IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 6, December 2025, pp. 4933~4942
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i6.pp4933-4942

Classifying classical music's therapeutic effects using deep learning: a review

Angelin, Samuel Ady Sanjaya, Dinar Ajeng Kristiyanti
Department of Information Systems, Faculty of Engineering and Informatics, Universitas Multimedia Nusantara, Tangerang, Indonesia

Article Info

Article history:

Received Nov 21, 2024
Revised Aug 6, 2025
Accepted Sep 7, 2025

Keywords:

Deep learning
Feature selection
Mental health
Music classification
Music elements
Music therapeutic effects
Music therapy

ABSTRACT

Mental health issues are the leading cause of global disability, increasing the need for treatment options. While there is much research on the emotional recognition of music in general, there is a gap in studies that directly connect musical features with their therapeutic effects using deep learning. This systematic literature review explores the use of deep learning in classifying the therapeutic effects of classical music for mental health. Following the preferred reporting items for systematic reviews and meta-analyses (PRISMA) framework, a total of 15 papers were reviewed. This review synthesized studies on the role of musical elements that affect mental states. Different feature extraction methods, including mel-frequency cepstral coefficients (MFCCs), spectral contrast, and chroma features, are discussed for their roles in classifying these therapeutic effects. This review also looks at deep learning algorithms like convolutional neural network (CNN), deep neural network (DNN), long short-term memory (LSTM) network, and combined models to assess their effectiveness. Common evaluation methods are also assessed to measure the performance of these models in audio classification. This review highlights the potential of combining deep learning with classical music for mental health support, and to future possibilities for applying these methods in the real world.

Corresponding Author:

Samuel Ady Sanjaya
Department of Information Systems, Faculty of Engineering and Informatics
Universitas Multimedia Nusantara
Tangerang, Indonesia
Email: samuel.ady@umn.ac.id

1. INTRODUCTION

Mental disorders have been identified as the leading cause of global disability [1]–[3]. Mental health is becoming an increasingly important issue worldwide, with more and more people experiencing anxiety, depression, and other emotional or psychological challenges. As a result, there is a growing interest in finding effective treatments that can help improve mental health [4]. Among the various methods explored, music therapy has attracted significant interest [5]. For many years, music has been used and researched to help mental health, from boosting mood and reducing stress, to treating clinical conditions [6]–[8]. In particular, the use of classical music. Studies have shown that classical music can do more than just entertain, it can reduce stress, enhance mood, and even positively affect cognitive functions like memory and attention [9]. Classical music stands out in mental health therapy because of its emotional depth and structured compositions. It's also been found to lower heart rates and reduce blood pressure, which makes it a tool for improving both mental and physical health [10]. Because it affects both mind and body, classical music is increasingly seen as a valuable tool for mental health treatments [11]. Recent studies have focused on how music's emotional and psychological effects can be classified using advanced computational techniques.

Studies have shown that different musical features, such as tempo, harmony, and rhythm, can elicit various emotional responses [12]. However, connecting these musical features with therapeutic outcomes remains underexplored [13]–[15], especially when aiming to automate this process through deep learning. Several studies have utilized machine learning and deep learning algorithms to explore and predict the therapeutic effects of music [15]. According to Modran *et al*. [16], deep learning was employed to identify the emotional effects of music by training a neural network to categorize music into emotions like happy, sad, calm, and energetic. The researcher used mel-frequency cepstral coefficients (MFCCs), which capture the spectral properties of the audio, to predict the therapeutic effect of the music. That study highlighted the importance of connecting audio features with emotional feelings. However, it focused on various music genres without thoroughly exploring classical music's therapeutic potential. Similarly, Dutta and Chanda [17] employed a hybrid convolutional neural network-long short-term memory (CNN-LSTM) deep neural network (DNN) to classify music emotions, showing high accuracy when applied to a dataset of Assamese songs and the Ryerson audio-visual database of emotional speech and song (RAVDESS) emotional song database. But like many previous studies, it is focused on emotion recognition rather than the therapeutic outcomes. Despite these studies, a significant gap remains in connecting musical features directly with therapeutic outcomes, particularly for classical music. Current research often emphasizes general emotion classification or music recommendation systems, not the specific therapeutic effects, such as stress reduction or mood enhancement. This gap opens an opportunity for exploring a more targeted approach that combines deep learning with feature extraction techniques to classify the therapeutic impacts of classical music on mental health, therefore addressing the therapeutic aspects.

This systematic literature review uses the preferred reporting items for systematic reviews and meta-analyses (PRISMA) framework to evaluate the use of feature extraction methods, deep learning algorithms, model evaluation, and musical elements in classifying audio. Audio signal processing commonly uses feature extraction techniques to capture key characteristics like timbre, pitch, amplitude, and rhythm [18]. These feature extraction methods can make raw audio data into a structured format so that it can be analyzed computationally [19], [20]. Common feature extraction techniques include MFCCs, chroma features, spectral contrast, and zero-crossing rate [21]–[23]. Musical features are also essential in understanding how musical compositions can impact someone and their mental state [24], [25]. The researcher aims to identify which specific elements of music play a significant role in contributing to its effects on mental state. This review also compares feature extraction techniques to determine the most common and effective methods for extracting musical features.

Deep learning offers an efficient approach for handling the complex data produced by the feature extraction method [26]. Models such as CNNs and LSTM networks are particularly suited for classifying and analyzing time-series data, like music [27], [28]. Moreover, this review provides an original contribution by systematically synthesizing and critically analyzing existing studies on the use of feature extraction methods and deep learning algorithms in classifying the therapeutic effects of music on mental health, with a focus on classical music. By following the PRISMA guidelines and well-defined inclusion criteria, this systematic literature review ensures a rigorous and transparent approach. This review also aims to be a useful reference for future research, emphasizing the methods that can be utilized in this area.

