

<https://doi.org/10.31891/2219-9365-2023-76-4>

УДК 004

СУХОЛЕЙСТЕР Олександр

Національний Університет «Львівська Політехніка»

<https://orcid.org/0000-0001-7831-522X>

e-mail: oleksandr.sukholeyster@gmail.com

НАКОНЕЧНИЙ Адріан

Національний Університет «Львівська Політехніка»

<https://orcid.org/0000-0002-1873-6337>

e-mail: admakon@gmail.com

СТРАТЕГІЇ ЗЛИТТЯ ФІЗІОЛОГІЧНИХ СИГНАЛІВ ДЛЯ РОЗПІЗНАВАННЯ ЕМОЦІЙ

Дана стаття зосереджена на аналізі та класифікації стратегій розпізнавання емоцій на основі фізіологічних сигналів, а також на дослідженні доступних баз даних для мультимодального аналізу психологічних станів з використанням інформації фізіологічних сигналів. У статті представлено класифікацію методів розпізнавання, проведено аналіз впливів факторів модальності і стратегії злиття на результати розпізнавання та визначено важливість вибору фізіологічних сигналів для точного розпізнавання емоцій. Також досліджено переваги моделей глибокого навчання та практичні застосування розпізнавання емоцій у реальних сценаріях.

Ключові слова: розпізнавання емоцій, бази даних емоцій, фізіологічні сигнали, біосенсори, злиття ознак

SUKHOLEYSTER Oleksandr, NAKONECHNYI Adrian

Lviv Polytechnic National University

BIOSENSOR FUSION STRATEGIES FOR EMOTION RECOGNITION

This article focuses on the analysis and classification of emotion recognition strategies based on physiological signals, as well as on the study of available databases for multimodal analysis of psychological states using information from physiological signals. The article presents a classification of recognition methods, examines the influence of factors such as modalities and fusion strategies onto recognition performance, and determines the importance of choosing physiological signals for accurate emotion recognition. Advantages of deep learning models and practical applications of emotion recognition in real scenarios are also discussed.

Although affective computing systems using unimodal or multimodal data have made significant breakthroughs, there are only a few reliable and efficient algorithms for emotion prediction and sentiment recognition in diverse and complex scenarios. Based on the analysis, the following recommendations can be made for future research in the field of efficient computing.

The development of new and larger basic databases, in particular multimodal databases consisting of different modalities, will be very useful. Conditions should include both spontaneous and nonspontaneous scenarios, providing annotation data in both discrete and dimensional emotion models.

Currently, there is a significant need to improve fusion strategies, especially using rule-based or statistical knowledge, to implement mutual fusion of different modalities that can take into account the role and importance of each modality in emotion recognition.

Since there is a significant need to develop new features that can more accurately reflect complex psychological states, the use of multimodal analysis to create new features or descriptors that take into account the interaction of different physiological signals is particularly desirable.

The use of deep learning methods (for example, neural networks) to create models that can automatically determine psychological state based on multimodal analysis of physiological signals is particularly desirable and necessary.

Keywords: emotion recognition, emotion databases, physiological signals, biosensors, feature fusion

Постановка проблеми у загальному вигляді

та її зв'язок із важливими науковими і практичними завданнями

Дослідницький інтерес до розпізнавання людських емоцій через фізіологічні сигнали тісно пов'язаний з прискореним розвитком технологічних рішень, особливо в галузі охорони здоров'я. Вважається, що сфера розпізнавання емоцій бере свій початок з написаної статті Розалінд Пікард у 1995 році, де обговорюються неврологічні дослідження людських емоцій і можливості комп'ютерів розпізнавати їх шляхом класифікацій вираження цих емоцій [1] та книги, виданої в 1997 році [2], яка започаткувала розпізнавання емоцій як окрему галузь комп'ютерних наук. Розпізнавання емоцій - це міждисциплінарне дослідження на перетині комп'ютерних наук, психології, та фізіології [2]. Багато дослідників вважають, що афективне обчислення є ключем до підвищення та розвитку орієнтованого на людину штучного інтелекту (ШІ) та людського інтелекту. Мультимодальні системи взаємодії людина-комп'ютер забезпечують більш схожу взаємодію між машинами та людьми на взаємодію між людьми. Створення однозначного обміну інформацією між ними робить ці системи більш надійними, ефективними, менш схильними до помилок і здатними вирішувати складні завдання. Ідея використання афективного обчислення в електронному навчанні, маркетингу, безпеці, науках про здоров'я тощо збільшило попит на високоточні системи

розпізнавання емоцій. Використання методів машинного навчання (МН) в останні роки дозволило покращити результати розпізнавання, а наявність відкритих анованих баз даних суттєво знизило поріг для досліджень.

Системи розпізнавання емоцій мають багато потенційних застосувань, охоплюючи охорону здоров'я, розваги, електронне навчання, маркетинг, моніторинг людей і безпеку. Згідно з [3], існувало три основних застосування систем розпізнавання емоцій конкретно за допомогою сигналів ЕКГ:

Моніторинг емоцій людини під час виконання певних завдань та оцінка реагування в критичних ситуаціях. Наприклад, у праці [4] система розпізнавання емоцій фокусується на вивченні продуктивності водія під час гонки.

Клінічне застосування для моніторингу психологічного стану пацієнтів з метою призначення ліків або лікування. У роботі [5] розпізнавання емоцій реалізовано в медичних закладах для сприяння розслабленню та зменшенню стресу.

Нарешті, розпізнавання емоцій можна використовувати для маркетингу. Наприклад, розпізнавання емоцій може використовуватися для оптимізації веб-сайту [6] чи програми, де система може збирати інформацію про те, яка реклама привертає найбільшу увагу, що може дозволити доставку відповідного контенту відповідно до демографії аудиторії.

Фізіологічний підхід, який базується на аналізі сигналів сенсорів, став кращою альтернативою методам, які ґрунтуються на зміні міміки, жестів і голосових особливостей. Системи на основі машинного зору не є стійкими до фальшивих емоцій і ними легко маніпулювати [7].

Саме тому на сьогоднішній день так багато досліджень зосереджено на аналізі фізіологічних сигналів, у тому числі мультимодальних підходах, поєднуючи різні фізіологічні сигнали від біосенсорів, таких як електрокардіограми (ЕКГ), електроенцефалограми (ЕЕГ), електроміограми (ЕМГ), електро-шкірні активності (ЕША) або шкірно-гальванічні реакції (ШГР), фотоплетизмограми (ФПГ) або кров'яний тиск (КрТ), дихальні ритми (ДР), температура шкіри (ТШ). Незважаючи на це, що мультимодальний підхід до розпізнавання емоцій зазвичай ефективніший, унімодальний підхід має перевагу завдяки меншому часу обробки та простішому збору даних.

