Human Emotion Recognize Using Convolutional Neural Network (CNN) and Mel Frequency Cepstral Coefficient (MFCC)

ISSN: 1533 - 9211

Mannar Mannan J[1], Srinivasan L[2], Maithili K[3], Ramya C[1]
[1]Associate Professor, Dept. of ISE, CMR Institute of Technology, Bengaluru, Karnataka, India. [2]Associate Professor, Dept. of CSE, Dr NGP Institute of Technology, Coimbatore, India.
3Assistant Professor, Dept. of CSE, KG Reddy College of Engineering and Technology, Hyderabad, India
*man.endeavour6381@gmail.com

CORRESPONDING AUTHOR:
Mannar Mannan J.
man.endeavour6381@gmail.com

KEYWORDS:
Convolutional Neural Network, Emotion State Recognition, Speech, Facial Emotion.

Received: 26 May 2023
Accepted: 12 June 2023
Published: 16 June 2023

TO CITE THIS ARTICLE:
Mannar Mannan J, Srinivasan L, Maithili K, & Ramya C (2023). Human Emotion Recognize Using Convolutional Neural Network (CNN) and Mel Frequency Cepstral Coefficient (MFCC). *Seybold Report Journal, 18(4)*, 49-61. https://seybold-report.com

Abstract

Developing an application for definitive human emotion state recognition is yet to optimize to resolve problems in mental health diagnosis. Facial expression together with voice processing are essential to unify the features to enhance the machine learning (ML) model for human emotion state recognition. The facial expression fusion with speech emotion recognition using ML emerges to finalize the emotion state of person. This research paper proposes human emotion state recognition using facial expression fusion with voice recognition from the real time video. In this work, the real time video frames are extracted, segregated into image frames and audio blocks. The CNN for facial expression recognition on the video frames and MFCC for recognize speech emotions is used. Both models executes concurrently and the outcome convergence to produce combined results with 75:25 ratio on face and voice respectively to recognize human emotion state. The result shows that this new model performs better, and the multiple test cases ensured the same compared to individual CNN for facial recognition and MFCC for speech emotion recognition.

INTRODUCTION

Present decade of application development using machine learning (ML) takes a vital role in facial recognition for emotion state observations. Image processing and facial recognition using ML becomes an emerging field of computing sciences which can utilize for resolving the unsolved problems in security, forensic sciences, and health care sector. Applications of merging facial emotion with speech recognition can be used to research on telemedicine to diagnosis mental health. Human trends to expose their emotions by the facial gestures and speech slang. These facial gestures are possible to detect using the application with ML model [6]. Certain facial expressions that humans express on their face and way they talk combinedly can help us to determine their emotion state accurately. These facial expressions can store in digital form in three-dimensional matrix, and possible to retrieve the pattern in the facial expression. These patterns are datasets used to train the ML model for detecting the emotion state of a speaker. Similarly, extracted properties from the speech data is useful to train the ML model to extract the speaker's emotions. Facial expression accumulated with voice recognition can use to optimize ML model to finalize the human emotion state.

Facial expression recognitions and emotion classification achieved using Gobor filters [31], Histogram of Oriented Gradients [32], and Local Binary Pattern [12]. Similarly, for classification, logistics regression analysis, random forest, support vector machine, and K-nearest Neighbour algorithm can also be used to classify the facial expression state [23]. CNN Applications reached almost in all fields in the computing sciences. The CNN model for intrusion detections [29] and electric network analysis [30] are notable applications in non-image processing. Alternatively, in image processing, CNN applications are wide spread in almost all sectors such that, space sciences, geographic information system, transportation, security, and healthcare sectors. The CNN for human emotion state recognition can use in various fields like customer care, security, health care etc., Applications of human emotion recognition used to customer care services [24], health care [25], forensic, security, and intelligence bureau [26][27], etc., Deep learning for facial emotion recognition using CNN [28] and optimization using facial features using facial acting coding system (FACS) achieved the considerable improvement in performance.

In parallel to the facial image processing, speech processing applications are also concurrently increased to recognise emotion state of a speaker. Different flavours of ML models are proposed for processing audio, sound, or speech data such that CNN [20], MFCC, LSTM [17], or SVM. MFCC is one among the finest method for processing sound to classify heart sound [14][15], smart home assistant [16], Multiscale deep convolutional long short-term memory (LSTM) framework for real time speech emotion recognition [17], and CNN for SER [18][19]. Fourier parameters contributed to recognize speech emotions [21], and other speech models [22] are also participated best in this fields.

Facial and voice recognition is the most essential applications of image processing. Human face creates a various expressions depending on their experiences. The objective is to improve feedback systems in hospitals and customer care services. Doctors are excepting to know how the patients are feeling about the treatment they are undergoing, whether they feeling happy or not about the treatment. Continues patient's emotion state monitoring is time, space and cost consuming process. This system can help the doctor with all these aspects. Whenever a patient feels discomfort at any situation, this human emotion state recognition system can be widely used in hospitals to improve the feedback system from the patients about their feeling towards the treatment they are undergoing to their doctor and help them to get immediate care and support during emergency situations and helps the doctors with monitoring their patients digitally from remote environment.

Efficiency of proposed model is verified with metrics precision and recall for quantitative evaluation. Empirical result prove that the proposed method performs better compared to other existing methods. This paper is organized into review of earlier research in this field.

