

DOI: <https://doi.org/10.15276/ict.03.2026.49>

УДК 004.451:502/504

Підвищення точності розпізнавання аудіо-емоцій людини на базі ансамблевої класифікації з адаптивною нечіткою агрегацією на основі функції Гомперца

Арсирій Олена Олександрівна¹⁾

ORCID: <https://orcid.org/0000-0001-8130-9613>; e.arsiriy@gmail.com. Scopus Author ID: 54419480900

Андронаті Олександр Кирилович¹⁾

ORCID: <https://orcid.org/0009-0009-1794-5864>; alex.andronati@gmail.com. Scopus Author ID: 58677655800

¹⁾ Національний університет “Одеська Політехніка”, пр. Шевченка, 1. Одеса, 65044, Україна

АНОТАЦІЯ

У роботі розглянуто задачу розпізнавання аудіо-емоцій людини та досліджено підхід до підвищення ефективності ансамблевої класифікації на основі адаптивної нечіткої агрегації. Запропоновано метод агрегації прогнозів класифікаторів машинного та глибокого навчання із використанням функції Гомперца, у якому кількість найбільш інформативних класів визначається адаптивно для кожного об'єкту на основі нормалізованої ентропії розподілу ймовірностей. На відміну від стандартного підходу з фіксованою кількістю найбільш інформативних класів, запропонований метод враховує рівень невизначеності конкретного прогнозу, що дає змогу динамічно змінювати кількість класів, які беруть участь у подальшій агрегації. Експериментальне дослідження проведено на наборі даних Ryerson Audio-Visual Database of Emotional Speech and Song. Для побудови гетерогенного ансамблю використано вісім базових класифікаторів, які працюють із різними представленнями аудіо-даних. Для табличних спектральних ознак додатково застосовано процедуру Backward Feature Elimination, що дозволило сформувати індивідуальні простори ознак для базових класифікаторів. У межах експерименту для кожного способу агрегації сформовано та оцінено двісті дев'ятнадцять ансамблевих класифікаторів. Стандартна нечітка агрегація на основі функції Гомперца з фіксованою кількістю найбільш інформативних класів забезпечила Accurasy вісімдесят одна ціла тридцять дев'ять сотих та F1-score вісімдесят одна ціла двадцять сотих відсотків, тоді як запропонований адаптивний метод досяг Accurasy вісімдесят п'ять цілих п'ятдесят шість сотих відсотків та F1-score вісімдесят п'ять цілих п'ятдесят п'ять сотих відсотків. Таким чином, адаптивний метод забезпечує підвищення Accurasy на чотири цілих сімнадцять сотих відсотків та F1-score на чотири цілих тридцять п'ять сотих відсотків порівняно зі стандартною нечіткою агрегацією.

Ключові слова: розпізнавання аудіо-емоцій; ансамблева класифікація; функція Гомперца, адаптивна нечітка агрегація; стекинг-ансамбль; відбір ознак; машинне навчання; глибоке навчання; налаштування гіперпараметрів.

Для цитування: Арсирій О. О., Андронаті О. К. “Підвищення точності розпізнавання аудіо-емоцій людини на базі ансамблевої класифікації з адаптивною нечіткою агрегацією на основі функції Гомперца”. *Інформатика. Культура. Техніка*. 2026; Том 3 № 2 (4): 647–664. DOI: <https://doi.org/10.15276/ict.03.2026.49>

Актуальність. Автоматичне розпізнавання емоційного стану людини є одним із актуальних напрямів розвитку інтелектуальних інформаційних систем. Технології автоматичного розпізнавання емоцій знаходять застосування у людино-машинній взаємодії, системах підтримки прийняття рішень, персоналізованих інформаційних системах, онлайн-навчанні та інших прикладних сферах. Одним із напрямів такого аналізу є автоматичне розпізнавання емоцій за візуальними характеристиками людини. Зокрема, у роботі [1] запропоновано адаптивну згорткову нейронну мережу для розпізнавання міміки людини, що демонструє перспективність застосування адаптивних методів глибокого навчання для аналізу емоційних станів.

У той час як розпізнавання емоцій за мімікою обличчя вже активно досліджується та широко впроваджується, саме розпізнавання аудіо-емоцій (Speech Emotion Recognition, SER) набуває все більшої актуальності в практичних завданнях. Мовлення містить не лише лінгвістичну інформацію, а й широкий спектр акустичних характеристик, пов'язаних з емоційним станом мовця. Тому SER є важливим напрямом дослідження, спрямованим на автоматичне виявлення та класифікацію емоцій за мовленнєвим сигналом [2], [3]. Але точність існуючих SER-систем, які побудовано з використанням базових класифікаторів на основі методів опорних векторів, К-найближчих сусідів, екстремального градієнтного бустінгу, випадкового лісу, логістичної регресії або багатопараметричних перцептронів не

відповідає сучасним вимогам до практичного розпізнавання. Спроби використовувати ансамблі із базових класифікаторів в SER-системах дають позитивний результат щодо підвищення точності розпізнавання аудіо-емоцій, але залишається недостатньо дослідженим використання ансамблів з динамічним складом базових класифікаторів та нестандартних методів агрегації прогнозів, у тому числі використання нечіткої агрегації з використанням функції Гомперца. На подолання таких недоліків спрямовано матеріал даного дослідження

Огляд сучасних публікацій за тематикою дослідження. Особливостями задачі SER є динамічна природа мовленнєвого сигналу та складні взаємозв'язки між акустичними ознаками. Сучасні дослідження охоплюють широкий спектр методів – від використання традиційних акустичних ознак і класичних алгоритмів машинного навчання до глибоких нейронних мереж та моделей [2], [3]. Сучасний розвиток глибокого навчання також сприяє переходу від розпізнавання базових емоцій до аналізу складніших емоційних станів та використання різних представлень мовленнєвої інформації [3].

Водночас висока варіативність емоційних проявів суттєво ускладнює задачу автоматичного розпізнавання. Один і той самий емоційний стан може проявлятися по-різному у різних мовців, тоді як різні емоції можуть характеризуватися схожими акустичними властивостями. Додатковою проблемою є відмінність наборів даних, на яких виконуються навчання та тестування моделей. Модель, навчена на одному наборі, може демонструвати зниження ефективності під час роботи з невідомими даними, оскільки емоційні характеристики та умови формування різних наборів даних можуть суттєво відрізнятися [4].

Одним із підходів до підвищення надійності класифікації в таких умовах є використання ансамблевих моделей, у яких підсумкове рішення формується на основі результатів декількох базових класифікаторів. Різні моделі можуть використовувати різні представлення даних і виявляти різні закономірності, унаслідок чого їхні рішення доповнюють один одного. Об'єднання таких рішень дає змогу компенсувати окремі помилки базових класифікаторів та підвищити стійкість підсумкової класифікації. В роботі [5] автори демонструють ефективність використання гетерогенного стекінг-ансамблю для класифікації аудіо-даних різної природи, в тому числі для класифікації аудіо-даних.

Водночас результат ансамблю значною мірою залежить від способу агрегації рішень окремих класифікаторів. Традиційні стратегії агрегації, такі як hard voting та soft voting, передбачають фіксовані правила об'єднання результатів. При цьому ваги класифікаторів можуть бути однаковими або визначатися заздалегідь, що не дає змоги повною мірою враховувати особливості конкретного екземпляра. Більш гнучким підходом є використання нечітких методів агрегації, які дають змогу враховувати ступінь впевненості та невизначеність класифікаційних рішень.

Одне з сучасних рішень, а саме застосування нечіткої агрегації на основі функції Гомперца для побудови ансамблю було запропоновано в дослідженні [6]. У цій роботі розроблено метод, у якому значення впевненості декількох CNN-класифікаторів перетворюються на нечіткі ранги за допомогою модифікованої функції Гомперца. Після цього ранги та значення впевненості використовуються для формування підсумкового рішення ансамблю. Автори зазначають, що такий підхід дає змогу враховувати впевненість кожного класифікатора для конкретного тестового зразка, на відміну від традиційних правил hard voting та soft voting. Запропонований метод продемонстрував високу ефективність на задачі класифікації COVID-19 за КТ-зображеннями та показав перевагу над окремими базовими моделями.

В роботі [7] аналогічний підхід було використано для розпізнавання емоцій за даними електроенцефалографії. Автори запропонували нечіткий ансамбль на основі функції Гомперца, який об'єднує результати трьох моделей: 1D-CNN+LSTM, 1D-CNN+GRU та 1D-CNN. Запропонований ансамбль досліджувався на наборах даних DEAP та AMIGOS для визначення валентності та рівня збудження емоцій. Результати показали, що застосування

нечіткої агрегації на основі функції Гомперца перевищує окремі базові моделі за Accurasy та F1-score.

