International Journal of Autonomous Robotics and Intelligent Systems
Vol. 1, No. 1, June 2025 [13-24]

Implementation of Music Emotion Classification using Deep Learning

Qing Xiang Sow[1*] and Swee Kheng Eng[1,2]

[1]Faculty of Electrical Engineering & Technology, Universiti Malaysia Perlis, Pauh Putra Campus 02600 Arau, Perlis, Malaysia
[2]Centre of Excellence for Intelligent Robotics and Autonomous System (CIRAS), Universiti Malaysia Perlis, 02600 Arau, Perlis, Malaysia

Received 24 June 2025, Revised 30 June 2025, Accepted 12 July 2025

ABSTRACT

Music plays a crucial role in shaping emotions and experiences, making its classification an important area of research with applications in therapy, recommendation systems, and affective computing. This study develops a deep learning-based system to classify music into three emotional categories: "Angry," "Happy," and "Sad." The dataset, consisting of 22 audio files collected from YouTube, was manually labelled, segmented into 30-second clips, and augmented using pitch shifting and time stretching to enhance diversity. Features were extracted using Mel-Frequency Cepstral Coefficients (MFCC) and spectral contrast to analyse the harmonic and timbral characteristics of the audio. Three deep learning models, CNN, CNN-LSTM, and CNN-GRU, were evaluated. CNN-GRU achieved the highest weighted accuracy of 99.10%, demonstrating superior performance. Future work includes adding more emotion categories, diversifying the dataset, exploring advanced architectures like transformers, optimising hyperparameters, implementing real-time applications, and conducting user studies to assess effectiveness. This research successfully developed and evaluated a music emotion classification system, contributing to advancements in the field.

Keywords: CNN, CNN-LSTM, CNN-GRU, Deep Learning, MFCC Extraction, Music Emotion Classification, Spectral Contrast

1. INTRODUCTION

Music is one of the fundamental pillars of human culture and life, shaping emotions and enhancing experiences. It has long been recognised for its ability to reflect and influence human emotions, serving as a source of entertainment, memory building, and emotional expression. The introduction of technology in music analysis has transformed it into a blend of art and science.

The rise of digital music and streaming platforms has redefined how people engage with music, bringing about the need to explore its emotional content. While traditional systems categorise music by genre, artist, or beats per minute (BPM), they fail to address its emotional impact [1]. This gap has inspired researchers to investigate music classification methods that focus on emotional dimensions, paving the way for improved listening experiences, music therapy, and personalised recommendations.

The global Artificial Intelligence (AI) in music market is projected to grow significantly, from USD 3.9 billion in 2023 to USD 38.7 billion by 2033, at a compound annual growth rate (CAGR) of 25.8%. Figure 1 shows this projected growth and highlights the increasing role of AI technologies in revolutionising music production, distribution, and consumption. AI technologies have revolutionised music production, distribution, and consumption, enabling applications such as personalised playlists, real-time composition, and immersive experiences through virtual and

*alexsowqxiang@gmail.com

augmented reality (VR/AR) [2]. These advancements have expanded the integration of AI into music therapy and user engagement.

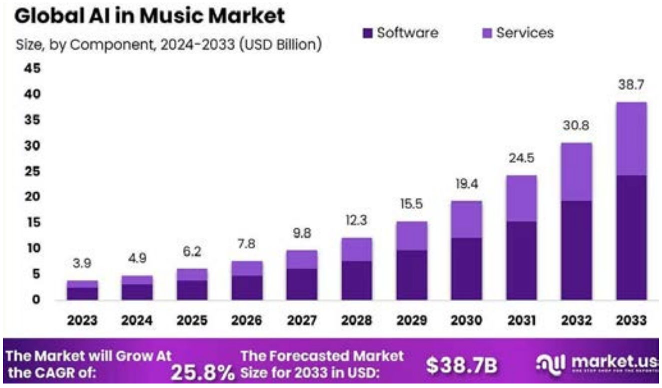


Figure 1. The projected growth of the global AI in music market [2]

Deep learning models have shown immense potential in addressing complex problems like music emotion classification. By learning from large volumes of data, these models can capture irregular patterns and interpret elements such as melodies and harmonies to predict emotional states [3].

This project aims to apply deep learning architectures, including Convolutional Neural Networks (CNN), CNN-Long Short-Term Memory (CNN-LSTM), and CNN-Gated Recurrent Unit (CNN-GRU), to classify music into three emotional categories: "Angry," "Happy," and "Sad." These models will be implemented and evaluated on large datasets to ensure accurate emotion classification. The ultimate goal is to develop a reliable, versatile system that enhances user interaction with music and provides practical applications in therapy, recommendation systems, and entertainment.