This method of human emotion state recognition is described in section 3, experimental outcomes are discussed in section 4, and conclusion at the end.

Review of Literature

Information and computing technology play a major role in forensic and health care sectors. FER is a growing field in healthcare industry, which is used to recognize the patient or elderly's emotion state. Machine intelligence for facial gesture extraction need large quantity of facial data set to train the model before going to implement. The data set for facial properties acquisition and evaluation is elaborately described [1] for Facial Emotion Recognition (FER). The survey covers latest deep learning models available for FER and their corresponding pros and cons. The existing system for human emotion detection uses only facial expressions which is ultimately not sufficient to accurately predict human emotion state. To overcome this issue, naturally, human emotions are expressing not only on the face, but also through body language and speech.

To improve performance, it is mandatory to incorporate speech data along with facial data. This system is highly required for health care sciences to assess emotion state of patient. Here, to classify the physically disabled and children with autism, facial points, and electroencephalograph (EEG) signals were identified to recognize emotional expressions. Here, identify facial points for CNN, long short-term memory used to classify and recognize facial emotion expressions [2]. Using this method, it is ensured that the real time emotion recognition is possible. EEG signal processing is providing comfortable data for FER. The complete review of literature regarding emotion recognition using EEG has been analysed and the role of deep learning and shallow machine learning is categorized. EEG data analysis and acquisition, EEG dataset management and data pre-processing are addressed to recognize FER [3] and human brain signals are trapped to recognize human emotion state [13].

Another AI method for facial recognising using deep learning is CNN. The process of background removal can provide better facial features vector extraction. The CNN used for two different purpose, such that, first one is to removes background from the picture, and the second is for facial features vector extraction. The expression vector used to recognize five different types of facial expressions [4]. Facial recognition can be achieved by the different approach from the CNN by adjusting the approach. Extracting local features from the face images using local gravitational force description and the retrieved descriptor is fed into a CNN. The DCNN with two branches which explores geometric features such as edges, curves, and lines whereas holistic features are extracted by the second branch. The score will fusion to final classification score [5]. A system uses a single recognition system cannot give accurate results for all situations. These systems use live face recognition to detect emotions from time to time on a lively basis. The emotions during live recognition might fluctuate without giving a steady emotion throughout. Many time the system fails to detect face for recognition.

Face recognition system dominating on mobile applications. Emotion-aware mobile applications are increased because of its smart features, attractions, and user usability. To use those applications, an emotion recognition system should be in real time and highly accurate. In our model, video frames are segregated and a face detection module is applied to extract the face regions in the frames. The dominant region is then fed into a Gaussian mixture model-based classifier for emotion classification [7]. Likewise, image processing for facial emotion recognition appended in the Humanoids. Robots are developed to identify emotion

ISSN: 1533 – 9211

state of a person to improve human–robot interaction (HRI)][8]. The robot with camera eyes can observe user's facial and recognize users' emotions, and respond appropriately.

The CNN model used to speech emotion recognition [9]. Spectral subtraction and linear predictive coding with MFCC voice denoising is achieved to recognise speech emotion recognition. Alternatively, A hybrid model [10] has been proposed to enhance the SER performance, which comprises advantages of different models to acquire better performance. In the model, deep neural network (DNN), CNN, and recurrent neural network (RNN) have been taken to finding out the best combination for SER. Integration of distinct methods for SER can be used to aggregate the different features for improving the performance compared to the individual SER methods with different parameters. DNN, CNN, and RNN has been aggregated to recognize four emotions such as angry, happy, neutral, and sad [10]. A common framework proposed to use interdependent features of DNN, CNN, and RNN to obtain generalized features for SER, and have achieved 57.1 % on weighted accuracy and 58.3% on unweighted accuracy over IEMOCAP dataset. SER can be achieved through various existing algorithms along with different parameters and features. RNN classifier with Mel frequency cestrum coefficients (MFCC), and modulation spectral features (MSFs) on Berlin and Spanish data set performs better compared to Multivariate Linear Regression (MLR) and SVM [11].

Human Emotion State Recognition

During the digital era, machine intelligence is ubiquitous in almost all the fields. Deep learning in artificial intelligence (AI) reached the next stage in advancement of emotion state recognition of a speaker in real-time basis. The emotion state of a person can be recognized by the machine in two ways, such that, either facial or speech emotion recognition. To enhance the performance of the AI system for human emotion state recognition, it is necessary to fusion the both facial and speech recognition together to get high precision outcomes. The prominent work of this model is to fusion the outcomes of CNN for facial expression and MFCC for speech emotion recognition in certain ratio to finalize the emotion state. The independent features of CNN and MFCC has been utilized without altering its functional properties for quantitative and qualitative feature extraction from the dataset.

General Considerations

1. Offline videos used to test the system.
2. Only English language is considered testing.
3. Facial and Speech expressions are independently handled.

Initially, video data is segregated into image frames and voice data. Here, for *"Speech Recognition"*, *"MoviePy"* used for extracting speech data from the video and *"OpenCV"* used extract video frames. The CNN model applied to recognize facial expressions on video frames and the MFCC applied to recognize speech emotions states. Architecture of this model given in figure 1.