2. METHOD

The PRISMA framework is applied in this systematic review to guide the comprehensive and transparent identification and assessment of relevant literature. The use of PRISMA also helps reduce the likelihood of reporting inaccuracies in systematic reviews and enhances the clarity and transparency of the review process [29], [30]. The researcher uses the PRISMA framework closely to ensure a systematic, transparent, and consistent selection process, which strengthens the reliability and validity of the findings presented in this review. The aim is to also effectively investigate the papers regarding the theory, tools, methods, algorithms, and evaluation that are used to classify audio-based datasets.

2.1. Research questions

To give this review a clear focus, the research questions were chosen based on research gaps and limitations that were found in previous studies. While many studies explore the emotional effects of music on the mental state, there is still a lack of research that directly connects specific musical elements with measurable therapeutic outcomes. This gap led to these research questions for this systematic literature review, which aim to explore the less explored areas. Many previous studies have also focused on classifying general emotions or music recommendation systems, rather than classifying the specific therapeutic effects of the music. The limitations helped to guide the research questions to focus on which musical elements, feature extraction methods, and deep learning algorithms that are commonly used and effective in classifying the therapeutic effects of music. To make sure that the classification performs well, the research question also aims to evaluate how well various evaluation methods are used to assess the performance of deep learning models in audio-based classification. Table 1 shows the research questions for this systematic literature review.

Table 1. Research questions

Index	Research questions
RQ1	Which elements of music are usually used to determine the effects on mental state?
RQ2	What are the most common and effective feature extraction methods for classifying the impact of classical music on mental health?
RQ3	What deep learning algorithms are most effective for classifying the therapeutic effects of classical music?
RQ4	What methods are employed for evaluating the performance of deep learning models in audio-based classification?

2.2. Search strategy

Publish or perish (PoP) is a widely used tool for academics and researchers to conduct and retrieve academic papers from various online databases [31]. It simplifies the process of gathering and analyzing research papers by collecting results from sources such as Google Scholar, Scopus, and IEEE Xplore, providing citation metrics like the h-index, g-index, and total citations for selected papers [32]. Thus, for this systematic literature review, the researcher identified relevant literature using PoP software to conduct systematic searches across electronic databases, including Google Scholar, Scopus, and IEEE Xplore. Figure 1 shows the PRISMA framework flow diagram, illustrating from the initial identification of papers collected, to the relevant studies that are included in this review.

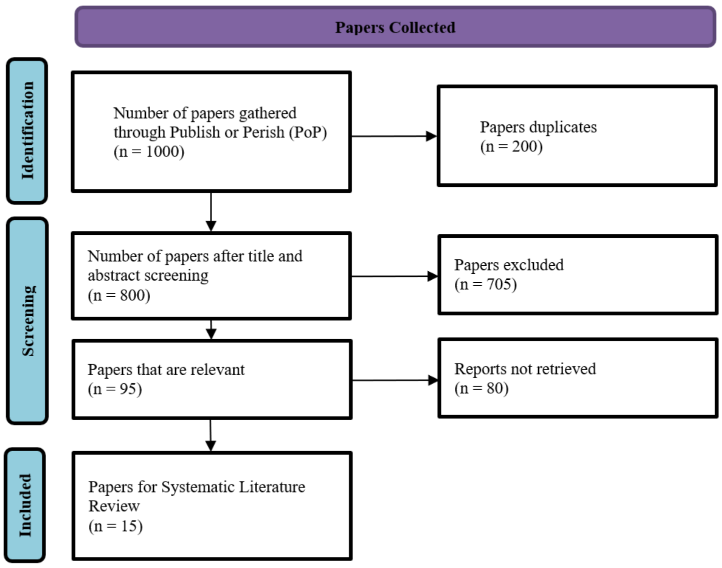

Figure 1. PRISMA framework flow diagram

The search terms included keywords such as “classical music” AND “mental health”, “music for mental health”, “deep learning” AND “music”, “music” AND “classification”, “sound classification”, “audio classification”, “feature extraction” AND “therapeutic effects”, “feature extraction”, “music therapy”, “music emotion classification” AND “deep learning”. The searches focused on journal articles published in English within the last 5 years. The search results are reviewed using the PRISMA method, which involves several steps [33]. It starts by identifying all relevant papers, and then narrowing them down based on specific criteria to select the studies that are most useful for this review.

The inclusion criteria for this systematic literature review are designed to prioritize the selection of publications that are both recent and relevant. The review focuses on peer-reviewed articles, or conference papers published in English from 2019 onwards, to make sure the research is recent. Studies must focus on topics such as the effects of music on mental state, or how deep learning is used in music analysis, and or feature extraction methods like MFCC for understanding music’s emotional or therapeutic effects. These criteria help to make sure that the selected studies are directly related to the review’s aim and research questions.

After the searches or identification, titles and abstracts will be screened for relevance to the research questions, and a full-text review will be conducted for papers meeting the inclusion criteria. The last stage involves analyzing and synthesizing to answer the research questions. A narrative synthesis is used to compare the findings of the reviewed studies. The results of the review are structured, summarizing key findings and highlighting gaps in current research. The limitations of the reviewed studies are discussed, along with recommendations for future research.

