Automated Personalized Mood-Based Song Selector

Nilay Jain
Department of Computer Science
California State University, Fullerton
Fullerton, USA
njain12@csu.fullerton.edu

Kanika Sood
Department of Computer Science
California State University, Fullerton
Fullerton, USA
kasood@fullerton.edu

2024 International Conference on Computing, Sciences and Communications (ICCSC) | 979-8-3503-5364-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICCSC62048.2024.10830371

***Abstract*— Music and weather significantly influence individuals' moods, playing pivotal roles in daily life. This research explores integrating music and weather data to enhance the driving experience by dynamically adjusting the music playlist based on the driver's mood and the current weather conditions, considering the potential hazards of manual song selection while driving. The research addresses the challenge of maintaining a consistent mood during long drives, which can be significantly influenced by the music played. We propose a solution that involves an AI-integrated application utilizing facial emotion recognition to gauge the driver's mood and weather data to infer the appropriate mood for the music. This mood detection system aims to select songs matching the combined mood, enhancing the overall driving experience. We employ machine learning models for emotion detection from facial expressions and weather data, employing binary and multiclass classification techniques. The study contributes to personalized music recommendation systems by introducing a novel approach that considers external environmental factors, demonstrating the potential for further advancements in mood-based music recommendation systems.**

***Keywords*—Artificial Neural Network, Face Emotion Detection, Convolutional Neural Network, Image Processing, Computer Vision, AI-Integrated Music Recommendation**

I. INTRODUCTION

In today's rapidly evolving technological landscape and the incessant rhythm of modern life, the demand for seamless, personalized experiences has reached new heights. Human emotions play a pivotal role in enriching these experiences, with musical preferences intricately intertwined with individual personality traits and emotional states. Emotion, serving as a vital means of communication, manifests through various channels, including vocal inflections, body language, and facial expressions. Notably, facial expressions serve as powerful indicators of non-verbal emotions, categorizable into nuanced states such as neutrality, energy, happiness, and sadness [1].

The emotive potential of music lies in its diverse properties, wherein distinct song features can evoke varied moods. Factors such as danceability, tempo, energy, and instrumentalness exert tangible effects on human emotions, either amplifying or attenuating mood states. Danceability denotes a track's suitability for dancing, while tempo defines its pace in beats per minute [2], [3]. Energy, that is quantified on a scale between 0.0 to 1.0, gauges intensity and activity levels, with energetic tracks evoking sensations of rapidity and vigor. Furthermore, instrumentalness discerns whether a track contains vocals, thereby influencing its emotive resonance. Beyond facial expressions, environmental factors, notably weather conditions, wield considerable influence over human mood states. Sunny weather often engenders feelings of happiness and vitality, while overcast skies or rainy conditions can evoke sensations of melancholy or tranquility [4]. Thus, an intricate interplay of musical attributes and environmental stimuli contributes to the multifaceted tapestry of human emotion. In the realm of artificial intelligence, Convolutional Neural Networks (CNNs) emerge as powerful tools for classification tasks. Leveraging Conv2D layers with suitable activation functions, CNNs facilitate both binary and multi-class classification, enabling the categorization of detected faces into diverse emotional states like energetic, happy, neutral, and sad. Furthermore, these models extend their utility beyond facial expressions, demonstrating proficiency in mood detection based on weather data [5]. Thus, CNNs stand poised to revolutionize mood recognition across various domains, offering robust solutions for understanding human emotions in nuanced contexts.

The main contributions of this work are as follows: (1) we leverage Convolutional Neural Networks (CNNs) for facial emotion recognition and RandomForestClassifier (RFC) for weather-based mood detection, (2) this research pioneers an AI-driven approach to enhance driving experiences, (3) by integrating these technologies with the Spotify API, the system dynamically adjusts music playlists based on drivers' moods and weather conditions, (4) this innovative fusion not only ensures safer driving by eliminating manual song selection but also underscores the significance of considering external factors in personalized music recommendation systems and (5) finally drawing from datasets, AffectNet, FER2013, FER+, and weather data, this study highlights the potential of AI and machine learning in optimizing user experiences and sets the stage for future advancements in mood-based music selection algorithms.

This paper is structured as follows. Section II presents the similar work done in the past. In the next section we present the technique used in this work. Section IV presents our results. In Section V we conclude this work with some future directions.

II. RELATED WORK

The intersection of music recommendation, facial expressions, and live weather conditions presents a compelling problem with numerous potential applications. This section provides an overview of relevant research conducted by experts in the fields of music recommendations and face detection.

A. Personalized Facial Dataset Creation:

In the research paper [1], the authors proposed a novel approach on the development of a specialized dataset for the identification of facial expressions, utilizing the Facial Emotion Recognition using CNN (FERC) model to classify expressions across five distinct image categories. The dataset comprises 10,000 images collected from 154 individuals, ensuring a broad representation of facial expressions. To enhance the dataset's usability and address potential challenges such as variations in lighting and camera distance, a preprocessing step involving background removal is implemented. This meticulous approach to dataset creation and preprocessing aims to improve the accuracy and reliability of facial expression recognition models.

B. Facial Emotion Recognition and Music Recommendation System:

In the realm of facial emotion recognition, the research paper [2], [6] stands out for its innovative approach to emotion detection using Convolutional Neural Networks (CNNs). The study leverages the OAHEGA and FER-2013 datasets, which are rich sources of facial expression images, to train a CNN model capable of predicting six primary emotions: anger, fear, joy, neutrality, sadness, and surprise. This work is significant as it demonstrates the potential of deep learning techniques in accurately identifying and classifying human emotions based on facial expressions. The use of TensorFlow for constructing and training the CNN model, coupled with the application of the face recognition Python library, showcases the integration of advanced machine learning algorithms with practical applications in real-time facial recognition systems.