2. METHODOLOGY

2.1 Data Collections

The dataset comprises raw audio files sourced from publicly available content on YouTube, featuring music tracks labelled with emotional categories. These tracks were manually curated to correspond to the targeted emotional labels: Angry, Happy, and Sad. The diversity of these tracks ensures a representative dataset for robust training and evaluation of the deep learning models.

While the files were selected to use royalty-free or Creative Commons content, some may be copyrighted. They were used strictly for academic, non-commercial purposes. As the original durations varied, all audio was standardised into 30-second clips for model training. Table 1 shows the source links by emotion.

Table 1 YouTube source links by emotion

No	Angry	Happy	Sad
1	https://youtu.be/m23Cl-E9Qkc	https://youtu.be/rCHMqrkoyFA	https://youtu.be/PLQvm6mbg9U
2	https://youtu.be/EuaCVa_ET7U	https://youtu.be/bn-ERrsr_wk	https://youtu.be/_3noGaPV3UQ
3	https://youtu.be/IPGekm9GZok	https://youtu.be/00H4NcrSK1o	
4	https://youtu.be/1rxpUma1ROE	https://youtu.be/gIyGfaYGOfE	
5	https://youtu.be/TqNQJUMhook	https://youtu.be/PnXFz-01ONo	
6	https://youtu.be/NoFXCaFMgOs	https://youtu.be/sxieXFrKByw	
7	https://youtu.be/XZVygeHwN5M	https://youtu.be/TpMsGTu9kAI	
8	https://youtu.be/mSiWEjZj2aI	https://youtu.be/X11EC849J78	
9	https://youtu.be/_SkUsk07Z-I	https://youtu.be/8TIpMzTQpMg	
10	https://youtu.be/y6LhtZABxJQ	https://youtu.be/WRudRSS7o2E	

2.2 Data Pre-processing

Data pre-processing is a crucial step in preparing raw audio for MFCC extraction, spectral contrast, and model training. It involves several key stages that enhance the quality and consistency of the data while introducing variability to make the dataset more robust.

The first stage is segmentation, which splits audio tracks into smaller, fixed-length segments with overlapping windows [4]. This helps the model better capture temporal patterns and variations in the music. Each segment is labelled with its related emotional category.

Following segmentation, the dataset undergoes augmentation, a process aimed at enhancing the diversity and robustness of the audio data. Two augmentation techniques are applied:

2.2.1 Pitch Shifting

Pitch shifting changes the pitch of an audio track by adjusting its frequency, either higher or lower. It can be done without affecting the track's length, preserving emotional and timing qualities [5]. This adds variation to the dataset, helping models generalise better while keeping the music's original emotional feel.

2.2.2 Time Stretching

Time stretching is a technique that changes the duration of an audio sample without affecting its pitch. It uses algorithms to separate the time and frequency components of the sound. By slowing down or speeding up the audio, it creates timing differences while preserving the original sound quality. This adds variety to the dataset and helps the model learn and adapt to different rhythms and timing patterns [6]. Figure 2 below shows the flowchart of the pre-processing techniques.

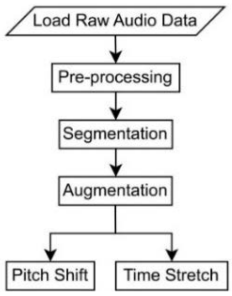


Figure 2. Data pre-processing steps

2.3 MFCC Extraction

Mel-Frequency Cepstral Coefficients (MFCC) extract important audio features by imitating how humans hear. The process begins by changing the audio from time to frequency using the Fourier Transform. Then, the frequency data is mapped to the Mel scale using triangular filters, which focus on sounds the human ear hears best. After that, the logarithm is applied to match how we hear loudness. Finally, the Discrete Cosine Transform (DCT) reduces the data into a few key numbers that capture the sound's tone and quality [7].

2.4 Spectral Contrast

Spectral Contrast measures the difference in energy between peaks (high energy) and valleys (low energy) in a sound spectrum across frequency bands. High contrast indicates clear, narrow-band signals like harmonic tones, while low contrast reflects broadband noise or less defined sounds. This feature helps capture harmonic structure and timbral characteristics, making it useful for identifying musical genres and expressing emotions in audio [8].