Vol 18. No. 04. 2023

ISSN: 1533 – 9211

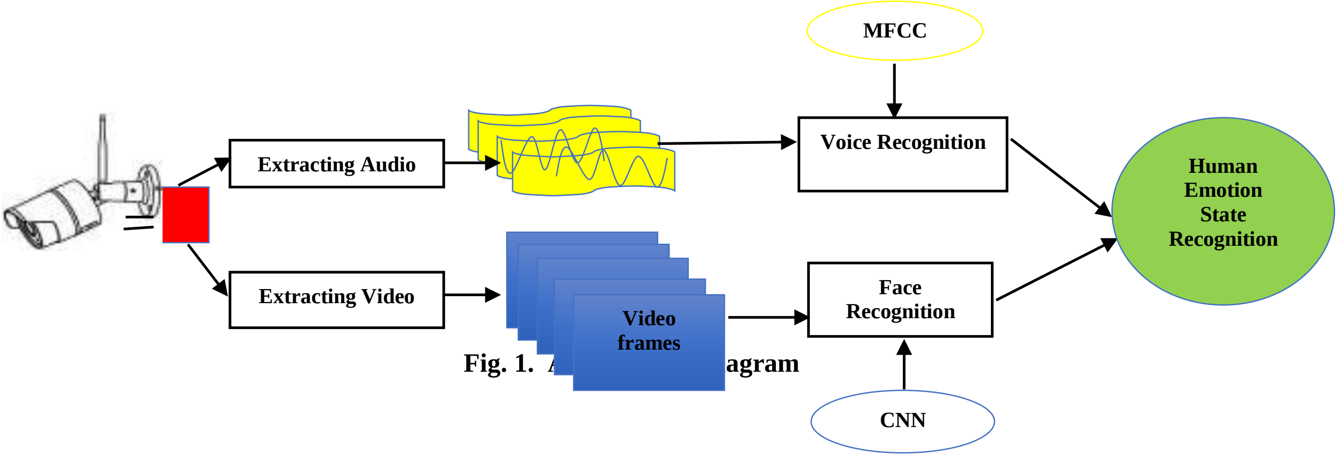


Fig. 1. agram

Facial Emotion Recognition using CNN

The outperformance features of CNN help the machine to learn and classify the images. The CNN consists of multiple convolutional layers followed by pooling layers to extract facial properties in local pattern, and fully connected layers for global pattern recognition. The output layer with neurons produces the features corresponding to the emotions to be recognized.

Initially the video frame images are converted into gray-scale images, later extracting important pixels from the image and summarize using filters/kernels. To speed up the process, the max-pooling layer used to reduce the dimensionality. The max-pooling filter passed through the data and highest value is extracted as final feature. The final step in CNN model is flattening which is extracting various emotion features from the image and classification.

Here, the input is defined in *'f'* and the trained pattern is *'h.'* The results matrix marked as *G[m,n]* shown in equation 1 below.

$$G[m,n] = (f * h)[m,n] = \sum_{j} \sum_{k} j[j,k]\, f[m-j,\; n-k] \qquad ---(1)$$

The output matrix dimension considering of padding and stride, and is is calculated using given equation

$$n_{out} = \left[\frac{n_{in} + 2p - f}{s} + 1\right] \qquad ---(2)$$

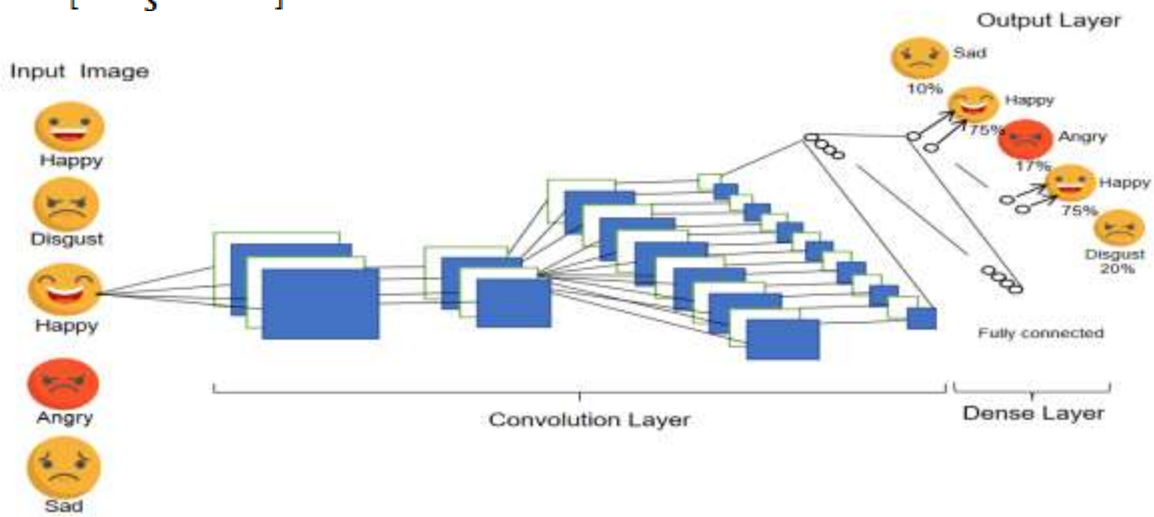


Fig. 2. CNN model for facial emotion recognition

The resultant matrix calculated using the equation 3, given below. It classifies the different facial emotions and compare with the labeled facial emotions stored in the matrix to finalize the emotion state.

$$dA += \sum_{m=0}^{n_h} \sum_{n=0}^{n_w} W.dZ[m,n] \qquad ---(3)$$

ISSN: 1533 – 9211

The CNN model developed to classify emotion state of speaker. The output matrix ensures that proposed CNN model completes the design objectives by classifying emotion state of speaker.

