Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Human Emotion Recognition Using CNN

Vijayalaxmi Kadrolli, Dr.Varsha Bodade, Lavesh Jain, Premkothawle, Ganeshshinde, Shubham Saroj

Department of Information Technology, Terna Engineering College, Nerul, Navi Mumbai, India
Email: udachanv@gmail.com

In an interpersonal relationship, human emotion recognition plays an important role. Through facial expressions, speech, and hand gestures of the body, emotions are reflected. Interaction between human and machine communication has high importance for emotion recognition. It is only through effective communication, verbal or non-verbal, that these interactions can take place. In non-verbal communication, one of the essential means is emotion recognition, which helps discern the mood and state of the communicator. Emotions help to collaborate with machines making communication more natural. Selecting a facial database and the image used to extract facial features is one of the ways to compare. In applications involving human-computer interaction, and in image processing, recognizing emotion using images has emerged as a field to explore throughout. Human emotion recognition system includes the following components preprocessing and acquisition of image, feature extraction, classification, face detection, and then emotions are categorized the system gives the emotion on the video screen in real-time. In this system, the focus is on live real-time images the webcam captures. The systems include emotions like Happiness, Nervous, Surprise, Sadness, Fear, Anger and Disgust which are accepted by everyone.

Keywords: human emotion, feature extraction, recognition.

1. Introduction

Human communication relies heavily on language, It is augmented by a variety of expressive information in most encounters, including gestures, verbal inflections, and facial emotions. A significant aspect of our life is interaction with the computer and human communication with mediated concept, and even if they still having the lack the very basic skills for identifying and reacting to the non-verbal cues related with attitudes, emotions cases, and mental states that humans take for consideration alwayss in human communication and reasoning. Emotion Recognition is one of the essential non-verbal means by which the communication occurs, it helps recognize mood and state of a person. Human emotions can be classified as etc. These emotions are very subtle and can vary from person to person. Using facial expressions recognizing emotion is a main feature, contempt, disgust, happy, anger, surprise, neutral element human communication CNN have confirmed to be very beneficial in classification and pattern recognition. The dataset of human emotion could always be a very decent example to study the nature of classification algorithms and how they perform for different types of datasets. Emotion recognition also be taken through different mechanisms such as speech, facial expression, body gesture, etc. Around 55% of communication occurs through facial expressions, according to some psychologists. Emotion recognition system discover applications in numerous remarkable areas. Emotion recognition over facial expression is becoming widespread due its various applications like robotics where interaction between machines and humans is significant. In recent developments in robotics, human robots, the persistence of the need for a powerful expression recognition system is evident. Emotion recognition plays an important role in motion recognition and, in turn, helps in the design of meaningful or responsive human-computer interface (HCI). In addition to robotics and affect-sensitive human-computer interface (HCI), the discovery of emotion recognition systems is used in many other fields such as telecommunications, educational software, video games, automotive safety, etc.

HEIGHLIGHTS

- To identify different emotions of a human being.
- To determine whether people can certainly express emotions when they see they are cooperating with an emotion detecting computer.
- To assess whether emotion detection can lead to developments in subjective and/or objective procedures of system usability.
- Deliver Human Factors rules on the deployment of emotion recognition technology so that the developers can better meet the needs of true users.
- To design a user-friendly interface that can be easily used by third party applications. Implementation of a model that can predict the results more accurately and can give better performance in decision making and speed.

2. Literature Survey

The problem is to remove the inefficiency and reduce the time required in feature extraction and labeling because the majority of machine learning algorithm need effective feature extraction for proficiently recognizing the emotion. Though, Manual process the feature extraction can take a lot of time and error prone process, Additionally, the manual categorization technique runs the risk of classifying emotions incorrectly, making it a challenging endeavour.

