та спів. (2019) було розглянуто два різні підходи до класифікації дихальних звуків із використанням технік глибокого вилучення ознак і переносного навчання. У першому підході використовували попередньо навчену модель глибокої згорткової нейронної мережі (CNN) для вилучення ознак із даних дихальних звуків. Ці вилучені ознаки потім подавалися в класифікатор Support Vector Machine (SVM) для виконання завдання класифікації звуків легенів. У другому підході попередньо навчена модель глибокої CNN пройшла налаштування (переносне навчання) з використанням спектрограм, пов'язаних із даними звуків легенів. Переносне навчання передбачає адаптацію попередньо навченої моделі до нового завдання або набору даних шляхом точного налаштування її параметрів. У цьому випадку попередньо навчена модель CNN була налаштована з використанням спектрограм для кращого виконання завдання класифікації звуків легенів.

Обидва підходи оцінювалися на точність за допомогою перехресної валідації з десяти розбиттів. Перший підхід досяг точності 65,5%, тоді як другий підхід досяг точності 63,09%. Ці результати свідчать про те, що обидва підходи досягли середніх показників точності в класифікації дихальних звуків. Однак існує потенціал для подальшого покращення

класифікації дихальних звуків із використанням технік глибокого вилучення ознак і переносного навчання.

Лікви Ши та спів. запропонували алгоритм розпізнавання звуків легенів на основі переносного навчання (TL). Вони підкреслили важливість переносного навчання для підвищення здатності моделі до узагальнення, особливо в тих випадках, коли є нестача аудіозаписів дихальних звуків. Автори використали попередньо навчену мережу VGGish, відому своєю здатністю обробляти аудіодані, та рекурентну нейронну мережу (RNN), яка відзначається успіхами у роботі з часовими рядами даних. Об'єднуючи просторові та часові домени, вони запропонували модель для розпізнавання респіраторних захворювань. Ця модель об'єднує згорткову нейронну мережу VGGish із двонаправленою рекурентною нейронною мережею BiGRU. Структура цієї об'єднаної моделі показана на рисунку 3.4.3.3, що демонструє злиття архітектур CNN і RNN для ефективного розпізнавання звуків легенів. Цей підхід має на меті використання сильних сторін обох типів нейронних мереж для покращення точності та надійності класифікації респіраторних захворювань.

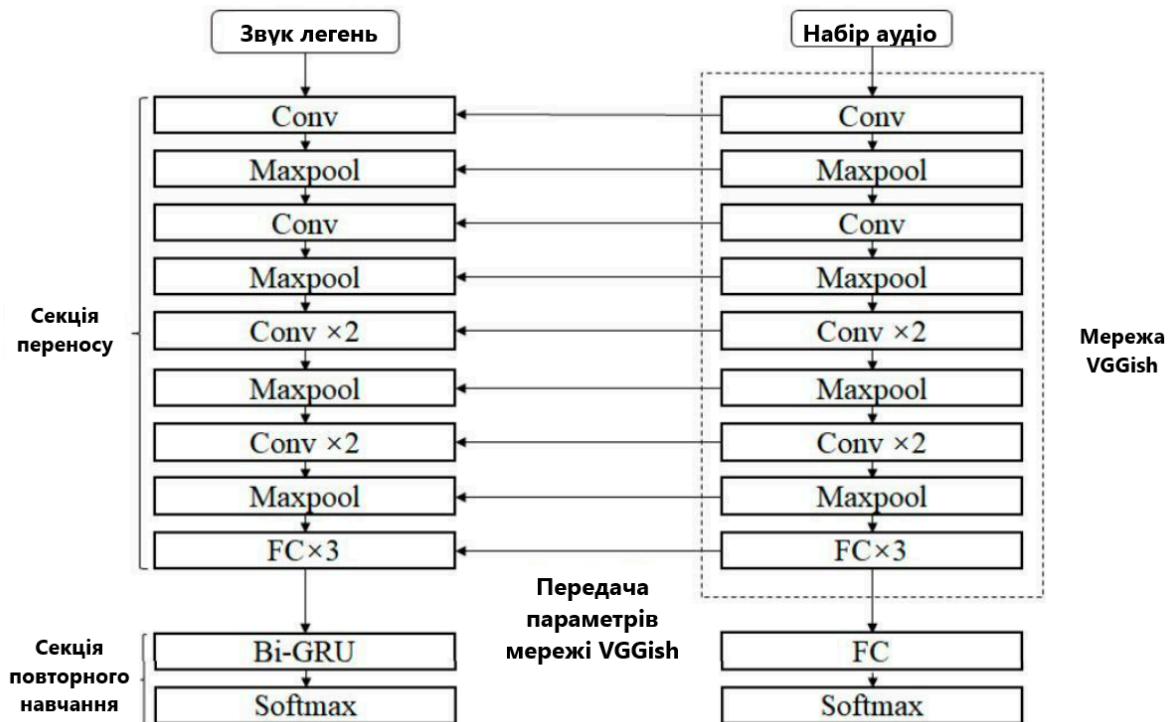


Рис. 3.4.3.3. Модель перенесення навчання, запропонована Лікві Ши та спів. (2019), заснована на VGGish і BiGRU.

