

**Могилевич Д.І.**

Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»

**Хміль Р.В.**

Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»

## АНАЛІЗ ЗАСТОСУВАННЯ НЕЙРОННИХ МЕРЕЖ ДЛЯ РОЗПІЗНАВАННЯ АВТОМОБІЛЬНИХ АВАРІЙ

*Система автоматичного екстреного виклику e-Call стала обов'язковою в Європейському Союзі з 2018 року. Ця система автоматично викликає служби екстреної допомоги в разі аварії, але існує проблема з хибними викликами, які призводять до невиправданих витрат ресурсів аварійних служб. Більшість виробників авто вирішує первинну задачу визначення аварійної ситуації за допомогою аналізу інформації, що надходить від вбудованих датчиків-акселерометрів. В результаті, досить часто різке гальмування, яке дозволяє уникнути аварії, помилково ідентифікується як аварійна ситуація і призводить до хибного виклику аварійних служб.*

*В статті запропоновано підхід до розпізнавання автомобільних аварій на основі використання згорткових нейронних мереж для класифікації звукових доріжок, що записуються всередині автомобіля при виникненні аварійних ситуацій на дорогах.*

*Мета дослідження полягає у проведенні аналізу існуючого математичного апарату згорткових нейронних мереж та розробці підходу до вирішення задачі розпізнавання та верифікації автомобільних аварій на основі їх використання. Основними етапами дослідження є аналіз методів обробки даних, вдосконалення алгоритмів навчання та використання передових технологій у галузі машинного навчання.*

*Проведено аналіз архітектури нейронних мереж для ефективного вирішення поставленої задачі з використанням мобільних пристроїв, що включає дослідження різних конфігурацій нейронних мереж, їхню адаптацію до умов обмежених обчислювальних ресурсів мобільних пристроїв та вибір моделі з мінімальним використанням ресурсів при забезпеченні високої точності розпізнавання аварій.*

**Ключові слова:** штучний інтелект, нейронні мережі, обробка звукових сигналів.

**Постановка проблеми.** Актуальність пошуку інноваційних підходів до вирішення задачі розпізнавання та класифікації автомобільних аварій зумовлена економічною доцільністю підвищення точності існуючих методів. Кожен додатковий відсоток точності дасть можливість зекономити мільярди доларів прямих витрат на хибні виклики, що надходять до аварійних служб. Тому актуальним є підхід щодо побудови системи розпізнавання та верифікації автомобільних аварій на основі використання математичного апарату загорткових нейронних мереж (ЗНМ), яка може бути інтегрована у мобільні пристрої для швидкого та точного реагування на небезпечні ситуації на дорозі.

**Аналіз останніх досліджень і публікацій.** Проблематика розпізнавання образів та сигналів на основі нейронних мереж досліджена у наукових працях багатьох вітчизняних та закордонних вчених.

Лебідь О., Кіпоренко С., Вовк В. здійснили аналіз методів машинного навчання та нейромережевих технологій, які використовуються для

виявлення аномалій. Вченими доведено доцільність використання нейромережевих технологій у системах інформаційної безпеки та розроблено метод вирішення цього завдання на основі архітектурних нейромереж [1].

Мясіщев О., Ленков Є., Білик О. глибинно вивчили алгоритми та інструменти розпізнавання об'єктів з використанням нейронних мереж. Корисним для нашого дослідження є класифікація штучних мереж, з використанням яких доцільно проводити діагностику об'єктів на основі формування образу-еталону [2].

Мірошник О. запропонував для розпізнавання вейвлет-образів несинусоїдальних сигналів на базі адаптивної резонансної теорії використовувати нейронну мережу. Вчений розробив програмний продукт, в якому закладено алгоритм роботи нейронної мережі та описані його основні функціональні можливості [3].

Мірошник О., Святобатько А. запропонували спосіб прогнозування фізичних параметрів з використанням нейронних мереж. Авторами наведені

математичні вирази для наглядного розуміння структури нейронних мереж та їх роботи [4].

О. Генчоглу (O. Gencoglu), Т. Віртанен (T. Virtanen) та Х. Хуттунен (H. Hattunen) дослідили проблеми класифікації аудіо сигналів та розпізнавання акустичних подій за допомогою глибинних нейромереж [5].

С. Абдолі (S. Abdoli), П. Кардинал (P. Cardinal), А. Керіх (A. Koerich) описують підходи до класифікації звуків з використанням згортових нейронних мереж прямого поширення, які навчаються на одновимірних зразках звукових подій [6].

М. Пачорек (M. Paciorek), А. Клусек (A. Klusek), П. Ваврика (P. Wawryka) аналізують існуючі підходи до розпізнавання автомобільних аварій за допомогою виключно мобільних телефонів та даних їх акселерометрів та гіроскопів, використовуючи 53 набори даних, отриманих за допомогою різних телефонів під час реальних (тестових) аварій [7].

Машинному навчанню та різним підходам до розпізнавання образів присвячені видання Кутковецького В.Я. [8].

Що таке глибинне навчання, глибинні нейронні мережі як засоби класифікації об'єктів та особливості застосування згорткових нейронних мереж для розпізнавання зображень детально викладено в працях Субботіна С.О. [9] та Новотарського М.А. [10].

Зазначені наукові дослідження демонструють підвищену увагу вчених до проблематики методів розпізнавання сигналів та образів на основі нейронних мереж.

**Постановка завдання.** Метою роботи є проведення аналізу існуючого математичного апарату згорткових нейронних мереж та розробці підходу до вирішення задачі розпізнавання та верифікації автомобільних аварій на основі їх використання.

**Виклад основного матеріалу.** Для розпізнавання сигналів та образів найбільш популярними та ефективними є згорткові нейронні мережі [1; 2]. Їх можна застосувати для обробки будь-якого сигналу, чи то даних з датчиків, аудіосигналів, образів тощо.

Цей вид нейромереж є багато-шаровим перцептроном, що складається з багатьох рівнів вузлів, прихованих і вихідного шарів, та має односпрямований інформаційний потік. Функцією активації для вузлів прихованого шару зазвичай обирається монотонна нелінійна  $S$ -подібна функція, в той час як для вузлів вихідного шару достатньо використовувати лінійну функцію.