Водночас використання цього підходу в [6], [7] має певне обмеження, пов'язане з формуванням множини найбільш інформативних класів (Тор-К). У підході враховуються ранги $1, 2, \dots, k$, що належать до множини Тор-К, а для класів, які не потрапляють до цієї множини, застосовується штраф [6]. Значення k при цьому є фіксованим і не змінюється залежно від конкретного вхідного зразка.

Таким чином, хоча ранжування класифікаторів та їхні значення впевненості визначаються окремо для кожного зразка, кількість рангів k , що враховуються під час агрегації, залишається сталою. Це може бути обмеженням для задачі розпізнавання емоцій за мовленням, оскільки для різних екземплярів базові класифікатори можуть мати дуже різну ступінь впевненості. Для зразка з чітко вираженою емоцією достатнім може бути врахування лише 1-2 найбільш впевнених рішень, тоді як для неоднозначного зразка доцільним може бути залучення більшої кількості рішень базових класифікаторів.

Отже, рішенням цього недоліку є адаптивне визначення кількості рангів Тор-К залежно від характеристик конкретного екземпляра. Такий підхід дає змогу пов'язати процес агрегації з рівнем невизначеності класифікаційного рішення: для більш впевнених прогнозів використовувати меншу кількість найбільш інформативних рангів, а для невизначених – розширювати множину рішень, що беруть участь в агрегації.

Мета та задачі дослідження. На основі аналізу літературних джерел сформулюємо мету та задачі дослідження. Метою даної роботи є підвищення точності розпізнавання аудіо-емоцій людини за рахунок розробки методу побудови ансамблевих класифікаторів аудіоданих з використанням агрегації на основі функції Гомперца з адаптивним визначенням кількості найбільш інформативних класів. Запропонований метод спрямований на підвищення ефективності ансамблевої класифікації за рахунок динамічного визначення кількості рангів, що беруть участь у формуванні підсумкового рішення, відповідно до ступеня невизначеності конкретного екземпляра в наборі аудіо-даних.

Для досягнення поставленої мети в роботі вирішено наступні задачі:

- 1) попередня обробка аудіо-даних з наявними інформаційними інтонаціями на основі запропонованої тривірневої моделі їх представлення;
- 2) обґрунтування вибору та побудова базових класифікаторів для розпізнавання аудіо-емоцій людини;
- 3) побудова мультиагрегаційного ансамблевого класифікатору з використанням методу адаптивної нечіткої стекинг-агрегації на основі функції Гомперца;
- 4) експериментальне дослідження та аналіз результатів методу адаптивної нечіткої стекинг-агрегації на основі функції Гомперца.

1. Попередня обробка аудіо-даних

У дослідженні використано набір даних RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song) [8]. Набір містить аудіозаписи 24 акторів, які вимовляють фрази з різними емоційними інтонаціями. Загальна кількість використаних аудіофайлів становить 1440. Аудіозаписи представлені у форматі WAV. Задача розглядається як багатокласова класифікація емоційних станів. RAVDESS містить вісім категорій емоцій: neutral – нейтральний стан; happy – радість; sad – сум; angry – гнів; fearful – страх; disgust – відраза; surprised – здивування.

У даному дослідженні початкові вісім категорій було перетворено на сім класів шляхом об'єднання категорій calm та neutral в один клас neutral. Таким чином, перетворення міток здійснювалося за правилом: calm стає neutral, тоді як мітки happy, sad, angry, fearful, disgust та surprised залишалися без змін. Об'єднання calm та neutral застосовано для формування єдиного класу, що характеризує відсутність вираженого емоційного стану.

У початковому наборі категорія neutral містить 96 аудіофайлів, а категорія calm – 192 аудіофайли. Після їх об'єднання сформований клас neutral містить 288 аудіофайлів ($96 +$

192). Кожен із шести інших класів містить по 192 аудіофайли. Таким чином, набір даних має помірний дисбаланс класів. RAVDESS обрано як експериментальний набір даних, оскільки він безпосередньо відповідає задачі дослідження – класифікації емоційного мовлення – та містить багатокласову структуру, що дає змогу оцінити здатність запропонованого ансамблевого методу працювати з декількома емоційними станами.

Аудіосигнал у вихідному вигляді є часовою послідовністю значень амплітуди та не має фіксованої табличної структури, безпосередньо придатної для більшості класичних алгоритмів машинного навчання. Тому перед класифікацією виконується його перетворення у числові представлення, що дозволяють виділити різні характеристики аудіосигналу.

У дослідженні використовується трирівнева модель представлення аудіоданих, у межах якої один і той самий аудіосигнал трансформується у три різні простори:

- 1) табличне представлення спектральних характеристик;
- 2) спектральне представлення у вигляді зображення;
- 3) часовий ряд спектральних характеристик на основі мел-частотних кепстральних коефіцієнтів (MFCC).

Такий підхід дає змогу отримувати різноманітну інформацію з одного аудіосигналу.

Перед формуванням зазначених представлень виконується попередня обробка аудіосигналів. Зокрема, здійснюється підготовка сигналу до подальшого спектрального аналізу та обчислення відповідних характеристик. Для отримання частотно-часового представлення використовується короточасне перетворення Фур'є, при якому аудіосигнал розділяється на короткі часові фрейми та для кожного фрейму визначається спектр.

Перший рівень представлення аудіо-даних передбачає перетворення аудіо-файлів у вектор спектральних ознак фіксованої довжини.

Згідно з дослідженнями [9], [10], [11], найперспективнішими спектральними характеристиками для класифікації аудіо-даних є спектральний центроїд, спектральний спад, спектральна пологість, частота переходів через нуль, мел-частотні кепстральні коефіцієнти (MFCC) та спектральний контраст. Усі зазначені спектральні характеристики було використано для формування вхідного вектору для навчання базових класифікаторів, що працюють із табличними даними. Для MFCC у дослідженні використовується 20 коефіцієнтів.

Другий рівень представлення аудіо-даних передбачає представлення аудіо-сигналу у вигляді мел-спектрограми. У межах дослідження мел-спектрограма після масштабування подається у вигляді квадратного зображення розміром 128x128 пікселів. Таке представлення використовується для подальшого аналізу аудіо-даних згортковою нейронною мережею.

Третій рівень представлення аудіо-даних базується на збереженні часової послідовності MFCC. На відміну від табличного представлення, де MFCC усереднюються за часом, у цьому випадку зберігається повна послідовність значень для окремих часових фреймів. Таке представлення використовується для подальшого аналізу аудіо-даних рекурентною нейронною мережею.

Після перетворення аудіо-даних набір RAVDESS було розділено на навчальну, валідаційну та тестову вибірки. Навчальна та валідаційна вибірки використовувалися для формування та налаштування моделей та відбору простору ознак, тоді як незалежна тестова вибірка використовувалася виключно для оцінювання вже сформованих і зафіксованих конфігурацій моделей та просторів ознак. Тестова вибірка була відокремлена на початковому етапі експерименту та не використовувалася під час відбору ознак, налаштування гіперпараметрів базових класифікаторів або вибору складу ансамблю. Такий підхід дає змогу уникнути використання тестових даних під час побудови моделей і отримати більш об'єктивну оцінку їхньої прогностичної здатності.

Відповідно до схеми, використаної в дослідженні, 75 % даних спочатку формують повну навчальну вибірку, тоді як 25 % залишаються для тестування конфігурацій моделей. Повна навчальна вибірка, у свою чергу, розділяється у співвідношенні 75/25 на власне навчальну та валідаційну вибірки. Таким чином, від загального обсягу RAVDESS приблизно 56% даних

використовуються для навчання, 19 % для валідації та 25 % для незалежного тестування. Розподіл виконується із застосуванням стратифікації, що забезпечує збереження розподілу емоційних класів у кожній із вибірок.

При цьому під час формування вибірок ідентифікатор актора (Actor ID) не використовувався як критерій розділення даних. Отже, аудіозаписи одного й того самого актора могли одночасно потрапляти до навчальної, валідаційної та тестової вибірок. У зв'язку з цим отримані результати характеризують сценарій *speaker-dependent* розпізнавання емоцій, за якого моделі під час навчання можуть зустрічатися із записами мовців, представлених також у тестовій вибірці. Це є обмеженням проведеного дослідження, оскільки результати, отримані за такого протоколу, не дають змоги безпосередньо оцінити здатність запропонованого методу узагальнювати на повністю нових мовців і можуть відрізнятися від результатів у *speaker-independent* сценарії. Оцінювання із розділенням даних за Actor ID, за якого записи кожного актора належать лише до однієї вибірки, є перспективним напрямом подальших досліджень.

