

МОГИЛЕВИЧ ДМИТРО

КПІ ім. Ігоря Сікорського

<https://orcid.org/0000-0002-4323-0709>e-mail: mogilev11@ukr.net**ХМІЛЬ РОМАН**

КПІ ім. Ігоря Сікорського

<https://orcid.org/0009-0001-6839-4152>e-mail: rkhamil@gmail.com

МЕТОДОЛОГІЯ КЛАСИФІКАЦІЇ ЦИФРОВИХ АУДІО-СИГНАЛІВ НА ОСНОВІ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ

Розвиток засобів телекомунікації, мережі інтернет, соціальних мереж, та поширення різноманітних мобільних пристроїв запису аудіо- та відео-інформації призводить до вибухового зростання кількості мультимедійних даних, зокрема, аудіо-файлів. Прикладом звукових даних є записи звуків різноманітного походження, відео з камер спостереження та відео-реєстраторів, мобільних телефонів тощо. Тому виникає потреба у автоматизації пошуку, розпізнавання, ідентифікації та класифікації аудіо-візуальної інформації.

У даній статті описується методологія ідентифікації та класифікації автомобільних аварій на основі їх аудіо-сигналів та математичного апарату згорткових нейронних мереж. Основною задачею розробленої методології є підвищення точності класифікації аварій, та покращення енерго-ефективності методів та алгоритмів, що для цього використовуються, з метою забезпечення їх роботи у режимі реального часу на мобільних та периферійних пристроях з обмеженими обчислювальними потужностями. Актуальність даної задачі пов'язана з сучасними вимогами до безпеки автомобілів та запровадження країнами Європейського Союзу та США вимог до обладнання усіх нових автомобілів системами автоматичного сповіщення про аварії (ACN – Automatic Crash Notification.). Оскільки реагування екстрених служб на такі повідомлення надзвичайно дорого-вартісне, виникає завдання підвищення точності методик аналізу дорожніх подій.

На сьогоднішній день більшість авто-виробників використовують для цього датчики акселерометрів. Проте, використання виключно даних про прискорення часто призводить до помилкових класифікацій: ями на дорозі визначаються як аварія, тоді як удар ззаду може бути класифікований як нормальне прискорення. В результаті середня точність таких методик не перевищує 85%, що прийнятно, однак вимагає подальшого удосконалення. Останніми роками з'явилася значна кількість сучасних досліджень, котрі пропонують використовувати згорткові нейронні мережі для класифікації аудіо-сигналів аварійних подій. Ефективність таких підходів суттєво вища та знаходиться на рівні 85%-95%. Найкращі результати у 96% було досягнуто авторами [9], котрі використали ансамбль з трьох згорткових нейронних мереж. Проте такий підхід є досить енерго-витратним і не може бути використаний для класифікації аварій у режимі реального часу на периферійних пристроях з обмеженими обчислювальними можливостями.

У даній статті описується методологія класифікації автомобільних аварій на основі використання згорткових нейронних мереж шляхом класифікації аудіо-сигналів, котра забезпечує високу точність класифікації на рівні 98% та є при цьому енерго-ефективною у порівнянні з іншими існуючими підходами.

Ключові слова: штучний інтелект, згорткові нейронні мережі, обробка аудіосигналів.

MOGYLEVYCH DMYTRO, KHMIL ROMAN

Igor Sikorsky Kyiv Polytechnic Institute

AUDIO-SIGNALS CLASSIFICATION METHODOLOGY USING CONVOLUTIONAL NEURAL NETWORKS

The development of telecommunications, the Internet, social networks, and various mobile devices for recording audio and video information leads to an explosive growth in the number of multimedia data, in particular, audio files. An example of sound data is recordings of sounds of various origins, videos from surveillance cameras and video recorders, mobile phones, etc. That is why there is a need to automate the search, recognition, identification and classification of audio-visual information.

This article describes the methodology for identifying and classifying car accidents based on their audio signals and the mathematical apparatus of convolutional neural networks. The main task of the developed methodology is to improve the accuracy of accident classification and improve the energy efficiency of the methods and algorithms used for this purpose in order to ensure their operation in real time on mobile and peripheral devices with limited computing power. The relevance of this task is driven by new car safety standards and the introduction by the countries of the European Union and the United States of requirements for equipping all new cars with automatic accident notification systems (Eng. Automatic Crash Notification, ACN). Since the response of emergency services (police, ambulance, etc.) to such reports is extremely expensive, the task of increasing the accuracy of methods of analyzing road events arises.

Nowadays, most automakers use accelerometer sensors for this purpose. However, the use of acceleration data alone often leads to misclassifications: pits on the road are defined as an accident, while a rear impact can be classified as normal acceleration. As a result, the average accuracy of such techniques does not exceed 85%, which is acceptable, but obviously requires further improvement. In recent years, a significant number of modern studies have appeared that propose the use of convolutional neural networks for classifying audio signals of emergency events. The effectiveness of such approaches is significantly higher and is at the level of 85%-95%. The best results of 96% were achieved by authors [9], who used an ensemble of three convolutional neural networks. Nevertheless, this approach requires considerable energy and cannot be applied for real-time accident detection on peripheral devices with restricted processing power.

This article describes the methodology of car crash classification based on the crash audio signals and the use of convolutional neural networks as a classifier, achieving higher accuracy of 98% and better energy-efficiency comparing with other existing methodologies.

Keywords: artificial intelligence, convolutional neural networks, audio signal processing.

Стаття надійшла до редакції / Received 22.10.2025

Прийнята до друку / Accepted 15.11.2025

Постановка проблеми у загальному вигляді та її зв'язок із важливими науковими чи практичними завданнями

Дослідження, проведені Всесвітньою організацією охорони здоров'я (ВООЗ) свідчать, що в середньому 1,3 мільйона людей помирають щороку через дорожньо-транспортні пригоди, від 20 до 50 мільйонів отримують травми [10]. Беручи до уваги, що шанси на виживання безпосередньо залежать від своєчасної медичної допомоги, стає критично важливим покращити існуючі системи ACN, щоб скоротити час надання екстреної допомоги.

Актуальність розробки інноваційних підходів для виявлення та класифікації автомобільних аварій визначається економічною вигодою підвищення точності існуючих підходів. Кожне додаткове підвищення точності дозволяє заощаджувати мільярди доларів на прямі витрати, пов'язані з помилковими викликами екстрених служб.

У цій статті описується інноваційний підхід до підвищення точності виявлення аварій на основі згорткових нейронних мереж та можливі напрямки подальшого вдосконалення запропонованого підходу. Автоматично виявляючи аномальні звуки дорожнього руху, а саме автомобільні аварії та заноси, запропонована методологія допомагає зменшити помилкові спрацювання та пропущені тривоги, значно покращуючи точність розпізнавання аварій.