Нещодавні досягнення в розпізнаванні емоцій сприяли випуску загальнодоступних еталонних баз даних, які в основному складаються з унімодальних баз даних (тобто текстових, аудіо, візуальних і сигналів) і мультимодальних баз даних (що об'єднують два і більше сигналів). Наявність таких баз даних спонукало наукову спільноту до ефективних методів класифікації на основі машинного навчання (МН) і глибокого навчання (ГН).

Дослідження наведені в роботі [8] показали, що людські емоції виражаються головним чином через міміку (55%), голос (38%) і мову (7%) у щоденному людському спілкуванні. Зазвичай текстові, звукові та візуальні сигнали разом називаються фізичними даними. Розпізнавання емоцій на основі фізичних даних вже використовується на практиці, хоча воно може бути неефективним, оскільки люди можуть мимоволі або навмисно приховувати свої справжні емоції (так зване соціальне маскування) [9]. На відміну від фізичних, фізіологічні сигнали (наприклад, ЕЕГ та ЕКГ) не підпадають під дію цих обмежень, оскільки спонтанну фізіологічну діяльність, пов'язану з емоціями, майже не можливо контролювати чи підміняти. Таким чином, розпізнавання емоцій на основі ЕЕГ або ЕКГ може генерувати більш об'єктивні прогнози в реальному часі та надавати надійні характеристики емоційних станів.

Формулювання цілей статті

Метою даної роботи є аналіз та класифікація підходів до розпізнавання емоцій на основі інформації фізіологічних сигналів, стратегій злиття в багатомодальному розпізнаванні, а також аналіз наявних баз даних, які використовуються для мультимодального розпізнавання психологічних станів на основі фізіологічних сигналів. Вивчаються різноманітні джерела та типи даних, які використовуються в таких базах даних, а також методи аналізу та класифікації, що застосовуються для визначення психологічних станів на їхній основі. Проведення такого аналізу має велике значення, оскільки дозволяє визначити, які методи даних є найбільш перспективними для подальших досліджень та які бази даних можна та варто використати для таких досліджень.

В статті наведені наступні наукові результати

Подано класифікацію методів розпізнавання емоцій з фізіологічних сигналів. На верхньому рівні методи розпізнавання емоцій можна розділити на два основних класи – унімодальне та мультимодальне розпізнавання. Подальша класифікація базується на джерелах сигналів, способі класифікації та методах злиття для мультимодального розпізнавання емоцій.

Відібрано і класифіковано наявні бази даних за основними сигналами, та іншими ключовими характеристиками.

Проаналізовано ефект модальностей, стратегій злиття та інших факторів на точність розпізнавання та можливості прикладного застосування.

Основні бази даних для мультимодального розпізнавання психологічних станів

В процесі дослідження вибір шуканих баз даних для аналізу здійснюється з урахуванням кількох критеріїв. Початково у наукових журналах та матеріалах конференцій проводиться пошук інформації, де опубліковані результати досліджень, пов'язаних із мультимодальним розпізнаванням психологічних станів на основі фізіологічних сигналів. Такий пошук включає в себе ключові слова, такі як "мультимодальне розпізнавання", "психологічні стани", "фізіологічні сигнали", тощо.

Після пошуку, отримані бази даних об'єктивно оцінюються згідно певних критеріїв, які визначені дослідницьким колективом. Критерії включають такі аспекти.

Мультимодальність. Бази даних повинні містити принаймні два різних фізіологічних сигнали, такі як ЕЕГ, ЕКГ, ШГР та інші, що надають можливість аналізу взаємодії різних систем організму.

Розмір вибірки. Бази даних мають містити достатню кількість учасників для отримання статистичної значущості результатів. В роботі обмежені бази, де число учасників становить не менше 10 осіб, оскільки такий обсяг вважається мінімальним для надійної статистичної обробки.

Анотації емоційного стану. Оскільки мета роботи спрямована на мультимодальне розпізнавання емоцій, бази даних повинні містити анотації, які описують психологічний стан учасників в термінах емоцій.

Доступність і можливість використання. Важливим критерієм є можливість доступу до баз даних та їх використання для подальших досліджень.

У більшості праць з використанням наведених баз даних приводяться повідомлення про збільшення частоти розпізнавання зі збільшенням кількості модальностей даних, а саме даних ЕЕГ, ЕКГ, ШГР, ЕМГ, ДИХ та ТШ однак досі немає чітких доказів того, які комбінації функцій фізіологічних сигналів є найбільш актуальними.

Деякі набори даних для розпізнавання емоцій за допомогою фізіологічних сигналів є загальнодоступними (ТАБЛИЦЯ 1). Вони дозволяють порівнювати алгоритми розпізнавання емоцій, полегшуючи пряме порівняння результатів різних методів розпізнавання емоцій.

Таблиця 1.

Загальнодоступні набори даних щодо особливостей мультимодального розпізнавання емоцій на основі інформації фізіологічних сигналів

База даних	Фізіологічні Модальності (біосигнали)	Метод виявлення	К-сть учасників	Анотації	Зауваження
AMIGOS [10]	ЕЕГ, ЕКГ	ШГР, Відео (фрагменти фільму)	40	Самооцінка: Валентність, збудження, домінування, симпатії, знайомство та базові емоції (радість, сум, злість, огида, подив, страх і нейтральний) Зовнішня оцінка: Валентність та збудження.	В наявності. Дані були записані в індивідуальних і групових налаштуваннях для спостереження та аналізу впливу соціального контексту на емоційний стан людини.
ASCERTAIN [11]	ЕЕГ, ШГР	ЕКГ, Фрагменти фільмів	58	Самозвіти (для 36 відео по 58 суб'єктів у кожному): збудження, валентність, залучення, симпатія, знайомство; Особистісні оцінки для великої 5 рис особистості: екстраверсія, приємність, сумлінність, емоційна стабільність, відкритість	В наявності. Перша база даних, яка пов'язує риси особистості та емоційні стани через фізіологічні реакції
BioVid [12]	Емо ЕКГ, ЕМГ	Відео (фрагменти фільму)	94	5 емоцій: веселість, смуток, гнів, огида і страх.	Спочатку було 94 учасники, але дані є лише для 86 через пропущені або пошкоджені записи.
CLAS [13]	ФПГ, ШГР	ЕКГ, Перцептивні та інтерактивні завдання	62	Низьке/високе збудження, валентність, стан негативної валентності високого збудження (HANV) і низька/висока концентрація.	Використання інтерактивних задач
DEAP [14]	ЕКГ, ШГР, ЕОГ, ТШ	ЕЕГ, ЕМГ, ДИХ, Музичні відео	32	Інтенсивність збудження, валентність уподобання до відео, домінування та знайомість.	В наявності. З 32 учасників було знято фейс-відео 22 осіб.

