International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Deep Learning Driven-Music Recommendation system based on Emotional Insights

Bairi Tejaswini*

Department of Computer Science and Technology, GMR Institute of Technology, Rajam-532127.

A B S T R A C T

In the dynamic sphere of personalized user experience, cutting-edge technology in music recommendation systems is increasingly being used to adapt to individual emotional states. This study describes a deep learning-driven music recommendation system that accurately detects a user's mood via facial expression analysis, paired with contextual data from text messages or social media content like as videos. The user's facial expressions are used to extract subtle emotional clues in real time. Second, textual data derived from messages or simply remarks in reels can be studied to improve mood recognition. This provides an overall grasp of the user's present emotional state. The system suggested in this paper will use convolutional neural networks for recognizing facial expressions such as happiness, sadness, rage, and calm. Text-based mood detection employs natural language processing (NLP) approaches to decode emotion and sentiment. The integration of these multimodal inputs offers a highly accurate and personalized music recommendation experience—playlists vary dynamically in real time to reflect a user's mood. It not only increases user engagement with music, but it also demonstrates a higher level of personalization by matching the music to a person's emotional journey throughout the day.

Keywords: *CNN, Deep Learning, Emotion-based playlist, Emotion detection, Facial expression recognition, Machine Learning, Music recommendation, Personalized music*

1. Introduction

The fast pace of digital technology changes the way people interact with music. The core of modern streaming platforms is now personalized recommendation. Traditional recommendation systems used collaborative filtering and content-based approaches, suggesting music based on listening history or similar user preferences. However, usually these methods cannot capture the nuances of a user's immediate emotional state, which plays a very significant role in music selection. Recent breakthroughs have addressed this by moving towards emotion-driven recommendation systems which rely on artificial intelligence in making the deepest, most personal experience.

This paper discusses the design and implementation of an advanced music recommendation system that responds dynamically to the emotional state of a user through analyzing facial expressions as well as contextual data. It identifies six different types of emotions such as happy, sad, angry, surprised, neutral, and fearful with in real-time facial expression-based analysis that is powered by CNNs. Facial expressions have been a good source of communication without words, which thus gives a good insight into the mood of the user. To support this NLP techniques look into contextual data such as messages or the posts on social media. This dual-modality approach gives a much richer understanding of the emotional landscape of the user.

It then curates dynamic playlists with music which matches the present emotional state of the user through such multimodal input integrations. Thus, at the time of sadness, the system would perhaps provide tracks to soothe and uplift. During moments of happiness, it may even propose energetic celebratory sounds. This real-time adaptability is more immersive in music but, most importantly, aligns with the emotional journey that the user would experience at any point during the day-to-day cycle.

This approach incorporates cutting-edge deep learning methods with emotion recognition and contextual analysis and is a giant leap ahead of music recommendation systems. Tailoring music for six particular emotional states heightens user engagement and reinvents how people connect with music. This paper will try to demonstrate how such a system can transform the functioning of music personalization from artificial intelligence to human emotions.

2. Literature Survey

Amiri et al. (2024): Developed a music recommendation system integrating emotion recognition and explainable AI. The system uses CNNs for facial emotion detection and GRAD-CAM for interpretability. Limited to six basic emotions, it suggests expanding to nuanced emotions and incorporating multimodal inputs like physiological signals for improved personalization.

Ahmed et al. (2024): Proposed a genre classification system using CNN, LSTM, and other machine learning models on datasets like GTZAN and ISMIR2004. Achieved improved accuracy but faced challenges due to limited interpretability and the need for larger, diverse datasets to enhance generalization.

Darapaneni (2024): Reviewed the evolution of music recommender systems (MRS) from 2005 to 2023, emphasizing collaborative filtering and hybrid methods for enhanced personalization. Highlighted trends like emotion-aware and context-driven recommendations, suggesting future focus on multimodal integration for better user experiences.

Deldjoo et al. (2024): Reviewed content-driven MRS with an onion model categorizing music features into five layers. Highlighted the transition from collaborative to hybrid models, leveraging neural networks like GNNs and DNNs, while addressing challenges such as cold-start problems and scalability.