Аналіз досліджень та публікацій

Найефективнішими для вирішення задачі класифікації цифрових аудіо-сигналів на сьогодні є згорткові нейронні мережі (CNN). Вони придатні для застосування до будь-якого сигналу, будь то дані з датчиків, аудіо-сигналів, зображень тощо [8]. Цей тип нейронної мережі є багатошаровим перцептроном, що складається з багатьох рівнів вузлів, прихованих і вихідних шарів, і має односторонній потік інформації. Універсальна теорема наближення стверджує, що нейронна мережа з одним прихованим шаром і кінцевою кількістю вузлів може наблизити будь-яку безперервну функцію до будь-якого ступеня точності. У завданнях розпізнавання образів цей тип нейронної мережі з нелінійною S-подібною функцією активації та кількома шарами здатен розпізнавати об'єкти з високою точністю. Особливості багатошарової нейронної мережі прямого поширення створюють теоретичну основу для використання багатошарових перцептронів у моделюванні та діагностиці помилок розпізнавання образів. Помилки можуть визначатися двома способами: шляхом програмування моделі для корекції розпізнавання образів або вибором відповідного класифікатора шаблонів [1]. Якість розпізнавання зображень нейронними мережами визначається ефективністю їх навчання на певній вибірці даних із використанням великої кількості тренувальних пар «вхід-вихід». За результатами навчання формується функція помилки або відхилення (функція втрат). Мета навчального процесу нейронної мережі полягає в мінімізації цих помилок, що дозволяє штучному інтелекту самостійно підбирати оптимальні значення ваг з'єднань між нейронами. Ключова особливість і відмінність між CNN і стандартним перцептроном полягає в тому, що шарові нейрони не мають індивідуальних ваг, але використовують спільні ваги: матриці ваг, які також називають ядрами згортки. Отже, згорткова мережа містить значно менше параметрів, ніж повністю зв'язана мережа, що забезпечує її вищу продуктивність та економію обчислювальних ресурсів.

Формулювання цілей статті

Метою цієї статті є аналіз існуючого математичного апарату згорткових нейронних мереж, опис інноваційного підходу до вирішення проблеми розпізнавання та перевірки автомобільних аварій на основі їх використання, а також вибір оптимальної архітектури нейронної мережі для ефективного вирішення проблеми на основі мобільних пристроїв, а також інших джерел звуку, які можуть бути доступні всередині автомобіля під час аварії.

Для досягнення цієї мети необхідно вивчити звукові властивості автомобільних аварій, запропонувати методи звукової фільтрації, здатні зменшити нерелевантні шуми, провести навчання згорткових нейронних мереж на вибраних зразках аварій та оцінити якість класифікації. Для того, щоб навчання нейронних мереж відбувалося на достатній та репрезентативній вибірці навчання, необхідно вивчити методи збільшення набору навчальних та валідаційних даних.

Виклад основного матеріалу

Найбільш ефективними для вирішення задачі класифікації цифрових аудіо-сигналів на сьогодні є згорткові нейронні мережі (ЗНМ). Їх застосування можливе для будь-якого сигналу, будь то дані з датчиків, аудіо-сигналів, зображень тощо [8]. Даний тип нейронної мережі є багатошаровим перцептроном, що складається з чисельних рівнів вузлів, прихованих і вихідних шарів, і має односторонній потік інформації.

Універсальна теорема наближення стверджує, що нейронна мережа з одним прихованим шаром і кінцевою кількістю вузлів може наблизити будь-яку безперервну функцію до будь-якого ступеня точності. У завданнях розпізнавання образів цей тип нейронної мережі з нелінійною S-подібною функцією активації та кількома шарами здатен до високоточного розпізнавання об'єктів. Особливості багатошарової нейронної мережі прямого поширення формують теоретичну основу для використання багатошарових перцептронів у моделюванні та діагностиці помилок розпізнавання образів. Виявлення помилок може здійснюватися двома шляхами: шляхом програмування моделі для корекції процесу розпізнавання або через вибір відповідного класифікатора шаблонів. [1].

Якість розпізнавання зображень нейронними мережами визначається ефективністю навчання на певній вибірці даних із використанням великої кількості тренувальних пар «вхід-вихід». На основі результатів навчання формується функція помилки або відхилення (функція втрат). Мета навчального процесу нейронної

мережі полягає у мінімізації цих помилок, що дозволяє штучному інтелекту автоматично підбирати оптимальні значення ваг з'єднань між нейронами.

Ключова особливість і відмінність між ЗНМ і стандартним перцептроном полягає в тому, що шарові нейрони не мають індивідуальних ваг, але використовують спільні ваги: матриці ваг, які також називають ядрами згортки. Таким чином, згорткова мережа має значно меншу кількість параметрів у порівнянні з повністю зв'язаною мережею, з чого слідує її вища продуктивність та економія у використанні обчислювальних ресурсів.

Архітектура MobileNet. На сьогодні найбільш передовою архітектурою ЗНМ, що забезпечує високу швидкість для роботи в режимі реального часу на мобільних пристроях, є MobileNet-V3.

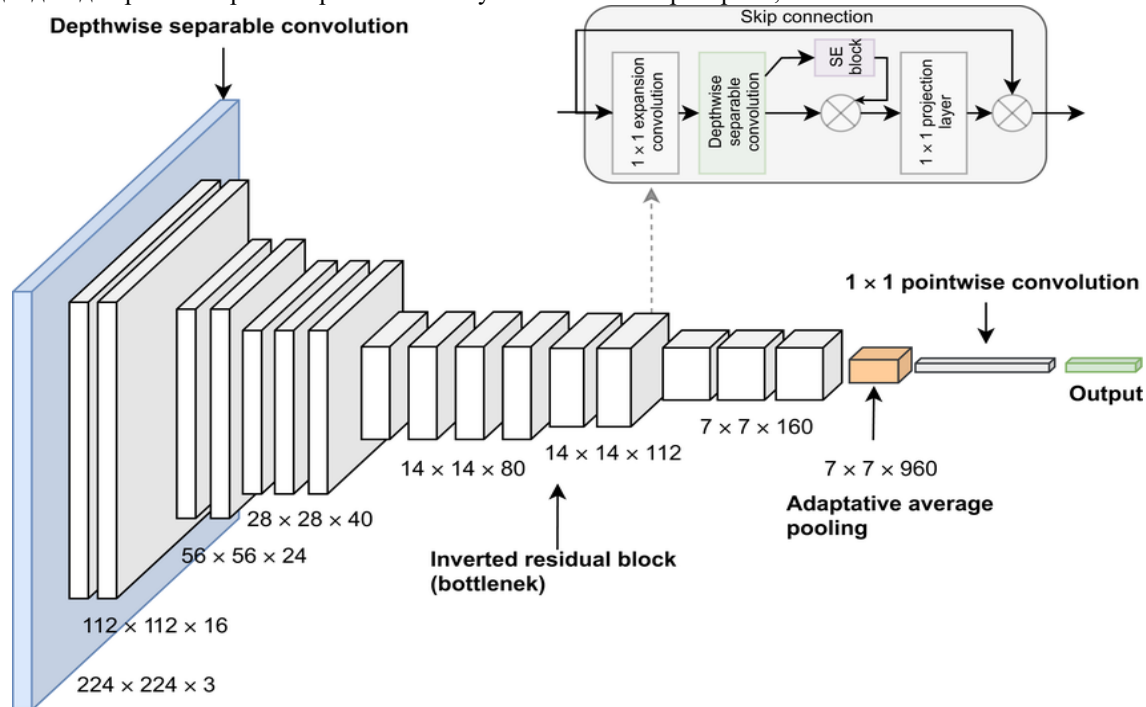


Рис. 1. Архітектура MobileNet-V3 ЗНМ

Основна концепція архітектури MobileNet полягає в досягненні високої точності класифікації та виявлення об'єктів при мінімальному споживанні обчислювальних ресурсів і пам'яті. MobileNet-V1 використовує двохшарову згортку, що дозволяє значно скоротити кількість параметрів і обчислень, зберігаючи при цьому високу точність розпізнавання. Крім того, за допомогою налаштування коефіцієнта ширини мережі та зміни розміру вхідного зображення можливо регулювати ефективність розпізнавання та обсяг пам'яті, необхідний для розгортання мережі. Вагомою перевагою такої архітектури є те, що згортки 3×3 вимагають в 8-9 разів менше операцій, з подальшим зменшенням кількості параметрів. Наявність меншої кількості параметрів також дозволяє зменшити кількість даних, необхідних для навчання. Це дуже важлива і незначна особливість при вирішенні проблеми розпізнавання звукових подій, оскільки кількість даних, доступних для навчання, дуже обмежена.

