Proceedings of the 4th International Conference on Computing and Data Science (CONF-CDS 2022)
DOI: 10.54254/2755-2721/2/20220656

Emotion-Based Music Movie and TV Series Recommendation System Using Deep Learning Algorithm

D.Deepa*, D.M.Vijayalakshmi , B.Shanmathi , A.Sherlip Evelin and K.Tamil Elakkiya

Department of Computer Science and Engineering, Kongu

deepadresearch@gmail.com

Abstract. It can be challenging to decide which music or movies to listen to from a huge number of options. The major purpose of our music and movie recommendation system is to provide clients with selections that fit their tastes. An assessment of a user's facial expression may provide insight into their current emotional or mental state. More than 60% of users anticipate that the number of songs in their music collection will grow to the point where they will be unable to find the song they need to play at some point in the future. It is feasible to assist a user in picking which music or movie to listen to or watch, by building a suggestion system. The face of the user is detected using the webcam. The snapshot of the user is taken based on their mood or feeling. It recognizes six facial expressions: angry, sad, fearful, joyful, surprised, and neutral. Based on the expression classification, the users are given three categories of recommendation as movies, music, or series based on their feelings. Seven different human facial expressions are classified using the Convolutional Neural Network (CNN) model. The Haar Cascade is an Object Detection Algorithm for recognizing faces in images and real-time video.

Keywords: Convolutional Neural Network, H5 Model, Haar Cascade Frontal Face Classifier.

1. Introduction

Music, movies, and other TV series can reduce stress and provide relaxation to people. Everyone enjoys these activities because they can switch their mood and provide a sense of relief in their everyday lives. Every type of music or movie evokes diverse feelings since the user may quickly associate them with their surroundings. Also, it can connect with and affect people in ways that other forms of communication do not. Many people turn to music and movies to connect with others, express themselves, or discover common ground among peers. But, the user still has to actively browse through the songs and movies and select songs based on their current mood and behavior. A user had to carefully scroll through their playlist and select tracks that match their mood. This was a time consuming procedure, and people frequently struggled to come up with a suitable choice of songs or movies. Deep learning based facial expression recognition is one of the techniques for detecting human emotion states such as anger, happiness, fear, neutral, sad, and surprise. This technology seeks to recognize facial expressions automatically and accurately identify emotional states. In this method, CNN is trained by sending annotated facial photographs from a facial expression dataset to it. After that, the proposed CNN model determines

DOI: 10.54254/2755-2721/2/20220656

which facial expression is used, and based on the facial expression detected, music, movies, and TV series will be suggested through YouTube.

2. Related Work

A two-stage emotion recognition utilizing frame level and video level information, contrasts a seven-class classifier with a two-step classification for categorical emotion recognition. [1] They used the FG2020 Multimodal Emotion Recognition (MER) Dataset, including skeleton data obtained with a Microsoft Kinect and video. They compare the performance of numerous unimodal features as well as various multimodal feature combinations. They also compare features at the frame and video levels. Changes in the curvatures of the face and the brightness of the pixels that corresponded [2]. The author used Artificial Neural Networks (ANN) to classify emotions. Several playlist approaches were also suggested by the author.

Authors Published a paper that will show us the recommended songs when a specific song is processed using libraries like NumPy and Pandas [3]. Music service providers need a useful system for categorizing recordings and helping their customers find music by providing outstanding suggestions[4]. Using Support Vector Machine (SVM), a database of 714 face emotion images was published. It was created by taking two digital photos of seven different facial expressions on 51 different persons. Later, utilizing 476 training and 238 testing to identify seven emotions, the performance of four SVM kernels for face emotion identification was evaluated. By eliminating the manual selection, a feature extraction method based on a deep convolutional neural network and It reconstructs the conventional local binary pattern (LBP) feature operator for facial expression images and fuses it with the abstract expression in the full connection layer[5]，[6]. A paper describing the Face expression recognition method and the test sample image's expression that is recognized and classified by utilizing the Softmax Classifier and Convolutional Neural Network has been published (CNN).