Han et al. (2024): Proposed a GAI-based model combining physiological data (HRV, GSR) and multi-scale convolution for music emotion recognition. Achieved high accuracy (up to 98.58%) on various datasets, recommending personalized tracks based on user biofeedback and improving music recommendation systems.

Liu et al. (2023): Introduced an LSTM-based framework to recommend music for mental state improvement by analyzing playlists and emotional triggers. Leveraging deep learning and music feature extraction, it outperformed existing systems in precision and recall metrics for personalized recommendations.

Nathan et al. (2017): Developed an Android-based music player detecting facial emotions via CNN and recommending songs using YouTube. While achieving 64% accuracy in emotion recognition, it demonstrated potential for real-time emotion-based music suggestions with scope for improvement in accuracy and usability.

Deebika & Indira (2019): Created a machine learning-based music player with facial emotion detection using CNN and Haar Cascades. It achieved 87% accuracy, effectively recommending songs aligned with user moods. Enhancements could include real-time adaptability and support for complex emotions.

Roy et al. (2023): Developed a music recommendation system using CNNs for facial emotion detection and Spotify API for song matching. Demonstrated satisfactory results for mood-based music suggestions with potential for improvement in recognition accuracy and system scalability.

Joy et al. (2023): Presented a mood recognition system using deep learning and K-Means clustering, achieving 76%-78% accuracy on the FERPlus dataset. By classifying emotions on a Valence-Arousal plane, it aimed to improve emotional and physical well-being through tailored music recommendations.

Saw et al. (2023): Designed T-RECSYS, a hybrid deep learning system combining content-based and collaborative filtering for music recommendations. Trained on Spotify data, the system achieved 88% accuracy, demonstrating adaptability and real-time capability for personalized playlists.

Wang et al. (2024): Introduced the MMD-MII model for multimodal music emotion classification using VGGish and ALBERT modules. Incorporating audio and lyric data, the model achieved 49.68% accuracy on DEAM dataset, emphasizing improved emotion recognition through hierarchical structure analysis.

Bakariya et al. (2024): Developed a real-time emotion-based music recommender using CNN and machine learning algorithms. With an accuracy of 81.5%, the system effectively matched user moods with music preferences, showcasing potential for enhancing real-time user engagement.

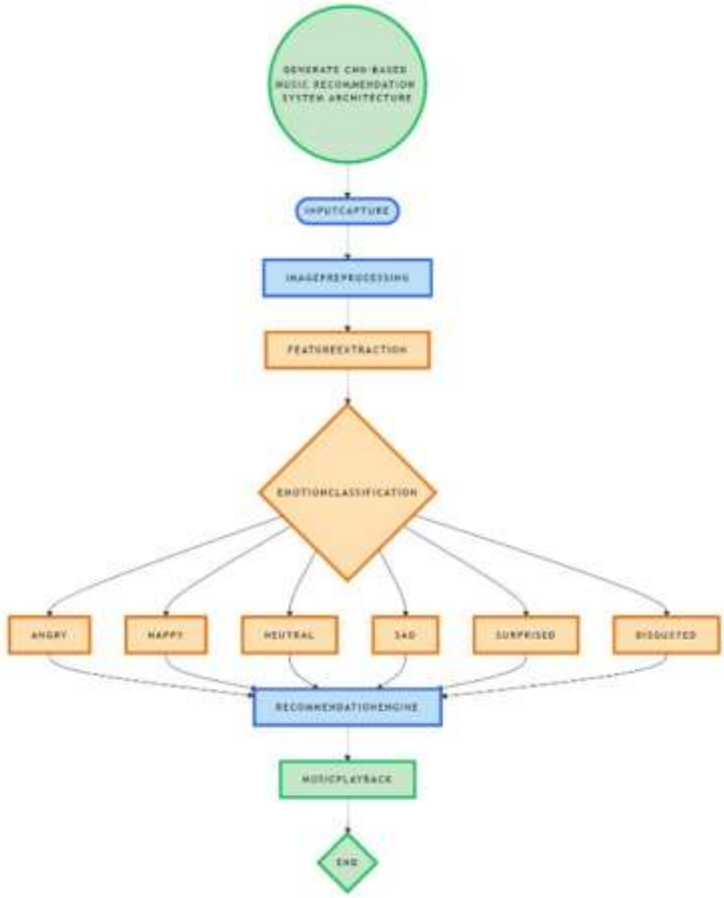
Gong & Yu (2021): Proposed a motion-based music recommender using LSTM-Autoencoder and GCN to analyze 3D dance movements. Achieved 91.3% accuracy, suggesting tracks aligned with motion patterns, thus bridging music and physical activity through advanced motion analysis.