У запропонованому процесі перенесення навчання мережа VGGish спочатку проходить попереднє навчання на наборі даних. Потім параметри навчання, отримані мережею VGGish, передаються на відповідні шари цільової мережі. Після цього часові ознаки звуків легенів вилучаються шляхом повторного навчання двонаправленої рекурентної нейронної мережі (BiGRU). Під час цієї фази перенавчання параметри в шарах VGGish залишаються фіксованими, щоб зберегти вивчені аудіопредставлення. Тим часом параметри мережі BiGRU піддаються точному налаштуванню, щоб адаптуватися до специфічних характеристик дихальних звуків у цільовій області.

Цей алгоритм перенесення навчання дозволяє моделі ефективно використовувати попередньо навчену мережу VGGish для вилучення аудіо ознак, при цьому персоналізуючи можливості обробки часових даних

шляхом повторного навчання мережі BiGRU. Процес детально проілюстровано на Рисунку 3.4.3.4, який демонструє послідовні кроки, що беруть участь у перенесенні знань з вихідної на цільову область.



Рис. 3.4.3.4. Алгоритм перенесення навчання, запропонований Лікві Ши та спів. (2019)

У проведеному експерименті Лікві Ши та спів. (2019) порівняли продуктивність запропонованої моделі на основі перенесення навчання з кількома традиційними моделями машинного навчання (МН) та згортковою нейронною мережею (CNN) для задачі трьохфакторної класифікації. Серед моделей МН були класифікатор найближчих сусідів (KNN), метод опорних векторів (SVM) та випадковий ліс. Ці моделі часто використовуються для задач класифікації в машинному навчанні. Результати експерименту показали, що модель на основі перенесення навчання перевершила традиційні моделі МН і CNN за точністю та оцінкою в задачі трьохфакторної класифікації. Це свідчить про те, що підхід до перенесення навчання, використаний Ши та ін., забезпечив кращу продуктивність порівняно з традиційними методами МН і CNN у

конкретній класифікаційній задачі, яку було оцінено в їхньому експерименті.

Щоб перевірити вплив перенесення навчання на результати класифікації, дослідники провели порівняння між моделлю, що включає перенесення навчання, та моделлю без нього. В обох випадках збір даних, попередня обробка, метод вилучення ознак і розміри вхідних даних були однаковими. Для моделі без перенесення навчання використовували архітектуру VGGish-BiGRU. У цьому випадку параметри VGGish ініціалізувалися випадковим чином і не передавалися з результатів, отриманих на AudioSet. Результати цього порівняння, представлені в термінах загальної точності, наведені в Таблиці 3.4.3.1.

Таблиця 3.4.3.1

Результати порівняння традиційного та перенесеного навчання

| | Пневмонія | Астма | Стандарт | Середнє значення |
|--------------------------------|------------------|--------------|-----------------|-------------------------|
| Звичайне навчання | 84,62% | 78,4% | 90,27% | 84,64% |
| Навчання з перенесенням | 86,75% | 83,33% | 91,94% | 87,41% |

З результатів, наведених у Таблиці 3.4.3.1, можна зробити висновок, що перенесення навчання ефективно покращує експериментальні результати. Точність розпізнавання патернів покращилася у всіх категоріях, зокрема в категорії «Астма». Дослідження Лікві Ши та спів. (2019) доводить, що класифікація звуків легенів і респіраторних захворювань може бути ефективно покращена шляхом використання перенесення навчання, перенісши параметри з моделі вихідної області. Однак автори

зазначають, що хоча набір даних AudioSet має кращу узагальнюваність, він не містить звуків легенів, що знижує загальну точність моделі. Таким чином, наявність більш відповідного джерела даних, такого як велика база даних звуків дихання, могла б оптимізувати модель на основі перенесення навчання.

3.4.4. Інтерпретація результатів машинного навчання

Основна увага в розробці автоматизованих систем для розпізнавання та класифікації дихальних шумів і респіраторних захворювань сьогодні зосереджена на розробці методів і моделей машинного навчання та досягненні кращої продуктивності. Для більшості дослідників кінцевою метою є встановлення моделі машинного навчання з прийнятною продуктивністю. Дійсно, для деяких систем, які мають низький ризик і не передбачають значного впливу або серйозних наслідків у разі отримання неправильних результатів, цього цілком достатньо, оскільки в таких випадках потрібно лише знати, що було передбачено.

Однак у деяких галузях, які мають глибокий вплив на життя окремих людей і суспільства, таких як охорона здоров'я, важливо знати, як система дійшла свого рішення, тобто мати можливість інтерпретувати результати, оскільки точність класифікації сама по собі є неповним описом більшості реальних проблем.

На сьогодні більшість точних систем підтримки прийняття рішень на основі МН, а особливо на основі GN, залишаються складними чорними ящиками, внутрішня логіка яких прихована від користувача. Наприклад, деякі глибокі резидуальні мережі (ResNets) мають глибину більше ніж 200 рівнів і перевищують людський рівень продуктивності в завданнях розпізнавання об'єктів. Крім того, висока складність притаманна іншим моделям МН, включаючи найефективніші, такі як ансамблеві класифікатори або згорткові нейронні мережі [68].

