УДК 004.421:004.85:681.84:78.02 **DOI:** 10.31673/2412-9070.2026.055608

Є. М. ЛАБУН, аспірант;
ORCID: 0009-0006-0054-4915
Б. О. ХУДІК, PhD, доцент;
ORCID: 0009-0005-1914-6418
Н. А. ТРІНТІНА, канд. техн. наук, доцент,
ORCID: 0000-0001-6827-4030
Державний університет інформаційно-комунікаційних технологій, Київ

АВТОМАТИЗАЦІЯ ПРОЦЕСУ МАСТЕРИНГУ АУДІО З ВИКОРИСТАННЯМ АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ

У статті розглянуто проблему автоматизації процесу мастерингу аудіо з використанням алгоритмів машинного навчання. Актуальність дослідження зумовлена тим, що сучасні цифрові засоби створення музики значно спростили виробництво аудіоконтенту, однак якісний мастеринг і надалі потребує спеціальних знань, тренованого слуху, технічного контролю та значних часових витрат. Це створює бар'єр для незалежних музикантів, невеликих студій і користувачів, які не мають доступу до професійного звукорежисера. У роботі узагальнено основні етапи аналізу аудіосигналу, зокрема спектральний аналіз, еквалізацію, динамічну обробку, контроль гучності та оцінювання властивостей сигналу за допомогою об'єктивних і суб'єктивних критеріїв. Окрему увагу приділено можливостям згорткових нейронних мереж, рекурентних нейронних мереж і мереж LSTM у задачах обробки аудіо, класифікації музичних жанрів та аналізу часової структури звукового сигналу. Проаналізовано наукові підходи до оцінювання якості звуку, автоматичного мастерингу та використання еталонних або професійно підготовлених аудіозаписів як бази для навчання моделей. Для ілюстрації експериментального підходу розглянуто використання набору даних GTZAN, що містить музичні фрагменти різних жанрів. Наведено порівняння типів нейронних мереж за точністю класифікації, кількістю епох навчання та орієнтовним часом тренування. Результати узагальнення показують, що CNN доцільно використовувати для аналізу спектрограм і виділення частотних ознак, тоді як RNN та LSTM краще придатні для роботи з послідовностями й контекстом композиції. Зроблено висновок, що машинне навчання може бути ефективною основою для створення інструментів підтримки автоматизованого мастерингу, однак повна заміна людської експертизи є передчасною через важливість художнього задуму, жанрового контексту та суб'єктивного сприйняття слухачем.

Ключові слова: машинне навчання, мастеринг, звукозапис, штучний інтелект, обробка аудіо, саунддизайн, нейронні мережі.

Вступ

Музика є невід'ємною частиною людської культури, а якість звучання музичного твору значною мірою впливає на його сприйняття слухачем. У цифрову епоху створення музики стало доступним значно ширшому колу користувачів: для запису, редагування та поширення композицій достатньо персонального комп'ютера, цифрової аудіостанції та базового набору програмних інструментів. Водночас доступність інструментів створення музики не гарантує професійної якості фінального звучання, оскільки останній етап підготовки аудіозапису — мастеринг — потребує одночасного врахування технічних, психоакустичних і художніх чинників.

© Лабун Є. М., Худік Б. О., Трінтіна Н. А., 2026

Мастеринг спрямований на досягнення тонального балансу, контроль динамічного діапазону, корекцію гучності, усунення небажаних артефактів і забезпечення коректного відтворення композиції на різних акустичних системах. Традиційно ці завдання виконуються фахівцем, який має професійний слух, досвід роботи з жанровими особливостями та доступ до якісного моніторингу. Розвиток машинного навчання створює передумови для часткової автоматизації таких процесів, однак автоматизовані системи повинні враховувати не лише числові характеристики сигналу, а й особливості сприйняття аудіо людиною.

Сучасні дослідження у сфері автоматичної обробки звуку демонструють перспективність використання глибинних нейронних мереж для аналізу спектрограм, класифікації жанрів, розпізнавання артефактів і моделювання окремих рішень звукорежисера. Тому актуальним є узагальнення підходів, які можуть бути використані як основа для подальшого розроблення методу оцінки та автоматизованого покращення якості мастерингу аудіо.