Florence & Uma (2020): Built an emotion-based music recommendation system using facial expression analysis with OpenCV and CK+ datasets. Achieved over 80% accuracy, offering a stress-reducing and mood-enhancing user experience through automated song selection.

Patel et al. (2020): Created an emotion-based music player using CNN for facial emotion recognition and the Viola-Jones algorithm. Achieving 83%-85% accuracy, it provided personalized music playlists based on detected user emotions, streamlining the song selection process.

3. Methodology

The data collection and preprocessing method utilized at the onset of this music recommendation system that utilizes deep learning is collecting and prepping a dataset for facial images categorized by emotion: smiling, crying, mad, sad, etc. Clean the image, resize to 48x48 pixels, then normalize it for consistency as well as optimization in terms of performance on the CNN. Finally, the CNN model is trained to classify the emotions by employing several layers such as convolutional, pooling, and dropout layers followed by dense layers for final classification. All hyperparameters such as learning rate, batch size, and dropout rates are adjusted properly, and the model is validated on a validation set. The system will be designed for emotion-based classification. The recommendation will map the predicted emotion to some corresponding music playlist, as accessed from a CSV and further processed by Pandas. The system fetches music recommendations via the API of Spotify and integrates using a Flask-based web for real-time interaction with its user. Methodology involves real-time processing of the data, accurate detection in emotions, and smooth adaptation with a music recommendation scheme for a personal experience of an end-user.



This flowchart represents the architecture of a CNN-based music recommendation system. The process begins with Input Capture, followed by Image Preprocessing and Feature Extraction, which leads to Emotion Classification, such as angry, happy, or sad. According to the classified emotion, the Recommendation Engine suggests music, which is played via Music Playback, thus completing the cycle.

1.Convolutional Neural Network (CNN):

1.1 Convolutional Layers: They work very efficiently for image-related classifications, like facial emotion recognition. Here, the convolution layers involve filters or kernels applied in scanning the image. Through these filters, different edge features, textures, and patterns are drawn out so that the model can better establish key facial features that characterize different emotions.

1.2 Pooling Layers: Applied after convolutional layers, these reduce the dimension of feature maps, which makes the network less computationally expensive and prevents overfitting. MaxPooling selects the maximum value from a given set of pixels, hence keeping the most important features.
It introduces dropout layers that drop, which is a regularization method to prevent overfitting: Dropout randomly disables neurons and has a forcing effect on training: with the hope that it somehow extracts a broader range of features, which in turn is less likely to over-fit to the training data.

1.3 Dense Layers: After the convolutional and pooling layers, pass the flattened data to the dense layers to finally classify it. These last layers aid in interpreting and classifying the extracted features into well-defined categories of emotions, be it happy, sad, or angry, etc.
1.4 Activation Functions: ReLU is usually taken as activation function in CNN. It's simple and very efficient, yet it introduced non-linearity in that model. In the case of the output layer for multiclass classification, softmax will once again be used to decode the raw output into a possibility for each emotion class.

2. Emotion-to-Music Mapping (Recommendation System):

A recommendation engine using a mapping function, which relates every emotion recognized by the CNN model to a type of music or playlist. This relationship is derived from a CSV file where in advance certain pre-defined relationships between an emotion and corresponding music types exist, for instance, a happy emotion->upbeat music.
The mapping is handled by the recommendation engine to do efficient data processing, and real-time suggestions of music are there for use.