3. RESULTS AND DISCUSSION

The researcher conducted a screening and selection process on 800 papers to finally include only 15 papers that met the inclusion criteria. By analyzing specific parts of the selected papers, the researcher aims to answer the (RQ1, RQ2, RQ3, and RQ4) questions. The analysis looks at the data of musical elements such as melody, tempo, and rhythm, that can affect the mental state. Also, the musical features can be produced from raw audio using different feature extraction methods to do this processing. The review also analyzes the specific deep learning models used (e.g., CNN, recurrent neural networks (RNNs)) to classify music. Lastly, this review also examines evaluation metrics like accuracy, precision, and recall, evaluating the performance of the models. Table 2 provides a summary of six selected research papers that explore the relationship between specific musical features and their therapeutic effects. Table 2 presents the features extracted from the music, the mental or emotional benefits reported (such as relaxation, focus, or stress reduction), and the corresponding findings and classification results. This overview helps highlight which features are most commonly associated with certain emotional or therapeutic outcomes and how effective different models have been in recognizing them.

Table 2. Research paper summaries of key musical features and their corresponding therapeutic effects on mental states

Reference	Features of music	Therapeutic effects	Result
[16]	i) Tone class profile ii) Clarity of the key iii) Harmonic change iv) Musical module v) Beat histogram vi) Medium tempo	Calm, happy, energetic, and sad	The researcher can classify calm, happy, and energetic songs with 94% accuracy, but the accuracy for sad songs is a bit lower at 85%.
[34]	i) Tempo ii) Timbre iii) Amplitude	Relaxing	The researcher found that changing one part of the musical element from 'relaxing' music affected people's heart rate, breathing, and even skin reactions.
[35]	Beats	Relaxing	The researcher found that songs can be an effective tool for reducing acute stress.
[36]	i) Tempo ii) Beats	Happy, relax, sad	The researcher successfully classified 'happy' and 'sad' with an accuracy of 80%-90%, but for 'relax', the accuracy is lower, around 60%-70%, so it is still in the poor category.
[37]	i) Tempo ii) Timbre	Focusing	Although features like tempo and timbre of the music can influence focus, the familiarity of the song has a significantly greater impact on enhancing attention and maintaining focus.
[38]	Tempo	Stress reduction, motivating	Slow background music successfully reduced the stress experienced by students during tough dental training. It is also found to increase their motivation to learn and practice.

3.1. RQ1: which elements of music are usually used to determine the effects on mental state?

The research paper summaries reveal several musical features that significantly influence emotions and cognitive responses. Table 3 shows the overall comparison of the key musical features and the effects on mental states. Tempo is a versatile element, faster tempos tend to enhance motivation and alertness, making listeners feel more active, while slower tempos are associated with relaxation and stress reduction [39]. Similarly, timbre, or the unique tonal quality of an instrument or voice, affects how engaging or soothing the music feels. Softer timbres are often associated with relaxation, whereas brighter, sharper timbres can help maintain attention and focus. Amplitude influences the volume or intensity of music, which directly impacts physiological responses like heart rate and stress levels. Softer volumes are often associated with a calming effect. A steady beat can affect the sense of stability that can help to relax and ease acute stress [40]. Also, certain harmonic patterns can evoke calmness, sadness, to nostalgia.

It is found that the effects on the mental state do not just come from one element of music. Rather, the combination of tempo, timbre, amplitude, beats, and harmonic structures creates the effects on the mental state. For example, the mix of tempo and timbre can really affect how people feel. Studies have shown that faster tempos combined with softer sounds can affect someone's motivation and happiness, while slower tempos combined with sharper sounds can improve focus and clarity [34], [36]–[38]. Analyzing these combinations of musical features helps researchers understand how music affects the mind. This makes the process of modeling or classifying the emotional impact of music easier. Table 4 summarizes various research papers that utilize feature extraction techniques and classification algorithms to analyze an audio-based dataset. This table includes information on the types of machine learning or deep learning algorithms used, tools and methods for feature extraction such as MFCC, and evaluation metrics such as accuracy, precision, recall, and F1-score.

Table 3. Comparison of key features and the effects

Musical feature	Indicators and the effect on mental state	Studies supporting
Tempo	Faster tempos enhance motivation and focus, and slower tempos enhance relaxation	[16], [34], [36]–[38]
Timbre	Softer tones promote relaxation and sharper tones promote focus	[34], [37]
Amplitude	Lower volumes calm and reduce stress	[34]
Beats	Steady rhythms reduce acute stress	[35]
Harmonic patterns	Evokes emotions such as calmness, nostalgia, or sadness	[16]

Table 4. Research paper summaries of feature extraction and classification techniques