MobileNet-V2 оснащена додатковим шаром, а саме інвертованим залишковим блоком з лінійними вузькими місцями, що зменшують розмір вхідних даних та ефективно розширюють та стискають функції; цей шар приймає на вхід низько-вимірне стиснене представлення, яке розширюється, а потім фільтрується глибинним шаром згортки, зменшуючи кількість параметрів моделі у $\sim 8-9$ разів порівняно зі стандартними згортками. Іншими словами, кількість карт функцій збільшується (розширення), а потім зменшується (стиснення); дані розсіюються у просторі вищого виміру, щоб нелінійна потужність функції активації ReLU могла бути використана без втрати інформації.

MobileNet-V3 - це передова архітектура для моделей глибокого навчання, розроблена для мобільних та периферійних пристроїв, яка забезпечує хороший компроміс між точністю, кількістю операцій з плаваючою комою та кількістю параметрів, необхідних для навчання мережі. Вона впроваджує нові вдосконалення архітектури, такі як пошук нейронної архітектури з оптимізацією платформи (NAS) для оптимізації архітектури для мобільних процесорів, NetAdapt для тонкого налаштування мережових шарів з метою підвищення ефективності, модулі Squeeze-and-Excite (SE) для покращення навчання та функцію активації H-Swish для підвищення обчислювальної ефективності моделі.

Одним із ключових удосконалень є застосування мобільної оберненої згортки вузького місця (MBConv), представлені в MobileNet-V2, яка в MobileNet-V3 додатково удосконалена шляхом інтеграції легких механізмів уваги, відомих як блоки Squeeze-and-Excitation (SE). Ці SE-блоки адаптивно перекалібрують каналні реакції на ознаки, дозволяючи моделі підсилювати більш інформативні характеристики та пригнічувати менш значущі. Кінцевий класифікаційний шар MobileNet-V3 розрахований на 1000 класів ImageNet, і його потрібно замінити шаром, відповідно до кількості класів конкретного набору даних.

Функція активації. MobileNet-V3 використовує функцію активації Hard-Swish, яка є обчислювально ефективним наближенням функції Swish, та обчислюється наступним чином:

$$\text{Hard-Swish}(x) = x \frac{\text{ReLU6}(x+3)}{6}, \text{ReLU6}(x) = \min(\max(0, x), 6).$$

Hard-Swish транслює вхідне значення, якщо воно знаходиться між 0 і 6, і стискає вихід до 0 або 6 в іншому випадку. Вона використовується для створення швидшої та більш ефективної альтернативи функції *Swish*, оскільки використовує ReLU6 замість більш складної сигмоїдної функції, зберігаючи при цьому більшість її переваг, таких як покращена точність моделі в складних задачах розпізнавання зображень. Як і *Swish*, *Hard-Swish* не є монотонною, тобто може опускатися нижче нуля для невеликого діапазону негативних входів. Ця властивість може допомогти моделям вивчити складні закономірності. Це особливо корисно, коли потрібен баланс між точністю та обчислювальною ефективністю, що стосується мобільних/периферійних пристроїв.

Функція втрат. Функція втрат — це функція, яка характеризує величину втрат у разі неправильного прийняття рішень на основі спостережуваних даних. Вона є методом оцінки того, наскільки точно алгоритм моделює заданий набір даних. Основною метою функції втрат у нейронній мережі є оцінка похибки та оновлення ваг нейронів для покращення прогнозів на наступному кроці. Функція втрат FFF може бути реалізована через середню абсолютну похибку, середньо-квадратичну логарифмічну похибку, перехресну ентропію, фокусну втрату, Softmax, Cosine та інші методи.

В даній роботі використовується фокусна втрата, яка виявилася кращою альтернативою ніж перехресна ентропія у випадку наявності дисбалансу класів. Це спостерігається, коли інформація, пов'язана з певним класом у наборі даних, що використовуються в навчанні, представлена частіше, ніж інші класи. Навчання мережі на незбалансованому наборі даних робить мережу упередженою до вивчення більшої кількості уявлень про клас, який домінує у навчальних зразках, тоді як інші класи будуть недооцінені.

Функцію фокусної втрати можна розрахувати наступним чином:

$$\text{FocalLoss}(p_i) = -\frac{1}{N} \sum_{i=1}^N \alpha_i (1 - p_i)^\gamma \log_b(p_i),$$

де N – кількість зразків для навчання;

p_i – прогнозована ймовірність моделі для правильного класу, при наближенні p_i до 1, приклад вважається «легким», і його внесок у втрати зменшується;

γ – параметр фокусування, настроюваний гіпер-параметр, який контролює, наскільки зважені «легкі» зразки ($\gamma = 0$ робить фокусну втрату еквівалентною стандартній перехресній ентропії);

α_i (зазвичай використовується 0.25) – ваговий коефіцієнт, подібний до того, що використовується в збалансованій перехресній ентропії, щоб вручну збалансувати важливість позитивних/негативних класів;

$\log(p_i)$ – стандартний компонент перехресної ентропії.

Чим більше значення γ буде надано хибно класифікованим прикладам, тим менше втрат буде поширюватися від «легких» зразків. Згідно з дослідженням Lin та ін. [23], значення $\gamma = 2$ дає найкращі результати.

Цей механізм дозволяє запобігти перевантаженню моделі величезною кількістю «легких» зразків під час навчання, що особливо корисно в таких завданнях, як виявлення об'єктів, де часто домінує певний клас.

Методика класифікації аудіо-сигналів аварій на основі ЗНМ. Загальна ідея використання нейронних мереж для вирішення задачі класифікації аудіо-сигналів автомобільних аварій полягає в наступному:

використовуючи смартфон та/або інші пристрої з мікрофонами, постійно (в циклічному режимі) записуються аудіо-сигнали всередині автомобіля. Постійно записуються лише останні 10 секунд, чого в разі аварії достатньо для аналізу;

записані звукові доріжки аналізуються згортковою нейронною мережею в режимі реального часу, щоб класифікувати акустичні події, що відбуваються.

Для того, щоб значно зменшити навантаження на процесор, робота нейронної мережі може обмежуватися випадками, коли аварійна ситуація попередньо ідентифікується алгоритмами на основі даних акселерометра та гіроскопа, встановлених у автомобілі або мобільному телефоні, що слугують джерелом сигналів. У такому випадку результати попереднього аналізу виконують роль тригера для запуску роботи штучної нейронної мережі. Подібний підхід був описаний у роботі Расіогек та ін. [6].

Для того, щоб точність виявлення аварій була вищою, ніж забезпечують алгоритми на основі даних акселерометрів, необхідно повністю замінити дані акселерометрів безперервним аналізом звукової доріжки за допомогою ШНМ в режимі реального часу. Беручи до уваги, що постійне використання процесора мобільного пристрою може швидше розряджати його акумулятор, смартфон потрібно підключати до зарядного пристрою.

Основні завдання при розробці архітектури ШНМ є наступними:

максимізувати Accuracy: правильно ідентифікувати аварійні події;

максимізувати Recall: мінімізувати False Negative (пропущені аварії);

максимізувати F1-miry: хороший баланс високих Precision та Recall;

високий рівень Precision: мінімізація помилкових спрацьовувань (False Positives);

енерго-ефективність: скорочення часу та ресурсів для обчислень;

надійність: добре узагальнює події для різних умов та ситуацій;

інтерпретованість: зрозуміти, які особливості сприяють найбільше.