Speech Emotion Recognition MFCC

Speech data is converted to digital form at 44.1 kHz using sampling process shown in figure 1. To extract various emotion patterns, the voice data divided into frame. Each frame is windowed using Hamming window.

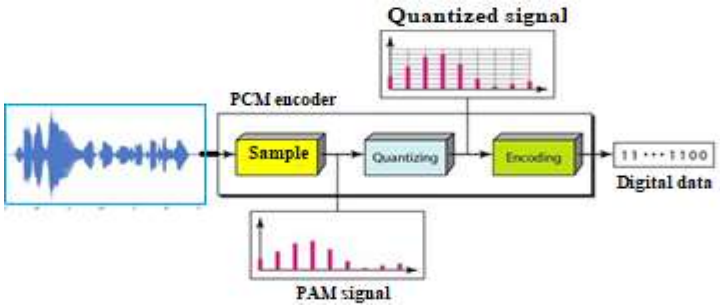


Fig. 3. Analog data encoding

Later, all frames are converted to frequency domain using a short time Fourier Transform. Meanwhile, a certain number of sub-band energies are calculated using a Mel filter bank, which is a nonlinear- scale filter bank that imitates a human's aural system and sub-band energies are calculated. Finally, the MFCC is computed by an inverse Fourier Transform. The process of MFCC shown in figure 4.

$$mel(f) = 1127 \ln\left(1 + \frac{f}{700}\right) \quad ---(4)$$

$$energy = \sum_{t=t1}^{t2} x^2[t] \quad ---(5)$$

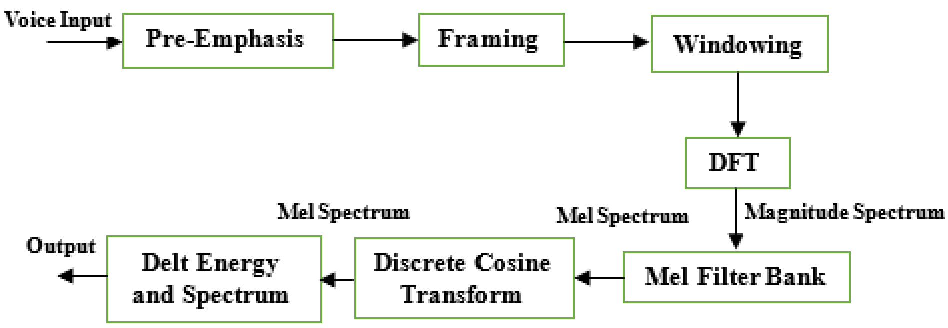


Fig. 4. MFCC Process

First the voice data is divided into frame. Each frame is windowed using Hamming window

$$W(n) = 0.54 - 0.46 \cos\frac{2\pi * n}{N-1} \quad ---(6)$$

Where *W(n)* – value of the window at sample index *'n'*, N – length of window, and π – is a constant, its value equal to 3.14. In continue to this, the data frames are converted to frequency domain using a short time Fourier Transform (STFT), The mathematical formula for the STFT is as follows:

$$STFT\,(t,\omega) = \int \{-\alpha\}^{\{\alpha\}} \mathrm{X}\,(\tau - \mathrm{t}) e^{\wedge}(-j\omega\tau)\}\, \mathrm{d}\tau \quad ---(7)$$

Where, *'t'* is the time index of the window, *'ω'* is the frequency index, and *'x(τ)'* is the original signal *'w(τ - t)'* is a window function centered at time *'t'*, *'j'*, is the imaginary unit.

ISSN: 1533 – 9211

Significant number of '*sub-band'* energies are calculated using a '*Mel filter bank'*, which is a nonlinear- scale filter bank that imitates a human's aural system.

$$Mel\ (f) = 2595 * Log\ 10\left(1 + \frac{f}{700}\right) \quad - - - \quad (8)$$

Where *'f'* is the frequency in Hz.

Convergence of CNN and MFCC

The prominent work of this research is to improve performance of human emotion recognition system, which is integrating facial expression and voice data to get a unified, enhanced emotion state recognition system. The unique features of CNN and MFCC pair are summed, and the highest value for the corresponding emotion state is finalized.

Table 1. Weight Based Emotion State Recognition

Emotion State	Facial Expression CNN	Value	Emotion State	Voice Expression MFCC	Value	Augmented Maximum Weight	Human Emotion State
1	Happy	62%	1	Happy	14%	76%	
2	Sad	34%	2	Sad	12%	46%	Neutral
3	Angry	28%	3	Angry	10%	38%	
4	Neutral	54%	4	Neutral	23%	77%	

Table 1 shows the human emotion state for final emotion weight calculation.

To finalize the ultimate emotion state of a speaker, two sets labelled emotions states for speech and facial expressions are kept in two different arrays.

Facial Emotion State (FES) = [' Happy' , ' Sad' ,' Angry', 'Neutral' , …..]

Speech Emotion State(SES) = [' Happy' , ' Sad' ,' Angry', 'Neutral', …..]