Таким чином, сучасні моделі машинного навчання не мають належної інтерпретації того, як здійснюється прогнозування. Це потенційно може мати серйозні наслідки, оскільки той факт, що МН вже підтримує прийняття важливих рішень, не означає, що він не схильний до помилок. Наприклад, коли медичні дані (наприклад, звуки легенів), які потенційно містять викривлені розподіли (більше здорових, ніж хворих, більше одного захворювання, ніж іншого тощо), вводяться в модель, вона вчиться на цих шаблонах і повертає прогнози, які враховують цей ухил, що може призвести до дискримінації рідкісних захворювань серед інших речей.

У системах класифікації дихальних звуків відсутність належної інтерпретації збільшує ймовірність включення зайвої та шкідливої інформації (звуки серця, зовнішні шуми, мова тощо) серед інформативних маркерів, що призводить до неправильної класифікації.

Крім того, алгоритми МН не можуть самостійно адаптуватися до змін, наприклад, поширення атипових захворювань, які мають викривлену клінічну картину і, зокрема, змінені моделі дихальних звуків. Вони кодують минуле, не маючи змоги еволюціонувати чи вигадати майбутнє, що означає, що алгоритмічні перевірки є вирішальними для виявлення дискримінації та упередженості. Дійсно, прозорість є важливим аспектом надійних систем штучного інтелекту (ШІ). Здатність розуміти та пояснювати рішення, які приймають системи ШІ, є важливою для забезпечення відповідальності, справедливості та етичного використання технологій ШІ. Рекомендації, надані Групою високого рівня експертів з питань штучного інтелекту (AI HLEG), підкреслюють прозорість як одну з ключових вимог для надійного ШІ.

Прозорість у ШІ охоплює кілька компонентів, зокрема відстежуваність, пояснення та комунікацію. Надаючи пріоритет прозорості у розробці та впровадженні ШІ, зацікавлені сторони можуть гарантувати, що системи ШІ є відповідальними, справедливими та відповідають

етичним принципам. Такий підхід сприяє зміцненню довіри та впевненості у технологіях ШІ та допомагає зменшити ризики, пов'язані з непрозорими процесами прийняття рішень. [69]

Відповіддю на цю проблему став пояснювальний штучний інтелект (EAI), який враховує інтерпретацію результатів машинного навчання. Основною метою EAI є створення набору інтерпретованих моделей і методів, тобто зрозуміліших моделей, які зберігатимуть високий рівень прогностичної продуктивності [70]. Однак, оскільки ця сфера ще перебуває на ранній стадії розвитку, у поточній літературі немає прикладів її практичного застосування, а теоретичні основи часто плутаються в термінології та визначеннях.

Повертаючись до поточних досліджень, можна стверджувати, що для клінічного використання автоматизованих систем класифікації дихальних звуків і діагностики респіраторних захворювань дуже важливо мати чітке уявлення про те, які фізіологічні маркери використовуються, щоб уникнути упереджених рішень і медичних помилок.

Це потребує постмодельної інтерпретації з комплексною оцінкою розроблених моделей. На відміну від внутрішньої інтерпретації, яка відповідає на питання "як працює модель" і досягається шляхом накладення певних обмежень на складність моделі, постмодельна інтерпретація використовує методи, які аналізують модель після навчання, тобто відповідає на питання "що модель може нам розповісти".

Метою використання постмодельної інтерпретації є зрозуміти прогнози моделей МН, що досягається за допомогою пояснень, згенерованих певним алгоритмом. Загальний принцип постфактумної інтерпретації представлений на наступній схемі.