В результаті для вирішення поставленої задачі було обрано та дорацьовано архітектуру *MobileNet-V3*, яка максимізує точність обчислень на мобільних пристроях за допомогою глибинних згорток, та оптимізованого розміщення блоків, що дозволяє прискорити обробку аудіо-сигналів, вчитися безпосередньо з аудіо-сигналу та в результаті забезпечує найкращу ефективність класифікації аудіо-сигналів в режимі реального часу.

Підготовка набору даних. Запропонований підхід полягає у використанні унікального набору даних з 18000+ аудіо-зразками аварій, який був створений на основі відео автомобільних аварій, доступних в YouTube. Для цього було використано автоматичний пошук відповідних зразків аварій, та алгоритм їх попередньої обробки та розмітки. Цей набір даних використовується для навчання нейронної мережі, тоді як її продуктивність оцінюється на загальнодоступному наборі даних *MIVIA Road Events*, який зазвичай використовується вченими для порівняння продуктивності моделі.

Набір даних *MIVIA* структурований таким чином, щоб представити кожну аудіо-подію на шести рівнях співвідношення сигнал/шум (5 дБ, 10 дБ, 15 дБ, 20 дБ, 25 дБ і 30 дБ), з різноманітними комбінаціями звуків навколишнього середовища для імітації різних умов навколишнього середовища. Він складається з 57 аудіо-файлів тривалістю 60 секунд, оцифрованих з частотою 32 кГц і розрізняє автомобільні аварії та занос шин: 200 подій позначені як автомобільна аварія, а 200 - як занос/скрежет шин.

Попередня обробка даних. Оскільки довжина вхідних зразків, що подаються на вхід ЗНМ, має бути фіксованою, було експериментально визначено оптимальну довжину сегмента аудіо-зразків. Застосування процедури «ковзного вікна», яка сегментує вхідні аудіо-доріжки на кадри, які перекриваються на стиках, дозволяє обробляти звукові доріжки будь-якої довжини. Ширина ковзного вікна залежить від частоти вибірки сигналу. Також послідовні аудіокадри можуть мати деякий відсоток перекриття, тому оптимальний відсоток перекриття було визначено експериментально.

Запропонована ЗНМ оцінювалася за різними довжинами аудіо-кадрів (100 мсек., 200 мсек., 500 мсек., 1 сек., 2 сек., 3 сек.) та відсотками перекриття (0%, 25%, 50% та 75%), щоб оцінити вплив вхідної довжини та перекриття на ефективність класифікації. Експерименти показують, що короткі сегменти не містять достатньо інформації для належного навчання ЗНМ, а найвища точність досягається при використанні 1-сек. кадрів з 50% перекриттям. Процес сегментації аудіосигналу у відповідні кадри проілюстровано на рис. 2.

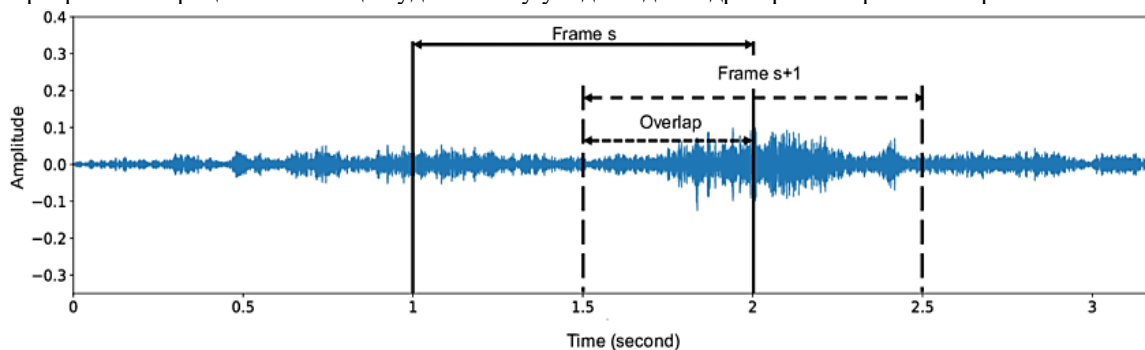


Рис. 2. Сегментація вхідного аудіо-сигналу на кадри з 50% перекриттям

Частота дискретизації аудіосигналів безпосередньо впливає на розмір вхідного зразка і, зрештою, на обчислювальну вартість моделі. Для звуків автомобільних аварій частоту дискретизації 32-41 кГц можна вважати хорошим компромісом між якістю вхідного зразка та обчислювальною вартістю моделі.

Попередня обробка даних включає як вихідні аудіосигнали, так і їх графічне представлення. В основному вона спрямована на підготовку даних для їх використання запропонованим методом глибокого навчання шляхом і) збереження важливих ознак, таких як потужність сигналу та ii) зменшення діапазону значень для прискорення обробки.

Різні мікрофони мобільних або стаціонарних пристроїв можуть мати різну чутливість, тому амплітуда звукових сигналів, що подаються на вхід нейронної мережі, може суттєво відрізнятися навіть у однакових аварійних ситуаціях. Отже, при попередній обробці необроблених даних потрібна нормалізація даних: для кожного вхідного саундтреку аудіосигнал було масштабовано таким чином, щоб довести найвищий пік амплітуди (в абсолютному значенні) до максимально можливого.

Отже, для того, щоб живити нейронну мережу входами фіксованого розміру, 1-секундні аудіо-кадри послідовно сегментуються за допомогою розсувного вікна, з 0.5-секундним зміщенням між ними так, щоб кожен наступний кадр містив 50% інформації попереднього, щоб запобігти втраті інформації і полегшити виявлення подій, які можуть відбуватися в крайніх точках окремих кадрів.

Отримані кадри використовуються для створення візуального представлення аудіо, тобто його спектрограми. Вона представляє собою спектр сигналу як функцію частоти та часу, отримані шляхом застосування короткочасного перетворення Фур'є (англ. STFT):

$$X[k, m] = \sum_{n=0}^{N-1} w[n]x[n + mH]e^{-\frac{2i\pi}{N}kn},$$

де $x[n]$ – вхідна послідовність дискретних відліків сигналу;

k – індекс частотного біна;

m – індекс часового кадру (положення вікна) ;

N – довжина вікна (кількість відліків у кадрі) ;

H – крок зсуву вікна у відліках;

$w[n]$ – дискретна віконна функція (наприклад, вікно Гана, прямокутне вікно), яка «вирізає» короткий сегмент сигналу.

Підсумовуючи, сигнал розбивається на сегменти, після чого до них застосовується SFFT. Внаслідок цього перетворення кожен аудіокадр переходить з одновимірного вектора в 2D-матрицю, до якої вже можна застосовувати методи обробки зображень за допомогою ЗНМ.

З практичної сторони, створення спектрограм з набору вхідних даних передбачало використання бібліотеки *matplotlib* зі значеннями за замовчуванням, включаючи вікно Гана з STFT з 128 зразків і перекриттям 64 зразків, застосованих на 1-секундних кадрах, відібраних на частоті 44,1 кГц. Отримані спектрограми були стиснуті до розміру 224x224 пікселів, щоб прискорити наступні етапи попередньої обробки та зменшити кількість параметрів, необхідних нейронній мережі.