The use of deep learning for face expression identification is now common these days [7]. The dataset consists of 35,887 face grayscale images, and the batch size options are 8, 40, 50, 55, 88, 100, and 128. The final proposed batch size in the training model is 50, which indicates that the proposed deep convolutional neural net-work is the most effective model to use. The diagonal characteristics are also apparent in the confusion matrix, indicating that the method of this paper successfully achieves the effect of expression classification. Researchers proposed work on multimodal emotion identification from voice and expression was published [8]. It has around 12 hours of audiovisual content, including video, audio, speech text, and facial expressions from 10 actors in either scripted or impromptu scenes. To extract face expression components from this data, several small scale kernel convolution blocks were created [9]. With some performance loss, the dimensionality of the feature vector was reduced in this research using a distance metric learning approach. They used 5 and 10 dim feature vectors to classify genres without losing speed. The bulk of music recommendation algorithms uses CF and CBF to find common patterns. One of the characteristics that are most frequently used for this filtering is genre [10]. This paper focused on a neural network based personalized music recommendation system that was created based on research on personalized music recommendation systems that solely use tag information [11][12].

DOI: 10.54254/2755-2721/2/20220656

3. Methodology

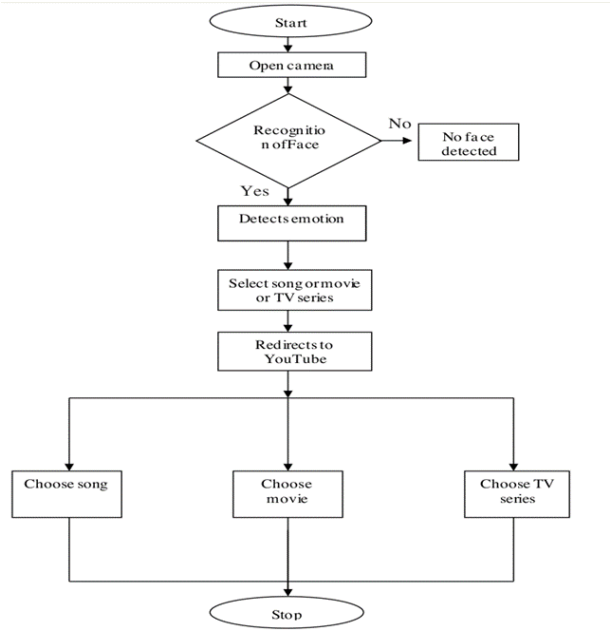

Fig. 1. Workflow of Recommendation System.

The proposed system also suggests movies and TV series along with song suggestions. This project focuses on the user's emotions as sensed by the webcam, and music, movies, and series are offered based on facial expressions. Following the recognition of their facial expression, the desired movie, song, or series is selected based on their mood. The major goal of this system is to create a sophisticated recommendation system that can improve the user's moods and rejuvenate them. Fig1. Show the work flow of the recommendation system. One of the strategies for recognizing human emotional states is deep learningbased facial expression identification such as anger, happiness, fear, neutral, sad, and surprise. This technique aims to reliably determine emotional states by automatically recognizing facial expressions. CNN is trained using this method by feeding it annotated facial pictures from a facial expression dataset. Following that, the proposed CNN model identifies which facial expression is employed, and music, movies, and TV shows are offered via YouTube based on the facial emotion detected [13].

4. Convolutional Neural Network

Convolutional neural networks are a particular kind of deep neural network that is used in deep learning to analyze visual vision. It uses a process called convolution. By using a mathematical technique called convolution, it is possible to create a function that expresses how the shape of one is influenced by the other. Convolutional neural networks are composed of many layers of neurons [14]. Similar to their biological counterparts, artificial neurons are mathematical functions that compute the weighted sum of numerous inputs and produce an activation value. When an image is an input, each layer of a convolutional neural network generates many activation functions that are then transferred to the following layer [15]. Horizontal and diagonal edges are taken from the first layer. The subsequent layer, which is in charge of identifying more intricate properties like corners and combinational edges, receives this information. Even more intricate things like objects, faces, and other things are recognized by it [16]. The Pooling layer, like the Convolutional Layer, is in charge of shrinking the spatial size of the Convolved Feature. The amount of computing power needed to process the data is decreased by reducing its size [17]. Max Pooling is used to determine which pixel from a kernelcovered area of the image has the biggest value. It gets rid of all clamorous activations.