2. Обґрунтування вибору та побудова базових класифікаторів для розпізнавання аудіо-емоцій людини

Для подальшої побудови ансамблевого класифікатора було обґрунтовано використання набору базових класифікаторів, що відрізняються принципами побудови та способом обробки аудіоданих. Використання моделей різної природи дає змогу отримувати різні класифікаційні рішення для одного й того самого аудіозразка, що є важливою передумовою ефективної ансамблевої агрегації.

Було розроблено шість базових класифікаторів для першого рівня представлення даних: метод опорних векторів (SVM), К-найближчих сусідів (KNN), екстремальний градієнтний бустінг (XGBoost, XGB), випадковий ліс (RF), логістична регресія (LR) та багатоваріантний перцептрон (MLP). Для другого та третього рівнів представлення даних було розроблено згорткову нейронну мережу (CNN) та довгострокову пам'ять (LSTM) відповідно.

LR використовується як еталонний статистичний метод для моделювання ймовірностей класів на основі логіт-перетворення. SVM застосовується для побудови оптимальної розділяючої гіперплощини в просторах високої розмірності з використанням ядерних функцій.

RF – ансамблевий алгоритм на основі бегінгу дерев прийняття рішень, що демонструє високу стійкість до перенавчання на малих вибірках.

XGBoost – реалізація градієнтного бустингу з використанням регуляризації та апроксимації другого порядку, що дозволяє виявляти складні нелінійні зв'язки.

KNN – метричний алгоритм, що базується на гіпотезі близькості об'єктів у просторі ознак.

MLP – повнозв'язна нейронна мережа прямого поширення.

CNN – спеціалізована архітектура для автоматичного виділення просторових ознак із спектрограм аудіо-сигналів.

LSTM – тип рекурентної мережі для моделювання довготривалих часових залежностей у послідовностях ознак аудіо-даних.

Процес підготовки базових моделей складався з декількох послідовних етапів. На першому етапі виконується первинне налаштування архітектури та гіперпараметрів моделей на повному просторі ознак.

Для базових класифікаторів для першого рівня представлення даних додатково виконується процедура Backward Feature Elimination (BFE), яка формує власну підмножину найбільш інформативних спектральних ознак для кожного базового класифікатора. Таким чином, різні класифікатори працюють не з однаковим, а з індивідуально оптимізованим простором ознак.

Після формування оптимізованих підмножин ознак для базових класифікаторів для першого рівня представлення даних виконується повторне налаштування гіперпараметрів.

Така двоетапна схема дає змогу спочатку визначити базову конфігурацію моделі, після чого оптимізувати її з урахуванням скороченого простору ознак.

Окремою вимогою до базових класифікаторів є отримання ймовірнісних оцінок належності екземплярів до кожного емоційного класу. Це необхідно для подальшого застосування запропонованого методу нечіткої агрегації на основі функції Гомперца.

Якщо такі методи, як логістична регресія або глибокі нейронні мережі (NN, CNN, LSTM), формують ймовірнісні оцінки безпосередньо через свою природу, то інші класифікатори потребують спеціальних адаптацій. Для алгоритму KNN ймовірнісні оцінки формуються на основі часток класів серед k сусідів. Для Random Forest та XGBoost оцінка ймовірності обчислювалася як пропорція елементів відповідного класу у кінцевих листках дерев. Зокрема, для методу опорних векторів (SVM), застосовується процедура калібрування Платта [12].

Після завершення оптимізації базові класифікатори навчаються на сформованій навчальній вибірці. Отримані моделі надалі використовуються для формування набору прогнозів, які є вхідними даними для запропонованого ансамблевого методу.

3. Побудова мультиагрегаційного стекінг-ансамблю з використанням методу адаптивної нечіткої агрегації на основі функції Гомперца

Наступним кроком після індивідуального налаштування базових класифікаторів є об'єднання їх передбачень у гетерогенний мультиагрегаційний стекінг-ансамбль

Для визначення топології ансамблю здійснювався вичерпний комбінаторний перебір усіх можливих підмножин базових алгоритмів в діапазоні від $k = 3$ до $k = 8$ моделей.

Загальна кількість унікальних комбінацій базових класифікаторів N визначається як сума сполук із n елементів по k ($n = 8, k = 3 \dots 8$):

$$N = C_8^3 + C_8^4 + C_8^5 + C_8^6 + C_8^7 + C_8^8 = 219.$$

На відміну від традиційних методів агрегування, таких як Hard Voting та Soft Voting, які використовують фіксовані правила об'єднання прогнозів базових класифікаторів, запропонований метод враховує не лише кінцеві рішення моделей, а й ступінь їхньої впевненості та рівень невизначеності прогнозів.

Запропонований метод агрегування базується на нелінійному перетворенні оцінок впевненості класифікаторів із використанням функції Гомперца. На відміну від раніше запропонованих алгоритмів рангово-нечіткого агрегування [6, 7], у запропонованому методі кількість класів, які беруть участь у подальшому агрегуванні, визначається адаптивно на основі ентропії розподілу ймовірностей для кожного окремого екземпляра. Це дозволяє використовувати різний механізм агрегування для об'єктів із високою та низькою невизначеністю прогнозу, зменшуючи втрати інформації під час прийняття остаточного рішення та підвищуючи стійкість ансамблю до неоднозначних випадків класифікації.

Після отримання прогнозів від базових класифікаторів формується матриця ймовірностей належності кожного об'єкта до всіх можливих класів.

На першому етапі виконується нелінійне перетворення отриманих ймовірностей за допомогою функції Гомперца. Таке перетворення забезпечує побудову нечітких рангів класів, що дозволяє підсилити відмінності між більш і менш достовірними прогнозами класифікаторів без жорсткого відкидання інформації.

На другому етапі для кожного прогнозу оцінюється рівень невизначеності за допомогою ентропії розподілу ймовірностей. Якщо розподіл характеризується низькою ентропією, тобто один із класів має суттєво більшу ймовірність порівняно з іншими, прогноз вважається достатньо впевненим. У разі високої ентропії ймовірності між декількома класами розподілені більш рівномірно, що свідчить про неоднозначність прогнозу. На основі нормалізованої ентропії для кожного об'єкта автоматично визначається кількість найбільш інформативних класів (Тор-К), які братимуть участь у подальшому агрегуванні. Таким

чином, на відміну від класичного алгоритму з фіксованим значенням K , запропонований метод адаптується до складності кожного окремого прикладу.

Після визначення кількості домінуючих класів для кожного класифікатора виконується відбір Тор-К класів із найкращими значеннями нечітких рангів. Для відібраних класів зберігаються обчислені значення рангу Гомперца, тоді як усім іншим класам призначається фіксоване штрафне значення. Одночасно формується маскована матриця ймовірностей, у якій значення відкинутих класів замінюються нулями. Такий підхід дозволяє усунути вплив малої ймовірності альтернатив під час подальшого агрегування.

На наступному етапі для кожного класу обчислюються два інтегральні показники. Перший показник характеризує сумарний нечіткий ранг, отриманий від усіх базових класифікаторів, і відображає ступінь узгодженості їхніх прогнозів. Другий показник визначається на основі скоригованих апостеріорних ймовірностей та характеризує сумарну впевненість ансамблю щодо відповідного класу.

На завершальному етапі обидва інтегральні показники об'єднуються в єдину функцію агрегування. Для кожного об'єкта значення цієї функції обчислюється для всіх можливих класів, після чого остаточним рішенням ансамблю вважається клас, для якого значення інтегральної функції є мінімальним. Такий підхід забезпечує комплексне врахування як впевненості окремих класифікаторів, так і рівня їхньої узгодженості, що підвищує якість прийняття рішення порівняно з традиційними методами голосування.

Наведемо математичну формалізацію розробленого методу

Нехай ансамбль складається з N базових класифікаторів, які для об'єкта x повертають оцінки ймовірностей за C класами.

Позначимо вектор передбачень i -го базового класифікатора як:

$$P_i = [p_{i,1}, p_{i,2}, \dots, p_{i,C}], \quad (1)$$

де $p_{i,k}$ – ймовірність належності об'єкта x до k -го класу, отримана i -м базовим класифікатором.

$$\sum_{k=1}^C p_{i,k} = 1, i = 1, \dots, N.$$

Для підсилення впливу найбільш достовірних прогнозів виконується нелінійне перетворення кожної ймовірності за допомогою функції Гомперца:

$$R_{i,k} = 1 - e^{-e^{-\alpha p_{i,k}}}. \quad (2)$$

Функція перетворює ймовірності на ранги штрафу.