Зазначений вид нейронної мережі можна апроксимувати до будь-якого неперервного відоб-

раження, якщо кількість прихованих шарів достатньо велика. В умовах розпізнавання образів цей вид нейронних мереж з нелінійною  $S$ -подібною функцією та декількома шарами може впізнавати об'єкти з досить високою точністю. Ці характеристики багатошарової нейронної мережі прямого поширення закладають теоретичну основу для застосування багатошарових перцептронів для процесу моделювання та діагностики похибок з розпізнавання образів. Похибки визначаються двома способами – програмують модель процесу виправлення розпізнавання образів або обирають класифікатор шаблонів.

Якість розпізнавання образів нейронними мережами залежить від ефективності проведеного навчання на прикладі певної вибірки даних з використанням значної кількості навчальних пар (вхід-вихід). За результатами навчання нейронної мережі визначається функція помилки або відхилення (функція втрат). Процес навчання нейронної мережі направлений на мінімізацію помилки, що дозволяє штучному інтелекту самостійно корегувати показники допустимих ваг зв'язків між нейронами.

Загальний вигляд багатошарового перцептрону наведено на рис. 1.

Ми ж будемо використовувати архітектуру згорткової нейронної мережі, що наведена на рис. 2. Ключовою особливістю і відмінністю згорткової нейромережі від перцептрона є те, що нейрони шару мають не індивідуальні вагові коефіцієнти, а використовують розділені ваги: матриці ваг невеликого розміру, які ще називають ядрами згортки. Таким чином, згорткова мережа має значно меншу кількість параметрів у порівнянні з повнозв'язною мережею, з чого слідує її більш висока продуктивність і економічність у використанні пам'яті.

Згорткова мережа має багатошарову структуру. Приховані шари зазвичай складаються зі згорткових, агрегувальних, повноз'язних шарів та шарів нормалізації.

Після вхідного шару сигнал проходить декілька шарів згортки де на кожному шарі виконується послідовне чергування згортки і пулінгу. Чередування шарів дозволяє скласти «карти ознак» з яких, на кожному з наступних шарів карта зменшуватиметься в розмірах, але збільшуватиметься кількість каналів. В практичному застосуванні це буде означати можливість розпізнавання складних підлеглих ознак.

Після згорткових шарів додатково додають перцептрон, на вхід котрого будуть подаватися

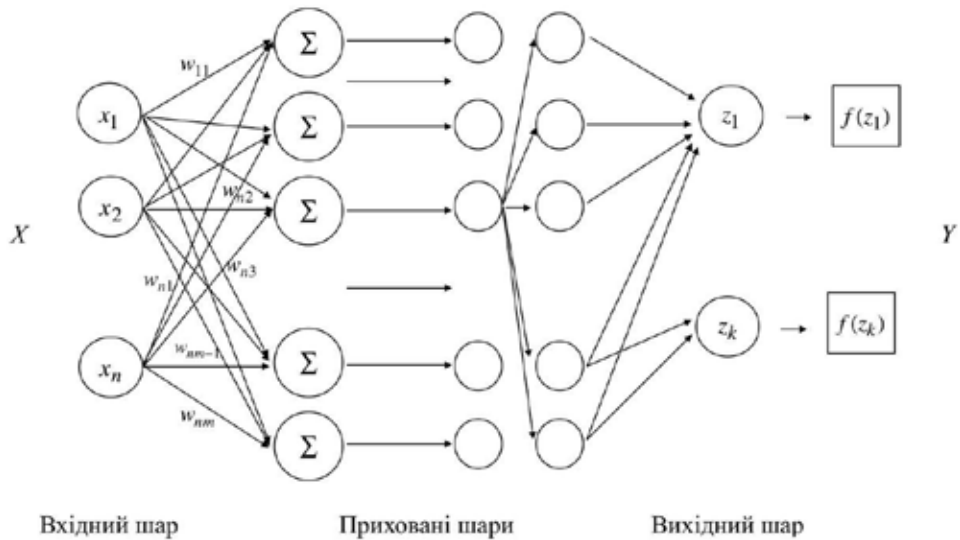


Рис. 1. Структура багатшарового перцептрон

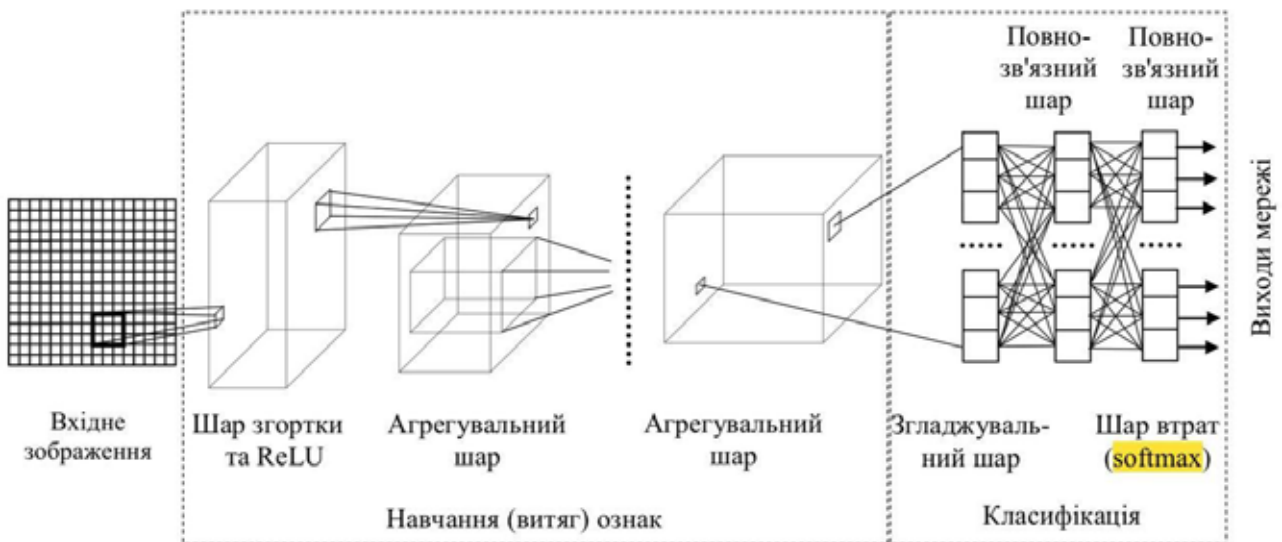


Рис. 2. Структура згорткової нейронної мережі

кінцеві карти ознак. Перші два типи шарів, згортковий та субдискретизований, чергуючись між собою, формують вхідний вектор ознак для багатшарового перцептрон.

