APTIKOM Journal on Computer Science and Information Technologies
Vol. 1, No. 3, 2016, pp. 92~108
ISSN: 2528-2417
DOI: 10.11591/APTIKOM.J.CSIT.24

A Review on Deep Learning Algorithms for Speech and Facial Emotion Recognition

Charlyn Pushpa Latha, Mohana Priya
Faculty of Engineering, Karpagam University, Coimbatore, India
*Corresponding author, email: charlyn.latha@gmail.com, mohanapriyaasaithambi@gmail.com

Abstract

Deep Learning is the recent machine learning technique that tries to model high level abstractions in data by using multiple processing layers with complex structures. It is also known as deep structured learning, hierarchical learning or deep machine learning. The term "deep learning" indicates the method used in training multi-layered neural networks. Deep Learning technique has obtained remarkable success in the field of face recognition with 97.5% accuracy. Facial Electromyogram (FEMG) signals are used to detect the different emotions of humans. Some of the deep learning techniques discussed in this paper are Deep Boltzmann Machine (DBM), Deep Belief Networks (DBN), Convolutional Neural Networks (CNN) and Stacked Auto Encoders respectively. This paper focuses on the review of some of the deep learning techniques used by various researchers which paved the way to improve the classification accuracy of the FEMG signals as well as the speech signals.

Keywords*: Deep Learning; Facial Electromyography; Emotions; Deep Boltmann Machine; Deep Belief Networks; Convolutional Neural Networks; Stacked Auto Encoders*

1. Introduction

Emotions are experienced from an individual point of understanding. It is mostly related with mood, temperament, personality, and disposition. Emotions and their corresponding expressions are important elements in social interactions and they are used as mechanisms for signaling, directing, attention, motivating and controlling interactions, situation assessment etc. Emotional assessment has recently attracted the attention of many researchers from different fields. Research on emotion recognition consists of facial expressions, vocal, gesture and physiological signal recognition and so on. Of all these emotion recognition factors, facial emotion reaction is the recent and blooming technique chosen by the researchers. Emotion recognition in humans is an important research area. As per Robert Plutchik's theory, the basic emotions are classified into 8 different types. The emotions listed are fear, anger, sadness, joy, disgust, trust, anticipation and surprise. The Book Two of Aristotle's Rhetoric suggests that there are totally 9 emotions. The emotions according to his listing are anger, friendship, fear, shame, kindness, pity, indignation, envy and love. Lojban classified emotions into three categories namely simple emotions, complex emotions and pure emotions. Simple emotions include anger, envy, wonder, happiness, amusement, courage and pity. Complex emotions include pride, closeness, pain, caution, patience, relaxation and envy. Pure emotions include fear, togetherness, respect, appreciation, love, hope and confusion. Facial Emotion Recognition using deep learning technique is the recent research area.

Research using deep learning techniques could make better representations and create innovative models to learn these representations from large-scale unlabeled data. Some of the deep learning techniques like Convolutional Deep Neural Networks, Deep Boltzmann Machine, Deep Belief Networks, recurrent neural networks, deep neural networks and stacked auto encoders are applied to practical applications like pattern analysis, audio recognition, computer vision, natural language processing, automatic speech recognition, bioinformatics, vehicle, pedestrian and landmark identification for driver assistance, image recognition, customer relationship management, speech recognition and translation and life sciences where they produce challenging results on various tasks. Scientists use deep learning techniques to solve highly practical problems in all traits of business such as:

a. Payment systems providers use deep learning to identify doubtful transactions in real time.

Received Aug 18, 2016; Revised Oct 20, 2016; Accepted Oct 25, 2016

b. Organizations with large data centers and computer networks use deep learning to mine log files and identify threats.
c. Vehicle manufacturers and fleet operators use deep learning to mine sensor data to foretell part and vehicle failure.
d. Deep learning helps companies with large and complex supply chains forecast delays and holdups in production.

Some of the advantages of deep learning technique are its ability to detect complex interactions among features, capability to learn low-level features from minimally processed raw data, easy to work with high-cardinality class memberships and its competence to work with unlabeled data. This paper gives a brief introduction on the deep learning techniques used. It also summarizes the different deep learning approaches adopted by scientists in the previous years with their advantages, drawbacks and future works.

2. Deep Learning Techniques Used

Deep learning is a fast growing area which models high-level patterns in data as complex multilayered networks. It is mainly used in machine learning and artificial intelligence. Leading companies like Microsoft and Google use deep learning techniques to solve challenging problems in crucial areas like speech recognition, image recognition, 3-D object recognition, and natural language processing. Some of the deep learning algorithms discussed in this section are Deep Boltzmann Machine (DBM), Deep Belief Networks (DBN), Convolutional Neural Networks (CNN) and Stacked Auto Encoders.

2.1. Deep Boltzmann Machine (DBM)

Deep Boltzmann Machine is a type of binary pairwise Markov Random field with multiple layers of hidden random variables with a network of symmetrically coupled stochastic binary units. It also comprises a set of visible units $v_i \in \{0,1\}^c$ and a series of hidden units $h_i^{(1)} \in \{0,1\}^{c_1}, h_i^{(2)} \in \{0,1\}^{c_2}, \ldots . h_i^{(L)} \in \{0,1\}^{c_1}$. No connection exists between the units of the same layer as in the case of Restricted Boltzmann Machine. The probability assigned to vector *v*, is $p(v_i) = \frac{1}{Z} \sum_{h_i} e^{\sum abWab^{(1)} v_i hb^{(1)} + \sum bcWbc^{(2)} h_b^{(1)} h_i^{(2)} + \Sigma_{cd} W_{cd}^{(3)} h_c^{(2)} h_d^{(3)}}$ where $h = \{h_i^{(1)}, h_i^{(2)}, h_i^{(3)}\}$ the set of hidden units are and $\theta = \{W_i^{(1)}, W_i^{(2)}, W_i^{(3)}\}$ are the corresponding model parameters representing visible–hidden and hidden–hidden interactions. If $W_i(2)=W_i(3)=0$, then the network is known as the restricted Boltzmann Machine. Figure 1 represents the graphical architecture of a Boltmann Machine.

From Figure 1, it is evident that each undirected edge represents dependency. Here there are three hidden units and four visible units. This is not a restricted Boltzmann machine. Figure 2 represents the pictorial representation of a Restricted Boltzmann Machine. From Figure 2 it is inferred that the four blue units represent hidden units, and three red units represent visible units. This proves that the restricted Boltzmann machine has connections or dependencies only between the hidden units and the visible units, and there exists no connection between the hidden-hidden units or the visible-visible units.