База даних	Фізіологічні Модальності (біосигнали)	Метод виявлення	К-сть учасників	Анотації	Зауваження
DECAF [15]	ЕКГ (3 канали), ЕМГ, ЕОГ, МEG	Музичні відео	30	Оцінки валентності та збудження за кожним учасником	В наявності. DECAF також містить анотації безперервних емоцій для відеокліпів від семи користувачів
DREAMER [16]	ЕКГ (2 канали), ЕЕГ (14 каналів)	Фрагменти фільмів	23	Валентність/збудження оцінюється за дискретною шкалою цілих чисел від 1 до 5	В наявності. Сигнали були захоплені за допомогою портативного, носимого, бездротового, недорогого та готового обладнання.
МАННОВ-НСІ [17]	32 канали ЕЕГ, ЕКГ, ШГР, ЕМГ, ДИХ, ТШ	Емоційні відео та фото	27	Валентність, збудження, домінування, симпатії, знайомство та базові емоції (радість, сум, злість, огида, подив, страх і нейтральний)	В наявності. Суб'єкти з різних культур.
MPED [18]	ЕКГ, ЕЕГ, ШГР, ДИХ	Емоційні відео	23	7 емоцій: радість, смішно, гнів, огида, страх, сум і нейтральність	В наявності. 28 відео для виявлення були ановані вручну та були вибрані з групи з 1500 відео, щоб зменшити вплив культури
RECOLA [19]	ЕКГ, ШГР	Спільні завдання, що включали соціальну та емоційну поведінку	46	Безперервна шкала валентності та збудження.	В наявності. 34 учасники з 46 дали згоду на передачу своїх даних.

Методи розпізнавання емоцій

Розпізнавання емоцій на основі фізіологічних сигналів можна розділити на два основні рівні. На РИС 1 наведена запропонована класифікація методів розпізнавання емоцій на основі фізіологічних сигналів. На верхньому рівні пропонується класифікація в два основних класи – унімодальне та багатомодальне розпізнавання емоцій. На нижньому рівні методи класифікуються за сигналами, або способом класифікації / методом злиття.



Рис. 1. Класифікація методів розпізнавання емоцій на основі фізіологічних сигналів

Фізіологічні джерела сигналу, пов'язані з емоційним станом

Сучасні дослідження в літературі вказують на те, що емоції виникають в результаті роботи кори головного мозку і підкіркових нервів, де підкіркові ядра в середньому мозку відіграють основну роль, а

вегетативні ядра і кардіореспіраторні ядра в довгастому мозку відіграють вторинну роль. Крім того, емоції також регулюються гіпоталамусом, лімбічною системою, ретикулярною структурою, нейроендокринною системою тощо. Оскільки емоції по суті походять від активності мозку, сигнали ЕЕГ повинні мати найбільш тісний зв'язок з емоціями серед усіх фізіологічних сигналів.

Людське серце пронизується як симпатичними, так і парасимпатичними нервами, які зазвичай антагоністичні один одному. Симпатичне збудження серця сприяє серцево-судинній діяльності, тоді як парасимпатичне збудження серця пригнічує серцево-судинну діяльність. Більшість раптових емоційних реакцій відображаються у вигляді відносної гіперактивності симпатичної нервової системи. Коли симпатичний нерв збуджений, норадреналін, який вивільняють закінчення симпатичних нервів, так і катехоламін, посилений у циркулюючій крові, діють на адренергічні рецептори клітинної мембрани міокарда, що призводить до збільшення частоти серцевих скорочень, підвищення провідності в атріовентрикулярному (передсердно-шлуночковому) з'єднанні, а також до підвищення систолічної здатності м'язів передсердь і шлуночків. Деякі емоційні реакції проявляються у вигляді відносної гіперактивності парасимпатичної нервової системи. У цій ситуації ацетилхолін, який вивільняють парасимпатичні нервові закінчення, діє на холінергічні рецептори клітинної мембрани міокарда, що призводить до зниження частоти серцевих скорочень і провідності, а також сповільнення кровообігу та пульсу. Таким чином, такі сигнали як ЕКГ, ФПГ, частота серцевих скорочень (ЧСС), варіабельність серцевого ритму (BCP) також тісно пов'язані з емоціями.

М'язи відрізняються високою збудливістю. Загалом м'язові волокна скорочуються під контролем нервової системи. Коли якась емоція стимулює моторні нейрони, вони проводять нервові імпульси вздовж аксонів до терміналів і вивільняють ацетилхолін як передавач для реалізації передачі збудження на з'єднанні рухового нерва і м'яза. Деякі емоції також можуть безпосередньо викликати активність великих виличних м'язів і м'язів-гофраторів, які також можуть генерувати ЕМГ. Також існує взаємозв'язок між емоціями і легенями. Емоційні стани високого збудження пов'язані зі стимулюючим механізмом мозку, який призводить до посилення адренергічної секреції тіла, що, як наслідок, призводить до високої інтенсивності дихання для посилення газообміну крові і навпаки, емоційні стани з низьким збудженням пов'язані з гальмівним механізмом мозку, який призводить до зниження адренергічної секреції організму, що, відповідно, призводить до низької інтенсивності дихання для зниження газообміну крові. Тому дихальну активність можна також використовувати для розпізнавання емоцій.

Головним центром регуляції температури тіла є гіпоталамус. Емоційні реакції можуть впливати на гіпоталамус, а гормони гіпоталамуса додатково викликають зміни температури тіла. Зокрема, інтенсивні імпульсивні емоції підвищують температуру тіла, а ті, що знаходяться в протилежних ситуаціях, таких як смуток і розчарування, знижують температуру тіла. Крім того, різна температура тіла може приводити до виділення різної кількості поту, який містить різну щільність електролітів, що може впливати на стійкість шкіри. Отже, сигнали ТШ та ЕША корисні для розпізнавання емоцій.