Повноз'єднані нейронні мережі прямого поширення можна застосовувати як для навчання ознак, так і для класифікації даних, але застосування цієї архітектури до зображень є непрактичним. Наприклад, повнозв'язний шар для невеликого зображення розміром  $100 \times 100$  має 10 000 ваг. Операція згортки дає змогу розв'язати цю проблему, оскільки вона зменшує кількість вільних параметрів, дозволяючи мережі бути глибшою за меншої кількості параметрів. Наприклад, незалежно від розміру зображення, області розміру  $3 \times 3$ , кожна з одними й тими ж спільними

вагами, вимагають лише 9 вільних параметрів. Таким чином, це розв'язує проблему зникання або «вибуху» градієнтів у тренуванні традиційних багатшарових нейронних мереж з багатьма шарами за допомогою зворотного поширення.

Згортковий шар являє собою набір карт (інша назва – карти ознак або матриці), у кожній карти є синаптичне ядро (в різних джерелах його називають по-різному: скануюче ядро або фільтр). Кількість карт визначається вимогами до задачі, якщо взяти велику кількість карт, то підвищиться якість розпізнавання, але збільшиться обчислювальна складність. Виходячи з аналізу наукових статей, в більшості випадків пропонується брати співвідношення один до двох, тобто кожна карта попереднього шару (наприклад, у першого

згорткового шару, попереднім є вхідний) пов'язана з двома картами згорткового шару [11].

Розміри у всіх карт згорткового шару однакові і обчислюються за формулою:  $(x,y)=(mW-kW+1, mH-kH+1)$ , де  $(x,y)$  – обчислюваний розмір згорткової карти,  $mW$  – ширина попередньої карти,  $kW$  – ширина ядра,  $mH$  – висота попередньої карти,  $kH$  – висота ядра.

Ядро являє собою фільтр або вікно, яке «ковзає» по всій області карти і знаходить певні ознаки об'єктів. Розмір ядра зазвичай беруть в межах від  $3 \times 3$  до  $7 \times 7$ . Якщо розмір ядра маленький, то воно не зможе виділити будь-які ознаки, якщо занадто велике, то збільшується кількість зв'язків між нейронами. Також розмір ядра вибирається таким, щоб розмір карт згорткового шару був парним, це дозволяє не втрачати інформацію при зменшенні розмірності в підвибірному шарі.

Використання процедури «ковзного вікна» (sliding window), котра «нарізає» звукові доріжки, що надходять, на сегменти, котрі перекриваються на стиках (overlap), дозволяє обробляти звукові доріжки будь-якої довжини.

Ядро являє собою систему поділюваних ваг або синапсів, це одна з головних особливостей згорткової нейронної мережі. У звичайній багатшаровій мережі дуже багато зв'язків між нейронами, тобто синапсів, що вельми уповільнює процес детектування. У згортковій нейронній мережі – навпаки, загальні ваги дозволяє скоротити число зв'язків і дозволити знаходити одну і ту саму ознаку по всій області зображення [3]. Тут застосовується математична функція згортки, котра є особливим видом лінійної операції, в якій дві функції перемножуються, щоб отримати третю функцію, яка виражає, як форма однієї функції змінюється іншою. Простіше кажучи, два зображення, які можна представити у вигляді матриць, перемножуються, щоб отримати вихід, який використовується для вилучення ознак із зображення. Операція згортки обчислюється за формулою:

$$(fg)[m,n] = \sum_{kl} f[m-k, m-l]g[kl],$$

де  $f$  – вихідна матриця зображення,  $g$  – ядро згортки [11].

Отже, згортка – це операція обчислення нового значення обраного пікселя, що враховує значення оточуючих його пікселів. Для обчислення значення використовується матриця, що називається ядром згортки або матрицею згортки. Зазвичай ядро згортки є квадратною матрицею  $n \times n$ , де  $n$  – непарне число. Під час обчислення нового значення обраного пікселя ядро згортки «прикла-

дається» своїм центром до даного пікселя. Навколишні пікселі так само накриваються ядром. Далі вираховується сума, де складовими є помножені значення пікселів на значення комірки ядра, що накрила даний піксель. Сума ділиться на коефіцієнт нормування (div) або на суму всіх елементів ядра згортки. Отримане значення якраз і є новим значенням обраного пікселя. Якщо застосувати згортку до кожного пікселя зображення, то в результаті вийде певний ефект, що залежить від обраного ядра згортки [11].

Іншим важливим поняттям ЗНМ є агрегування (субдискретизація), яке є різновидом нелінійного зниження дискретизації. Існує декілька нелінійних функцій для реалізації агрегування, серед яких найпоширенішою є максимізаційне агрегування (max pooling). Воно розділяє вхідне зображення на набір прямокутників без перекриттів, і для кожної такої під-області виводить її максимум. Ідея полягає в тому, що точне положення ознаки не так важливе, як її грубе положення відносно інших ознак. Агрегувальний шар слугує поступовому скороченню просторового розміру представлення для зменшення кількості параметрів та об'єму обчислень у мережі, і відтак також для контролю перенавчання. В архітектурі ЗНМ є звичним періодично вставляти агрегувальний шар між послідовними згортковими шарами. Операція агрегування забезпечує ще один різновид інваріантності відносно паралельного перенесення.