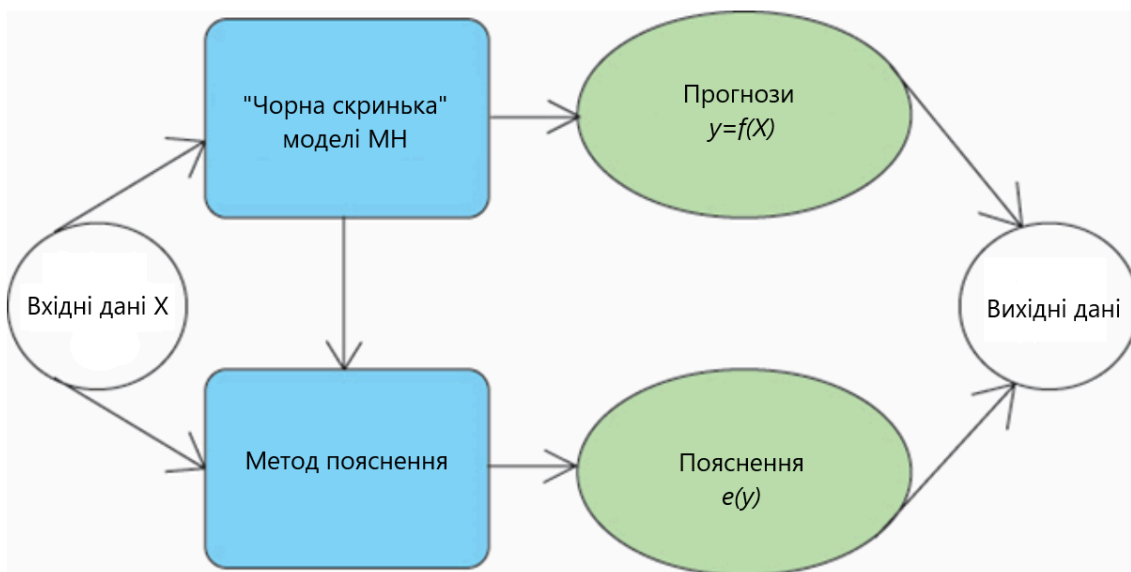


Рис. 3.4.5.1. Алгоритм постмодельної інтерпретації МН.

Щодо пояснення роботи автоматизованих систем класифікації дихальних звуків і діагностики респіраторних захворювань для методів МН та ГН, необхідно визначити значущість акустичних характеристик, щоб знайти їх значення, які відповідають асоційованим респіраторним станам у реальному світі. Крім того, корисно простежити значущість різних сегментів спектрограми та визначити, які з них підкреслюють наявність захворювання. Це потенційно дозволить експертам визначати нові зрозумілі симптоми певних захворювань, а системам – виявляти кілька захворювань у тому ж зразку.

Таким чином, пояснення важливі для забезпечення цілісності алгоритмів і виявлення упереджень у навчальних даних. Можливість інтерпретації дозволяє виявляти некоректну поведінку моделі та допомагає зрозуміти причину помилки, що підвищує надійність системи в цілому.

Ще одним критично важливим аспектом, який підсилюється завдяки інтерпретованості, є безпека. Здатність інтерпретувати, тестувати,

перевіряти та налагоджувати моделі МН є способом підвищення їхньої безпеки, особливо для галузей, де помилки можуть мати серйозні наслідки.

Зрештою, інтерпретованість слугує засобом задоволення людської цікавості, надаючи пояснення несподіваним подіям і підвищуючи довіру та прийняття машинного навчання людиною. Моделі машинного навчання, рішення яких можна легко інтерпретувати, з більшою ймовірністю будуть прийняті людьми.

3.4.5. Впровадження системи автоматичної класифікації дихальних звуків та респіраторних захворювань у клінічну практику

Архітектура автоматизованої системи класифікації респіраторних захворювань, запропонована в цьому дослідженні, та деякі інші моделі, розроблені іншими авторами, мають перспективну ефективність для автоматичного респіраторного скринінгу, дистанційного догляду та телемедицини, моніторингу хронічних респіраторних захворювань або підтримки діагностичних рішень. Примітно, що точність таких систем є вищою або принаймні порівнянною з точністю діагностики на основі результатів класичної аускультації. Наприклад, Юнжу Кім та спів. (2021) [71] оцінили, що результати класифікації дихальних звуків різними групами медичних працівників продемонстрували різний рівень точності. Показник становив 60,3% для студентів-медиків, 53,4% для інтернів, 68,8% для резидентів і 80,1% для асистентів.

Однак для практичного впровадження автоматизованих систем класифікації дихальних звуків або респіраторних захворювань необхідна додаткова валідація продуктивності моделей на клінічно перевірених даних. Цей крок забезпечує перевірку фактичної продуктивності моделей і допомагає запобігти перенавчанню на експериментальних даних. Не можна виключати, що дані, зібрані в лабораторії, та моделі глибокого

навчання, які зазвичай використовуються і демонструють перспективні результати, можуть бути погано налаштовані, що призводить до надто впевнених прогнозів, які можуть не відображати справжню діагностичну точність. Фахівці з обробки даних відмінно володіють моделюванням, тоді як лікарі мають практичні знання. Тому важливо поглиблювати співпрацю між цими двома сферами для розвитку високоефективних систем респіраторного скринінгу.

Масштабні дослідження автоматизованих систем класифікації дихальних звуків і діагностики допоможуть як підтвердити або уточнити існуючі результати, так і збільшити їхнє прийняття медичною спільнотою, а також вирішити ризики, пов'язані з клінічним впровадженням таких систем.

Глибоке навчання, особливо в контексті незнайомих даних, таких як нові типи звуків, нові респіраторні захворювання або шумові аудіосигнали, може демонструвати непередбачувану поведінку. Це непередбачуваність може мати значні наслідки в клінічних умовах. Потенційним вирішенням цієї проблеми може бути кількісна оцінка невизначеності прогнозу акустичних моделей, що дозволить попереджати про незнайомі вхідні сигнали та невизначені прогнози для пацієнтів. У будь-якому випадку ризики, які можуть призвести до помилкової діагностики, слід добре контролювати шляхом відстеження ситуацій, коли штучний інтелект не справляється і потребує втручання людини.

4. Висновки

В рамках даного дослідження була запропонована автоматична модель діагностики респіраторних захворювань як однієї з найважливіших проблем охорони здоров'я. Звуки дихання можуть бути оброблені та проаналізовані в цифровому форматі для полегшення діагностики захворювань, покращення управління клінічними рішеннями або надання послуг дистанційної охорони здоров'я.