- *Proposed Solution*: The proposed solution in the research paper by Bakariya et al. [1] is a comprehensive system designed for real-time facial emotion recognition and music recommendation. This system is divided into several key components, including Face Detection, Face Emotion Prediction, Music Recommendation, and Face Recognition, with additional functionalities such as a search and removal button for previously uploaded faces. The system's architecture is designed to operate in real-time, making it suitable for a wide range of applications where immediate emotion recognition is crucial. The use of the Pygame Python package for building the music recommendation component further underscores the system's versatility and applicability in enhancing user experiences through personalized music recommendations.

- *DetectMultiScale Function*: This function is instrumental in identifying faces in images by detecting objects of various sizes and returning their coordinates. The paper further explores the application of the cv2.rectangle method to draw rectangles around detected faces, thereby mapping facial features for emotion recognition.

C. Different Networks for Mood Detection

The paper by Mahadik [3], titled "Mood based music recommendation system," utilizes numerous classification models to identify correct mood by facial expressions. The implementation of a mood-based music recommendation system presents a novel approach to real-time mood detection and personalized music selection. This system comprises two main modules: Facial Expression Recognition/Mood Detection and Music Recommendation. The Mood Detection Module is instrumental in identifying facial expressions and classifying emotions, utilizing advanced techniques such as MobileNet, a lightweight CNN architecture suitable for resource-constrained environments. The integration of diverse datasets, including the FER 2013 and MMA Facial Expression Recognition datasets, ensures robust training and validation of the emotion classification model, achieving an accuracy of approximately 75%.

Additionally, leveraging the face mesh model provided by MediaPipe [17], our research explores an alternative technique for facial expression analysis. This model facilitates the plotting of a mesh structure with boundary coordinates delineating various facial features like eyes, nose, mouth, and cheeks. By utilizing these coordinates, we conducted experiments to generate a dataset encompassing coordinates corresponding to different facial expressions, including Calm, Energetic, Happy, and Sad. This approach not only contributes to a more nuanced understanding of facial expression dynamics but also offers a cost-effective means of training machine learning models, particularly in scenarios where large image datasets might pose resource constraints.

- *Proposed Solution*: The Mood Detection Module encompasses two key components: Face Detection and Mood Detection. Face Detection involves identifying the location of faces within images, employing the FaceDetector class in Java to streamline integration with Android applications. Mood Detection further categorizes emotions detected on faces, utilizing MobileNet to classify emotions into seven distinct categories efficiently. Meanwhile, the Music Recommendation Module leverages datasets of songs classified by mood and language, stored and retrieved via Firebase, a cost-effective and easily integrable cloud storage solution. The mp3 versions of songs are manually uploaded to Firebase storage and linked in the Real-Time database, enabling seamless retrieval and recommendation based on user preferences.

- *Methodology*: The integration of the Facial Expression Recognition and Music Recommendation modules into an Android application involves several steps. Firstly, the trained MobileNet model is saved and converted into a TensorFlow Lite (.tflite) file for efficient deployment on mobile devices. This model is then integrated into the Android application alongside the Firebase backend, The application's user interface is meticulously designed to provide a seamless user experience, with intuitive controls for music playback and mood-based recommendations. Throughout the development process, rigorous testing is conducted to identify and rectify any potential bugs, ensuring the reliability and performance of the final product. Additionally, user feedback is solicited and incorporated to enhance usability and satisfaction.

III. METHODOLOGY

Presented by this research, are the detailed steps involved in creating a comprehensive end-to-end machine learning web application for detecting mood from human faces and real-time weather and selecting a song based on the detected mood. This includes preprocessing image data, designing and training convolutional neural networks for mood detection, integrating weather APIs for real-time weather data retrieval, and implementing algorithms for mood-based song recommendation, all seamlessly orchestrated into a user-friendly web interface:

A. Mood Detection based on Facial Expressions

This section discusses the process of detecting mood from human faces.

1) Dataset Collection and Preprocessing: The dataset collection for this research encompasses two distinct methodologies, each contributing to the creation of a comprehensive and diversified dataset for facial expression recognition. The first step involves the acquisition of existing datasets of human facial expressions, categorized into four distinct emotional states: energetic, happy, calm, and sad. These datasets, sourced from open-sourced repositories, AffectNet, FER2013, and FER+ [5], are instrumental in providing a broad spectrum of facial expressions for model training [3], [4]. The diversity of these datasets ensures that the model is exposed to a wide range of expressions, enhancing its robustness and generalization capability. Furthermore, careful preprocessing techniques are applied to standardize and optimize the data for effective model learning.

The second methodology involves the creation of a personalized dataset through the manual download of images from Google, utilizing a Chrome extension designed for bulk image downloading. This method not only expands the dataset's size but also introduces a level of personalization, catering to specific research needs or application scenarios. The dataset is divided into training and testing subsets, totaling 30,000 images. Fig. 1 shows a sample of images of different human faces representing different emotions, demonstrating the richness and variability of the dataset collected. This personalized approach allows for a finer granularity in the training data potentially improving the model's performance in recognizing subtle variations in facial expressions. Additionally, data augmentation techniques such as rotation, scaling, and flipping are applied to further enrich the dataset and enhance model robustness.