Reference	Algorithm	Tools and methods	Evaluation	Results
[16]	DNN	Librosa, MFCC	Confusion matrix	An overall accuracy of more than 94% was achieved by the model
[41]	CNN, gated linear units (GLU), recurrent convolutional neural networks (RCNN), residual gated linear units (RGLU)	Librosa, MFCC	Accuracy (%), Confusion matrix	The highest model accuracy achieved was 87%.
[17]	Decision tree, random forest, K-nearest neighbors, multilayer perceptron (MLP), LSTM neural network, CNNs, and CNN-LSTM hybrid	MFCC, Mel Spectrogram, Chroma features	Accuracy (%)	Using MFCC, Mel spectrogram, and chroma features with decision tree, random forest, MLP, and K-nearest neighbors, the model accuracy is 89.66%. With LSTM, CNN, and CNN-LSTM models, it achieves 85% accuracy.
[18]	MLP, sequential minimal optimization (SMO) for SVM, random forest, DNN	MFCC, Librosa	Confusion matrix, accuracy, AUC, F1-score, precision, recall	The SMO model achieved an AUC of 0.969, an accuracy of 0.860, an F1-score of 0.861, a precision of 0.865, and a recall of 0.860. The random forest model had an AUC of 0.987, an accuracy of 0.855, an F1-score of 0.849, a precision of 0.863, and a recall of 0.855. The neural network model performed the best, with an AUC of 0.989, an accuracy of 0.889, an F1-score of 0.889, a precision of 0.891, and a recall of 0.889.
[42]	CNN-LSTM	MFCC	Accuracy	Using the Million Song Dataset, the combined audio features network classifier achieves an audio classification accuracy of 68%, a lyrics classification accuracy of 74%, and an average multimodal accuracy of 78%, which is better than using a single mode.
[43]	CNN, RNN	MFCC	Confusion matrix, precision, recall, F1-score	The model achieves an accuracy of 85.36% on a 10-genre classification task. When trained and tested on the MagnaTagATune dataset with 18,476 29-second clips, it achieves an accuracy of 86.06%.
[44]	CNN	Music emotion classification system (MECS) spectrogram	Accuracy (%), loss	The MECS models have varying results: MECS 1 has better accuracy and lower loss than MECS 2, while MECS 3 has accuracy results of 82.3%, 82%, and 81.6% across three tests.
[27]	LSTM-DNN	MFCC	Accuracy, precision, recall, F-measure	The overall accuracy is 99.19%, using 10-fold cross-validation.
[45]	CNN	MFCC	Accuracy, loss	The researcher found that the recognition accuracy of the proposed method is 92.06%.

Tables 2 and 3 show the overall comparison of the research papers and provide an analysis of the various techniques used to classify music or audio-based classification. One of the focuses of these studies is on the methods of feature extraction, with MFCC showing as one of the most commonly used techniques. Due to their ability to represent audio in a way that aligns with human perception of sound, MFCCs are widely applied in audio signal processing [46]. In audio processing, this method is crucial for transforming raw audio into a form that is suitable for classifying emotional responses. Besides MFCC, previous researchers have also used methods like the Mel spectrogram and chroma features to explore the frequency and harmony of music more closely. These techniques help them to understand the musical elements that affect emotions, helping them to classify using machine learning or deep learning algorithms. To better illustrate these relationships, Figure 2 presents a research mind map that summarizes the connections between feature extraction techniques, classification models, and evaluation methods. This visualization provides a clearer understanding of how different components work together in music-based emotion classification.

In summarized studies, classification methods often use deep learning algorithms, with CNNs being the most common choice. CNNs are good at capturing spatial features, which makes them effective for analyzing complex patterns in music data [47]. Because music happens over time, it often needs models that can work with this kind of data. As a result, CNNs are often combined with LSTM networks to create hybrid models, like CNN-LSTM, which can better understand parts of music [48]. Other deep learning algorithms, like DNN, are also often used for their ability to capture patterns within large datasets by passing data through multiple layers of neurons. For the same reason that CNN is combined with LSTM, DNN is frequently combined with LSTM to handle the temporal parts of the music. RNNs are also used for their strength in processing sequences.

More advanced models like GLU, RCNN, and RGLU are also designed to capture both spatial and temporal features, making them versatile for detailed music classification tasks [49]. Meanwhile, MLP is used as a simpler, fully connected neural network, and although they don't specialize in sequential data, they are still applied for certain types of music classification to analyze the data [50]. These models are evaluated using different measures, like accuracy, precision, recall, and the F1-score, to ensure that the classification is reliable. Some studies also take a multi-modal approach by combining audio and lyrics, which helps to give more understanding of how different aspects of music influence emotion.

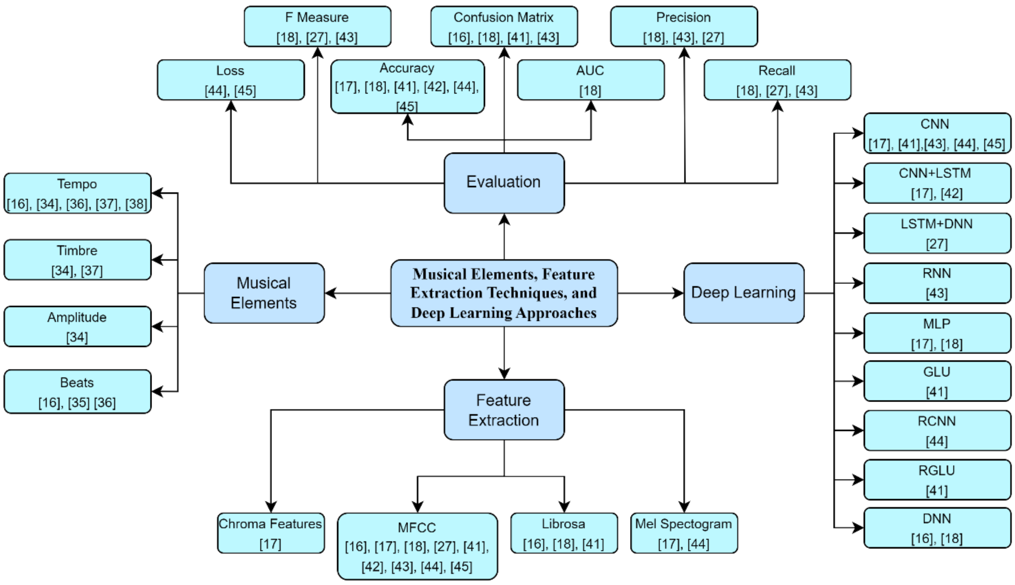

Figure 2. Research mind map

3.2. RQ2: what are the most common and effective feature extraction methods for classifying the impact of classical music on mental health?

