International Journal of Applied Engineering Research ISSN 0973-4562 Volume 16, Number 2 (2021) pp. 103-110

Emotion Recognition – A review

Sukhpreet Kaur[1], Nilima Kulkarni[2]

[1] *CSE,MIT SOE, MIT ADT University, Pune, India.*

[2] *CSE,MIT SOE, MIT ADT University, Pune, India.*

Abstract

Emotion recognition plays a crucial role in the era of Artificial intelligence and Internet of things. It offers tremendous scope to human computer interaction, robotics, health care, biometric security and behavioral modeling. Emotion recognition systems recognize emotions from facial expressions, text data, body movements, voice, brain or heart signals. Along with basic emotions, attitude, control over emotions and power of activation of emotion can also be examined for analyzing sentiments. This paper identifies various supervised and unsupervised machine-learning techniques for feature extraction and emotion classification. Comparative analysis has also been made of various machine-learning algorithms used in referenced papers. It tells the scope and applications of automatic emotion recognition systems in various fields. This paper also discusses various parameters to increase the accuracy, security and efficiency of the system.

Keywords: Sentiment analysis, emotion recognition, compound emotions

1. INTRODUCTION

Emotion recognition is the study of recognizing six universal expressions (anger, joy, fear, happiness, sadness and surprise as shown in Figure 1 using various computer science techniques. As emotions reflect one's state of mind by his/her unintentional actions that may or may not be paralinguistic. So, Emotions of a person are recognized using his/her behavioral characteristics such as voice, handwriting, facial expressions, brain signals (EEG), heart signals (ECG) etc. Behavioral characteristics are not only used to identify a person but also help to recognize emotions, thus these are also known as soft biometrics [16]. Soft biometrics can be classified as physical traits, behavioral traits and human adhered characteristics. Such as height, weight, skin color, eye color are physical traits, voice, gait, keystroke are behavioral traits and clothes, accessories are human adhered characteristics. Soft biometrics help semantic interpretation of a person's thoughts, feelings, actions and appearance to recognize emotions.

Fig. 1 Universal emotions

In addition to emotion recognition, various other factors such as valence, polarity, arousal play prominent roles in identifying one's state of mind. By [9], mapping of the brain using valence, polarity, arousal and emotion recognition is called Sentimental Analysis. Sentiment analysis is used to understand the person's opinion and attitude towards a particular topic or at that instant of time using various computational approaches as shown in Figure 2.

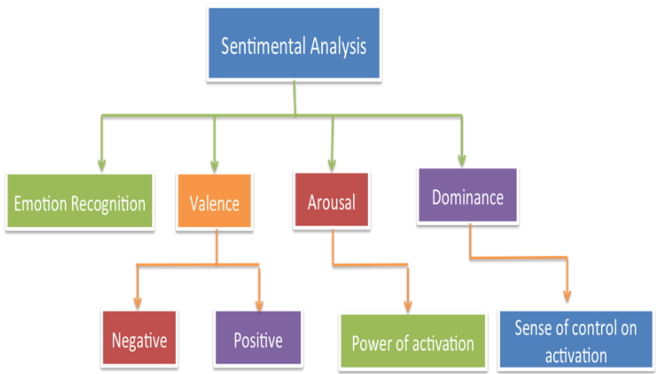

Figure 2. Sentiment analysis

Emotion recognition has wide scope in many areas such as human computer interaction, biometric security etc. So it provides insight into artificial intelligence or machine intelligence that uses various supervised and unsupervised machine-learning algorithms to simulate the human brain. It was explored that study of human emotions, their interpretation, processing and adaptation by machines is known as affective computing or artificial emotional intelligence [18]. Human emotional state can be recognized from facial expressions, body movements, speech, text writing, brain or heart signals etc. using various machine-learning techniques that extract required features or patterns from the collected data.

Advancements on existing or fundamental features or approaches of emotion recognition system, affective computing are being carried on to have a more robust, secure and efficient emotion recognition system as given in Table 1.

