ORIGINAL RESEARCH ARTICLE

A multimodal deep learning algorithm for polyphonic music applied to music sentiment analysis and generation

Qidi Sun, Hyuntae Kim[*]

Department of Applied Music, Sejong University, Seoul 05006, *Korea*

* Corresponding author: Hyuntae Kim, kimht@sejong.ac.kr

ABSTRACT

The acronym for "polyphonic music" (PM) is employed when referring to music which includes different melody lines that are performed sequentially. Integrating PM with Sentiment Analysis (SA) and music composition involves evaluating and creating tunes with several tunes played simultaneously. Advanced techniques, usually centred on Deep Learning (DL) methods, have been employed to achieve the aim. The intention of this study is to provide an innovative framework for monitoring and developing SA in good management. Initially, a particular system of analysis is created, employing sophisticated DL methods to enhance the accuracy and sensitivity of PM detection of sentiments. The research addresses the intricate functioning of sound features like Mel Frequency Cepstral Coefficients (MFCC) and Chroma. This research project investigates whether dimension reduction approaches like Stacked Autoencoders (SAE) enhance PM-SA models. To address computationally demanding issues. The recommended SA system MDL is thoroughly evaluated compared to traditional techniques. Accuracy, precision, recall, and F1-score examine the MDL framework's potential to detect and classify PM sentiment states.

Keywords: sentiment analysis; polyphonic music; accuracy; deep learning; classification and performance; Mel frequency cepstral coefficients

1. Introduction

The music industry has experienced a significant change due to the rapid advancement of communication technologies, intelligent devices, and the proliferation of digital data[1,2]. The emergence of the digital age has not only changed how a person listens to music but has also required a complete reconsideration of the methods used to manage and access the rapidly growing amount of music data. Recently, metadata content has been crucial in finding music[3,4]. However, due to the significant increase in music data, there is a considerable need to transition towards content-based retrieval[5,6].

The proliferation of smart devices, radio stations, and Recommender Systems (RS) has brought about a time when music services are now ubiquitous. In order to improve customer satisfaction and involvement, music service providers have undertaken a mission to intelligently suggest songs, with sentiments being identified as a vital element in this context[6,7].

The process of obtaining music data, which used to rely on genre, artist, and musical instruments primarily, has now expanded to incorporate sentiments as a crucial factor[8,9]. The learning environment for kids is the setting where music's impact is most

apparent[10–12]. Scientific research indicates that listening to music significantly backs kids' psychological well-being in virtual educational environments[13–15]. This emphasises the requirement for Music Information Retrieval (MIR) platforms to present data-driven decision-making and customised features that can be adapted to people's interests[16–18].

Music-based SA systems concentrate on music with instruments, which may trigger sentiments without lyrics. When identifying sentiments from musical instruments, there are plenty of issues to consider. The ambiguous nature of sentiments in musical instruments, the diversity in how identical sentiments are conveyed via various verbs, and a shortage of datasets for detecting sentiments in musical instruments make music-based Sentiment Analysis (SA) problematic[19–21]. Based on the outcomes of this examination, a MuSe sample[22–25] has been determined to be available to the public at large.

Feature Extraction (FE) and classification, with the help of Deep Learning (DL) methods, are the main fields of emphasis for the recognition of patterns used in the context of music research. Furthermore, Mel Frequency Cepstral Coefficients (MFCC) have been the subject of thorough investigation in the context of music-related studies[26–28]. These parameters are frequently used as sound features in voice and music applications. In music-based unique SA proof of identity, chroma features and traits that are associated with their use are leveraged. Additionally, the research into the possibility of recognising sentiments from musical instruments remains an arena that has not been thoroughly investigated because it is still an emerging area.

By describing an extensive technique for recognising SA in musical instruments, this article attempts to deal with the flaws that have been highlighted. The experiment involves recourse to a method for decreasing dimensionality utilising a Stacked Autoencoder (SAE), which includes MFCC and Chroma Energy Normalised Statistics (CENS) data[29–31]. This approach is frequently used in the evaluation of chromatic musical instrument sources. A practical classification method for Polyphonic Music (PM) is feasible using a multimodal Deep Learning (DL) method. These video clips are grouped based on an array of SA, and an SAE is used to minimise the total number of variables in the data that is input. Furthermore, to provide an impact on advancing the field of music SA, the main goal of this study is to provide an understanding of the complex link that exists between musical instruments, sentiments, and developing computer technology.