Чим вище впевненість моделі, тобто чим ближче ймовірність $p_{i,k}$ до 1, тим ближчим є ранг до мінімального значення $R_{\min}$. Значення $R_{\min}$ залежить від параметру α , наприклад при $\alpha=3$, $R_{\min} \approx 0.048$.

Чим нижче впевненість моделі, тобто чим ближче ймовірність $p_{i,k}$ до 0, тим ближчим є ранг до максимального значення $R_{\max} = 1 - e^{-e^0} \approx 0.632$.

Для вимірювання ступеня невпевненості передбачень i -го класифікатора обчислюється нормалізована інформаційна ентропія Шеннона $E_i \in [0,1]$:

$$E_i = - \frac{\sum_{k=1}^C p_{i,k} \cdot \ln(p_{i,k} + \varepsilon)}{\ln(C)}, \quad (3)$$

де $\varepsilon = 10^{-12}$ – малий регуляризаційний додатак для запобігання обчисленню $\ln(0)$, а $\ln(C)$ – максимально можлива ентропія для C класів.

На відміну від вже запропонованих рішень, які використовують фіксоване значення Тор-К, тобто кількості класів, що братимуть участь у подальшому агрегуванні (для кожного об'єкта x), у запропонованому методі Тор-К визначається автоматично.

Максимально допустиме значення Тор-К визначається як:

$$K_{max} = \max(2, \log_2(C)). \quad (4)$$

Тоді для кожного прогнозу обчислюється Тор-К:

$$K_i = \text{round}(E_i \cdot (K_{max} - 1)) + 1, \quad K_i \in [1, K_{max}] \quad (5)$$

Таким чином, кількість класів автоматично збільшується зі зростанням невизначеності прогнозу.

Позначимо через T_i множину індексів що мають найкращі (мінімальні) значення функції Гомперца. Для кожного об'єкту їхня кількість дорівнює K_i .

Формується скоригована матриця рангів:

$$\widehat{R}_{i,k} = \begin{cases} R_{i,k}, & \text{якщо } k \in T_i \\ R_{max}, & \text{якщо } k \notin T_i \end{cases} \quad (6)$$

Таким чином, у подальшому агрегуванні беруть участь лише найбільш інформативні класи, тоді як внесок інших суттєво зменшується.

Аналогічно формується скоригована матриця ймовірностей:

$$\widehat{p}_{i,k} = \begin{cases} p_{i,k}, & \text{якщо } k \in T_i \\ 0, & \text{якщо } k \notin T_i \end{cases} \quad (7)$$

На основі скоригованої матриці нечітких рангів визначається сумарний нечіткий ранг для кожного класу:

$$RS_k = \sum_{i=1}^N \widehat{R}_{i,k}. \quad (8)$$

Також обчислюється показник узгодженої впевненості ансамблю:

$$CFS_k = 1 - \frac{1}{N} \sum_{i=1}^N \widehat{p}_{i,k}. \quad (9)$$

Фінальна агрегована оцінка для класу k є добутком сумарного нечіткого рангу та показнику узгодженої впевненості ансамблю.

$$FS_k = RS_k \times CFS_k. \quad (10)$$

Оскільки кращі класи отримують найменші значення штрафів Гомперца та найменші значення CFS, остаточне класифікаційне рішення $\hat{y}$ приймається за правилом мінімізації підсумкової оцінки:

$$\hat{y} = \arg \min_{c_k \in C} (FS_k). \quad (11)$$

Розглянемо приклад розрахунку для одного екземпляру з набору даних RAVDESS для ансамблю, який складається з $n = 3$ (MLP, KNN, SVM) базових класифікаторів, кількість класів емоційних станів $c = 7$ класів: angry (0), disgust (1), fearful (2), happy (3), neutral (4), sad (5), surprised (6).

Розглянемо передбачення базових класифікаторів:

MLP: $p_1 = [0.0128, 0.1901, 0.2995, 0.0214, 0.0751, 0.3537, 0.0474]$

KNN: $p_2 = [0, 0, 0.3333, 0, 0, 0, 0.6667]$

SVM: $p_3 = [0.0208, 0.1132, 0.4007, 0.0553, 0.0681, 0.2546, 0.0874]$

При цьому справжній клас $y = 2$ (fearful).

Крок 1. Розрахунок ентропії та автоматичний вибір Top-K

Для $C=7$ маємо $\ln(7) \approx 1.9459$ та $K_{max} = \max(2, \log_2(7)) = 3$.

Для MLP: $E \approx 1.5213/1.9459 \approx 0.7818$.

Нормалізована ентропія висока.

$$K = \text{round}(0.7818 \cdot 2) + 1 = 2 + 1 = 3$$

Вибрано Top-3 класи: клас 5 (0.3537), клас 2 (0.2995), клас 1 (0.1901).

Для KNN: $E \approx 0.6365/1.9459 \approx 0.3271$

Нормалізована ентропія висока.

$$K = \text{round}(0.3271 \cdot 2) + 1 = 1 + 1 = 2$$

Вибрано Top-2 класи: клас 6 (0.6667) та клас 2 (0.3333).

Для SVM: $E \approx 1.5983/1.9459 \approx 0.8214$

Нормалізована ентропія висока.

$$K = \text{round}(0.8214 \cdot 2) + 1 = 2 + 1 = 3$$

Вибрано Top-3 класи: клас 2 (0.4007), клас 5 (0.2546), клас 1 (0.1132).

Крок 2. Перетворення Гомперца та матриця штрафів

Обчислюємо ранги Гомперца для вибраних класів:

Для MLP: $R_{1,5}(0.3537) = 0.293$, $R_{1,2}(0.2995) = 0.335$, $R_{1,1}(0.1901) = 0.432$

Для KNN: $R_{2,6}(0.6667) = 0.127$, $R_{2,2}(0.3333) = 0.308$.

Для SVM: $R_{3,2}(0.4007) = 0.260$, $R_{3,5}(0.2546) = 0.372$, $R_{3,1}(0.1132) = 0.509$

Усі інші класи (відкинуті ентропійним фільтром) отримують базовий штраф 0.632.

Обчислюємо суму рангів для класів-лідерів:

Клас 2: $RS_2 = 0.335_{NN} + 0.308_{KNN} + 0.260_{SVM} = 0.903$

Клас 5: $RS_5 = 0.293_{NN} + 0.632_{KNN} + 0.372_{SVM} = 1.297$

Клас 6: $RS_6 = 0.632_{NN} + 0.127_{KNN} + 0.632_{SVM} = 1.391$

Клас 1: $RS_1 = 0.432_{NN} + 0.632_{KNN} + 0.509_{SVM} = 1.573$

Крок 3. Розрахунок показнику узгодженої впевненості ансамблю

Клас 2: $CFS_2 = 1 - \frac{0.2995+0.3333+0.4007}{3} = 1 - 0.3445 = 0.6555$

Клас 5: $CFS_5 = 1 - \frac{0.3537+0+0.2546}{3} = 1 - 0.2028 = 0.7972$

Клас 6: $CFS_6 = 1 - \frac{0+0.6667+0}{3} = 1 - 0.2222 = 0.7778$

Клас 1: $CFS_1 = 1 - \frac{0.1901+0+0.1132}{3} = 1 - 0.1011 = 0.8989$

Крок 4. Підсумкова оцінка FS

Клас 2: $FS_2 = 0.903 \times 0.6555 = 0.592$

Клас 5: $FS_5 = 1.297 \times 0.7972 = 1.034$

Клас 6: $FS_6 = 1.391 \times 0.7778 = 1.082$

Клас 1: $FS_1 = 1.573 \times 0.8989 = 1.414$

Найменше значення фінальної агрегованої оцінки отримав клас 2. Отже, результуючий прогноз ансамблю – клас 2 – fearful.

Цей приклад яскраво доводить стійкість розробленого методу.

Наприклад, для класу 6 модель KNN видала високу ймовірність (0.6667). У класичному методі Soft Voting цей голос мав би велику вагу, через що Клас 6 створив би небезпечну конкуренцію правильному рішенню (Класу 2). Проте розроблений адаптивний нечіткий агрегатор діє інакше. Оскільки моделі MLP та SVM взагалі не розпізнали патернів Класу 6 (їхні ймовірності 0.04 та 0.08 не пройшли ентропійний фільтр), вони автоматично призначили Класу 6 штрафний ранг 0.632. У результаті сумарний ранг Класу 6 зріс до 1.391, що надійно відкинуло його з числа лідерів (показник узгодженої впевненості ансамблю=1.082). Отже, запропонований метод, який визначає кількість найбільш

інформативних класів на основі нормалізованої ентропії у поєднанні з функцією Гомперца математично ізолює та пригнічує «впевнені помилки» окремих класифікаторів, забезпечуючи стійке фінальне рішення.