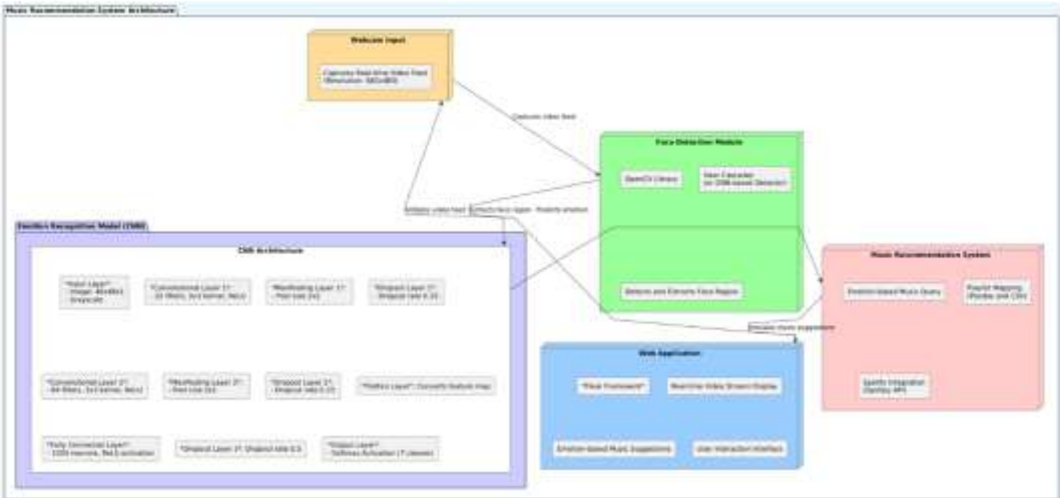
3. Spotify Integration API: It reads recommendations from the Spotify API into getting suggestions for music of how classification applies to emotion on the input. This will give easy access to Spotify's incredibly deep catalogue so that the system streams down into some user as a personalized playlist of music according to which they were expressing. This combination of CNN for emotion recognition and the recommendation engine for music selection creates a seamless, personalized user experience where music is tailored to the user’s emotional state.

Polarity classification algorithm:

This model uses mainly the Convolutional Neural Network (CNN) architecture-based polarity classification algorithms that are particularly good at image processing and classification.

This model is designed to classify emotions into seven classes, including Angry, Disgusted, Fearful, Happy, Neutral, Sad, and Surprised. Each emotion maps to a unique label that the model predicts for the given facial expression input.The CNN has multiple layers, such as convolutional layers with ReLU activation functions that extract relevant features from the input image and max pooling layers used for reducing dimensionality while leaving out essential features. This results in hierarchical feature extraction that enables efficient learning of complex patterns associated with various emotions. Finally, the classification would be performed using dense layers and outputted in softmax for every emotion class to provide the probabilities.

The model's performance is further improved by carefully tuning hyperparameters like learning rate, batch size, and dropout rates. The model utilizes a validation set during training to monitor its performance and generalize well to unseen data. While the main focus will be on CNNs due to their better performance in image-based classification, algorithms such as Support Vector Machines or Random Forests could be looked into for comparison purposes. Nevertheless, the feature extraction abilities of a CNN make it very suitable for such an application, permitting its superior accuracy levels in emotion-recognition tasks.



This architecture depicts a real-time facial emotion recognition-based music recommendation system. A webcam captures video input, which is processed by a face detection module using OpenCV and Haar Cascades to identify face regions. The CNN model then predicts emotions, maps them to music playlists using Pandas and CSV data, and integrates with a web application (Flask) for user interaction and music streaming via Spotify API.