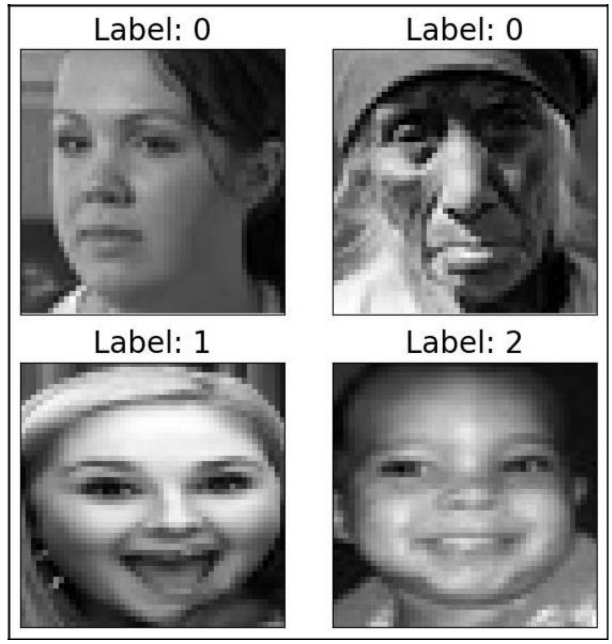
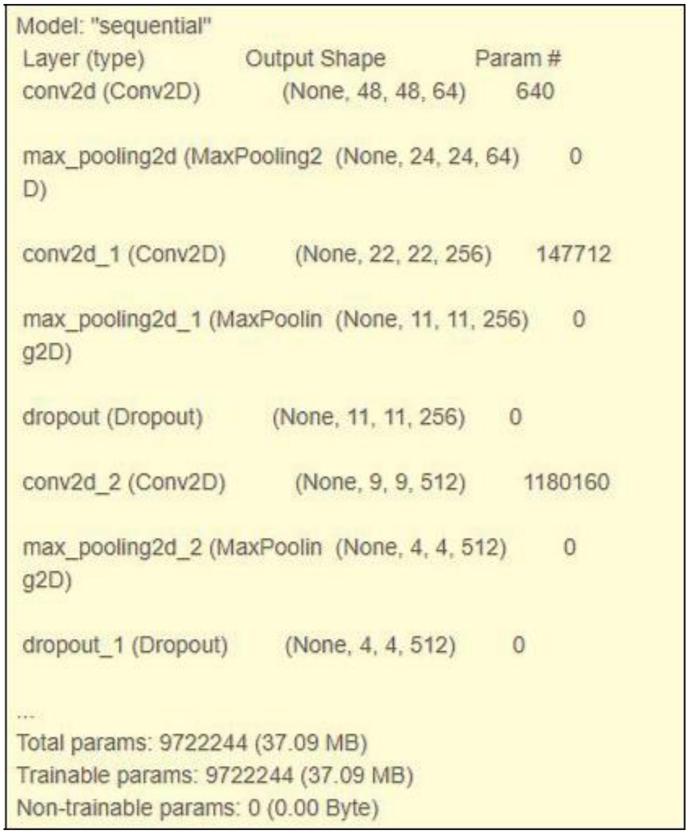


Fig. 1. Sample of different human faces.

In the preprocessing phase, images are converted to grayscale and resized to 48x48 pixels for efficient model training. Additionally, a rescaling factor of 1/255 is applied to normalize pixel values between 0 and 1, enhancing neural network convergence and stability during training. This normalization ensures consistent feature scaling, aligning input data with activation functions for effective learning and faster convergence. Moreover, techniques such as histogram equalization can be employed to further enhance image contrast and improve model performance.

2) CNN Model Architecture for mood detection by face: The mood_detection_model presents a convolutional neural network architecture specifically tailored for mood detection from facial expressions. It comprises several convolutional layers with increasing filter sizes, followed by max-pooling layers for downsampling to capture essential features effectively. Dropout layers are strategically integrated to mitigate overfitting, ensuring better generalization. Furthermore, the model is concluded with fully connected layers for classification, providing a comprehensive understanding of the input data. The final layer employs SoftMax activation, yielding probabilities for four distinct mood categories: energetic, happy, calm, and sad. This architecture is optimized for accurate mood classification while maintaining computational efficiency, suitable for real-time applications on resource-constrained devices.



Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 48, 48, 64)	640
max_pooling2d (MaxPooling2D)	(None, 24, 24, 64)	0
conv2d_1 (Conv2D)	(None, 22, 22, 256)	147712
max_pooling2d_1 (MaxPooling2D)	(None, 11, 11, 256)	0
dropout (Dropout)	(None, 11, 11, 256)	0
conv2d_2 (Conv2D)	(None, 9, 9, 512)	1180160
max_pooling2d_2 (MaxPooling2D)	(None, 4, 4, 512)	0
dropout_1 (Dropout)	(None, 4, 4, 512)	0

...
Total params: 9722244 (37.09 MB)
Trainable params: 9722244 (37.09 MB)
Non-trainable params: 0 (0.00 Byte)

Fig. 2. CNN architecture for mood detection by face.