1.1. Research motivation

The integration of musical instruments, sentiments, and advanced statistical techniques is what motivated the invention of a concept for studying SA in PM. Recognising the mental and sentiment elements that are incorporated into these complex rhythmic patterns is an exciting challenge in the contemporary musical type, which is defined by an extensive range of harmonic tunes that are simultaneously varied and complex. The application of SA is prevalent in the study of text and speech, where the current study provides an intriguing approach to understanding the sentiments aspects that prevail in PM. In the scenario of PM, which is defined by owning a great deal of distinct melodies, there are specific difficulties related to identifying the SA. Traditional methods contingent on metadata are insufficient for understanding the scope and complexity of the SA that mesh musical elements can communicate. The requirement to create solid analytical techniques with the goal of determining the sentiment elements that can be found in PM is what motivates the current study. Online streaming services are the force that drives behind the musical entertainment industry, and they demand user relationships that are both customised and engaging. The introduction of SA into PM, search and recommendation systems is following the increasing need for intelligent music streaming services that accommodate the SA needs of users as individuals. The composition of PM and the absolute need of SA are the two factors considered in this study.

1.2. Research objective

The key objectives of this research are:

a) Create and execute an efficient analytical framework that integrates DL algorithms designed explicitly for the PM to enhance precision and sensitivity in SA.

b) Examine the function of various acoustic features, such as MFCC and Chroma features, in capturing sentiment expression in PM.

c) Evaluate the influence of dimensionality reduction methods, such as Stacked Auto Encoders, on enhancing the efficiency of SA models for PM, thereby reducing computational complexity.

d) Evaluate the SA framework created in this research by thoroughly experimenting and comparing it to existing approaches. This evaluation will be based on accuracy, precision, recall, and F1 score metrics. The goal is to determine the framework's effectiveness in accurately identifying and categorizing sentiments states in PM.

The paper is organized as follows: Section 2 presents the methodology, Section 3 presents the evaluation of the model, Section 4 describes the experimental result, and Section 4 concludes the work.

2. Related works

The section provides an overview of the developing area of Music Information Retrieval (MIR) in the context of pattern recognition, with a particular emphasis on Music Emotion Recognition (MER). Previous studies predominantly focus on MER, specifically analyzing music clips incorporating instrumental and vocal components. However, when exploring the recognition of sentiments in instrumental music, the analysis focuses on instrumental portions and involves two crucial steps: i) FE and ii) recognizing and classifying them. Existing research primarily focuses on acoustic aspects essential for developing SA systems. The sentiment content of audio signals is greatly influenced by timbre, which is efficiently represented by MFCC which plays a prominent role in numerous works[32,33].

Krishnamohan et al.[34] used MFCC as a feature and leverages auto-associative neural networks in their investigation, and the efficient average recognition rate is 94.4%. In their study, Raju et al.[35] utilized spectral roll-off, brightness, energy, rhythm features, and Support Vector Machines (SVM) to reach a SA rate of 87.27%. Suneetha et al.[36], Suneetha et al.[37] and Wagdarikar and Senapati[38] employ a hierarchical SVM model that integrates rhythm features, tempo, articulation, and variation in note length. The accuracy rate of this model is 92.33%.

Janarthanan et al.[39] employ Artificial Neural Networks (ANN) and a 35-dimensional feature vector to categorize music signals into four distinct sentiment groups. Their accuracy rate reached 67.0%.

Machine Learning (ML) techniques are essential for optimizing the effectiveness of SA. Priyadharshini and Gomathi[40] present a regression technique in their work that uses Gaussian Process (GP) regressors to assess the SA of a musical composition precisely. The authors attained remarkable results by utilizing adaptive aggregation approaches. Ali et al.[41] proposed a system that combines MFCC with residual phase features. The results demonstrate that SVM provides superior performance in this system. Pande and Chetty[42] proposed a pattern recognition system that utilizes Recurrent Neural Networks (RNN) with Long Short-Term Memory (LSTM) layers to accurately extract sentiments-specific information from Chromogram features.

Exploring DL techniques, specifically Convolutional Neural Networks (CNN), has proven effective in music SA. The researchers attained an accuracy of 88.5% by utilizing a CNN to classify six sentiments based on music data[43]. This was achieved by transforming the signal into a Chroma spectrogram. Furthermore, Ali et al.[44] introduced a pattern recognition approach that employs RNN to analyze monophonic instrumental music. Meanwhile, a study presents a method that improves the backpropagation algorithm by including an Artificial Bee Colony (ABC) algorithm to boost music SA.

The study by Rani et al.[45] employ a convolutional extended short-term memory model for SA recognition. Utilizing Mel filter banks and MFCCs outperforms the performance of conventional machine

learning techniques. Nana classifies music SA into two categories, joyful or sad, using gated recurrent neural networks[46]. When implemented on a ten-song model, this technique achieves an accuracy of 73.4%. Nevertheless, there are several domains within the literature where the potential of Chroma features remains untapped, and the application of autoencoders in Music Information Retrieval (MIR) research is not thoroughly investigated[47]. The proposed work addresses these deficiencies by conducting a comparative inquiry into the efficacy of instrumental SA from PM samples. The task will be accomplished by employing MFCC and Chroma Energy Normalized Statistics (CENS) features using stacked autoencoders.