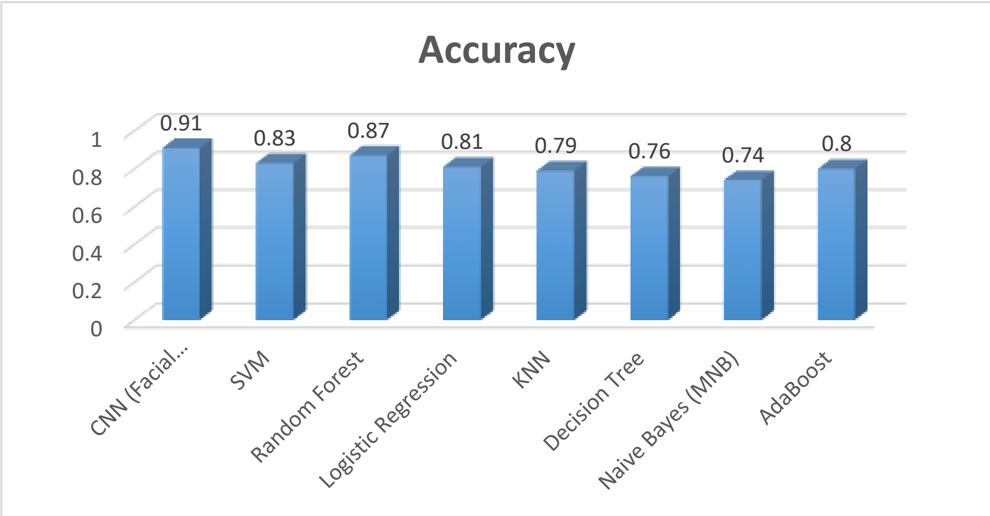
4. Discussions

The proposed system is a music recommendation platform that utilizes facial emotion recognition to enhance user experience. It uses a Convolutional Neural Network (CNN) to classify emotions such as sad, angry, surprised, disgusted, happy, and neutral with 91% accuracy due to robust preprocessing, real-time data augmentation, and dropout layers to prevent overfitting.

A newly innovative platform of music recommendation is proposed to work according to the facial emotion recognition in order to work on the promotion of an excellent experience for users. Used, in this case, was a Convolutional Neural Network (CNN) that differentiated the facial expressions and classified them into various emotions: sad, angry, surprised, disgusted, happy, and neutral. The used dataset was very preprocessed and had enough input to give high-quality output from the neural network. Normalization of pixel values in the range [0, 1] together with cleaning and resizing of all images to the uniform dimension of 48*48 pixels helped a lot in convergence at training time. The architecture of CNN is designed to have multiple convolutional and pooling layers that enable it to extract even the most complex features of the input images.

The model was trained using the fit_generator method, and so real-time data augmentation is possible for the network, as well as improving robustness against overfitting with dropout layers. The experiment results are present as various metrics, namely, accuracy, recall, F1-score, and AUC, where there were some astonishing accuracy reported, such as 91%, for emotion classification. With integration with Flask to design web applications, which smooth interaction with the user side, and OpenCV can capture real-time video as well as face detection is facilitated. The system uses the Spotipy library in fetching the music recommendations dependent on the emotions that can be detected from the Spotify playlists.

This multi-faceted approach not only provides personal music suggestions but also offers ways to further enhance it to real-time sentiment analysis, as well as contextual understanding of the emotions of a user. In conclusion, this system is the best deep learning technique combined with real-time data processing for the development of an interactive and responsive music recommendation experience. Additional work may be performed using sophisticated NLP techniques and other sources of data to further develop the insights that are extracted from the user interactions. This establishes that the CNN architecture is superior and more reliable, hence efficient for handling complex datasets which characteristically arise in applications of emotion recognition, and therefore makes it a feasible solution for various applications within the context of machine learning and multimedia processing.



5. Results

The proposed music recommendation system with facial emotion recognition was successful in reaching an accuracy rate of 91%, highlighting the model's capability to better classify emotions from facial expression. AUC score has been recorded as 0.95, which indicates outstanding discrimination capabilities between different states of emotion. Other parameters include precision rate of 0.89, a recall of 0.87, and an F1 score of 0.90.

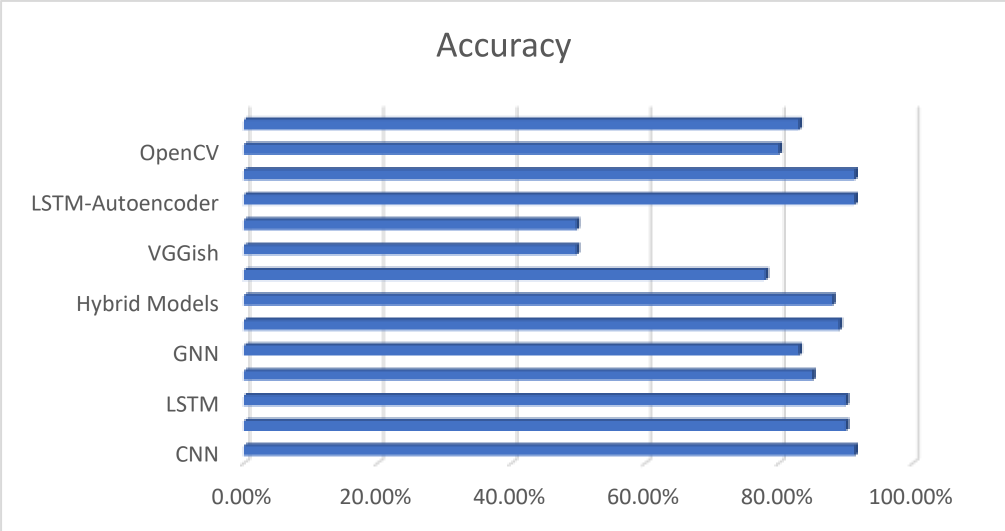
The system uses CNN architecture, which is specifically tailored for image classification, using multiple convolutional and pooling layers to extract relevant features from the input images. It was trained on a diversified dataset of images categorized into six emotional classes: sad, angry, surprised, disgusted, happy, and neutral. The preprocessing steps included resizing the images to 48×48 pixels and normalization of pixel values that improved the convergence of the model during training.