$$Wout = W - F + \frac{2P}{S} + 1 \quad (1)$$

DOI: 10.54254/2755-2721/2/20220656

4.1 ReLU Layer

After each convolution operation, a nonlinear layer is added. It has a nonlinear feature due to its activation function. The function () =max (0,) is computed [19]. To put it another way, the activation is simply set to zero

4.2 Pooling Layer

It reduces the size of the image by down sampling it. After the convolutional, nonlinear, and pooling layers have been completed, a fully connected layer must be added.

$$W2 = \frac{W1 - F}{S} + 1 \quad (2)$$

$$H2 = \frac{H1 - F1}{S} + 1 \quad (3)$$

$$D2 = D1 \quad (4)$$

The method above can be used to calculate the output volume's size, where W2, H2, and D2 represent the output's width, height, and depth [20].

4.3 Fully-Connected Layer

The output data from convolutional networks is received by the fully connected layer. When a completely linked layer is attached to the end of the network, an Ndimensional vector is created, where N is the number of classes from which the model chooses the necessary class [21].

$$g(Wx + b) \quad (5)$$

Dimensions of the output tensor can be determined from the input tensor by using the above formula where g is the activation function, x is the given input vector, W is the weight matrix and b is the bias vector [22].

4.4 Dropout Layer

When every attribute is linked to the Fully Connected Layer, the training dataset is susceptible to overfitting occurs when a model performs so well on training data that it has a negative effect on its performance when applied to new data [23]. In order to tackle this issue, a dropout layer is utilized, which causes a tiny model to be created by removing a few neurons from the neural network during training. Thirty percent of the nodes in the neural network are lost out randomly with a dropout of 0.3.

4.5 Activation Function

One of the most important components of the CNN model is the activation function. They are used to learn and approximation any type of continuous and complex net-work variable to variable association. It adds nonlinearity to the network. ReLU, Softmax, tanH, and Sigmoid are the activation functions used in deepa learning models. There is a specific usage for each of these functions. Sigmoid and softmax functions are preferred for a CNN model for binary classification, although softmax is commonly used for multi-class classification [24].

5. HAAR Cascade

Haar Cascade Detection is an extensive face detection approach that has been around for quite some time. Faces, eyes, and lips were all recognized using Haar Features. It's important to remember that, like other machine learning models, this one takes a lot of positive photos of faces and negative photos of non faces to train the classifier. This algorithm consists of four stages. Firstly, it calculates haar feature [25].

5.1 Calculating Haar Features

The initial step is to gather the Haar characteristics. Haar feature performs a set of calculations on neighboring at a given location, rectangular parts of a detection window. The total of the pixel intensities in each section is then subtracted from the total. Below are a few examples of Haar characteristics [26].

DOI: 10.54254/2755-2721/2/20220656

5.2 Creating Integral Images

It constructs subrectangles and array references for each of those subrectangles instead of computing at each pixel. The Haar features are then computed using them. It's vital to remember that while doing object detection, practically all of the Haar characteristics are meaningless because the only features that matter are those of the object. Adaboost is useful in choosing the finest characteristics from hundreds of thousands of Haar features to represent an object [27].

5.3 Adaboost Training

Adaboost generally selects the most useful features and trains the classifiers with which they should be used It combines weak classifiers used by the algorithm to detect things to build a powerful classifier. Weak learners are produced by moving a window over the input image and computing Haar characteristics for each region of the image. This difference is compared to a learned threshold for distinguishing between non-objects and objects. Because these are poor classifiers, creating a strong classifier is essential [28].

5.4 Implementing Cascade Classifiers

Each stage of the cascade classifier has a collection of weak learners. Weak students are taught by boosting, which generates a highly accurate classifier based on the average prediction of all weak students. The classifier then selects whether to mark an object as found or to go on to the next region based on this prediction. Because the bulk of the windows does not contain anything of interest, stages are designed to eliminate negative samples as quickly as feasible [29].