Запропонований метод включає попередню обробку даних, яка базується на фільтрації неінформативних частот та перетворенні аудіозаписів у спектральну форму. Незбалансований набір конфіденційних медичних даних доповнюється тренувальним набором за допомогою синтезу записів із використанням варіаційного автокодера згортки, який тренується на цільових класах. Для виконання завдання класифікації була використана архітектура згорткової нейронної мережі, що характеризується швидким навчанням та прогнозуванням, а також здатністю швидко адаптувати архітектуру до поставленого завдання.

Класифікація проводилася за п'ятьма категоріями, базуючись не на окремих патологічних звуках дихання, а на патернах респіраторних захворювань, включаючи пневмонію, ХОЗЛ, інфекції верхніх дихальних шляхів, бронхоектатичну хворобу чи бронхіоліт, а також нормальні звуки дихання у здорових людей.

Експеримент проводився у два етапи, що відрізнялися розподілом даних. Перший етап, що включав стратифікований поділ на навчальні та тестові вибірки без групування за суб'єктами, показав високу ефективність запропонованої моделі. Середні показники продуктивності були такими: чутливість — 98,44%, специфічність — 99,61%, правильність — 98,44%, каппа Коена — 98,05%, і F1 — 98,45%. За останнім параметром модель, запропонована в цьому дослідженні, перевершила всі релевантні

дослідження, що також класифікували респіраторні захворювання з використанням згорткових нейронних мереж.

Аналіз якості додаткових даних, проведений з використанням таких показників, як Фрешева аудіовідстань, кореляція та аналіз основних компонент, у поєднанні з загальними результатами першого етапу експерименту свідчить про високу специфічність класу, що вказує на успішне дотримання патернів патології на етапі доповнення даних із використанням варіаційного автокодера згортки. Цей підхід до доповнення даних добре проявив себе при десятиразовій крос-валідації та дозволив узагальнити характеристики для менш значних класів. Це свідчить про доцільність застосування методу доповнення даних із використанням варіаційних автокодерів згортки в поєднанні з згортковими нейронними мережами або іншими архітектурами класифікаторів.

Другий етап, в якому навчальні та тестові серії були розділені для забезпечення відсутності збігу між тестовим та навчальним наборами, показав суттєво відмінні результати від попередніх. Ця відмінність свідчить про витік даних із навчального набору в тестовий на першому етапі експерименту, який не було скореговано крос-валідацією. Основною причиною цього витоку даних було те, що записи, отримані від одного й того самого пацієнта за короткий період, формально вважалися різними, але мали дуже схожі акустичні характеристики через індивідуальні особливості пацієнта, технічні та амплітудно-частотні характеристики записуючого пристрою та шумові умови навколишнього середовища, в яких проводився запис.

Відносно високі $F1$ і чутливість до класу ХОЗЛ, отримані на другому етапі експерименту, свідчать про достатню кількість пацієнтів із цим захворюванням та кількість зразків цього класу. Гірші результати в інших категоріях свідчать про те, що збільшення менш значущих класів з відносно вузьким розподілом було недостатнім для узагальнення їх

характеристик і покращення класифікації, в першу чергу через невелику кількість пацієнтів із цими захворюваннями.

Аналіз інших досліджень щодо автоматизованої респіраторної діагностики показав, що автори зазвичай не враховують проблему витоку даних або не вирішують її повністю, що призводить до нерепрезентативних результатів.

Таким чином, основний результат дослідження включає розробку нового підходу до поділу набору даних для тренування та тестування моделі класифікації респіраторних захворювань, включаючи стратифікований поділ на основі належності записів (спектрограм) певним пацієнтам, що запобігає витоку корисних характеристик між наборами.

Можна припустити, що при використанні запропонованого розподілу даних ефективність інших моделей автоматизованої діагностики респіраторних захворювань, які демонструють занадто оптимістичні результати через витік даних, суттєво погіршиться, і тому пошук методів покращення продуктивності таких систем залишається актуальним питанням.

Збільшення баз даних конфіденційних медичних даних, що містять записи аускультатії дихальних звуків, значно прискорить прогрес у покращенні алгоритмів. Одним із найбільш перспективних методів для досягнення цієї мети є федеративне навчання з використанням диференційованої конфіденційності для захисту персональних даних.

Зважаючи на складний характер дихальних звуків, які формують патерни певних респіраторних захворювань, слід також звернути увагу на технічні методи покращення продуктивності моделей штучного інтелекту, що використовуються для їх діагностики. Зокрема, ансамблеве навчання, глибока екстракція ознак і трансферне навчання, які використовують моделі, попередньо навчені на інших більших базах даних інших звуків, можливо, є перспективними підходами в цій галузі.

Щоб зменшити помилки, які можуть мати важливі наслідки для громадського здоров'я, і запобігти упередженості моделей щодо рідкісних або атипових захворювань, необхідно трансформувати використовувані в них моделі машинного навчання з «чорних скриньок» у прозорі алгоритми, тобто забезпечити належну інтерпретацію класифікації респіраторних захворювань. Це також допоможе підвищити довіру медичних працівників до таких систем і сприяти їхньому швидшому впровадженню в клінічну практику з належним тестуванням продуктивності моделей на клінічно перевірених даних і впровадженням ефективної системи управління ризиками.

