IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

Music Recommendation System Based on Human Emotion Recognition

Dr. Akila K[1], Srivatsan V[2], Gunalan S[3]
Assistant Professor, SRM Institute of Science and Technology, Vadapalani, Chennai, India[1,2,3]

Abstract: *This project presents human feelings recognition can be solved by analysing one or more of these features. Choosing to follow the lexical features would require a transcript of the face which would further require an additional step of feature extraction from face if one wants to predict emotions from face reaction is the act of attempting to recognize human feelings and affective states from face. This is also the phenomenon that animals like dogs and horses employ to be able to understand human feelings, we will use the libraries Tensorflow and keras to build a model using an Artificial Neural Network. This will be able to recognize feelings from face data set. We will load the data, extract features from it, then split the dataset into training and testing sets. Then, we'll initialize an Artificial Neural Network and train the model. Finally, we'll calculate the accuracy of our model*

Keywords: object detection, Emotion recognition, CNN, ANN

I. INTRODUCTION

In the sectors of aviation and transportation, object detection for employing images— which is regarded as the categorization of human feelings—plays a significant role. We can now acquire clearer and better quality photographs thanks to advancements in the technologies used to use images. - High- resolution photos can offer more in-depth information, which can improve our ability to recognize and find the target items. However, the scenario of human sentiments for images is often a complex item in most instances. The accuracy and rate of object detection using the CNN algorithm are now quite good. Therefore, this strategy makes it simple to discern human emotions.

II. LITERATURE REVIEW

The task of assessing emotions in written text is a recently developed area of study in computational linguistics. This involves analyzing and categorizing text into specific emotional classes, using dimensional models of emotions. In the study, a set of social media data was classified into five different emotion categories using machine learning techniques.[2] dentifying and categorizing emotions and sentiments in dialogues is a difficult task that requires considerable effort and expertise it requires multi-label emotion detection and sentiment analysis. The Multimodal Multi-label Emotion, Intensity, and Sentiment Dialogue dataset (MEISD), which has a substantial amount of data and is balanced, has been introduced by the authors.

In addition to providing a starting point for benchmarking future research, this dataset also serves as a foundation for it.[3]The study aimed to evaluate various classification techniques for recognizing facial emotions based on a 5-dimensional principal component reduction and a Pyramid of Histogram of Orientation Gradients. The results demonstrated that a decision tree- based method was more successful in mitigating overfitting compared to a deep neural network.[4]This study focuses on using facial recognition software to recognize people, analyze their facial expressions, and forecast their emotional condition. This strategy may be useful for protecting data in circumstances involving security. The face of the person must first be recognized in order to undertake emotion analysis. After that, several algorithms, including CNN, are used. While the testing data is acquired by translating a recorded image into pixels, the training data for these algorithms consists of a dataset with pixel values and accompanying emotions. The emotional condition of the subject in the shot is then predicted using an algorithm.[5]The proposed emotion detection system uses a Convolutional Neural Network (CNN) model to recognize emotions and combines two modalities, namely speech and facial expressions. This system's goal is to categorize seven basic emotions from a given video using the audio and visual frames that are contained in the video. The RAVDESS dataset was used in this work to carry

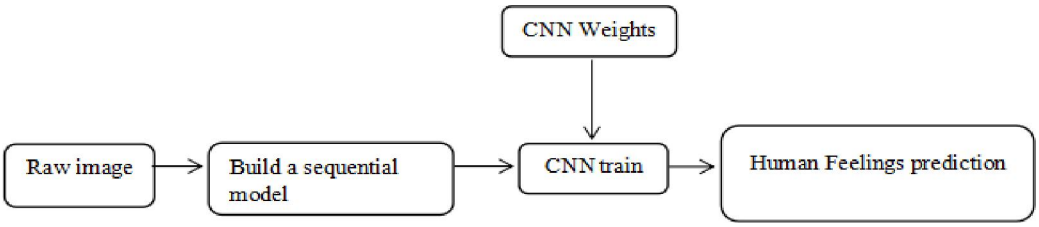
IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