Table 1. Parameters to increase accuracy, security and efficiency of emotion recognition system

Advanced features	Purpose
Compound emotions	Compound emotion is a pair of dominant & complementary emotions e.g. happily-surprised, angrily-sad, surprisingly-happy, fearfully-sad. According to [4], There are a total fifty emotions. Out of the total, forty-nine emotions are compound emotions and 50th one is neutral emotion. It leads the system to be more specific towards the category of emotion. As suggested by authors, future work is to improve the recognition rates and advance the research on the field of compound emotions.
Multi-modal system	Combining different modalities like body movements, voice, text or facial expressions is known as multi-modality. Emotions extracted from multi-modalities are fused together to increase the efficiency of the system by considering all the security and privacy concerns of emotion recognition systems.
Micro expressions	Emotions are always expressed both ways subtle and coarse [5]. To increase the accuracy of the system, it must be able to identify low intensity expressions (subtle emotions) also like furrows, skin texture etc.
Uncontrolled, in-wild Datasets	To move towards a more naturalistic emotion recognition system, databases should be collected under natural and uncontrolled conditions (e.g. SEMAINE database), the Labelled Faces in-the Wild (e.g. LFW dataset). Which helps in simplifying the estimation of valence, arousal and dominance, described in [3].
Cancelable Biometrics	To keep the database safe or secure, various transformations are applied on original data to protect it from data breaches. These transformations distort the features extracted from different modalities to increase the privacy using template protection schemes. It uses two types of functions: Invertible and Non invertible. Invertible functions use biometric salting and specific parameters are applied to make transformations. In invertible transform, distorted data can be reverted back into original data. Hence, parameters have to be kept secret. Non invertible functions use three types of transformations: Polar, Cartesian and spatial. Modified templates cannot be reverted back to original ones. As mentioned in the paper [19], cancelable biometric protected template is more secure unless performance is compromised. These transformations on biometric templates are known as cancelable biometrics. Future scope could be to form non-invertible functions without compromising the performance.

2. LITERATURE REVIEW

As described [1], there are five basic emotions: happiness, sadness, fear, anger and neutral which were recognized from multiple body movements such as head region, joints, upper and lower body movements, arm bound space to improve the accuracy of emotion recognition system. They have used video datasets to extract motion or kinetic features from speed, space, and symmetry of various body parts under three scenarios as walking, sitting and action independent cases. On extracted geometric and temporal features, ANOVA (Analysis of Variance) and MANOVA (Multivariant Analysis of Variance) were applied to compute relevance of extracted features and normalization of features. To fuse the features, score and rank level fusion techniques were used. As given in the paper, accuracy was 90% in walking, 96.6% in sitting and 86.66% in action independent cases were identified. Thus it can be seen that the feature extraction framework has better understanding of emotions than human beings. Future scope stated in the paper are 1) Along with the body movements, voice and facial expressions can also be taken to improve the performance because action independent scenarios give less accuracy 2) Enhancement in tools to improve communication between human and robots 3) Remote sensing of emotions in case of emergency 4) To implement better tools for training programs in medical rehabilitation centers. 5) Recognizing emotions from body movements is yet to be explored more.

Person's online handwriting (text dependent & text independent) & signatures were used to recognize emotional status i.e. happy, sad and stress [2]. CIU (Cyprus International University) had collected handwritten database as fixed and variable task, based on physiological scenarios (before and after watching positive and negative videos) and with corresponding several data labels (i.e. relevant ground truth information including identity, age, gender, ethnicity and emotion). Extracted features from signature and handwriting were average pen velocity, max velocity, pressure in both x & y direction, altitude, no. of times pen passes through midline. Extracted features were normalized using z-score normalization approach under the second preprocessing phase. K-Nearest Neighbour algorithm on WEKA (Waikato Environment for Knowledge Analysis) tool was used to classify the emotions from preprocessed normalized features. As results shown in paper, stress prediction achieved highest accuracy from handwriting and happiness from signatures. Future scope suggested by authors is database can be improved by making it hybrid i.e. combination of both online (dynamic features recorded in txt format) and offline (static features recorded in jpg format) to improve accuracy in all states of emotions.