6. Hierarchical Data Format 5

HDF5 is gaining popularity because of its portability. Other programming languages that may handle HDF5 files include Python, MATLAB, Fortran, and C. It is widely utilized in the scientific world to store huge datasets, as Simon mentioned. It is a file format for storing structured data rather than a model. Because the weights and model configuration may be easily stored in a single file, Keras saves models in this format. It is a file format for storing structured data rather than a model.

6.1 Data Set

The dataset taken for this experiment covers six different face emotions (Happy, Angry, Surprise, Fear, Neutral). These are normally collected images of size 1920*2560 pixels in size. The size of the training images is 28710 photos. These photos are divided into the following categories: Happy with 5300 samples, Sad with 5110 samples, Angry with 4500 samples, Surprise with 4600 samples , Fear with 4200 samples and Neutral with 5000 samples.

7. Performance Metrics

Accuracy can be used to assess the performance of the model. Accuracy refers to the percentage of correct predictions made by the model. It is calculated based on the

$$Accuracy = \frac{(TP + TN)}{TP + TN + FP + FN} \quad (6)$$

TP stands for "true positives," FN for "false negatives," TN for "true negatives," and FP for "false positives," respectively.

DOI: 10.54254/2755-2721/2/20220656

8. Result

The proposed system suggests music, movie, and series based on the user's emotions as captured by the webcam. To detect the face in real-time video, the CNN Algorithm is used. Depending on the facial expression and mood the user will be redirected to YouTube which suggests songs, movies, or TV series. Comparing the training data of the mobile net and the H5 model, the H5 model gives slightly higher accuracy than that of the mobile net. Even though convolutional neural network models like mobile net, and Restnet recognize emotions and characters there are few contradictions in detection. For suggesting music, movies, and series the pre-trained modified architecture gives a loss of 4% and accuracy of 98% shown in Fig.2

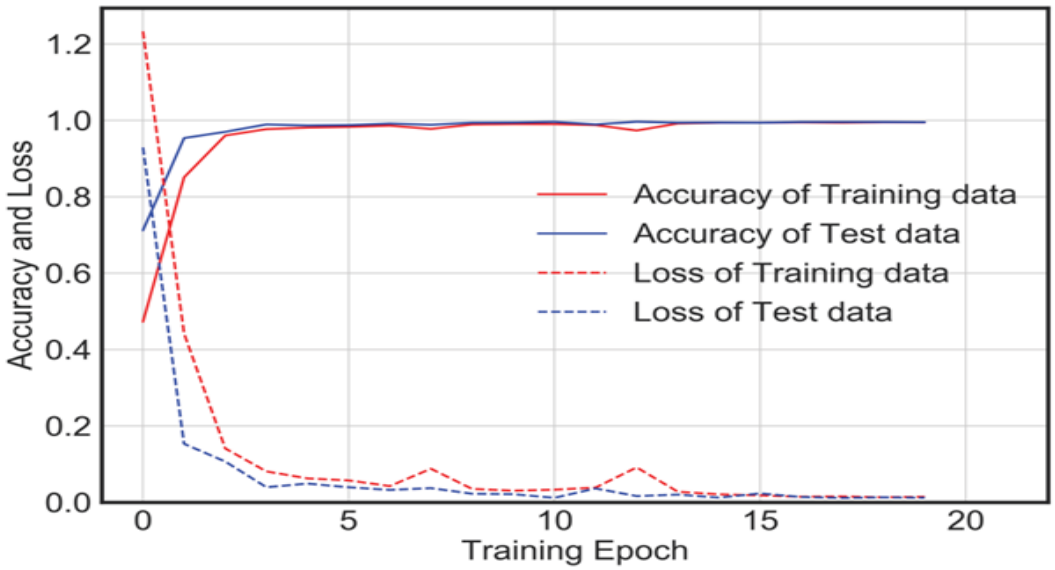

Fig. 2. Accuracy of Mobile Net Model vs. H5Model.