out bimodal emotion identification.[6] The characteristics of facial expressions are extracted by the study using a neural network to recognize different facial emotions. It achieves 97 percent accuracy with an existing simulator and comparisons of existing and new systems are made.[7] The paper proposes a prototype system that, by combining image processing with a neural network-based method, automatically recognizes emotions shown on a person's face. Frontal face photos in color serve as the system's input.[8] Emotion recognition is an important part of interpersonal relationships, and there are many techniques for recognizing it. A wide range of emotions are communicated through facial expressions, which are also the main factor in their identification. The paper suggests a machine learning-based real-time emotion identification system to overcome this problem.[9]The technique for emotion detection makes use of both biophysical and facial expression analysis to determine a person's emotional state. The paper suggests a method that uses feature extraction and convolutional neural networks to identify emotions. According to the testing data, HOG and CNN together perform better than cutting-edge models.[10] The study highlights that employing neural networks for image pre-processing can enhance the precision and effectiveness of Facial Emotion Recognition (FER) applications in individuals with Autism Spectrum Disorder (ASD). The study created a model using Transfer Learning, based on the AlexNet architecture, and utilized three distinct techniques for picture inputs. The findings indicate that the accuracy levels of up to 99.90% were achieved through this approach.

III. PROPOSED SYSTEM

Our goal was to create a deep learning approach to face emotion classification. We put out a system that anticipates face expressions by gathering more examples of images and comparing them across multiple categories. The shape and texture-based elements of the image make up the majority of its attributes. Initially we are preprocessing our dataset and implementing more than two CNN architecture. Each architecture gives a different kind of accuracy, so we compare each one architecture. Finally, build a hierarchical model for saving our trained model. Once the model is saved then we can deploy it in any web browser by using Django Framework.

Architecture:



Data Collection

The first step in building an emotion recognition system is to collect a dataset of human emotions. This dataset should contain a variety of emotions expressed by different individuals in various contexts.

Data Pre-processing

Pre-processing is necessary once the data has been gathered. In order to do this, the data must be cleaned, noise removed, and normalized.

Development of Model

Building a music recommendation system based on human emotion recognition requires careful consideration of data collection, pre- processing, feature extraction, emotion classification, recommendation, validation, and deployment. It is important to choose appropriate algorithms and techniques at each stage to ensure that the system is accurate and effective in recommending music that matches the user's emotional state.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

Conv2d:

A small weight matrix, referred to as a kernel, is used in the 2D convolution technique to iteratively pass over the 2D input data. A single output pixel is produced by multiplying the values in the input data by the corresponding weights in the kernel and then adding the resulting products. Every area of the input that the kernel traverses receives this operation again, creating a fresh 2D feature matrix. The weighted sum of the input features, with the kernel weights serving as the coefficients, determines the output features. It is a useful tool for extracting useful features from 2D data in various computer vision and deep learning applications because the input features are placed roughly.

The location of an input feature in relation to the kernel determines if it falls within the "roughly same location" or not. The number of input features that are combined to produce an output feature is proportional to the size of the kernel.

MaxPooling2D layer

Convolutional neural networks (CNNs) employ the Max Pooling 2D layer type to condense the spatial dimensions of the feature maps that are generated by convolutional layers. The basic goal of Max Pooling is to minimize the number of parameters in the network and downsample the feature maps, which can assist avoid overfitting and increase computational efficiency. The MaxPooling 2D layer operates on 2D feature maps produced by the convolutional layer. It slides a 2D window over the feature map and selects the maximum value in each window. The size of the window (pool size) and the stride (the step size of the sliding window) are hyperparameters that can be set by the user. We would slide the window over the feature map and choose the highest value in each 2x2 subregion, producing a 2x2 down sampled feature map, for instance, If we applied a 2x2 window with a stride of 2 to a 4x4 feature map. The original feature map's spatial dimensions would be halved in this downsampled version, which might result in fewer parameters and more effective network operation. The MaxPooling 2D layer is often used in CNNs for image classification tasks.

After several convolutional layers, the feature maps can become very large and the number of parameters can become prohibitively high. Adding a MaxPooling 2D layer after each convolutional layer can help downsample the feature maps and reduce the number of parameters, while preserving the most salient features of the input image.