Based on the research summaries in Table 3, it is evident that the most common and effective feature extraction method used for classifying audio is MFCCs. In almost all of the studies, MFCC is used as the feature extraction method. MFCCs can effectively extract audio features like timbre, pitch, and rhythm [19], [20]. Its effectiveness is also demonstrated in numerous studies, where models using MFCC consistently achieve high accuracy. For example, in the study that uses LSTM-DNN for music emotion recognition, MFCC helped the model to reach an accuracy of 99.19%. Besides MFCC, other feature extraction methods, such as the Mel spectrogram and chroma features, are also used, but less frequently. Mel spectrogram and chroma features are also used in studies, combining them with advanced deep learning models like CNN and CNN-LSTM. The Mel spectrogram emphasizes change over time, which is useful for music classification, while chroma captures the pitch and the harmonic content [51]. These methods provide additional representations of the audio's frequency and harmonic structure, further enhancing the classification process. The combination of these features with deep learning algorithms significantly enhances the accuracy and effectiveness of models aimed at assessing the emotional and therapeutic effects of classical music.

3.3. RQ3: what deep learning algorithms are most effective for classifying the therapeutic effects of classical music?

Table 5 shows that the studies use several deep learning algorithms, but CNN are shown as one of the most commonly used for music classification. CNNs, which are usually used for image and pattern recognition tasks [52], adapt well to music classification by identifying complex, hierarchical patterns within the audio features extracted from the music [53]. In several studies, CNNs have been used both as standalone models and in combination with other architectures such as LSTM networks, which are particularly good at handling time-sequenced data [54]. Although CNNs are the most commonly used, it is shown that CNN, as a standalone model, only reaches 92.06% accuracy. The hybrid model CNN-LSTM approach has been shown to achieve results with an overall accuracy of 85%, not as high as LSTM-DNN, which reaches an accuracy of up to 99.19%. Other deep learning algorithms, such as MLP and RNN, also demonstrate good performance, but CNN and DNN-based algorithms generally achieve higher accuracy, making them the preferred choice for music classification tasks. It concludes that CNN-LSTM models are usually used for audio-based classification, even though their accuracy varies across studies, but LSTM-DNN has proven to have the highest accuracy.

Table 5. Comparison of deep learning accuracy algorithms

Algorithms used	Accuracy (%)	Studies reference
DNN	94	[16]
CNN, GLU, RCNN, RGLU	87	[41]
MLP, LSTM neural network, CNNs, and CNN-LSTM hybrid	89.66, 85	[17]
MLP, DNN	88.9	[18]
CNN-LSTM	68	[42]
CNN, RNN	86.06	[43]
CNN	82.3	[44]
LSTM-DNN	99.19	[27]
CNN	92.06	[45]

3.4. RQ4: what methods are employed for evaluating the performance of deep learning models in audio-based classification?

Researchers commonly use accuracy metrics alongside other evaluation techniques to comprehensively assess the performance of deep learning models in audio-based classification. Accuracy (%) is the most frequently used metric, indicating how accurately the model's predictions match the true classification outcomes [55]. However, models are often further evaluated using a confusion matrix, which provides insight into specific prediction errors across different classes. Besides accuracy, other metrics like precision, recall, F1-score, and area under the curve (AUC) are often used to evaluate how well a model can classify [56]. Precision measures the accuracy of positive predictions, while recall assesses the model's effectiveness in detecting all of the true positive cases. The F1-score combines precision and recall into one value, which makes it useful when there is an uneven number of classes. AUC, on the other hand, measures how well the model separates the classes at different thresholds, with higher AUC meaning better performance [57]. One of the reviewed studies employs cross-validation techniques, such as 10-fold cross-validation, to ensure that the model's performance is not biased by any particular subset of the data [58].

It is found that not all evaluation methods work effectively in all contexts. Even though accuracy is commonly used, it can be misleading with imbalanced data, where one class is much more common than the other. In that case, a model might seem accurate while missing important predictions for smaller classes. This is when precision, recall, and F-measure can be used effectively because the F-measure balances precision and recall.

From all of the findings, it is concluded that CNN, DNN, and LSTM algorithms can really be used to classify the therapeutic effects of classical music on mental health effectively. These models can be designed to analyze audio data to reveal how different musical elements influence emotions and mental states. CNNs are good at recognizing patterns like rhythm and tone, while the combination of CNNs and LSTM allows for a deeper understanding of how music is over time. DNNs can discover complex patterns in large datasets. Using MFCCs as a feature extraction method is beneficial, incorporating features like chroma or spectral contrast can enhance the model's ability to understand important aspects of music, such as harmony and tonality. The results from various studies support the idea that these methods can effectively classify the impact of classical music or its therapeutic effects on mental health. By not only focusing on accuracy but also considering metrics like F1-score, accuracy, and receiver operating characteristic (ROC)-AUC, researchers can create more reliable models that accurately assess the mental health benefits of classical music.

4. CONCLUSION

This systematic literature review highlights key elements and methods used in classifying the effects of music, particularly classical music, on mental health. Musical features such as tempo, timbre, amplitude, and harmonic patterns significantly influence emotional and cognitive states, with faster tempos and brighter timbres boosting alertness, while slower tempos and softer tones affect relaxation. For feature extraction, MFCC is the most commonly employed and effective method. CNN and hybrid DNN-LSTM models are identified as the most effective for classifying music's therapeutic effects, achieving high accuracy. To assess these models, evaluation methods like accuracy, confusion matrices, and metrics such as precision, recall, F1-score, and AUC are used with cross-validation. These findings point to the importance of both musical features and advanced deep learning methods in studying music's impact on mental health. This review highlights useful approaches, but there are some limitations. Deep learning models often need a lot of data, but finding enough research papers on classical music's therapeutic effects is challenging. Also, many studies focus on general music features and might miss the specific therapeutic qualities of classical music. Future research could help by reviewing more papers that explore these unique therapeutic features. It could also look into using these models in real-time, especially in clinical settings where quick responses are needed.