4. Експериментальне дослідження та аналіз результатів методу адаптивної нечіткої агрегації на основі функції Гомперца для ансамблевої класифікації аудіоданих емоцій людини

У цьому розділі наведено результати експериментального дослідження запропонованого методу адаптивної нечіткої агрегації на основі функції Гомперца для ансамблевої класифікації аудіоданих емоцій людини. Експерименти проведено на наборі даних RAVDESS.

Для попередньої обробки аудіосигналів використовувалася бібліотека Librosa. На основі аудіозаписів формувалося табличне представлення спектральних характеристик. Далі виконувалась Min-Max нормалізація.

Для забезпечення відтворюваності експериментів під час формування вибірок використовувалося фіксоване значення `random_state = 5`.

Після підготовки даних виконувалося первинне налаштування гіперпараметрів базових класифікаторів. Метою цього етапу було визначення робочих конфігурацій моделей перед проведенням відбору ознак. До складу базових моделей включено SVM, KNN, XGBoost, Random Forest, Logistic Regression, MLP, CNN та LSTM. Отримані конфігурації використовувалися для оцінювання впливу окремих ознак на якість класифікації під час подальшої процедури BFE.

Після первинного налаштування базових класифікаторів виконувалася процедура BFE для формування оптимізованого простору спектральних ознак.

Процедура BFE застосовувалася до базових моделей, що працюють із табличним представленням аудіоданих: SVM, MLP, XGBoost, Random Forest, KNN та Logistic Regression. CNN та LSTM не включалися до процедури BFE, оскільки вони використовують відповідно спектральне представлення у вигляді зображень та часові послідовності MFCC.

Важливо, що незалежна тестова вибірка не використовувалася для визначення складу оптимізованого простору ознак. Остаточний набір ознак для кожного базового класифікатора визначався виключно на навчально-валідаційній частині даних. Фінальні набори ознак, сформовані процедурою BFE для кожного базового алгоритму, наведено в Таблиці 1. Кожній комбінації {ознака, алгоритм} відповідає значення 1, якщо ознака залишилась після проведення BFE та 0, якщо була видалена.

Після завершення процедури BFE та фіксації оптимізованих просторів ознак виконувалося фінальне налаштування гіперпараметрів базових класифікаторів на тренувально-валідаційному сегменті даних.

Фінальні параметри KNN: `n_neighbors = 3`, `weights = uniform`.

Фінальні параметри SVM: `C = 2`, `kernel = poly`, `degree = 2`

Фінальні параметри XGBoost: `n_estimators = 40`, `min_child_weight = 5`, `max_depth = 4`, `learning_rate = 0,1`, `gamma = 0,4`

Фінальні параметри Random Forest: `n_estimators = 500`, `max_depth = 15`, `min_samples_leaf = 4`, `min_samples_split = 8`, `max_features = sqrt`, `class_weight = balanced_subsample`.

Фінальні параметри Logistic Regression: `solver = liblinear`, `penalty = l1`, `C = 16`, `class_weight = None`.

Фінальна архітектура MLP: включала три повнозв'язні приховані шари з 256, 128 та 64 нейронами відповідно. Після кожного прихованого шару застосовувалися ReLU та Dropout із коефіцієнтом 0,2. Для ваг прихованих шарів використовувалася L2-регуляризація з коефіцієнтом 0,0001. Вихідний шар містив 7 нейронів із softmax. Навчання виконувалося оптимізатором Adam із функцією втрат `sparse categorical crossentropy`, розміром пакета 64 та максимальною кількістю 200 епох. Для запобігання перенавчанню застосовувалася рання зупинка за `val_loss` із `min_delta = 0,001`, `patience = 15` та відновленням найкращих ваг.

Фінальна архітектура CNN: складалася з трьох послідовних згорткових блоків із 32, 64 та 128 фільтрами розміром 3×3 . У кожному блоці після згортки застосовувалися Batch Normalization, ReLU та MaxPooling2D із розміром вікна 2×2 . Після згорткової частини використовувався повнозв'язний блок із шарами на 128 та 64 нейрони, Batch Normalization, ReLU та Dropout із коефіцієнтами 0,5 та 0,4 відповідно. Вихідний шар містив 7 нейронів із softmax. Навчання CNN виконувалося також оптимізатором Adam із функцією втрат sparse categorical crossentropy, розміром пакета 32 та максимальною кількістю 100 епох. Рання зупинка виконувалася за val_loss із patience = 15 та відновленням найкращих ваг.

Таблиця 1. Перелік спектральних ознак, відібраних процедурою BFE для базових класифікаторів

№	Назва предиктора	SVM	XGBoost	KNN	Random Forest	Logistic regression	Multilayer Perceptron
1	mfcc_0	1	1	1	1	1	1
2	mfcc_1	1	1	1	1	1	1
3	mfcc_2	1	1	1	1	1	1
4	mfcc_3	1	1	1	1	1	1
5	mfcc_4	1	1	1	1	1	1
6	mfcc_5	1	1	1	1	1	1
7	mfcc_6	1	1	0	1	1	1
8	mfcc_7	1	1	1	1	1	1
9	mfcc_8	1	1	1	1	1	1
10	mfcc_9	1	1	1	1	1	1
11	mfcc_10	1	1	1	1	1	1
12	mfcc_11	1	1	1	1	1	1
13	mfcc_12	1	1	1	1	1	1
14	mfcc_13	1	1	1	1	1	1
15	mfcc_14	1	1	0	1	1	1
16	mfcc_15	0	1	1	1	1	1
17	mfcc_16	1	1	0	1	1	1
18	mfcc_17	1	1	0	1	0	1
19	mfcc_18	1	1	1	1	1	1
20	mfcc_19	1	1	1	1	1	1
21	spectral_contrast_0	1	1	1	1	1	1
22	spectral_contrast_1	0	1	1	1	1	1
23	spectral_contrast_2	1	1	1	1	1	1
24	spectral_contrast_3	0	1	1	1	1	1
25	spectral_contrast_4	1	1	1	1	1	1
26	spectral_contrast_5	1	1	0	1	1	1
27	spectral_contrast_6	1	1	1	1	1	0
28	spectral_centroid	1	1	0	1	0	0
29	spectral_flatness	1	1	1	1	1	1
30	spectral_rolloff	1	1	0	1	1	1
31	spectral_bandwidth	1	1	1	1	1	1
32	zero_crossing_rate	0	1	0	1	1	1
33	gender	1	0	0	1	1	1

Джерело: розроблено авторами

Фінальна архітектура LSTM: складалася з Batch Normalization для вхідної послідовності розміром 125×20 та шар Gaussian Noise із параметром 0,1. Для зменшення часової розмірності застосовувався AveragePooling1D із pool_size = 4. Основний рекурентний блок складався з двонапрявленого LSTM-шару із 48 нейронами, return_sequences = True, dropout = 0,2, recurrent_dropout = 0,2 та L2-регуляризацією з коефіцієнтом 0,002. Після нього застосовувалися Batch Normalization, Dropout 0,4 та GlobalAveragePooling1D. Повнозв'язний шар містив 32 нейрони з ReLU та L2-регуляризацією 0,002, після чого застосовувалися Batch Normalization та Dropout 0,3. Вихідний шар містив 7 нейронів із softmax. Навчання виконувалося також оптимізатором Adam із швидкістю навчання 0,0007, функцією втрат sparse categorical crossentropy, розміром пакета 32 та максимальною кількістю 150 епох. Для запобігання перенавчанню також використовувалася рання зупинка за val_loss із min_delta = 0,001, patience = 15 та відновленням найкращих ваг.

Оцінювання ефективності запропонованого підходу виконано у два етапи. На першому етапі оцінюється вплив процедури Backward Feature Elimination (BFE) на якість роботи окремих базових класифікаторів. На другому етапі оцінюється ефективність ансамблевої класифікації шляхом порівняння запропонованої адаптивної агрегації на основі функції Гомперца з окремими базовими класифікаторами та стандартною агрегацією на основі функції Гомперца з фіксованим значенням Тор-К.

На першому етапі проведено порівняння якості класифікації окремих моделей машинного навчання до та після застосування процедури Backward Feature Elimination.

Порівняння виконано для класифікаторів, що працюють із табличним представленням аудіоданих: SVM, MLP, XGBoost, Random Forest, KNN та Logistic Regression. CNN та LSTM не включалися до порівняння впливу BFE, оскільки ці моделі працюють відповідно зі спектром та часовими послідовностями MFCC, а не з табличним представленням спектральних ознак.