2.5 Encode Labels

Label encoding changes category labels into numbers so the model can use them. Each label gets a different number. This method is simple, but it doesn't show any order unless the data has one. It's useful when numbers are needed as input, but should be used carefully with labels that don't have a clear order to avoid confusion [9].

2.6 Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNN) are commonly used in music signal analysis because they can extract important features from spectrograms and mel-spectrograms. CNN has convolutional, pooling, and fully connected layers, and each layer is trained using backpropagation to identify edges, shapes, and textures in the input signal. These patterns are thought to correspond with timbral and structural features which impact emotional perception in music. Early studies have explored the use of CNN for music emotion classification, proving that the model can learn high-level representations directly from audio input without relying on handcrafted features [10]. Figure 3 shows a typical CNN architecture.

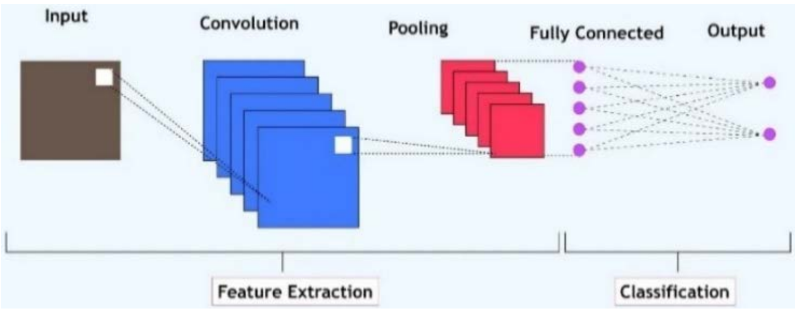


Figure 3. CNN architecture [11]

2.7 CNN-LSTM

It has been observed by researchers that CNN with Long Short-Term Memory (LSTM) works well in combining the spatial and temporal parts of music signals [12]. This architecture uses CNN layers to extract spatial features from spectrogram inputs and capture local pitch, timbre, and rhythm time-frequency patterns. These high-level features are sent to LSTM layers to model music's temporal progression and sequential structure. Furthermore, by incorporating both audio features and lyrical content, the CNN-LSTM model enhances its ability to classify music emotions with greater accuracy and contextual understanding. Figure 4 shows the structure of a CNN-LSTM model.

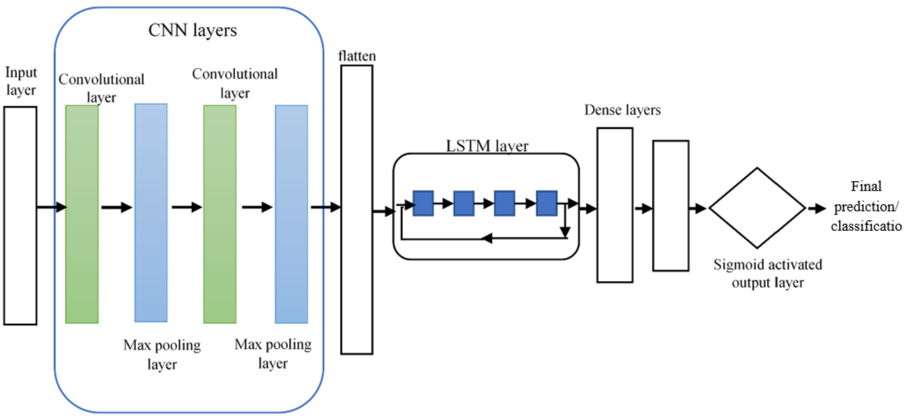


Figure 4. CNN-LSTM architecture [13]

2.8 CNN-GRU

Similar to the CNN-LSTM structure, the CNN-GRU model combines CNN with Gated Recurrent Units (GRU) to capture spatial and temporal features in music signals. In this model, CNN layers extract spatial features from spectrograms, identifying patterns like timbral textures and spectral shapes. These features are then passed to GRU layers, which learn the timing and sequence of changes in the audio. GRU use a simpler structure than LSTM, allowing faster training and lower computational cost while still performing well in sequence tasks. This model has proven effective in classifying music emotions by learning how emotional cues change over time [14]. Figure 5 shows an example of the CNN-GRU architecture.