Research gap

Modern Music Emotion Recognition (MER) study emphasises vocal and instrumental audio clips, investigating sound qualities and MFCC. However, experiments primarily concentrated on MFCC features and overlooked Chroma features' possibility of SA recognition in instrumental music. Autoencoders' robust feature detectors will probably be substantially investigated for Music Information Retrieval (MIR). SVM, ANN, RNN, and CNN have been frequently employed for music emotion proof of identity, but there is not much research comparing instrumental SA[48]. Combining MFCC and Chroma features with SAE is accurate. In order to better comprehend musical instruments and create reliable SA recognition systems, the study will investigate how well musical methods can identify sentiments in PM tests. In particular, it will use Chroma features and autoencoders in MIR.

3. Proposed methodology: A multimodal DL for PM-based SA

This section explains the dataset and the process of SA from the PM using a multimodal DL algorithm.

3.1. Dataset description

The MuSe dataset includes SA data for 90,001 songs encompassing musical history. Since the social keywords that each music has on tunes. It is possible to assume its tone of voice. As evidence of the database's ability to identify SA along three key parameters, these tags have been created using the database. According to the initial definition of value, it is "the enjoyment of an event" which represents the sentiments and mood of the music as well as its positivity or negativity. In addition, involvement is an indicator of "the level of sentiments caused by a challenge" which is a metric of the sense of excitement that is evoked by the musical composition. Finally, power is an analysis that considers "the level of control exerted by an event," which indicates the sentiments of control or impact that the person listening perceives the music to have upon individuals. For the aim of presenting relevant data about the sentiments nature of music compositions, the MuSe dataset enables an in-depth and complex view of the sentiments that are communicated over an extensive spectrum of 90,001 songs. This is achieved through integrating recordings.

3.2. Features for SA recognition

Autoencoders aid in constructing complex shapes from simpler ones, serving as proficient feature detectors. Both continuous and discrete data forms are fused for learning. This method employs a feed-forward artificial neural network perceptron, where weight values multiply input data. The process involves adding the bias value 'a' to the total inputs and weights, initiating through functions like linear, log-sigmoid, hard limit, and hyperbolic tangent with potential saturation, Equation (1)

$$B = f(wp + a) \tag{1}$$

Usually, the perceptron employs a general function for the estimation process, with the commonly chosen function represented by Equation (2).

$$f(p) = \frac{1}{(1 + e^{-x})} \tag{2}$$

Accurate weight estimation minimizes the error between the output and the expected value within the training dataset. Multiple perceptions are organized into layers, with the output of one layer serving as the

input for the next. This multilayer network enables the solution of complex linear separable classification problems by transmitting input data through the initial weights.

3.3. Classification of sentiment

The proposed Multimodal Deep Learning (MDL) framework uses the MFCC and chroma features. The MDL is responsible for developing a complicated and distinct learning structure.

The framework of MDL is signified in Equations (3) and (4).

$$H_{MFCC} = Layer_{MFCC}(X_{MFCC}) \tag{3}$$
$$H_{Chroma} = Layer_{Chroma}(X_{Chroma}) \tag{4}$$

The features are merged and given in Equation (5).

$$H_{Concatenated} = Concatenated(H_{MFCC}, H_{Chroma}) \tag{5}$$

The output layer in MDL models' SA is given in Equation (6).

$$Y = OutputLayer(H_{Concatenated}) \tag{6}$$

where the derived features from MFCC and Chroma are signified as X_{MFCC} and X_{Chroma} respectively. The output information is given in the sentiment label Y, and the feature processors in the MDL layer are signified as in X_{MFCC} and X_{Chroma} for the MFCC and Chroma features, respectively.

During training, the MDL framework is optimized for SA by utilizing the MuSe dataset. The proposed MDL framework is fine-tuned by the loss function and optimization technique. The disparity between the predicted and true SA label Y is measured using the loss function $\mathcal{L}$. It is determined using an appropriate categorical cross-entropy-based classification given in Equation (7).

$$\mathcal{L} = \sum_{i} Y_i \cdot \log(\widehat{Y_i}) \tag{7}$$

The loss function is optimized to be relevant to the parameters modelled in the MDL framework. The optimization is accomplished using Stochastic Gradient Descent (SGD), which is given in Equation (8).

$$\theta \leftarrow \theta - \alpha \cdot \nabla_\theta \mathcal{L} \tag{8}$$

where a model parameter is signified by θ, the learning rate is signified by α, and gradient loss is signified in $\nabla_\theta \mathcal{L}$. The training process optimizes the biases and weight in the MDL framework, which can effectively enhance the accuracy by utilizing the PM features for the classification model. The overall process of the MDL is given in **Figure 1**.
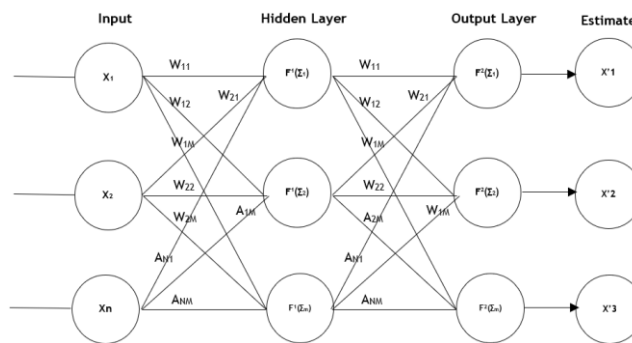

Figure 1. Process of MDL.