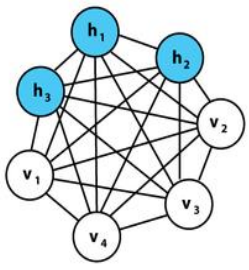

Figure 1. Graphical Representation of a Boltzmann Machine

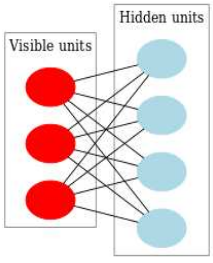

Figure 2. Graphical Representation of a Restricted Boltzmann Machine

DBM learns complex and abstract internal representations of the input in various models like object recognition or speech recognition, using sufficient labeled data to fine-tune the representations

built using a large supply of unlabeled sensory input data. DBMs also adopt the inference and training procedure in both directions namely bottom-up and top-down pass, which allows the DBMs to better disclose the representations of the ambiguous and complex input structures. The speed of DBMs limits their performance and functionality. The advantages of the Deep Boltzmann Machine are their capability to learn efficient representations of complex data, [1] with efficient pre – training technique layer by layer. The most benefit of DBMs is that it could be trained even with unlabeled data and fine-tuned with the possible limit data for a specific application. DBMs could also predict the uncertainty of the ambiguous input by the way of analyzing the approximate inference procedure found in DBMs. By applying the approximate gradient procedure to all the layers, the parameters in it could be optimized which in turn facilitates for the learning of better generation of models. The disadvantages of the Deep Boltzmann Machine are works well for theoretical purpose rather than a general computational medium and it stop learning correctly when the machine is scaled up to anything larger than a minor machine. The approximate inference procedure followed in DBMs is nearly 50 times slower than which is followed in DBNs. Hence DBMs is not suitable for larger databases.

2.2. Deep Belief Networks (DBN)

Deep belief networks [2] are highly complex directed acyclic graph, which are formed by a sequence of restricted Boltzmann Machine (RBM) architectures. DBN could be trained by training RBMs layer by layer from bottom to top. Since RBM could be trained rapidly through layered contrast divergence algorithm, the training avoids a high degree of complexity of training DBNs which inturn simplifies the process to train each RBM. Studies on DBN illustrated that it can solve low convergence speed and local optimum problems in traditional back propagation algorithm in training multilayer neural network. Figure 3 represents the architecture of the Deep Belief Network in which the RBMs are trained layer by layer from bottom to top.

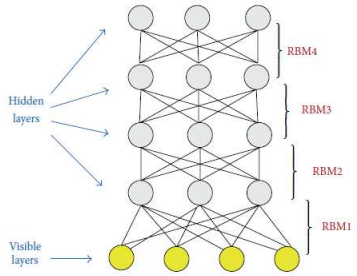

Figure 3. Graphical Representation of a Deep Belief Network

The advantages of the Deep Belief Network model include the ability to learn an optimum set of parameters quickly even for the models which contain many large number of parameters and the layers with nonlinearity by way of the greedy layer–by layer algorithm. DBNs use an unsupervised pre training method even for very large unlabeled databases. DBNs could also compute the output values of the variables in the lowest layer using approximate inference procedure. The disadvantages of DBNs include the limitation of the approximate inference procedure to a single bottom-up pass. The greedy procedure learns only the features of one layer at a time and it never readjusts with the other layers or parameters of the network. The wake-sleep algorithm proposed by Hinton for DBNs is very slow and inefficient though it fine tunes globally.

2.3. Convolutional Neural Networks (CNN)

A Convolutional Neural Network is a type of Feed forward neural network architecture in machine learning. CNNs have a collection of small neurons in multiple layers that process the input image in portions called as the receptive fields. The output of these collections are lined in such a way that there is an overlapping of the input regions that gives a clear representation of the original input image. The process is repeated for all the layers. CNNs are mostly used in video and image recognition, natural language processing and recommender systems. Figure 4 shows the architectural diagram of the CNN in which the subsampling is done for each and every layer and the connections are lined up to form the fully connected representation of the image.

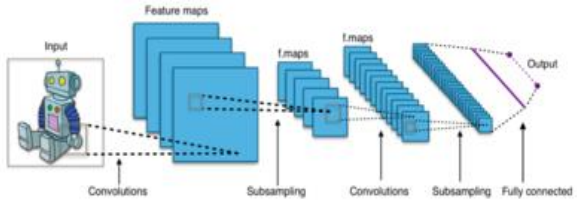

Figure 4. Graphical Representation of a Convolutional Neural Network

The advantages of the convolutional neural network are as follows. First, the usage of shared weights in convolutional layers paves the way to use the same filter for each pixel in the layer. Next, CNNs use relatively little pre-processing which means that the CNN network is responsible for learning the filters where the traditional algorithms are hand- engineered. Thirdly, CNNs are easy to train and are less dependent on the human understanding and effort and also on the previous knowledge in designing the features for the model. CNNS can also design 2D structure of the input image by using local connections and weights followed by pooling technique which in turn results in translation invariant features. The main advantage of CNNs is that it has only fewer parameters when compared to the fully connected networks with the same number of hidden units. The most distinguishing feature of the CNN is that it has 3D volume of neurons in which the neurons are arranged in 3 dimensions namely weight, height and depth. The disadvantage of CNNs is that CNNs require a huge amount of memory requirement to hold all the intermediate results of the convolutional layer for giving as the input to the backpropogation layer.

2.4. Stacked Auto Encoders (SAE)

The idea behind the auto encoder is based on the concept of a well-built representation of any model. An encoder is a mapping f_θ that transforms an input vector i into a hidden layer representation *h,* where $\theta = \{W_t, c\}, W_t$, is the matrix weight and c is the bias/the offset vector. The decoder maps the hidden representation h to the reconstructed input d through e_0. The auto encoder process is to compare the reconstructed input with the original by minimizing the error so as to make the reconstructed value as close as possible to the original value. In this technique, the output which is partially corrupted is cleaned. After the encoding function $f_\theta$of the first denoising auto encoder is trained, it is used to uncorrupt the corrupted input where the second level can be trained. After all the layers in the stacked auto encoder is trained, the output can be used as an input to any supervised learning algorithms like Support Vector Machine, Multiclass logistic regression etc. Figure 5 represents the diagram of a stacked auto encoder.

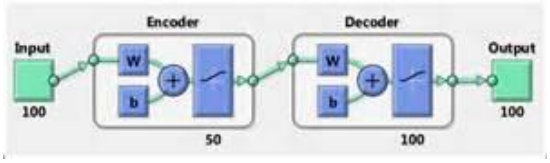

Figure 5. Graphical Representation of a Stacked Auto Encoder

The advantages of stacked encoders are as follows: First, it is a layer wise training method. SAE works very much compatible with Artificial Neural Networks. SAE considers all the real number inputs, binary inputs, probabilistic distribution only by simply changing the loss function and the activation function. SAE can also fine tune by itself using the back propagation algorithm which reduces the total reconstruction loss.