Субдискретизований шар можна описати формулою  $x^l = f(a^l \text{subsample}(x^{l-1}) + b^l)$  де  $x^l$  – вихід шару  $l$ ,  $f()$  – функція активації,  $a^l, b^l$  – коефіцієнти зсуву шару  $l$ ,  $\text{subsample}()$  – операція вибірки локальних максимальних значень.

Агрегувальний шар діє незалежно на кожен зріз глибини входу, і зменшує його просторовий розмір. Найпоширенішим видом є агрегувальний шар із фільтрами розміру  $2 \times 2$ , що застосовуються з кроком 2, який знижує дискретизацію кожного зрізу глибини входу в 2 рази як за шириною, так і за висотою, відкидаючи 75 % збуджень. В цьому випадку кожна операція взяття максимуму діє над 4 числами. Розмір за глибиною залишається незмінним.

На додачу до максимізаційного агрегування, агрегувальні вузли можуть використовувати й інші функції, такі як усереднювальне агрегування (average pooling) та  $L^2$  – нормове агрегування. Історично усереднювальне агрегування застосовувалася часто, але останнім часом втратило популярність у порівнянні з дією максимізаційного агрегування, робота якого на практиці

виявилася кращою. Через агресивне скорочення розміру представлення, тенденція йде до менших фільтрів, або відмови від агрегувального шару взагалі [12].

Субдискретизований шар також, як і згортковий має карти. Мета субдискретизованого шару – зменшення розмірності карт попереднього шару. Якщо на попередній операції згортки вже були виявлені деякі ознаки, то для подальшої обробки настільки докладне зображення вже не потрібно, і воно ущільнюється до менш докладного.

У процесі сканування ядром субдискретизованого шару (фільтром) карти попереднього шару, скануюче ядро не перетинається на відміну від згорткового шару. Зазвичай, кожна карта має ядро розміром 2x2, що дозволяє зменшити попередні карти згорткового шару в 2 рази. Вся карта ознак поділяється на осередки 2x2 елементи, з яких вибираються максимальні за значенням.

Нейрони кожної карти попереднього субдискретизованого шару пов'язані з одним нейроном прихованого шару. Таким чином, число нейронів прихованого шару дорівнює числу карт субдискретизованого шару, але зв'язки можуть бути не обов'язково такими, наприклад, тільки частина нейронів будь-якої з карт субдискретизованого шару може бути пов'язана з першим нейроном прихованого шару, а частина, що залишилася з другим, або всі нейрони першої карти пов'язані з нейронами 1 і 2 прихованого шару. Обчислення значень нейрона можна описати формулою

$$x_j^l = f \left( \sum_i x_i^{l-1} w_{i,j}^{l-1} + b_j^{l-1} \right),$$

де  $x_j^l$  – карта ознак  $j$  (вихід шару  $l$ ),  $f()$  – функція активації,  $b^l$  – коефіцієнт зсуву (bias) шару  $l$ ,  $w_{i,j}^{l-1}$  – матриця вагових коефіцієнтів шару  $l$  [12].

Зазвичай, в субдискретизованому шарі застосовується функція активації *ReLU* (від англ. *Rectified Linear Units*, шар зрізаних лінійних вузлів). Цей шар застосовує ненасичувальну функцію активації  $f(x) = \max(x, 0)$ . Він посилює нелінійні властивості функції ухвалення рішення і мережі в цілому, не зачіпаючи рецептивних полів згорткового шару. Зрізаному лінійному вузлові *ReLU* часто віддають перевагу перед іншими функціями, оскільки він тренує нейронну мережу в декілька разів швидше без значної розплати точністю узагальнення [12].

Використаємо функцію активації *ReLU* для всіх шарів нейронної мережі, за виключенням вихідного шару, для якого застосуємо функцію активації *softmax* (нормовану експоненціальну

функцію). Функція *softmax* часто використовується в останньому шарі класифікаторів на основі нейронних мереж. Такі мережі зазвичай навчаються за допомогою перехресної ентропії, що на виході дає нелінійний варіант поліноміальної логістичної регресії.

Дана функція «стискає»  $K$ -вимірний вектор  $Z$  із довільним значенням компонент до  $K$ -вимірного вектору  $\sigma(Z)$  з дійсними значеннями компонент в області  $[0, 1]$ , що в сумі дають одиницю, та задається наступним чином:

$$\sigma(Z)_i = \frac{e^{z_i}}{\sum_{k=1}^K e^{z_k}}, j = 1, \dots, K.$$

Останній з типів шарів – це шар звичайного багатошарового перцептрона, або, як його ще називають – повнозв'язний шар. Мета шару – класифікація. Повнозв'язний шар моделює складну нелінійну функцію, оптимізуючи яку, покращується якість розпізнавання.

Функція втрат – це функція, яка характеризує втрати при неправильному прийнятті рішень на основі спостережених даних. Тобто це метод оцінки того, наскільки добре алгоритм моделює вказаний набір даних, наскільки гарно алгоритм працює з заданим набором. Метою функції втрат в нейронній мережі є оцінка та оновлення ваг нейронів з метою поліпшення оцінки на наступному кроці.

У якості функції втрат  $F$  можна використовувати середню абсолютну похибку, середньо-квадратичну логарифмічну похибку, завісні втрати (hinge loss), перехресну ентропію (cross entropy), softmax та cosine.

Якщо  $y_i^{pred}$  – передбачуваний клас,  $y_i^{true}$  – справжній клас,  $n$  – кількість зразків для навчання, то дані функції втрат обчислюються наступним чином:

середня абсолютна похибка – сума абсолютних різниць між цільовими значеннями та прогнозованими змінними:

$$F = \| y^{true} - y^{pred} \|_1 = \frac{\sum_{i=1}^n |y_i^{true} - y_i^{pred}|}{n};$$

середньо-квадратична похибка – сума квадратів відстаней між цільовими значеннями та прогнозованими змінними:

$$F = \| y^{true} - y^{pred} \|_2^2 = \frac{\sum_{i=1}^n (y_i^{true} - y_i^{pred})^2}{n};$$

середньо-квадратична логарифмічна похибка:

$$F = \frac{\sum_{i=1}^n \log \left( \frac{y_i^{pred} + 1}{y_i^{true} + 1} \right)^2}{n};$$

завісні втрати (*hinge loss*) – функція втрат, яка використовується для максимізації розділової класифікації і має таке представлення:

$$F = \sum_{i=1}^n \max \left( 0, \frac{1}{2} - y_i^{true} y_i^{pred} \right),$$

де  $y_i^{true}$  приймає значення 0 або 1; перехресна ентропія (*cross entropy*) має такий вигляд:

$$F(P, Q) = - \sum_x P(x) \log Q(x),$$

де  $P(x)$  – розподіл правильних відповідей, а  $Q(x)$  – розподіл ймовірностей прогнозів моделі.