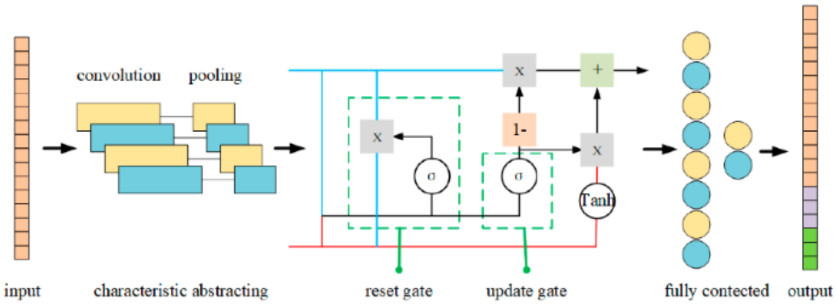


Figure 5. CNN-GRU architecture [15]

2.9 Model Training, Validation and Testing

In developing the music emotion classification system, the dataset is split into 70% for training, 15% for validation, and 15% for testing. This ensures the model learns effectively from training data while using the validation set for fine-tuning and the test set for final performance evaluation.

During training, the goal is to optimise the model to minimise errors and maximise weighted accuracy. The categorical cross-entropy loss function measures the difference between true and predicted labels, making it ideal for multi-class classification. The Adam optimiser adjusts the learning rate dynamically for stable and efficient convergence. Key training parameters include epochs, batch size and learning rate.

2.10 Model Evaluation

The performance of each model is evaluated using the following metrics:

2.10.1 Accuracy

Accuracy is a basic yet widely used metric that measures the proportion of correct predictions made by a classification model. It considers both true positives (TP), where the model correctly identifies a positive case, and true negatives (TN), where it correctly identifies a negative case. Conversely, false positives (FP) occur when a negative case is incorrectly predicted as positive, and false negatives (FN) occur when a positive case is wrongly classified as negative. The formula is presented in Equation (1):

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \tag{1}$$

2.10.2 Weighted Accuracy

Weighted accuracy addresses class imbalance by assigning a weight to each class based on its sample size. It calculates the accuracy for each class and then computes a weighted average. In this context, Total TP refers to the sum of true positives across all classes, and Total Support refers to the total number of actual samples from all classes. The formula is shown in Equation (2):

$$\text{Accuracy}_{\text{Weighted}} = \frac{\text{Total TP}}{\text{Total Support}} \tag{2}$$

2.10.3 Precision

Precision evaluates the correctness of positive predictions. It is the proportion of predicted positive cases out of all predicted positives. High precision indicates fewer false positives, which is important in scenarios where false alarms are costly. This is shown in Equation (3):

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{3}$$

2.10.4 Recall

Recall measures the model's ability to identify all actual positive cases. It is the ratio of correctly predicted positives to all actual positives. A high recall means the model misses fewer relevant results, making it important in applications where missing positive cases is critical. Recall is shown in Equation (4):

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{4}$$

2.10.5 F1-score

The F1-score is the harmonic mean of precision and recall. It balances both metrics and is especially useful when the dataset is imbalanced or when both false positives and false negatives must be minimised. It is calculated using Equation (5):

$$\text{F1-score} = 2 \times \frac{\text{Precision}\times\text{Recall}}{\text{Precision+Recall}} \tag{5}$$

3. RESULTS AND DISCUSSION

This section discusses the results and discussion of three deep learning models: CNN, CNN-LSTM, and CNN-GRU, for music emotion classification.

3.1 Data Pre-processing

The pre-processing phase focused on preparing raw audio data collected from YouTube by enhancing its quality and introducing variability, ensuring it was suitable for MFCC extraction, spectral contrast and effective model training. This process comprised two main steps: segmentation and data augmentation.

3.1.1 Segmentation

The audio files were first converted to mono format with a sampling rate of 16 kHz to standardise the input. Large files over 4GB were split into smaller chunks to avoid system issues. All audio was then divided into uniform 30-second segments, expanding the dataset from 22 original files to 1,726 segments. This ensured consistent input and supported effective model training. Table 2 shows the number of segments for each emotion, "Angry," "Happy," and "Sad", with "Happy" and "Sad" having more segments due to their longer recordings.

Table 2 Segmentation results for each emotion category

Emotion	Initial Files	Segment Files	Total Duration (HH:MM:SS)
Angry	10	237	01:55:21
Happy	10	767	06:21:06
Sad	2	722	06:00:07
Total	22	1726	14:16:35

3.1.2 Augmentation

The augmentation process was used to expand the dataset and improve its diversity, allowing the models to generalise more effectively. By increasing variability in the audio samples, the models could better recognise different emotional patterns.

One of the techniques used was pitch shifting. This method changes the pitch of an audio file by ±2 semitones without affecting its duration. It introduces subtle differences in tonal quality while keeping the emotional tone intact, making the dataset more varied and robust.