3) Model Working using OpenCV Real-Time Face Detection:

- *Face Recognition:* In the realm of human face detection, various methodologies are employed, including the Haar Cascade Classifier, Histogram of Oriented Gradients (HOG), Feature-Based Methods, and Eigenfaces. For the purpose of this project, the Haar Cascade Classifier stands out due to its superior speed compared to other algorithms. This classifier operates through a series of boosted classifiers to identify objects, showcasing its high accuracy [3]. The underlying principle of this technique is rooted in the edge or line detection methodology introduced by Viola and Jones [16] in their 2001 paper titled "Rapid Object Detection Using a Boosted Cascade of Simple Features." The 'detectMultiScale' function is instrumental in this process, as it identifies objects of varying sizes within the input image, returning a list of rectangles that represent the detected items, with their coordinates specified as (x, y, w, h) [5]. Fig. 3 [18] shows how the Haar Cascade detects the face by extracting features from the image. The versatility of the 'detectMultiScale' function lies in its ability to adapt to diverse image compositions, scaling efficiently across different resolutions and aspect ratios, thus enhancing its applicability in real-world scenarios. Additionally, its robustness against variations in lighting conditions and facial orientations further solidifies its effectiveness in practical implementations.

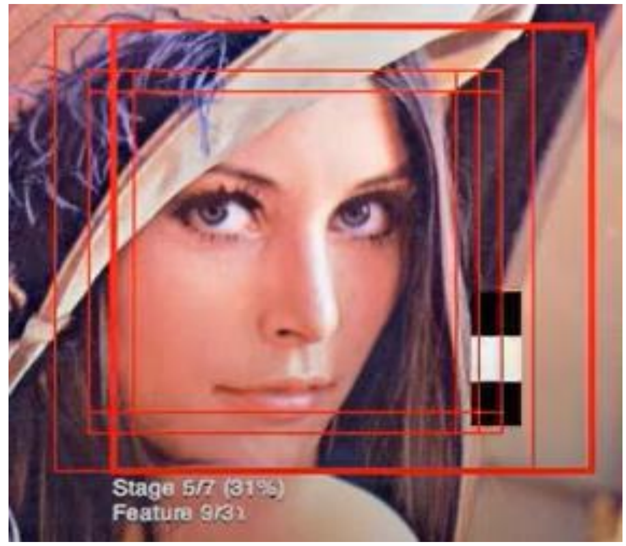


Fig.3. Haar cascade extracting features from the image.

The flowchart in Fig. 4, depicts a face detection system that utilizes OpenCV to capture a user's face via webcam. A machine learning model is then loaded to analyze the user's mood based on their facial expressions within the captured frame. If no face is detected, the system restarts the process to continuously read frames until a face is found. Once a face is identified, the system predicts the user's mood using the loaded model.

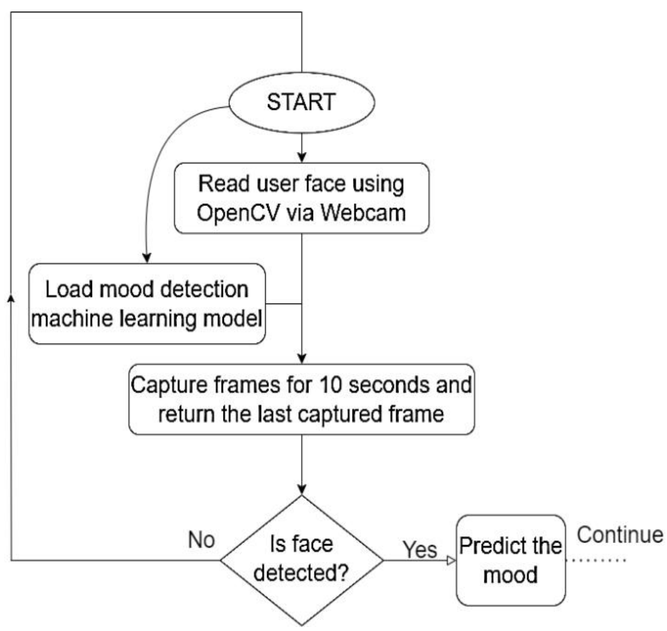


Fig.4. Flowchart to detect mood from human face

4) Model Working using MediaPipe Real-Time Face Detection:

Apart from training a custom CNN model, we explored the option of using MediaPipe FaceMesh for facial expression detection [21]. MediaPipe FaceMesh leverages a pre-trained model to map 468 facial landmarks in real-time, making it a compelling alternative for expression analysis [22]. This approach offers significant advantages over custom CNNs, particularly in terms of computational efficiency and ease of deployment. While the custom CNN achieved a commendable accuracy of 94% on our dataset, MediaPipe outperformed it with a 97% accuracy in detecting facial expressions. The pre-trained nature of MediaPipe eliminates the need for extensive training data, making it faster and more resource-efficient. Our results indicate that MediaPipe not only simplifies the process of facial expression detection but also enhances accuracy, proving to be a robust solution for real-time applications. This comparison highlights the effectiveness of MediaPipe as a viable alternative to traditional deep learning models for face detection tasks.

B. Mood Detection based on Weather Conditions:

Weather conditions exert a profound influence on the human mood, shaping our emotional states in subtle yet significant ways. Sunny days often elevate spirits and energize individuals, while overcast skies and rainy nights can induce feelings of sadness and lethargy. Factors such as cloud cover, precipitation levels, and the distinction between day and night play crucial roles in this dynamic [8]. These environmental cues interact with individual physiological responses to weather, contributing to the complex interplay between weather and mood. Below steps discusses how the mood can be determined by the current weather conditions.

1) *Dataset Creation:* To construct a dataset for training a model capable of predicting mood based on weather conditions, this research incorporates three primary features: 'is_day' (binary, 0 for night, 1 for day), 'cloud_cover' (percentage, 0-100%), and 'precipitation' (amount, 0-2). These features are pivotal as they significantly influence human mood. The 'is_day' feature distinguishes between day and night, 'cloud_cover' quantifies the extent of cloud coverage, and 'precipitation' measures the amount of rain or snow [8], [9]. By analyzing thousands of variations of these three features, a mood is associated as the target value, thereby creating a comprehensive dataset for mood prediction based on weather conditions.