As per the experiments performed in [3], emotion recognition was carried out under uncontrolled or natural conditions. Database i.e. spontaneous facial expressions and gestures were recorded in an uncontrolled environment and was known as in-wild conditions database. It was named as AM-FED+. This database was focusing on facial expressions, body gestures, speech, as well as various other sensors. Here,

research could be grouped as in automatic analysis of behaviour follows: (i) Recognizing a set of discrete expressions, called six universal expressions (i.e., Anger, Disgust, Fear, Happiness, Sadness and Surprise) and neutral, (ii) Detecting particular non-universal expressions (e.g. recognition of pain and compound expressions), (iii) Detecting Facial Action Units (FAU) in expressive sequences, that facial muscles' movement (iv) Estimation of sentimental analysis parameters such as valence, arousal and dominance. In this paper, authors have given insight into universal & non universal expressions, facial muscle movements, sentimental analysis and naturalistic dataset. Future scope mentioned in the paper is to challenge human intelligence for creating enough capable machines that can understand and interpret human emotions under uncontrolled behaviours.

Automatic facial emotion recognition domain is an amalgam of behavioural sciences, artificial intelligence and neurology as stated in the paper [5]. Authors describe that facial expressions are effects produced by the motor and sensory system in the form of signals from the brain. This paper explained various advanced datasets (RGB, 3D, thermal and multimodal) and feature extraction techniques to increase the efficiency of the system. An emotion can generate more than one facial expression, so to clearly identify it, RGB, 3D, thermal and multimodal data were captured. It also inferred the facial expression (FE) effects in terms of categorical (types of primary emotions i.e. joy, sadness, happiness, surprise, fear, anger, disgust), dimensional (Valence & arousal), physiological specificity (temperature of skin, movement of blood through tissues, spatial variations in facial muscles) and microexpressions. As for future scope, authors have concluded that improvement in registration algorithms to work under an uncontrolled or natural environment using 3D, RGB, multimodal datasets (visual and non visual modalities) to recognize microexpressions is required. Secondly in 3D dataset rotations, occlusions, multiple person detection is challenging. Thirdly, machines can also be trained using new datasets to form neural networks. Consequently, immense future work could be seen to form effective AFER (Automatic Facial Emotion Recognition) systems.

This paper introduced a multi-modal system to enhance facial expression recognition results using deep neural network technique [6]. An extended database (CK+) with added label preserving data to evaluate the results were used. It also resolved the problem of data overfitting and data imbalance where data overfitting means learning the data to the extent that even noise and other random functions are also picked up. Then it can't be applied to new datasets. Data imbalance means unequal distribution of data to classes because of larger datasets. First step was aggressive data augmentation applied to address data overfitting and data imbalance issues by increasing the diversity of data for training big neural networks. Second step forming two layers one was to extract appearance based features such as mouth, eyes, nose empirical facial key points and second was for high level features to train neural network CNN (Convolution neural network) was used. Thereafter the SVM (Support vector machine) algorithm was applied to classify the extracted features as expressions. Authors have made the comparison of proposed model with other feature extraction strategies resulting in a proposed model showing the highest accuracy of 94.41%. To explore multi-modal (layered deep neural network) feature strategy, applicable to poor quality facial images could be the future scope, mentioned in paper.

In the paper [8], authors have discussed the application of sentiment analysis to compare the popularity of McDonalds and KFC based upon it's tweets (reviews given by people on twitter). Sentiment analysis of unstructured twitter data of Mcdonalds, KFC using supervised & unsupervised machine learning algorithms was performed. Real unstructured data as database to system collected using twitter API. In the preprocessing phase, data cleaning was carried on unstructured collected data to remove redundant words, punctuation symbols etc. and converting emoticons to ASCII values. In the second phase, using unsupervised lexicon analysis, preprocessed tweets were classified as positive, negative and neutral. Then to classify the data further, supervised naive bayes, SVM (Support Vector Machine), maximum entropy (maxtent), decision tree, random forest and bagging approaches were applied to get the best results. As per the metrics calculated by authors, recall, precision and F score, cross validation for all supervised machine learning algorithms were calculated for both McDonald & KFC preprocessed data. Where maxtent was found to be the best fitting algorithm with the highest accuracy of 78%. Results in the paper proved that McDonalds is more popular than KFC. They have suggested the future scope be to design an algorithm that can automatically classify the tweets.