Another technique applied was time stretching. This adjusts the speed of the audio playback, slower or faster typically by factors of 0.8x and 1.2x, without changing the pitch. It creates variations in rhythm and timing, helping the model learn from different temporal patterns in the audio.

A summary of the waveform effects for each method is shown in Table 3, which compares the original audio with the pitch-shifted and time-stretched versions.

Table 3 Comparison of original, pitch shifted and time stretched audio waveforms

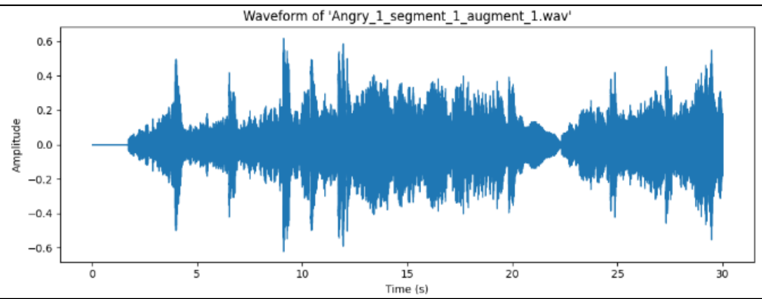
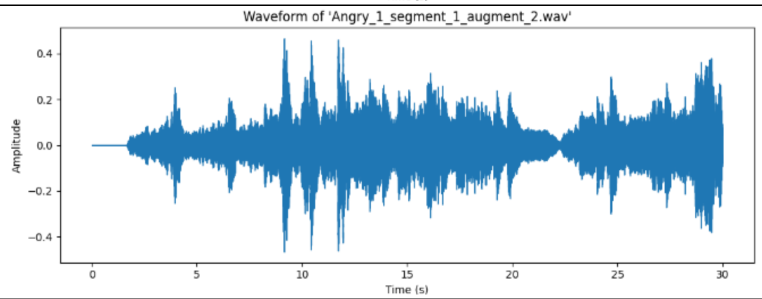
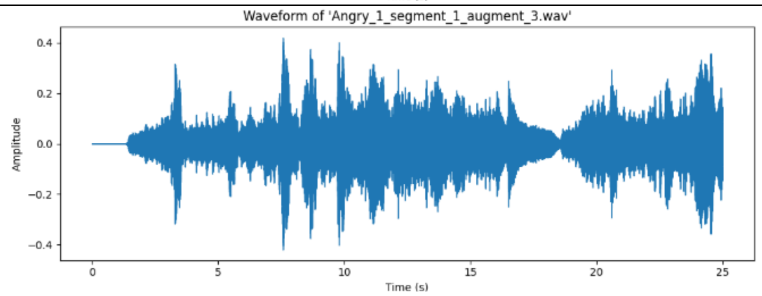
Description	Waveform
Original Audio Segment	
After Pitch Shifting	
After Time Stretching	

Table 4 shows the augmented dataset, detailing the number of files and total duration across emotion categories. This process increased the dataset size to 5,178 audio files, significantly improving the training diversity.

Table 4 Augmentation results for each emotion category

Emotion	Augmented Files	Total Duration (HH:MM:SS)
Angry	711	05:49:43
Happy	2301	19:18:38
Sad	2166	18:16:27
Total	5178	43:24:49

3.1.3 Data Splitting and Class Balancing

The dataset was divided into three subsets: 70% for training, 15% for validation, and 15% for testing. This ensured most data was used for training, with enough samples reserved for validation and testing. The number of samples in each subset was calculated using the following formula:

$$\text{Samples in Subset } = \text{Total Samples } \times \text{Subset Percentage} \tag{6}$$

Table 5 shows the dataset distribution after splitting into training, validation, and testing subsets.

Table 5 Dataset splitting results

Subset	Samples
Training Set	3624
Validation Set	777
Testing Set	777
Total	5178

Class balancing addressed the dataset's natural imbalance, where "Angry" had the fewest samples. To mitigate this, inverse frequency-based class weights were calculated and applied during training. Table 6 shows the data distribution and class weights, with the "Angry" class assigned a weight of 242.76% to increase its influence. This adjustment helps reduce bias towards more frequent classes and improves overall model performance.