4. Result and discussion

The concepts of "Pitches Across Time" and "Timbres Across Time" depict the dynamic differences in the features of timbre and pitch across the musical piece. The heatmaps give significant insights into the melody, sonic attributes, tonal variations, and musicians and assist academics in comprehending the dynamic audio file. Each heatmap in **Figure 2** depicts distinct audio features.
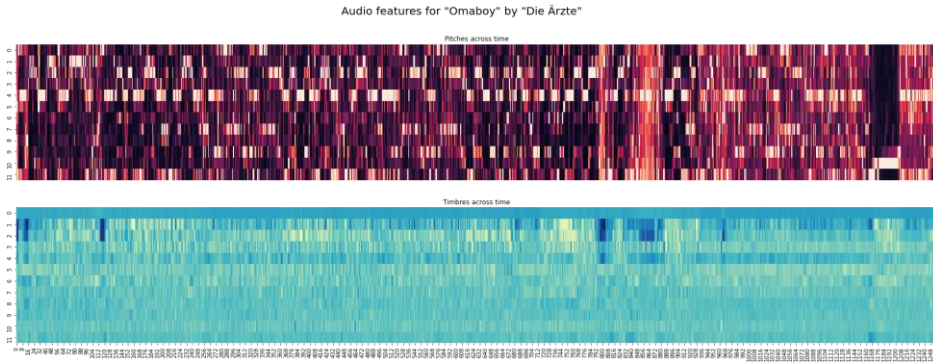
Figure 2. Heat map.

The high-level features represent musical attributes extracted from a Spotify track with ID '63GEJ5qD9Eu4DKq0EYBGVO'. These features include 'Danceability' (0.29), 'Energy' (0.624), 'Key' (9), 'Loudness' (-8.243), 'Mode' (1), 'Speechiness' (0.0611), 'Acoustics' (0.426), 'Instrumentalness' (0.0114), 'liveness' (0.891), 'valence' (0.246), 'Tempo' (201.37), 'Type' ('Audio_Features'), 'ID', 'URI', 'Track_HREF', 'Analysis_URL', 'Duration_ms' (402707), and 'Time_Signature' (3). These metrics provide insights into the track's features, such as danceability, energy level, key, and tempo. The data is retrieved through Spotify's audio features API.

The classification is evaluated using the Accuracy, Precision, Recall, and F1-score performance metrics. The classification metrics and comparative analysis are detailed in a subsequent section.

Accuracy in MDL-SA classification quantifies the model's overall accuracy in predicting SA across several modalities. The evaluation metric considers both TP (sentiments successfully detected) and true negatives (non-sentiments correctly identified) in relation to the total number of instances. The experimental outcome is discussed in **Table 1** and illustrated in **Figure 3**.

Table 1. Comparison of accuracy.

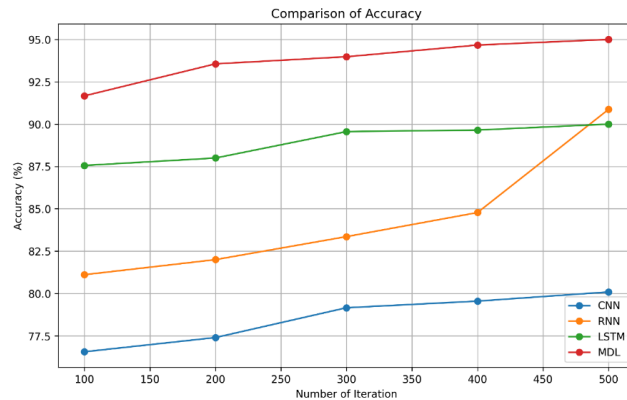| Algorithm	CNN	RNN	LSTM	MDL
100	76.56	81.11	87.56	91.67
200	77.4	82	88	93.56
300	79.16	83.36	89.56	93.98
400	79.55	84.78	89.65	94.67
500	80.09	90.87	90	95

Figure 3. Comparison of accuracy.

In the MDL context, precision measures the model's ability to predict positive attitudes accurately. In the context of SA, accuracy refers to the ratio of accurately recognized positive sentiments (TP) to all instances

6

that were projected as positive (TP and FP). The experimental outcome is discussed in **Table 2** and illustrated in **Figure 4**.