FES[i]=max(CNN(IS))

SES[i]=max(MFCC(VS))

$$\boldsymbol{HESR = max\{(FES1 + SES1), (FES2 + SES2), \ldots\ldots(FEEn + SESn)\}} \quad - - - (9)$$

$$\boldsymbol{HESR = \max\left(\sum_{i}^{n}(FESi + SESi)\right)} \quad - - - - \quad (10)$$

The highest augumented score or weight of the FES and SES considered for select the appropriate emotion state lable to finalize.

$$\mathbf{HESR = max(CNN(Image_i) + MFCC(Audio_i)} \text{ - - - - - - } \mathbf{(11)}$$

This research and experiments gives the highest importance of facial expression because human emotions will immediately reflect on the face, and which is difficult to control however voice emotion control.

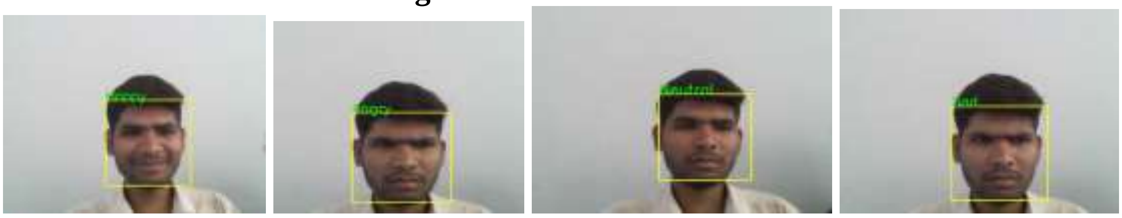
Results and Discussions

OpenCV is a computer vision object that is implemented as a library files, which is used in this implementation using Python language. We segregated the video frames and speech data separately. Later, facial feature extraction and classification, removal of irrelevant visual components is aided by facial detection are accomplished, which is key step in emotion identification. The facial components are used to extract emotion detection characteristics. Sample code for facial emotion detection is given in figure 5. Speech emotion recognition achieved using MFCC by implementing *'librosa'* ,*'sklearn'*, and *'Kerosene'*. All the library files along with MFCC function used to classify emotions state of a speaker. The convergence of the results from both the models detected the emotion state of a person. To experiment the proposed model, OpenCV has been used. The OpenCV provides set of predefined library files for classification. The *'keras'* , *'librosa'*, and *'tensorflow'* has been used here to recognize voice emotion.

ISSN: 1533 – 9211

```
face_calssifier=cv2.CascadeClassifier("path")
Classifier = load_model(model)
Emotion_labels=['Angry', 'Disgust', 'Fear', 'Happy', 'Neutral', 'Sad', 'Surprise']
Cap=cv2.videoCapture(0)
Prediction= classifier.predict(x)[0]
```

Fig. 5. Video Frames Extraction



X. (a) Happy x. (b) Angry x.(c) Neutral x.(d) Sad

Fig. 6. Facial Emotion Recognition

Figure 6 shows the various facial emotion states while testing our proposed system.

```
mfcc=pd.Series(demo_mfcc)
pit = pd.Series(demo_pitch)
mag=pd.Series(demo_mag)
c=pd.Series(demo_chrom)
```

Fig. 7. Sample Code for Speech Emotion State Extraction

For speech emotion recognition, we build the model using ***'librosa'*** and ***'sklearn'*** and ***'RAVDESS'*** data used to train and test for different speech modulation irrespective of male, female, or children.

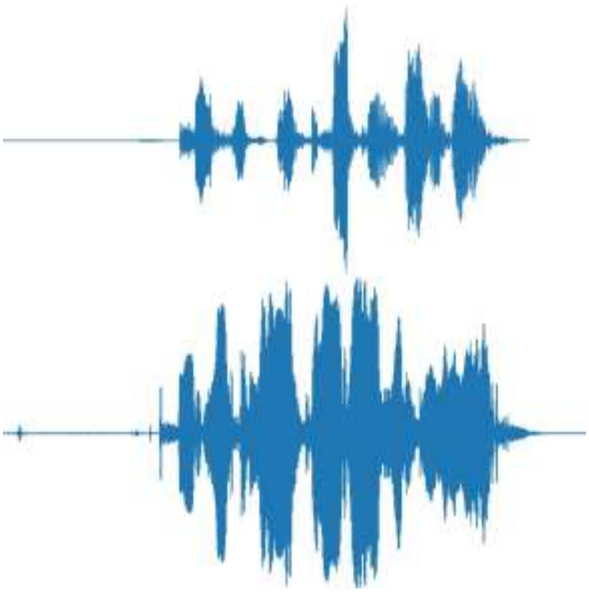


Fig. 8.a. Happy Emotion State **Fig. 8.b Neutral Emotion State**

Fig. 8. Speech Emotion State Extraction

ISSN: 1533 – 9211

The score for the emotion state using CNN FES = [60, 35, 40, 50] and the MFCC is SES= [10, 20, 10, 25]. The addition of two outcomes from CNN and MFCC combined result shows that CNN conveys *'Happy'* but the MFCC conveys *'Neutral.'* To integrate the result, the maximum probability is classified into *'Neutral.'* Figure 9 shows the difference between the training data set and the validation.