Flatten layer

After convolution, the flatten layer is used to shrink the image's size. A hidden layer called the Dense layer is used to build a fully linked model. The Dropout approach is used to avoid overfitting the dataset, and the output layer is sparse since there is just one neuron that determines which category each picture belongs to.

Dense layer

The key components of the Dense layer are the activation function, referred to as "Activation," the weight matrix generated by the layer, called "Kernel," and the bias vector produced by the layer, which is only applicable if use_bias is set to True, known as "Bias." These three terms collectively define the functionality of the Dense layer.

Dropout layer

In neural networks, the Dropout layer is employed to prevent overfitting. In each training phase, part of the input units are randomly set to 0, which helps it achieve this. The rate determines how often something happens. The non-zero inputs are scaled up by a factor of 1/(1-rate) in order to maintain a constant total of inputs.

IV. EXPERIMENTAL ANALYSIS

1st Phase of Our Implementation:

IMPORT THE GIVEN IMAGE FROM DATASET:

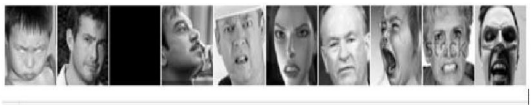
We must first import our dataset and set its size, rescale factor, range, zoom range, and horizontal flip before using the Keras preprocessing image data generator function. The data generator tool can then be used to set the proper size, batch size, and class-mode for the train, test, and validation phases. We can add CNN layers to train the network after importing our image dataset from a folder. This function assists in preprocessing the data and getting it ready for use in network training with the proper settings.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

Human Feelings: Angry

```
Trainned data for angry:

 ====== Images in:  Dataset/train/angry
images_count:   3995
min_width:      48
max_width:      48
min_height:     48
max_height:     48
```



Disgust:

```
Trainned data for disgust:

 ====== Images in:  Dataset/train/disgust
images_count:   436
min_width:      48
max_width:      48
min_height:     48
max_height:     48
```

Happy:

```
Trainned data for happy:

 ====== Images in:  Dataset/train/happy
images_count:   7215
min_width:      48
max_width:      48
min_height:     48
max_height:     48
```

Sad:

```
Trainned data for sad:

 ====== Images in:  Dataset/train/sad
images_count:   492
min_width:      48
max_width:      48
min_height:     48
max_height:     48
```

2nd Phase of Our Implementation:

TO TRAIN THE MODULE BY GIVEN IMAGE DATASET:

The total number of epochs, the number of training steps for each epoch, the validation data, and the validation steps must all be specified in order to train our dataset. Once we know this, we can actually train the dataset using the classifier and fit generator functions. These procedures use the parameters that we have provided to train the network

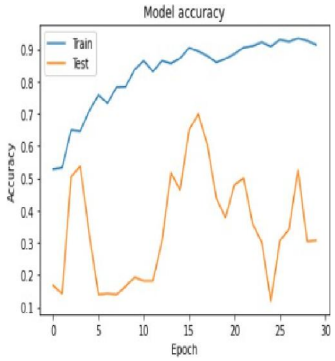
IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

appropriately. Overall, we can efficiently train our dataset and produce precise predictions by using these parameters and functions.

Convo Layer:

The feature extractor layer connects a portion of the input image to the Convo layer in order to extract features from the image. The dot product between the filter and a receptive field is calculated during convolution operations as a result, which are possible. The receptive field is a local region with the same size as the filter in the input image.This operation produces an integer that represents the output volume.After performing the convolution process, the filter is advanced to the following receptive field of the same input image based on the Stride value. Up till the entire image has been processed, this procedure is repeated. The output of this layer is subsequently sent into the following layer as its input.

CNN model trained dataset accuracy



Pooling Layer:

To decrease the spatial volume of the input picture, a pooling layer is added after the convolution layer. When utilizing a fully linked layer following the convolution layer, this is required to reduce computing costs. Max pooling is commonly employed for this, where a single depth slice is collected and the maximum value is calculated using a stride of two. As a result, the input's dimensions are reduced from four to two.