Емоції також пов'язані з електроокулограмою (ЕОГ). Коли люди перебувають у гнівному або депресивному емоційному стані, організм виділяє речовину під назвою «катехоламін», що діє на центральну нервову систему. Така речовина може впливати на діяльність очей і змінювати форму зіниць.

В таблиці 2 наведені основні атрибути, що впливають на використання тих чи інших фізіологічних сигналів для розпізнавання емоцій.

Таблиця 2.

Використання фізіологічних сигналів для досліджень емоційних станів

Джерело сигналу (орган)	Сигнал	Використання в існуючих дослідженнях	Складності прикладного застосування
Головний мозок	ЕЕГ	Дуже часто	Необхідність стійкого електричного контакту та великої кількості контактів для високої точності
Серце	ЕКГ,	Дуже часто	Вплив завад на якість сигналу, необхідність стійкого контакту в області торсу. Спрощене ЕКГ з рук є досить чутливим до завад.
Серце, кровеносна система	ФПГ	Часто	ФПГ сенсори є дуже розповсюдженими, проте вони дають суттєво меншу кількість ознак порівняно з ЕКГ
М'язи	ЕМГ	Часто	Необхідність використання носимих пристроїв в певних областях (наприклад, на плечі)
Легені	Дихальна активність	Нечасто	Дихальна активність досить легко реєструється сенсором акселерометра, проте точність сигналу є досить низькою
Гіпоталамус	Температура	Нечасто	Вплив температури навколишнього середовища
Гіпоталамус	ЕША	Часто	Вплив температури, вологості навколишнього середовища
Очі	ЕОГ	Нечасто	Складність використання з носимими пристроями
Інші органи (язик, шлунок тощо)		Рідко	Висока інвазивність

Унімодальне розпізнавання емоцій

Діаграма фізіологічного розпізнавання емоцій, як показано на рис. 2, зазвичай включає наступні п'ять аспектів: 1) Стимулювання емоцій суб'єктів за допомогою зображень, музики та відео; 2) Запис фізіологічних сигналів, які в основному включають ЕЕГ, провідність шкіри, дихання, частоту серцевих скорочень, ЕМГ та ЕКГ; 3) Вилучення ознак шляхом попередньої обробки фізіологічних сигналів, аналізу ознак, вибору та зменшення ознак; 4) Навчання моделі класифікації, такої як опорно-векторні машини (ОВМ, англ. support vector machines, SVM), метод k-найближчих сусідів (КНС), випадковий ліс (англ. random forest, RF) тощо; і 5) розпізнавання емоцій на основі дискретної моделі емоцій або багатомірної моделі емоцій.

Серед перерахованих вище фізіологічних сигналів емоцій, ЕЕГ або ЕКГ найчастіше використовуються на практиці і надають прості, об'єктивні та надійні дані для розпізнавання емоцій.

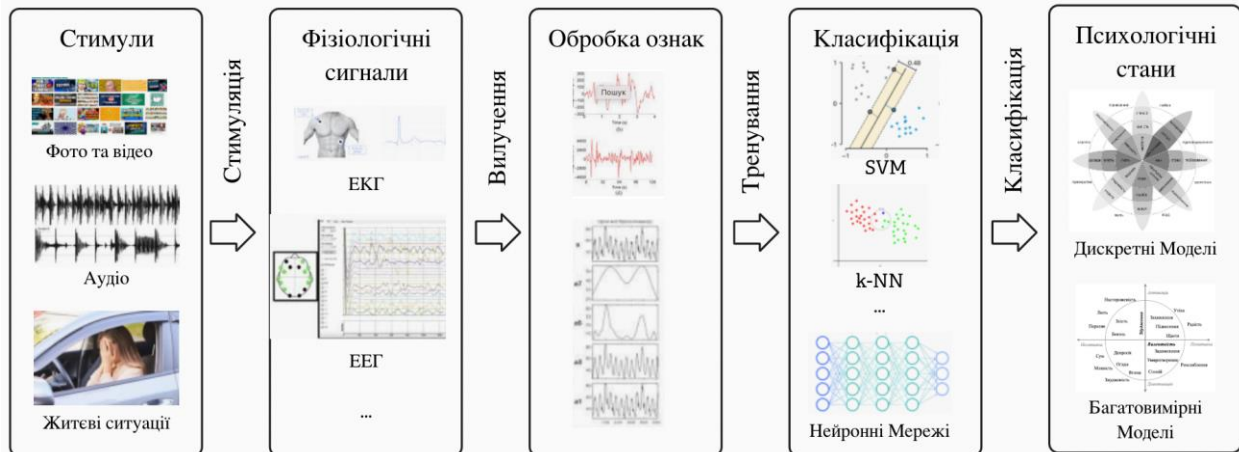


Рис 2. Основні етапи розпізнавання емоцій

Унімодальне розпізнавання емоцій з ЕЕГ

Порівняно з іншими периферійними фізіологічними сигналами, ЕЕГ може безпосередньо вимірювати зміни діяльності мозку, що відображає особливості емоційних станів. Крім того, ЕЕГ з високою часовою роздільною здатністю дає можливість відслідковування емоційного стану в режимі реального часу.

Продуктивність розпізнавання емоцій на базі ЕЕГ на основі машинного навчання МН залежить від того, як правильно спроектувати вилучення ознак, зменшення розмірності ознак (або вибір ознак) і методи класифікації.

Основною метою виділення ознак є виділення важливих характеристик ЕЕГ, які містять особливості часової області, частотної області та частотно-часової області представлень. При цьому аналіз швидкого перетворення Фур'є (ШПФ) часто використовується для перетворення ЕЕГ у спектр потужності. Зменшення розмірності ознак є важливим кроком у розпізнаванні емоцій на основі ЕЕГ через надмірність ЕЕГ. Зазвичай використовуваним класифікатором на основі МН для розпізнавання емоцій на основі ЕЕГ є ОВМ або його варіації.

На відміну від розпізнавання емоцій на основі МН на базі ЕЕГ, в роботі [20] використано згорткову нейронну мережу (ЗНМ) і обмежену машину Больцмана (ОМБ) з трьома шарами, щоб одночасно вивчати особливості та класифікувати емоції на основі ЕЕГ. Розпізнавання емоцій на основі ЗНН із протоколом, прив'язаним до об'єкта, дозволяє досягнути точність до 68,4%. В працях [21–23] запропоновано різні мережі розпізнавання емоцій на основі ЕЕГ, зокрема такі як двопівкульна доменна нейронна мережа [21], адаптивна мережа графів [22] та міркування варіаційного шляху [23].