Список використаної літератури

1. Pasterkamp H., Kraman S.S., Wodicka G.R. (1997). Respiratory sounds. Advances beyond the stethoscope. *Am J Respir Crit Care Med*. Sep;156 (3 Pt 1):974-87. doi: 10.1164/ajrccm.156.3.9701115.
2. Qamar S, Tekin A, Taweeseedt PT, Varon J, Kashyap R, Surani S. (2021) Stethoscope - An essential diagnostic tool or a relic of the past? *Hosp Pract*. Oct;49(4):240-244. doi: 10.1080/21548331.2021.1949170.
3. Nowak L.J., Nowak K.M. (2018). Sound differences between electronic and acoustic stethoscopes. *BioMedical Engineering OnLine*, vol. 17, no. 1, p. 104, 2018.
4. Gross V., Dittmar A., Penzel T., Schüttler F., von Wichert P. (2000) The relationship between normal lung sounds, age, and gender. *Am J Respir Crit Care Med*. Sep;162(3 Pt 1):905-9. doi: 10.1164/ajrccm.162.3.9905104.
5. Moriki D., Koumpagioti D., Kalogiannis M., Sardeli O., Galani A., Priftis K.N., Douros K. (2023) Physicians' ability to recognize adventitious lung sounds. *Pediatr Pulmonol*. Mar; 58(3):866-870. doi: 10.1002/ppul.26266.
6. Khan S. I., Jawarkar N. P., Ahmed V. (2012) Cell phone based remote early detection of respiratory disorders for rural children using modified stethoscope. *International Conference on Communication Systems and Network Technologies*, p. 936-940.
7. Kraman S.S., Wodicka G.R., Oh Y., Pasterkamp H. (1995) Measurement of respiratory acoustic signals. Effect of microphone air cavity width, shape, and venting. *Chest*. 108:1004–8.
8. Xia T., Han J., Mascolo C. (2022) Exploring machine learning for audio-based respiratory condition screening: A concise review of databases, methods, and open issues. *Exp Biol Med (Maywood)*. Nov;247(22):2053-2061. doi: 10.1177/15353702221115428.

9. Saldanha J., Chakraborty S., Patil S., Kotecha K., Kumar S., Nayyar A. (2022) Data augmentation using Variational Autoencoders for improvement of respiratory disease classification. *PLoS One*. Aug 12;17(8):e0266467. doi: 10.1371/journal.pone.0266467.
10. Sfayyih A.H., Sabry A.H., Jameel S.M., Sulaiman N., Raafat S.M., Humaidi A.J., Kubaiaisi Y.M.A. (2023) Acoustic-Based Deep Learning Architectures for Lung Disease Diagnosis: A Comprehensive Overview. *Diagnostics*. Basel. May 16;13(10):1748. doi: 10.3390/diagnostics13101748.
11. Kim, Y., Hyon, Y., Lee, S. et al. (2022). The coming era of a new auscultation system for analyzing respiratory sounds. *BMC Pulm Med* 22, 119. <https://doi.org/10.1186/s12890-022-01896-1>.
12. Sovijarvi A., Dalmaso F., Vanderschoot J., Malmberg L., Righini G., Stoneman S. (2000). Definition of terms for applications of respiratory sounds. *European Respiratory Review*, 10(77), 597-610.
13. Emmanouilidou D., McCollum E.D., Park D.E., Elhilali M. (2018). Computerized Lung Sound Screening for Pediatric Auscultation in Noisy Field Environments. *IEEE Trans Biomed Eng*. Jul;65(7):1564-1574. doi: 10.1109/TBME.2017.2717280.
14. Singh D., Singh B. K. and Behera A. K. (2020) Comparative analysis of Lung sound denoising technique. In: *First International Conference on Power, Control and Computing Technologies (ICPC2T)*. Apr 2020; pp. 406–410, doi: 10.1109/ICPC2T48082.2020.9071438
15. Shi L., Du K., Zhang C., Ma and H., Yan W. (2019) Lung Sound Recognition Algorithm Based on VGGish-BiGRU. *IEEE Access*, vol. 7, pp. 139438-139449, 2019, doi: 10.1109/ACCESS.2019.2943492.
16. Fraiwan M., Fraiwan L., Alkhodari M., Hassanin O. (2022) Recognition of pulmonary diseases from lung sounds using convolutional neural networks and long short-term memory. *J Ambient Intell Humaniz Comput*. 13(10):4759-4771. doi: 10.1007/s12652-021-03184-y.

