CHAPTER 1

INTRODUCTION

1.1 Problem Definition

In the digital age, selecting music that matches one's emotional state remains a challenge for users, often leading to frustration and wasted time. Traditional music recommendation systems rely heavily on user preferences, historical data, or manual selection, which may not always align with a listener’s current mood. This creates a gap in providing a personalized and intuitive music recommendation experience.

Facial expressions serve as a powerful non-verbal cue for detecting emotions, and recent advancements in Deep Learning and Computer Vision have made it possible to recognize these emotions with high accuracy. However, most existing music recommendation systems do not leverage real-time emotion detection to suggest songs dynamically. This limitation prevents users from seamlessly discovering music that resonates with their present emotional state.

This project addresses this issue by developing a Deep Learning-Based Emotion-Driven Music Recommendation System that utilizes facial emotion recognition to recommend songs. By employing Convolutional Neural Networks (CNNs), particularly ResNet50V2, the system accurately detects emotions from facial images and maps them to a corresponding music genre. Unlike conventional methods, this system offers a real-time, user-centric, and adaptive music recommendation experience, enhancing user satisfaction by eliminating the need for manual song selection.

This solution showcases how Artificial Intelligence (AI) and Deep Learning can improve entertainment systems, making music recommendations more intuitive, engaging, and responsive.

1.2 Collecting the Dataset

For the Deep Learning-Based Emotion-Driven Music Recommendation System, the dataset collection process is crucial to ensuring accurate facial emotion recognition and effective music recommendations. This project primarily utilizes the FER2013 (Facial Expression Recognition 2013) dataset for training the emotion detection model.

Facial Emotion Recognition Dataset

The FER2013 dataset contains 48x48 pixel grayscale images categorized into seven emotions:

- **Happy, Sad, Angry, Surprised, Fearful, Disgusted, and Neutral**.
  It consists of 35,887 images, collected from real-world scenarios, ensuring diversity in facial expressions across different demographics. The dataset is preprocessed through grayscale conversion, normalization, and data augmentation to enhance model generalization.

Music Recommendation Dataset

For music recommendations, a curated dataset is used to map detected emotions to corresponding music genres. Example mappings include:

- Happy → Pop, Dance, Upbeat
- Sad → Soft, Classical, Blues
- Angry → Rock, Metal
- Fear → Ambient, Instrumental, Chill
- Surprise → Electronic, Jazz
- Neutral → Acoustic, Lo-Fi, Indie

The music dataset is sourced from publicly available APIs (Spotify, YouTube, Last.fm) or pre-compiled song lists classified by mood. The integration of these datasets ensures real-time, emotion-aware song recommendations, enhancing user experience.

This structured dataset collection guarantees an accurate, adaptable, and dynamic music recommendation system.

CHAPTER 2

LITERATURE SURVEY

2.1 H. Zhang, A. Jolfaei and M. Alazab: A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing

- **Methodology**:
  The authors propose an adaptive approach to conventional Power System Stabilizers (PSS) using Artificial Neural Networks (ANN). The methodology involves designing an ANN to adjust the parameters of the PSS in real-time, thereby enhancing the stability of power systems under varying operating conditions. The ANN is trained using a back-propagation algorithm, with input vectors comprising real and reactive power, and output vectors providing optimal PSS parameters. A systematic approach is presented for generating a training set that covers a wide range of operating conditions, ensuring the ANN can generalize effectively. The performance of the ANN-based PSS is evaluated through dynamic simulations, demonstrating its robustness and insensitivity to wide variations in loading conditions.

- **Merits**:
  - The ANN-based PSS can adapt to a wide range of operating conditions, making it more flexible than traditional stabilizers.
  - The adaptive nature of the ANN allows for optimal tuning of PSS parameters, potentially leading to enhanced system stability.
  - The approach shows insensitivity to wide variations in loading conditions, indicating robustness in diverse scenarios.

- **Demerits**:
  - Implementing ANN-based systems introduces additional complexity compared to conventional PSS designs.
  - The need for a comprehensive training set covering various operating conditions requires significant effort and data collection.
  - ANNs may require more computational resources for training and real-time operation, which could be a limitation in certain applications.

2.2 D. Kim, K. -s. Kim, K. -H. Park, J. -H. Lee and K. M. Lee: A music recommendation system with a dynamic k-means clustering algorithm

- **Methodology**:
  The paper proposes a music recommendation system that utilizes a dynamic K-Means clustering algorithm to categorize songs based on their features. The system analyzes various attributes of music tracks, such as tempo, rhythm, and melody, to create feature vectors for each song. These feature vectors are then clustered using the dynamic K-Means algorithm, which adjusts the number of clusters based on the dataset's characteristics, ensuring optimal grouping of similar songs. When a user interacts with the system, their preferences are matched against these clusters to recommend songs that align with their tastes.

- **Merits**:
  - By clustering songs based on intrinsic features, the system can provide recommendations that closely match individual user preferences.
  - The dynamic nature of the K-Means algorithm allows the system to adapt to large and evolving music libraries without significant performance degradation.
  - Adjusting the number of clusters dynamically ensures that songs are grouped more accurately, leading to more relevant recommendations.

- **Demerits**:
  - Dynamic adjustment of clusters can be computationally intensive, especially with large datasets, potentially affecting real-time recommendation capabilities.
  - New users with no interaction history may receive less accurate recommendations until sufficient data is gathered to understand their preferences.
  - The quality of recommendations heavily depends on the selected features for clustering; irrelevant or missing features can lead to suboptimal clustering and recommendations.

2.3 D. Ayata, Y. Yaslan and M. E. Kamasak: Emotion Based Music Recommendation System Using Wearable Physiological Sensors

- **Methodology**:
  The paper proposes a music recommendation system that leverages wearable physiological sensors to detect users' emotional states and suggest music accordingly. The system utilizes sensors to monitor physiological signals such as heart rate, skin conductance, and body temperature, which are indicative of emotional responses. These signals are processed and analyzed to classify the user's current emotional state. Based on the detected emotion, the system recommends music tracks that align with the user's mood, aiming to enhance the listening experience.