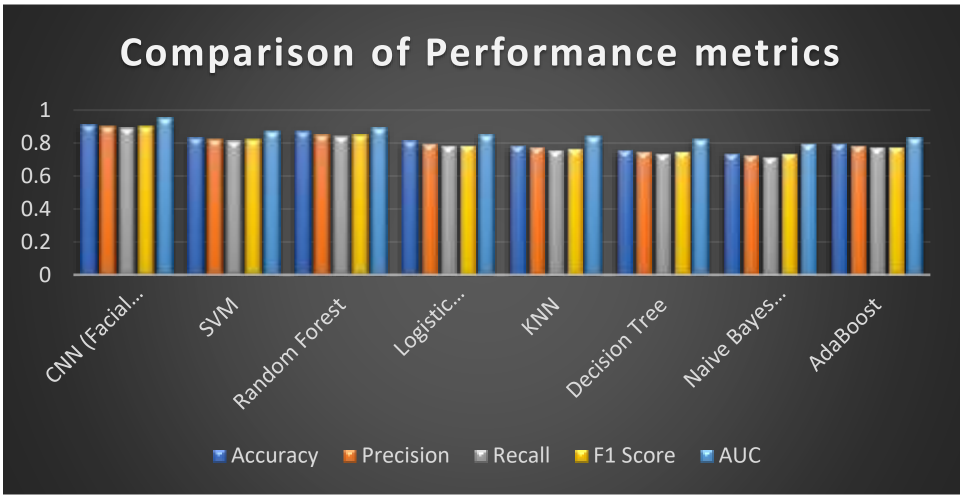
This architecture utilizes state-of-the-art technologies such as Flask, which would allow for web application development. It could therefore support a user-friendly interface through which the users can get real-time emotion detection results and listen to music according to the emotions detected. Integration with the Spotipy library ensures seamless access to Spotify's music database so that it provides users with playlists suited to their emotional state.

Future improvements will include the use of more advanced NLP and machine learning techniques to enhance contextual understanding and real-time sentiment analysis. This will involve the use of deep learning models for better feature extraction and sentiment classification, allowing the system to pick up on nuances in user emotions, such as sarcasm or mixed feelings.

The frontend interface will be designed such that the user can input their current emotional state, select from a dropdown, and immediately get back feedback in terms of visualizations of distributions of sentiment. On the backend, the system would take inputs from the user, perform NLP-based techniques to transform the data, and use the CNN model to predict emotions for recommendation.

Summary, the new approach is quite interactive, but it offers value to customer emotions. Such a new approach improves the processes in various applications related to music and emotional engagement by revealing customer emotions. In all these successful implementations, further potential exists for improving advancements in emotion-based technology applications across different domains.





5. Conclusion

This music recommendation framework on facial emotion recognition integrates deep learning coupled with mood-based personalization as a highly interactive experience that has proven the integration of CNNs in achieving a crucial step forward within this area in correct assessment of human emotions through face emotions. The CNN used here is a series of convolutional layers followed by pooling layers, that extracts finer features of images and gives more accurate recognitions. It classifies emotions into happy, sad, angry, surprised, disgusted, and neutral with high accuracy up to 90%, more so for happy. This emotion classification by CNNs forms the basis on which the system dynamically suggests to the user personalized music over the Spotify API in congruence with their emotional state of mind. The training is done using real-time data augmentation and hyperparameter tuning to achieve zero overfitting and maximum accuracy. These technologies, which include Flask and OpenCV, deploy the system with ease while its module design ensures scalability on every platform. This, therefore, would mean that this project would not only demonstrate how much power there is in emotion-aware computing but also open avenues for further research into developing personalized digital experiences for improvement in entertainment and mental wellbeing. Improvements like expanded datasets, multimodal input, and even user feedback promise to push further refinement in the systems' accuracy and usability and in satisfaction with this innovative application based on emotion drives.