У випадку бінарної класифікації функція втрат перехресної ентропії буде мати наступний вигляд:

$$F_p(Q) = - \frac{1}{n} \sum_{i=1}^n y_i \log(p(y_i)) + (1 - y_i) \log((1 - p(y_i))).$$

У цьому випадку для точки  $y=1$  додаємо  $\log(p(y))$  (логарифмічну ймовірність того, що  $y=1$ ) до втрат, а у випадку коли  $y=0$  додаємо  $\log(1 - p(y))$ . Ця функція сприяє наближенню розподілу прогнозування до цільового, штрафуючи не лише за помилкові прогнози, але і за невпевнені.

У випадку категоріальної класифікації функція втрат перехресної ентропії буде мати наступний вигляд:

$$F = - \sum_i y_i^{true} \log(p(y_i^{pred})),$$

де  $p$  – оцінка імовірності.

Категоріальна перехресна ентропія визначає міру того, наскільки два дискретних розподілу ймовірностей відрізняються один від одного. Знак мінуса гарантує зменшення втрат, коли розподіли стають ближчими один до одного.

Функція втрат *softmax* – це нормовані експоненційні втрати, які обчислюються як сума значення функції активації *softmax* та значення функції втрат перехресної ентропії:

$$F = - \frac{1}{N} \sum_{i=1}^N \left( \sum_{k=1}^K y_i \log \left( \frac{e^{w_k y_i + b_k}}{\sum_{k=1}^K e^{w_k y_i + b_k}} \right) \right).$$

Дана функція застосовуються для передбачення єдиного класу з  $S$  взаємно виключних класів.

Функція втрат *cosine* має наступне представлення:

$$F(x, y) = 1 - \sigma_{\cos}(f(\theta), \phi(y)).$$

Косинусова подібність двох векторів  $x$  і  $y$  базується на куті між цими двома векторами:

$$\sigma_{\cos}(x, y) = \cos(x \leq y) = \frac{\langle x, y \rangle}{\|x\|_2 \times \|y\|_2},$$

де  $\langle \cdot \rangle$  – добуток векторів,  $\| \cdot \|_p$  –  $L^p$  норма.

$$\begin{aligned} F(x, y) &= 1 - \sigma_{\cos}(f(\theta), \phi(y)) = \\ &= 1 - \frac{\sum_{i=1}^n (y_i^{true} \times y_i^{pred})}{\sqrt{\sum_{i=1}^n (y_i^{true})^2} \times \sqrt{\sum_{i=1}^n (y_i^{pred})^2}} \end{aligned}$$

Чим вище значення косинусової подібності, тим вища значення точності моделі. Цілком протилежний вектор має значення косинусової подібності  $-1$ , цілком ортогональний вектор має значення косинусової подібності  $0$ , а повністю ідентичні вектори мають значення  $1$ .

Проведений аналіз показав, що в наш час існує багато архітектур згорткових нейронних мереж. До ЗНМ, що широко використовуються можна віднести такі:

*LeNet*. Перше успішне застосування згорткової нейронної мережі вдалося розробити Яну Лекуну в 1990-і роки. Архітектура *LeNet* застосовувалася для зчитування поштових індексів, цифр і т.д.

*AlexNet*. Розробники *AlexNet* Алекса Крижевський, Ілля Суцкевер і Джефф Хинтон. Ця розробка зіграла істотну роль в популяризації ЗНМ в області комп'ютерного зору. Архітектура *AlexNet* була представлена на *ImageNet ILSVRC Challenge* в 2012 році і зайняла перше місце (16% помилок проти 26% у архітектури, яка зайняла друге місце).

*ZF Net*. Розробники Метью Зеллера і Роба Фергюса. Архітектура *ZF Net* перемогла у *ILSVRC 2013*. Дана архітектура є удосконаленою версією *AlexNex* (збільшено розміри середніх згорткових шарів і зменшено крок і розмір фільтра на першому шарі).

*GoogLeNet*. У 2014 році конкурс *ILSVRC 2014* виграла ЗНМ розробки Шегеда і інших – співробітників корпорації *Google*. Основною особливістю даної архітектури є розробка та впровадженні вхідного модуля (*Inception Module*), що дозволило різко скоротити число параметрів до 4 млн з 60 млн. Скорочення параметрів відбувається також завдяки заміні повнозв'язних шарів у верхній частині мережі шарами середнього пулінгу.

*VGGNet*. Розробники Карен Симонян і Ендрю Ціссерман. Архітектуру *VGGNet* представлено на *ILSVRC 2014*. Розробникам вдалося наочно продемонструвати, що глибина є ключовим фактором

для продуктивності. Їх мережа містить 16 згорткових і повнозв'язних шарів і має надзвичайно однорідну архітектуру, яка виконує згортку  $3 \times 3$  і пулінг  $2 \times 2$  від початку до кінця. Вихідна модель доступна в режимі *Plug and Play* у фреймворку для глибокого навчання *Caffe*. Недоліком *VGGNet* є те, що потрібно оцінювати і використовувати набагато більше пам'яті і параметрів (140M). Більшість цих параметрів знаходяться в першому повнозв'язному шарі, і з тих пір було виявлено, що ці *FC*-шари можуть бути видалені без зниження продуктивності, значно зменшуючи кількість необхідних параметрів.