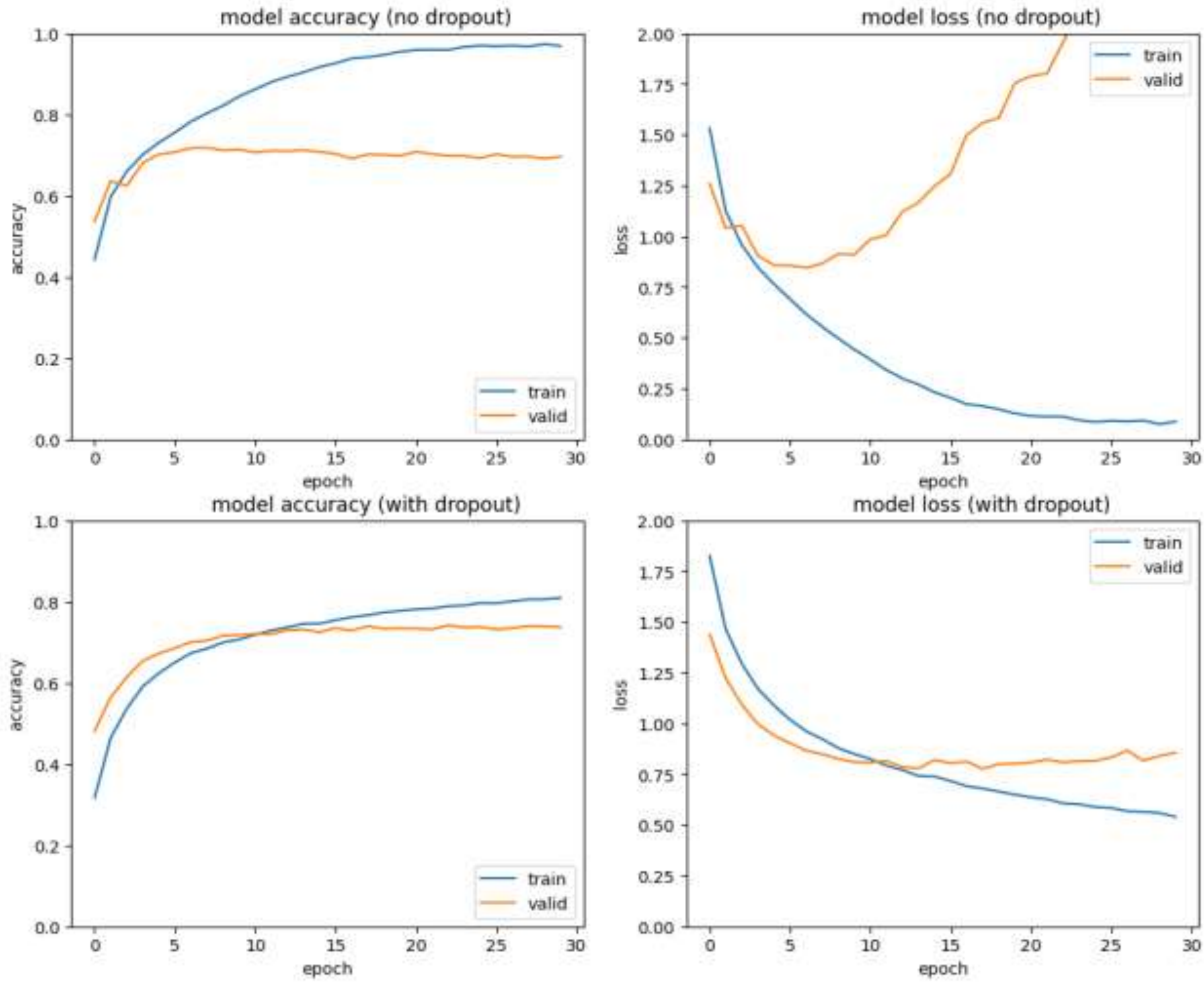


Fig. 9 Epoch history of CNN+MFCC for human emotion state recognition

Conclusion

Machine Learning can address unsolved problems in human computer interface specifically in the health and elderly care. In this proposed research work, the performance of machine learning for human emotion state recognition is improved by convergence of CNN for facial expression recognition and MFCC for voice recognition. Here, the fusion of CNN and MFCC have provided accurate results with the combination of 75:25 ratio. The facial data stored on the vector and the voice data stored in the linear array. The outcome of the proposed work ensures that the proposed model performs better to recognize the emotion state of a person compare to other existing models.

ISSN: 1533 – 9211

COMPETING INTERESTS

The authors have no competing interests to declare.

Authors Affiliation

Dr. J. Mannar Mannan, working as an Associate Professor, Department of Information Science and Engineering, CMR Institute of Technology, Bengaluru, Karnataka, India. He has around 18 years of experience in various reputed engineering colleges and University. He has published 24 articles in reputed journals and 13 international conferences. His research interest are image processing, semantic web, ontology and information retrieval, natural language processing, IoT, sensor and ad-hoc networks, man machine learning.

Dr. L. Srinivasan, is working as an Associate Professor at Dr N.G.P Institute of Technology, Coimbatore, Tamil Nādu. He Received his Ph.D., from Anna University Chennai, MTech (IT) from Anna University Coimbatore and B.E(IT) from V.L.B Janakiammal College of Engineering & Technology, Coimbatore in 2020,2011 and 2004, respectively. His current research includes Text mining, Artificial Intelligence & Machine Learning and IOT. He has published 7 papers in reputed journals with high impact factor leading journals, conference proceedings. He had Filed 6 Indian Patents and been granted 1 Australian Patent. He had Published 2 books titled 'Learn Python the Easy Way 'and 'Mobile App Development for Beginners'. He has more than fifteen years of professional experience in Various Engineering Colleges in Tamil Nādu, Bengaluru India.