3. Literature Review on the Techniques for Deep Learning

Many algorithms are available for the understanding of deep learning technique to generate classification models. This section typically gives an overview of the techniques used by different researchers over the years.

3.1. Survey on Deep Boltzmann Machine (DBM)

DBM is a type of network model comprising of stochastically binary units. Table 1 reviews the deep Boltzmann technique used by the researchers in the previous years.

Table 1. State of art on Deep Learning using DBM technique

S.No	Year	Authors	Title	Deep Learning Technique Used	Contributions and Classification Accuracy Obtained	Drawbacks
1		Ankit Awasthi	Facial Emotion Recognition using Deep Learning	Restricted Boltzmann Machine Deep Belief Networks	i) Model develops important insights for the features learnt by varying the number of hidden layers and units. ii) Develops quantitative ways for evaluation of the cognitive importance of features through which the DBNs could be used as the good model of the visual system.	
2	2015 Jan	Labmaster	Automatic Emotion Recognition system using Deep Belief Network	Deep Learning Network (DBN) Restricted Boltzmann Machine		
3	2014	E. M. Albornoz et al	Spoken emotion recognition using deep learning	Restricted Boltzmann machines and deep belief networks	Method achieves an improvement of 8.67% over the baseline in a speaker independent scheme	The deep classifier should be tested with noisy signals.
4	2015	Zheng-wei HUANG et al	Speech emotion recognition with unsupervised feature learning	the sparse auto-encoder, and sparse restricted Boltzmann machines	Evaluated the effects of the important parameters namely number of hidden nodes and the size of the content window on the eNTERFACE database which proved that larger window size and more number of hidden nodes can increase the performance rate.	Searching or assembling large emotional data is mandatory to prepare sufficient amount of data for this deep network model.
5		Xiaowei Jia et al	A Novel Semi-supervised Deep Learning Framework for Affective State Recognition on EEG Signals	Generative Restricted Boltzmann Machine (RBM) model	Results reveal that this model exceeds extensive baseleines in classification and it also outperforms the training process by a decent rate.	
6		Kihyuk Sohn et al	Improved Multimodal Deep Learning with Variation of Information	restricted Boltzmann machines	The method could be extended to deep networks with repeated coding structure to fine tune the entire network.	
7		John Goddard	Spoken Emotion Recognition using Restricted Boltzmann Machines and Deep Learning	restricted boltzmann machines (rbm) and deep belief networks (dbn)		
8	March 2013	Branislav Popović et al	Deep Architectures for Automatic Emotion Recognition Based on Lip Shape	Boltzmann Machines Deep Belief Network	Comparing the results obtained by human classification, as well as the results achieved using the previously developed algorithms, this method achieved more encouraging results.	Over fitting limitation could be minimized.

Ankit Awasthi [3] used the combined algorithms namely Restricted Boltzmann Machine and the Deep Belief Networks and developed a model that attracted the important features learnt by the model by changing the number of hidden layers and units. He also proposed some quantitative ways inorder to evaluate certain cognitive features of the model through which the DBN could be proposed. In 2015,

Labmaster [4] proposed an automatic emotion recognition model using the combined algorithms namely DBN and RBM. In 2014, E.M. Albornoz et al [5], developed a model for spoken emotion recognition. They used the two techniques namely RBM and DBN and achieved an improvement of 8.67% over the existing baseline speaker independent scheme. This model could be further enhanced by testing the classifier with noisy signals.

In 2015, Zheng–Wei Huang et al [6], developed a model for Speech emotion recognition using unsupervised feature learning in an eNTERFACE database and proved that the performance of the system could be improved by larger window size and more number of hidden nodes. They used the combined models namely the SAE and the sparse restricted DBM They also insisted that the speech data should be fine tuned using labeled data inorder to improve the performance of the network model.Also large number of emotional data should be collected inorder to prepare enough data for the network model. Since the deep learning method is sensitive to small changes in the input features, the features affecting the speaker variations or environmental distortions could be analyzed and removed.

Xiaowei Jia et al [7] developed a Novel Semi-supervised Deep Learning framework using EEG signals for affective state recognition .They used the model generative restricted RBM and the results revealed that the model outperformed the baselines in classification and it improved in training using a decent margin. Kihyuk Sohen et al [8] developed an improved multimodel deep learning with variation of information. They used the RBM technique and found that this method could be extended to deep networks with recurrent coding structure to fine tune the whole network. John Goddard [9] developed a model for spoken emotion recognition. He used the models namely RBM and DBN. In March 2013, Brainslav Popovix et al [10] developed a model for Automatic Emotion Recognition based on lip shape. They used both the models namely Boltzmann Machine and DBN. Results indicated that the model achieved encouraging classification when compared with the results achieved using human comparison even with small amount of training data. This model could be made efficient by minimizing the limitation of overfitting. On the whole, survey on DBMs indicated that in most of the studies DBMs are combined with either SAE or DBN.

3.2. Survey on Deep Belief Networks (DBN)

From Table 2, it is analyzed that DBN technique could be used separately for classification purpose unlike DBM technique. In March 2016, Hiranmayi Ranganathan et al [11] proposed a multimodal emotion recognition system using the 2 deep learning techniques namely DBN and the convolutional DBN (CDBN) models. In 2016, Xuanyang Xia et al [12], discussed a biologically inspired model using the Cascaded CNN model, CDBN model and the hierarchical max pooling (HMAX) models. Using this model, they were able to estimate the center locations of the facial components, locations of the most discriminative feature learning and selection and also to utilize a new memory formation by integrating the preliminary facts, expression modulation and final decision process. Yelin Kim et al [13] proposed an audio–visual emotion recognition system using DBN models which could generate audio-visual features for emotion classification even in an unsupervised manner. Yan Zhang et al [14] developed a model for speech recognition using the DNN and DBN models. In 2014, Kevin Terusaki et al[15] developed an emotion detection model using DBN which learns human- interpretable features even when it is pretrained with unlabeled data. In 2006, Geoffrey E.Hinton et al [16] developed a model using DBN which can learn the networks one layer at a time. In 2014, Bu Chen et al [17], developed a model for chinese speech emotion recognition in which the speech emotion recognition rate of the system reached 86.5% than the SVM method discussed in the earlier study. Wei- Long Zheng et al [18] proposed an EEG based emotion classification using DBN and obtained the highest average accuracy of 86.91% with the DBN- HMM model. Ping Liu et al [19] developed a Facial Expression Recognition model using Boosted DBN which performed the three training stages namely feature learning, selection and classifier construction. Chenchen Huang et al in August 2014 developed a speech emotion recognition model through which the emotional characteristic parameter from the speech signals could be extracted accurately and automatically thereby improving the recognition rate. They used SVM and DBN models. Erik M.Schimdt et al [20] proposed modeling and predicting emotion in music by using the deep learning techniques namely regression based DBNs. Table 2 summarizes the state of art for DBNs on facial emotion recognition