На наступних кроках значення інтенсивності пікселів були нормалізовані в діапазоні [0,1] та стандартизовані. На останньому етапі попередньої обробки до спектрограм було застосовано знешумлення за допомогою фільтру Гауса з розміром ядра 3x3.

Для збереження характерних сигнатур аварій та покращення стійкості до шумів було використано аугментацію даних (помірно агресивна форма хвилі з конвеєром SpecAugment).

Дизайн архітектури ШНМ. Враховуючи обрану довжину кадру аудіосигналу дорожньо-транспортної пригоди у 1 сек. та з урахуванням всіх задач дизайну нейронної мережі, описаних вище, було адаптовано архітектуру MobileNet-V3, що складається з 19 згорткових шарів: роздільні глибинні згортки, інвертовані залишкові блоки, блоки SE на більш глибоких стадіях.

Вхід представлений 10 каналами і кодує ключові репрезентативні особливості, які характеризують акустичні події: MEL спектрограма, короткострокова часова динаміка (дельта/швидкість, дельта-дельта/прискорення), 5 ознак MFCC, спектральна енергія (2 проекції інтенсивності спектрограм).

Спеціальний шар уваги вхідного каналу застосовується перед основною частиною мережі, щоб збалансувати вхідні сигнали. Кілька згорткових шарів застосовуються для захоплення точної часової структури сигналу і служать фільтрами, які виконують завдання класифікації аварій. Це також дозволяє відмовитися від додаткового модуля обробки сигналів, оскільки така нейронна мережа достатньо потужна для виділення як низькорівневої, так і високорівневої інформації з вхідних аудіосигналів.

Оскільки обсяг вхідних даних для навчання мережі обмежений, застосування глибших архітектур нейронних мереж є недоцільним, тому що це може призвести до явища «перенавчання». При перенавчанні нейронна мережа правильно розпізнає дані тренувальної вибірки, але демонструє низьку точність на нових даних. Це відбувається тоді, коли модель «запам'ятовує» тренувальні дані замість того, щоб узагальнювати закономірності. Як наслідок, перенавчена модель стає менш ефективною через надмірну чутливість до відхилень у тренувальних даних. Архітектура запропонованої нейронної мережі зображена нижче на рисунку 3.

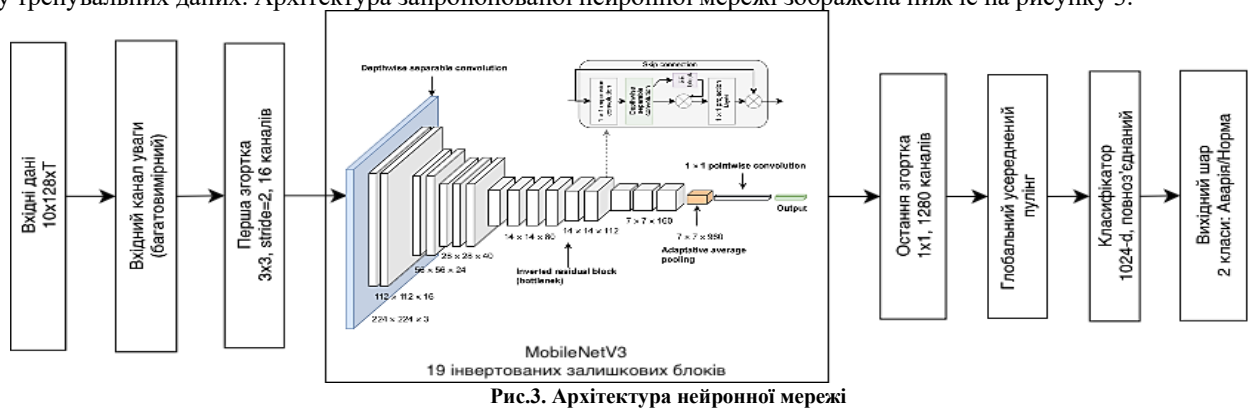


Рис.3. Архітектура нейронної мережі

Функція активації *Hard-Swish* використовується для всіх шарів нейронної мережі, яка виявилася найбільш ефективною з архітектурою *MobileNet-V3*. 1280-d шар глобального усередненого пулінгу і легкий повнозв'язний шар класифікації (щільний перцептрон з 1024 нейронів з 0,32 дроп-аутом) забезпечують компактність моделі з метою підвищення енерго-ефективності та роботи в режимі реального часу на мобільних пристроях.

Дизайн нейронної мережі, описаний вище, забезпечує необхідний компроміс між точністю та енерго-ефективністю, що дозволяє класифікувати аудіо-сигнали в режимі реального часу. Результати вимірів показали усереднений час класифікації нових зразків у 145 мсек., що є більш ніж достатнім для вирішення поставленої задачі.

Навчання моделі. Нейронна мережа навчається з використанням фокусних втрат як функції втрат з наступними параметрами, оптимізованими експериментально: $\alpha_{\text{crash}}=3.7$, $\gamma=2.5$ з адаптивним згладжуванням міток (0→0.09 over 10 epochs). Адаптивне згладжування міток разом з оптимізатором AdamW та планувальником косинусів фокусується на досягненні високої точності (Accuracy) у випадку незбалансованих аварійних та нормальних даних.

Також були реалізовані механізми уваги: спеціальний модуль уваги вхідного каналу (avg/max pooling + тимчасова увага + FC re-weighting) поєднує мульти-масштабне об'єднання з тимчасовою увагою для адаптивного зважування кожного каналу на зразок.

Навчання моделі можна зупинити після 29-ї ітерації, як показано на рисунку 4

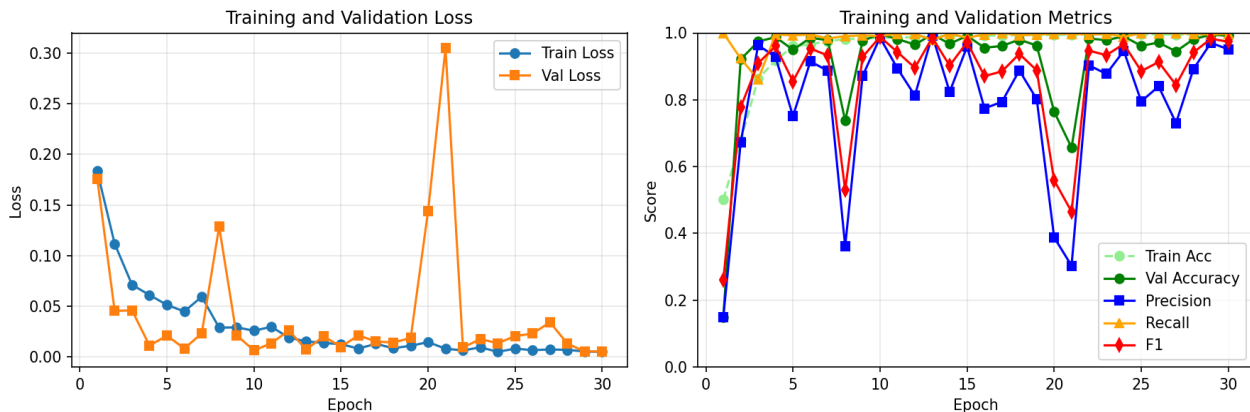


Рис. 4. Навчальні ітерації протягом 30 епох

Для оцінки точності моделі під час навчання застосовувалася 10-кратна перехресна перевірка. Для кожного класу функція втрат визначалася як мінімальне значення втрат серед усіх тренувальних ітерацій (епох) у процесі перевірки. Аналогічно, точність для кожного класу обчислювалася на основі найкращого значення класифікації, отриманого в кожному епоху.