Dr Maithili, Working as an Associate Professor, KG Reddy College of Engineering and Technology, Hyderabad, India. She as completed PhD and has around 7 years of teaching experience at various reputed colleges in Pondicherry and Telungana. Her research towards machine learning and image processing. She as published 4 research papers in various reputed international journals.

Dr. Ramya C, working as an Assistant Professor, Dept. of Information Science and Engineering, CMR Institute of Technology, Bengaluru. She has completed the research (Ph.D) on swarm optimization for information retrieval at VTU, Belagavi, Karnataka. Her research interests are information retrieval, IoT, and machine learning. She has published 20 research articles in various reputed international journals.

ISSN: 1533 – 9211

HOW TO CITE THIS ARTICLE:

Mannar Mannan J, Srinivasan L, Maithili K, & Ramya C (2023). Human Emotion Recognize Using Convolutional Neural Network (CNN) and Mel Frequency Cepstral Coefficient (MFCC). *Seybold Report Journal, 18(4),* 49-61. https://seybold-report.com

REFERENCES

[1] Shan Li and Weihong Deng, "Deep Facial Expression Recognition: A Survey", IEEE Transactions, May 4, 2020.

[2] Aya Hassouneh , A.M. Mutawa, and M. Murugappan, "Development of a Real-Time Emotion Recognition System Using Facial Expressions and EEG based on machine learning and deep neural network methods", Informatics in Medicine Unlocked 20 (2020) 100372. doi.org/10.1016/j.imu.2020.100372.

[3] Rabiul Islam M, Ohammad Ali Moni, Milon Islam, Rashed-Al-Mahfuz, Saiful Islam, Kamrul Hasan, Sabir Hossain, Mohiuddin Ahmad, Shahadat Uddin, Aam Azad, Salem A. Alyami, Atiqur Rahman Ahad, and Ietro Lio, "Emotion Recognition From EEG Signal Focusing on Deep Learning and Shallow Learning Techniques" ,IEEE Open access Journal, VOLUME 9, 2021.

[4] Ninand Mehendale, "Facial Emotional Recognition Using Convolutional Neural Networks (FERC)", Springer, Applied Sciences, 2020.

[5] Karnati Mohan, Ayan Seal, Ondrej Krejcar, and Anis Yazidi, "Facial Expression Recognition using Local Gravitational Force Descriptor based Deep Convolution Neural Networks", IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, 2020.

[6] CHIRAG DALVI, MANISH RATHOD, SHRUTI PATIL, SHILPA GITE, AND KETAN KOTECHA, "A Survey of AI-Based Facial Emotion Recognition: Features, ML & DL Techniques, Age-Wise Datasets and Future Directions", IEEE Multidisciplinary, Open Access Journal, Dec. 23, 2021. DOI:0.1109/ACCESS.2021.3131733

[7] Dhwani Mehta, Mohammad Faridul Haque Siddiqui, and Ahmad Y. Javaid, "Recognition of Emotion Intensities Using Machine Learning Algorithms: A Comparative Study", sensors, April 2019.

[8] D. Yangm Abeer Alsadoon, P.W.C. Prasad, School of Computing and Mathematics, Charles Sturt University, Sydney, Australia, "An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment", International Conference on Smart Computing and Communications

[9] Kalyanapu Jagadeeshwar, T. Sreenivasarao, Padmaja Pulicherla, K. N. V. Satyanarayana, K. Mohana Lakshmi, and Pala Mahesh Kumar, "ASERNet: Automatic speech emotion recognition system using MFCC-based LPC approach with deep learning CNN", International Journal of Modeling, Simulation, and Scientific Computing, Vol. 11, 2022. doi.org/10.1142/S1793962323410295.

[10] Zengwei Yao, Zihao Wang, Weihuang Liu, Yaqian Liu, Jiahui Pan, "Speech emotion recognition using fusion of three multi-task learning-based classifiers: HSF-DNN, MS-CNN and LLD-RNN", Speech Communication, Vol. 120, pp-11-19, 2020.

[11] Leila Kerkeni, Youssef Serrestou, Mohamed Mbarki, Kosai Raoof, and Mohamed Ali Mahjoub, "Speech Emotion Recognition: Methods and Cases Study", In Proceedings of the 10th International Conference on Agents and Artificial Intelligence (ICAART 2018) - Volume 2, pages 175-182.

[12] Md. Abdur Rahim, Md. Najmul Hossain, Tanzillah Wahid and Md. Shafiul Azam, "Face Recognition using Local Binary Patterns (LBP)", Global Journal of Computer Science and Technology Graphics & Vision, Vol. 13, Issue 4, 2013.

[13] Essam H. Houssein, Asmaa Hammad, and Abdelmgeid A. Ali, "Human emotion recognition from EEG-based brain–computer interface using machine learning: a comprehensive review", Neural Computing and Applications (2022) 34:12527–12557. doi.org/10.1007/s00521-022-07292- 4.

[14] Jinghui Li, Li Ke, and Qiang Du, "Classification of Heart Sounds Based on the Wavelet Fractal and Twin Support Vector Machine", Entropy (Basel). 2019 May 6;21(5):472. doi: 10.3390/e21050472. PMID: 33267186; PMCID: PMC7514961.