Table 2. State of art on Deep Learning using DBN technique

S.No	Year	Authors	Title	Deep Learning Technique Used	Contributions and Classification Accuracy Obtained	Drawbacks
1	2016	Xuanyang Xia et al	A biologically inspired model mimicking the memory and two distinct pathways of face perception	Cascaded-CNN (convolutional neural networks convolutional deep belief networks (CDBN) hierarchical max-pooling (HMAX)	i) Identification of Center locations of facial components. ii)Easy use of locating the places containing the most discriminative information and parallel scheme to complete the feature learning and selection iii)Encoding important characteristics of the facial components by combining the integration modulation and final decision process.	
2		Yelin Kim et al	Deep learning for robust feature generation In audiovisual emotion recognition	Deep Belief Network models	DBNs can be used to generate audio-visual features for emotion classification, even in an unsupervised context.	i) Investigation on the comparative advantage of deep learning techniques with additional emotion corpora and the investigation of deep modeling could be done in the context of dynamic feature generation. ii) Visualization of complex dependencies between either features or weights between hidden nodes of the DBNs will open a new gateway for the interpretation of audio-visual emotion data.
3		Yan Zhang et al	Speech Recognitio n Using Deep Learning Algorithms	deep neural networks (DNN) and deep belief networks (DBN)	Applying typical deep learning algorithms, including deep neural networks (DNN) and deep belief networks (DBN), for automatic continuous speech recognition.	DBN-based speech recognition system beats other two speech recognition systems.
4	2014	Kevin Terusaki and Vince Stigliani	Emotion Detection using Deep Belief Networks	Deep Belief Network	DBN architectures autonomously learns representationally useful and human-interpretable features of the input, and this method is still effective when pretrained with unlabeled data.	i) Difficult to visualize the pre-training filters beyond the first layer and hence further research can be devoted to finding a good method to visualize the contribution of deeper hidden units. ii) Additional time should be devoted to a real-time emotion detection system that would use the trained DBN to recognize an emotion through a webcam.
5	2006	Geoffrey E. Hinton et al	A Fast Learning Algorithm for Deep Belief Nets	Deep Belief Network	Proposed algorithm learns deep, directed belief networks one layer at a time, provided the top two layers form an undirected associative memory	i) Use of top –down feedback during perception is limited to the associative memory in the top two layers. ii) No systematic way of dealing with perceptual invariances. iii) Does not learn to sequentially attend to the objects when discrimination is difficult.
6	2014	Bu Chen et al	A Study of deep belief network	deep belief network (DBN)	The speech emotion recognition rate of the system reached 86.5%,	By training the data set, the study of speech emotion recognition based on DBNs

S.No	Year	Authors	Title	Deep Learning Technique Used	Contributions and Classification Accuracy Obtained	Drawbacks
			based Chinese speech emotion recognition		which was 7% higher than the original SVM method.	is done and the recognition rate is improved further.
7		Wei-Long Zheng et al	EEG-BASED EMOTION CLASSIFICATION USING DEEP BELIEF NETWORKS	deep belief network (DBN)	i) Average accuracies of DBN-HMM, DBN, GELM, SVM, and KNN in the experiments are 87.62%, 86.91%, 85.67%, 84.08%, and 69.66%, respectively. ii)The reliability of classifications achieved suggests that such neural signatures associated with positive or negative emotions do exist.	
8		Ping Liu et al	Facial Expression Recognition via a Boosted Deep Belief Network	Boosted Deep Belief Network (BDBN)	A novel Boosted Deep Belief Network (BDBN) for performing the three training stages namely feature learning, feature selection, and classifier construction are performed iteratively in a unified loopy framework.	
9	August 2014	Research Article A Research of Speech Emotion Recognition Based on Deep Belief Network and SVM Chenchen Huan Chenchen Huang	A Research of Speech Emotion Recognition based on Deep Belief networks and SVM	Deep Belief Network	Extraction of the emotional characteristic parameter from emotional speech signals is done automatically and accurately, improving the recognition rate of emotional speech recognition obviously.	The time cost for training DBNs feature extraction model is higher compared to other neural networks.
10		Erik M. Schmidt et al	Modeling and Predicting Emotion in Music	Regression based deep belief networks (DBNs)	i) Models relationships using a conditional random field (CRF), a powerful graphical model that is trained to predict the conditional probability for a sequence of labels. ii) The application of regression based (DBNs) to learn features directly from magnitude spectra are analyzed.	Models should be found that could incorporate multiple spectral time-scales to derive musical emotion

3.3. Survey on Convolutional Neural Networks (CNN)

Table 3 summarizes the literature review on the convolutional neural network technique. In March 2016, Shashank Jaiswal et al [21], proposed a dynamic appearance for the shape of facial action unit using the convolutional and bi-directional long short-term memory neural networks (CNN- BLSTM) method which learned the dynamic appearance and shape of the facial regions. In 2016, Quanzeng You et al [22] developed a model for Joint Visual Textual Sentiment analysis of social multimedia using the cross-modality consistent regression technique which fine tunes the CNN model. In May 2016, Quanzeng You et al [23] analyzed a method to build large scale dataset for image emotion recognition using the Deep CNN model.

In May 2016, Steven K.Esser et al [24] developed a model for efficient Neuromorphic computing using the Deep CNN technique. In Feb 2016, Carlos Argueta [25] developed a facial emotion recognition using the techniques namely simple softmax regression deeper code (MNIST) and a 2D CNN. Hong–Wei Ng et al [26] developed a model for emotion recognition on small datasets using Deep CNN technique. The results achieved showed that the cascading fine tune approach achieved better results than a single stage tuning. Victor Emil Neagoe et al [27] developed a model for Subject independent emotion recognition from facial expressions using combined CNN and DBN and achieved higher accuracy.

