International Journal On Engineering Technology and Sciences – IJETS
ISSN (P):2349-3968, ISSN (O): 2349-3976 Volume XI - Issue III, March - 2024

Music Player Based On Real Time EmotionDetection Using CNN

[1]Dr Nandha Gopal S M, [2]Samreen Maryam, [3]Sharanya M, [4]Syeda Ayesha Ruzaina, [5]Vineet Singh,
[1]Professor, [2,3,4,5]UG Student [1,2,3,4,5]Department of Computer Science and Engineering,
HKBK College Of Engineering, Bangalore, India

Abstract—A person's music preferences are not only dependenton their current emotional state but also by their historical musical choices. This research presents a robust music recommendation system that proposes music aligned with the user's present mood. Recognizing the profound impact of harmonious soundson emotions, the system goes a step further by dynamically tailoring playlist creation to the listener's experience, leveraging real-time emotional cues. The incorporated CNN (Convolutional Neural Network) model adeptly examines visual and audio content from radio waves, discerning emotions such as pleasure, sadness, enthusiasm, and tranquility. Users actively contribute to refining the system through feedback, fostering an interactive and evolving emotional connection between the listener and theircurated collection. This endeavor marks a significant advancement in cultivating a more empathetic and immersive playback experience, strengthening the emotional bond between people and their musical selections. The music classification engine achieves noteworthy accuracy by employing audio elements for precise categorization, collectively enhancing the user's musical journey through the seamless integration of mood-based recommendations and accurate music classification.

Index Terms—CNN model, Mobile net architecture, Convolutional Neural Network

I. INTRODUCTION

The integration of human emotion detection finds application in various domains that demand heightened security or detailed insights into an individual. Acting as a supplementary phase to facial recognition, it becomes imperative in scenarios necessitating an additional layer of security. In this context, beyond identifying the face, the system also discerns the emotional state. This functionality proves valuable in ensuring that the individual in the camera's view is not merely a flat, two-dimensional representation. .Classifying human emotions involves categorizing them into distinct states. These emotional states are characterized by subtle nuances, making their identification a complex task. Facial muscle contortions associated with these emotions are often minimal, posing a significant challenge in detecting the subtle differences. The slightest of the variations can result to distinct expressions, adding to the intricacy of the process. In the contemporary era marked by continuous advancements in multimedia and tech- nology, a plethora of music players have emerged, equipped with features like fast forward, reverberation, variable playing speed, local and multicast playback, volume transitions, and type separation. Despite meeting fundamental user require- ments, these players still necessitate manual song selection based on the user's current disposition. Users often grapple with the task of browsing playlists to align with their emotions and feelings.Recognizing this need, the application integrates the Haar cascade classifier for face detection and facial feature extraction. Facial expressions serve as reliable indicators of an individual's mental state, providing a natural means of expressing emotions. Given the strong connection between music and emotions, the algorithm is designed to enable users to navigate playlists effortlessly based on their feelings. The algorithm aims for optimal performance by minimizing

memory usage, reducing processing time, and eliminating the need for costly EEG hardware or sensors. The advancement of digital music technology underscores the importance of developing a customized music recommendation system to cater to individual user preferences. Navigating the extensive information online offers an opportunity to provide meaningful recommendations poses a considerable challenge. E-commerce giants such as Amazon and eBay leverage users' tastes and histories to offer personalized suggestions, while companies like Spotify and Pandora employ Machine Learning and Deep Learning techniques for more refined recommendations. The realm of personalized music recommendation has seen ef- forts to tailor suggestions based on user preferences. Two primary approaches have emerged in this domain. The firstis the content-based filtering approach, which delves into the analysis of the content of music that users have enjoyed in the past. Recommendations are then made by identifying music with similar or relevant content. The surge in multimedia content accessibility has made music readily available for daily consumption. However, individual preferences for music genres vary significantly based on emotions or moods. To enhance user convenience, technologies are emerging that pro- vide music recommendations tailored to specific atmospheres.With the advent and commercialization of A.I. speakers can now select music based on their moods. This developmentis expected to drive the expansion of music-related services. Furthermore, businesses across diverse industries can

leverage this technology to target customers more effectively.

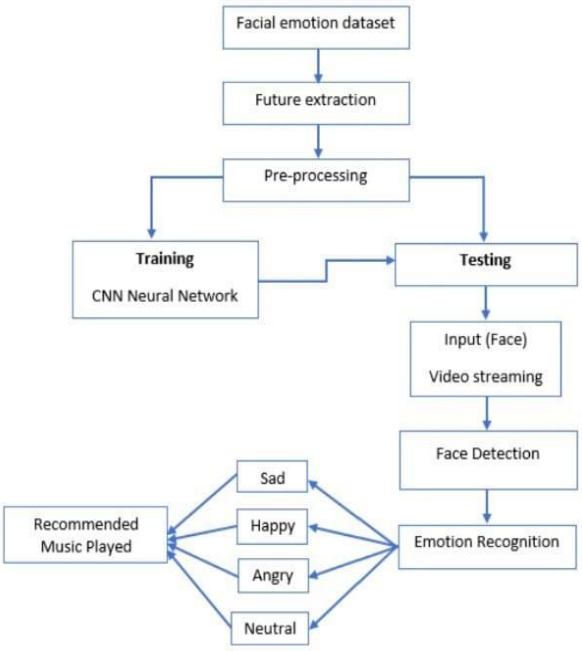


Fig 1 : Flowchart

Emotions in music can be represented as continuous quanti- ties mapped into real numbers. Several models, such as the circumflex model and Thayer's model, have been proposed to describe emotions in music based on dimensions like arousal and valence. In conclusion, AI has made significant stridesin various domains, including face recognition and music analysis. The integration of facial emotion recognition with music recommendation systems shows promise in enhancing user experiences and enabling more sophisticated applications.