17. Emmanouilidou D., McCollum E.D., Park D.E., Elhilali M. (2015). Adaptive Noise Suppression of Pediatric Lung Auscultations With Real Applications to Noisy Clinical Settings in Developing Countries. *IEEE Trans Biomed Eng.* Sep;62(9):2279-88. doi: 10.1109/TBME.2015.2422698.
18. Levine O.S., O'Brien K.L., Deloria-Knoll M., Murdoch D.R., Feikin D.R., DeLuca A.N., Driscoll A.J., Baggett H.C., Brooks W.A., Howie S.R., Kotloff K.L., Madhi S.A., Maloney S.A., Sow S., Thea D.M., Scott J.A. (2012). The pneumonia etiology research for child health project: a 21st century childhood pneumonia etiology study. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, vol. 54 Suppl 2, pp. S93–101.
19. Nelson G., Rajamani R., Erdman A. (2014) Noise control challenges for auscultation on medical evacuation helicopters. *Applied Acoustics*. Volume 80, 2014, Pages 68-78. <https://doi.org/10.1016/j.apacoust.2014.01.006>.
20. Nersisson R., Noel M.M. (2016). Heart sound and lung sound separation algorithms: a review. *J Med Eng Technol.* Jan;41(1):13-21. doi: 10.1080/03091902.2016.1209589.
21. Hossain I., Zahra M. (2003). An overview of heart-noise reduction of lung sound using wavelet transform based filter. *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439)* 1: 458-461 Vol.1.
22. Semmad A., Bahoura M. (2021). Long Short Term Memory Based Recurrent Neural Network for Wheezing Detection in Pulmonary Sounds. *IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)*, pp. 412-415, doi: 10.1109/MWSCAS47672.2021.9531784.
23. Iyer V.K., Ramamoorthy P.A., Fan H. et al. (1986). Reduction of heart sounds from lung sounds by adaptive filterng. *IEEE transactions on biomedical engineering.* 33:1141-1148.

24. Kompis M., Russi E. (1992). Adaptive heart-noise reduction of lung sounds recorded by a single microphone. 14th Annual International Conference of the IEEE. IEEE. 1992;2:691-692
25. Hadjileontiadis L.J., Panas S.M. (1997) Adaptive reduction of heart sounds from lung sounds using fourth-order statistics. IEEE Transactions on Biomedical Engineering. 44:642-648.
26. Hadjileontiadis L.J., Panas S.M. (1998). A wavelet-based reduction of heart sound noise from lung sounds. Int J Med Inform. Oct-Dec;52(1-3):183-90. doi: 10.1016/s1386-5056(98)00137-3.
27. Al-Naggar N.Q., Al-Udyni M.H. (2018). Performance of Adaptive Noise Cancellation with Normalized Last-Mean-Square Based on the Signal-to-Noise Ratio of Lung and Heart Sound Separation. J Healthc Eng. Jul 12;2018:9732762. doi: 10.1155/2018/9732762.
28. Li T., Tang H., Qiu T., Park Y. (2013). Heart sound cancellation from lung sound record using cyclostationarity. Med Eng Phys. Dec;35(12):1831-6. doi: 10.1016/j.medengphy.2013.05.004.
29. Flores-Tapia D., Moussavi Z.M., Thomas G. (2007). Heart sound cancellation based on multiscale products and linear prediction. IEEE Trans Biomed Eng. Feb;54(2):234-43. doi: 10.1109/TBME.2006.886935.
30. Meng F., Shi Y., Wang N., Cai M., Luo Z. (2020). Detection of Respiratory Sounds Based on Wavelet Coefficients and Machine Learning. IEEE Access, vol. 8, pp. 155710-155720, 2020, doi: 10.1109/ACCESS.2020.3016748.
31. Wang W., Wang S., Qin D., Fang Y., Zheng Y. (2023). Heart-lung sound separation by nonnegative matrix factorization and deep learning. Biomedical Signal Processing and Control, Volume 79, Part 2, 104180. doi: 10.1016/j.bspc.2022.104180.
32. Chen H., Yuan X., Pei Z., Li M., Li J. (2019). Triple-Classification of Respiratory Sounds Using Optimized S-Transform and Deep Residual

Networks. IEEE Access, vol. 7, pp. 32845-32852. doi: 10.1109/ACCESS.2019.2903859.

33. Nuckowska M.K., Gruszecki M., Kot J. et al. (2019). Impact of slow breathing on the blood pressure and subarachnoid space width oscillations in humans. *Sci Rep* 9, 6232. <https://doi.org/10.1038/s41598-019-42552-9>.

34. Petmezas G., Cheimariotis G.-A., Stefanopoulos L., Rocha B., Paiva R.P., Katsaggelos A.K., Maglaveras N. (2022). Automated Lung Sound Classification Using a Hybrid CNN-LSTM Network and Focal Loss Function. *Sensors* 22, 1232. <https://doi.org/10.3390/s22031232>.

35. Niu J., Cai M., Shi Y., Ren S., Xu W., Gao W., Luo Z., Reinhardt J.M. (2019). A Novel Method for Automatic Identification of Breathing State. *Sci Rep*. Jan 14;9(1):103. doi: 10.1038/s41598-018-36454-5.

36. García-Ordas M.T., Benitez-Andrades J.A., Garcia-Rodriguez I., Benavides C., Alaiz-Moreton H. (2020). Detecting Respiratory Pathologies Using Convolutional Neural Networks and Variational Autoencoders for Unbalancing Data. *Sensors*. Basel. Feb 22;20(4):1214. doi: 10.3390/s20041214.