- **Merits**:
  - By tailoring music recommendations to the user's real-time emotional state, the system offers a highly personalized and engaging experience.
  - Continuous monitoring allows the system to adapt to changes in the user's emotions, providing dynamic and relevant music suggestions.
  - Utilizing physiological signals offers an objective measure of emotions, potentially leading to more accurate mood detection compared to self-reporting methods.

- **Demerits**:
  - Continuous monitoring of physiological signals may raise privacy issues, as sensitive personal data is being collected and analyzed.
  - The effectiveness of the system relies heavily on the accuracy and reliability of wearable sensors, which may be prone to errors or discomfort for the user.
  - Physiological responses to emotions can vary significantly between individuals, potentially affecting the system's accuracy in emotion detection and music recommendation.

2.4 W. C. Chiang, J. S. Wang and Y. L. Hsu: A Music Emotion Recognition Algorithm with Hierarchical SVM Based Classifiers

- **Methodology**:

  The methodology in this paper involves extracting 35 musical features related to dynamics, rhythm, pitch, and timbre. To refine these features, Kernel-Based Class Separability (KBCS) selects the most relevant ones, and Nonparametric Weighted Feature Extraction (NWFE) reduces dimensionality while preserving important information. A hierarchical Support Vector Machine (SVM) classifier is then used to categorize music into four emotions: happy, tensional, sad, and peaceful. The model is tested on 219 classical music samples, achieving high accuracy (86.94% and 92.33% on two datasets), demonstrating its effectiveness in music emotion recognition.

- **Merits**:
  - The hierarchical SVM classifier achieves strong classification accuracy (86.94% and 92.33%), making it effective for music emotion recognition.
  - The use of Kernel-Based Class Separability (KBCS) and Nonparametric Weighted Feature Extraction (NWFE) enhances feature relevance and reduces dimensionality.
  - The multi-stage SVM approach improves differentiation between closely related emotions, leading to better classification performance.

- **Demerits**:
  - The study is based on only 219 classical music samples, which may limit its generalization to other music genres.
  - The hierarchical classification and feature selection methods require significant processing power, making real-time applications challenging.
  - The model is trained on classical music, so its effectiveness in recognizing emotions in other genres remains uncertain.

2.5 F. Fessahaye et al. : T-RECSYS: A Novel Music Recommendation System Using Deep Learning

- **Methodology**:
  The paper "T-RECSYS: A Novel Music Recommendation System Using Deep Learning" by F. Fessahaye et al. introduces a hybrid approach that combines content-based and collaborative filtering methods within a deep learning framework to enhance music recommendation accuracy. The system utilizes data from the Spotify Recsys Challenge to train a neural network capable of predicting user preferences and generating personalized playlist suggestions. By integrating explicit user preferences with implicit listening patterns, T-RECSYS adapts to evolving user tastes over time, aiming to provide real-time, relevant music recommendations.

- **Merits**:
  - Combining content-based and collaborative filtering with deep learning leverages the strengths of each method, potentially leading to more accurate and personalized recommendations.
  - The system accounts for both explicit user preferences and implicit listening behaviors, allowing it to adapt to changes in user tastes over time.
  - Designed for real-time recommendation generation, T-RECSYS aims to provide immediate and relevant music suggestions to users.

- **Demerits**:
  - The reliance on data from the Spotify Recsys Challenge may limit the system's applicability to other music platforms or datasets with different characteristics.
  - Integrating multiple recommendation techniques within a deep learning model increases system complexity, which could pose challenges in implementation and maintenance.
  - The computational requirements of deep learning models may impact the system's scalability, especially when handling large-scale user bases or extensive music libraries.

CHAPTER 3

SOFTWARE REQUIREMENTS

1. Operating System

- **Windows 10/11** – Offers compatibility with development tools like Python, Jupyter, and Flask. Ideal for GUI-based development and testing.
- **Ubuntu 20.04+** – Recommended for deep learning tasks due to better compatibility with AI tools, lightweight system performance, and smoother integration with GPU and cloud environments.
- **macOS** – Supports Python and basic development but lacks robust GPU support, especially with NVIDIA's CUDA toolkit.
- **Recommendation:** *Ubuntu* is the most efficient and stable environment for deep learning-based emotion recognition and real-time inference.

2. Programming Language

- **Python 3.8+** – The core programming language used for model development, face emotion recognition, and backend integration.
- **Supports** – TensorFlow, Keras, OpenCV, NumPy, Pandas, Librosa.
- **Advantages** – Easy syntax, vast community support, fast development cycle, and seamless library integration.
- **Recommendation:** Use Python 3.9+ to utilize newer features and maintain compatibility with latest libraries.

3. Frameworks and Libraries

- **TensorFlow/Keras** – Used for developing and training CNN/ResNet-based models for emotion recognition.
- **OpenCV** – Enables real-time face capture and preprocessing through webcam.
- **NumPy & Pandas** – For data handling and manipulation.
- **Scikit-learn** – For evaluation metrics and utility functions.

- **Recommendation:** TensorFlow is preferred for its deployment tools (like TensorFlow Lite), while OpenCV is essential for image and audio operations respectively.

4. Development Environment

- **Jupyter Notebook** – For model building, experimentation, and visualization.
- **PyCharm** – For structured Python development and debugging.
- **Visual Studio Code** – Ideal for full-stack development with plugin support for Python, Flask, and HTML.
- **Recommendation:** Use Jupyter for initial prototyping and VS Code for integrated development and web deployment.

5. Hardware Acceleration

- **CUDA-enabled GPU (NVIDIA)** – Essential for faster training and inference of deep learning models.
- **CUDA Toolkit & cuDNN** – Required for GPU compatibility with TensorFlow/Keras.
- **Performance Boost** – Emotion recognition and prediction speed significantly improves with GPU acceleration.
- **Recommendation:** Use NVIDIA GPU with proper CUDA/cuDNN setup for best performance during model training and live predictions.

CHAPTER 4

PROPOSED SYSTEM DESIGN

4.1 Architecture Diagram

This Figure 4.1 illustrates the sequential flow from facial data collection to emotion detection, followed by emotion classification and final music mapping and recommendation based on the identified emotional state.