II. RELATED WORKS

The Research paper titled "Emotion Detection and Characterization using Facial Features" explains that detecting the face in the input image is the starting phase in the process,which is then followed by identifying facial features like the nose and lips etc.. These features are then subjected to specific filters and transformations. The resulting outputs are fed into classifiers that utilize trained data to classify the emotions. The Fischerface classifier employs LDA and PCA techniques, which contribute to its accuracy. PCA is used to reduce dimensionality, minimizing the number of variables in the dataset. Gabor filters, on the other hand, excel at recognizing texture in an image, including differentiating the mouth and eyes from the surrounding skin. This texture- based approach is advantageous for effective feature extraction [1]. The research paper titled "Facial Emotion Detection Using Deep Learning" presents a convolutional neural network (CNN) that makes use of the "Keras" deep learning toolkit to recognise face emotions. There are two steps involved: feature extraction and categorization The study employs two datasets containing images with facial expressions representing seven emotions and trains the proposed network. Evaluation includes validation accuracy and loss. The method achieves reduced computation time, increased validation accuracy, and decreased loss compared to existing models. Performance evaluations are conducted on the JAFFE and FERC-2013 datasets, which encompass the seven main emotional domains [2]. The research paper titled "Emotion Recognition Basedon Facial Expressions Using Convolutional Neural Network (CNN)" discusses issues with emotion detection datasets and investigates CNN architectures and parameters for identifying seven different human facial emotions. It utilizes the iCVMEFED dataset for its novelty and difficulty. With a total of 899,718 parameters, the original CNN design consists of 4 Convolutional Layers, 2 Fully Connected Layers ,one Dropout Layer and 4 Max Pooling Layers . The model processes images through Conv2D filters with ReLu activation, followed by MaxPooling2D for dimension reduction, and ends with Flatten and Dropout layers [3]. The research paper titled "Automatic Human Emotion Recognition System using Facial Expressions with Convolution Neural Network" presents an automated face emotion categorization system that uses SURF for extracting features and Convolutional Neural Networks (CNN) for feature extraction. It achieves a 91 percent ac- curacy in emotion recognition from facial expressions, fa-cilitating effective emotion tracking. High boost filtering is used to reduce image noise while preserving low-frequency components. The SURF feature extraction process involves computing the integral image, determining the Hessian Ma- trix, normalizing the determinant response, applying Non- Maximal Suppression and thresholding, computing interestpoints, and generating Interest Point Descriptors [4]. The research paper titled " Emotion Detection using Deep Facial Features" explores Face Emotion Recognition (FER) through pre-processing, feature extraction, and classification phases. Using Deep Facial Features via Transfer Learning using pre- trained models like VGG16, ResNet152V2, InceptionV3, and Xception, it analyses various deep learning architectures in

Keras for emotion detection. Bottleneck features are created, and a dataset containing CK+ and JAFFE is used to assess model performance. Pre-processing steps include image transparency, scaling, contrast correction, and enhancement to improve emotion frame quality [5]. The research paper titled ”The Application of Artificial Intelligence in Emotion Recognition” explains how artificial intelligence is being used to recognise emotions, emphasizing facial recognition as an automated identification technology utilizing facial features. It highlights the utilization of digital image processing, video processing, pattern recognition, and related technologies. Algorithms for machine learning and deep learning, especially ANNs, mimic neural functions to develop intricate networks for automatic face recognition systems. The paper is to give a general overview of emotion recognition and the ways in which it is used in different fields and insights into future directions in emotion recognition technology, serving as a reference for researchers in the domain [6]. The research paper titled ”Facial Emotion Recognition Using Deep Convolutional Neural Network” suggests Deep learning methods have completely changed the field of face emotion recognition. The goal of this work was to create a DCNN model that could correctly categorize five distinct facial emotions. In order to prevent overfitting, the model employs two convolution layers and dropout layers. The input picture is first downsized to 32 x 32 and processed via these layers before ReLU activation and pooling. The neural network’s dense layers receive the flattened 2-dimensional array of feature values produced by the second convolution layer. The output layer has five units, each representing a different facial emotion, and uses the softmax activation function to produce a probabilistic output. The model achieved an accuracy rate of 90.04 percent in classifying the five facial emotions and was evaluated using test samples. Although it struggled with Class 1 (happy), it demonstrated excellent prediction results for Classes 0 (angry) and 3 (neutral). The model has several advantages, including accurate classification of five different facial emotions, well- fit and generalized training and validation accuracy, use of the Adam optimizer to improve performance, and potential for real-time applications like feedback analysis. For efficient control, the system can potentially be integrated with other electronic equipment. However, applications requiring higher accuracy rates may find the current accuracy of 78.04 percent insufficient [7].The paper ”Facial Emotion Detection using Action Units” presents a system that is aimed at identify- ing the following seven emotions: surprise, wrath, disgust, fear, sadness, contempt, and happiness. It employs an al- gorithm tested on various databases, dividing images into matrices, adjusting grayscale and scale, and applying facial feature descriptions using an AAM model and Gabor wave conversion. The process involves facial recognition, feature extraction, and grading to determine expressions. It starts with image capture, preprocessing, and uses a database for training. The system emphasizes high-speed analysis, real-time changes in expressions, and employs Bayesian Network Classifier for modeling. Visualization techniques focus on key areas like the mouth and eyebrows. The method involves image operations, face detection, lip region segmentation, emotion template comparison to identify emotions.[8].The paper ”Emotion Detection of Contextual Text using Deep Learning” addresses the challenge of extracting emotions from social media textual data. The proposed system, named Aimens, utilizes the LSTM model with word2vec and doc2vec embeddings to detect emotions such as sad, angry, and happy, achieving an F-score of 0.7185. The model undergoes fine- tuning with hyperparameters and utilizes Bi-LSTM for im- proved performance. Enhanced preprocessing steps include resolving character encoding and correcting spelling. An an- notated corpus aids in labeling emotions, and word embed- ding techniques are applied for dataset mapping. Despite the advantages of pre-trained embeddings and hyperparameter tuning, there’s a reduction in accuracy when employing a large emotion-based dataset. Additionally, limitations include reduced accuracy with stop word removal and applicability restricted to 3-turn conversations. Future work may involve integrating emotion lexicons and emoticon handling for enhanced performance.[9].The paper ”Speech Interactive Emo- tion Recognition System Based on Random Forest” aims to develop a system using Random Forest Classifier and Support Vector Machines for speech emotion recognition. It utilizes the Berlin Emotional Database, explaining preprocessing steps to reduce noise and enhance signal clarity. Features are extracted using OpenSMILE software, employing statistical functions on acoustic features. The Random Forest Classifier achieves an 89 percent recognition accuracy. The system collects speech signals through a user-friendly interface, processes them, and returns emotional categories. Using WeChat for data collection enhances system usability. The paper contributes to speech emotion recognition advancement, suggesting potential future improvements in the field.[10].The paper