Постановка проблеми

Проблема, якій присвячене дослідження, полягає у складності формалізації якості мастерингу аудіо. З одного боку, аудіосигнал може бути описаний об'єктивними характеристиками: рівнем гучності, піковими значеннями, спектральним балансом, динамічним діапазоном, наявністю кліпінгу, шуму або спотворень. З іншого боку, кінцева оцінка якості залежить від суб'єктивного сприйняття слухача, жанрових очікувань, художнього задуму та умов прослуховування.

На практиці це ускладнює створення універсальної системи, яка могла б автоматично визначати якість мастерингу та пропонувати коректні зміни для різних музичних жанрів. Особливої уваги потребує поєднання об'єктивних метрик із моделями машинного навчання, здатними виявляти закономірності у спектральних і часових ознаках аудіосигналу. Задача дослідження полягає в аналізі підходів до автоматизації мастерингу, визначенні придатних типів нейронних мереж і формуванні узагальненого бачення їх застосування для оцінювання та покращення аудіо.

Аналіз останніх досліджень і публікацій

У науковій літературі якість аудіо розглядається як поєднання технічних і перцептивних властивостей. Дослідження Wilson та Fazenda показує, що сприйняття якості популярної музики пов'язане з характеристиками сигналу, зокрема гучністю та динамічною компресією, водночас вподобання композиції може відрізнятися від оцінки її технічної якості [1]. Це важливо для автоматизованого мастерингу, оскільки модель не повинна ототожнювати суб'єктивну симпатію до музичного твору з його технічною якістю.

Питання динамічного діапазону та надмірної компресії розглядаються у роботах, присвячених явищу hyper-compression і "loudness war". Зокрема, Taylor та Martens досліджують слухацькі переваги щодо різних рівнів динамічної компресії та підкреслюють, що збільшення гучності не завжди приводить до кращого сприйняття [2]. Campbell, Toulson і Paterson також зазначають, що надмірна компресія може спричиняти спотворення, втому слухача та зниження психоакустичної якості [3].

Методи стандартизованого вимірювання та суб'єктивного оцінювання аудіо описані у рекомендаціях ITU. Рекомендація ITU-R BS.1770-5 визначає алгоритми вимірювання програмної гучності та true-peak рівня [4]. Рекомендація ITU-R BS.1116-3 регламентує суб'єктивне оцінювання малих спотворень у високоякісних аудіосистемах [5], а ITU-R BS.1534-3 описує метод MUSHRA для оцінювання проміжного рівня якості аудіо [6]. Для статистичного порівняння об'єктивних моделей якості може бути використана рекомендація ITU-T P.1401 [7].

Окремий напрям становлять роботи з використання машинного навчання в аудіоінженерії. Shtern, Casas і Tzerpos запропонували підхід до оцінювання якості мастерингу за допомо-

гою моделей, навчених розпізнавати відхилення від професійно підготовленого еталона [8]. Pardo та співавтори розглядають інтерфейси аудіопродакшену, що використовують машинне навчання для зближення технічних параметрів обробки з природними способами комунікації музикантів і звукорежисерів [9]. Сучасні дослідження автоматичного мікшування та мастерингу також демонструють потенціал глибинного навчання для моделювання еквалізації, динамічної обробки, просторового балансу та реверберації [10], [11].

Незважаючи на значний прогрес, невирішеною залишається проблема інтеграції об'єктивних метрик, суб'єктивного сприйняття та жанрового контексту в єдиний метод оцінювання якості мастерингу. Саме ця невирішена частина загальної проблеми визначає подальшу спрямованість дослідження.

Мета і задачі дослідження

Метою статті є огляд сучасних підходів до автоматизації процесу мастерингу аудіо з використанням алгоритмів машинного навчання та аналіз можливостей застосування різних типів нейронних мереж для обробки музики різних жанрів.

Для досягнення мети поставлено такі задачі: охарактеризувати основні методи аналізу аудіосигналу, що можуть бути використані в автоматизованому мастерингу; розглянути особливості застосування CNN, RNN та LSTM у задачах аудіообробки; проаналізувати використання жанрових наборів даних, зокрема GTZAN, для експериментальної перевірки моделей; порівняти типи нейронних мереж за базовими показниками точності та обчислювальної складності; окреслити перспективи подальших досліджень у напрямі поєднання об'єктивних метрик і суб'єктивної оцінки якості звуку.