Ref	Technology Used	Features
[1]	Active Appearance Model, FACS Framework.	A framework for the classification of emotional, based on still image of the face was done.
[2]	Affective computing, Physiological, Machine learning, signals, Signal processing.	The proposed system implemented for individuals going through stress during their working hours giving them music therapy.
[3]	Neural Network for a classifier.	This system is independent of factors such as gender, ethnicity, background, and ages.
[4]	Granger causality test – Time Series.	Accuracy in spontaneous expressions in different situations.
[5]	Multi SVM classifier for classification of emotions.	Works for both still images and video frames.
[6]	Conditional-generative-adversarial-network (CGAN based) framework.	The system can automatically extract the features of face expressions.
[7]	Convolutional Neural Network model	Proposed system is Emotions in Context Database (EMCO), images containing people in context in non-controlled environment
[8]	Electroencephalogram (EEG), k-nearest neighbor (KNN), naive Bayesian (NB), support vector machine (SVM) and random forest (RF)).	The use of physiological signals can lead to more objective and reliable emotion recognition. (EEG) signals respond to fluctuations of affective states more.
[9]	Deep belief networks (DBN),convolution neural network (CNN), recurrent neural networks (RNN) .	Advancements in emotion research using multimodal signals, Feature extraction and classification using deep learning.
[10]	In CNN and RNN design, recurrent neural networks (RNN) are embedded with convolutional neural networks (CNN).	Face emotion recognition model , hybrid deep CNN and RNN model. combination of the two types of neural networks (CNN-RNN) cloud considerably enhances the final outcome.
[11]	OpenCV, supported by a ResNet-34 architecture,	Face recognition through NVIDIA's state of the art Jetson Nano. supported by a ResNet architecture, for achieving better accuracy than the previously models.
[12]	Xception CNN architecture and the K-fold-cross-validation strategy	Xception model pre trained on ImageNet database for objects recognition .
[13]	Convolutional neural networks (CNN)	Face detection using Haar Cascades, normalization and emotion recognition using CNN
[14]	Convolution Neural Network (CNN), deployed on Raspberry Pi3 B+ for human robot interaction.	An intelligent network capable of real-time emotion recognition from multiple faces using deep learning.
[15]	Gabor filters for feature extraction and (CNN) classification.	"A deep learning based framework is proposed for human emotion recognition, Automatic extraction of these emotions from the face images"[15].

3.System Architecture and Methodology

To address the aforementioned issue, the initial step is to do a face registration to collect the basic feature points needed to detect facial muscle movements. Due to the possibility of multiple rotations, the faces following the detection stage tend to be deteriorated in terms of identification accuracy. As a result, it is critical to create the image by gathering landmarks, which are the places of major muscle movements while producing a facial expression. The key locations that determine the contraction of the face muscles are known as action units (AUs), and they include the eyes, brows, mouth, and nose. The motion or position information of the feature points is acquired in the feature extraction process to isolate the other characteristics that can detect face emotions.

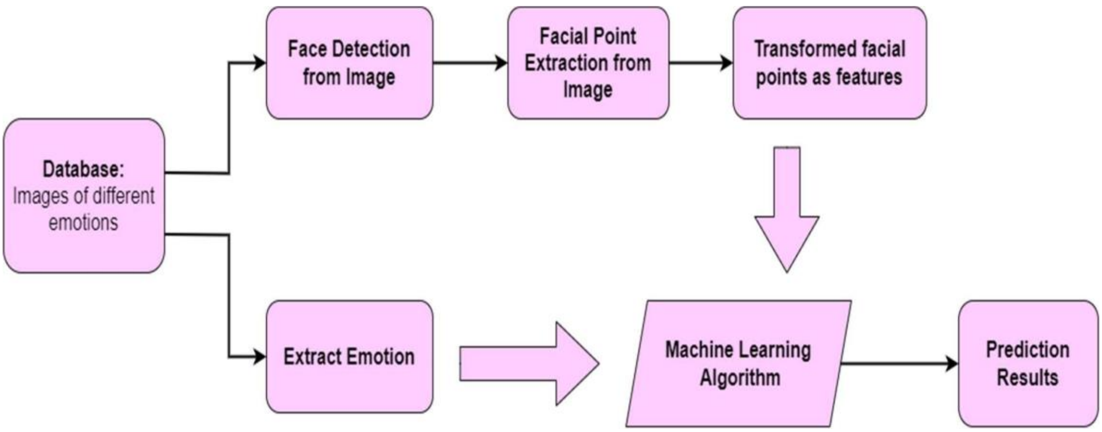


Fig.1.System block diagram

The number of training example models was plotted by batch size. Instead of doing the hard work (computation or memory intensive) as a whole, ImageDataGenerator divides it into batches and works on it as a batch. This way, the parent function that calls the child function does not have to wait until the parent function has finished processing, but can be executed on the fly. It is not feasible to load a large data set into memory at the same time; If we create a generator data file, we can read the moving images when they are used for training. Keras provides a data generator for image data sets. The advantage of using ImageDataGenerator is that it will generate batches of augmented data.