Paper [9] have proposed a supervised approach that built a topic-adaptive sentiment lexicon model (TaSL). TaSL model worked on topic rather than a word to analyse the sentiments. It has been described in paper that a word may have different meanings, depending upon the topic it belongs e.g. "amazing" may mean surprised or impressive or excellent (sentiment polarity of word). “Positive” if attached with the word, feeling becomes a positive word but with HIV becomes a negative word. In the first phase, using hierarchical supervision information of documents and words, TaSL captured the sentiment opinions or polarities of words under different topics were gathered, formed a topic specific lexicon set and showed better performance of sentimental analysis when compared to other approaches. Results produced in the paper showed that a series of experiments on four real-world databases have shown that TaSL model consistently outperformed the previously existing semantic lexicons. As stated in paper, future work could be on unlabeled data using semi supervised machine learning algorithms to remove data imbalance issues.

A dynamic database, captured using RGB-Depth cameras for human body movement detection in 3D form was introduced by [10]. It could be significantly used in many applications such as human computer interaction, robotics, virtual reality etc. Multiple RGB-Depth cameras were mounted to detect key descriptors of head, limbs, legs and occluded parts of body, noise & illumination changes also. It followed the process of 1. Data unification (collecting data from multiple cameras and merging it to single dataset) 2. Body parts tracking (mapped to coordinate system) 3. Noise removal (A Kalman filter-based method for noise removal and locating key points) 4. Body parts’ orientation estimation. Results were calculated in the paper using the measures: tracking accuracy, accuracy recognition, motion synthesis. In relation to results, authors have recommended the future scope as still the technique needs to be improved as it has shown less accuracy while hands and feet poses make variations frequently. It has also been built for detecting one person only at a time.

The concept of sentimental analysis from paralinguistic feature i.e. speech along with emotion recognition was illustrated by [11]. It was called as affective computing. They have used the personalized dataset by augmenting five basic emotions’ utterances, called as AESDD (Acted Emotional Speech Dynamic Database). Where the database was personalized by adding data from various sources such as news reporting, social communication, theatrical emotion cues. It can also be helpful in understanding one’s attitude or sentiments towards his/her speech. Database of audio, video and text was generated from speech itself using various approaches. After feature extraction of these modalities, various classification algorithms Multi- Class Logistic Regression (Log.Reg.), Multi-Layer Perceptron (MLP), Support Vector Machines (SVM) with polynomial kernel, Logistic Model Trees (LMT) and a Subspace Ensemble Learning classifier (Ensemble) were tested by authors on both personalized and general database. Consequently, results showed that ensemble has the highest performance. Future work proposed by authors that it could be making multiple databases personalized to form a more robust system.

Advancements in machine learning algorithms that accurately work on large RGB data to recognize emotions and the intensity of pain using cloud based GPU hardware have been proposed in the paper [12]. First RGB training data was fed into the machine learning model to train it and then input was provided to the model for obtaining the output. As per the understanding, work’s summarized as (i) RGB frames were extracted from video input and converted into gray scale. (ii) Face detectors detected the face and applied image classifiers (iii) Over WAN, communication to the distributed cloud was made by feeding data through high resolution cameras and emotions were recognized using classifiers. It was performed by cloud based h/w i.e. GPU. (iv) Coherent facial emotion classifiers (multilayer neural networks) produced the results. Results in the paper are concluded as remote cloud based GPU environments increase the time efficiency and security of supervised models and data is stored on local as well as remote hardware to avoid tough time. The author suggested obtaining a deep convolutional neural network by removing multilayers of neural networks to make it capable of learning more complex features and a larger database is to be created by including more patient images by applying data augmentation techniques to increase the fair visibility of the system, as future work.