”Speech Emotion Recognition using Convolution Neural Networks and Deep Stride Convolutional Neural Networks” investigates the use of deep learning methods in Speech Emotion Recognition (SER) improvements. In order to emphasise computational economy without sacrificing accuracy, a modified DSCNN architecture is introduced. After being trained and assessed on spectrograms from the SAVEE dataset containing neutral, happy, sad, and angry emotions, the DSCNN model performs better than the CNN model, with an accuracy of 87.8 percent as opposed to 79.4 percent. Spectrograms serve as rich inputs, enhancing emotion recognition by extracting discriminative features over time and frequencies. The DSCNN’s unique strides for dimension reduction improve efficiency without pooling layers. Further optimization of the DSCNN model could lead to even better emotion recognition performance, suggesting avenues for future research [11]. The research paper titled ”Investigation voice features for Speech emotion recognition based on four kinds of machine learning methods” covers the following topics in its introduction to speech emotion recognition: dimensionality reduction of feature quantities, emotional feature extraction, and model-based speech emotion recognition. SVM, Random Forest, NN, and KNN arethe four speech classification models that are systematically compared and analysed in this work. With an overall accuracy of 81.11 percent, the results depict that the SVM model-based speech recognition effect is the most noticeable one. The next is the NN model, which reaches 80.56 percent. There are obvious shortcomings in the KNN model and random forest model, whose accuracy is only 55.56 percent and 58.89 percent respectively. Speech emotional recognition is a kind of recognition of emotional speech signals, which can identify the emotional changes of each other through voice to understand the emotional state of each other. In this paper, we propose to compare the recognition effects of four emotional modes according to different speech features [12]. The research paper titled ”Design of a Convolutional Neural Network for Speech Emotion Recognition” is The accuracy of speech emotion recognition (SER) with voice improves with the amount of data used. Large amounts of data are especially necessaryfor deep learning. Nevertheless, when utilizing an already- existing data collection, the data set’s size is constrained and its constituent data points may have varying lengths. The audio recordings with utterances of varying lengths make up the data set used in this work. In this work, deep learning methods including a MLP and CNN were used to extract one-dimensional data from audio recordings and train two-dimensional Mel- spectrogram images. Additionally, audio files were shortened to less than two seconds in order to increase test accuracy [13].The paper titled ”Speech Emotion Recognition based on Interactive Convolutional Neural Network” presents an improved Speech Emotion Recognition (SER) system using an Interactive Convolutional Neural Network (ICNN). Through the use of Mel Frequency Cepstral Coefficients (MFCC), which take into account the interaction of various frequencies, the ICNN considerably enhances classification performance. While the Music Classification Module uses audio features to categorise songs into mood classes with 98 percent accuracy, the Emotion Module uses face photos to determine the user’s mood. The Module of Recommendation suggests songs based on user emotions and preferences, aiming to create an inter- active music player that uplifts the user’s mood. The paper underscores the superiority of ICNN over traditional CNN in SER through interactive convolution, enhancing feature representations for emotion recognition [14]. The research paper titled ” Emotion Recognition Through Speech Signal Using Python” voice analysis is used to identify emotional states and human emotions using voice recognition. This experiment takes into account the following emotions: neutral, joy, and melancholy. Physical attributes including blood pressure, heart rate, breathing, speech, skin suppleness, and muscle tension can all affect an individual’s emotions. The analysis makes use of Python libraries, and Keras is a high-level TensorFlow- based neural network API. The sequential model and the functional model are the two primary models that Keras provides. Because it depicts layers in a linear stack, the sequential model is appropriate for developing basic categorization networks and encoder-decoder models. In contrast, the functional model produces accurate results in 95 percent of the cases and allows shared layer multi-input and multi-output structures. Using Gaussian Naive Bayes as a classifier[15].