Процедура BFE виконувалася виключно на навчально-валідаційній частині даних. На кожній ітерації оцінювалася доцільність вилучення окремих ознак за результатами на валідаційній вибірці, після чого формувався оптимізований простір ознак. Незалежна тестова вибірка не використовувалася для визначення списку відібраних ознак. Після завершення процедури відбору сформований простір ознак фіксувався, на незалежній тестовій вибірці було проведено порівняння результатів базових класифікаторів до та після застосування BFE (тобто з повним та оптимізованим простором ознак).

Для кожного класифікатора якість класифікації оцінювалася за метриками Accuracy та F1-score. Результати порівняння до та після застосування BFE на незалежній тестовій вибірці наведено у Таблиці 2.

Таблиця 2. Accuracy та F1-score базових класифікаторів до та після BFE

Алгоритм	Accuracy			F1-score		
	Повний набір ознак, %	Ознаки після BFE, %	Різниця, %	Повний набір ознак, %	Ознаки після BFE, %	Різниця, %
SVM	66.94	65	-1.94	66.55	64.9	-1.65
MLP	64.17	65.83	1.67	64.03	65.66	1.63
XGB	57.78	58.89	1.11	57.08	58.19	1.11
RF	0.55	58.89	3.89	53.19	58.05	4.85
KNN	61.67	64.72	3.06	60.59	63.91	3.32
LR	60.83	58.89	-1.94	59.81	57.81	-2.00

Джерело: розроблено авторами

Отримані результати демонструють вплив процедури відбору ознак на якість роботи окремих класифікаторів. Для більшості моделей видалення надлишкових та менш інформативних ознак призводить до підвищення значень Accuracy та F1-score. Це підтверджує, що інформативність окремих спектральних ознак залежить від алгоритму класифікації, а використання єдиного повного простору ознак не обов'язково є оптимальним для всіх базових моделей.

Покращення, отримане після застосування BFE, використовується як основа для подальшого формування ансамблю. На наступному етапі дослідження оптимізовані конфігурації базових класифікаторів використовуються для отримання прогнозів, які передаються до процедури ансамблевої агрегації.

На другому етапі проведено оцінювання ефективності запропонованого ансамблевого підходу. На відміну від першого етапу, метою якого було дослідження впливу відбору ознак, на цьому етапі основна увага приділяється формуванню ансамблів та оцінюванню запропонованого способу агрегації їхніх прогнозів.

Для проведення експерименту було сформовано множину ансамблевих класифікаторів, що відрізнялися складом базових моделей. Для кожного типу агрегації було побудовано та оцінено 219 ансамблевих класифікаторів. Таким чином, окремо досліджено 219 ансамблів зі

стандартною нечіткою агрегацією на основі функції Гомперца з фіксованим Тор-К та 219 ансамблів із запропонованою адаптивною нечіткою агрегацією на основі функції Гомперца з адаптивним Тор-К. Значення ($K=3$) було обрано як фіксовану базову конфігурацію з урахуванням компромісу між концентрацією на найбільш імовірних класах та збереженням альтернативних варіантів класифікації. За наявності семи класів врахування трьох найвищих за рангом класів відповідає $\sim 43\%$ класів: на відміну від ($K=1$), такий підхід не обмежується лише найбільш імовірним класом, водночас значення ($K>3$) призводять до врахування більшої частини класів і, відповідно, до меншої концентрації на найбільш релевантних альтернативах. Тому ($K=3$) використовується як фіксоване базове значення для порівняння зі запропонованим адаптивним механізмом визначення Тор-К.

Відбір найкращого ансамблевого класифікатора для кожного типу агрегації здійснювався виключно на валідаційній вибірці. Для кожного з 219 сформованих ансамблів оцінювалися відповідні значення метрик якості, після чого обирався найкращий ансамбль.

Після завершення процедури відбору склад найкращого ансамблю та параметри відповідного методу агрегації фіксувалися. Тестова вибірка не використовувалася для вибору найкращого ансамблю або налаштування його параметрів. Фінальне оцінювання відібраних ансамблів виконувалося на незалежній тестовій вибірці.

Такий підхід дає змогу уникнути оптимістичного зміщення результатів, пов'язаного з вибором найкращої моделі безпосередньо за результатами тестування. Отже, наведені нижче результати характеризують якість ансамблів, попередньо відібраних за валідаційною вибіркою, на незалежних тестових даних.

Для кожного аудіо-зразка базові класифікатори формують вектори ймовірнісних оцінок належності до емоційних класів. Отримані значення використовуються для формування підсумкового рішення ансамблю.

У межах експерименту порівнюються три типи результатів:

- 1) окремі базові класифікатори після BFE;
- 2) найкращий ансамбль зі стандартною нечіткою агрегацією на основі функції Гомперца з фіксованим Тор-К;
- 3) найкращий ансамбль зі запропонованою адаптивною нечіткою агрегацією на основі функції Гомперца з адаптивним Тор-К.

Стандартна агрегація на основі функції Гомперца використовується як базовий метод для оцінювання ефективності запропонованої модифікації. В обох випадках використовується однакова загальна концепція нечіткого ранжування на основі функції Гомперца, проте відрізняється спосіб визначення кількості рангів, що беруть участь у подальшій агрегації.

Таке порівняння дає змогу оцінити внесок запропонованого механізму адаптивного визначення Тор-К, а не лише загальний ефект від використання ансамблю.

Якість усіх розглянутих методів оцінюється за метриками Ассурасу та зваженого F1-score.

Серед усіх побудованих ансамблів зі стандартною нечіткою агрегацією на основі функції Гомперца з фіксованим Тор-К найкращий результат показав ансамбль з KNN, SVM, XGB, CNN, LSTM.

Результати порівняння базових класифікаторів після BFE та ансамблю із стандартною агрегацією на основі функції Гомперца наведені в Таблиці 3.

Згідно з отриманими результатами, найкращий ансамблевий класифікатор зі стандартною агрегацією на основі функції Гомперца отримав на тестовій вибірці наступні значення метрик: Ассурасу = 81.39 %, F1-score = 81.20 %. Таким чином, ансамблеве рішення перевищує результати кожного з окремих базових класифікаторів.

Найкращий результат серед окремих базових класифікаторів демонструє LSTM, для якого Ассурасу = 80.28 %, а F1-score = 80.17 %. Перевага стандартного ансамблю над LSTM становить відповідно 1.11 % та 1.03 %.

Отримані результати підтверджують доцільність об'єднання прогнозів гетерогенних базових моделей. Водночас відносно невелика перевага стандартної агрегації над LSTM показує, що саме використання ансамблю ще не гарантує суттєвого покращення порівняно з найсильнішою окремою моделлю. Тому наступним етапом є оцінювання запропонованого механізму адаптивної агрегації.

Таблиця 3. Порівняння базових класифікаторів та ансамблю із стандартною агрегацією на основі функції Гомперца

Алгоритм	Метрики для ансамблю			
	Accuracy, %	F1-score, %	Різниця Accuracy, %	Різниця F1-score, %
Ансамбль	81.39	81.2		
SVM	65	64.9		
MLP	65.83	65.66	15.56	15.52
XGB	58.89	58.19	22.50	22.99
RF	58.89	58.05	22.50	23.13
KNN	64.72	63.91	16.67	17.28
LR	58.89	57.81	22.50	23.38
CNN	70.00	69.63	11.39	11.55
LSTM	80.28	80.17	1.11	1.02

Джерело: розроблено авторами

Серед усіх побудованих ансамблів зі запропонованою адаптивною нечіткою агрегацією на основі функції Гомперца з адаптивним Тор-К найкращий результат показав ансамбль з KNN, SVM, CNN, LSTM.

Результати порівняння базових класифікаторів із найкращим ансамблем із застосуванням запропонованої нечіткої агрегації на основі функції Гомперца з адаптивним Тор-К, наведено в Таблиці 4.

Таблиця 4. Порівняння базових класифікаторів та ансамблю із запропонованою адаптивно нечіткою агрегацією на основі функції Гомперца з адаптивним Тор-К

Алгоритм	Метрики для ансамблю			
	Accuracy, %	F1-score, %	Різниця Accuracy, %	Різниця F1-score, %
Ансамбль	85.56	85.55		
SVM	65	64.9	20.56	20.65
MLP	65.83	65.66	19.72	19.89
XGB	58.89	58.19	26.67	27.36
RF	58.89	58.05	26.67	27.50
KNN	64.72	63.91	20.83	21.64
LR	58.89	57.81	26.67	27.74
CNN	70.00	69.63	15.56	15.92
LSTM	80.28	80.17	5.28	5.38

Джерело: розроблено авторами

Згідно з отриманими результатами, найкращий ансамблевий класифікатор зі запропонованою агрегацією на основі функції Гомперца отримав на тестовій вибірці наступні значення метрик: Accuracy = 85.56 %, F1-score = 85.55 % на незалежній тестовій вибірці. Таким чином, запропонований ансамблевий класифікатор перевищує результати всіх окремих базових моделей. Порівняно з LSTM, перевага ансамблю по Accuracy та F1-score складає 5.28 % та 5.38 %.