*ResNet*. Розробники Каймінгом Хе та ін. (команда дослідників з *Microsoft Research*). Особливістю цієї нейронної мережі є використання «залишкових з'єднань» або “*skip connections*”, які дозволяють передавати інформацію з вхідного шару безпосередньо до вихідного шару. В кінці архітектури відсутні повнозв'язні шари. Сама архітектура *ResNet* складається з блоків з функцією активації *ReLU* та «залишковими з'єднаннями», які дозволяють нейронній мережі навчатися ефективно навіть при значній глибині. *ResNet* став популярним архітектурним варіантом для різноманітних завдань комп'ютерного зору та інших областей машинного навчання.

Загалом ідея використання нейронних мереж для вирішення задачі класифікації звуків автомобільних аварій полягає в наступному:

- за допомогою додатку для смартфона, та наявності відповідної згоди водія, вести постійний запис звукової доріжки того, що відбувається в автомобілі, у циклічному режимі. Постійно записуються лише останні 5 секунд, яких у випадку аварії достатньо для аналізу;

- альтернативно, виробники автомобілів можуть забезпечити запис звукової доріжки в авто за допомогою вбудованих мікрофонів, а подальші алгоритмічні обчислення проводити за допомогою процесорів, які наявні в усіх нових автомобілях;

- у випадку надходження від акселерометрів, вбудованих у смартфон/авто, даних про екстрене гальмування, останні 5 секунд записаної звукової доріжки передаються на аналіз згорткової нейронній мережі для бінарної класифікації події як «аварія/не аварія»;

- звуковий трек ділиться на сегменти, які передаються для аналізу на вхід нейронної мережі, котра попередньо навчалась на відповідних зразках інших автомобільних аварій.

З метою суттєвого зменшення навантаження на мобільні пристрої, робота нейронної мережі

може бути обмежена випадками, коли аварійна ситуація попередньо ідентифікована за допомогою алгоритмів, що базуються на даних акселерометрів та гіроскопів, встановлених в автомобілі або мобільному телефоні, що використовується як джерело сигналів. В такому випадку результати попереднього аналізу слугуватимуть тригером для запуску роботи ЗНМ, яка в такому випадку виконує додаткову верифікацію такої первинної класифікації. Для цього можна скористатись методиками, наведеними в роботі [10].

Для того ж, щоб точність розпізнавання аварій нейронною мережею була вищою, ніж алгоритмів, що базуються на інформації від акселерометрів, варто повністю замінити аналіз даних від акселерометрів на постійний аналіз звукової доріжки за допомогою ЗНМ в режимі реального часу, використовуючи дані від акселерометрів та гіроскопів лише як додаткову інформацію для аналізу, а не як тригер. Хоч це і призведе до постійного навантаження на процесорні потужності авто, а у випадку використання смартфона може вести до швидкого розряду батареї. В такому випадку можна рекомендувати роботу мобільного телефону з постійним підключенням до зарядного пристрою.

Для вирішення задачі розпізнавання аварійних ситуацій ми обрали архітектуру *ResNet*. Метою запропонованої архітектури є швидка обробка аудіо-сигналів, навчання безпосередньо з аудіо-сигналу, дискримінаційного представлення, яке досягає хорошої класифікаційної продуктивності на різних звуках автомобільних аварій.

Однією з проблем використання ЗНМ в обробці звуку є те, що довжина вхідного зразка повинна бути фіксована, але звук, отриманий з навколишнього середовища, може мати різну тривалість. Тому необхідно адаптувати ЗНМ для використання зі звуковими сигналами різної довжини. Ми вирішуємо цю задачу шляхом обмеження тривалості звукових треків 5 секундами.

Також різні мікрофони мобільних чи стаціонарних пристроїв можуть мати різну чутливість, в результаті чого амплітуда звукових треків, що подаються на вхід нейронної мережі, може суттєво відрізнятись при одних і тих самих аварійних ситуаціях. В подальших дослідженнях ми плануємо дослідити вплив амплітуди звукового сигналу на здатність ЗНМ коректно його класифікувати, а також запропонувати підхід до нормалізації амплітуди вхідних даних, що дозволить підвищити точність роботи ЗНМ.

Один із способів обійти обмеження, накладене вхідним шаром ЗНМ, це розділити аудіо-сигнал на



кілька кадрів фіксованої довжини за допомогою розсувного вікна відповідної ширини. Ширина вікна залежить в основному від частоти дискретизації сигналу. Крім того, послідовні аудіо-кадри також можуть мати певний відсоток перекриття, метою якого є максимальне використання інформації. Це, природно, збільшує кількість зразків, оскільки деякі частини аудіо-сигналу повторно використовуються, і це можна розглядати як явесь збільшення даних. Процес нарізки аудіо-сигналу у відповідні кадри проілюстрований на рис. 3.

У випадку, коли вхідна аудіо-форма сигналу  $X$  розділена на  $S$  кадрів, що позначаються як  $X_1, X_2, \dots, X_S$ , під час класифікації нам потрібно агрегувати прогнози ЗНМ, щоб прийти до рішення щодо  $X$ . Для цього для досягнення остаточного рішення можуть бути використані різні правила злиття, такі як більшість голосів або правило суми:

$$y_i = \sum_{j=1}^S a_{ji}, \text{ або } y_i = \frac{1}{S} \sum_{j=1}^S a_{ji},$$

де  $a_{ji}$  – прогноз ЗНМ для  $j=1, \dots, S$  сегмента аудіо-сигналу  $X$ ,  $i = 1, \dots, K$  – це передбачуваний клас.  $S$  – це кількість кадрів, а  $K$  – кількість класів.

Коли є  $K$  класів, ми генеруємо  $K$  значень і далі для відповідного аудіо-входу обираємо клас з максимальним значенням  $y_i = \max_{k=1}^K y_k$ .

Зважаючи на обрану довжину аудіо-сигналу дорожньої аварії в 5 секунд, нами було побудовано згорткову нейронну мережу, що складається з п'яти згорткових шарів.

Декілька згорткових шарів використовуються для того, щоб вловлювати точну часову структуру сигналу та слугувати фільтрами, які виконують поставлену задачу класифікації аварійних ситуацій. Це також дасть можливість позбутись використання додаткового модулю обробки сигналів, оскільки така нейронна мережа є достатньо

потужною, щоб «втягувати» релевантну інформацію низького та високого рівня з аудіо-сигналів, які надходять.