PROPOSED WORK

A. System Architecture

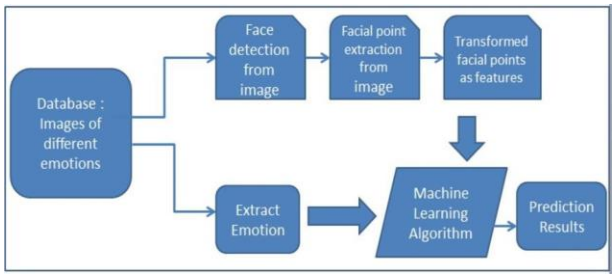


Fig 2 : System Architecture Stages of Recognition of Facial Expressions:

Using pictures of different facial expressions, the supervised learning method is used to train the facial expressionrecognition system. The system is divided into two stages: training and testing. During the training phase, the following processes are carried out: image acquisition, face detection, preprocessing, feature extraction, and classification. Thefollowing is a summary of the procedure: • Obtaining Images Facial expression-containing images are either taken in real time using a camera or gathered from a dataset. • Recognition of faces Pre-processing of images: Noise reduction and normalisation against pixel location or brightness variations are two aspects of image pre-processing. a) Normalisation of Colours b) Normalisation of Histograms • Feature extraction: In a pattern classification task, choosing the feature vector is crucial. The representation of.

Convolution Neural Network:

CNNs are vital in Deep Learning for Computer Vision tasks. They use fully connected, pooling, and convolutional layers toanalyse and extract features from images. This enables them to learn hierarchical representations of visual data, allowing them to classify images, detect objects, and segment images. CNNs revolutionized Computer Vision by capturing spatial dependencies and hierarchical patterns, leading to remarkable success in diverse applications.

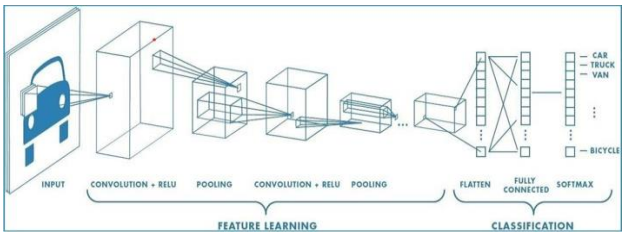


Fig 3 : Neural network with many convolutional layers

Convolution Layer:

The initial layer in feature extraction from an input image is this layer. Convolution uses input data to learn attributes, saving the link between pixels. A filter or kernel and an image matrix are the two inputs used in this mathematical technique.

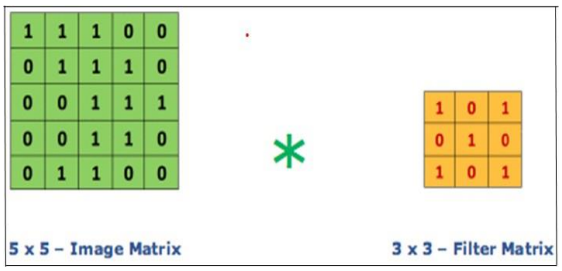


Fig 4 : Image matrix multiplies kernel or filter matrix

Examine a 5 x 5 with a filter matrix of 3 x 3 with picturepixel values of 0 and 1, as displayed below.

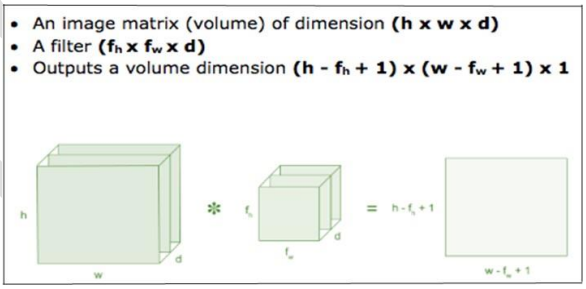


Fig 5 : continuation of Fig 4

Next, as the output below illustrates, the convolution of a 5 x5 image matrix multiplies with a 3 x 3 filter matrix to createa ”Feature Map.”

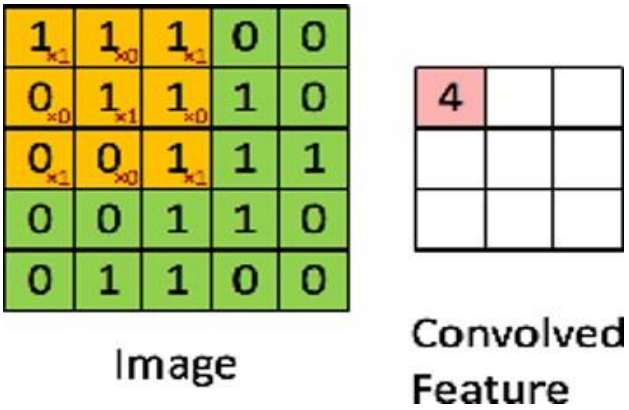


Fig 6 :3 x 3 Output matrix

It is possible to perform operations like edge detection, blurring, and sharpening by convolutioning an image with several filters. These operations are represented by distinct convolutional images that we can obtain by applying different kernels. Strides: In convolution the number of times pixel shifts are done to the input matrix is called stride. The filters move 1 pixel at a time when they have a stride of one, 2 pixels at a time when they have a stride of 2, and so on. During the convolution process, the filters skip every other pixel when a stride of two is used. With a stride of two pixels, convolution is demonstrated in the following graphic.

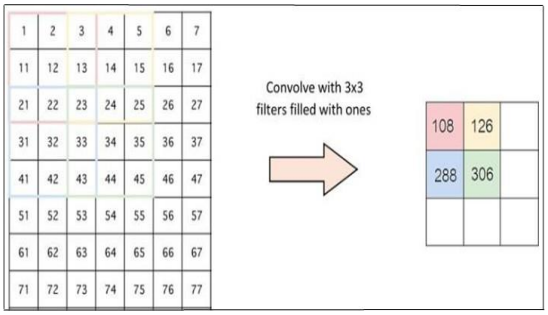


Fig 7 : Stride of 2 pixels

The area of the image is removed if filter does not fit leaving only the area that is legitimate. This indicates that only the areas of the image where the filter completely overlaps without going beyond the borders are subjected to the filter. Because only the portions of the picture that can be fully convolved with the filter are taken into consideration, valid padding guarantees that the output size is smaller than the input size. ReLU (Non Linearity):