Перехід від фіксованого Тор-К до адаптивного Тор-К дозволив підвищити Accuracy найкращого ансамблю з 81.39 % до 85.56 %, тобто на 4.17 %. Відповідні значення F1-score збільшились з 81.20 % до 85.55 %, тобто на 4.35 %.

Для оцінювання статистичної надійності отриманого приросту точності класифікації застосовано метод парного бутстрепу. На основі прогнозів стандартного та адаптивного методів агрегації на одному тестовому наборі сформовано 1000 бутстреп-вибірок із поверненням. Для кожної вибірки обчислювалася різниця значень Accuracy та F1-score між методами. Статистичну невизначеність оцінювали за допомогою 95 % довірчих інтервалів, визначених на основі 2,5-го та 97,5-го процентилів емпіричного розподілу різниць метрик.

За результатами експерименту значення Accuracy стандартного методу становило 81,39%, тоді як адаптивного – 85,56 %, що відповідає приросту на 4,17 %. 95 % довірчий інтервал для різниці Accuracy становив [1,94 %; 6,94 %]. Значення F1-score зросло з 81,18 % до 85,55%, а приріст склав 4,37 % за 95 % довірчого інтервалу [1,99 %; 6,94 %]. Оскільки обидва довірчі інтервали не містять нульового значення, отримані результати свідчать про статистично значущу перевагу адаптивного методу за обома метриками на досліджуваному тестовому наборі. Такий підхід дав змогу оцінити стійкість виявлених відмінностей без повторного навчання моделей.

Для детального аналізу результатів класифікації окремих класів найкращого ансамблевого класифікатора із запропонованою агрегацією на основі функції Гомперца побудовано confusion matrix (Таблиця 5), яка відображає розподіл правильних і помилкових прогнозів для кожного із семи класів емоцій: Angry, Disgust, Fearful, Happy, Neutral, Sad та Surprised. Для кожного класу визначено значення Precision, Recall та F1-score, що дає змогу оцінити ефективність розпізнавання окремих емоцій (Таблиця 6).

Таблиця 5. Confusion Matrix

Клас	Angry	Disgust	Fearful	Happy	Neutral	Sad	Surprised
Angry	42	2	1	1	1	0	1
Disgust	1	41	1	2	2	1	0
Fearful	0	1	39	2	3	3	0
Happy	0	0	6	38	1	1	2
Neutral	0	0	0	0	71	1	0
Sad	0	0	1	1	8	37	1
Surprised	0	1	1	1	5	1	39

Джерело: розроблено авторами

Таблиця 5. Індивідуальні метрики для класів

Клас	Precision, %	Recall, %	F1-score, %
Angry	97.7	89.6	93.5
Disgust	91.3	87.5	89.4
Fearful	76.5	81.2	78.8
Happy	82.6	79.2	80.9
Neutral	80.5	97.2	88.1
Sad	84.8	81.2	83.0
Surprised	92.5	77.1	84.1

Джерело: розроблено авторами

Отримані результати свідчать про високу ефективність запропонованого методу адаптивної нечіткої агрегації не лише за узагальненими метриками, а й під час розпізнавання окремих класів емоцій. Значення F1-score для всіх семи класів перевищує 78 %, а для чотирьох класів становить 83 % і вище. Особливо високі результати отримано для класів Angry 93.5 %, Disgust 89.4 % та Neutral 88.1 %.

Для оцінювання впливу запропонованого механізму адаптивного визначення Тор-К було додатково проведено порівняння стандартної та адаптивної нечіткої агрегації на однаковому складі базових класифікаторів. Такий підхід дає змогу виключити вплив зміни складу ансамблю та оцінити саме відмінність між фіксованим і адаптивним визначенням кількості класів, що беруть участь в агрегації. Порівняння було проведено на незалежній тестовій вибірці.

В обох випадках використовувався ансамбль, що складався з KNN, SVM, XGB, CNN та LSTM (найкраща комбінація серед ансамблів зі стандартною нечіткою агрегацією на основі функції Гомперца з фіксованим Тор-К). Таким чином, склад ансамблю, навчені моделі та їхні ймовірнісні прогнози залишалися незмінними, а відмінність між двома варіантами полягала лише у способі визначення кількості класів, що враховуються під час агрегації. Для ансамблю зі складом KNN, SVM, XGB, CNN та LSTM адаптивна агрегація забезпечила значення Accuracy 83.89 % та F1-score 83.77 %. Перехід від фіксованого Тор-К до адаптивного Тор-К дозволив підвищити Accuracy на 2.5 %, а F1-score на 2.57 %.

Таким чином, результати експерименту демонструють послідовний ефект від використання запропонованого підходу. Спочатку формування ансамблю гетерогенних класифікаторів забезпечує покращення порівняно з окремими моделями. Подальша заміна стандартної агрегації Гомперца з фіксованим Тор-К на запроповану адаптивну агрегацію забезпечує додаткове підвищення як Accuracy, так і F1-score.

Важливо, що всі наведені в Таблиці 3 та в Таблиці 4 результати отримано на незалежній тестовій вибірці після завершення процедури вибору ансамблю. Таким чином, тестова вибірка не використовувалася для визначення того, який із сформованих ансамблів є найкращим. Це забезпечує більш коректну оцінку узагальнювальної здатності запропонованого методу та зменшує ризик оптимістичного зміщення результатів унаслідок post-hoc вибору моделі за тестовими даними.

Висновки. У роботі досліджено підхід до ансамблевої класифікації аудіоданих емоцій людини на основі адаптивної нечіткої агрегації з використанням функції Гомперца. Експериментальне дослідження проведено на наборі даних RAVDESS та включало два послідовні етапи: оцінювання впливу процедури BFE на базові класифікатори та дослідження ефективності ансамблевої агрегації.

Застосування BFE продемонструвало, що відбір ознак по-різному впливає на окремі алгоритми. Для MLP, XGBoost, Random Forest та KNN спостерігалось покращення Accuracy та F1-score, тоді як для SVM та Logistic Regression після скорочення простору ознак якості дещо знизилася. Це підтверджує залежність ефективності простору ознак від використовуваного алгоритму класифікації та обґрунтовує застосування процедури відбору ознак під час побудови ансамблевих моделей.

На другому етапі для кожного з двох способів агрегації було сформовано по 219 ансамблевих класифікаторів, що відрізнялися складом базових моделей. Найкращий ансамбль для кожного типу агрегації визначався результатами на валідаційній вибірці. Після вибору структура ансамблю та параметри агрегації фіксувалися, а його остаточна оцінка проводилася на незалежній тестовій вибірці. Така схема експерименту дозволила уникнути вибору найкращого ансамблю безпосередньо за результатами навчання та зменшити ризик оптимістичного post-hoc зміщення.

Порівняння на тестовій вибірці показало, що стандартна нечітка агрегація на основі функції Гомперца з фіксованим Тор-К забезпечила Accuracy = 81.39 % та F1-score = 81.20 %. При цьому найкращий окремий базовий класифікатор LSTM продемонстрував 80.28 % Accuracy та 80.17 % F1-score, що свідчить про обмежений, але позитивний ефект.

Застосування запропонованої адаптивної нечіткої агрегації на основі функції Гомперца дозволило підвищити якість класифікації до Accuracy = 85.56 % та F1-score = 85.55 %. Таким чином, порівняно зі стандартною агрегацією приріст становив відповідно 4.17 % та 4.35 %. Порівняно з найкращим окремим базовим класифікатором LSTM приріст склав 5.28 % та 5.38%. Для оцінювання статистичної надійності отриманого приросту точності класифікації застосовано метод парного бутстрепу. Також було побудовано Confusion Matrix та проаналізована якість класифікації на рівні окремих класів.

Також було проведено експеримент з порівнянням ансамблю зі стандартною агрегацією та адаптивною на однаковому складі ансамблів (KNN, SVM, XGB, CNN та LSTM). Перехід від фіксованого до адаптивного Тор-К підвищив Accuracy з 81.39% до 83.89% (+2.5 %), а F1-score – з 81.20 % до 83.77 % (+2.57 %).