ACKNOWLEDGEMENTS

Thanks to the Big Data Lab for providing the essential resources and infrastructure.

FUNDING INFORMATION

This research is supported by the Research, Innovation, and Sustainability Division of Universitas Multimedia Nusantara.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Angelin	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Samuel Ady Sanjaya		✓	✓	✓			✓			✓		✓	✓	✓
Dinar Ajeng Kristiyanti		✓	✓	✓			✓			✓		✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest related to this work.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

[1] T. L. Golden *et al.*, "The use of music in the treatment and management of serious mental illness: a global scoping review of the literature," *Frontiers in Psychology*, vol. 12, Mar. 2021, doi: 10.3389/fpsyg.2021.649840.

[2] A. Fonseca and J. Osma, "Using information and communication technologies (ICT) for mental health prevention and treatment," *International Journal of Environmental Research and Public Health*, vol. 18, no. 2, Jan. 2021, doi: 10.3390/ijerph18020461.

[3] B. Wies, C. Landers, and M. Ienca, "Digital mental health for young people: a scoping review of ethical promises and challenges," *Frontiers in Digital Health*, vol. 3, Sep. 2021, doi: 10.3389/fdgth.2021.697072.

[4] N. R. Provenza *et al.*, "The case for adaptive neuromodulation to treat severe intractable mental disorders," *Frontiers in Neuroscience*, vol. 13, Feb. 2019, doi: 10.3389/fnins.2019.00152.

[5] B. Foster, S. Pearson, A. Berends, and C. Mackinnon, "The expanding scope, inclusivity, and integration of music in healthcare: recent developments, research illustration, and future direction," *Healthcare*, vol. 9, no. 1, 2021, doi: 10.3390/healthcare9010099.

[6] D. L. Bowling, "Biological principles for music and mental health," *Translational Psychiatry*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41398-023-02671-4.

[7] E. Liebman and P. Stone, "Utilizing mood-inducing background music in human-robot interaction," *arXiv:2308.14269*, 2023.