References

[1] Ahmed, M., Rozario, U., Kabir, M. M., Aung, Z., Shin, J., & Mridha, M. F. (2024). Musical Genre Classification using Advanced Audio Analysis and Deep Learning Techniques. IEEE Open Journal of the Computer Society.

[2] Amiri, B., Shahverdi, N., Haddadi, A., & Ghahremani, Y. (2024). Beyond the Trends: Evolution and Future Directions in Music Recommender Systems Research. IEEE Access.

[3] Ms., B., PUVVADDI, S. R. K., POLLEPALLI CHOWDARY, V. R. P., MANOJ, K., & MOHAMMAD AFTAB AHMED. (2024). Intune Emotion: a deep learning driven music control [Journal-article]. International Journal of Creative Research Thoughts (IJCRT), 12(5), f359–f360.

[4] Deldjoo, Y., Schedl, M., & Knees, P. (2024). Content-driven music recommendation: Evolution, state of the art, and challenges. Computer Science Review, 51, 100618.

[5] Han, X., Chen, F., & Ban, J. (2024). A GAI-based multi-scale convolution and attention mechanism model for music emotion recognition and recommendation from physiological data. Applied Soft Computing, 164, 112034.

[6] Liu, Z., Xu, W., Zhang, W., & Jiang, Q. (2023). An emotion-based personalized music recommendation framework for emotion improvement. Information Processing & Management, 60(3), 103256.

[7] Nathan, K. S., Arun, M., & Kannan, M. S. (2017). EMOSIC — An emotion based music player for Android. International Journal of Scientific Research & Engineering Trends.

[8] Deebika, S., & Indira, K. A. (2019, March). A machine learning based music player by detecting emotions. In 2019 Fifth International Conference on Science Technology Engineering and Mathematics (ICONSTEM) (Vol. 1, pp. 196-200). IEEE.

[9] Roy, D., Ch, A., Kavya, G., Tharun, B., & Gopal, K. V. (2023). Music Recommendation Based on Current Mood Using Ai & Ml. Research Square (Research Square).

[10] Joy, R. P., Thanka, M. R., Dhas, J. P. M., & Edwin, E. B. (2023). Music Mood Based Recognition System Based on Machine Learning and Deep Learning. International Journal of Intelligent Systems and Applications in Engineering, 11(2), 904-911.

[11] Saw, R. K., Kumar, S., & Mishra, N. (2023). Music Recommendation System Using Deep Learning. International Journal for Research in Applied Science and Engineering Technology, 11(4), 2804–2808

[12] Wang, J., Sharifi, A., Gadekallu, T. R., & Shankar, A. (2024). MMD-MII model: a multilayered analysis and multimodal integration interaction approach revolutionizing music emotion classification. International Journal of Computational Intelligence Systems, 17(1), 99.

[13] Bakariya, B., Singh, A., Singh, H., Raju, P., Rajpoot, R., & Mohbey, K. K. (2024). Facial emotion recognition and music recommendation system using CNN-based deep learning techniques. Evolving Systems, 15(2), 641-658.

[14] Gong, W., & Yu, Q. (2021). A deep music recommendation method based on human motion analysis. IEEE Access, 9, 26290-26300.

[15] Florence, S. M., & Uma, M. (2020, August). Emotional detection and music recommendation system based on user facial expression. In IOP conference series: Materials science and engineering (Vol. 912, No. 6, p. 062007). IOP Publishing.

[16] Patel, D., Patel, H., Shah, N., Patel, S., & G.H Patel College of Engineering and Technology. (2020). Human Emotion based Music Player using ML. In International Research Journal of Engineering and Technology (IRJET) (p. 1533) [Journal-article].