Результати дослідження

Методи аналізу аудіо

Аналіз аудіо зазвичай здійснюється за допомогою кількох взаємопов'язаних методів: спектрального аналізу, еквалізації, динамічної обробки, вимірювання гучності, контролю true-peak рівня та виявлення небажаних артефактів. Спектральний аналіз дозволяє візуалізувати частотні компоненти аудіосигналу та визначити зони, у яких можливі тональні дисбаланси. Еквалізація використовується для регулювання амплітуди частотних діапазонів, що дає змогу підкреслити або послабити окремі елементи звучання. Динамічна обробка, зокрема компресія та лімітування, допомагає контролювати динамічний діапазон сигналу, але за надмірного застосування може погіршувати природність звучання.

З розвитком машинного навчання ці методи можуть бути доповнені автоматичним виділенням ознак із сирого аудіосигналу або його спектрограм. Зокрема, спектрограма дає змогу подати аудіо у формі, зручній для аналізу згортковими нейронними мережами. Приклад спектрограми музичного треку представлено на рис. 3.

Огляд різних підходів та видів нейромереж

Глибинні нейронні мережі демонструють високу ефективність у задачах аудіообробки, оскільки здатні виявляти складні залежності у великих наборах даних. Згорткові нейронні мережі (CNN) доцільно застосовувати для аналізу спектрограм, оскільки вони ефективно виділяють локальні частотно-часові ознаки. Це робить CNN корисними для класифікації жанрів, виявлення характерних тембрових патернів і розпізнавання спектральних відхилень.

Рекурентні нейронні мережі (RNN), зокрема LSTM, орієнтовані на роботу з послідовними даними. Аудіосигнал є часовою послідовністю, тому RNN та LSTM можуть зберігати інформацію про попередні стани сигналу й використовувати її для аналізу контексту композиції. Це корисно для задач, у яких важливо враховувати розвиток музичного матеріалу в часі, наприклад, зміну динаміки, структури аранжування або переходи між частинами композиції.

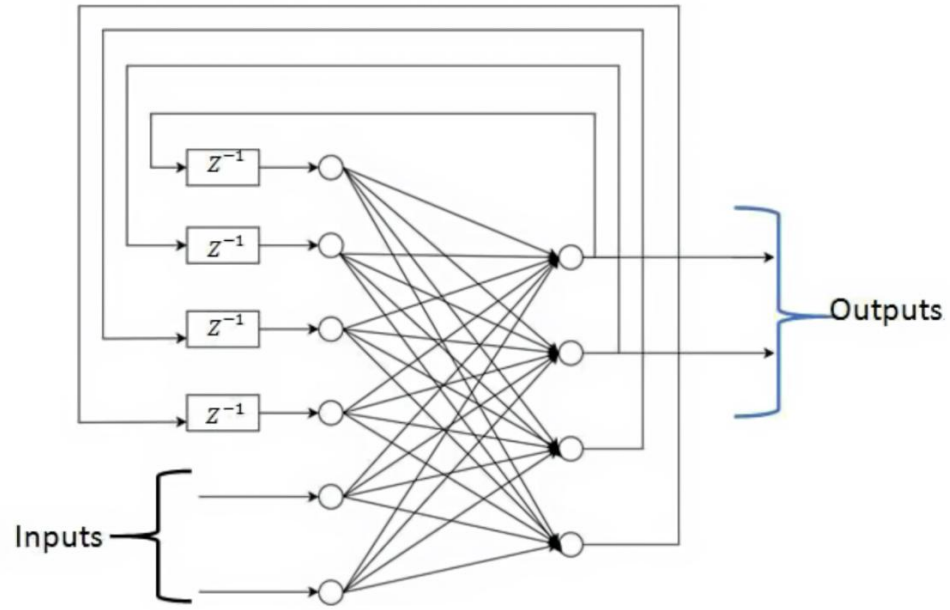


Рис. 1. Рекурентна нейронна мережа з прихованими нейронами (RNN)