To facilitate the machine learning model's interpretation of mood data, the target values, which represent moods in the dataset, undergo a transformation process to convert them into numerical form [10]. This conversion is crucial as most machine learning algorithms operate on numerical data, necessitating the translation of categorical mood labels into a numerical format that the algorithms can interpret. Specifically, the moods are encoded numerically as follows: 0 for energetic, 1 for calm, 2 for happy, and 3 for sad. This numerical encoding facilitates the model's ability to learn and predict moods based on the numerical representation of moods, thereby enhancing the model's predictive capabilities and accuracy.

	is_day	cloud_cover	precipitation	mood
4280	1	4.572192	1.357517	Sad
151	1	0.084067	0.014964	Energetic
2896	1	25.255036	0.231827	Calm
765	0	1.369262	0.029750	Energetic
4115	0	55.633292	0.253274	Sad
3854	1	61.978400	0.259238	Sad
2724	1	74.194989	0.131484	Calm
1091	0	23.355546	0.077193	Energetic
980	1	10.930933	0.106426	Energetic
2976	0	23.400318	0.283350	Calm

Fig.5. Dataframe of weather features with target mood

2) *RandomForestClassifier (RFC) for mood detection by weather:*

The RFC is a machine learning model renowned for its efficacy in classification tasks. Unlike single decision trees, it constructs a multitude of decision trees during training, utilizing a technique known as ensemble learning [15]. Through this process, the classifier aggregates the predictions of these individual trees, ultimately outputting the class that is the mode of the classes predicted by the constituent trees. This approach imbues the RandomForestClassifier with resilience against overfitting and the capacity to discern complex relationships within high-dimensional datasets.

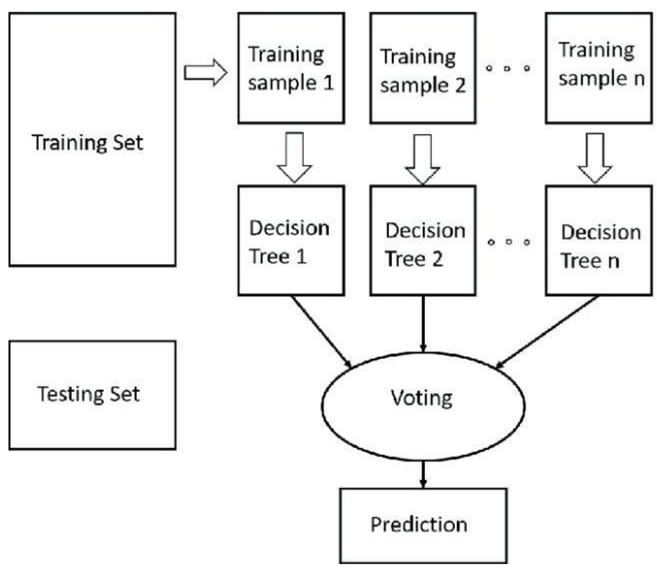


Fig.6. RFC model to predict mood based on weather features.

3) *Model Working on Current Weather Conditions:* The process of predicting mood based on weather conditions begins with initializing a Random Forest Classifier (RFC) model, a versatile machine learning technique renowned for its accuracy. Following this, the system utilizes the Open-Meteo weather API to gather an array of current weather features. These encompass not only basic meteorological data such as

'is_day' (a binary indicator distinguishing between night and day), 'cloud_cover' (expressed as a percentage ranging from 0 to 100%), and 'precipitation' (measured in millimeters), but also more nuanced variables like 'wind_speed' and 'temperature'. Once these weather data points are collected, they are meticulously processed and transmitted to the RFC model, which has been trained to classify mood among four distinct categories: energetic, happy, calm, and sad. The RFC model, an ensemble learning method comprising multiple decision trees, evaluates the input data to make its predictions. What sets this system apart is its adaptability; as users' mood preferences evolve, the model continuously refines its understanding, ensuring that recommendations remain personalized and relevant. By leveraging the power of machine learning algorithms, this system continually enhances its proficiency in discerning the subtle interplay between mood and weather patterns, thus refining the accuracy of its recommendations with each interaction.

Fig. 7 shows the complete process of determining the mood based on the weather data extracted from the Open-Meteo weather API using the random forest classifier model.

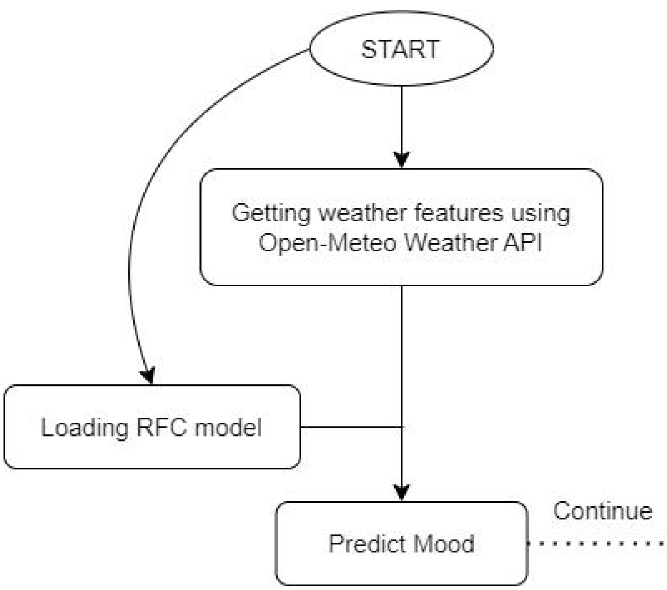


Fig.7. Workflow to predict mood from weather data.