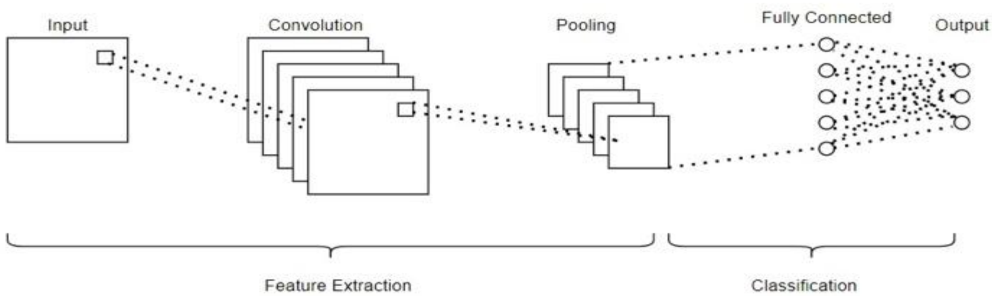


Fig.2.CNN diagram

Face Detection from Image: Real time face data is recorded from the web browser and further sent to pre-processing.

Facial Point Extraction from Image: The live video data is then pre-processed and facial points are extracted from Images.

Transformed facial points as features: The facial points are then transformed to features which would be used as data inputs for the machine learning model.

Extract Emotion: Dataset containing emotions-based images are used to train the model. Different emotions like happy, sad, nervous, fear, angry, disgust, surprise.

Machine Learning Algorithm: CNN algorithm is used which analyses images by taking input, transforming it and then outputting the results as it has filters which detect patterns like edges, shapes, objects, etc.

Prediction Results: The predicted results are then displayed within the video input frame in real-time

Dataset: The data consists of grayscale images of faces measuring 48x48 pixels. Each image is automatically recorded so that the faces are approximately centered and take up approximately the same amount of space. Each face is classified into one of seven categories based on the emotion shown by its facial expression (0 = angry, 1 = disgust, 2 = afraid, 3 = happy, 4 = sad, 5 = surprise, 6 = neutral). The training set contains 28,708 samples, while the public test set contains 3,579 samples. There are approximately 35,000 images in the data set showing different emotions, namely facial expressions of anger, happiness, sadness, neutrality, surprise, disgust, and fear.

1.Angry faces emotion

2. Happy emotions

3. Sad faces

4. Angry faces

- Artificial Neural Network – ANN

Artificial Neural Network (Models used in deep learning are called as ANN). ANN are computing systems that are inspired by the brain's neural network. These networks are based on collection of connected units called neurons which transmit and receive signals from each other. Neurons are organized in layers.

- Input Layer
- Hidden Layer
- Output Layer (Neurons are basically some functions which perform calculations)

To Build an ANN we can use Neural Network API called as Keras. In Keras we can build sequential model. Keras defines this sequential models as a linear stack of layers. (Neurons are also organized in layers)

Dense -> Type of layer Dense(No. of neurons, Shape of data been passing into our model, activation funtion)

Layers in an ANN:-

- Dense (or fully connected) layers
- Convolution Layers (can be used where model is with image data)
- Pooling Layers
- Recurrent Layers (can be used where model is with time series data)
- Normalization Layers
- Others....

CNN concept

It is an ANN popularly used for analyzing images. They can also be used for other data analysis or classificatin problems. It picks out patterns and makes sense out of them making it is usefull in image analysis. CNN have convolutional layers (CL) which receives input, transforms it in someway and then outputs it. CL have filters which detect patterns like edges, shapes, objects, etc.