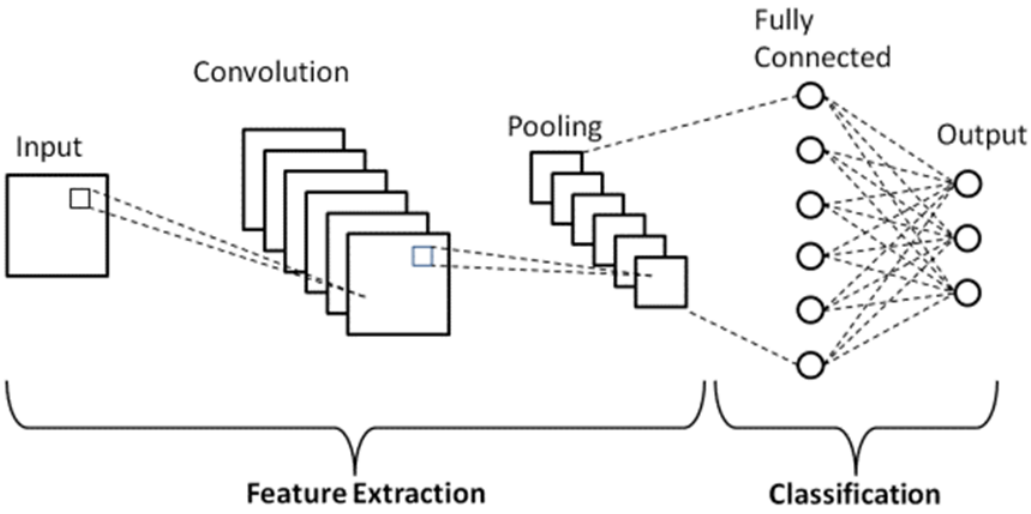


Рис. 2. Згорткова нейронна мережа (CNN)

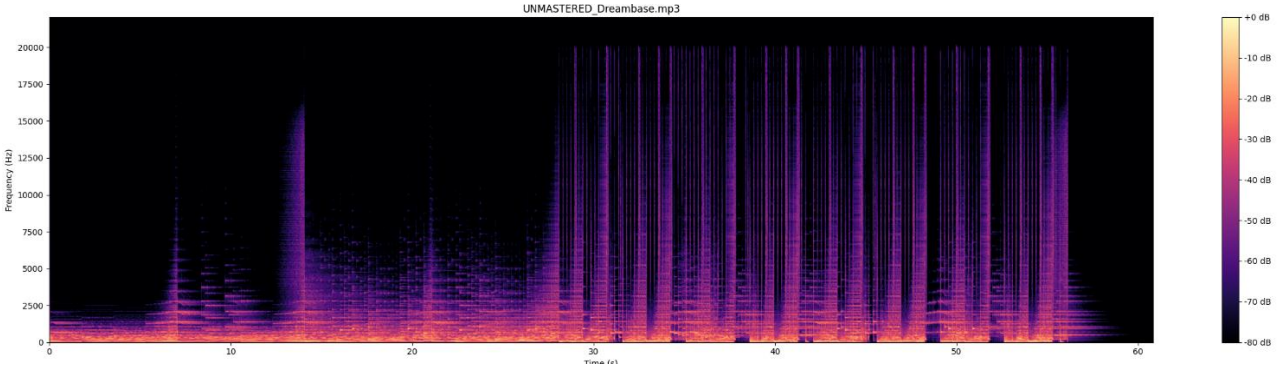


Рис. 3. Приклад спектрограми музичного треку

Приклади обробки музики різних жанрів

Однією з основ для експериментів із автоматизованою обробкою музики може бути набір даних GTZAN Dataset, який містить понад тисячу музичних фрагментів десяти жанрів: блюз, класика, кантрі, диско, хіп-хоп, джаз, метал, поп, реггі та рок [12]. Такий набір даних дає змогу перевіряти, наскільки моделі здатні узагальнювати властивості різних жанрів і чи можуть вони виявляти типові спектральні та ритмічні особливості композицій.

Для задач автоматичного мастерингу жанровий контекст є важливим, оскільки очікуваний тональний баланс, рівень компресії та суб'єктивне сприйняття гучності можуть суттєво відрізнятися для класичної, рок-, поп- або електронної музики. Наприклад, надмірне зменшення динамічного діапазону може бути критичним для класики та джазу, тоді як у деяких стилях електронної музики високий середній рівень гучності може сприйматися як естетично прийнятний. Тому, універсальна модель мастерингу повинна враховувати не лише числові параметри сигналу, а й жанрові норми та очікування слухача.