Отримані результати свідчать, що заміна фіксованого Тор-К на адаптивне визначення кількості рангів дозволяє ефективніше використовувати ймовірнісні оцінки базових класифікаторів під час формування підсумкового рішення. Таким чином, запропонований механізм адаптивного Тор-К підвищує ефективність нечіткої агрегації на основі функції Гомперца та є перспективним підходом для ансамблевої класифікації аудіоданих емоцій людини.

Конфлікт інтересів: Автори декларують, що не мають конфлікту інтересів, зокрема фінансового, особистого, авторського чи будь-якого іншого характеру, який міг би вплинути на дослідження, а також на результати, опубліковані в цій статті

Фінансування: Дослідження проводилося без фінансової підтримки

Доступність даних: Усі дані доступні в цифровій або графічній формі в основному тексті рукопису

Використання засобів штучного інтелекту: Під час підготовки рукопису використовувався інструмент Gemini 3.5 Flash-літ штучного інтелекту (ШІ) для граматичної та стилістичної перевірки тексту. Усі фрагменти, опрацьовані за його допомогою, були перевірені й відредаговані автором. ШІ не використовувався для генерації тексту, формул і таблиць, формування наукових положень, отримання результатів або формулювання висновків. Наукові положення, результати та висновки є власним інтелектуальним внеском авторів

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Arsirii, O. O.; Petrosiuk, D. V. “An adaptive convolutional neural network model for human facial expression recognition”. *Herald of Advanced Information Technology*. 2023; 6 (2): 128–138. DOI: <https://doi.org/10.15276/hait.06.2023.8>.
2. Kakuba, S., & Han, D. S. “Addressing data scarcity in speech emotion recognition: A comprehensive review”. *ICT Express*. 2025; 11 (1): 110–123. DOI: <https://doi.org/10.1016/j.icte.2024.11.003>.
3. Gudivaka, S. R., Polipogu, P., Kodali, R., Dhulipalla, V. R., & Tatavarty, V. S. K. “Speech emotion recognition in adults and children: a comprehensive review of traditional features and raw waveform models”. *International Journal of Speech Technology*. 2026; 29: 21, <https://www.scopus.com/pages/publications/105027451592>. DOI: <https://doi.org/10.1007/s10772-025-10229-6>.
4. Ahn, Y., Han, S., Lee, S., & Shin, J. W. “Speech emotion recognition incorporating relative difficulty and labeling reliability”. *Sensors*. 2024; 24 (13): 4111. DOI: <https://doi.org/10.3390/s24134111>.
5. Andronati, O., Antoshchuk, S., Babilunha, O., Arsirii, O., Nikolenko, A. & Mikhalev, K. “A method of constructing ensemble classifiers for recognizing audio data of various nature”. *14th International Conference on Advanced Computer Information Technologies (ACIT)*. Ceske Budejovice, Czech Republic. 2024. p. 758–761, <https://www.scopus.com/authid/detail.uri?authorId=58677655800>. DOI: <https://doi.org/10.1109/ACIT62333.2024.10712469>.
6. Kundu, R., Basak, H., Singh, P. K., Ahmadian, A., Ferrara, M., et al. “Fuzzy rank-based fusion of CNN models using Gompertz function for screening COVID-19 CT-scans”. *Scientific Reports*. 2021; 11: 14133. DOI: <https://doi.org/10.1038/s41598-021-93658-y>.
7. Dhara, T., Singh, P.K. & Mahmud, M. “A fuzzy ensemble-based deep learning model for EEG-based emotion recognition”. *Cogn Comput*. 2024; 16: 1364–1378. DOI: <https://doi.org/10.1007/s12559-023-10171-2>
8. “RAVDESS”. – Available from: <https://www.kaggle.com/datasets/uwrfkaggler/ravdess-emotional-speech-audio>. – [Accessed: Jun 2026].
9. McFee, B., Raffel, C., Liang, D., Ellis, D. P. W., McVicar, M., Battenberg, E. & Nieto, O. “librosa: Audio and music signal analysis in python.” In *Proceedings of the 14th Python in Science Conference*. 2015. p. 18–25. DOI: <https://doi.org/10.25080/Majora-7b98e3ed-003>.
10. Thukroo, I. A., Bashir, R. & Giri, K. J. “A comparison of cepstral and spectral features using recurrent neural network for spoken language identification”. *Comput. Artif. Intell*. 2024; 2 (1). DOI: <https://doi.org/10.59400/cai.v2i1.440>.
11. Sharma, G., Umapathy, K. & Krishnan, S. “Trends in audio signal feature extraction methods”. *Applied Acoustics*. 2020; 158: 107020. DOI: <https://doi.org/10.1016/j.apacoust.2019.107020>.

12. Platt, J. “Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods”. *Adv. Large Margin Classif.* 2000. – Available from: https://www.researchgate.net/publication/2594015_Probabilistic_Outputs_for_Support_Vector_Machines_and_Comparisons_to_Regularized_Likelihood_Methods.

Отримано 27.08.2026

Отримано після перегляду 02.10.2026

Прийнято 08.10.2026

DOI: <https://doi.org/10.15276/ict.03.2026.49>

UDC 004.451:502/504

Improving the accuracy of speech emotion recognition based on ensemble classification with adaptive fuzzy aggregation using the Gompertz function

Olena O. Arsirii¹⁾

ORCID: <https://orcid.org/0000-0001-8130-9613>; e.arsirii@gmail.com. Scopus Author ID: 54419480900

Oleksandr K. Andronati¹⁾

ORCID: <https://orcid.org/0009-0009-1794-5864>; alex.andronati@gmail.com. Scopus Author ID: 58677655800

¹⁾ Odesa Polytechnic National University, 1, Shevchenko Ave. Odesa, 65044, Ukraine

ABSTRACT

The paper addresses the problem of speech emotion recognition and investigates an approach to improving the efficiency of ensemble classification based on adaptive fuzzy aggregation. A method is proposed for aggregating predictions from machine learning and deep learning classifiers using the Gompertz function, in which the number of the most informative classes is determined adaptively for each object based on the normalized entropy of the probability distribution. Unlike the standard approach with a fixed number of the most informative classes, the proposed method takes into account the level of uncertainty of a specific prediction, which allows for dynamically changing the number of classes involved in further aggregation. An experimental study was conducted on the Ryerson Audio-Visual Database of Emotional Speech and Song dataset. Eight base classifiers, each operating on different representations of audio data, were used to construct a heterogeneous ensemble. For tabular spectral features, the Backward Feature Elimination procedure was additionally applied, which allowed to form individual feature spaces for the base classifiers. As part of the experiment, two hundred and nineteen ensemble classifiers were formed and evaluated with each aggregation method. Standard fuzzy aggregation based on the Gompertz function with a fixed number of the most informative classes achieved an accuracy of eighty-one point three nine percent and an F1-score of eighty-one point two zero percent, whereas the proposed adaptive method achieved an accuracy of eighty-five point five six percent and an F1-score of eighty-five point five five percent. Thus, the adaptive method resulted in a four point one seven percent increase in accuracy and a four point three five percent increase in the F1-score compared to standard fuzzy aggregation.

Keywords: audio emotion recognition; ensemble classification; Gompertz function; adaptive fuzzy aggregation; stacking ensemble; feature selection; machine learning; deep learning; hyperparameter tuning

ІНФОРМАЦІЯ ПРО АВТОРІВ



Арсірій Олена Олександрівна - доктор технічних наук, професор, завідувачка кафедри Інформаційних систем. Національний університет “Одеська Політехніка”, пр. Шевченка, 1. Одеса, 65044, Україна

Галузь досліджень: інформаційні технології; штучний інтелект; системи підтримки прийняття рішень; машинне навчання; нейронні мережі

Olena O. Arsirii - Doctor of Engineering Sciences, Professor, Head of Department of Information Systems. Odesa Polytechnic National University, 1, Shevchenko Ave. Odesa, 65044, Ukraine

ORCID: <https://orcid.org/0000-0001-8130-9613>; e.arsirii@gmail.com. Scopus Author ID 54419480900

Research field: Information technology; artificial intelligence; decision support systems; machine learning; neural networks



Андронаті Олександр Кирилович - аспірант кафедри Інформаційних систем. Національний університет “Одеська Політехніка”, пр. Шевченка, 1. Одеса, 65044, Україна

Галузь досліджень: ансамблеве навчання, машинне навчання, глибоке навчання, нейронні мережі

Oleksandr K. Andronati - graduate student, Department of Information Systems. Odesa Polytechnic National University, 1, Shevchenko Ave. Odesa, 65044, Ukraine

ORCID: <https://orcid.org/0009-0009-1794-5864>; alex.andronati@gmail.com. Scopus Author ID 58677655800

Research field: Data science; machine learning; ensemble classification