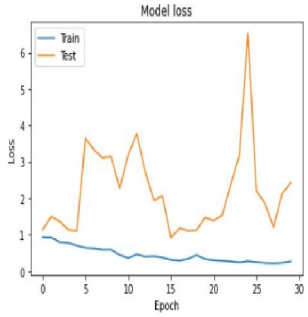
Fully Connected Layer (FC):

A fully connected layer is made up of neurons, weights, and biases that connect to neurons in the layer beneath. Through training, these layers are used to classify photos into numerous categories.

Softmax / Logistic Layer:

Sometimes referred to as the layer after that, the Softmax or Logistic layer is the final layer in a CNN and follows the fully linked layer. Softmax is used for classification issues requiring multiple classes as opposed to logistic, which is used for binary classification.

CNN model trained dataset loss values



IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

Output Layer:
Label in one-hot encoding form is present in the output layer. By this point, you know a lot about CNN.
3rd Phase of Our Implementation
Using the Django Framework to deploy the model and forecast results
Our Django framework uses this module to convert the trained deep learning model into a hierarchical data format file (.h5 file), which improves user interaction and prediction accuracy.
Django:
Django, a web framework for Python, simplifies the process of creating fast and reliable websites that are also secure. Experienced programmers designed Django to handle many of the tedious tasks involved in web development. With Django, developers can focus on building their applications without starting from scratch.

V. CONCLUSION

The project's goal was to use deep learning techniques to classify images of human emotions. This is a challenging issue that has previously been addressed using a variety of techniques, with feature engineering producing positive outcomes. Although feature learning is one of the promises of deep learning, this study concentrated on it. Although it is not required, feature engineering, picture pre-processing improves classification accuracy. As a result, it lessens input data noise. Software used today to identify human emotions uses feature engineering. Due to a significant constraint, a solution entirely based on feature learning does not appear to be close yet. Using deep learning algorithms, it is possible to classify human emotions.

REFERENCES

[1]. Sonia Xylina Mashal , Kavita Asnani,Emotion Analysis of Social Media Data using Machine Learning Techniques,National Conference On Advances In Computational Biology, Communication, And Data Analytics (ACBCDA 2017)

[2]. Mauajama Firdaus∗ , Hardik Chauhan, AsifEkbal and Pushpak Bhattacharyya,MEISD: A Multimodal Multi- Label Emotion, Intensity and Sentiment Dialogue Dataset for Emotion Recognition and Sentiment Analysis in Conversations Proceedings of the 28th International Conference on Computational Linguistics, Barcelona, Spain (Online), December 8-13, 2020

[3]. Andrew Koch,Emotion recognition classification methods,Research School of Computer Science, Australian National University u5371834@anu.edu.au

[4]. Rahul Mahadeo Shahane, Ramakrishna Sharma.K, Md.Seemab Siddeeq,Emotion Recognition using Feed Forward Neural Network & Naïve Bayes,International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278- 3075, Volume-9 Issue-2, December 2019

[5]. Manisha S, Nafisa Saida H, Nandita Gopal, Roshni P Anand,Bimodal Emotion Recognition using Machine Learning,International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249- 8958, Volume-10 Issue-4, April 2021

[6]. Dilbag Singh,Human Emotion Recognition System,I.J. Image, Graphics and Signal Processing, 2012, 8, 50-56 Published Online August 2012 in MECs (http://www.mecs- press.org/)DOI: 10.5815/ijigsp.2012.08.07

[7]. Recognition based on Image Processing and Machine Learning,International Journal of Computer Applications (0975 – 8887),Volume 139 – No.11, April 2016

[8]. Monisha.G.S,Yogashree.G.S, Baghyalaksmi.R ,Haritha.P,Enhanced Automatic Recognition of Human Emotions Using Machine Learning Techniques,International Conference on Machine Learning and Data EngineeringProcedia Computer Science 218 (2023) 375–382

[9]. Chahak Gautam, Seeja K.R,Facial emotion recognition using Handcrafted features and CNN,International Conference on Machine Learning and Data Engineering/ Procedia Computer Science 218 (2023) 1295–1303

[10]. H. Arabian*. V. Wagner-Hartl.**J. Geoffrey Chase***. K. Möller*,Image Pre- processing Significance on Regions of Impact in a Trained Network for Facial Emotion Recognition H. Arabian et al. / IFAC PapersOnLine 54-15 (2021) 299–303