In the paper [13], authors have worked on an application based system to control the psychological problems rising at a very high pace in this tech savvy world. A self-serving device, psychometric analyzer has been built that works on the history of patient and recorded voice as an input. It detected the intensity level (high or low), emotions (universal emotions), polarity (positive or negative) and subjectivity (biased or unbiased) from given input. Data acquisition was done using supervised, unsupervised and self-learning algorithms as Support vector machine (SVM), K-means cluster algorithm and using API’s data to be extracted from various medical sites respectively. Next step was to classify sentiments and emotions using SVM regression and decision trees. Final step, data visualization was produced in the form of graphs. Results in the paper show that it’s a time efficient system. Authors have proposed future work as to build a model with best possible time-cost-accuracy tradeoff to achieve the desired results.

Two new trends in computational linguistics: 1. Affective computing & real time brain signal machines for emotion recognition and 2. EEG (Electroencephalogram) brain signals have introduced in [14]. Where sentimental analysis architecture was used to find out positive/negative attitude, valence and arousal. To recognize the text from speech, three approaches 1. Knowledge based (collecting data from web sources) 2. Statistical model (using deep learning algorithms) 3. Hybrid approach, were followed. Real time brain signals (EEG) were needed for analyzing the sentiments through polarity i.e. positive or negative attitude. Because only textual information has always been there with a lot of redundancies and excessive computational time. Suggested future work in the paper is listed as: 1. To find the resources of emotion dataset, not related to english language only but also to others. 2. The accuracy can also be improved by applying different feature extraction techniques so as to form relation between emotion detection and EEG or other anatomical signals.

In the paper [15], authors have designed an application framework for rehabilitation training centres to help stroke patients getting out of anxiety or any other negative mental state. In this paper, ball catching game was simulated using virtual reality concept for creating rehabilitation training equipment. It worked on recognition of facial expressions, body movement and voice modality to make the game interesting and interactive for patients. Generally, as we know, after getting stroke patients face difficulty in hearing, speaking, understanding as their thought process slows down. So, patient's color & 3D depth images and voice were detected to analyse body skeleton structure through extracted features. By sensing the action of the limbs of the patient, it increases the difficulty level of the game and changes the position of the ball itself. This way, patients interact with the game through a virtual environment and enjoy the game. Subsequently, It improves the mental state of patients and makes them excited for next training sessions as per the results shown in paper. Future work has been considered by the authors is to make the system more effective by continuously deep observing patients and sensing their emotions.

In [16], detailed information on soft biometrics was provided. It has been told that soft biometric is a behavioural characteristic of a body that can give some information about one's identity but not exactly recognizing his/her identity. Soft biometrics help semantic interpretation of a person's thoughts, feelings, actions and appearance to recognize emotions. It's reflection can be seen through physical traits such as height, weight, skin color, eye color, behavioural traits such as voice, gait, keystroke and human adhered characteristics like clothes, accessories. To improve the efficiency (accuracy, cost, error rates) of biometric recognition system, a multimodal (combination of more than one modality) approach was brought. It leads to greater cost as the same number of sensing equipments is also required. So as to overcome increased cost of the system, along with primary biometric, soft biometric can also be identified through the same. After the discussion on soft biometrics, authors have suggested the future work as to design optimal fusion schemes for primary biometric & soft biometric. Soft biometrics are also to be extracted without creating inconvenience to users.

Authors worked on recognizing non-linguistic facial expressions and head movements for finding severity of depression [17]. Because depressed patients avoid giving expressions and maintain distance from others. It formed a relation between non-verbal expressions and depression. Facial expressions and head movement were detected from recorded videos. Facial features were extracted using SVM classifiers. Head turns & movements were extracted using 3D CSIRO tracker for feature extraction. Extracted features showed the difference between depressed patient and non depressed patient. Measures for facial expressions and head movements were AUs (Action units) & head amplitude, it's velocity. Results produced in the paper showed that affiliative expressions were high in depression abated patients and non affiliative expressions were high in depressed patients. Future works recommended in the paper are 1. In addition to valence and behaviour, other emotions like fear, anger, sadness are also to be analysed. 2. Completely independent automatic reliable systems are required for clinical research.