Table 2. Comparison of precision.

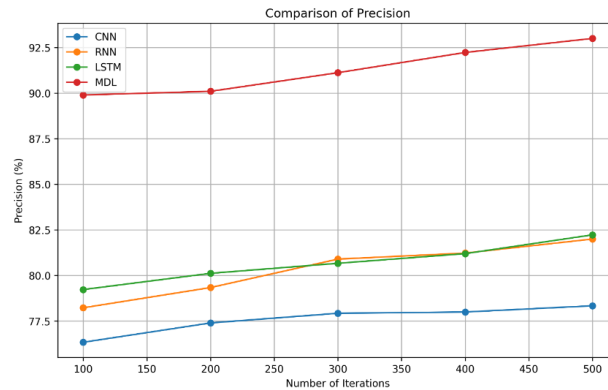| Algorithm	CNN	RNN	LSTM	MDL
100	76.34	78.23	79.23	89.9
200	77.4	79.34	80.12	90.1
300	77.93	80.9	80.67	91.12
400	78	81.23	81.2	92.23
500	78.34	82	82.23	93

Figure 4. Comparison of precision.

In MDL-SA, recall measures the model's capacity to identify all pertinent occurrences of positive attitudes. The metric calculates the proportion of correctly identified positive sentiments (TP) out of the total number of genuine positive sentiments (TP and FN). The experimental outcome is discussed in **Table 3** and illustrated in **Figure 5**.

Table 3. Comparison of recall.

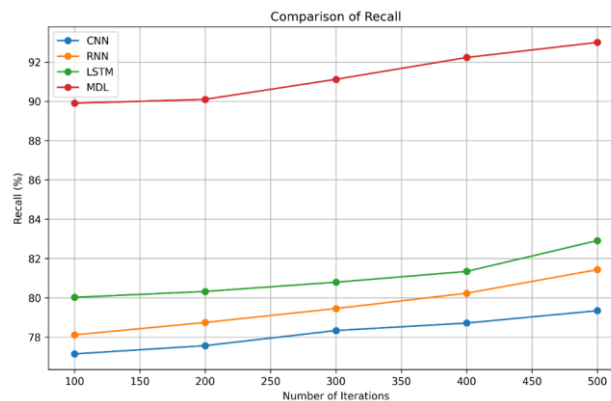| Algorithm	CNN	RNN	LSTM	MDL
100	77.14	78.11	80.02	89.9
200	77.56	78.74	80.32	90.1
300	78.33	79.45	80.79	91.12
400	78.71	80.23	81.34	92.23
500	79.34	81.43	82.91	93

Figure 5. Comparison of recall.

7

As it gives an accurate assessment of both precision and recall, the F1-score is particularly significant in the framework of SA. In the MDL system, the F1-score is employed to determine the harmonic mean of precision and recall. This guarantees that a balanced compromise can be achieved between accurately identifying positive sentiments and avoiding false negatives. The conclusions of the research are given in **Table 4**, and **Figure 6** presents a visual representation of the results that were obtained.

Table 4. Comparison of F1-score.

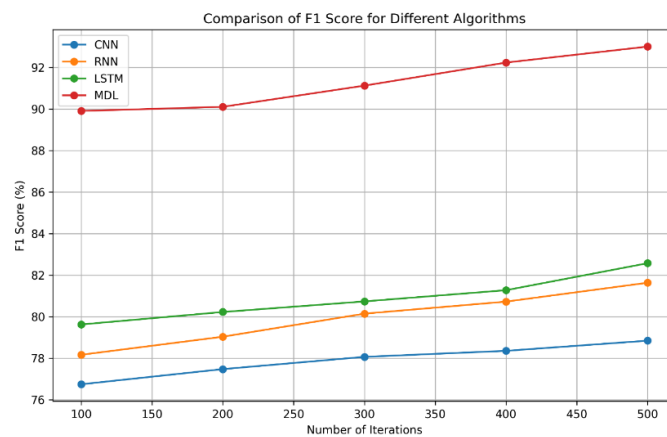| Algorithm	CNN	RNN	LSTM	MDL
100	76.74	78.16	79.62	89.9
200	77.47	79.03	80.22	90.1
300	78.06	80.14	80.73	91.12
400	78.35	80.72	81.27	92.23
500	78.84	81.63	82.57	93

Figure 6. Comparison of F1 score.

Compared with different methods, the MDL-SA approach stands out by its reliability and effectiveness across various dataset sizes. It shows better accuracy, precision, recall, and F1 score. The results of this study shed clarification on the potential of MDL in SA tasks, which necessitate achieving a balance between precision and recall in order to make reliable forecasts about sentiment.