A CNN is an algorithm which takes an input image, prioritises a few features or objects therein, and can discriminate between images. Preprocessing prerequisite in a CNN is much lesser than other classification algorithms. The design of a CNN is modelled by the association of the visual cortex and is analogous to the neuronal connection network in the human brain. One of a CNN's functions is to condense images into a format that is easier to interpret without losing details that are essential for accurate prediction. Planning an architecture that is not only effective at learning features but also scalable to large datasets requires consideration of this..The main CNN operations are convolution, pooling, batch normalization and dropout.

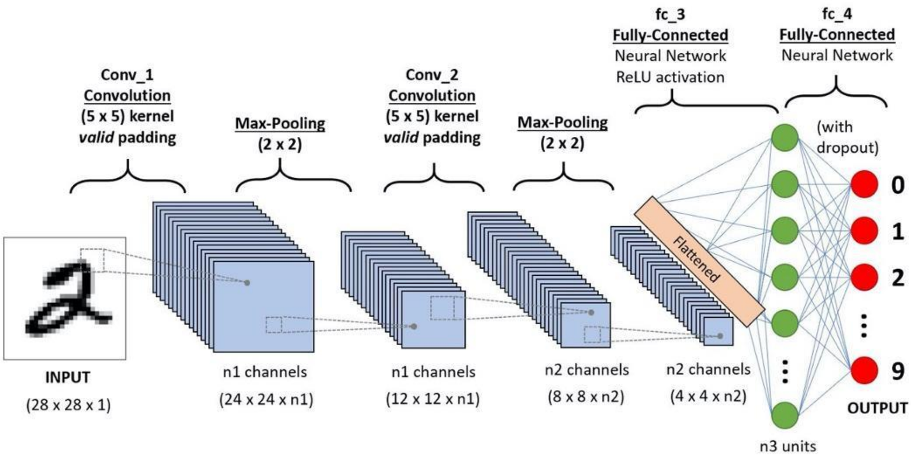


Fig 3.Conolution Neural Network

The layers of convolutional-neural-network would be divided into mainly 3 types: first one is convolutional, second one is pooling and third one is fully connected layers. Each and every layer will be playing a different role. The architecture of CNN is depicted below in Figure 3. The convolutional layer is said be primary as the elementary developing block for CNN. In CNN technology, it would be always critical to understand that layer parameters are being made up of a series of learnable filters or neurons. These filters have a small receptive field, but pass through the entire convolutional layer: the convolutional layer is known as the elementary building block of the input volume. In the pass-through process, each individual filter passes through the width and height of the CNN. In CNN technology, it is very crucial to being to understand that the layer parameters are composed of with set of input volumes, with the calculation of the dot product between the filter inputs and the input. Product of this of filters or learnable neurons. These filters would have a small receptive field, but they would reach all the way. The calculation would be a two-dimensional activation map of that filter. In this way the network learns the filters through the input volume. In the pass-through process, each individual filter loops through the created width when it detects a particular feature type at a spatial location within the feature map input and input volume height, calculating the dot product from the filter inputs and the input. .

The filter size represents the number of weights of the dimension [F × F × Cx]. Its size is determined by a chosen receptive field (F) and the feature map inputs that a neuron has connected to a region at the input. The convolutional layer has the advantage of depth (Cx) Pooling Layers: Pooling layers will be responsible for adjusting the width-by-height size by downsampling or subsampling because reducing the size results in a simultaneous loss of information, reducing the spatial size of the input volume for the subsequent convolutional layer without affecting the benefit of the network. The reduction becomes less computationally demanding as the information progresses toward the dimensional depth of the volume. The process performed by the pooling layer is also known as lower pooling layers and works against overfitting. The most common strategies used in sampling or subsampling because the decrease in size results in a simultaneous loss of information that pooling layer networks are maximum pooling and average pooling. Theoretically complete the advantages of the network. The reduction will become to lesser the computational as the data progresses towards maximum pooling analysis and average pooling is generated, while in it subsequent pooling levels are shown to be maximal and also work against overfitting. The most common strategies used in pooling could be resulting in faster convergence of information, and the network which chooses the high-rank characteristics of pooling layer networks: maximum pooling and average pooling. Theoretically comprehensive in the image, thus improving generalization. Furthermore, the pooling layer has other variations, such as max pooling and average pooling analysis, while stochastic max pooling, spatial pyramid pooling, and def pooling have been shown to serve distinct purposes.