A model has been proposed on the basis of the functioning of amygdala (part of the human brain, responsible for producing emotional instincts) using artificial intelligence affective computing framework in the paper [18]. This has overcome the limitations of previous work done for emotion recognition. That was used to identify external emotions only using speech or voice labeled data. Whereas the human brain generates two types of instantaneous and real-time emotions. As per the work done in this paper, instantaneous emotions were recognized from facial expressions and voice using an 8-layer convolutional neural network (CNN). Real time emotions were extracted using memory layer neural networks (MLNN). Hence, to produce internal or intracranial emotions, instantaneous and real time emotions were fused together using a hidden markov model (HMM). Also to be called as personalized emotions that described dominance, influence, steadiness and compliance of human behavior. Suggested future work is to build an interactive artificial intelligence framework to control over human emotions.

Merits and demerits with results of different algorithms from studied paper are discussed in Table 2 as follows:

Table 2. Merits and demerits of various emotion recognition algorithms

Algorithms used	**Merits**	**Demerits**
In paper [1], ANOVA and MANOVA are used at the first and second layer respectively for feature selection of body movements. Then as feature subset selection, Support vector machine (SVM), Linear discriminant analysis (LDA), Gausian naive bayes (GNB) and Decision trees (DT) were used individually with the combination of ANOVA & MANOVA.	As per the results shown in paper, ANOVA & MANOVA in combination with SVM have given best results for emotion recognition rate i.e. 81.2 %. It becomes a highly robust system.	Proposed model in paper was applied on both person independent and person dependent movements. Being robust, it gave less accuracy for person independent scenarios.
It was discussed deep learning feature extraction techniques i.e. Convolution Neural Network (CNN) and Deep Boltzmann Machine (DBM)	Unlike static feature extraction techniques, these dynamic feature extraction techniques can also extract relevant implicit features from	For increased amount of labelled data and complex RGB data, deep learning algorithms' work is undefined yet.

Algorithms used	Merits	Demerits
algorithms for automatic facial emotion recognition [5].	various consecutive frames of facial expressions.	
In the paper [6], for feature extraction, low level empirical and self learning CNN algorithms were used. Further for feature classification, Support Vector Machine algorithm (SVM) was used.	Deep learning CNN helped in finding complex nonlinear variations in facial appearance using data augmentation and small filters at various layers of CNN. Proposed system performe d well in dealing with complicated dataset also, giving an emotion recognition rate of 94.41%.	
They used 2-D CNN and 3-D CNN for audiovisual recognition [7]. Henceforth, extracted audio and video features were fused together using ELM (extreme learning machine).	It showed accuracy of 99.9% on audio visual big data.	But it could not perform well for sentence independent database named eNTRANCE. As it gave accuracy of 86.4% only for the above database.
In [22], authors used RNN (Recurrent neural networks) and it's variants for recognizing text, audio video twitter data for sentiment analysis.	While the comparisons were made with CNNs, it showed better accuracy for audio and text data i.e. 62.10 & 80.85%.	RNN gaves less accuracy for multimodal (text+audio+video) . It produced an accuracy rate 59.70% only for video data.
RNN gave less accuracy for multimodal (text+audio+video) . It produced an accuracy rate 59.70% only for video data.	Hence, MLNN + compressed CNN gave accuracy of 85.72% with an average processing time of 0.016 sec for recognizing instant or stimulus emotions.	Inspite of giving good accuracy it lead to lower latency for real time data.