George Trigeorgis et al [28] proposed an end-to-end speech emotion recognition using combined CNN and Long Short Term memory (LSTM) networks. In 2015, Samira Ebrahimi Kahou et al [29] proposed a Recurrent Neural Network platform for Emotion recognition in video using the CNN hybrid RNN (CNN-RNN) model. In 2015, Gil Levi et al [30] proposed emotion recognition in the wild through CNN and Mapped Binary patterns. Amogh Gudi et al [31] used Deep CNN model to promote deep learning based FACS action unit occurrence and Intensity estimation.

In 2014, Amogh Gudi [32] also developed a method to recognize semantic features in faces using the CNN and the maximum pooling local contrast normalization technique. In December 2015, Pablo Barros et al [33] developed a multimodal emotional state recognition using CNN model and achieved higher accuracy of 91.3%. Yilin Yan et al [34] used CNN with bootstrapping technique for imbalanced multimedia data classification. Li Wang et al [35] used CNN for a creative application to video analytics for a smart city. They created an advanced hardware, which reduced the training time for deep neural networks. In 2011, Moez Baccouche et al [36] formulated a sequential deep learning for human action recognition using the 3D CNN and improved the recognition accuracy from 92.17% to 94.39%. Claudia Aracena [37] developed a repository for EEG- based Emotion classification using the CNN model which could efficiently learn to represent the data.

Table 3. State of art on Deep Learning using CNN technique

S.No	Year	Authors	Title	Deep Learning Technique Used	Contributions and Classification Accuracy Obtained	Drawbacks
1	2016 March	Shashank Jaiswal Michel Valstar	Deep Learning the Dynamic Appearance and Shape of Facial Action Units	Convolutional and Bi-directional Long Short-Term Memory Neural Networks (CNN-BLSTM)	i) Learns the dynamic appearance and shape of facial regions for Action Unit detection. ii) Each component of the system contributes towards an improvement in the model performance.	To increase the performance with more challenging results
2	2016	Quanzeng You Jiebo Lou Hailin Jin Jianchao Yang	Cross-modality Consistent Regression for Joint Visual-Textual Sentiment Analysis of Social Multimedia	cross-modality consistent regression (CCR) model	The developed (CCR) model utilizes both the state-of-the-art visual and textual sentiment analysis techniques.	Sentiment analysis results encourage further research on online user generated multimedia content.
3	2016 May	Quanzeng You Jiebo Lou Hailin Jin Jianchao Yang	Building a Large Scale Dataset for Image Emotion Recognition: The Fine Print and The Benchmark	Deep Convolutional Neural Network	i)Development of a new data set which started from 3+ million weakly labeled images of different emotions and ended up 30 times as large as the current largest publicly available visual emotion data set. ii) Evaluation of the deep visual features extracted from differently trained neural network models is exhibited.	i) Visual emotion analysis results can encourage further research on online user generated multimedia content in the wild. ii)Better understanding of the relationship between emotion arousals and visual stimuli could be developed and further extension of understanding the valence are the primary future directions for visual

S.No	Year	Authors	Title	Deep Learning Technique Used	Contributions and Classification Accuracy Obtained	Drawbacks
						emotion analysis
4	2016 May	Steven K. Esser et al	Convolutional Networks for Fast, Energy-Efficient Neuromorphic Computing	deep convolution networks	The algorithmic power of deep learning can be merged with the efficiency of neuromorphic processors, bringing the promise of embedded, intelligent, brain-inspired computing one	Co design between algorithms and future neuromorphic architectures could be implemented which promises better accuracy and efficiency.
5	2016 Feb	Carlos Argueta	Facial Emotion Recognition: Single-Rule 1–0 DeepLearning	simplest code (simple MNIST), a modest softmax regression deeper code (deep MNIST), a two-layers Convolutional Network		
6		Hong-Wei Ng et al	Deep Learning for Emotion Recognition on Small Datasets Using Transfer Learning	deep Convolutional Neural Network (CNN)	Experimental results show that cascading fine-tuning approach achieves better results, compared to a single stage fine tuning with the combined datasets with an overall accuracy of 48.5% obtained in the validation set and 55.6% in the test set, which compares favorably to the respective 35.96% and 39.13% of the challenge baseline	The difficulty in assigning correct labels to faces depicting some of the more nuanced emotions, and how that can affect the performance of our models could be solved.
7		VICTOR-EMIL NEAGOE et al	A Deep Learning Approach for Subject Independent Emotion Recognition from Facial Expressions	CNN and DBN	Two deep learning models namely DBN and CNN models were proposed and the efficiency obtained were 65.22% and 95.71% respectively for both person independent and person dependent approaches.	
8		George Trigeorgis et al	ADIEU FEATURES? END-TO-END SPEECH EMOTION RECOGNITION USING A DEEP CONVOLUTIONA L RECURRENT NETWORK	Convolutional Neural Networks (CNNs) with LSTM networks	A solution is proposed to the problem of 'context-aware' emotional relevant feature extraction, by combining Convolutional Neural Networks (CNNs) with LSTM networks, in order to automatically learn the best representation of the speech signal directly from the raw time representation	
9	2015	Samira Ebrahimi Kahou et al	Recurrent Neural Networks for Emotion Recognition in Video	convolutional neural networks (CNNs) hybrid CNN-RNN architecture	A complete system for the 2015 Emotion Recognition in the Wild (EmotiW) Challenge is proposed	

S.No	Year	Authors	Title	Deep Learning Technique Used	Contributions and Classification Accuracy Obtained	Drawbacks
10	2015	Gil Levi et al	Emotion Recognition in the Wild via Convolutional Neural Networks and Mapped Binary Patterns	Convolutional Neural Networks (CNN)	Model is designed with the goal of simplifying the problem domain by removing confounding factors from the input images, with an emphasis on image illumination variations which in an effort reduces the amount of data required to effectively train deep CNN models.	
11		Amogh Gudi et al	Deep Learning based FACS Action Unit Occurrence and Intensity Estimation	a deep convolutional neural network (CNN)	A novel application of a deep (CNN) to recognize AUs.	
12	2014	Amogh Gudi	Recognizing Semantic Features in Faces using Deep Learning	Convolution Neural Netwoks Maximum Pooling Local Contrast Normalization	Explores the effectiveness of the system to recognize the various semantic features like emotions, age, gender, ethnicity etc present in faces.	i) A limiting factor in the conducted experiments is the available computational resources. ii) Another limiting factor is the size of the dataset and the quality of annotations
13	Decemb er 2015	Pablo Barros et al	Multimodal emotional state recognition using sequence-dependent deep hierarchical features	Convolutional Neural Networks	Experiments show that a significant improvement of recognition accuracy is achieved when hierarchical features and multimodal information is used in which the model improves the accuracy of state-of-the-art approaches from 82.5% reported in the literature to 91.3% for a benchmark dataset on spontaneous emotion expressions.	i) Further analysis of the features learned by the architecture in each experiment would help to visualize the complex features extracted by the model. Ii) Implementation of the model in a real-world scenario will be explored, to extend the model to real-time continuous recognition.
14		Yilin Yan et al	Deep Learning for Imbalanced Multimedia Data Classification	Convolutional Neural Networks (CNNs) with Bootstrapping Technique	CNNs are computationally expensive in which the extended CNN can work effectively on low-level features, which greatly reduces the required training time in deep learning.	
15		Li Wang et al	Deep Learning Algorithms with Applications to Video Analytics for A Smart City: A Survey	Convolutional Neural Networks	New advanced hardware (e.g. GPU) has largely reduced the training time for deep networks.	
16		Moez Baccouche et al	Sequential Deep Learning for Human Action Recognition	3D Convolutional Neural Networks	The neural-based deep model is developed to classify sequences of human actions, without a priori modeling, but relying only on automatic learning from training examples with accuracies of 94.39% and 92.17%.	Main limitation is the adaptation of the training algorithm, especially when calculating the retro-propagated error