The results of the experiments, which are laid out in **Tables 1** and **4** and represented in **Figures 3** and **5**, present an empirical examination of the different methods employed in SA. The Multi-Modal Deep Learning (MDL) algorithm consistently outperforms its counterparts, CNN, RNN, and LSTM, across multiple metrics. Regarding accuracy (**Table 1** and **Figure 3**), MDL exhibits a marked superiority, achieving an accuracy of 91.67% for a dataset size of 100, escalating to an impressive 95% for a dataset size of 500. Comparatively, CNN, RNN, and LSTM demonstrate lower accuracies across all dataset sizes. The precision results (**Table 2** and **Figure 5**) reinforce MDL's dominance, with precision values ranging from 89.9% to 93% across different dataset sizes, outpacing the other algorithms consistently.

In recall (**Table 3** and **Figure 4**), MDL again demonstrates its prowess, achieving recall values ranging from 89.9% to 93%, surpassing CNN, RNN, and LSTM. With F1-scores that range from 89.9% to 93%, MDL continually outperforms the rest of the algorithms. This level of performance is further reinforced by the F1 score, which can be observed in **Table 4** and **Figure 5**. These numerical analyses prove that MDL performs well in the classification of SA, showing that it is better over a wide range of dataset sizes and demonstrating its promise as an accurate tool in the field of SA. After that, it is of tremendous significance to accept the possibility of boundaries and examine the different paths of research that could be explored to recognize these conclusions better.

5. Conclusion and future work

The research presents a cutting-edge method for analyzing and generating sentiments in the complex field of Polyphonic Music (PM), where frequent melodies intersect. The anticipated Multimodal Deep Learning (MDL) background uses innovative Deep Learning (DL) methods, such as Mel Frequency Cepstral Coefficients (MFCC) and Chroma features, to attain high precision and sensitivity in SA recognition. By combining dimensionality-saving approaches, mainly Stacked Autoencoders (SAE), one can successfully manage computational complexities while sustaining high performance. After systematic testing and assessment, the MDL model exceeds current practices by finding the maximum level of accuracy in precisely identifying and classifying sentiment states in PM. This investigation makes a valuable input to the Music Information Retrieval (MIR) field and creates a state-of-the-art basis for refining Sentiment Analysis (SA) and generation in the complex background of PM compositions. The results highlight the use of the MDL model in increasing the boundaries of accuracy, precision, recall, and F1 score, representing a distinguished improvement in the computational knowledge of sentiment refinements in PM. In order to ensure the validity of the projected method, it is vital to conduct empirical studies in future work. Conducting thorough experiments on a standard of PM tests, encompassing many types and sentiments subtleties, will improve the strength and pertinence of the results.

Furthermore, the study can be expanded to thoroughly investigate the influence of autoencoders, examining different structures and setups to get the most effective Feature Extraction (FE). Incorporating further sophisticated Machine Learning (ML) techniques, such as Deep Neural Networks (DNN), can significantly augment the effectiveness of SA systems.

Author contributions

Conceptualization, HK; methodology, HK; software, HK; validation, HK; formal analysis, QS; investigation, QS; resources, HK; data curation, HK; writing—original draft preparation, HK and QS; writing—review and editing, HK, QS and HK; visualization, HK; supervision, HK; project administration, HK; funding acquisition, QS and HK. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

Conflict of interest

The authors declare no conflict of interest.

References

1. Çano E. Text-based sentiment analysis and music emotion recognition. arXiv. 2018; arXiv:1810.03031.
2. Abboud R, Tekli J. Integration of nonparametric fuzzy classification with an evolutionary-developmental framework to perform music sentiment-based analysis and composition. Soft Computing. 2019; 24(13): 9875-9925. doi: 10.1007/s00500-019-04503-4
3. Xi C. Music Emotion Analysis Based on PSO-BP Neural Network and Big Data Analysis. Ahmed SH, ed. Computational Intelligence and Neuroscience. 2021; 2021: 1-9. doi: 10.1155/2021/6592938
4. Rajesh S, Nalini NJ. Polyphonic Instrument Emotion Recognition using Stacked Auto Encoders: A Dimensionality Reduction Approach. Procedia Computer Science. 2023; 218: 1905-1914. doi: 10.1016/j.procs.2023.01.167
5. Nalini NJ, Palanivel S. Emotion recognition in music signal using AANN and SVM. International Journal of Computer Applications. 2013; 77(2).
6. Pouyanfar S, Sameti H. Music emotion recognition using two-level classification. In: Proceedings of the 2014 Iranian Conference on Intelligent Systems (ICIS 2014).
7. Chiang WC, Wang JS, Hsu YL. A music emotion recognition algorithm with hierarchical SVM-based classifiers. In: Proceedings of the 2014 International Symposium on Computer, Consumer and Control.

8. Lin C, Liu M, Hsiung W, Jhang J. Music emotion recognition is based on two-level support vector classification. In: Proceedings of the 2016 International Conference on Machine Learning and Cybernetics (ICMLC 2016).