Для того, щоб кількісно перевірити ефективність запропонованого методу та адекватно порівняти його з існуючими, ми використовуємо такі показники: Accurasy (Точність), Precision (Влучність), Recall/TPR (частота вірно-позитивних висновків), TNR (частота вірно-негативних класифікацій), FPR (частота хибно-позитивних висновків), MR/FNR (частота хибно-негативних висновків), ER (Рівень Помилки, $= 1 - \text{Accuracy}$), F1-міра, та EDR (Event Detection Rate).

Експериментальні результати демонструють надзвичайно високу точність класифікації запропонованої моделі на рівні 98%+, що краще, ніж інші відомі на сьогодні методики (табл. 1), які базуються на обробці сигналів від акселерометрів/гіроскопів автомобіля (80%-85%) [5], а також ніж ті моделі, які використовують аудіо-сигнали/спектрограми для класифікації аварійних подій (85%-96%) [6, 7, 12 – 22].

Запропонована методика забезпечує найкращі результати за усіма показниками ефективності: Accurasy 99,2%, Precision 95,1%, Recall (TPR) 99,8%, FPR 3,1%, MR 0,2%, ER 0,8%, F1-міра 97,7% та EDR 100%. Перевірка проводилась на публічному наборі даних MIVIA Road Events, що використовується іншими вченими. Швидкодія розробленої моделі також показує результат у 145 мсек. при класифікації нових зразків, що є більш ніж прийнятним для забезпечення класифікації аудіо-сигналів у режимі реального часу.

Таблиця 1

Порівняння продуктивності сучасних моделей класифікації

Рік	Автори	Архітектура	Вхідні дані	Accuracy	Precision	F1-Score	TPR	FPR	MR	ER
2018	Zoph	CNN, NASNet	Spectrogram 2D	92.8			84.6	5.5	8.2	7.2
2018	Mansar	CNN	Wave Form 1D	85.9±2.9	75.0	82.5	96.1±2.7	2.9±3.7	2.9±2.2	1.0±1.0
2019	Abdoli	CNN	Wave Form 1D	87.5±3.3	78.0	84.0	92.2±3.5	1.0±1.0	5.4±1.6	2.4±2.1
2019	Foggia	CNN, AReN	Gemmatonegram	79.0			88.9	3.8	7.0	4.2
2020	AENet	CNN, VGG	Spectrogram 2D	95.4			85.2	6.1	10.1	4.6
2019	Strisciuglio	CNN, COPE	Spectrogram 2D	94.0±4.3	94.8	94.7	95.2±4.9	3.9±1.8	4.8±4.9	1.3±1.3
2019	Choi	CNN, Ensemble	Spectrogram 2D	87.8			99.5	3.8	0.5	0.5
2019	Sammarco	CNN	Spectrogram+T/F	93.3	91.1	90.3	89.5		10.5	6.7
2020	Chitale	CNN, InceptionV3	Spectrogram 2D	93.8						
2022	Balia	CNN, FCNN	Spectrogram 3D	96.0			92.0	0.0	8.0	4.0
2024	Podda	CNN, Ensemble	Spectrogram+IPs	90.6±2.7	93.0	96.4	99.0±1.0	1.0±1.7	1.0±1.0	0.0±0.0
2024	Khnil	CNN, MobileNetV2	Spectrogram 2D	97.9	92.7	94.9	97.3			
2025	Khnil	CNN, MobileNetV3	Spectrogram 3D	99.2±1.0	95.1	98.0	99.8±1.0	3.1±1.0	0.2±1.0	0.8±1.0

Подальше покращення точності виявлення аварій. Беручи до уваги, що кожен додатковий відсоток точності виявлення аварій рятує життя і призводить до значної економії витрат, має сенс ще більше вдосконалити цю методологію.

У подальших дослідженнях має сенс дослідити ефект розширеного набору даних, вдосконалену попередню обробку та фільтрацію вхідних даних, агрегацію вхідних даних з декількох аудіо-джерел, а також застосування методів ансамблювання декількох мереж на здатність ЗНМ ще більше підвищувати точність класифікації.

Розширення тренувального набору даних. У більшості випадків кількість навчальних зразків навчального набору даних обмежена та/або недостатня для належного навчання мережі ЗНМ. Одним із способів розширити набір даних є створення додаткових копій аудіодоріжок шляхом деформації існуючих. Іншим варіантом було б партнерство з компаніями з виробництва автомобілів, які проводять аварійні випробування автомобілів як звичайний захід безпеки та збирають різноманітні дані, включаючи аудіозаписи.

Попередня обробка та фільтрація даних. Фільтри, навчені в проміжних згорткових шарах запропонованої ЗНМ, не проявляють виразних домінуючих частот і сприймаються шумними. Має сенс відфільтрувати шумові звуки під час попередньої обробки вхідних саундтреків і в той же час зосередитися на відповідних звуках, щоб правильно навчити ШНМ.

Декілька джерел звуків. Точність виявлення може бути додатково покращена за допомогою декількох джерел звуку. Як обговорювалося вище, сучасні автомобілі можуть бути оснащені різними пристроями, здатними записувати аудіо всередині автомобіля, а саме: вбудований мікрофон системи гучного зв'язку, мобільні телефони водія та/або пасажирів, пристрої запису відео-реєстратора, розумні дзеркала заднього виду тощо. Саундтреки з усіх цих джерел звуку можуть бути проаналізовані одночасно за допомогою паралельної обробки сигналів кількома нейронними мережами, при цьому результати агрегуються в кінці через правила злиття (більшість голосів або правило суми).

Ансамблі з нейронних мереж. Методи ансамблювання ШНМ, які поєднують кілька моделей ШНМ, можуть призвести до покращення загальної продуктивності моделі, поєднуючи прогнози з декількох моделей з хорошою точністю замість тонкого налаштування окремих моделей, зменшуючи дисперсію прогнозів і зменшуючи похибку через узагальнення. Також за допомогою ансамблів ШНМ можна подолати численні недоліки налаштування гіпер-параметрів, такі як неправильний вибір параметрів, пере-навчання та інші неефективні стратегії оптимізації.

Подальші дослідження спрямовані на оцінку того, чи можуть ці підходи забезпечити кращу ефективність при класифікації аварій. Також буде досліджено, як можна скорегувати запропоновані методи, щоб використовувати додаткову інформацію, надану поліфонічними звуковими подіями, записаними кількома джерелами звуку в транспортному засобі.

Висновки

У статті представлено методику класифікації автомобільних аварій на основі аналізу спектрограм аудіо-сигналів та відповідних похідних ознак із застосуванням згорткових нейронних мереж. Результати дослідження показали, що згорткові нейронні мережі, розроблені для розпізнавання зображень, можуть ефективно застосовуватися для класифікації аудіо-сигналів. Розроблена методологія забезпечує високу точність і енергоефективність класифікації аварійних подій за їх аудіо-сигналами.

Розроблено оригінальну багато-канальну архітектуру ЗНМ з шаром уваги, оптимізовану для вирішення поставленої задачі та запропоновано метод навчання такої ЗНМ. Для навчання нейронної мережі було зібрано та розмічено унікальний набір навчальних даних з більш ніж 18000 аудіо-зразків.

Результати оцінки ефективності розробленої методики показують, що з її допомогою можна суттєво покращити точність класифікації у порівнянні з існуючими підходами, забезпечивши 98%+ точності класифікації автомобільних аварій.

Запропонований підхід показує вищу продуктивність порівняно з моделями, які використовують необроблений аудіо-сигнал, спектрограми звукових доріжок, а також методи, що базуються на даних вбудованих акселерометрів транспортного засобу, на наборі даних MIVIA Audio Road Events.