S.No	Year	Authors	Title	Deep Learning Technique Used	Contributions and Classification Accuracy Obtained	Drawbacks
17		Claudia Aracena	Towards An Unified Replication Repository for EEG-based Emotion Classification	Convolutional Neural Network (CNN)	Model automatically learns to represent the data that can be competitive in comparison with manually crafted features.	Study on how the hyperparameter setup affects the performance of the classification.

3.4. Survey on Stacked Auto-Encoders (SAE)

In Feb 2016, Wei Liu et al [38] used the deep learning techniques namely Deep Auto Encoder for unimodal task. Bimodal deep encoder for multimodal task and developed a multimodal emotion recognition system. The authors achieved the best recognition accuracies of 82.11% for unimodal tasks and 91.01% and 83.25% for bimodal tasks on SEED and DEAP datasets. For cross modal learning task, the accuracy of 66.34% is obtained. In March 2016, Otkrist et al [39], used autoencoder and predictor architectures and designed a deep video gesture recognition using illumination variants. The authors concluded that this model could outperform all other models with a reduced size labeled dataset.

In 2014, Suwicha Jirayucharoesak et al [40], used SAE approach using hierarchical feature learning approach and designed an EEG based Emotion Recognition which overcomed the fitting problem using Principal Component Analysis. Classification accuracies obtained are 49.52% and 46.03% respectively. In 2012, Salah Rifai et al [41], used the multiscale contractive convolutional network (CCNET) Contactive Discriminant Analysis and Contractive Autoencoder obtained the accuracy of 85%. In 2015, Zheng–Wei Huang et al [42] used the combined techniques of the SAE and sparse RBM and developed a model for speech emotion recognition. The authors evaluated the effects of the two parameters namely number of hidden nodes and content window size and concluded the fact that larger window and more hidden nodes increases the performance of the model.

Jun Deng et al [43] used SAE method and designed a SAE based Feature Transfer Learning for Speech emotion recognition. Hector.P et al [44] used the combined deep learning techniques namely auto encoders and CNN, researched on Learning Deep Physiological Models of Affect.
Table 4 summarizes the state of art for SAE on facial emotion recognition

Table 4. State of art on Deep Learning using SAE technique

S.No	Year	Authors	Title	Deep Learning Technique Used	Contributions and Classification Accuracy Obtained	Drawbacks
1	2016 Feb	Wei Liu, Wei-Long Zheng, Bao-Liang Lu	Multimodal Emotion Recognition Using Multimodal Deep Learning	Deep AutoEncoder (DAE) for unimodal task Bimodal Deep AutoEncoder (BDAE) for multimodal task	i)Unimodal enhancement task- 82.11% ii)Multimodal facilitation tasks- 91.01% and 83.25% on SEED and DEAP datasets, respectively. iii)Cross-modal learning task- 66.34%.	i) Clear explanation of confusing matrices is found. ii) Performance of the DAE and BDAE networks should be further identified when the parameters change. iii) More experiments should be conducted in order to study the stability of the DAE and BDAE networks.
2	2016 Marc h	Otkrist Gupta Dan Raviv Massachu setts Ramesh Raskar	Deep video gesture recognition using illumination invariants	autoencoder and predictor architectures using an adaptive weighting scheme	Combines autoencoder and predictor architectures using an adaptive weighting scheme coping with a reduced size labeled dataset, while enriching our models from enormous unlabeled sets.	i) Other modalities such as sound could be taken into account. ii) System was built and tested using only frontal perspectives thereby imposing a constraint on the input facial orientations. iii) Though 7 emotions are tested, system could not be able to recognize neutral emotion. iv)Training data samples require more time.

3	2014	Suwicha Jirayucharoensak et al	EEG-Based Emotion Recognition Using Deep Learning Network with Principal Component Based Covariate Shift Adaptation	stacked auto encoder (SAE) using hierarchical feature learning approach	i) Overfitting problem is alleviated using the principal component analysis which extracts the most important components of initial input features. ii)Using covariate shift adaptation of the principal components, the nonstationary effect of EEG signals is minimized. iii) Results indicate that the DLN is capable of classifying three different levels of valence and arousal with an accuracy of 49.52% and 46.03%, respectively. iv) Principal component based covariate shift adaptation enhances the respective classification accuracy by 5.55% and 6.53% which provided better performance compared to SVM and naive Bayes classifiers.	The major limitations for performing EEG-based emotion recognition algorithm is dealing with the problem of intersubject variations in their EEG signals. The common features of transferable nonstationary information can be investigated to alleviate the intersubject variation problems.
4	2012	Salah Rifai et al	Disentanglin g Factors of Variation for facial expression recognition	Multiscale contractive convolutional network (CCNET) Contactive Discriminative Analysis Contractive Auto - Encoder	The system shifted the accuracy from 82.4% to 85.0% improving the accuracy of a standard CAE by 8%.	
5	2015	Zheng-wei HUANG et al	Speech emotion recognition with unsupervise d feature learning	the sparse auto-encoder, and sparse restricted Boltzmann machines	Evaluation of the effects of the two important parameters, number of hidden nodes and content window size, on the eNTERFACE database which proved the fact that a larger content window and more hidden nodes can contribute to better performance.	i) Consideration using labeled speech data to fine-tune the parameters of the network is done to further improve the performance. ii) It is necessary to search or assemble large emotional data to prepare enough data for a deep network. iii)Salient features could be found which can be robust to environmental distortion or speaker variation by adding some penalty terms, since a deep learning method may be sensitive to small perturbations in the input features.
6		Jun Deng et al	Sparse Autoencoder -based Feature Transfer Learning for Speech Emotion Recognition	sparse autoencoder method		

7	Hector P et al	Learning Deep Physiological Models of Affect	auto encoders, convolutional neural networks	i) Application of deep learning (DL) to the construction of reliable models of affect built on physiological manifestations of emotion in which DL outperforms manual ad-hoc feature extraction as it yields significantly more accurate affective models. ii) Results suggest that the method is powerful when fusing different type of input signals and it is expected to perform equally well across multiple modalities.	