Оскільки біологічна топологія між різними ділянками мозку може охоплювати як локальні, так і глобальні зв'язки між різними каналами ЕЕГ, в роботі [24] розроблено регуляризовану графову нейронну мережу (РГНН), яка складається як з операторів регуляризації змагального навчання по вузлах домену, так і з навчання розподілу з урахуванням емоцій.

Унімодальне розпізнавання емоцій з ЕКГ

ЕКГ фіксує фізіологічні зміни серця людини в різних ситуаціях через діяльність автономної нервової системи. Зі зміною емоцій або почуттів людини ЕКГ виявляє відповідну трансформацію хвилі, яка може надати достатньо інформації для розпізнавання емоцій.

Характеристики ЕКГ можуть бути безпосередньо виділені в часовій області. Найпоширеніші характеристики часової області включають частоту серцевих скорочень, середню пікову амплітуду R та середні інтервали R-R. Поширеним способом є перетворення функцій ЕКГ у часовій області на функції інших областей. Наприклад, використання швидкого перетворення Фур'є (ШПФ), дискретного вейвлетного перетворення (ДВП) і перетворення Гілберта-Хуанга для перетворення сигналів ЕКГ у характеристики частотної області.

У багатьох випадках корисним є поєднання різних типів функцій ЕКГ таких як лінійні характеристики, нелінійні характеристики, особливості часової області та характеристики частотно-часової області з ЕКГ та її похідної варіабельності серцевого ритму.

Останнім часом все частіше дослідники звертають увагу на методи глибокого навчання, що пропускають етапи вилучення та вибору ознак і використовують сирі дані сигналу для класифікації. Так автори у своїй роботі [25] запропонували структуру на основі емоційної генеративної змагальної мережі (Емоційна ГЗМ, англ. Emotional GAN) для підвищення здатності узагальнення розпізнавання емоцій шляхом включення розширених зразків ЕКГ, створених Emotional GAN. Порівняно з використанням лише оригінальних даних, приблизно 5% покращення середньої точності свідчить про важливість моделей на основі GAN для завдання розпізнавання емоцій. В роботі [26] запропонований самоконтрольований підхід до навчання мережі розпізнавання трансформації сигналу (МРТС) для вивчення просторово-часових особливостей і абстрактних зображень ЕКГ. При цьому ваги згорткових шарів у МРТС заморожуються, а потім тренуються два щільні шари для класифікації збудження та валентності.

Мультимодальне розпізнавання емоцій

Сьогодні більшість підходів мультимодального розпізнавання емоцій зосереджуються на мультимодальних стратегіях злиття та класифікують їх як злиття на рівні ознак (або раннє злиття), злиття на рівні прийняття рішень (або пізнє злиття), злиття на рівні моделі та злиття на гібридному рівні. Однак мультимодальний афективний аналіз також можна варіювати за допомогою комбінацій різних модальностей. Таким чином, мультимодальне розпізнавання емоцій можна класифікувати на основі чотирьох типів стратегій злиття.

1. Злиття на рівні ознак об'єднує ознаки, отримані з мультимодальних вхідних даних і формує один загальний вектор ознак, який потім надсилається в класифікатор.
2. Злиття на рівні прийняття рішень з'єднує усі вектори рішень, незалежно згенеровані з кожної модальності, в один вектор ознак.
3. Злиття на рівні моделі виявляє кореляційні властивості між функціями, які отримані з різних модальностей, і використовує або проектує модель злиття з розслабленими та повільними типами, такими як прихована марковська модель (ПММ) та двоетапна машина екстремального навчання (МЕН).
4. Гібридне злиття поєднує злиття на рівні функцій і на рівні прийняття рішень.

З удосконаленням технологій носимих пристроїв, автоматичний афективний аналіз на основі фізіологічних модальностей привертає все більше уваги. Однак через складність емоцій і значні індивідуальні відмінності у фізіологічних реакціях важко досягти задовільного прогнозу за допомогою розпізнавання емоцій на основі лише одного сигналу ЕЕГ або ЕКГ. Таблиця 3 містить інформацію про репрезентативні приклади використання різних стратегій злиття.

Злиття на рівні функцій

У роботі [27] наведений збір сигналів емоцій, викликаних музикою, зображенням або відео, і виділені низькорівневі дескриптори та статистичні характеристики з хвиль сигналу. Далі проводилось об'єднання цих функцій з метою прогнозування емоцій за допомогою класифікаторів на основі МН із груповою моделлю СІР (специфіки індивідуальної реакції). Комбінація МН з СІР на основі групи дала можливість досягти вищої точності, приблизно 90%, у дисбалансі бази даних емоцій. Подібним чином в роботі [28] попередньо проводилось оброблення, виділення та об'єднання часо-частотних характеристик ЕЕГ і ФПГ, після чого здійснювалась передача їх у мережу довгої короткочасної пам'яті (ДКЧП), гіперпараметри якої були оптимізовані за допомогою диференціальної еволюції (ДЕ). На такому наборі даних загальна точність складала 77,68% для чотириквadrантних вимірів емоцій.

Використовуючи біосигнали ЕКГ, ШГР, ТШ, ФПГ, ДИХ, ЕМГ та ЕОГ, вибрані з бази даних DEAP [14], автори в роботі [29] виділили характеристики з різною роздільною здатністю на основі DWT і використали їх для оцінки емоцій валентності-збудження-домінування. У праці [30] виділено 9 статистичних характеристик, 9 характеристик спектральної щільності потужності та 46 характеристик DBN EDA, фотоплетизмограми та ЕМГ. Функціональні залежності цих характеристик були об'єднані, з метою проведення навчання тонкого SVM Гауса розпізнавати 5 основних емоцій. Крім того в роботі [31] наведена розроблена мультимодальна залишкова мережа ДКЧП для вивчення залежності високорівневих часових характеристик ЕЕГ, ЕОГ та ЕМГ з метою прогнозування емоцій. Така мережа дає можливість досягати точності класифікації 92,87% і 92,30% для збудження і валентності відповідно.