Архітектура ЗНМ, запропонована у цьому дослідженні, має приблизно 2,3 млн параметрів, що значно менше за більшість існуючих моделей (~4–6 млн) і не потребує жодного модуля обробки сигналу для класифікації звуку, що робить модель придатною для використання у мобільних додатках розпізнавання звуку та вбудованих автомобільних системах.

Резюмуючи, основними внесками цієї роботи є:

запропоновано методологію, яка забезпечує вищу точність класифікації аудіо-сигналів аварій, ніж існуючі на сьогодні моделі;

запропоновано методологію (архітектура ЗНМ, метод підготовки навчального набору даних та його попередньої обробки, метод навчання ЗНМ), яка забезпечує високу точність класифікації аварій;

підтверджено ефективність розробленої методології на загальнодоступному наборі даних MIVIA Audio Road Events: покращено точність та енерго-ефективність класифікації у порівнянні з іншими відомими методиками.

Література

1. Convolutional neural network, part 1: structure, topology, activation functions and training set [Electronic resource] // Habr. – Mode of access: <https://habr.com/ru/post/348000/> (date of access: 01.10.2025). – Title from screen.
2. Lyashchinsky P. Automated synthesis of convolutional neural network structures / P. Lyashchinsky //

Problèmes et perspectives d'introduction de la recherche scientifique innovante. – 2019.

3. Dashkevych A. Research of multilayer neural networks for automatic selection of features for pattern recognition / A. Dashkevych // Scientific Bulletin of TDATU. – 2016. – №6 (2). P. 21–37.

4. Gencoglu O. Recognition of acoustic events using deep neural networks [Electronic resource] / O. Gencoglu, T. Virtanen, H. Huttunen. – Tampere University of Technology, Finland: 2014. – Mode of access: https://homepages.tuni.fi/tuomas.virtanen/papers/dnn_eusipco2014.pdf. (date of access: 01.10.2025). – Title from screen.

5. Abdoli S. End-to-end environmental sound classification using a 1D convolutional neural network [Electronic resource] / S. Abdoli, P. Cardinal, A. Koerich. – Montreal, Canada: 2019. – Mode of access: <http://dx.doi.org/10.1016/j.eswa.2019.06.040>.

6. Paciorek M. Effective Car Collision Detection with Mobile Phone Only [Electronic resource] / M. Paciorek, A. Klusek, P. Wawryka. – AGH University of Science and Technology, Krakov, Poland: 2021. – Mode of access: http://dx.doi.org/10.1007/978-3-030-77980-1_24.

7. Sammarco M. Car Accident Detection and Reconstruction Through Sound Analysis with Crashzam [Electronic resource] / M. Sammarco, M. Detyniecki. – Polish Academy of Sciences, Poland: 2020. – Mode of access: http://dx.doi.org/10.1007/978-3-030-26633-2_8.

8. Kutkovetskyi V. Image recognition: a tutorial / V. Kutkovetskyi. – Mykolaiv: Publishing House of CHNU named after P. Mogyla, 2017.

9. Podda A. CARgram: CNN-based accident recognition from road sounds through intensity-projected spectrogram analysis [Electronic resource] / A. Podda, R. Balia, L. Pompianu, S. Carta, G. Fenu, R. Saia // Digital Signal Processing. – 2024. – Vol. 147. – Mode of access: <https://doi.org/10.1016/j.dsp.2024.104431>.

10. World Health Organization. Road traffic injuries report [Electronic resource] / WHO. – 2021. – Mode of access: https://www.who.int/health-topics/road-safety/children-and-young-people#tab=tab_1 (date of access: 13.10.2025). – Title from screen.

11. Salamon J. A Dataset and Taxonomy for Urban Sound Research [Electronic resource] / J. Salamon, C. Jacoby, J. Bello. – 2014. – Mode of access: <http://dx.doi.org/10.1145/2647868.2655045> (date of access: 13.10.2025). – Title from screen.

12. Champion H., Automatic crash notification and the urgency algorithm: its history, value, and use [Electronic resource] / H. Champion // Adv. Emerg. Nursing J. – 2004. – 26 (2). – P. 143–156. – Mode of access: <https://www.careforcrashvictims.com/wp-content/uploads/2021/07/URGENCYAlgorithm.pdf> (date of access: 15.10.2025). – Title from screen.

13. Haria S. Car crash prevention and detection system using sensors and smart poles [Electronic resource] / S. Haria Anchaliya, V. Gala, T. Maru // Second International Conference on Intelligent Computing and Control Systems (ICICCS), IEEE. – 2018. – P. 800–804. – Mode of access: <http://dx.doi.org/10.1109/ICCONS.2018.8663017>.

14. Huang M., Environmental sound classification framework based on L-mHP features & SE-ResNet50 network [Electronic resource] / M. Huang, M. Wang, X. Liu, R. Kan, H. Qiu // Symmetry. – 2023. – 15 (5). – P. 1045. – Mode of access: <https://doi.org/10.3390/sym15051045>.

15. Rovetta S., Detection of hazardous road events from audio streams: an ensemble outlier detection approach [Electronic resource] / S. Rovetta, Z. Mnasri, F. Masulli // IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS). – 2020. – P. 1–6. – Mode of access: <https://doi.org/10.1109/EAIS48028.2020.9122704>.

16. Arslan Y., Canbolat H. Performance of deep neural networks in audio surveillance [Electronic resource] / Y. Arslan, H. Canbolat // 6th International Conference on Control Engineering & IT (CEIT), IEEE. – 2018. – P. 1–5. – Mode of access: <https://doi.org/10.1109/CEIT.2018.8751822>.

17. Zhang X., Acoustic traffic event detection in long tunnels using fast binary spectral features [Electronic resource] / X. Zhang, Y. Chen, M. Liu, C. Huang // Circuits Syst. Signal Process. – 2020. – 39. – P. 2994–3006. – Mode of access: <https://doi.org/10.1007/s00034-019-01294-9> (date of access: 27.10.2025). – Title from screen.

18. Kumar B., Road accident detection using machine learning [Electronic resource] / B. Kumar, A. Basit, M. Kiruba, R. Giridharan, S. Keerthana // International Conference on System, Computation, Automation and Networking (ICSCAN), IEEE. – 2021. – P. 1–5. – Mode of access: <https://doi.org/10.1109/ICSCAN53069.2021.9526546>.

19. Foggia P., Detecting sounds of interest in roads with deep networks [Electronic resource] / P. Foggia, A. Saggese, N. Strisciuglio, M. Vento, V. Vigilante // International Conference on Image Analysis and Processing. – Springer. – 2019. – C. 583–592. – Mode of access: https://doi.org/10.1007/978-3-030-30645-8_53.

20. Foggia P., Reliable detection of audio events in highly noisy environments [Electronic resource] / P. Foggia, M. Vento, N. Petkov, A. Saggese, N. Strisciuglio // Pattern Recognit. Lett. – 2015. – 65. – P. 22–28. – Mode of access: <http://dx.doi.org/10.1016/j.patrec.2015.06.026>.

21. Abdoli S. End-to-end environmental sound classification using a 1D convolutional neural network [Electronic resource] / S. Abdoli, P. Cardinal, A. Koerich // Expert Syst. Appl. – 2019. – 136. – C. 252–263. – Mode of access: <https://doi.org/10.1016/j.eswa.2019.06.040>.