Table 6 Class balancing weights

Emotion	Augmented Files	Total Duration (HH:MM:SS)	Class Weight (%)
Angry	711	05:49:43	**242.76**
Happy	2301	19:18:38	75.01
Sad	2166	18:16:27	79.69

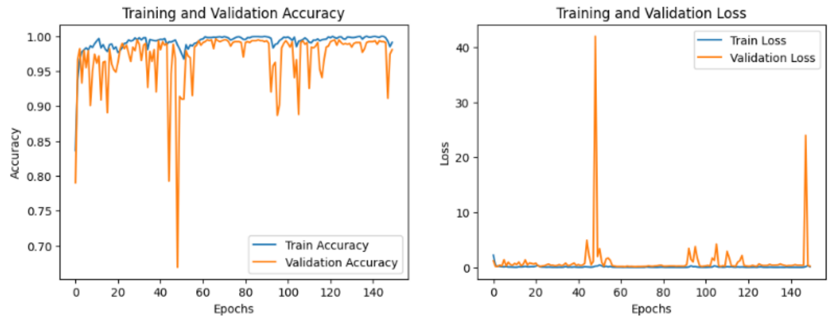
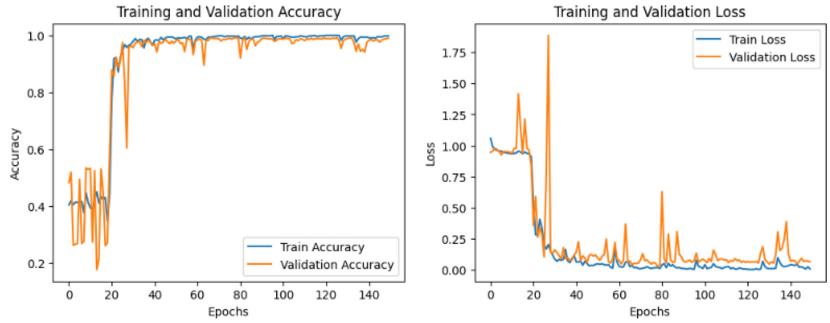
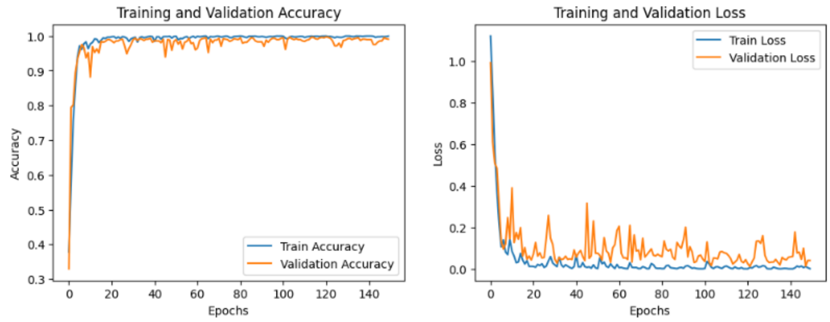
3.2 Classification Result for Each Model

This section shows the results of the three deep learning models, each trained for 150 epochs with a batch size of 64. Their performance is measured using weighted accuracy, precision, recall, and F1-score. Visuals and data summaries, such as training/validation accuracy, loss, and confusion matrices, show the strengths and weaknesses of each model.

3.2.1 Training and Validation Accuracy and Loss

The accuracy and loss curves show how well each model learned during training and how well it performed on unseen data. All models were trained under the same settings. The results are shown in Table 7.

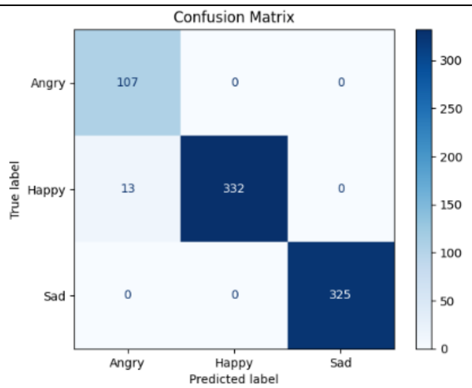
Table 7 Training and validation accuracy and loss for each model

Model	Training and Validation Accuracy and Loss
CNN	
CNN-LSTM	
CNN-GRU	

3.2.2 Confusion Matrix and Classification Report

The confusion matrix shows how well each model predicted the emotions "Angry," "Happy," and "Sad," including both correct and incorrect results. The classification report includes precision, recall, and F1-score. Table 8 presents the confusion matrices for all three models.