Числові дані та порівняння моделей

Для тренування моделей використовуються великі обсяги даних і достатні обчислювальні ресурси. У прикладному сценарії для навчання CNN на наборі GTZAN може бути використано 80% даних для тренування і 20% для тестування. Після 100 епох тренування модель CNN продемонструвала точність 92% для задачі класифікації музичних жанрів. RNN показала точність 89% на тестовому наборі після 150 епох, а LSTM — 90% після 200 епох. Наведені значення слід розглядати як ілюстративні для порівняння типів моделей у межах оглядового дослідження, оскільки фактична точність залежить від підготовки даних, архітектури мережі, параметрів навчання та способу валідації.

Таблиця 1

Порівняння точності обробки аудіо різних видів нейромереж

Метрика	**CNN**	**RNN**	**LSTM**
Точність	92%	89%	90%
Кількість епох	100	150	200
Час тренування	10 годин	12 годин	14 годин

Порівняння показує, що CNN є ефективними для роботи зі спектральними представленнями аудіо, тоді як RNN і LSTM краще враховують часовий контекст. У задачах автоматизованого мастерингу доцільним може бути комбінований підхід, у якому CNN аналізує спектральні ознаки, а рекурентні або трансформерні компоненти враховують часову структуру музичного матеріалу. Така архітектура потенційно здатна поєднати переваги частотно-часового аналізу і моделювання довготривалих залежностей.

Висновки та перспективи подальших досліджень

У статті розглянуто підходи до автоматизації процесу мастерингу аудіо з використанням алгоритмів машинного навчання. Показано, що якість мастерингу не може бути зведена лише до однієї об'єктивної метрики, оскільки вона формується під впливом спектрального балансу, динамічного діапазону, гучності, наявності артефактів, жанрового контексту та суб'єктивного сприйняття слухача. Згорткові нейронні мережі є перспективними для аналізу спектрограм і виділення частотно-часових ознак, тоді як RNN та LSTM доцільно використовувати для обробки послідовностей і врахування контексту композиції.

Впровадження нейронних мереж у системи автоматичного мастерингу може знизити вартість і підвищити доступність якісної аудіообробки для незалежних музикантів. Водночас повна автоматизація цього процесу залишається складною через необхідність урахування художнього задуму, жанрових норм і перцептивних особливостей слухача. Перспективами подальших досліджень є формування набору об'єктивних метрик якості мастерингу, розроблення датасету пар "неякісний/якісний мастер", поєднання стандартів ITU з результатами суб'єктивного прослуховування, а також створення гібридної моделі, яка поєднуватиме машинне навчання з експертною оцінкою звукорежисера.

Внесок авторів

Євген ЛАБУН – концептуалізація, аналіз джерел, підготовка огляду літератури, опис методів аналізу аудіо, підготовка тексту статті; Богдан ХУДІК – наукове керівництво, методика, перевірка структури дослідження, редагування теоретичних положень, узгодження висновків і перспектив подальших досліджень.

Декларація про штучний інтелект

Під час підготовки редакційної версії статті інструменти штучного інтелекту використовувалися для структурування тексту відповідно до вимог журналу, перекладу окремих фрагментів англійською мовою та перевірки логічної послідовності викладу. Остаточна відповідальність за зміст, добір джерел, наукові твердження та інтерпретацію результатів належить авторам.

Конфлікт інтересів

Автори заявляють про відсутність конфлікту інтересів та підтверджують, що під час підготовки цієї роботи не існувало жодних комерційних, фінансових чи інших взаємовідносин, які могли б бути розцінені як такі, що здатні вплинути на результати дослідження або їх інтерпретацію. Робота виконана відповідно до принципів академічної доброчесності, етичних норм проведення наукових досліджень та вимог редакційної політики щодо запобігання конфлікту інтересів.

Список використаної літератури

1. Wilson, A., & Fazenda, B. M. (2016). Perception of audio quality in productions of popular music. Journal of the Audio Engineering Society, 64(1/2), 23-34. https://doi.org/10.17743/jaes.2015.0090

2. Taylor, R. W., & Martens, W. L. (2014). Hyper-compression in music production: Listener preferences on dynamic range reduction. Audio Engineering Society Convention Paper, 136th AES Convention.