[8] J. Huang and X. Li, "Effects and applications of music therapy on psychological health: a review," in *Proceedings of the 2021 International Conference on Public Art and Human Development*, 2022. doi: 10.2991/assehr.k.220110.186.
[9] L. Du, "Classical music styles and piano therapy for patients with depressive disorders," *Revista Música Hodie*, vol. 21, Nov. 2021, doi: 10.5216/mh.v21.66519.
[10] O. Tanriverdi, T. Karaoglu, and F. N. Aydemir, "Can music and medicine be effective on anxiety, depression, and chemotherapy-related nausea and vomiting? (PEGASUS study)," *Indian Journal of Cancer*, vol. 60, no. 2, pp. 282–291, 2023, doi: 10.4103/ijc.IJC_387_20.
[11] K. Devendran, S. K. Thangarasu, P. Keerthika, R. M. Devi, and B. K. Ponnarasee, "Effective prediction on music therapy using hybrid SVM-ANN approach," *ITM Web of Conferences*, vol. 37, Mar. 2021, doi: 10.1051/itmconf/20213701014.
[12] T. Theorell and E. B. Horwitz, "Emotional effects of live and recorded music in various audiences and listening situations," *Medicines*, vol. 6, no. 1, Jan. 2019, doi: 10.3390/medicines6010016.
[13] Y. Ding *et al.*, "Uncovering potential distinctive acoustic features of healing music," *General Psychiatry*, vol. 36, no. 6, Dec. 2023, doi: 10.1136/gpsych-2023-101145.
[14] M. Hartmann *et al.*, "Musical interaction in music therapy for depression treatment," *Psychology of Music*, vol. 51, no. 1, pp. 33–50, Jan. 2023, doi: 10.1177/03057356221084368.
[15] A. Raglio *et al.*, "Machine learning techniques to predict the effectiveness of music therapy: a randomized controlled trial," *Computer Methods and Programs in Biomedicine*, vol. 185, Mar. 2020, doi: 10.1016/j.cmpb.2019.105160.
[16] H. A. Modran, T. Chamunorwa, D. Ursuțiu, C. Samoilă, and H. Hedeșiu, "Using deep learning to recognize therapeutic effects of music based on emotions," *Sensors*, vol. 23, no. 2, Jan. 2023, doi: 10.3390/s23020986.
[17] J. Dutta and D. Chanda, "Music emotion recognition and classification using hybrid CNN-LSTM deep neural network," *Bangladesh Journal of Multidisciplinary Scientific Research*, vol. 9, no. 3, pp. 21–32, Aug. 2024, doi: 10.46281/bjmsr.v9i3.2230.
[18] U. Ayvaz, H. Gürüler, F. Khan, N. Ahmed, T. Whangbo, and A. A. Bobomirzaevich, "Automatic speaker recognition using Mel-frequency cepstral coefficients through machine learning," *Computers, Materials & Continua*, vol. 71, no. 3, pp. 5511–5521, 2022, doi: 10.32604/cmc.2022.023278.
[19] G. Yoo, S. Hong, and H. Kim, "Emotion recognition and multi-class classification in music with MFCC and machine learning," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 14, no. 3, pp. 818–825, 2024, doi: 10.18517/ijaseit.14.3.18671.
[20] S. Carvalho and E. F. Gomes, "Automatic classification of bird sounds: using MFCC and Mel spectrogram features with deep learning," *Vietnam Journal of Computer Science*, vol. 10, no. 01, pp. 39–54, Feb. 2023, doi: 10.1142/S2196888822500300.
[21] Z. K. Abdul and A. K. A.-Talabani, "Mel frequency cepstral coefficient and its applications: a review," *IEEE Access*, vol. 10, pp. 122136–122158, 2022, doi: 10.1109/ACCESS.2022.3223444.
[22] A. Koduru, H. B. Valiveti, and A. K. Budati, "Feature extraction algorithms to improve the speech emotion recognition rate," *International Journal of Speech Technology*, vol. 23, no. 1, pp. 45–55, Mar. 2020, doi: 10.1007/s10772-020-09672-4.
[23] S. MSK, G. V. K. J, M. V. U, H. K. Vege, and S. Kuchibhotla, "MLP model for emotion recognition using acoustic features," *International Journal of Emerging Trends in Engineering Research*, vol. 8, no. 5, 2020, doi: 10.30534/ijeter/2020/34852020.
[24] B. Xie, J. C. Kim, and C. H. Park, "Musical emotion recognition with spectral feature extraction based on a sinusoidal model with model-based and deep-learning approaches," *Applied Sciences*, vol. 10, no. 3, Jan. 2020, doi: 10.3390/app10030902.
[25] J. Sun, J. Yang, G. Zhou, Y. Jin, and J. Gong, "Understanding human-AI collaboration in music therapy through co-design with therapists," in *CHI Conference on Human Factors in Computing Systems*, 2024, pp. 1–21, doi: 10.1145/3613904.3642764.
[26] A. Aggarwal *et al.*, "Two-way feature extraction for speech emotion recognition using deep learning," *Sensors*, vol. 22, no. 6, Mar. 2022, doi: 10.3390/s22062378.
[27] S. Hizlisoy, S. Yildirim, and Z. Tufekci, "Music emotion recognition using convolutional long short term memory deep neural networks," *Engineering Science and Technology, an International Journal*, vol. 24, no. 3, pp. 760–767, 2021, doi: 10.1016/j.jestch.2020.10.009.
[28] K. Bagastio, R. S. Oetama, and A. Ramadhan, "Development of stock price prediction system using Flask framework and LSTM algorithm," *Journal of Infrastructure, Policy and Development*, vol. 7, no. 3, Oct. 2023, doi: 10.24294/jipd.v7i3.2631.
[29] A. Liberati *et al.*, "The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration," *Journal of Clinical Epidemiology*, vol. 62, no. 10, 2009, doi: 10.1016/j.jclinepi.2009.06.006.
[30] H. Liu, X. Zhou, G. Yu, and X. Sun, "The effects of the PRISMA statement to improve the conduct and reporting of systematic reviews and meta-analyses of nursing interventions for patients with heart failure," *International Journal of Nursing Practice*, vol. 25, no. 3, Jun. 2019, doi: 10.1111/ijn.12729.
[31] N. C. P. Susanto, S. J. Hartati, and R. E. Standsyah, "Systematic literature review: application of dynamic geometry software to improve mathematical problem-solving skills," *Mathline : Jurnal Matematika dan Pendidikan Matematika*, vol. 8, no. 3, pp. 857–872, Aug. 2023, doi: 10.31943/mathline.v8i3.458.
[32] K. Ariyanto, "Rural development research trends: bibliometric analysis using publish or Perish and Vosviewer," *Athena: Journal of Social, Culture and Society*, vol. 1, no. 4, pp. 169–179, Sep. 2023, doi: 10.58905/athena.v1i4.121.
[33] J. Guerrero and I. Alsmadi, "Synthetic text detection: systemic literature review," *arXiv:2210.06336*, Oct. 2022.
[34] E. H.-Ruiz, B. James, J. Noll, and E. G. Chrysikou, "What makes music relaxing? an investigation into musical elements," *Psychology of Music*, vol. 48, no. 3, pp. 327–343, May 2020, doi: 10.1177/0305735618798027.
[35] D. Shepherd, M. J. Hautus, E. Giang, and J. Landon, "'The most relaxing song in the world'? a comparative study," *Psychology of Music*, vol. 51, no. 1, pp. 3–15, Jan. 2023, doi: 10.1177/03057356221081169.
[36] T. B. Santoso and T. Dutono, "Instrumental music emotion recognition with MFCC and KNN algorithm," *Indonesian Journal of Computer Science*, vol. 12, no. 1, Feb. 2023, doi: 10.33022/ijcs.v12i1.3152.
[37] L. Kiss and K. J. Linnell, "The effect of preferred background music on task-focus in sustained attention," *Psychological Research*, vol. 85, no. 6, pp. 2313–2325, Sep. 2021, doi: 10.1007/s00426-020-01400-6.
[38] S. Felszeghy *et al.*, "Influence of background music on stress reduction and impact on performances during students' simulation exercises," *Journal of Dental Education*, vol. 87, no. 8, pp. 1170–1179, Aug. 2023, doi: 10.1002/jdd.13235.
[39] A. Raglio, "A novel music-based therapeutic approach: the therapeutic music listening," *Frontiers in Human Neuroscience*, vol. 17, Jul. 2023, doi: 10.3389/fnhum.2023.1204593.
[40] Q. Ding, "Evaluation of the efficacy of artificial neural network-based music therapy for depression," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–6, Aug. 2022, doi: 10.1155/2022/9208607.
[41] J. Zhang, "Music feature extraction and classification algorithm based on deep learning," *Scientific Programming*, vol. 2021, pp. 1–9, May 2021, doi: 10.1155/2021/1651560.
[42] C. Chen and Q. Li, "A multimodal music emotion classification method based on multifeature combined network classifier," *Mathematical Problems in Engineering*, vol. 2020, pp. 1–11, Aug. 2020, doi: 10.1155/2020/4606027.
[43] D. Bisharad and R. H. Laskar, "Music genre recognition using convolutional recurrent neural network architecture," *Expert Systems*, vol. 36, no. 4, Aug. 2019, doi: 10.1111/exsy.12429.