C. Integration of Models with Spotify API for Song Recommendation:

A detailed step-by-step flow is illustrated in Fig. 8, depicting a meticulously orchestrated sequence of steps aimed at seamlessly integrating multiple components: the predictive model, mood detection algorithms based on facial expressions and weather conditions, and the Spotify Developer API, to facilitate the precise selection of songs corresponding to detected moods. Initially, the predictive model receives inputs representing the predicted mood from both the facial expression analysis module and the weather-based mood detection module. These inputs, derived from sophisticated algorithms trained on diverse datasets, provide nuanced insights into the emotional states of users. Subsequently, leveraging the extensive capabilities of the Spotify Developer API, the model interfaces with the platform to identify song genres that best align with the detected moods. This interaction involves querying the API for genre recommendations tailored to the specific mood profiles, ensuring a personalized and contextually relevant song selection process. Upon receiving the API response containing genre recommendations, the model employs a randomized selection mechanism to choose a song from the recommended genre. This randomization strategy adds an element of spontaneity and diversity to the song selection process, enhancing user engagement and satisfaction. Finally, the chosen song is seamlessly integrated into the application, ready to be played and enjoyed by users, thereby encapsulating a sophisticated fusion of predictive modeling, data-driven insights, and API-driven functionality within the realm of mood-based song recommendation systems.

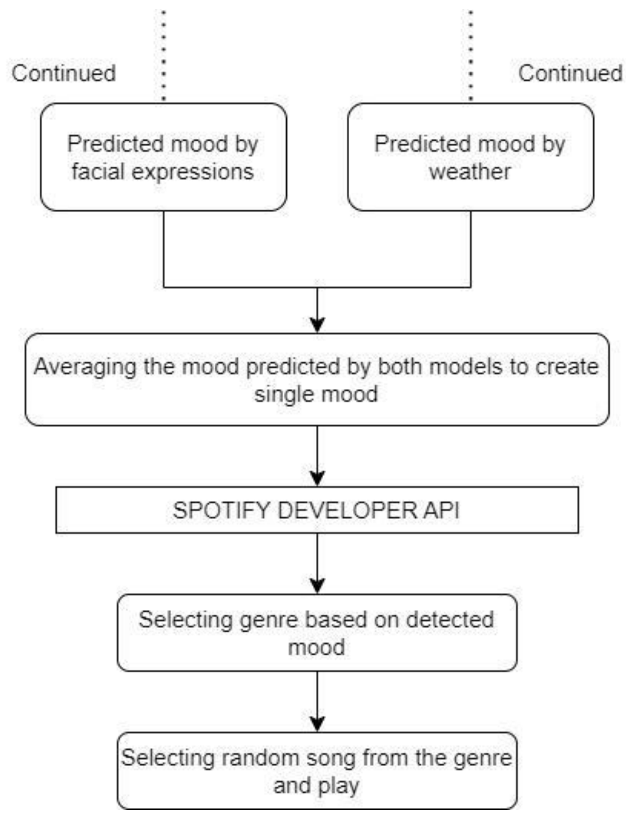


Fig.8. Steps to integrate models to select song using Spotify Developer API

IV. RESULTS

A. Results from real-time detection:

Upon the activation of the "Detect Mood" button on the front end, the OpenCV library initiates the webcam to capture frames for the purpose of facial expression analysis. This process involves capturing a continuous stream of frames for a duration of 10 seconds, with the final frame being selected for analysis. This frame is then processed through a Convolutional Neural Network (CNN) model, which specializes in image classification tasks. The CNN model analyzes the facial expression within the image and outputs a mood prediction based on the detected features [11].

Concurrently, the system also determines the user's current location and retrieves weather data features utilizing the Open-Meteo weather API. This data, encompassing various atmospheric conditions, is then fed into a Random Forest Classifier (RFC) [3]. The RFC, a machine learning algorithm. With the mood predictions obtained from both the facial expression analysis and the weather data, the system proceeds to select a song genre that aligns with the overall mood. This decision-making process is facilitated by the Spotify Developer API, which provides access to a vast database of music genres and playlists. Based on the mood predictions, the API identifies a genre that is most likely to match the user's current emotional state. Subsequently, a random song is selected from the identified genre and is streamed to the application for playback. This selection process ensures that the song chosen not only matches the genre but also aligns with the mood detected by the system.

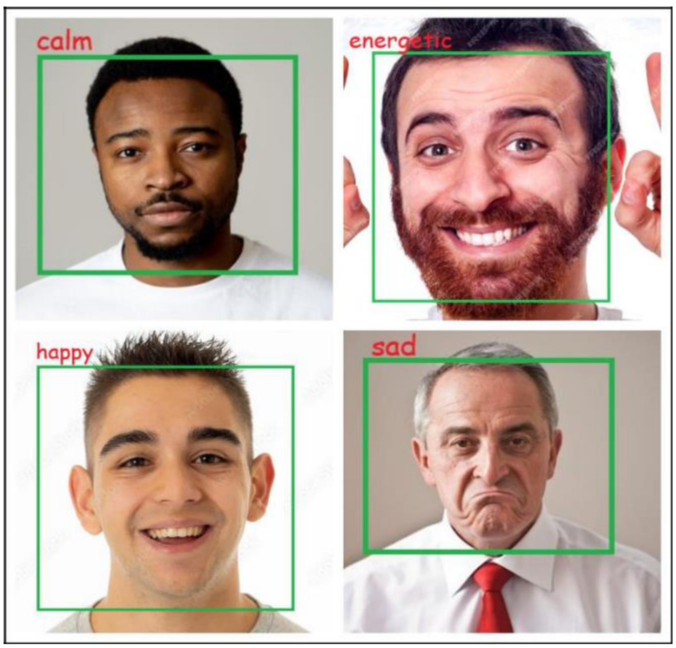


Fig.9. Emotion detected by face using CNN model and OpenCV

As illustrated in Fig. 9, the mood detected from the last frame captured by the webcam is identified as "happy" by the CNN model. In conjunction with the mood inferred from the weather data, which also indicates a happy mood, the system selects the 'Power-Pop' genre. This genre, a subcategory of rock music, is characterized by its high tempo and danceable nature.