9. Conclusion

An emotion based music, movie, and series recommendation system is demonstrated in this proposed project. Using the image collection, the model distinguishes six different facial expressions. It will be fascinating to see how the system reacts to the introduction of new emotions. It may be able to divide adults and children into categories and then recommend music, movies, and television shows based on those categories. Such an application might be valuable in supporting humans in relaxing and lowering stress in today's technological world.

References

[1] Carla Viegas "Two-stage emotion recognition using Frame level and Video level Features", 15th IEEE International Conference on Automatic Face and Gesture Recognition, pp.912-915,2020.

[2] Renuka R. Londhe, Vrushshen P. Pawar "Analysis of Facial Expression and Recognition Based on Statistical Approach", International Journal of Soft Computing and Engineering((IJSCE), vol.2, pp.391-394,2012.

[3] Varsha Verma, Ninad Marathe, Parth Sanghavi, Dr. Prashant Nitnaware “Music Recommendation System Using Machine Learning”, International Journal of Scientific Research, vol.7, no.6, pp.80-88, 2021.

[4] Ibrahim A. Adeyanju and Elijah “Performance Evaluation of Different Support Vector Machine Kernels for Facial Emotion Recognition”, SAI Intelligent Systems Conference, vol.6, pp.804-806, 2015.

DOI: 10.54254/2755-2721/2/20220656

[5] F. Kong, “Facial Expression Recognition Method based on Deep Convolutional Neural Network Combined with Improved LBP Features”, Personal and Ubiquitous Computing, vol.23, no.4, pp.531-539,2019.

[6] Hongli Zhang, Alireza Jolfaei and Mamoun Alazab. “A Face Emotion Recognition Method Using Convolutional Neural Network an Image Edge Computing”, 2949741,2019.

[7] Lu Lingling Liu, “Human Face Expression Recognition Based on Deep Learning-Deep Convolutional Neural Network”, 2019 International Conference on Smart Grid and Electrical Automation (ICSGEA).

[8] Sharmeen S. Saleem Abdullah,Siddeeq Y. Ameen,Mohammed A. M. Sadeeq, Subhi R. M.Zeebaree et al, “Multimodal Emotion Recognition Using Deep Learning ,” International Journal of Information Management, vol.02, No. 02, pp.52-58,2021.

[9] Lee, J., Shin, S., Jang, D., Jang, S. J., & Yoon, K. (2015, January). Music recommendation system based on usage history and automatic genre classification. In Consumer Electronics (ICCE), IEEE International Conference pp. 134-135,2015.

[10] S. H. Chang, A. Abdul, J. Chen, and H. Y. Liao, “A personalized music recommendation system using convolutional neural networks approach”. IEEE International Conference on Applied System Invention (ICASI), pp. 47-49,2018,

[11] Wahsheh, Heider AM, and Mohammed S. Al-Zahrani. "Secure real-time computational intelligence system against malicious QR code links." International Journal of Computers, Communications and Control 16.3 (2021).

[12] Sathishkumar V E, Changsun Shin, Youngyun Cho, “Efficient energy consumption prediction model for a data analytic-enabled industry building in a smart city”, Building Research & Information, Vol. 49. no. 1, pp. 127-143, 2021.

[13] Sathishkumar V E, Youngyun Cho, “A rule-based model for Seoul Bike sharing demand prediction using Weather data”, European Journal of Remote Sensing, Vol. 52, no. 1, pp. 166-183, 2020.

[14] Sathishkumar V E, Jangwoo Park, Youngyun Cho, “Seoul Bike Trip duration prediction using data mining techniques”, IET Intelligent Transport Systems, Vol. 14, no. 11, pp. 1465-1474, 2020.

[15] Sathishkumar V E, Jangwoo Park, Youngyun Cho, “Using data mining techniques for bike sharing demand prediction in Metropolitan city”, Computer Communications, Vol. 153, pp. 353-366, 2020.

[16] Sathishkumar V E, Yongyun Cho, “Season wise bike sharing demand analysis using random forest algorithm”, Computational Intelligence, pp. 1-26, 2020.