3.5. Survey on Other Deep learning Techniques

In 2014, Kun Han et al [45] developed a speech emotion recognition model using DNN and Extreme Learning Machine (ELM) and the experimental results revealed that the proposed DBN effectively learns the emotional features and leads to 20% increased accuracy. In 2015, Weihong Deng et al [46], developed a deep learning architecture, DeepEmo to recognize real world facial expression using the DL methods. Leimin Tian et al [47] used (LSTM – RNN) techniques to design a model for emotion recognition in spontaneous and acted dialogues. The results proved that the (LSTM- RNN) model gave better performance than the SVM model when the training data is sufficient.

Mohamed R.Amer et al [48] used the deep networks technique and designed a model for speech emotion detection. Leimen Tian et al [49] used the LSTM technique and developed a model for recognizing emotions in dialogues with acoustic and lexical features. The authors concluded that this model had the potential to improve the quality of emotional interactions in current dialogue systems. Natalia Neverova et al [50] used the multiscale and multimodal DL techniques to develop a model for gesture detection and localization. The authors proposed that the developed model can cope up with more spatial or temporal data. Table 5 summarizes the previous studies for other DL networks on facial emotion recognition.

Table 5. Previous studies on Deep Learning using other techniques

S.No	Year	Authors	Title	Deep Learning Technique Used	Contributions and Classification Accuracy Obtained	Drawbacks
1	2014	Kun Han et al	Speech Emotion Recognition using Deep Neural Network and Extreme Learning Machine	DNN, Extreme Learning Machine (ELM), simple and efficient single-hidden-layer neural network	i) Experimental results reveal that the proposed DBN approach effectively learns emotional information from low-level features which leads to 20% relative accuracy improvement compared to the state of- the-art approaches. ii)Boosts the performance of emotion recognition from speech signals which is very promising to use neural networks to learn emotional information from low-level acoustic features	

2	Leimin Tian et al	Emotion Recognition in Spontaneous and Acted Dialogues	Long Short-Term Memory Recurrent Neural Networks (LSTM-RNN)	LSTM-RNN model gives better performance than the SVM model when there is enough training data.	i) Complex structure of a LSTM-RNN model may limit its performance when there is less training data available, and may also risk over-fitting. ii)The predictive power of other knowledge-inspired acoustic features could be studied. iii) A hierarchical emotion recognition model could be build that combines different types of features at different levels based on their nature, such as whether the features are utterance-level features or frame-level features. iv) It is also necessary to examine whether the findings generalize to other databases of English dialogues annotated with dimensional emotion annotations, such as the Belfast naturalistic database
3	Leimin Tian et al	Recognizing Emotions in Dialogues with Acoustic and Lexical Features	Long Short-Term Memory Recurrent Neural Network (LSTM)	The model will have the potential to improve the quality of emotional interactions in current dialogue systems	To improve the performance of the emotion recognition models by including global prosodic features describing duration, speaking rate, pitch, energy, amplitude, and spectral features of the utterances is effected.
4	Natalia Neverova et al	Multi-scale deep learning for gesture detection and localization	multi-scale and multi-modal deep learning	i) General method for gesture and near-range action detection from a combination of depth and intensity video and articulated pose data is proposed. ii) Model can be extended by adding alternative sensory pathways without significant changes in the architecture. iii) Model can elegantly cope with more spatial or temporal scales	i) A deeper exploration into the dynamics of cross-modality dependencies could be identified. ii) Considering full signal reconstruction or explicit feedback connections as in the case of DBMs, the model would be helpful in the case when the input from one or more modalities is missing or noisy

4. Conclusion

Emotion recognition is a remarkable research area in the current scenario. Emotion identification using deep learning techniques is the blooming methodology used by the researchers to develop innovative models. This paper attempts to present various techniques that can be used to recognize the emotions using FEMG signals and also using speech signals. The review on different Deep learning algorithms are discussed inorder to classify the emotions. Though Deep Learning techniques offer several advantages like easy training, usage of shared weights etc, there are some limitations with the deep learning techniques.

The limitations are: Firstly, interpretation of the model using Deep learning is difficult since it has many layers with many nodes. Secondly, for analysis in explaining variance or to attribute outcomes

to treatments, deep learning technique is not the efficient method to propose. Only partial dependency plotting could be done to visualize the deep learning model. Thirdly, the Deep learning technique has the capability to overlearn the training data in which it memorizes the characteristics of the training data that may or may not generalize to the environment where the model will be used. Further the research could overcome the limitations of the deep learning techniques and make it useful to create models for the real time environment.

References

[1] Adarsh Pardhan, Neelam Agarwalla, "Deep Learning using Restricted Boltzmann Machines", *International Journal on Advanced Computer Theory and Engineering (IJACTE)*, ISSN (Print): 2319-2526, Volume -4, Issue -3, 2015, pp : 10-15.

[2] Chenchen Huang et al, "A Research of Speech Emotion Recognition Based on Deep Belief Network and SVM", *Mathematical Problems in Engineering Hindawi Publishing Corporation*, Volume 2014, http://dx.doi.org/10.1155/2014/749604, pp:1-7.

[3] Ankit Awasthi, "Facial Emotion Recognition using Deep Learning", Project Report submitted to Indian Institute of Technology Kanpur.

[4] Labmaster, "Automatic Emotion Recognition system using Deep Belief Network", Jan 2015.

[5] E. M. Albornoz et al, "*Spoken emotion recognition using deep learning*", 19th Iberoamerican Congress on Pattern Recognition (CIARP 2014), Nov, 2014.

[6] Zheng-wei HUANG et al, "Speech emotion recognitionwith unsupervised feature learning", *Frontiers of Information Technology & Electronic Engineering*, 2015, 16(5): pp: 358-366.

[7] Xiaowei Jia et al, "A Novel Semi-supervised Deep Learning Framework for Affective State Recognition on EEG Signals".

[8] Kihyuk Sohn et al, "Improved Multimodal Deep Learning with Variation of Information".

[9] John Goddard, "Spoken Emotion Recognition using Restricted Boltzmann Machines and Deep Learning".