Fig. 10 illustrates the process where the selected song, determined through the integration of machine learning models and the Spotify API, is redirected to the webpage for playback utilizing the Spotify player [12], [13]. This figure provides a visual representation of the final step in the recommendation system, showcasing how the system's output—a song chosen based on the user's detected mood and current weather conditions—is seamlessly integrated into the user's music listening experience on Spotify. Once the detect mood button is clicked, the backend triggers the webcam using the OpenCV framework in Python. The facial expressions are detected for 20 seconds, and the last frame captured is sent to the machine-learning model for mood detection. Meanwhile, the Open-Meteo weather API also gets triggered, and the weather details captured are sent to the mood detection model. Once the mood gets detected by both the models, an average mood is calculated and is sent to the Spotify API to get the song details like song name, song artist, and song genre. These are retrieved in JSON format to be displayed on the UI. Additionally, the system records user preferences over time, refining its recommendations to better suit individual tastes and moods, enhancing the overall user experience. This continuous learning process ensures that the system evolves dynamically, staying relevant and responsive to the user's changing preferences and moods. Through these mechanisms, the application not only provides immediate song recommendations based on mood and weather but also adapts over time to offer increasingly personalized suggestions, fostering long-term user engagement and satisfaction.

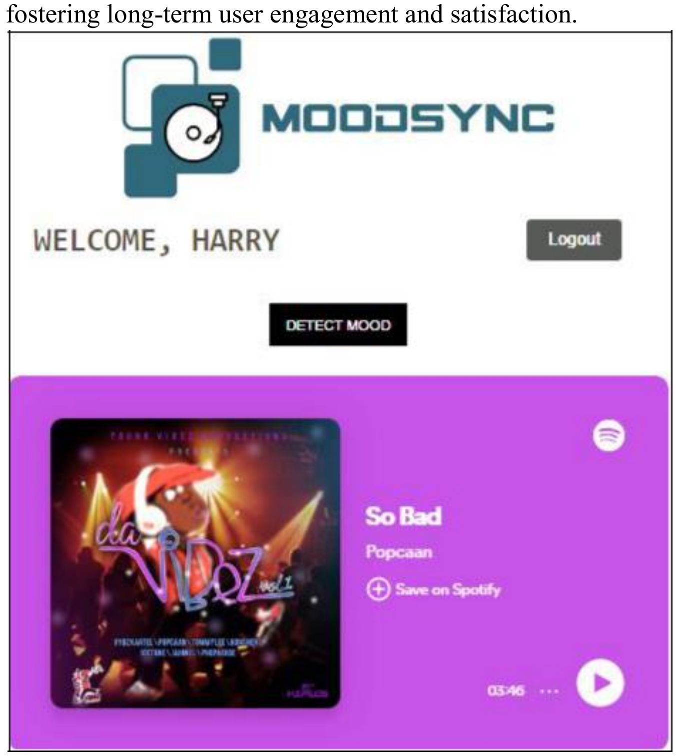


Fig.10. Song played from the power-pop genre due to happy mood.

Fig. 11 presents the outcomes achieved by the combined application of machine learning models and the Spotify API. This Fig. encapsulates the results of the system's mood detection and genre selection processes, demonstrating the effectiveness of the integrated approach in accurately predicting the user's musical preferences.

Fig.11. Mood detected by both models and genre selected based on mood.

B. Accuracy Comparison of pre-trained ML models vs Custom ML model:

The custom CNN model outperforms both VGG16 and VGG19 in mood detection from facial expressions on a dataset of 30,000 images, achieving an accuracy of 94% compared to VGG16's 76% and VGG19's 79% [19]. The table in Fig. 12 shows the evaluation metrics of the custom CNN model. While VGG16 and VGG19 have 16 and 19 layers respectively, with large model sizes of 138 million and 144 million parameters, the custom CNN is more compact and faster, making it more efficient in both memory usage and computation. The custom model's high accuracy and tailored design make it ideal for this task, although it requires careful design and lacks transfer learning benefits. In contrast, VGG16 and VGG19, though robust and benefiting from pre-training, are computationally expensive and less effective for this specific task.

	Precision	Recall	F1-Score	Support
1	1.00	0.93	0.96	100
2	1.00	0.98	0.99	100
3	1.00	0.85	0.92	100
4	0.81	1.00	0.89	100
Accuracy			0.94	400
Macro Avg.	0.95	0.94	0.94	400
Weighted Avg.	0.95	0.94	0.94	400

Fig.12. Custom CNN model evaluation metrics.