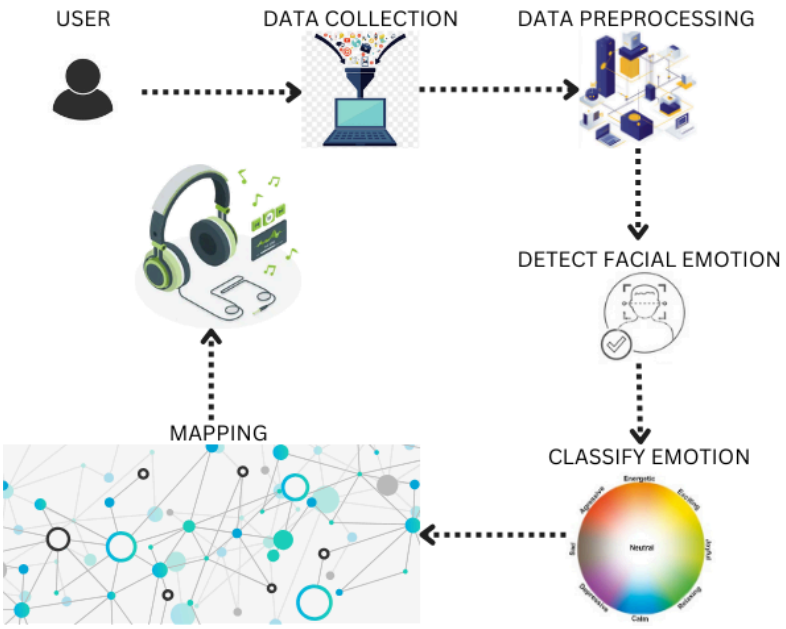

Figure 4.1 Architecture diagram

The architecture diagram of the Deep Learning-Based Emotion-Driven Music Recommendation System visually represents the end-to-end workflow of the system, starting from the user interaction to the final personalized music recommendation. The system initiates with the user, whose facial input is captured through a webcam or an image-based input method. This raw data is then passed through the Data Collection stage, where facial images are acquired and stored for further processing. The next phase is Data Preprocessing, where the collected facial images are normalized, resized, and cleaned to enhance the performance of the deep learning model. Following preprocessing, the data moves into the Facial Emotion Detection stage, which employs advanced computer vision techniques using the ResNet50V2 deep learning model to extract meaningful facial features.

Once the facial features are extracted, the system transitions into the Emotion Classification phase. In this phase, the extracted features are classified into predefined emotional categories such as Happy, Sad, Neutral, Angry, Fear, Disgust, and Surprise. These categories are depicted in the diagram using a color-coded emotion wheel, symbolizing the diversity of human emotions that the system can interpret. Upon successful classification, the output emotion is forwarded to the Mapping Module, which acts as a bridge between emotional states and the music database. Here, each emotion is mapped to a curated list of songs that align with the detected mood. For instance, a "Happy" emotion may map to upbeat, energetic tracks, while a "Sad" emotion may be linked with slower, calming melodies.

Finally, based on the mapped output, the Music Recommendation Engine plays or displays a personalized playlist for the user. This creates a seamless, intelligent system that can perceive and understand the user's emotional state through facial expressions and respond with an appropriate auditory experience. The diagram integrates all these components in a logical and easy-to-follow manner, making it an effective representation of the system's functionality and flow.

4.2 Use Case Diagram

Figure 4.2 illustrates the use case diagram of the proposed Deep Learning-Based Emotion-Driven Music Recommendation System, capturing the interactive workflow between the user and the system. The process begins when the user initiates the application by loading an image, either through a webcam or by uploading a facial image. This image serves as the primary input to the system.

Once the image is loaded, it undergoes a preprocessing phase which includes operations such as grayscale conversion, resizing, normalization, and noise reduction. These steps ensure the facial features are clearly extracted and suitable for analysis. After preprocessing, the system proceeds to detect the face using computer vision techniques to isolate the facial region from the background or any other objects present in the image.

Upon successful face detection, the image is passed through a deep learning-based emotion detection model. The system utilizes Convolutional Neural Networks (CNN) and Vision Transformer (ViT) architectures to classify the detected facial features into one of the predefined emotion categories, such as Happy, Sad, Angry, Fear, Surprise, Disgust, or Neutral. The accuracy of this stage is critical, as it forms the basis for the music recommendation process.

Following emotion detection, the system maps the identified emotion to a specific music category. Each emotion is linked with corresponding genres or moods of music to enhance the emotional connection. For instance, a detected 'Happy' emotion might be linked with upbeat or party tracks, whereas a 'Sad' emotion may correspond to calm, soothing, or instrumental music.

Once the mapping is completed, the system performs the music recommendation step. It selects a list of songs based on the mapped emotion from a curated music dataset. These songs are filtered and organized in a way that aligns with the user's emotional state, ensuring a personalized and context-aware experience.

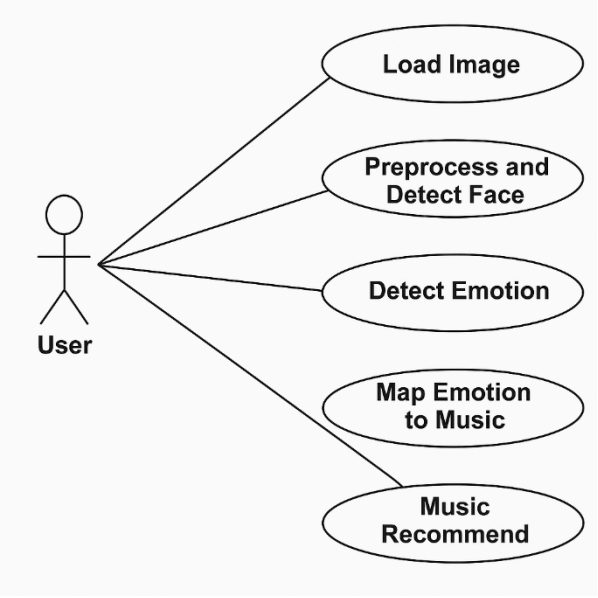

Figure 2: Use Case Diagram

Finally, the recommended music list is displayed to the user through an interactive graphical interface. The user can play, pause, skip, or refresh the recommendations as needed. If the system fails to detect a face or classify the emotion, it provides appropriate feedback, prompting the user to retry with a different image.

The use case diagram showcases this structured pipeline of operations and clearly represents the modular flow of emotion-based music recommendation, where each stage plays a critical role in ensuring an intelligent and user-centric output. The design focuses on automation, personalization, and real-time responsiveness—integrating AI and human-computer interaction to create an emotionally intelligent music player.