Оскільки обсяг вхідних даних для навчання мережі є обмеженим, використання більш глибоких архітектур нейронної мережі не є доцільним, оскільки може призвести до «перенавчання» (overfitting) – явища, при якому нейронна мережа коректно розпізнає дані, на яких відбувалось навчання, але невірно розпізнає нові дані. Перенавчання стається тоді, коли модель починає «запам'ятовувати» тренувальні дані, замість того, щоби «вчитися» узагальненню з тенденції. В результаті «перенавчана» модель має погану передбачувальну продуктивність, оскільки вона занадто сильно реагує на другорядні відхилення в тренувальних даних. Нашою ж метою є уможливити здійснення надійних передбачень на загальних даних, на яких не здійснювалося тренування [12].

Щоб оцінити точність моделі, ми використали 10-кратну перехресну перевірку. Для оцінки точності моделі ми використали такі показники як точність та втрата. Для функції втрат нами було обрано бінарну перехресну ентропію. Точність – це відсоток правильно класифікованих екземплярів. Для кожного класу втрата визначається як мінімальне значення втрат серед всіх епох (ітерацій в процесі навчання) в процесі перевірки. Аналогічно, точність для кожного класу розраховується шляхом отримання найкращого значення точності класифікації у кожену епоху.

Запропонований підхід був оцінений на основі набору тестових даних з 2000+ аудіо-зразків аварійних ситуацій. Експериментальні результати показали, що при використанні запропонованої ЗНМ досягається точність класифікації у 90 % + при розпізнаванні звуків нових аварій. Це краще, ніж моделі, які використовують необроблений аудіо-сигнал як вхід та набір даних UrbanSound8k (Salamon et al., 2014) для класифікації звуків навколишнього

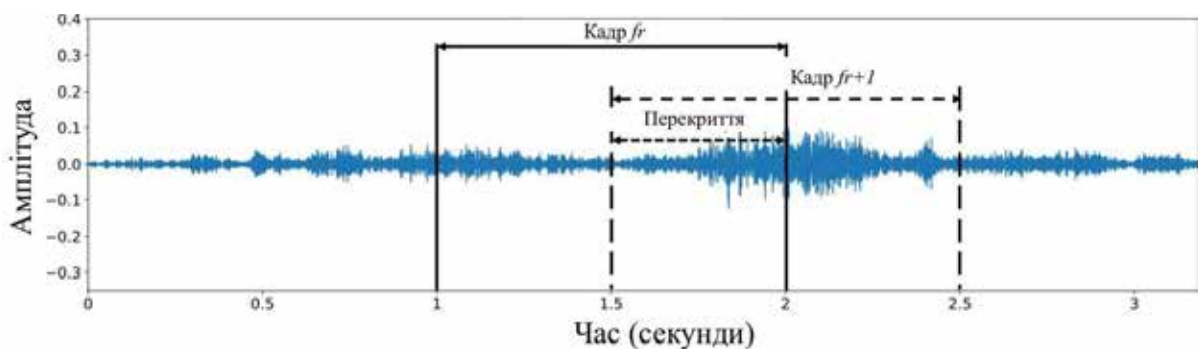


Рис. 3. Нарізка вхідного аудіо-сигналу в кілька кадрів ( $fr, fr + 1$ ) з відповідним відсотком перекриття (50 %)



середовища, а також підходи, що базуються на обробці даних, що надходять від вбудованих акселерометрів та гіроскопів автомобіля.

Проте, зважаючи на те, що кожен відсоток точності дає надзвичайний економічний ефект, має сенс продовжити вдосконалення даного підходу з метою подальшого підвищення точності класифікації аварій.

Одним з можливих напрямків підвищення точності розпізнавання аварій може стати збільшення навчального набору даних шляхом створення нових екземплярів аудіо-записів за допомогою деформації існуючих. За допомогою такого підходу ми можемо збільшити в рази кількість елементів навчального набору даних, щоб таким чином вирішити проблему його обмеженості.

Цілком імовірно, що загальну продуктивність даного підходу можна надалі покращити, об'єднавши підходи, які використовують не лише одновимірні, але і двовимірні представлення (спектрограми) звукових сигналів у якості вхідних даних. Для цього можна виконати перетворення кожного аудіо-запису вибірки у спеціальні зображення – спектрограми Мела, що є його компактним інформативним візуальним представленням. В ході подальших досліджень ми встановимо, чи може такий підхід призвести до кращої продуктивності при класифікації звуків автомобільних аварій.

Крім того, фільтри, навчені в проміжних згорткових шарах запропонованої ЗНМ, не демонструють домінуючих частот і, здається, шумні. Необхідне подальше дослідження, щоб з'ясувати, як обійти таку проблему і, можливо, ще більше покращити ефективність запропонованого підходу.

**Висновки.** В дослідженні запропоновано підхід до класифікації звуків автомобільних аварій за допомогою згорткових нейронних мереж.

Під час дослідження було встановлено, що згорткові нейронні мережі, які розроблені спеціально для розпізнавання зображень, можна успішно навчити класифікувати звуки аварійних ситуацій на дорогах.

Запропонований наскрізний алгоритм навчання ЗНМ вивчає представлення безпосередньо з аудіо-сигналу. Запропонований підхід був оцінений на основі набору даних з 2000+ аудіо-зразків аварійних ситуацій, і експериментальні результати показали, що він дозволяє перевершити точність класифікації існуючих підходів на основі акселерометрів, демонструючи 90 %+ достовірності при розпізнаванні звуків нових аварій. Він також працює краще, ніж моделі, які використовують необроблений аудіо-сигнал як вхід та набір даних *UrbanSound8k* (Salamon et al., 2014) для класифікації звуків навколишнього середовища.

Запропонована архітектура ЗНМ має менше параметрів, ніж більшість існуючих архітектур ЗНМ, що використовуються для класифікації звуків. Крім того, запропонований підхід не вимагає жодного модуля обробки сигналу для класифікації звуку, що робить цю модель цілком придатною для використання в мобільних додатках розпізнавання звуку, або у вбудованих системах автомобілів.