VGG16 and VGG19, despite their robust architectures and success in general image classification tasks, did not perform as well in mood detection from facial expressions due to their large size and complexity, which may lead to overfitting on a relatively smaller, specialized dataset like the 30,000 facial images used here [20]. Their deep architectures with 16 and 19 layers require extensive computational resources and may struggle to capture the nuanced features specific to facial expressions, especially without significant fine-tuning. In contrast, the custom CNN model, designed specifically for this task, is likely optimized to extract the most relevant features for mood detection, allowing it to achieve a much higher accuracy of 94%. The custom model's more efficient and targeted architecture makes it better suited for the specific characteristics of the dataset, leading to its superior performance.

V. CONCLUSION AND FUTURE WORK

In this work, we conducted experiments involving the development of image classification models by extending the training of established pretrained models (VGG16, VGG19, and MediaPipe), specifically tailored for mood detection from drivers' facial expressions. We enriched the training process by introducing supplementary image datasets, aimed at enhancing the quality and accuracy of the trained models. We quantified performance using a wide range of metrics: F1 Score, Precision, and Recall, providing a thorough assessment of their ability to categorize emotions into four distinct categories: Calm, Energetic, Happy, and Sad. Additionally, we developed models to infer mood from current weather, broadening the scope of mood detection algorithms.

However, the study has certain limitations. The reliance on pre-existing datasets may introduce biases that could limit the generalization of the models across diverse populations and driving conditions. The performance of the models is also influenced by the quality and diversity of the supplementary image datasets used during training. Furthermore, while the integration of weather data expands the capability of mood detection, the models may not fully capture the complex interplay between various environmental and personal factors affecting mood.

Despite these limitations, the research suggests promising directions for future work. Augmenting the current system with features such as personal voice assistance, implementing a feedback loop to iteratively enhance the performance of the machine learning models, and developing a more interactive user interface could further elevate the overall user experience. As technology evolves and datasets expand, the potential for creating even more sophisticated and personalized systems tailored to individual preferences and contexts continues to grow, promising an exciting future for AI-driven music recommendation systems in various domains beyond driving.

In the future, we aim to address the current limitations by incorporating more diverse and representative datasets, adding a user feedback module to further enhance ML model performance, and integrating voice assistance for seamless interaction and manual control over music selection. These enhancements will contribute to a more robust, user-centered system.

REFERENCES

[1] 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.

[2] Athavle, M. (2021). Music Recommendation Based on Face Emotion Recognition. Journal of Informatics Electrical and Electronics Engineering (JIEEE), 2, 1-11.

[3] Mahadik, A., Milgir, S., Patel, J., Kavathekar, V., & Jagan, V. B. (2021). Mood based music recommendation system.

[4] Bundo, M., Preisig, M., Merikangas, K., et al. (2023). How ambient temperature affects mood: an ecological momentary assessment study in Switzerland.

[5] Dhillon, A., & Verma, G. K. (2020). Convolutional neural network: a review of models, methodologies and applications to object detection.

[6] Mehendale, N. (2020). Facial emotion recognition using convolutional neural networks (FERC).

[7] Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep Face Recognition.

[8] Barbosa Escobar, F., Velasco, C., Motoki, K., Byrne, D. V., & Wang, Q. J. (2021). The temperature of emotions.

[9] Behnke, M., Overbye, H., Pietruch, M., & Kaczmarek, L. D. (2021). How seasons, weather, and part of day influence baseline affective valence in laboratory research participants?

[10] Dimolitsas, I., Kantarelis, S., & Fouka, A. (2023). SpotHitPy: A Study For ML-Based Song Hit Prediction Using Spotify.

[11] Thornton, A., Aliyeva, E., & Pande, T. (2021).

[12] Bhowmick, A., Shamkuwar, K., & Jayaseeli, J. D. (2022). Song Recommendation System based on Mood Detection.

[13] Medium Article - Music Genre Prediction.

[14] Apao, N. J., Feliscuzo, L. S., Sta. Romana, C. L. C., & Tagaro, J. A. S. (2020). Multiclass Classification Using Random Forest Algorithm.

[15] Apao, N. J., Feliscuzo, L. S., Sta. Romana, C. L. C., & Tagaro, J. A. S. (2020). Multiclass Classification Using Random Forest Algorithm.

[16] Viola, P., & Jones, M. (2001, December). Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001 (Vol. 1, pp. I-I). Ieee.

[17] B. Thaman, T. Cao and N. Caporusso, "Face Mask Detection using MediaPipe Facemesh," 2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, 2022, pp. 378-382, doi: 10.23919/MIPRO55190.2022.9803531.

[18] Shawky, Elham & El-Khoribi, Reda & Shoman, Mahmoud. (2014). Audio-Visual Speech Recognition for People with Speech Disorders. International Journal of Computer Applications. 96. 975-8887. 10.5120/16770-6337.

[19] A. S. Negi, A. Arora, S. Bisht, S. Devliyal, B. V. Kumar and G. Kaur, "Facial Emotion Detection using CNN & VGG16 Model," 2024 IEEE 9th International Conference for Convergence in Technology (I2CT), Pune, India, 2024, pp. 1-6, doi: 10.1109/I2CT61223.2024.10543618.

[20] Vignesh, S., Savithadevi, M., Sridevi, M. et al. A novel facial emotion recognition model using segmentation VGG-19 architecture. Int. j. inf. tecnol. 15, 1777–1787 (2023).

[21] Thaman, B., Cao, T., & Caporusso, N. (2022, May). Face mask detection using mediapipe facemesh. In 2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO) (pp. 378-382). IEEE.

[22] Al-Nuimi, A. M., & Mohammed, G. J. (2021, August). Face direction estimation based on mediapipe landmarks. In *2021 7th International Conference on Contemporary Information Technology and Mathematics (ICCITM)* (pp. 185-190). IEEE.