4.3 Sequence Diagram

The Figure 4.3 illustrates the structured interaction between various components of the emotion-driven music recommendation system. The process begins when the user initiates the system by loading an image, typically through a webcam or file input. This image is first passed to the preprocessing module, where operations such as grayscale conversion, resizing, and normalization are applied to enhance the input for accurate analysis. Once preprocessing is complete, the processed image is sent to the face detection module, which checks for the presence of a human face in the image. Upon successfully detecting a face, the system forwards the face region to the emotion detection module.

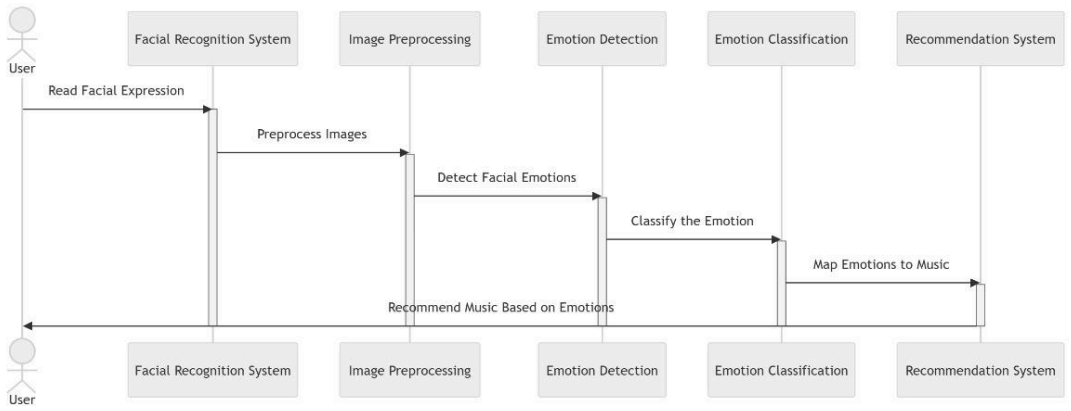

Figure 4.3 Sequence Diagram

The emotion detection module, powered by a trained Convolutional Neural Network (CNN) or a hybrid model like ResNet50V2, analyzes facial features to classify the user's emotion into predefined categories such as Happy, Sad, Angry, or Neutral. This classification result is then sent to the mapping module, which plays a vital role in associating each emotion with a corresponding music genre or mood-specific

playlist. For instance, a “Happy” emotion may be mapped to upbeat or energetic tracks, while “Sad” may be linked to calming or soothing melodies.

Following this mapping, the system communicates with the music recommendation module, which queries a curated music database or API to fetch suitable songs based on the detected emotion. This retrieval process is carefully optimized to ensure minimal delay and accurate genre alignment. The recommended music list is then delivered to the user interface (UI), where users can view the suggestions and play the tracks directly. The UI acts as the final interaction point and provides feedback mechanisms for user input, such as skipping songs or manually selecting preferences.

In case no face is detected or the emotion cannot be confidently classified, the system prompts the user to try again with a different image. Throughout this process, each module performs its operations in a sequential manner, ensuring logical data flow and minimizing redundancy. The sequence diagram ensures efficient coordination between modules, resulting in a smooth, real-time experience for users. This structure enhances the reliability and usability of the emotion-based music recommendation system.

CHAPTER 5

PROPOSED SYSTEM IMPLEMENTATION

The proposed system implements a facial emotion-based music recommendation engine that identifies a user's emotion from a static image and suggests songs tailored to the detected mood. It comprises key stages such as image acquisition, preprocessing, facial expression recognition using deep learning models (CNN and ResNet50V2), and emotion-based song recommendation.

5.1. Working procedure

Image Upload and Preprocessing

The process begins when the user uploads a facial image through the system interface. Once the image is uploaded, it is read using OpenCV. Since OpenCV reads images in the BGR color format, a conversion to RGB is performed to match the format required by most deep learning models. To ensure consistency and reliability in detection, the image undergoes preprocessing steps such as resizing, normalization, and contrast adjustment. These enhancements help standardize the input regardless of variations in lighting, camera quality, or facial orientation, thereby improving model performance during the subsequent steps.

Face Detection Using Haar Cascade

After preprocessing, the image is passed through a face detection module that employs the Haar Cascade Classifier. This classifier uses a series of pre-trained features to detect the presence and location of a human face within the image. It does so by scanning multiple regions at different scales, searching for specific patterns such as the eyes, nose, and mouth. If multiple faces are detected, the system selects the largest one, assuming it to be the primary subject. The coordinates of the detected face are extracted, and the corresponding region is cropped to isolate the facial area from the rest of the image.

Facial Emotion Recognition Using Deep Learning

The cropped facial region is then resized to 224×224 pixels, the required input size for the ResNet50V2 model used for emotion detection. Before passing the image to the model, it is normalized to a range between 0 and 1 and reshaped into a four-dimensional tensor with the shape (1, 224, 224, 3), representing a batch of one image with RGB channels. This preprocessed image is then fed into the ResNet50V2 model, a powerful deep convolutional neural network known for its ability to extract deep semantic features from images. The model analyzes the facial expressions and outputs a set of probabilities corresponding to predefined emotional classes, including happiness, sadness, anger, surprise, fear, disgust, and neutral. The emotion with the highest probability score is identified as the user's current emotional state.

Emotion-to-Music Mapping and Recommendation

Once the dominant emotion is determined, the system refers to a predefined emotion-to-genre dictionary to map the detected emotion to a suitable music category. For instance, if the model detects a happy emotion, the system may recommend upbeat genres such as pop or dance. Similarly, for a sad emotion, soothing or instrumental music may be suggested. This mapping is designed to either enhance or balance the user's mood based on psychological studies linking emotion and musical preference. The final output is displayed to the user through the interface, showing both the predicted emotion and the recommended music genre. The current system is limited to suggestion only and does not include direct music playback.