9. Fukayama S, Goto M. Music emotion recognition with adaptive aggregation of Gaussian process regressors. In: Proceedings of the 2016 IEEE international conference on acoustics, speech and signal processing (ICASSP 2016).

10. Nalini NJ, Palanivel S. Music emotion recognition: The combined evidence of MFCC and residual phase. Egyptian Informatics Journal. 2016; 17(1): 1-10. doi: 10.1016/j.eij.2015.05.004

11. Liu H, Fang Y, Huang Q. Music emotion recognition using a variant of recurrent neural network. In: Proceedings of the 2018 International Conference on Mathematics, Modeling, Simulation and Statistics Application (MMSSA 2018).

12. Er MB, Aydilek IB. Music emotion recognition is done using a chroma spectrogram and deep visual features. International Journal of Computational Intelligence Systems. 2019; 12(2): 1622-1634.

13. Rajesh S, Nalini NJ. Musical instrument emotion recognition using deep recurrent neural network. Procedia Computer Science. 2020; 167: 16-25.

14. Yang J. A Novel Music Emotion Recognition Model Using Neural Network Technology. Frontiers in Psychology. 2021; 12. doi: 10.3389/fpsyg.2021.760060

15. Hizlisoy S, Yildirim S, Tufekci Z. Music emotion recognition using convolutional long short term memory deep neural networks. Engineering Science and Technology, an International Journal. 2021; 24(3): 760-767. doi: 10.1016/j.jestch.2020.10.009

16. Niu N. Music Emotion Recognition Model Using Gated Recurrent Unit Networks and Multi-Feature Extraction. Mobile Information Systems; 2022.

17. Rahmany I, Dhahri N, Moulahi T, et al. Optimized Stacked Auto-Encoder for Unnecessary Data Reduction in Cloud of Things. In: 2022 International Wireless Communications and Mobile Computing (IWCMC). Published online 30 May 2022. doi: 10.1109/iwcmc55113.2022.9825372

18. Available online: https://paperswithcode.com/dataset/muse (accessed on 2 January 2024).

19. Shaik AA, Mareedu VDP, Polurie VVK. Learning multiview deep features from skeletal sign language videos for recognition. Turkish Journal of Electrical Engineering & Computer Sciences. 2021; 29(2): 1061-1076. doi: 10.3906/elk-2005-57

20. Ghuge A, Prakash VC, Ruikar SD. Systematic analysis and review of video object retrieval techniques. Control and Cybernetics. 2020; 49(4): 471–498.

21. Ghuge CA, Chandra Prakash V, Ruikar SD. Weighed query-specific distance and hybrid NARX neural network for video object retrieval. The Computer Journal. 2019; 63(11): 1738-1755. doi: 10.1093/comjnl/bxz113

22. Victoria DrAH, Manikanthan SV, H R DrV, et al. Radar Based Activity Recognition using CNN-LSTM Network Architecture. International Journal of Communication Networks and Information Security (IJCNIS). 2023; 14(3): 303-312. doi: 10.17762/ijcnis.v14i3.5630

23. Thota MK, Shajin FH, Rajesh P. Survey on software defect prediction techniques. International Journal of Applied Science and Engineering. 2020; 17(4): 331–344.

24. Appathurai A, Sundarasekar R, Raja C, et al. An Efficient Optimal Neural Network-Based Moving Vehicle Detection in Traffic Video Surveillance System. Circuits, Systems, and Signal Processing. 2019; 39(2): 734-756. doi: 10.1007/s00034-019-01224-9

25. Balamurugan D, Aravinth SS, Reddy PCS, et al. Multiview Objects Recognition Using Deep Learning-Based Wrap-CNN with Voting Scheme. Neural Processing Letters. 2022; 54(3): 1495-1521. doi: 10.1007/s11063-021-10679-4

26. Bhavana D, Kishore Kumar K, Bipin Chandra M, et al. Hand Sign Recognition using CNN. International Journal of Performability Engineering. 2021; 17(3): 314. doi: 10.23940/ijpe.21.03.p7.314321

27. Kumar EK, Kishore PVV, Kiran Kumar MT, et al. 3D sign language recognition with joint distance and angular coded color topographical descriptor on a 2 – stream CNN. Neurocomputing. 2020; 372: 40-54. doi: 10.1016/j.neucom.2019.09.059

28. Ghuge C, Prakash V, Ruikar S. An Integrated Approach Using Optimized Naive Bayes Classifier and Optical Flow Orientation for Video Object Retrieval. International Journal of Intelligent Engineering and Systems. 2021; 14(3): 210-221. doi: 10.22266/ijies2021.0630.19

29. Gullapelly A, Dr. BGB. Exploring the techniques for object detection, classification, and tracking in video surveillance for crowd analysis. Indian Journal of Computer Science and Engineering. 2020; 11(4): 321-326. doi: 10.21817/indjcse/2020/v11i4/201104064