В ході подальших досліджень необхідно встановити, як можна покращити загальну продуктивність даного підходу, об'єднавши підходи, які використовують не лише одновимірні, але і двовимірні представлення (спектрограми) звукових сигналів у якості вхідних даних.

Також фільтри, навчені в проміжних згорткових шарах запропонованої ЗНМ, не демонструють домінуючих частот і є зашумленими, тому необхідно у подальшому з'ясувати, як вирішити це завдання і, можливо, покращити ефективність запропонованого підходу.

#### Список літератури:

1. Лебідь О., Кіпоренко С., Вовк В. Виявлення кібератак та підвищення інформаційної безпеки на основі технології нейронних мереж в умовах кібервійни. *Наука і техніка сьогодні*. 2023. № 1 (15). С. 238–256. DOI: [https://doi.org/10.52058/2786-6025-2023-1\(15\)-238-256](https://doi.org/10.52058/2786-6025-2023-1(15)-238-256).
2. Мясіщев О.А., Ленков Є.С., Білик О.М. Розпізнавання графічних образів з використанням нейронних мереж. *Збірник наукових праць Військового інституту КНУ ім. Т. Шевченка*. 2016. № 54. С. 143–148. URL: [https://mil.knu.ua/files/223\\_1494110996.pdf](https://mil.knu.ua/files/223_1494110996.pdf). (дата звернення: 11.02.2024)
3. Мірошник О.О. Комп'ютерне моделювання нейронної мережі для розпізнавання вейвлет-образів. *Вісник Харківського національного технічного університету сільського господарства імені Петра Василенка*. 2014. Вип. 154. С. 57–58. URL: [http://nbuv.gov.ua/UJRN/Vkhdtusg\\_2014\\_154\\_22](http://nbuv.gov.ua/UJRN/Vkhdtusg_2014_154_22) (дата звернення: 8.01.2024).
4. Мірошник О.О., Святобатько А.В. Моделювання нейронної мережі для задач прогнозування фізичних параметрів. *Праці Таврійського державного агротехнологічного університету*. 2013. Т. 5. № 13. С. 34–40. URL: [https://nauka.tsatu.edu.ua/print-journals-tdatu/13-5/13\\_5/zmist.pdf](https://nauka.tsatu.edu.ua/print-journals-tdatu/13-5/13_5/zmist.pdf) (дата звернення: 9.02.2024).
5. Gencoglu O., Virtanen T., Huttunen H. Recognition of acoustic events using deep neural networks. Department of Signal Processing, Tampere University of Technology. EUSIPCO 2014. 506–510. 2014.

URL: [https://www.researchgate.net/publication/288213802\\_Recognition\\_of\\_acoustic\\_events\\_using\\_deep\\_neural\\_networks](https://www.researchgate.net/publication/288213802_Recognition_of_acoustic_events_using_deep_neural_networks) (дата звернення: 14.02.2024).

6. Abdoli S., Cardinal P., Koerich A. End-to-end environmental sound classification using a 1D convolutional neural network. *Machine Learning*. 2019. DOI: <https://doi.org/10.48550/arXiv.1904.08990>.

7. Paciorek M., Klusek A., Wawryka P. Effective Car Collision Detection with Mobile Phone Only. *Computational Science*. 2021. DOI:10.1007/978-3-030-77980-1\_24.

8. Кутковецький В. Я. Розпізнавання образів: навч. посіб. Миколаїв: Видавництво ЧНУ ім. Петра Могили, 2017. 420 с.

9. Субботін С.О. Нейронні мережі. Теорія та практика: навч. посіб. Житомир: Видавництво О.О. Євенок, 2020. 184 с.

10. Новотарський М.А., Нестеренко Б.Б. Штучні нейронні мережі: обчислення. *Праці Інституту математики НАН України. Математика та її застосування*. 2004. Т. 50. 408 с.

11. Лящинський П.Б. Синтез структур згорткових нейронних мереж для класифікації біомедичних зображень. Випускна кваліфікаційна робота. Тернопіль, 2019. 82 с. URL: [http://dspace.wunu.edu.ua/bitstream/316497/38265/1/Лящинський\\_Павло\\_МР\\_2019.pdf](http://dspace.wunu.edu.ua/bitstream/316497/38265/1/Лящинський_Павло_МР_2019.pdf). (дата звернення: 15.05.2024).

12. Яблонський С.В. Метод та алгоритми оцінювання емоційного стану людини на основі аналізу головних сигналів. Магістерська дисертація. Київ, 2018. 87 с. URL: <https://ela.kpi.ua/server/api/core/bitstreams/b2baa90b-fbed-49d9-955f-4f7cb9f2ea64/content> (дата звернення: 15.05.2024).

### **Mogylyevych D.I., Khmil R.V. ANALYSIS OF THE APPLICATION OF NEURAL NETWORKS FOR AUTOMOBILE ACCIDENTS RECOGNITION**

*The e-Call automatic emergency call system has been mandatory in the European Union since 2018. This system automatically calls emergency services in the event of an accident, but there is a problem with false calls that lead to unnecessary expenditure of emergency services resources. Most car manufacturers solve the primary task of determining an emergency by analyzing information coming from built-in accelerometer sensors. As a result, quite often a sharp braking that avoids an accident is mistakenly identified as an emergency and leads to a false call to emergency services.*

*This paper proposes an approach to recognizing car accidents based on the use of convolutional neural networks to classify audio tracks recorded inside a car in the event of road accidents.*

*The purpose of the study is to analyze the existing mathematical apparatus of convolutional neural networks and develop an approach to solving the problem of recognizing and verifying car accidents based on their use. The main stages of the research are the analysis of data processing methods, improvement of learning algorithms, and the use of advanced technologies in the field of machine learning.*

*The architecture of neural networks is analyzed to effectively solve the task using mobile devices, including the study of various configurations of neural networks, their adaptation to the conditions of limited computing resources of mobile devices, and the selection of a model with minimal resource use while ensuring high accuracy of accident recognition.*

**Key words:** artificial intelligence, neural networks, sound signal processing.