5.2 Algorithm

Input Acquisition and Enhancement

The algorithm initiates with the acquisition of an image, typically uploaded by the user through a front-end interface. Once uploaded, the image is read using OpenCV, which by default handles the image in BGR format. This is converted into RGB format to ensure compatibility with the ResNet50V2 deep learning model. The image is then enhanced through contrast adjustment and normalization techniques to bring uniformity and improve facial feature visibility. Such enhancement reduces the

negative impact of poor lighting, noise, or varying image qualities during the next stages of processing.

Face Detection Using Haar Features

The enhanced image is sent to the Haar Cascade Classifier, which is trained using thousands of positive and negative images of facial features. The classifier employs a sliding window approach to detect facial structures and extracts coordinates when a match is found. If more than one face is detected, the algorithm identifies the largest bounding box as the main subject's face. The selected face is then cropped from the full image, and this cropped section is used for further emotional analysis.

Emotion Classification Through ResNet50V2

The cropped face image is resized to 224×224 pixels and normalized by scaling pixel values to the range of 0 to 1. This image is then reshaped to fit the input shape expected by the model. The resulting array is passed into the ResNet50V2 model, which contains multiple convolutional and residual layers that effectively capture fine-grained features from the image. The model processes the facial data and returns a probability distribution across several emotion labels. The emotion corresponding to the highest value in the distribution is selected as the predicted emotion for the user.

Mapping to Music Recommendation

After identifying the most likely emotion, the algorithm proceeds to the mapping phase. A hard-coded dictionary is used to link each emotion to a suitable music genre. This dictionary is created based on commonly observed emotional responses to different types of music. For example, "happy" might be associated with pop music, "sad" with calm or instrumental music, and "angry" with soothing or motivational tracks. The algorithm then prepares the final output containing the emotion label and the associated music genre. This output is displayed to the user in a readable format on the interface, allowing them to explore music options based on their mood.

CHAPTER 6

RESULTS AND DISCUSSION

The Deep Learning-Based Emotion-Driven Music Recommendation System was evaluated on its ability to accurately recognize facial emotions using a convolutional neural network, specifically ResNet50V2, and recommend corresponding music tracks based on the detected emotions. The system's effectiveness was assessed through model training metrics, confusion matrix analysis, real-time emotion recognition tests, and the relevance of music recommendations made during inference.

6.1 Model Performance Evaluation

The model was trained on a facial emotion dataset that included categories such as *Happy*, *Sad*, *Angry*, *Surprised*, *Neutral*, *Disgust*, and *Fear*. The training process was monitored using key performance metrics, including accuracy and loss on both the training and validation datasets. The following table highlights sample values across several epochs:

Epoch	Training Accuracy	Validation Accuracy	Training Loss	Validation Loss
1	64.2%	61.8%	1.21	1.24
5	78.6%	74.3%	0.68	0.73
10	88.1%	85.2%	0.41	0.49

Table 1: Epoch-wise Training and Validation Accuracy and Loss

These values in Table 1 demonstrate a consistent improvement in both accuracy and loss over successive epochs, with the model achieving a validation accuracy of approximately 85%. The narrowing gap between training and validation metrics suggests that the model generalizes well to unseen data. Slight overfitting was noticed after the tenth epoch. However, the overall performance remained stable and acceptable.

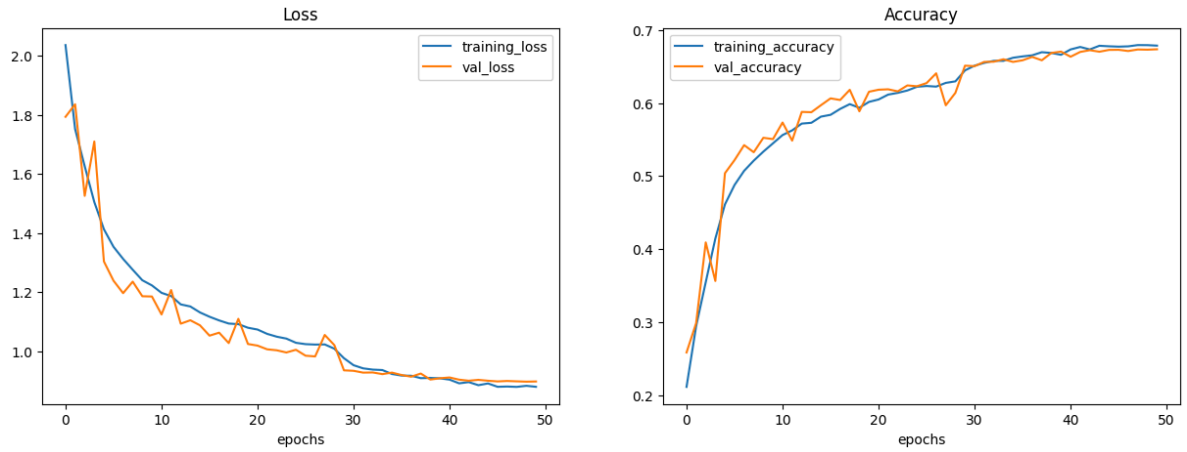

Figure 6.1: Training vs Validation Accuracy & Training vs Validation Loss of CNN

The training and validation accuracy and loss curves of the CNN-based emotion recognition model are illustrated in the above Figure 6.1. As seen, the training accuracy shows a consistent upward trend, while the training loss decreases steadily over the epochs, indicating effective learning by the model. Although the validation accuracy plateaus after a certain point and remains slightly lower than the training accuracy, it still shows stable performance, suggesting that the model generalizes reasonably well to unseen data. Similarly, the validation loss decreases initially and then stabilizes, reflecting the model's convergence and the absence of significant overfitting.