Злиття на рівні рішень

Використовуючи біосигнали ЕЕГ, ЕКГ, ДИХ і ШГР з бази даних MANNING-HCI [17] автори в роботі [32], з метою об'єднання вихідних даних з кількох класифікаторів ОБМ для прогнозування 5 основних емоцій, розробили лінійну вагову матрицю. При цьому найвища середня точність розпізнавання складала 84,6%. З метою дослідження емоційно-фізіологічних часових особливостей, в праці [33] проводилось

трансформування ЕЕГ, ЕКГ і ГСР, вибраних з AMIGOS [10], у зображення спектрограми для представлення їх часо-частотної інформації. Для автоматичного вивчення найкращих часових характеристик була розроблена модель двонаправленої ДКЧП на основі уваги для мультимодального розпізнавання емоцій, які далі вводилися в глибокої нейронної мережі (ГНН) для прогнозування ймовірності унімодальних емоцій. Кінцевий емоційний стан обчислювався на основі схеми рівних ваг або схеми змінних ваг. Оскільки особистісно специфічні люди часто виявляють різні фізіологічні реакції після стимуляції емоційними елементами, в роботі [34] запропоновано використовувати вбудований варіаційний автокодер із втратою інваріантності атрибутів для вивчення особистісно-інваріантних уявлень. З біосигналів ЕЕГ, ЕКГ і ШГР, які вибиралися з AMIGOS [10], різні характеристики були вилучені, а потім подані в різні класифікатори SVM для прогнозування результатів унімодальної класифікації. В праці [35] було розроблено 2-вимірну ЗНН для ЕЕГ, об'єднанню ДКЧП і 1-вимірну ЗНН для ЕКГ і ТШ. Завдяки використанню більшості голосів на основі рішень, прийнятих кількома класифікаторами, структура досягла найвищої загальної точності 99,0% і 90,8% для AMIGOS [10] і DREAMER [16] відповідно.

Злиття на рівні моделі

В роботі [36] вперше було попередньо опрацьовано та виділено 425 помітних фізіологічних особливостей семи сигналів, при цьому дані вибиралися з бази даних DEAP. Для отримання абстракції вищого рівня основних функцій в подальшому досліджувалися окремі глибоко приховані нейрони в стековому автокодері. З метою проведення розпізнавання емоцій на основі байєсівської моделі був розроблений ансамбль глибоких класифікаторів із суміжною ієрархічною мережею об'єднання ознак на основі графів.

В таблиці 3 наведене узагальнене представлення недавніх досліджень мультимодального розпізнавання емоцій.

Таблиця 3.

Огляд досліджень мультимодального розпізнавання емоцій

Публікація	Рік	Сигнали	Архітектура	Стратегія злиття	База даних	Результати З/В= збудження/валентність
Verma et al. [29]	2014	ЕЕГ, ШГР, ТШ, ФПГ, ДИХ, ЕМГ, ЕОГ	ДВП, Багато ядерна ОВМ	Злиття ознак	DEAP	13 класів: 85,00
Hassan et al. [30]	2019	ШГР, ФПГ, ЕМГ	ГМП, ОВМ	Злиття ознак	DEAP	5 класів: 89,53 З/В: 65,1/61,8
Yang, Lee [34]	2019	ЕЕГ, ЕКГ, ШГР	ОВМ	Злиття рішень	AMIGOS	З/В: 68,8/67,0
Li et al. [33]	2020	ЕЕГ, ЕКГ, ШГР	ДКЧП, ГНМ	Злиття рішень	AMIGOS	З/В: 83,3/79,4
Yin et al. [36]	2017	ЕЕГ, ЕОГ, ЕМГ, ТШ, ШГР, ФПГ, ДИХ	Глибокий стековий автокодер, Байєсівська Модель	Злиття моделей	DEAP	З/В: 77,19/76,17

Висновки проведеного дослідження і перспективи подальшої роботи

Проведено аналіз існуючих моделей емоцій і баз даних, які зазвичай використовуються для розпізнавання емоцій, а також унімодальне та багатомодальне розпізнавання емоцій та використання різних стратегій злиття. Окремі висновки можна сформулювати наступним чином.

Вплив вибору сигналів. Незважаючи на те, що фізіологічні сигнали, зібрані носимими сенсорами, отримати складніше, ніж фізичні сигнали, численні методи розпізнавання емоцій на основі ЕЕГ або ЕКГ дають зазвичай об'єктивніші та надійніші результати, ніж розпізнавання на основі тексту, відео чи аудіо.

Вплив комбінацій модальностей і стратегій злиття. Поєднання різних модальностей і стратегій злиття є двома ключовими аспектами мультимодального афективного аналізу. Дві основні стратегії злиття для мультимодального афективного аналізу – це злиття на рівні ознак і злиття на рівні прийняття рішень. Конкатенація або факторизоване білінійне об'єднання векторів ознак зазвичай використовується для злиття на рівні ознак. Більшість/середнє голосування часто використовується для злиття на рівні прийняття рішень. Лінійно-зважені обчислення можуть бути використані для об'єднання як на рівні ознак, так і на рівні рішень, використовуючи оператори суми або добутку для об'єднання функцій або класифікаційних рішень різних модальностей. Згідно з проведеним аналізом, злиття на рівні ознак є значно більш поширеним, ніж злиття на рівні прийняття рішень. Порівняно зі злиттям на рівні функцій, злиття на рівні прийняття рішень виконується легше, однак воно ігнорує релевантність між функціями різних модальностей. Злиття на гібридному рівні має на меті повне використання переваг стратегій злиття на основі функцій і стратегій злиття на основі рішень, а також подолання недоліків будь-якої з них. На відміну від трьох наведених вище стратегій злиття, злиття на рівні моделі використовує НММ або байєсовські мережі для встановлення кореляції між характеристиками різних модальностей.

Вплив моделей на основі МН і ГН. Більшість ранніх робіт використовували методи на основі машинного навчання [37]. Конвеєр на основі МН складається з попередньої обробки необроблених сигналів, ручного вибору ознак і добре розроблених класифікаторів. Незважаючи на те, що різні типи були розроблені для різних модальностей, методи розпізнавання емоцій на основі МН важко повторно використовувати для подібних проблем через дескриптори функцій, що стосуються конкретного завдання та домену. Зазвичай використовуваними класифікаторами на основі МН є ОВМ, ПММ, ВЛ, КВС і штучні нейронні мережі (ШНН), зокрема ЗНМ. Серед них класифікатор ОВМ є найефективнішим і використовується в більшості досліджень.