[10] Branislav Popović, Stevan Ostrogonac, Vlado Delić et al," Deep Architectures for Automatic Emotion Recognition Based on Lip Shape", *INFOTEH-JAHORINA* Vol. 12, March 2013,pp:939-943.

[11] Hiranmayi Ranganathan et al, "*Multimodal Emotion Recognition using Deep Learning Architectures*". EEE Winter Conference on Applications of Computer Vision (WACV) (2016).

[12] Xuanyang Xia et al," A biologically inspired model mimicking the memory and two distinct pathways of face perception".

[13] Yelin Kim et al," Deep Learning for Robust Feature Generation in Audiovisual Emotion Recognition".

[14] Yan Zhang et al," Speech Recognition Using Deep Learning Algorithms"

[15] Kevin Terusaki et al, "Emotion Detection using Deep Belief Networks", May 9, 2014.

[16] Geoffrey E. Hinton and Simon Osindero et al," A fast learning algorithm for deep belief nets", Neural Computation 2006.

[17] Bu Chen et al," A Study of Deep Belief Network Based Chinese Speech Emotion Recognition", *Computational Intelligence and Security (CIS)*, 2014 Tenth International Conference on 15-16 Nov. 2014, pp: 180 – 184, ISBN:978-1-4799-7433-7, IEEE publisher.

[18] Wei-Long Zheng et al," Eeg-Based Emotion Classification Using Deep Belief Networks",.

[19] Ping Liu et al," Facial Expression Recognition via a Boosted Deep Belief Network", Computer Vision Foundation, IEEE Explore.

[20] Erik M. Schmidt et al, "Modeling and Predicting Emotion in Music",

[21] Jaiswal et al, "*Deep learning the dynamic appearance and shape of facial action units*", In: Winter Conference on Applications of Computer Vision (WACV), 7-9 March 2016, Lake Placid, USA. (In Press).

[22] Quanzeng You and Jiebo Luo et al," Building a Large Scale Dataset for Image Emotion Recognition: The Fine Print and The Benchmark", Association for the Advancement of Artificial Intelligence.

[23] Quanzeng You and Jiebo Luo et al," *Cross-modality Consistent Regression for Joint Visual-Textual Sentiment Analysis of Social Multimedia*", WSDM'16, February 22–25, 2016, San Francisco, CA, USA,2016 ACM. ISBN 978-1-4503-3716-8/16/02.

[24] Steven K. Esser et al, "Convolutional Networks for Fast, Energy-Efficient Neuromorphic Computing", 2016, pp:1-7.

[25] Carlos Argueta," Facial Emotion Recognition: Single-Rule 1–0 DeepLearning", Feb 2016.

[26] Hong-Wei Ng et al," Deep Learning for Emotion Recognition on Small Datasets Using Transfer Learning", ICMI'15, November 09-13, 2015, Seattle, WA, USA, 2015 ACM. ISBN 978-1-4503-3912-4/15/11.

[27] VICTOR-EMIL NEAGOE et al," Deep Learning Approach for Subject Independent Emotion Recognition from Facial Expressions", *Recent Advances in Image, Audio and Signal Processing*", ISBN: 978-960-474-350-6, pp:93- 98.

[28] George Trigeorgis et al," Adieu Features? End-To-End Speech Emotion Recognition Using A Deep Convolutional Recurrent Network".

[29] Samira Ebrahimi Kahou et al," Recurrent Neural Networks for Emotion Recognition in Video", Nov 2015.

[30] Gil Levi," Emotion Recognition in the Wild via Convolutional Neural Networks and Mapped Binary Patterns", EmotiW'15, November 9, 2015, Seattle, WA, USA, 2015 ACM. ISBN 978-1-4503-3983-4/15/11 ...$15.00.
[31] Amogh Gudi et al," Deep Learning based FACS Action Unit Occurrence and Intensity Estimation",Vicarious Perception Technologies, Amsterdam, The Netherlands.
[32] Amogh Gudi et al, "Recognizing Semantic Features in Faces using Deep Learning", Master's Thesis, University of Amsterdam.
[33] Pablo Barros et al," Multimodal emotional state recognition using sequence-dependent deep hierarchical features", Neural Networks, 72 (2015),pp: 140–151.
[34] Yilin Yan et al," Deep Learning for Imbalanced Multimedia Data Classification".
[35] Li Wang et al," Deep Learning Algorithms with Applications to Video Analytics for A Smart City: A Survey".
[36] Moez Baccouche et al," Sequential Deep Learning for Human Action Recognition", HBU 2011, LNCS 7065, pp. 29–39, 2011, Springer-Verlag Berlin Heidelberg 2011.
[37] Claudio Aracena et al," Towards An Unified Replication Repository for EEG-based Emotion Classification",.
[38] Wei Liu et al," Multimodal Emotion Recognition Using Multimodal Deep Learning", Intelligent Interaction and Cognition Engineering.
[39] Otkrist Gupta et al," Deep video gesture recognition using illumination invariants", Massachusetts Institute of Technology, 21 Mar 2016.
[40] Suwicha Jirayucharoensak et al," EEG-Based Emotion Recognition Using Deep Learning Network with Principal Component Based Covariate Shift Adaptation", *The Scientific World Journal* Volume 2014, pp:1-10.
[41] Salah Rifai et al," Disentangling factors of variation for facial expression recognition".
[42] Zheng-wei HUANG et al, "Speech emotion recognition with unsupervised feature learning", *Frontiers of Information Technology & Electronic Engineering*, 2015 16(5),pp:358-366, ISSN 2095-9230.
[43] Jun Deng et al," Sparse Autoencoder-based Feature Transfer Learning for Speech Emotion Recognition".
[44] H´ector P et al," Learning Deep Physiological Models of Affect", *IEEE COMPUTATIONAL INTELLIGENCE MAGAZINE*".
[45] Kun Han et al," Speech Emotion Recognition Using Deep Neural Network and Extreme Learning Machine", INTERSPEECH 2014, Work at Research internship at Microsoft Research, ISCA 2014, 14-18 September 2014, Singapore.
[46] Weihong Deng et al," DeepEmo: Real-world facial expression analysis via deep learning".
[47] Leimin Tian et al," Emotion Recognition in Spontaneous and Acted Dialogues",.
[48] Mohamed R. Amer et al," Emotion Detection In Speech Using Deep Networks", Work done while being an intern at SRI International.ICASSP 2014.
[49] Leimin Tian et al," Recognizing Emotions in Dialogues with Acoustic and Lexical Features".
[50] Natalia Neverova et al," Multi-scale deep learning for gesture detection and localization",pp:1-16