Table 8 Confusion matrix for each model

Model	Confusion Matrix
CNN	

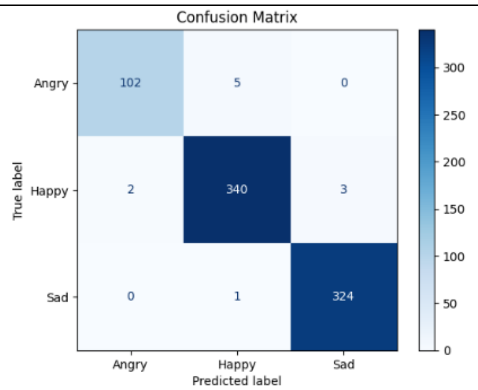
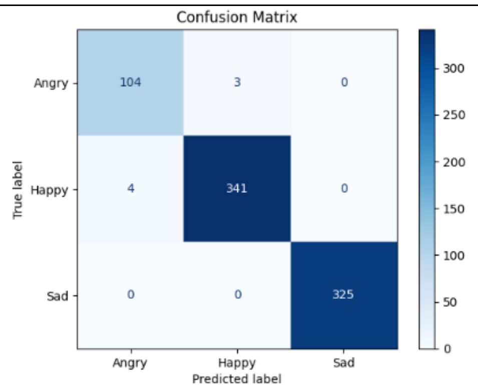
CNN-LSTM	
CNN-GRU	

Table 9 shows the classification report for each model, focusing on weighted accuracy across all emotional categories. CNN-GRU performs best with a weighted accuracy of 99.10%, effectively generalising across the dataset. CNN-LSTM follows with 98.58%, excelling in capturing spatial and temporal features. CNN achieves 98.33%, performing well in extracting spatial features but lagging behind the hybrid models.

Table 9 Classification report for each model

Model	Metric	Angry	Happy	Sad	Macro Avg	Weighted Avg	Weighted Accuracy (%)
CNN	Precision	0.89	1.00	1.00	0.96	0.99	98.33
	Recall	1.00	0.96	1.00	0.99	0.98	
	F1-Score	0.94	0.98	1.00	0.97	0.98	
CNN-LSTM	Precision	0.98	0.98	0.99	0.98	0.99	98.58
	Recall	0.95	0.99	1.00	0.97	0.98	
	F1-Score	0.97	0.98	0.99	0.97	0.98	
CNN-GRU	Precision	0.96	0.99	1.00	0.98	0.99	**99.10**
	Recall	0.97	0.99	1.00	0.99	0.99	
	F1-Score	0.97	0.99	1.00	0.99	0.99	

3.3 Discussion

This section discusses the performance of CNN, CNN-LSTM, and CNN-GRU models based on weighted accuracy, precision, recall, F1-score, and misclassification analysis. Each model's strengths and weaknesses were examined in addressing challenges such as class imbalance and overlapping audio features.

The CNN model, serving as the baseline, achieved a weighted accuracy of 98.33%, demonstrating its ability to extract spatial features effectively. However, it struggled to capture temporal dynamics, leading to a misclassification rate of 1.67%. Most errors involved misclassifying "Happy" segments as "Angry" due to overlapping tonal characteristics. Despite these limitations, CNN achieved perfect precision and recall for "Sad" segments.

The CNN-LSTM model improved performance by capturing both spatial and temporal features. It achieved a higher weighted accuracy of 98.58% and a lower misclassification rate of 1.42%. The model performed well in identifying "Happy" and "Sad" segments, but occasionally confused "Happy" with "Angry" or "Sad" and misclassified some "Angry" segments as "Happy." This indicates better generalisation compared to CNN, although challenges with overlapping rhythmic and tonal patterns remained.

The CNN-GRU model delivered the best results, achieving a weighted accuracy of 99.10% and the lowest misclassification rate of 0.90%. It effectively balanced spatial and temporal feature extraction, resulting in fewer errors. Most misclassifications occurred between "Angry" and "Happy" segments, but the model achieved higher precision, recall, and F1-scores across all categories, reflecting its robust handling of complex audio features and class imbalances.

Overall, the CNN-GRU model emerged as the most suitable for this project due to its superior accuracy and reduced misclassification rate. While CNN and CNN-LSTM performed well, their higher error rates in specific emotional categories limit their practicality. The misclassifications observed in all models highlight the inherent challenge of overlapping tonal and rhythmic features in audio classification.

4. CONCLUSION

This research explored the potential of deep learning models for music emotion classification, focusing on three models: CNN, CNN-LSTM, and CNN-GRU. The study successfully developed a system capable of classifying music into three emotional categories: "Angry," "Happy," and "Sad," by utilising features extracted through MFCC and spectral contrast techniques. The dataset was enhanced through segmentation and augmentation, ensuring robustness and diversity.