Сьогодні моделі на основі глибокого навчання стали популярними і перевершують моделі на основі МН у більшості областей. Це обумовлено їх суттєвою здатністю вивчати представлення функцій. Для інформації послідовності часових рядів фізіологічних сигналів рекурентні нейронні мережі (РНМ) їх варіанти часто використовуються для захоплення тимчасової динаміки. Моделі ЗНМ можуть виконувати глибоке виділення просторово-часових ознак. Змагальне навчання широко використовується для підвищення стійкості моделей шляхом доповнення даних і між-доменного навчання. Крім того, для покращення загальної продуктивності використовуються різні механізми уваги і автокодери, які інтегровані із застосуванням методів на основі ГН. Методи на основі ГН мають перевагу в автоматичному вивченні найбільш дискримінаційних ознак. Однак підходи, які базуються на ГН, ще не отримали великого впливу на розпізнавання фізіологічних емоцій, порівняно з моделями, заснованими на МН.

Застосування розпізнавання емоцій у сценаріях реального життя

Останніми роками все більше дослідницьких груп зосереджуються на застосуванні ефективних обчислень у сценаріях реального життя. Зокрема, фізіологічні сигнали часто використовуються для виявлення клінічної депресії та стресу завдяки надійності та стабільності сигналів емоцій і доступності фізіологічних сигналів від носимих пристроїв. Оскільки мультимодальне розпізнавання емоцій може підвищити надійність і продуктивність унімодального розпізнавання, все більше досліджень почали перетворювати їх у різні варіанти прикладного застосування, що робить його багатообіцяючим напрямком дослідження.

Напрямки подальших досліджень

Хоча афективні обчислювальні системи, що використовують унімодальні або мультимодальні дані, зробили значні прориви, існує лише кілька надійних і ефективних алгоритмів для прогнозування емоцій і розпізнавання настроїв у різноманітних і складних сценаріях. На базі проведеного аналізу, можна сформулювати наступні рекомендації для майбутніх досліджень у сфері ефективних обчислень.

Розроблення нових і більш масштабних базових баз даних, зокрема мультимодальних баз даних, що складаються з різних модальностей буде дуже корисним. Умови повинні включати як спонтанні, так і неспонтанні сценарії з наданням анотаційних даних як у дискретних, так і в розмірних моделях емоцій.

На даний час існує значна потреба для вдосконалення стратегій злиття, особливо з використанням знань, заснованих на правилах або статистичні, для реалізації взаємного злиття різних модальностей, які можуть враховувати роль і важливість кожної модальності в розпізнаванні емоцій.

Оскільки існує значна потреба у формуванні нових ознак, які можуть відображати складні психологічні стани більш точно, використання мультимодального аналізу для створення нових ознак або дескрипторів, які враховують взаємодію різних фізіологічних сигналів є особливо бажаним.

Використання методів глибокого навчання (наприклад, нейронні мережі) для створення моделей, які можуть автоматично визначати психологічний стан на основі мультимодального аналізу фізіологічних сигналів є особливо бажаним і необхідним.

References

1. Picard, R.W. *Affective Computing*; Technical Report 321; MIT Media Laboratory Perceptual Computing Section: Cambridge, MA, USA, 1995.
2. Strauss, M.; Reynolds, C.; Hughes, S.; Park, K.; McDarby, G.; Picard, R.; Tao, J.; Tan, T. *Affective Computing: A Review*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 699–706.
3. Nikolova, D.; Georgieva, P.; Petkova, P.; Manolova, A. ECG-based emotion recognition: Overview of methods and applications. In *Proceedings of the ANNA 2018—Advances in Neural Networks and Applications*, St. Konstantin and Elena Resort, Bulgaria, 15–17 September 2018; pp. 118–122.
4. Katsis, C.D.; Katertsidis, N.; Ganiatsas, G.; Fotiadis, D.I. Toward emotion recognition in car-racing drivers: A biosignal processing approach. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* 2008, 38, 502–512. [CrossRef]
5. Tivatansakul, S.; Ohkura, M. Healthcare system focusing on emotional aspects using augmented reality: Implementation of breathing control application in relaxation service. In *Proceedings of the 2013 International Conference on Biometrics and Kansei Engineering*, Tokyo, Japan, 5–7 July 2013; pp. 218–222.
6. Kołakowska, A.; Landowska, A.; Szwoch, M.; Szwoch, W.; Wróbel, M.R. *Emotion Recognition and Its Applications*. In *Human-Computer Systems Interaction: Backgrounds and Applications 3*; Hippe, Z.S., Kulikowski, J.L., Mroczek, T., Wtorek, J., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 51–62. ISBN 978-3-319-08491-6.
7. Ghali, A.L.I.; Bassam Kurdy, M.H.D. Emotion Recognition Using Facial Expression Analysis. *J. Theor. Appl. Inf. Technol.* 2018, 96, 6117–6129.