37. Abera Tessema B., Nemomssa H.D., Lamesgin Simegn G. (2022). Acquisition and Classification of Lung Sounds for Improving the Efficacy of Auscultation Diagnosis of Pulmonary Diseases. *Med Devices (Auckl)*. Apr 7;15:89-102. doi: 10.2147/MDER.S362407.

38. Minami K., Lu H., Kim H., Mabu S., Hirano Y., Kido S. (2019). Automatic Classification of Large-Scale Respiratory Sound Dataset Based on Convolutional Neural Network. 19th International Conference on Control, Automation and Systems (ICCAS), Jeju, Korea (South), 2019, pp. 804-807, doi: 10.23919/ICCAS47443.2019.8971689.

39. Demir F., Sengur A., Bajaj V. (2019). Convolutional neural networks based efficient approach for classification of lung diseases. *Health Inf Sci Syst*. Dec 23;8(1):4. doi: 10.1007/s13755-019-0091-3.

40. Jayalakshmy S., Lakshmi P., Gnanou F. (2021). Synthesis of respiratory signals using conditional generative adversarial networks from scalogram representation. Chapter 7. Generative Adversarial Networks for Image-to-Image Translation, Academic Press, 2021, p. 161-183. doi: 10.1016/B978-0-12-823519-5.00024-5.
41. Viet H., Ngoc H., Anh V., Quang H. (2022). Classification of lung sounds using scalogram representation of sound segments and convolutional neural network. Journal of Medical Engineering & Technology, 46:4, 270-279, doi: 10.1080/03091902.2022.2040624.
42. Abayomi-Alli O., Damaševičius R., Qazi A., Adedoyin-Olowe M., Misra S. (2022). Data Augmentation and Deep Learning Methods in Sound Classification: A Systematic Review. *Electronics* 11, no. 22: 3795. doi: 10.3390/electronics11223795.
43. Alqudah A.M., Qazan S., Obeidat Y.M. (2022). Deep learning models for detecting respiratory pathologies from raw lung auscultation sounds. *Soft comput.* 2022;26(24):13405-13429. doi: 10.1007/s00500-022-07499-6.
44. Bardou D., Zhang K., Ahmad S.M. (2018). Lung sounds classification using convolutional neural networks. *Artificial Intelligence in Medicine*, Volume 88, 2018, Pages 58-69, doi: 10.1016/j.artmed.2018.04.008.
45. Gairola S. (2020). A deep neural network for abnormal lung sound detection in limited data settings. 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC): 527-530.
46. Saldanha J., Chakraborty S., Patil S., Kotecha K., Kumar S., Nayyar A. (2022). Data augmentation using Variational Autoencoders for improvement of respiratory disease classification. *PLoS One*. Aug 12;17(8):e0266467. doi: 10.1371/journal.pone.0266467.
47. Yi M., Xinzi X., Yongfu L. (2020). LungRN+NL: An Improved Adventitious Lung Sound Classification Using Non-Local Block ResNet Neural Network with Mixup Data Augmentation. 10.21437/Interspeech.2020-2487.

48. Grønnesby M., Solis J.C., Holsbø E.J., Melbye H., Bongo L.A. (2017). Feature Extraction for Machine Learning Based Crackle Detection in Lung Sounds from a Health Survey. arXiv: Sound. May, 2017.
49. Jung S.-Y. (2020). Efficiently Classifying Lung Sounds through Depthwise Separable CNN Models with Fused STFT and MFCC Features. *Journal of Healthcare Engineering*, 2020, 1-12.
50. Bahoura M. (2009). Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes. *Computers in Biology and Medicine*, Volume 39, Issue 9, p. 824-843. doi:10.1016/j.compbimed.
51. Rocha B.M., Pessoa D., Marques A., Carvalho P., Paiva R.P. (2021). Automatic Classification of Adventitious Respiratory Sounds: A (Un)Solved Problem? *Sensors* 2021,21, 57. doi: 10.3390/s21010057.
52. Jácome C., Ravn J., Holsbø E., Aviles-Solis J.C., Melbye H., Ailo Bongo L. (2019). Convolutional Neural Network for Breathing Phase Detection in Lung Sounds. *Sensors Basel*. 2019 Apr 15;19(8):1798. doi: 10.3390/s19081798.
53. Hsu F.S., Huang S.R., Huang C.W., Huang C.J., Cheng Y.R. et al. (2021). Benchmarking of eight recurrent neural network variants for breath phase and adventitious sound detection on a self-developed open-access lung sound database. *HF Lung V1. PLOS ONE* 16(7): e0254134. doi: 10.1371/journal.pone.0254134.
54. Fraiwan L., Hassanin O., Fraiwan M., Khassawneh B., Ibnian A., Alkhodari M. (2020). Automatic identification of respiratory diseases from stethoscopic lung sound signals using ensemble classifiers. *Biocybernetics and Biomedical Engineering*. doi:10.1016/j.bbe.2020.11.003.
55. Mesaros A., Heittola T., Virtanen T. (2016). Metrics for Polyphonic Sound Event Detection. *Appl. Sci.* 6, 162. doi:10.3390/app6060162.