22. Francel A. Deep neural network models of sound localization reveal how perception is adapted to real-world environments [Electronic resource] / A. Francel, J. McDermott // Nature Human Behaviour. – 2022. – Vol. 6. – P. 111–133. – Mode of access: <https://doi.org/10.1038/s41562-021-01244-z>.

23. Lin T., Focal Loss for Dense Object Detection [Electronic resource] / T. Lin, P. Goyal, R. Girshick, K. He, P. Dollar. – FAIR: 2018. – Mode of access: <https://arxiv.org/pdf/1708.02002> (date of access: 29.10.2025). – Title from screen.

References

1. Convolutional neural network, part 1: structure, topology, activation functions and training set [Electronic resource] // Habr. – Mode of access: <https://habr.com/ru/post/348000/> (date of access: 01.10.2025). – Title from screen.
2. Lyashchinsky P. Automated synthesis of convolutional neural network structures / P. Lyashchinsky // Problèmes et perspectives d'introduction de la recherche scientifique innovante. – 2019.
3. Dashkevych A. Research of multilayer neural networks for automatic selection of features for pattern recognition / A. Dashkevych // Scientific Bulletin of TDATU. – 2016. – №6 (2). P. 21–37.
4. Gencoglu O. Recognition of acoustic events using deep neural networks [Electronic resource] / O. Gencoglu, T. Virtanen, H. Huttunen. – Tampere University of Technology, Finland: 2014. – Mode of access: https://homepages.tuni.fi/tuomas.virtanen/papers/dnn_eusipco2014.pdf. (date of access: 01.10.2025). – Title from screen.
5. Abdoli S. End-to-end environmental sound classification using a 1D convolutional neural network [Electronic resource] / S. Abdoli, P. Cardinal, A. Koerich. – Montreal, Canada: 2019. – Mode of access: <http://dx.doi.org/10.1016/j.eswa.2019.06.040>.
6. Paciorek M. Effective Car Collision Detection with Mobile Phone Only [Electronic resource] / M. Paciorek, A. Klusek, P. Wawryka. – AGH University of Science and Technology, Krakov, Poland: 2021. – Mode of access: http://dx.doi.org/10.1007/978-3-030-77980-1_24.
7. Sammarco M. Car Accident Detection and Reconstruction Through Sound Analysis with Crashzam [Electronic resource] / M. Sammarco, M. Detyniecki. – Polish Academy of Sciences, Poland: 2020. – Mode of access: http://dx.doi.org/10.1007/978-3-030-26633-2_8.
8. Kutkovetskiy V. Image recognition: a tutorial / V. Kutkovetskiy. – Mykolaiv: Publishing House of CHNU named after P. Mogyla, 2017.
9. Podda A. CARgram: CNN-based accident recognition from road sounds through intensity-projected spectrogram analysis [Electronic resource] / A. Podda, R. Balia, L. Pompianu, S. Carta, G. Fenu, R. Saia // Digital Signal Processing. – 2024. – Vol. 147. – Mode of access: <https://doi.org/10.1016/j.dsp.2024.104431>.
10. World Health Organization. Road traffic injuries report [Electronic resource] / WHO. – 2021. – Mode of access: https://www.who.int/health-topics/road-safety/children-and-young-people#tab=tab_1 (date of access: 13.10.2025). – Title from screen.
11. Salamon J. A Dataset and Taxonomy for Urban Sound Research [Electronic resource] / J. Salamon, C. Jacoby, J. Bello. – 2014. – Mode of access: <http://dx.doi.org/10.1145/2647868.2655045> (date of access: 13.10.2025). – Title from screen.
12. Champion H. Automatic crash notification and the urgency algorithm: its history, value, and use [Electronic resource] / H. Champion // Adv. Emerg. Nursing J. – 2004. – 26 (2). – P. 143–156. – Mode of access: <https://www.careforcrashvictims.com/wp-content/uploads/2021/07/URGENCYAlgorithm.pdf> (date of access: 15.10.2025). – Title from screen.
13. Haria S. Car crash prevention and detection system using sensors and smart poles [Electronic resource] / S. Haria Anchaliya, V. Gala, T. Maru // Second International Conference on Intelligent Computing and Control Systems (ICICCS), IEEE. – 2018. – P. 800–804. – Mode of access: <http://dx.doi.org/10.1109/ICICCS.2018.8663017>.
14. Huang M., Environmental sound classification framework based on L-mHP features & SE-ResNet50 network [Electronic resource] / M. Huang, M. Wang, X. Liu, R. Kan, H. Qiu // Symmetry. – 2023. – 15 (5). – P. 1045. – Mode of access: <https://doi.org/10.3390/sym15051045>.
15. Rovetta S., Detection of hazardous road events from audio streams: an ensemble outlier detection approach [Electronic resource] / S. Rovetta, Z. Mnasri, F. Masulli // IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS). – 2020. – P. 1–6. – Mode of access: <https://doi.org/10.1109/EAIS48028.2020.9122704>.
16. Arslan Y., Canbolat H. Performance of deep neural networks in audio surveillance [Electronic resource] / Y. Arslan, H. Canbolat // 6th International Conference on Control Engineering & IT (CEIT), IEEE. – 2018. – P. 1–5. – Mode of access: <https://doi.org/10.1109/CEIT.2018.8751822>.
17. Zhang X., Acoustic traffic event detection in long tunnels using fast binary spectral features [Electronic resource] / X. Zhang, Y. Chen, M. Liu, C. Huang // Circuits Syst. Signal Process. – 2020. – 39. – P. 2994–3006. – Mode of access: <https://doi.org/10.1007/s00034-019-01294-9> (date of access: 27.10.2025). – Title from screen.
18. Kumar B., Road accident detection using machine learning [Electronic resource] / B. Kumar, A. Basit, M. Kiruba, R. Giridharan, S. Keerthana // International Conference on System, Computation, Automation and Networking (ICSCAN), IEEE. – 2021. – P. 1–5. – Mode of access: <https://doi.org/10.1109/ICSCAN53069.2021.9526546>.
19. Foggia P., Detecting sounds of interest in roads with deep networks [Electronic resource] / P. Foggia, A. Saggese, N. Strisciuglio, M. Vento, V. Vigilante // International Conference on Image Analysis and Processing. – Springer. – 2019. – C. 583–592. – Mode of access: https://doi.org/10.1007/978-3-030-30645-8_53.
20. Foggia P., Reliable detection of audio events in highly noisy environments [Electronic resource] / P. Foggia, M. Vento, N. Petkov, A. Saggese, N. Strisciuglio // Pattern Recognit. Lett. – 2015. – 65. – P. 22–28. – Mode of access: <http://dx.doi.org/10.1016/j.patrec.2015.06.026>.
21. Abdoli S. End-to-end environmental sound classification using a 1D convolutional neural network [Electronic resource] / S. Abdoli, P. Cardinal, A. Koerich // Expert Syst. Appl. – 2019. – 136. – C. 252–263. – Mode of access: <https://doi.org/10.1016/j.eswa.2019.06.040>.
22. Francel A. Deep neural network models of sound localization reveal how perception is adapted to real-world environments [Electronic resource] / A. Francel, J. McDermott // Nature Human Behaviour. – 2022. – Vol. 6. – P. 111–133. – Mode of access: <https://doi.org/10.1038/s41562-021-01244-z>.
23. Lin T., Focal Loss for Dense Object Detection [Electronic resource] / T. Lin, P. Goyal, R. Girshick, K. He, P. Dollar. – FAIR: 2018. – Mode of access: <https://arxiv.org/pdf/1708.02002> (date of access: 29.10.2025). – Title from screen.