8. Mehrabian, Communicating Without Words, Psychol. Today. (1968) 53–55.
9. J. Zhang, Z. Yin, P. Chen, S. Nichele, Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review, Inf. Fusion. 59 (2020) 103–126.
10. Correa, J.A.M.; Abadi, M.K.; Sebe, N.; Patras, I. Amigos: A dataset for affect, personality and mood research on individuals and groups. IEEE Trans. Affect. Comput. 2018, 12, 479–493.
11. Subramanian, R.; Wache, J.; Abadi, M.K.; Vieriu, R.L.; Winkler, S.; Sebe, N. ASCERTAIN: Emotion and personality recognition using commercial sensors. IEEE Trans. Affect. Comput. 2016, 9, 147–160.
12. Zhang, L.; Walter, S.; Ma, X.; Werner, P.; Al-Hamadi, A.; Traue, H.C.; Gruss, S. “BioVid Emo DB”: A multimodal database for emotion analyses validated by subjective ratings. In Proceedings of the 2016 IEEE Symposium Series on Computational Intelligence (SSCI), Athens, Greece, 6–9 December 2016; pp. 1–6.
13. Markova, V.; Ganchev, T.; Kalinkov, K. CLAS: A Database for Cognitive Load, Affect and Stress Recognition. In Proceedings of the 2019 International Conference on Biomedical Innovations and Applications (BIA), Varna, Bulgaria, 8–9 November 2019; pp. 1–4.
14. Koelstra, S.; Muhl, C.; Soleymani, M.; Lee, J.S.; Yazdani, A.; Ebrahimi, T.; Pun, T.; Nijholt, A.; Patras, I. Deap: A database for emotion analysis; using physiological signals. IEEE Trans. Affect. Comput. 2012, 3, 18–31.”
15. M. K. Abadi, R. Subramanian, S. M. Kia, P. Avesani, I. Patras, and N. Sebe, “DECAF: MEG-based multimodal database for decoding affective physiological responses,” IEEE Trans. Affect. Comput., vol. 6, no. 3, pp. 209–222, Jul. 2015.
16. Katsigiannis, S.; Ramzan, N. DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices. IEEE J. Biomed. Health Inform. 2017, 22, 98–107.
17. M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, “A multimodal database for affect recognition and implicit tagging,” IEEE Trans. Affect. Comput., vol. 3, no. 1, pp. 42–55, Jan. 2012.
18. Song, T.; Zheng, W.; Lu, C.; Zong, Y.; Zhang, X.; Cui, Z. MPED: A multi-modal physiological emotion database for discrete emotion recognition. IEEE Access 2019, 7, 12177–12191.
19. Ringeval, F.; Sonderegger, A.; Sauer, J.; Lalanne, D. Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions. In Proceedings of the 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Shanghai, China, 22–26 April 2013; pp. 1–8.
20. Y. Gao, H.J. Lee, R.M. Mehmood, Deep learning of EEG signals for emotion recognition, in: 2015 IEEE Int. Conf. Multimed. Expo Workshop ICMEW, 2015; pp. 1–5. <https://doi.org/10.1109/ICMEW.2015.7169796>.
21. Y. Li, W. Zheng, Z. Cui, T. Zhang, Y. Zong, A Novel Neural Network Model based on Cerebral Hemispheric Asymmetry for EEG Emotion Recognition, in: Proc. Twenty-Seventh Int. Jt. Conf. Artif. Intell., International Joint Conferences on Artificial Intelligence Organization, Stockholm, Sweden, 2018; pp. 1561–1567. <https://doi.org/10.24963/ijcai.2018/216>.
22. T. Song, S. Liu, W. Zheng, Y. Zong, Z. Cui, Instance-Adaptive Graph for EEG Emotion Recognition, Proc. AAAI Conf. Artif. Intell. 34 (2020) 2701–2708. <https://doi.org/10.1609/aaai.v34i03.5656>.
23. T. Zhang, Z. Cui, C. Xu, W. Zheng, J. Yang, Variational Pathway Reasoning for EEG Emotion Recognition, Proc. AAAI Conf. Artif. Intell. 34 (2020) 2709–2716. <https://doi.org/10.1609/aaai.v34i03.5657>.
24. P. Zhong, D. Wang, C. Miao, EEG-Based Emotion Recognition Using Regularized Graph Neural Networks, IEEE Trans. Affect. Comput. (2020) 1–1. <https://doi.org/10.1109/TAFFC.2020.2994159>.
25. G. Chen, Y. Zhu, Z. Hong, Z. Yang, EmotionalGAN: Generating ECG to Enhance Emotion State Classification, in: Proc. 2019 Int. Conf. Artif. Intell. Comput. Sci., Association for Computing Machinery, New York, NY, USA, 2019; pp. 309–313. <https://doi.org/10.1145/3349341.3349422>.
26. P. Sarkar, A. Etemad, Self-Supervised Learning for ECG-Based Emotion Recognition, in: ICASSP 2020 - 2020 IEEE Int. Conf. Acoust. Speech Signal Process. ICASSP, 2020; pp. 3217–3221. <https://doi.org/10.1109/ICASSP40776.2020.9053985>.
27. C. Li, C. Xu, Z. Feng, Analysis of physiological for emotion recognition with the IRS model, Neurocomputing. 178 (2016) 103–111. <https://doi.org/10.1016/j.neucom.2015.07.112>.
28. B. Nakisa, M.N. Rastgoo, A. Rakotonirainy, F. Maire, V. Chandran, Long Short Term Memory Hyperparameter Optimization for a Neural Network Based Emotion Recognition Framework, IEEE Access. 6 (2018) 49325–49338. <https://doi.org/10.1109/ACCESS.2018.2868361>.
29. G.K. Verma, U.S. Tiwary, Multimodal fusion framework: A multiresolution approach for emotion classification and recognition from physiological signals, Neuro Image. 102 (2014) 162–172. <https://doi.org/10.1016/j.neuroimage.2013.11.007>.
30. M.M. Hassan, Md.G.R. Alam, Md.Z. Uddin, S. Huda, A. Almogren, G. Fortino, Human emotion recognition using deep belief network architecture, Inf. Fusion. 51 (2019) 10–18. <https://doi.org/10.1016/j.inffus.2018.10.009>.
31. J. Ma, H. Tang, W.-L. Zheng, B.-L. Lu, Emotion Recognition using Multimodal Residual LSTM Network, in: Proc. 27th ACM Int. Conf. Multimed., ACM, Nice France, 2019; pp. 176–183. <https://doi.org/10.1145/3343031.3350871>.
32. W. Wei, Q. Jia, Y. Feng, G. Chen, Emotion Recognition Based on Weighted Fusion Strategy of Multichannel Physiological Signals, Comput. Intell. Neurosci. 2018 (2018) e5296523. <https://doi.org/10.1155/2018/5296523>.
33. C. Li, Z. Bao, L. Li, Z. Zhao, Exploring temporal representations by leveraging attention-based bidirectional LSTM-RNNs for multi-modal emotion recognition, Inf. Process. Manag. 57 (2020) 102185. <https://doi.org/10.1016/j.ipm.2019.102185>.
34. H.-C. Yang, C.-C. Lee, An Attribute-invariant Variational Learning for Emotion Recognition Using Physiology, in: ICASSP 2019 - 2019 IEEE Int. Conf. Acoust. Speech Signal Process. ICASSP, 2019; pp. 1184–1188. <https://doi.org/10.1109/ICASSP.2019.8683290>.
35. M.N. Dar, M.U. Akram, S.G. Khawaja, A.N. Pujari, CNN and LSTM-Based Emotion Charting Using Physiological Signals, Sensors. 20 (2020) 4551. <https://doi.org/10.3390/s20164551>
36. Z. Yin, M. Zhao, Y. Wang, J. Yang, J. Zhang, Recognition of emotions using multimodal physiological signals and an ensemble deep learning model, Comput. Methods Programs Biomed. 140 (2017) 93–110. <https://doi.org/10.1016/j.cmpb.2016.12.005>.
37. M. El Ayadi, M.S. Kamel, F. Karray, Survey on speech emotion recognition: Features, classification schemes, and databases, Pattern Recognit. 44 (2011) 572–587. <https://doi.org/10.1016/j.patcog.2010.09.020>.