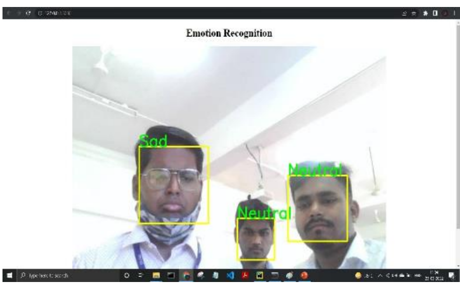
4. Results

As in the above image it can be can see the two human faces.

1) Depicting surprise emotion, sag, happy, neutral emtions

Our work system in this project is able to precisely identify the human emotions showed by the below faces

Accuracy of the proposed system: 70.71%



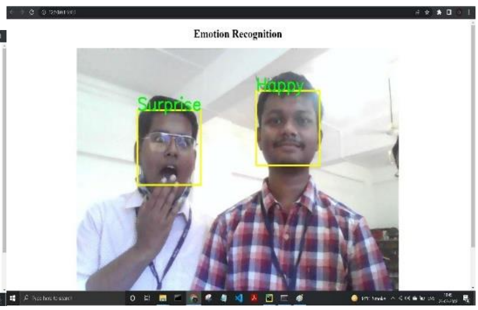
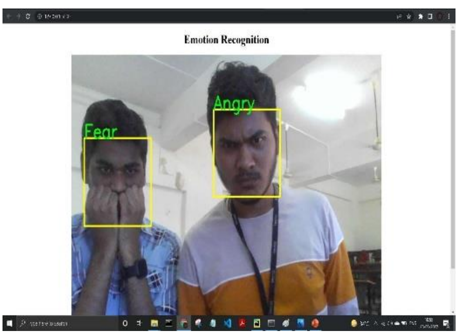


Fig 4. Differ types of emotions

Fig 5. Surprise and Happy emotions



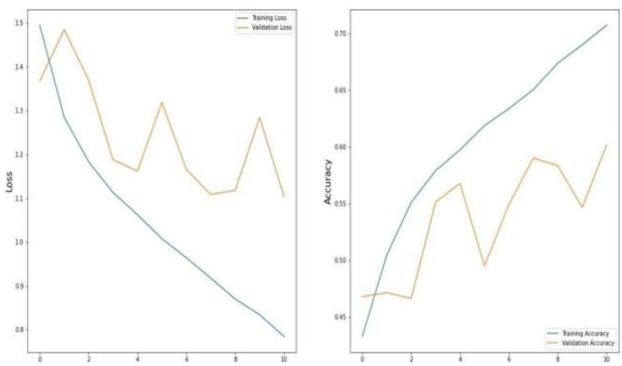


Fig 6. Fear and Angry faces

Fig 7 . graph of training loss and validation loss

As in the above image it can seem that the three human faces which are correctly recognized by the proposed system.

1) Depicting sad emotion.

2) Other two being neutral having normal facial expression / emotion.

Our proposed system is able to precisely identify the human emotions showed by the

above faces.The other two emotions i.e., fear and anger are depicted by the above faces which are also precisely identified by the proposed system.

5. Performance Analysis

The performance analysis would be accuracy, precision, recall, and f_measure. Precision was the fraction of predicted labels that were correct. Precision recall would be defined as

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP is the true positives rate, TN is the true negatives rate, FN is false negatives, and FP is false positives.

Recall is represented as the fact that the fraction is of true positive instances to the sum of true positives and false negatives. Recall would be defined as:

$$\text{Recall} = \frac{TP}{TP + FN}$$

Precisionn would be represented as fraction of true positive instances to all positive instances. Precision would be defined as:

$$\text{Precision} = \frac{TP}{TP + FP}$$

F_Measure represented the combination of precision and recall. F_Measure was defined as:

$$\text{F\_measure} = \frac{(\text{precision} * \text{recall})}{(\text{precision} + \text{recall})}$$

6. WEB UI

Fig:3 Landing Page

Fig: 4 About

5. Conclusion

As these computers and computer interfaces grow more embedded into our everyday lives, they emerge as a response to the need for computers to understand human speech as well as behavioral signs of emotions and mental states. Real-time face recognition allows for the identification of human faces, which may then be utilised for person identification and verification.