[44] D. Chaudhary, N. P. Singh, and S. Singh, “Development of music emotion classification system using convolution neural network,” *International Journal of Speech Technology*, vol. 24, no. 3, pp. 571–580, Sep. 2021, doi: 10.1007/s10772-020-09781-0.
[45] X. Jia, “A music emotion classification model based on the improved convolutional neural network,” *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–11, Feb. 2022, doi: 10.1155/2022/6749622.
[46] K. M. Rezaul *et al.*, “Enhancing audio classification through MFCC feature extraction and data augmentation with CNN and RNN models,” *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 7, 2024, doi: 10.14569/IJACSA.2024.0150704.
[47] E. K. Gyasi and S. Purushotham, “Soil-MobiNet: a convolutional neural network model base soil classification to determine soil morphology and its geospatial location,” *Sensors*, vol. 23, no. 15, Jul. 2023, doi: 10.3390/s23156709.
[48] M. Conner, L. Gral, K. Adams, D. Hunger, R. Strelow, and A. Neuwirth, “Music generation using an LSTM,” *arXiv:2203.12105*, 2022.
[49] S. Allamy and A. L. Koerich, “1D CNN architectures for music genre classification,” *arXiv:2105.07302*, May 2021.
[50] R. Refianti and F. Mahardi, “Comparison of music genre classification results using multilayer perceptron with chroma feature and Mel frequency cepstral coefficients extraction features,” *International Journal of Engineering, Science and Information Technology*, vol. 3, no. 2, pp. 53–59, Jun. 2023, doi: 10.52088/ijesty.v3i2.444.
[51] J. Li, L. Han, X. Li, J. Zhu, B. Yuan, and Z. Gou, “An evaluation of deep neural network models for music classification using spectrograms,” *Multimedia Tools and Applications*, vol. 81, no. 4, pp. 4621–4647, Feb. 2022, doi: 10.1007/s11042-020-10465-9.
[52] T. A. Hartanto and S. Hansun, “Comparative analysis of pre-trained CNN models on retinal diseases classification,” *International Journal of Industrial Engineering & Production Research*, vol. 35, no. 3, pp. 1–12, 2024, doi: 10.22068/ijiepr.35.3.1935.
[53] X. Chen, “The application of neural network with convolution algorithm in Western music recommendation practice,” *Journal of Ambient Intelligence and Humanized Computing*, Mar. 2020, doi: 10.1007/s12652-020-01806-5.
[54] Y. Xiao, H. Yin, Y. Zhang, H. Qi, Y. Zhang, and Z. Liu, “A dual-stage attention-based Conv-LSTM network for spatio-temporal correlation and multivariate time series prediction,” *International Journal of Intelligent Systems*, vol. 36, no. 5, pp. 2036–2057, May 2021, doi: 10.1002/int.22370.
[55] J. C. Obi, “A comparative study of several classification metrics and their performances on data,” *World Journal of Advanced Engineering Technology and Sciences*, vol. 8, no. 1, pp. 308–314, Feb. 2023, doi: 10.30574/wjaets.2023.8.1.0054.
[56] A. M. Carrington *et al.*, “Deep ROC analysis and AUC as balanced average accuracy, for improved classifier selection, audit and explanation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 329–341, Jan. 2023, doi: 10.1109/TPAMI.2022.3145392.
[57] V. Singh, M. Pencina, A. J. Einstein, J. X. Liang, D. S. Berman, and P. Slomka, “Impact of train/test sample regimen on performance estimate stability of machine learning in cardiovascular imaging,” *Scientific Reports*, vol. 11, no. 1, Jul. 2021, doi: 10.1038/s41598-021-93651-5.
[58] A. Ali and C. Gravino, “An empirical comparison of validation methods for software prediction models,” *Journal of Software: Evolution and Process*, vol. 33, no. 8, Aug. 2021, doi: 10.1002/smr.2367.

BIOGRAPHIES OF AUTHORS

Angelin holds an undergraduate Bachelor of Information Systems at Universitas Multimedia Nusantara, Indonesia. Her research interests include big data, data analysis, machine learning, deep learning, and art. She can be contacted at email: angelin@student.umn.ac.id.

Samuel Ady Sanjaya holds a Master of Informatics degree from Institut Teknologi Bandung, Indonesia, in 2019. He also received his Bachelor of Informatics from Atma Jaya Yogyakarta University, Indonesia, in 2016, respectively. He is currently a lecturer at Department of Information Systems, Faculty of Engineering and Informatics, Universitas Multimedia Nusantara, Indonesia. His research includes data analysis, machine learning, data mining, and social network analysis. He has published over 5 papers in international journals and conferences. He can be contacted at email: samuel.ady@umn.ac.id.

Dinar Ajeng Kristiyanti received her Master's degree in Computer Science and a Bachelor's degree in Information Systems from Nusa Mandiri University, Indonesia. Currently, she is a Ph.D. candidate in Computer Science at IPB University, Indonesia. She is a lecturer at Department of Information Systems, Faculty of Engineering and Informatics, Universitas Multimedia Nusantara, Indonesia. Her research interests include sentiment analysis, text mining, feature selection, optimization, data science, and machine learning. She has published over 13 papers in international journals and conferences. She can be contacted at email: dinar.kristiyanti@umn.ac.id.