Rectified Linear Unit, or ReLU for short, is a non-linear activation function that is frequently employed in CNNs, or convolutional neural networks. The function $f(x) = \max(0, x)$, where x is the input, is applied. ReLU is crucial to CNNs because it adds non-linearity to the system, enabling it to learn relationships that are complex

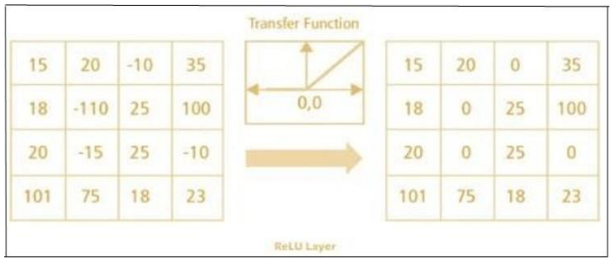


Fig 8 : ReLu operation

ReLU, tanh, and sigmoid are popular nonlinear activation functions in deep learning. While tanh and sigmoid functions can be used, ReLU is preferred by data scientists due to its computational efficiency and ability to mitigate the vanishing gradient problem. ReLU's sparsity property also aids inter- pretability and efficiency.

Pooling Layer:

Convolutional Neural Networks (CNNs) use pooling layersto lower the number of parameters when processing huge images. This method, sometimes referred to as downsampling or subsampling, aids in reducing each feature map's size while keeping important details. It can be max, average, and sum pooling. While average pooling uses the mean of the feature map's elements, max pooling chooses the biggest element from the feature map. Conversely, sum pooling determines the total of each element in the feature map. The network can concentrate on the most crucial features and ignore less crucial information by using these pooling processes to help reduce the spatial dimensions of the feature maps.

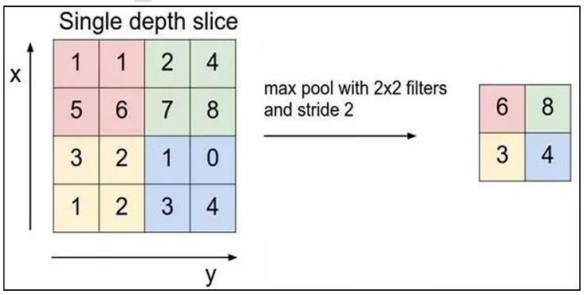


Fig 9 : Max Pooling

Completely Networked Layer: In a neural network, the layer where the input matrix is fed in after being flattened into a vector is known as the FC layer. This layer is comparable to the conventional neural network layer in that all neurons init are connected to all other neurons in the layers above and below. By taking into account the interactions between each feature in the flattened vector, the FC layer helps the network to discover intricate patterns and relationships.

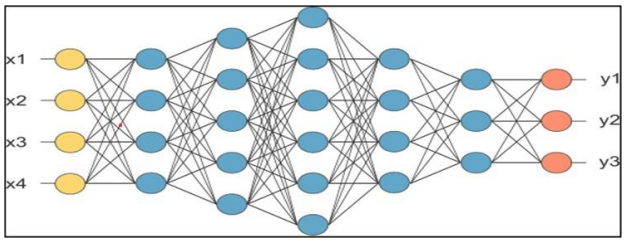


Fig 10: After pooling layer, flattened as FC layer

The feature matrix is transformed into a vector (x1, x2, x3,...) in the diagram above. We put these attributes together to makea model using the fully connected layers. Lastly, an activation function, such sigmoid or SoftMax, is used to categorise the outputs into groups like truck, automobile, dog, and so on.

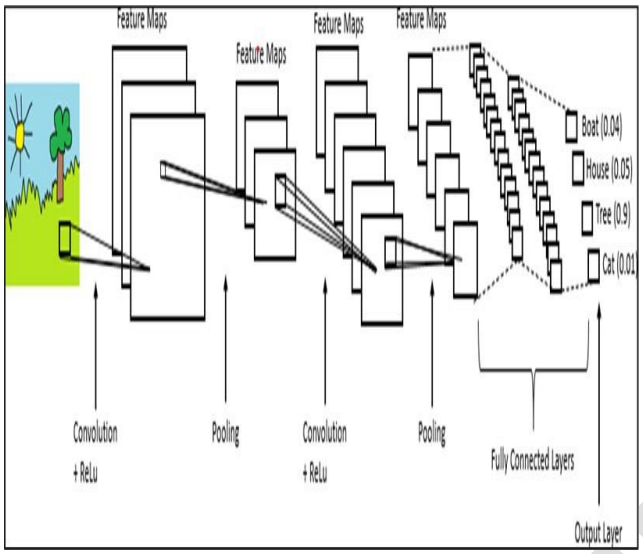


Fig 11: Complete CNN Architecture

B. Face Detection

The main objective of face detection techniques is to ac- curately identify and locate faces within images while mitigating external disturbances and other confounding factors. This process typically involves several key steps to achieve robust detection performance. Firstly, the use of an image pyramid facilitates multiscale analysis by decomposing theoriginal image into multiple resolutions. This hierarchical representation enables the extraction of features at different scales, aiding in the detection of faces across a range of sizes and orientations. Furthermore, the image is iteratively smoothed and down sampled within the pyramid, reducing noise and enhancing the saliency of facial features. Second, to describe local gradients in the image, the Histogram of Oriented Gradients (HOG) approach is applied as a feature descriptor. By quantifying the occurrences of gradient orientations within localized regions, HOG effectively captures the spatial distribution of gradient magnitudes, which are indicative of object edges and textures. When applied to face detection, HOG generates a descriptive representation of facial features based on intensity gradient distributions, enabling discriminative analysis for face localization. Based on the recovered HOG features, a linear classifier is then used to determine whether or not a certain area of the image contains a face. This classifier distinguishes between positive samples (regions containing faces) and negative samples (background regions) by learning a decision boundary in feature space.