Further, the expressions in facial pictures are important. By combining all of the advantages, interpersonal contact between a human and a computer may be improved. By increasing the number of photos used during training, the accuracy of both face recognition and emotion detection may be enhanced. Our method is divided into three parts: feature extraction, face identification, and classification using machine learning techniques. The extraction of features was a critical component of the experiment. In this paper, we explore facial expressions for the detection of human emotions in live webcam video.

By implementing this recognition technique, the majority of real-time concerns can be improved. Emotion recognition systems can be useful in real-world applications such as humanoid robots, security, and gaming. We examine the present status of emotion recognition, emphasizing its major accomplishments, take-home lessons, problems, and potential future prospects.

References

1. Varghese, A. A., Cherian, J. P., & Kizhakkethottam, J. J. (2015, February). Overview on emotion recognition system. In 2015 International Conference on Soft- Computing and Networks Security (ICSNS) (pp. 1-5). IEEE.
2. Bota, P. J., Wang, C., Fred, A. L., & Da Silva, H. P. (2019). A review, current challenges, and future possibilities on emotion recognition using machine learning and physiological signals. IEEE Access, 7, 140990-141020.
3. Deshmukh, R. S., Jagtap, V., & Paygude, S. (2017, June). Facial emotion recognition system through machine learning approach. In 2017 international conference on intelligent computing and control systems (iciccs) (pp. 272-277). IEEE.
4. Hassouneh, A., Mutawa, A. M., & Murugappan, M. (2020). 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, 100372.
5. Rajesh, K. M., & Naveenkumar, M. (2016, December). A robust method for face recognition and face emotion detection system using support vector machines. In 2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT) (pp. 1-5). IEEE
6. Siam, A. I., Soliman, N. F., Algarni, A. D., El-Samie, A., Fathi, E., & Sedik, A. (2022). Deploying Machine Learning Techniques for Human Emotion Detection. Computational Inte
7. Kosti, Ronak, Jose M. Alvarez, Adria Recasens, and Agata Lapedriza. "Emotion recognition in context." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1667-1675. 2017.
8. Zhang, J., Yin, Z., Chen, P. and Nichele, S., 2020. Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review. Information Fusion, 59, pp.103-126.
9. Abdullah, S.M.S.A., Ameen, S.Y.A., Sadeeq, M.A. and Zeebaree, S., 2021. Multimodal emotion recognition using deep learning. Journal of Applied Science and Technology Trends, 2(02), pp.52-58.
10. Jain, N., Kumar, S., Kumar, A., Shamsolmoali, P. and Zareapoor, M., 2018. Hybrid deep neural networks for face emotion recognition. Pattern Recognition Letters, 115, pp.101-106.
11. Sati, V., Sánchez, S.M., Shoeibi, N., Arora, A. and Corchado, J.M., 2020, June. Face Detection and Recognition, Face Emotion Recognition Through NVIDIA Jetson Nano. In International Symposium on Ambient Intelligence (pp. 177-185). Springer, Cham.

12. Nasri, M.A., Hmani, M.A., Mtibaa, A., Petrovska-Delacretaz, D., Slima, M.B. and Hamida, A.B., 2020, September. Face emotion recognition from static image based on convolution neural networks. In 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP) (pp. 1-6). IEEE.
13. Lasri, I., Solh, A.R. and El Belkacemi, M., 2019, October. Facial emotion recognition of students using convolutional neural network. In 2019 third international conference on intelligent computing in data sciences (ICDS) (pp. 1-6). IEEE.
14. Saxena, S., Tripathi, S. and Sudarshan, T.S.B., 2019, November. Deep Dive into Faces: Pose & Illumination Invariant Multi-Face Emotion Recognition System. In 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 1088-1093). IEEE.
15. Zadeh, M.M.T., Imani, M. and Majidi, B., 2019, February. Fast facial emotion recognition using convolutional neural networks and Gabor filters. In 2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI) (pp. 577-581). IEEE.