[15] Bahreini, Mahbubeh, Barati, Ramin and Kamaly, Abbas, "Heart Sound Classification Based on Fractal Dimension and MFCC Features Using Hidden Markov Model", 2021. 10.21203/rs.3.rs-1207404/v1.

[16] R. Chatterjee, S. Mazumdar, R. S. Sherratt, R. Halder, T. Maitra and D. Giri, "Real-Time Speech Emotion Analysis for Smart Home Assistants," in IEEE Transactions on Consumer Electronics, vol. 67, no. 1, pp. 68-76, Feb. 2021, doi: 10.1109/TCE.2021.3056421.

[17] S. Zhang, X. Zhao and Q. Tian, "Spontaneous Speech Emotion Recognition Using Multiscale Deep Convolutional LSTM," in IEEE Transactions on Affective Computing, vol. 13, no. 2, pp. 680-688, 1 April-June 2022, doi: 10.1109/TAFFC.2019.2947464.

[18] A. B. Abdul Qayyum, A. Arefeen and C. Shahnaz, "Convolutional Neural Network (CNN) Based Speech-Emotion Recognition," IEEE International Conference on Signal Processing, Information, Communication & Systems (SPICSCON), Dhaka, Bangladesh, 2019, pp. 122-125, doi: 10.1109/SPICSCON48833.2019.9065172.

[19] Xu Dong An and Zhou Ruan, "Speech Emotion Recognition algorithm based on deep learning algorithm fusion of temporal and spatial features", Journal of Physics: Conference Series, Vol. 1861. DOI 10.1088/1742-6596/1861/1/012064.

[20] S. Zhang, S. Zhang, T. Huang and W. Gao, "Speech Emotion Recognition Using Deep Convolutional Neural Network and Discriminant Temporal Pyramid Matching," in IEEE Transactions on Multimedia, vol. 20, no. 6, pp. 1576-1590, June 2018, doi: 10.1109/TMM.2017.2766843.

[21] K. Wang, N. An, B. N. Li, Y. Zhang and L. Li, "Speech Emotion Recognition Using Fourier Parameters," in IEEE Transactions on Affective Computing, vol. 6, no. 1, pp. 69-75, 1 Jan.-March 2015, doi: 10.1109/TAFFC.2015.2392101.

[22] Taiba Majid Wani, Teddy Surya Gunawan, Syed Asif Ahmad Qudri, Mira Kartiwi and Eliathamby Ambikairajah, “A Comprehensive Review of Speech Emotion Recognition Systems”, IEEE Access, 2021. DOI: 10.1109/ACCESS.2021.3068045.

[23] Murugappan M, Mutawa A.M, Sai Sruthi , Aya Hassouneh, Ali Abdulsalam, Jerritta Selvaraj, and R. Ranjana, “Facial Expression Classification using KNN and Decision Tree Classifiers” 4th International Conference on Computer, Communication and Signal Processing (ICCCSP), Pp. 1-6, 2020.. 10.1109/ICCCSP49186.2020.9315234.

[24] Renigier-Biłozor, M., Janowski, A., Walacik, M. et al. Human emotion recognition in the significance assessment of property attributes. J Hous and the Built Environ 37, 23–56 (2022). https://doi.org/10.1007/s10901-021-09833-0.

[25] Nazish Azam, Tauqir Ahmad and Nazeef Ul Haq, “Automatic emotion recognition in healthcare data using supervised machine learning”, PeerJ Computer Science, Dec. 2021. DOI 10.7717/peerj-cs.751.

[26] Lena Podoletz, “We have to talk about emotional AI and crime”, AI & Society, Springer, 2022.

[27] Mayank Kumar Rusia and Dushyant Kumar Singh “A comprehensive survey on techniques to handle face identity threats: challenges and opportunities”, Int. Journal of Multimedia Tools and Applications, Springer, June 2022.

[28] Tarun Kumar Arora,Pavan Kumar Chaubey, Manju Shree Raman, Bhupendra Kumar, Yagnam Nagesh, P. K. Anjani, Hamed M. S. Ahmed, Arshad Hashmi, S. Balamuralitharan , and Baru Debtera, “Optimal Facial Feature Based Emotional Recognition Using Deep Learning Algorithm”, Computational Intelligence and Neuroscience, Hindawi, 2022. doi.org/10.1155/2022/8379202.

[29] Zhiwei Gu,Shah Nazir, Cheng Hong,and Sulaiman Khan, “Convolution Neural Network-Based Higher Accurate Intrusion Identification System for the Network Security and Communication”, Security and Communication Networks, Hindawi, 2020. doi.org/10.1155/2020/8830903.

[30] Miguel Ramirez-Gonzalez, Felix Rafael Segundo Sevilla, Petr Korba, and Rafael Castellanos-Bustamante, “Convolutional neural nets with hyperparameter optimization and feature importance for power system static security assessment”, Electric Power Systems Research, Vol. 211, 2022.

[31] Kunika Verma and Ajay Khunteta, “Facial expression recognition using Gabor filter and multi-layer artificial neural network”, ICICIC, 2017.

[32] Chang Shu, Xiaoqing Ding and Chi Fang, “Histogram of the oriented gradient for face recognition”, IEEE Tsinghua Science and Technology, Vol. 16, Issue 2, 2011.