Through training on labeled datasets, the classifier learns to discriminate between facial and non-facial patterns, thereby enabling accurate face detection. By integrating these steps, the face detection technique aims to achieve precise and robustlocalization of faces within images while mitigating noise and other interferences. Face identification over a wide range of picture datasets is made possible by the combination of multiscale analysis, feature representation with HOG descriptors, and classification using linear classifiers. This allows for a multitude of applications in computer vision, biometrics, and surveillance. A flexible set of programming functions, OpenCV is made for real-time computer vision applications. The main language of OpenCV is C++, and it has interfaces in C++, Python, Java, and It is compatible with different platforms such as Windows, Linux, macOS, as well as mo- bile platforms like Android, iOS, and Blackberry. Numerous domains, including image segmentation, mobile robotics, item identification, gesture recognition, and facial recognition, find use for OpenCV. In order to achieve gesture-controlled cameraaccess, image capture, image-to-text translation, and speech conversion, our project makes use of OpenCV version 2. The reason behind image processing There are five groups into which image processing purposes can be divided:

• Visualisation: Increasing an object's visibility that is difficultto see.

• Image restoration and sharpening: enhancing the clarity and quality of photographs.

•Image retrieval: Searching for specific images based on predefined criteria.

•Measurement of patterns: Extracting measurements and fea- tures from images.

•Image recognition: Identifying and classifying objects within images.

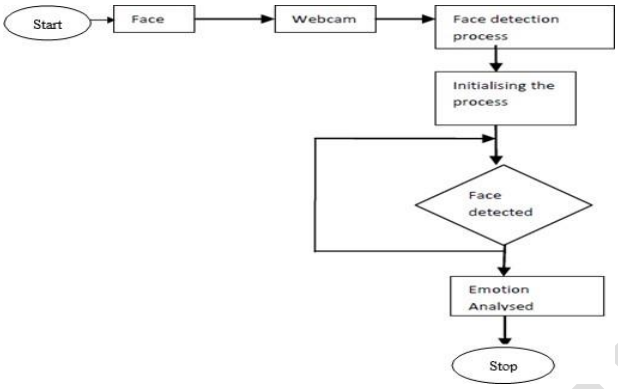
A small, open-source speech synthesiser for Linux and Win- dows that can speak English and eleven additional languages is called Espeak. It is employed for voice-to-text conversion.It is compact and supports a large number of languages. Rule files with feedback are used in Espeak software programming. It is SSML compatible. It is modifiable by voice variation. These are text files that have the ability to alter speech features like pitch range, add effects like echoes, whispers, and croaky voices, or

systematically alter formant frequencies to alter voice quality. Speaking at the default rate of 180 words per minute is too quick to understand. The text to voice signal conversion in our project is done using Espeak.

Fig 12: Flow Chart of Face Detection

C. Emotion Classification

Once a face has been identified in an image, the region of interest (ROI) that corresponds to the facial area is defined by the overlay of a bounding box. This ROI is then subjected to further processing using a specialized ”Predictor” scriptdesigned to extract 68 facial



landmark points, crucial for characterizing facial structure and expressions. These landmark points, saved in an array, serve as discriminative features for subsequent analysis. Subsequently, the feature array containing the landmark points undergoes Principal Component Analysis (PCA) reduction, a technique aimed at reducing data dimensionality and eliminating correlated coordinates. This process condenses the information into only the essential principal components, facilitating efficient analysis. The original 68x2 array representing the coordinates of each landmark point is transformed into a vector with dimensions 136x1, ready for further processing. The ”Predictor” code is trained on a dataset that consists of landmark maps and associated photos. Based on the pixel intensity values linked to each landmark point,the code learns to extract the facial landmark map from a particular face image using a supervised learning approach that uses regression trees and a gradient-boosting algorithm. This training enables the ”Predictor” to accurately predict the facial landmarks for new images, ensuring robust performance across diverse datasets. Following the PCA reduction, the resultant data is utilized for classification purposes. A multiclass Convolutional Neural Network (CNN) with a linear kernel is employed for this task. This CNN compares the input data, comprising the reduced landmark features, with stored data to determine the corresponding emotion class. In scenarios where the detected emotion is identified as anger, fear, or surprise, a safety precaution is executed. Specifically, a command is issued to decrease the speed of the wheelchair, prioritizing user safety and mitigating potential risks associated with height- ened emotional states. This integration of facial landmark extraction, dimensionality reduction, emotion classification, and safety measures underscores the system’s comprehensive approach to facial emotion recognition and user well-being.

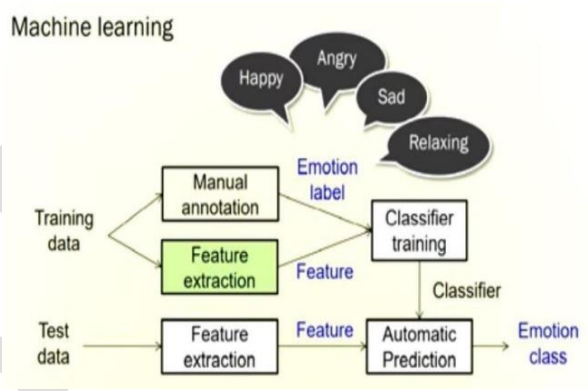


Fig 13 : Emotion Classification

D. Music Recommendation

The system is implemented using a camera feed as in- put. The camera captures video, which is then processed by dividing it into frames for analysis. Hidden Markov Model (HMM) classification is employed to process these framed images, considering all frames and pixel formats for emotion classification. This process involves calculating the value of each facial landmark and storing it for future reference. The classifier demonstrates high efficiency, reportedly achieving accuracy rates of 90-95 percent. Even in the presence of environmental changes affecting facial expressions, the system remains capable of accurately identifying emotions. Emotions are identified by comparing the calculated values from facial landmarks and pixel values to predefined thresholds in the system’s code. Upon identification, these emotional values are transferred to a web service for further processing. The system is equipped with a song selection feature where specific songs are associated with each emotion—happy, anger, sad, and surprise. When an emotion is detected, the system automati- cally selects and plays the corresponding song mapped to that emotion,

providing an interactive and responsive experience based on the poeple's emotional condition.