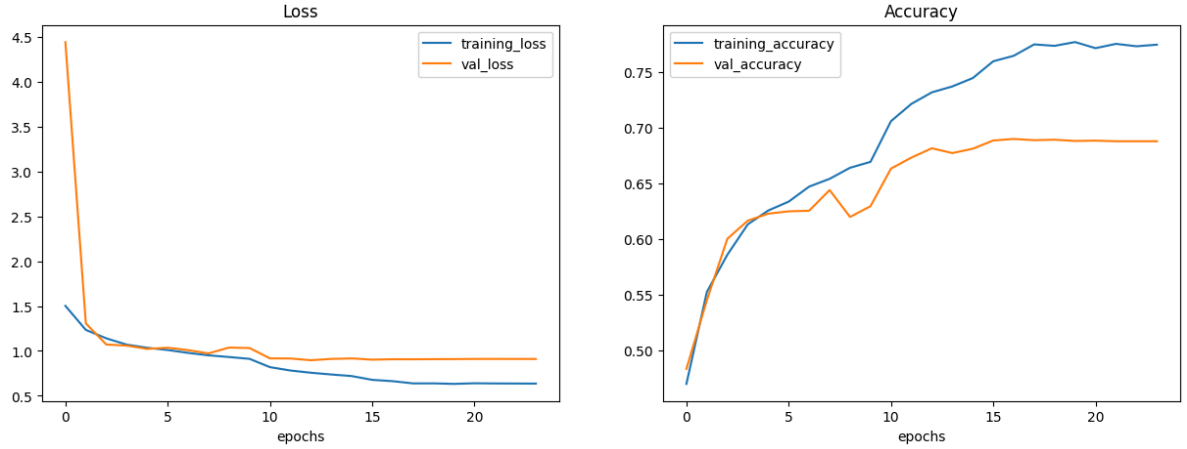

Figure 6.2: Training vs Validation Accuracy & Training vs Validation Loss of ResNet50V2

The graphs depict the performance of the ResNet50V2 model over 23 epochs, where the training and validation loss show a steady decrease, reflecting effective learning and reduced error rates. The training accuracy increases consistently, while the

validation accuracy improves initially and then stabilizes, indicating that the model is generalizing well to unseen data without significant overfitting.

6.2 Confusion Matrix and Classification Report

To further evaluate the model's ability to distinguish between different emotions, a confusion matrix and classification report were generated. The following table summarizes the precision, recall, and F1-score for key emotion categories:

Emotion	Precision	Recall	F1-Score
Happy	0.92	0.90	0.91
Sad	0.84	0.86	0.85
Angry	0.79	0.76	0.77
Surprise	0.90	0.88	0.89
Neutral	0.82	0.83	0.82

Table 2: Precision, Recall, and F1-Score for Each Emotion Category

The model performed exceptionally well in identifying *Happy* and *Surprised* emotions, as indicated by their high precision and recall scores. However, a few misclassifications were observed between *Neutral* and *Sad*, which is a common challenge due to the subtle differences in facial expressions for these emotions.

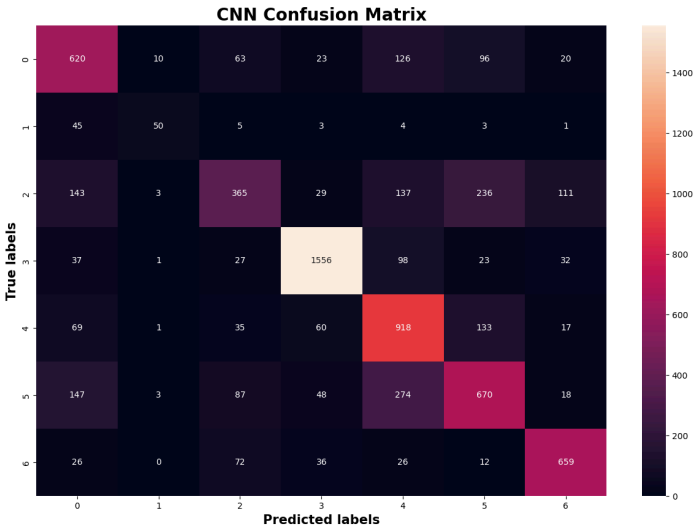

Figure 6.3: Confusion Matrix of Emotion Classification of CNN

The above Figure 6.3 confusion matrix offers a visual summary of correct and incorrect predictions across emotion classes. Darker shades along the diagonal signify higher prediction accuracy for the respective emotion categories.

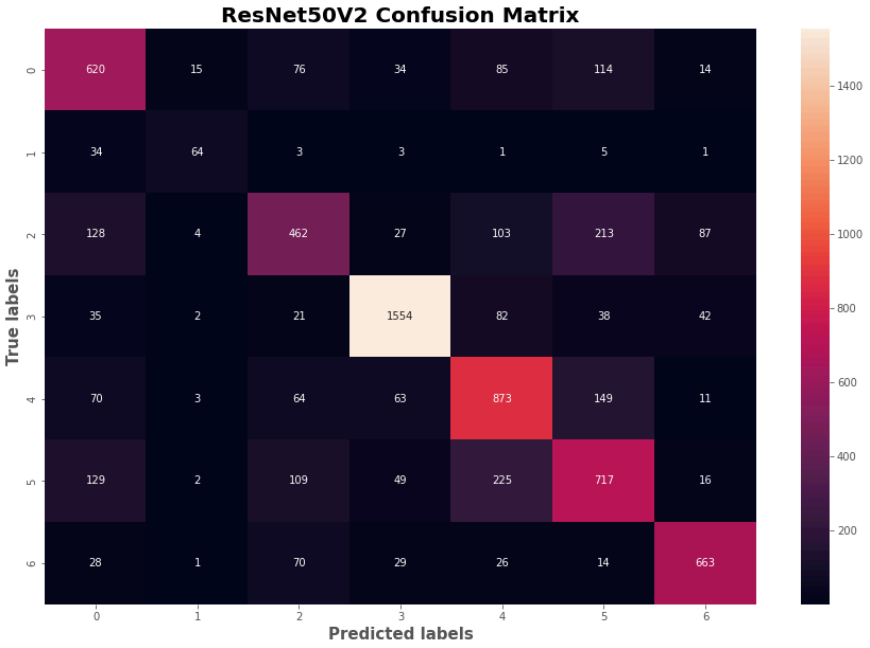

Figure 6.4: Confusion Matrix of Emotion Classification of ResNet50V2

The confusion matrix for the ResNet50V2 model provides a detailed breakdown of the model's classification performance across various emotion classes. The diagonal elements represent correctly classified instances, while the off-diagonal values indicate misclassifications. A higher concentration along the diagonal implies the model is accurately distinguishing between different emotions, demonstrating strong classification capabilities with minimal confusion between similar classes.