30. Saha J, Chowdhury C, Ghosh D, et al. A detailed human activity transition recognition framework for grossly labeled data from smartphone accelerometer. Multimedia Tools and Applications. 2020; 80(7): 9895-9916. doi: 10.1007/s11042-020-10046-w

31. Yadav J, Misra M, Rana NP, et al. Exploring the synergy between nano-influencers and sports community: behavior mapping through machine learning. Information Technology & People. 2021; 35(7): 1829-1854. doi: 10.1108/itp-03-2021-0219

32. Yadav J, Misra M, Rana NP, et al. Netizens' behavior towards a blockchain-based esports framework: a TPB and machine learning integrated approach. International Journal of Sports Marketing and Sponsorship. 2021; 23(4): 665-683. doi: 10.1108/ijsms-06-2021-0130

33. Mohan KK, Prasad CR, Kishore PVV. Yolo V2 with bifold skip: A deep learning model for video-based real-time train bogie part identification and defect detection. Journal of Engineering Science and Technology. 2021; 16(3): 2166–2190.

34. Krishnamohan K, Prasad ChR, Kishore PVV. Train rolling stock video segmentation and classification for bogie part inspection automation: a deep learning approach. Journal of Engineering and Applied Science. 2022; 69(1). doi: 10.1186/s44147-022-00128-x

35. Raju K, Sampath Dakshina Murthy A, Chinna Rao B, et al. A robust and accurate video watermarking system based on SVD hybridation for performance assessment. International Journal of Engineering Trends and Technology. 2020; 68(7): 19–24.

36. Suneetha M, Prasad MVD, Kishore PVV. Sharable and unshareable within class multi view deep metric latent feature learning for video-based sign language recognition. Multimedia Tools and Applications. 2022; 81(19): 27247-27273. doi: 10.1007/s11042-022-12646-0

37. Suneetha M, Prasad MVD, Kishore PVV. Multi-view motion modelled deep attention networks (M2DA-Net) for video-based sign language recognition. Journal of Visual Communication and Image Representation. 2021; 78: 103161. doi: 10.1016/j.jvcir.2021.103161

38. Wagdarikar AMU, Senapati RK. A secure communication approach in OFDM using optimized interesting region-based video watermarking. International Journal of Pervasive Computing and Communications. 2020; 18(2): 171-194. doi: 10.1108/ijpcc-05-2019-0044

39. Janarthanan P, Murugesh V, Sivakumar N, et al. An Efficient Face Detection and Recognition System Using RVJA and SCNN. Mathematical Problems in Engineering. 2022; 2022: 1-9. doi: 10.1155/2022/7117090

40. Priyadharshini B, Gomathi T. Naive Bayes classifier for wireless capsule endoscopy video to detect bleeding frames. International Journal of Scientific and Technology Research. 2020; 9(1): 3286–3291.

41. Ali SA, Prasad MVD, Kishore PVV. Ranked Multi-View Skeletal Video-Based Sign Language Recognition with Triplet Loss Embeddings. Journal of Engineering Science and Technology. 2022; 17(6): 4367–4397.

42. Pande SD, Chetty MSR. Linear Bezier Curve Geometrical Feature Descriptor for Image Recognition. Recent Advances in Computer Science and Communications. 2020; 13(5): 930-941. doi: 10.2174/2213275912666190617155154

43. Depuru S, Nandam A, Ramesh PA, et al. Human Emotion Recognition System Using Deep Learning Technique. Journal of Pharmaceutical Negative Results. 2022; 13(4): 1031–1035.

44. Ali SKA, Prasad MVD, Kumar PP, et al. Deep Multi View Spatio Temporal Spectral Feature Embedding on Skeletal Sign Language Videos for Recognition. International Journal of Advanced Computer Science and Applications. 2022; 13(4). doi: 10.14569/ijacsa.2022.0130494

45. Rani S, Ghai D, Kumar S. Reconstruction of Simple and Complex Three Dimensional Images Using Pattern Recognition Algorithm. Journal of Information Technology Management. 2022; 14: 235–247.

46. Rani S, Ghai D, Kumar S, et al. Efficient 3D AlexNet Architecture for Object Recognition Using Syntactic Patterns from Medical Images. Computational Intelligence and Neuroscience. 2022; 2022: 1-19. doi: 10.1155/2022/7882924

47. Kotkar VA, Sucharita V. Scalable Anomaly Detection Framework in Video Surveillance Using Keyframe Extraction and Machine Learning Algorithms. Journal of Advanced Research in Dynamical and Control Systems. 2020; 12(7): 395-408. doi: 10.5373/jardcs/v12i7/20202020

48. Li X, Manivannan P, Anand M. Task Modelling of Sports Event for Personalized Video Streaming Data in Augmentative and Alternative Communication. Journal of Interconnection Networks. 2022; 22(Supp01). doi: 10.1142/s0219265921410279