E. Movie Recommendation

Since the input is obtained in real time, the video is first captured by the camera, and then the framing is completed. The framed photos are processed using a hidden Markov model classification technique. For the purpose of classifying emotions, all frames and pixel formats are taken into consid- eration. Every landmark in the face has its value computed and saved for later use. The classifier has an efficiency of roughly 90–95 percent. in order for the system to recognise the face and the emotion being displayed, regardless of any changes in the face brought on by the surroundings. The values that are obtained and being set are then used to identify the emotions, and the value of the received pixel is compared to the values that are contained in the code as thresholds. Transferred to the web service are the values. The recognised emotion triggers the playing of the song. Each song has a designated set of feelings. The appropriate song will start playing when the sentiment is conveyed. The four emotions that might be employed are surprise, joyful, angry, and sad. The songs designated for each emotion are played whenthe happy emotion is identified. The same is true for the other emotions; the songs are played in accordance with the emotions that are detected.

F. Article Recommendation

Our platform utilizes NLP to analyze the emotional tone and subject matter of articles across various topics. By under- standing the sentiment and relevance of each piece, we deliver tailored article recommendations that cater to users' interests and current emotional states. Whether someone seeks uplifting stories, informative articles, or thought-provoking reads, our curated selection ensures they find content that resonates with them on a personal level

G. Healthcare Recommendation

We leverage data-driven insights and expertise in healthcare to provide personalized recommendations aimed at improving users' well-being. Whether it's suggesting mindfulness apps for stress relief, workout routines for physical fitness, or articles on mental health awareness, our recommendations promote holistic wellness and empower users to take proactive steps towards a healthier lifestyle.

H. Video Recommendation

With our video recommendation service, we utilize emo- tion detection technology and content analysis to suggest videos that align with users' preferences and emotional states. Whether it's entertaining comedy sketches, inspiring TED talks, or informative tutorials, our curated selection caters to diverse interests and ensures an engaging viewing experience for every user.

I. Podcast Recommendation

For podcast recommendations, we utilize a combination of user preferences, content analysis, and emotional detection to suggest podcasts that align with individual interests and moods. Whether someone is looking for educational content, entertainment, self-improvement, or simply a source of relax- ation, our recommendations cater to diverse tastes and pref-erences. By analyzing the topics, tone, and style of podcasts, we ensure that each recommendation resonates with users on a personal level, providing a curated selection of episodes that enhance their listening experience and keep them engaged. From thought-provoking interviews to captivating storytelling, our podcast recommendations offer something for everyone, fostering discovery and exploration within the vast world of podcasting.

RESULT AND DISCUSSION

We demonstrated a music recommendation system based on emotion recognition in this project. The system recognises facial emotions using a two-layer convolution network model. Seven distinct facial emotions are classified by the model using the image dataset. The training and validation accuracy indicate that the model is well-fitting and broadly applicable to the data. We acknowledge that we can yet do better. Analysing the system's performance when other emotions are taken into account might be fascinating. User preferences can be gathered and used with collaborative filtering to enhance the system as a whole. In subsequent work, we intend to address these concerns.

Algorithm	Accuracy
CNN	98%
Random Forest	96%
SVM	90%
RNN & LSTM	85%
JV8	50.80%

Fig 14 : Accuracy Comparisons

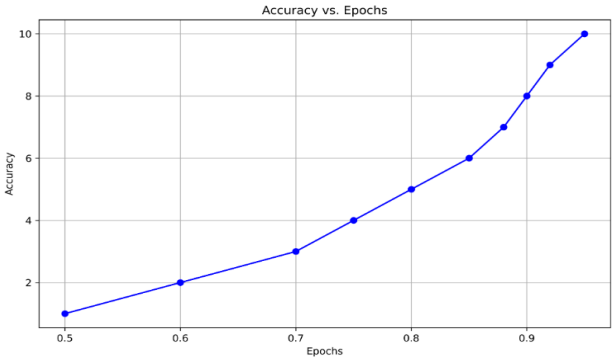
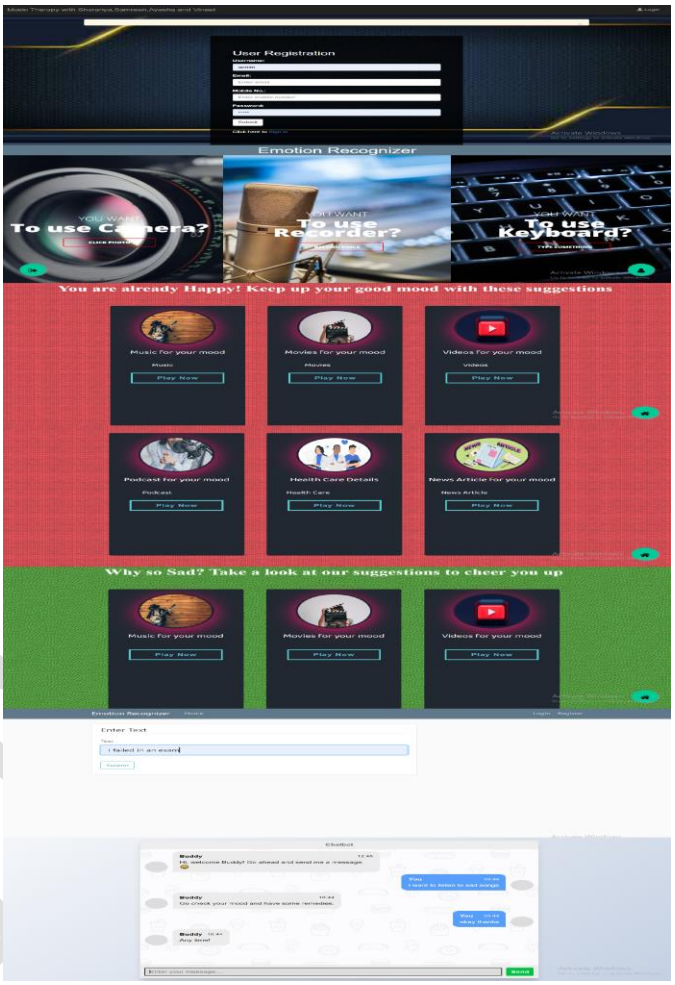