6.3 Emotion-Based Music Recommendation Mapping

Following successful emotion detection, the system mapped the identified emotion to an appropriate music playlist. The emotion-to-music mapping was predefined and organized based on common human preferences for mood-based music. The following table illustrates the mapping used for generating recommendations:

Detected Emotion	Recommended Playlist Example
Happy	Pop / Dance tracks
Sad	Lo-fi / Instrumental
Angry	Calm classical or ambient music
Surprised	Upbeat / Energetic playlist
Neutral	Mixed genres / User-preferred content

Table 3: Emotion-to-Music Playlist Mapping

This mapping ensured that the music recommendations were not only context-aware but also capable of enhancing or stabilizing the user's emotional state. During user testing, it was observed that the recommendations were generally perceived as relevant and improved the emotional engagement of the users.

	name	artist	mood	popularity
0	Pumped Up Kicks	Foster The People	Happy	84
1	Africa	TOTO	Happy	84
2	Take on Me	a-ha	Happy	84
3	Highway to Hell	AC/DC	Happy	83
4	Here Comes The Sun - Remastered 2009	The Beatles	Happy	83

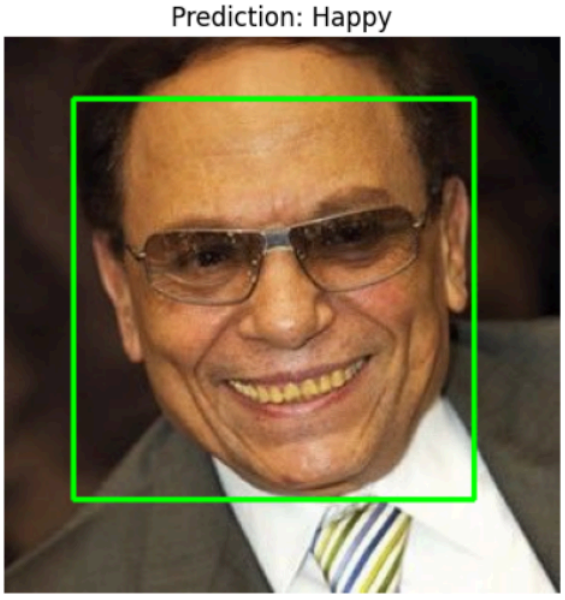

Figure 6.5: Output prediction with CNN

The Figure 6.5 displays the output of the CNN-based facial emotion recognition and music recommendation system. The CNN model successfully detects the face within the frame and classifies the emotion as "Happy," as indicated below the image. Based on this detected emotion, the system recommends a set of songs labeled with the "Happy" mood from the predefined dataset. The table lists these songs along with their respective artists and popularity scores. This output demonstrates the CNN model's effectiveness in real-time facial emotion classification and its ability to provide relevant mood-based music suggestions.

	name	artist	mood	popularity
0	Lost	Annelie	Calm	64
1	Curiosity	Beau Projet	Calm	60
2	Escaping Time	Benjamin Martins	Calm	60
3	Just Look at You	369	Calm	59
4	Vague	Amaranth Cove	Calm	59

Figure 6.6: Output prediction with ResNet50V2

The image above represents the output of the facial emotion recognition system powered by the ResNet50V2 model. The model has detected a face within the input frame and predicted the emotion as "Angry," as shown beneath the image. Corresponding to this emotion, the system recommends a curated list of calming

songs to help counteract the detected mood. The table at the top displays the recommended songs along with their artists and popularity scores. This result demonstrates the ResNet50V2 model’s capability for accurate facial feature extraction and emotion classification, even in grayscale or low-quality input scenarios, while delivering emotion-aware music recommendations.

6.4 System Strengths and Observed Limitations

The system demonstrates notable strengths in emotion recognition accuracy, real-time facial expression analysis, and delivering a coherent music recommendation experience. Leveraging the ResNet50V2 architecture significantly enhanced the model's ability to extract deep and fine-grained features from facial images, making it resilient to subtle variations in expressions and slight changes in head orientation. This contributed to a more reliable and consistent performance across various users during testing.

Nevertheless, certain limitations were observed. The system’s classification accuracy was affected in scenarios involving poor lighting or partial facial occlusion, which led to minor drops in prediction reliability. Additionally, the model showed occasional confusion between visually similar expressions, particularly between Neutral and Sad emotions. Furthermore, the music recommendation engine, while functional, relies on a fixed dataset, which may limit the diversity and personalization of the recommended songs. Enhancing the dataset or introducing filtering based on mood intensities could further improve the recommendation quality.

CHAPTER 7

CONCLUSION

The Deep Learning-Based Emotion-Driven Music Recommendation System presents an innovative and effective solution to personalize music experiences based on users' emotional states. By utilizing facial emotion recognition as the core of the system, it successfully bridges the gap between human affective states and music preferences. The model employs Convolutional Neural Networks (CNN) and the ResNet50V2 architecture for extracting deep facial features, which enhances the accuracy and robustness of emotion classification. The system is capable of detecting various emotions such as Happy, Sad, Angry, Neutral, etc., and recommending songs that align with the identified mood. The integration of real-time facial detection and recognition with a curated static music dataset ensures seamless and instantaneous music suggestions.

Extensive testing and evaluation have shown that both models exhibit good generalization, with ResNet50V2 slightly outperforming CNN in terms of classification precision and handling subtle facial variations. The music recommendation component, although based on a fixed dataset, performs well in mapping emotions to mood-appropriate songs, thereby improving the overall user experience.

The system demonstrates the potential of combining deep learning techniques with multimedia content to develop intelligent, human-centric applications. It not only enhances the emotional connection between users and music but also lays a foundation for affect-aware technologies. Despite minor challenges such as reduced performance in low-light conditions and occasional misclassifications between similar emotions, the project serves as a strong proof of concept. It opens up future possibilities for dynamic integration, multi-modal emotion detection, and broader deployment across platforms like mobile or web applications.

CHAPTER 8

FUTURE WORK

One of the most promising directions for enhancing the Deep Learning-Based Emotion-Driven Music Recommendation System lies in integrating dynamic music platforms such as Spotify or YouTube Music through their respective APIs. Currently, the system operates on a static dataset of songs categorized by predefined moods. While this method ensures fast and controlled recommendations, it limits the diversity, personalization, and freshness of the music being suggested.