3. Campbell, W., Toulson, R., & Paterson, J. (2010). The effect of dynamic range compression on the psychoacoustic quality and loudness of commercial music. INTER-NOISE 2010, Lisbon, Portugal.

4. International Telecommunication Union. (2023). Recommendation ITU-R BS.1770-5: Algorithms to measure audio programme loudness and true-peak audio level. ITU.

5. International Telecommunication Union. (2015). Recommendation ITU-R BS.1116-3: Methods for the subjective assessment of small impairments in audio systems. ITU.

6. International Telecommunication Union. (2015). Recommendation ITU-R BS.1534-3: Method for the subjective assessment of intermediate quality level of audio systems. ITU.

7. International Telecommunication Union. (2020). Recommendation ITU-T P.1401: Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models. ITU. https://handle.itu.int/11.1002/1000/14159

8. Shtern, M., Casas, P., & Tzerpos, V. (2018). Evaluating music mastering quality using machine learning. Proceedings of CASCON 2018, 126-135.

9. Pardo, B., Cartwright, M., Seetharaman, P., & Kim, B. (2019). Learning to build natural audio production interfaces. Arts, 8(3), 110. https://doi.org/10.3390/arts8030110

10. Martínez-Ramírez, M. A., Liao, W.-H., Fabbro, G., Uhlich, S., Nagashima, C., & Mitsufuji, Y. (2022). Automatic music mixing with deep learning and out-of-domain data. Proceedings of the 23rd International Society for Music Information Retrieval Conference, 411-418.

11. Canyakan, S. (2025). Machine learning in audio mastering: A comparative study. Journal for the Interdisciplinary Art and Education, 6(1), 47-65. https://doi.org/10.5281/zenodo.15074948

12. Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing, 10(5), 293-302. https://doi.org/10.1109/TSA.2002.800560

13. Müller, M. (2021). Fundamentals of music processing: Using Python and Jupyter notebooks (2nd ed.). Springer. https://doi.org/10.1007/978-3-030-69808-9

14. Melechovsky, J., Mehrish, A., Roy, A., & Herremans, D. (2025). SonicMaster: Towards controllable all-in-one music restoration and mastering. arXiv. https://arxiv.org/abs/2508.03448

Ye. Labun, B. Khudik, N. Trintina

AUTOMATION OF THE AUDIO MASTERING PROCESS USING MACHINE LEARNING ALGORITHMS

The article examines the problem of automating the audio mastering process using machine learning algorithms. The relevance of the study is determined by the fact that modern digital tools have made music production accessible to a wide range of users, while high-quality mastering still requires specialized knowledge, critical listening skills, technical control, and significant time resources. This creates a barrier for independent musicians, small studios, and creators who do not have access to professional audio engineers. The paper summarizes the main stages of audio signal analysis, including spectral analysis, equalization, dynamic processing, loudness control, and the evaluation of signal properties using objective and subjective criteria. Particular attention is paid to the capabilities of convolutional neural networks, recurrent neural networks, and LSTM networks in audio processing tasks, music genre classification, and the analysis of the temporal structure of an audio signal. The article analyzes scientific approaches to sound quality assessment, automatic mastering, and the use of reference or professionally mastered audio recordings as a basis for training models. The GTZAN dataset, which contains music fragments of different genres, is considered as an example of an experimental basis. A comparison of neural network types is presented in terms of classification accuracy, number of training epochs, and approximate training time. The summarized results indicate that CNNs are appropriate for spectrogram analysis and frequency feature extraction, while RNNs and LSTMs are better suited for sequential data and musical context modeling. The study concludes that machine learning can serve as an effective foundation for tools supporting automated mastering; however, complete replacement of human expertise remains premature due to the importance of artistic intention, genre context, and subjective listener perception.

Keywords: machine learning, mastering, sound recording, artificial intelligence, audio processing, sound design, neural networks.

Надійшла до редакції: 05.06.2026
Прийнята до друку: 25.06.2026
Опубліковано: 30.10.2026
© 2026 Лабун Є. М., Худік Б. О., Трінтіна Н. А.
Цей матеріал ліцензовано за умовами CC BY 4.0.https://creativecommons.org/licenses/by/4.0/