Among the models evaluated, CNN-GRU achieved the highest weighted accuracy of 99.10%, demonstrating superior performance in capturing both spatial and temporal features. While CNN and CNN-LSTM showed competitive results, CNN-GRU outperformed in terms of weighted accuracy, precision, recall, and F1-score. This highlights its robustness in handling complex emotional features in music.

The findings of this research contribute to advancements in music emotion classification, providing a foundation for future exploration. Potential improvements include expanding the dataset to include more emotional categories, exploring advanced architectures such as transformers, optimising hyperparameters, and enabling real-time applications.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the Faculty of Electrical Engineering & Technology, Universiti Malaysia Perlis for the continuous support for this research.

REFERENCES

[1] S. Dhanapala and K. Samarasinghe, “Music emotion classification: A literature review,” J. Res. Music, vol. 2, no. 2, pp. 14–28, 2024. [Online]. Available: https://doi.org/10.4038/jrm.v2i2.27

[2] Market.us, “Global AI in music market,” Oct. 2024. [Online]. Available: https://market.us/report/ai-in-music-market

[3] P. L. Louro, H. Redinho, R. Malheiro, R. P. Paiva, and R. Panda, "A comparison study of deep learning methodologies for music emotion recognition," Sensors, vol. 24, no. 7, p. 2201, 2024. [Online]. Available: https://doi.org/10.3390/s24072201
[4] S. Aggarwal, A. Sharma, R. Gupta, and A. Bhardwaj, "Audio segmentation techniques and applications based on deep learning," Scientific Programming, vol. 2022, Article ID 7994191, 2022. [Online]. Available: https://doi.org/10.1155/2022/7994191
[5] N. Titeux, "Everything you need to know about pitch shifting," Apr. 4, 2022. [Online]. Available: https://www.nicolastiteux.com/en/blog/everything-you-need-to-know-about-pitch-shifting
[6] J. Lavoie, "Time stretching and pitch shifting: How to make samples fit your beat," LANDR Blog, Sep. 27, 2024. [Online]. Available: https://blog.landr.com/time-stretching-pitch-shifting
[7] E. Deruty, "Intuitive understanding of MFCCs," Medium, Dec. 15, 2022. [Online]. Available: https://medium.com/@derutycsl/intuitive-understanding-of-mfccs-836d36a1f779
[8] Y. Verma, "A tutorial on spectral feature extraction for audio analytics," Analytics India Magazine, Sep. 19, 2021. [Online]. Available: https://analyticsindiamag.com/ai-trends/a-tutorial-on-spectral-feature-extraction-for-audio-analytics
[9] S. Kumar, "What is label encoding? Application of label encoder in machine learning and deep learning models," Medium, Jan. 13, 2024. [Online]. Available: https://medium.com/@sunnykumar1516/what-is-label-encoding-application-of-label-encoder-in-machine-learning-and-deep-learning-models-c593669483ed
[10] X. Liu, Q. Chen, X. Wu, Y. Liu, and Y. Liu, "CNN based music emotion classification," arXiv preprint, arXiv:1704.05665, 2017. [Online]. Available: https://arxiv.org/abs/1704.05665
[11] M. Gurucharan, "Basic CNN architecture: Explaining 5 layers of convolutional neural network," upGrad Blog, Dec. 7, 2020. [Online]. Available: https://www.upgrad.com/blog/basic-cnn-architecture
[12] C. Chen and Q. Li, "A multimodal music emotion classification method based on multifeature combined network classifier," Math. Probl. Eng., vol. 2020, p. 4606027, 2020. [Online]. Available: https://doi.org/10.1155/2020/4606027
[13] A. Pandey, P. K. Mannepalli, M. Gupta, and R. Dangi, "A deep learning-based hybrid CNN-LSTM model for location-aware web service recommendation," Neural Process. Lett., vol. 56, Article no. 234, 2024. [Online]. Available: https://doi.org/10.1007/s11063-024-11687-w
[14] X. Han, F. Chen, and J. Ban, "Music emotion recognition based on a neural network with an Inception-GRU residual structure," Electronics, vol. 12, no. 4, p. 978, 2023. [Online]. Available: https://doi.org/10.3390/electronics12040978
[15] Y. Xie, J. Wang, H. Li, A. Dong, Y. Kang, J. Zhu, Y. Wang, and Y. Yang, "Deep learning CNN–GRU method for GNSS deformation monitoring prediction," Appl. Sci., vol. 14, no. 10, p. 4004, 2024. [Online]. Available: https://doi.org/10.3390/app14104004