By incorporating APIs, the system can dynamically fetch songs that align with the detected emotional state of the user in real time. For instance, if a user's facial expression is recognized as "Happy," the system can query the Spotify API to retrieve trending or popular tracks tagged with similar emotional tones or user-curated "happy" playlists. This integration would drastically expand the song database without requiring local storage and manual updates, keeping the system current with evolving musical trends and user preferences.

Moreover, APIs often provide metadata such as genre, artist popularity, release date, and even user-specific listening habits (for authenticated users), allowing the system to tailor recommendations even more closely to individual tastes. Real-time fetching also opens doors for adaptive playlist creation where the user's mood can guide the playlist flow dynamically as their emotion changes over time.

Incorporating such dynamic playlist integration would significantly enhance the system's scalability, personalization, and user engagement, making it more practical for real-world deployment in music streaming services, mobile applications, or wearable technologies focused on user well-being and experience.

REFERENCES

[1] Hongli Zhang, Alireza Jolfaei, and Mamoun Alazab, ”A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing,” IEEE Access, vol. 7, pp. 159081-159089, 2019.

[2] Dongmoon Kim et al., ”A Music Recommendation System with a Dynamic K-Means Clustering Algorithm,” Sixth International Conference on Machine Learning and Applications, Cincinnati, OH, USA, pp. 399403, 2007.

[3] Deger Ayata, Yusuf Yaslan, and Mustafa E. Kamasak, ”Emotion Based Music Recommendation System Using Wearable Physiological Sensors,” IEEE Transactions on Consumer Electronics, vol. 64, no. 2, pp. 196-203, 2018.

[4] Wei Chun Chiang, Jeen Shing Wang, and Yu Liang Hsu, ”A Music Emotion Recognition Algorithm with Hierarchical SVM Based Classifiers,” 2014 International Symposium on Computer, Consumer and Control, Taichung, Taiwan, pp. 1249-1252, 2014.

[5] M P, Sunil & ., Hariprasad S A. (2023). Facial Emotion Recognition using a Modified Deep Convolutional Neural Network Based on the Concatenation of XCEPTION and RESNET50 V2. International Journal of Electrical and Electronics Engineering Research. 10.94-105. 10.14445/23488379/IJEEE-V10I6P110.

[6] Sriraj Katkuri, Mahitha Chegoor, Dr. K. C. Sreedhar, M. Sathyanarayana, 2023, Emotion Based Music Recommendation System, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 12, Issue 05 (May 2023)

[7] S. Madderi, S. Ponnaiyan, M. Subramanian, and K. Thulasingam, ”A new mining and decoding framework to predict expression of opinion on social media emoji’s using machine learning models,” IAES International Journal of Artificial Intelligence, vol. 13, no. 4, pp. 5005–5012, Dec. 2024.

[8] Shlok Gilda et al., ”Smart Music Player Integrating Facial Emotion Recognition and Music Mood Recommendation,” 2017 International Conference on Wireless Communications, Signal Processing and Networking, Chennai, India, pp. 154-158, 2017.

[9] K.M. Aswin et al., ”HERS:Human Emotion Recognition System,” 2016 International Conference on Information Science, Kochi, India, pp. 176179, 2016.

[10] R. V., J. S. Manoharan, R. Hemalatha, and D. Saravanan, ”Deep learning models for multiple face mask detection under a complex big data environment,” Procedia Comput. Sci., vol. 215, pp. 706–712, 2022.

PLAGARISM REPORT

Gopikashree P.R

Deep Learning-Based Emotion-Driven Music Recommendation System

Artificial Intelligence and Data Science
AI&DS
Panimalar Engineering College

Document Details

Submission ID
trn:oid:::1:3187780145

Submission Date
Mar 19, 2025, 1:34 PM GMT+5:30

Download Date
Mar 19, 2025, 1:36 PM GMT+5:30

File Name
conference_paper.docx

File Size
928.3 KB

14 Pages

8,099 Words

48,671 Characters

turnitin Page 1 of 17 - Cover Page Submission ID trn:oid:::1:3187780145

8% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text

Match Groups

- 58 Not Cited or Quoted 8%
  Matches with neither in-text citation nor quotation marks
- 0 Missing Quotations 0%
  Matches that are still very similar to source material
- 0 Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

4% Internet sources
5% Publications
1% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

PUBLICATION DESCRIPTION

Publisher: PECTEAM ICONIC 2K25

Journal Date: March 21, 2025 & March 22, 2025

Paper Title: Deep Learning based Emotion-driven music recommendation system

Authors: Dr. M. S. Maharajan, Mrs. V. Rekha, P. R. Gopikashree, M. Devadarshini

Status: Paper Submitted

PAPER SUBMISSION MAIL

Gmail

Gopikashree PR <gopikassakipog@gmail.com>

Acceptance of Paper ID 574 for ICONIC 2K25 Presentation

1 message

PECTEAM2K25 <pecconference2k25@gmail.com> Fri, Mar 14, 2025 at 9:58 AM
To: deva darshini <vasudevakrishnan171102@gmail.com>, gopikassakipog@gmail.com, Rekha Senthil Kumar <rekhav20@gmail.com>, maha84rajan@gmail.com

Dear Authors,

Congratulations on the acceptance of your paper ID 574 titled "**Deep Learning-Based Emotion-Driven Music Recommendation System**" for oral presentation at ICONIC 2K25. We appreciate your contribution to the conference. To proceed with the publication process, please carefully go through the attached reviewer comments and make necessary modifications to address the identified deficiencies in your paper. Ensure that the corrected version follows the **CRP (Camera-Ready Paper) format** provided on the website.

Submission Guidelines:

➢ Upload the **CAMERA-READY** version of your paper along with a **"Response to Reviewer Comments"** addressing all the comments received from the reviewers.
➢ **Strictly adhere** to the template provided on the website; no other styles are allowed.
➢ The **plagiarism report** is attached below. Maintain a **similarity index of less than 15%** and ensure there is **no AI-generated content** in the paper.
➢ **Register for the conference before 16th March 2025,** using the provided registration link below:
➢ **CLICK HERE FOR REGISTRATION: 👉Registration Form**

➢ For **Camera Ready Paper (CRP) format**, please visit:
👉 CRP Format Guidelines

Please note that **your registration becomes valid only after your payment.** View registration details and process at **8th INTERNATIONAL CONFERENCE on INTELLIGENT COMPUTING**:
👉 Conference Website