[17] Sathishkumar, V. E., Wesam Atef Hatamleh, Abeer Ali Alnuaim, Mohamed Abdelhady, B. Venkatesh, and S. Santhoshkumar. "Secure Dynamic Group Data Sharing in Semi-trusted Third Party Cloud Environment." Arabian Journal for Science and Engineering (2021): 1-9.

[18] Chen, J., Shi, W., Wang, X., Pandian, S., & Sathishkumar, V. E. (2021). Workforce optimisation for improving customer experience in urban transportation using heuristic mathematical model. International Journal of Shipping and Transport Logistics, 13(5), 538-553.

[19] Pavithra, E., Janakiramaiah, B., Narasimha Prasad, L. V., Deepa, D., Jayapandian, N., & Sathishkumar, V. E., Visiting Indian Hospitals Before, During and After Covid. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 30 (1), 111-123, 2022.

[20] Easwaramoorthy, S., Moorthy, U., Kumar, C. A., Bhushan, S. B., & Sadagopan, V. (2017, January). Content based image retrieval with enhanced privacy in cloud using apache spark. In International Conference on Data Science Analytics and Applications (pp. 114-128). Springer, Singapore.

[21] Sathishkumar, V. E., Agrawal, P., Park, J., & Cho, Y. (2020, April). Bike Sharing Demand Prediction Using Multiheaded Convolution Neural Networks. In Basic & Clinical Pharmacology & Toxicology (Vol. 126, pp. 264-265). 111 RIVER ST, HOBOKEN 07030-5774, NJ USA: WILEY.

DOI: 10.54254/2755-2721/2/20220656

[22] Subramanian, M., Shanmuga Vadivel, K., Hatamleh, W. A., Alnuaim, A. A., Abdelhady, M., & VE, S. (2021). The role of contemporary digital tools and technologies in Covid‐19 crisis: An exploratory analysis. Expert systems.

[23] Babu, J. C., Kumar, M. S., Jayagopal, P., Sathishkumar, V. E., Rajendran, S., Kumar, S., ... & Mahseena, A. M. (2022). IoT-Based Intelligent System for Internal Crack Detection in Building Blocks. Journal of Nanomaterials, 2022.

[24] Subramanian, M., Kumar, M. S., Sathishkumar, V. E., Prabhu, J., Karthick, A., Ganesh, S. S., & Meem, M. A. (2022). Diagnosis of retinal diseases based on Bayesian optimization deep learning network using optical coherence tomography images. Computational Intelligence and Neuroscience, 2022.

[25] Liu, Y., Sathishkumar, V. E., & Manickam, A. (2022). Augmented reality technology based on school physical education training. Computers and Electrical Engineering, 99, 107807.

[26] Sathishkumar, V. E., Rahman, A. B. M., Park, J., Shin, C., & Cho, Y. (2020, April). Using machine learning algorithms for fruit disease classification. In Basic & clinical pharmacology & toxicology (Vol. 126, pp. 253-253). 111 RIVER ST, HOBOKEN 07030-5774, NJ USA: WILEY.

[27] Sathishkumar, V. E., Venkatesan, S., Park, J., Shin, C., Kim, Y., & Cho, Y. (2020, April). Nutrient water supply prediction for fruit production in greenhouse environment using artificial neural networks. In Basic & Clinical Pharmacology & Toxicology (Vol. 126, pp. 257-258). 111 RIVER ST, HOBOKEN 07030-5774, NJ USA: WILEY.

[28] Sathishkumar, V. E., & Cho, Y. (2019, December). Cardiovascular disease analysis and risk assessment using correlation based intelligent system. In Basic & clinical pharmacology & toxicology (Vol. 125, pp. 61-61). 111 RIVER ST, HOBOKEN 07030-5774, NJ USA: WILEY.

[29] Kotha, S. K., Rani, M. S., Subedi, B., Chunduru, A., Karrothu, A., Neupane, B., & Sathishkumar, V. E. (2021). A comprehensive review on secure data sharing in cloud environment. Wireless Personal Communications, 1-28.