Fig 15 : Training Curve

Future Enhancement: In the contemporary environment, the facial recognition music player is an absolute necessity for every individual. This system is further improved with useful features for future upgrades. Facial expression detection is the technology used to improve the automatic song playing process. An RPI camera programming interface detects the expression on the face. At the moment, our system does not include feelings like fear and revulsion. But an other approach might use these feelings to improve music playback that is done automatically.

Snapshots:



CONCLUSION

We demonstrated a music recommendation system based on recognizing emotions in this project. The system recognises facial emotions using a two-layer convolution network model.Seven distinct facial emotions are classified by the model using the image dataset. The training and validation accuracy indicate that the model is well-fitting and broadly applicable tothe data. We acknowledge that we can yet do better. Analysingthe system's performance when other emotions are taken into account might be fascinating. User preferences can be gathered and used with collaborative filtering to enhance the system as a whole. In subsequent work, we intend to address these concerns.

REFERENCES

[1] C. Jain, K. Sawant, M. Rehman and R. Kumar, "Emotion Detection and Characterization using Facial Features," 2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE), Jaipur, India, 2018, pp. 1-6.

[2] A. Jaiswal, A. Krishna Raju and S. Deb, "Facial Emotion Detection Using Deep Learning," 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 2020, pp. 1-5.

[3] S. Begaj, A. O. Topal and M. Ali, "Emotion Recognition Based on

Facial Expressions Using Convolutional Neural Network (CNN)," 2020 International Conference on Computing, Networking, Telecommunica- tions Engineering Sciences Applications (CoNTESA), Tirana, Albania, 2020, pp. 58-63.

[4] Ram Kumar Madupu, Chiranjeevi Kothapalli, Vasanthi Yarra "Auto- matic Human Emotion Recognition System using Facial Expressions with Convolution Neural Network" Fourth International Conference on Electronics, Communication and Aerospace Technology (ICECA- 2020).

[5] Hari Kishan Kondaveeti, Mogili Vishal Goud "Emotion Detection using Deep Facial Features" IEEE International Conference on Advent Trends in Multidisciplinary Research and Innovation (ICATMRI) 2020.

[6] Charvi Jain, Kshitij Sawant, Mohammed Rehman "The Applicationof Artificial Intelligence in Emotion Recognition" 2020 International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI).

[7] E. Pranav, S. Kamal, C. Satheesh Chandran and M. H. Supriya, "Facial Emotion Recognition Using Deep Convolutional Neural Network," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2020, pp. 317-320.

[8] Monika Singh, Bharat Bhushan Naib, Amit Kumar Goel "Facial Emotion Detection using Action Units" Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020).

[9] U. Rashid, M. W. Iqbal, M. A. Sikandar, M. Q. Raiz, M. R. Naqvi,and S. K. Shahzad, "Emotion Detection of Contextual Text using Deep learning," 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), Istanbul, Turkey, 2020, pp. 1-5.

[10] Susu Yan, Liang Ye, Shuai Han, Tian Han, Yue Li and Esko Alasaarela "Speech Interactive Emotion Recognition System Based on Random Forest".

[11] T. M. Wani, T. S. Gunawan, S. A. A. Qadri, H. Mansor, M. Kar- tiwi and N. Ismail, "Speech Emotion Recognition using Convolution Neural Networks and Deep Stride Convolutional Neural Networks," 2020 6th International Conference on Wireless and Telematics (ICWT), Yogyakarta, Indonesia, 2020.

[12] H. Chen, Z. Liu, X. Kang, S. Nishide and F. Ren, "Investigating voice features for Speech emotion recognition based on four kinds of machine learning methods," 2019 IEEE 6th International Conference on Cloud Computing and Intelligence Systems (CCIS), Singapore, 2019, pp. 195- 199.

[13] K. H. Lee and D. H. Kim, "Design of a Convolutional Neural Network for Speech Emotion Recognition," 2020 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Korea (South), 2020, pp. 1332-1335.H. Cheng and X. Tang, "Speech Emotion Recognition based on In- teractive Convolutional Neural Network," 2020 IEEE 3rd International Conference on Information Communication and Signal Processing (ICI- CSP), Shanghai, China, 2020, pp. 163-167.M. A. Rohan, K. S. Swaroop, B. Mounika, K. Renuka and S. Nivas, "Emotion Recognition Through Speech Signal Using Python," 2020 International Conference on Smart Technologies in Comp