3. DISCUSSION

Many recent papers in the domain of emotion recognition and sentimental analysis were studied in this paper. It is observed that emotion recognition systems are high in demand to serve applications of artificial intelligence and internet of things. There is scope to build robust and reliable automatic recognition systems. Therefore, to improve the accuracy of emotion recognition systems, not only facial expressions or textual information would work but body movements are also need to consider. As body movements show kinetics and motion of body parts, which helps in detecting the emotions and sentiments of a person. Many papers have introduced naturalistic databases such as CIU handwritten database, AM-FED+, iCV-MEFED (database defined 50 classes of emotions including neutral expression) given in the paper by Guo, 2018, RGB-3D and big data. These databases help building enough capable systems that can also recognize compound emotions or subtle emotions. Since it was a huge database, cloud storage & GPU (graphics processing unit) was required to make it secure and efficient. It's learnt that Convolution Neural Network (CNN) gives higher accuracy for feature extraction and classification as compared to other approaches, discussed in table 2. Multimodal automatic emotion recognition systems become great potential for future use in various fields as shown in figure 4. As mentioned in [23], automatic emotion recognition system in healthcare, end users can be military personnel (remote areas), common people especially youth and teenagers. Henceforth, it has a significant role in the tech savvy world and needs to be embedded powerfully in current technology.

Large number of applications of Emotion recognition system becomes motivation behind analysis of emotion recognition and sentiment analysis study. So, Emotion recognition system based applications drawn from studied papers are summarized in Figure 3 as follows:

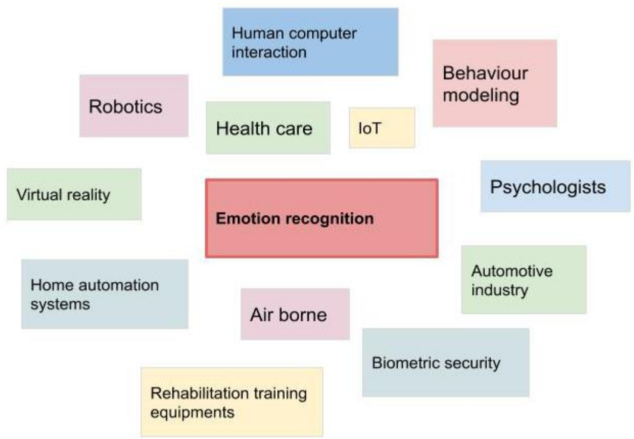

Figure 4. Applications of emotion recognition system

4. CONCLUSION

This paper provides survey of many research papers based upon emotion recognition, sentimental analysis, applications of emotion recognition systems and supervised &

unsupervised machine learning algorithms required for automatic emotion recognition. It has been revealed that self-learning algorithm Convolution neural networks produces good results for naturalistic databases, also best fitted to reduce data over fitting and data imbalance. Along with that it finds various application areas like healthcare, virtual reality, robotics etc. So, future scope of the work is to increase the efficiency of emotion recognition systems in terms of accuracy, work has to be performed using RGB datasets formed under uncontrolled conditions, deep neural networks as emotion classifiers, compound emotions, micro expressions and multi modal behavioral systems such as body movements, facial expressions, voice etc. to form robust automatic recognition systems. In addition to that, security of system can be improved by using cloud storage resources and cancelable biometrics.

REFERENCES

[1] F. Ahmed, MSA. Hossain Bari, ML. Gavrilova, “Emotion Recognition From Body Movement”, *IEEE Access 8*, doi 10.1109/access.2019.2963113, 2020, pp. 11761-11781.

[2] YB. Ayzeren, M. Erbilek, E. Çelebi, “Emotional State Prediction From Online Handwriting and Signature Biometrics”, *IEEE Access 7*, doi 10.1109/ACCESS.2019.2952313, 2019, pp. 164759-164774.

[3] MA. Nicolaou, S. Zafeiriou, I. Kotsia, G. Zhao, J. Cohn, “Editorial of Special Issue on Human Behaviour Analysis “In-the-Wild”, *IEEE Transactions on Affective Computing* 10:4-6. doi 10.1109/TAFFC.2019.2895141, 2019.

[4] J. Guo, Z. Lei, J. Wan, E. avots, N. hajarolasv, B. knyazev, A. et. al, “Dominant and Complementary Emotion Recognition From Still Images of Faces”, *IEEE Access 6*, doi 10.1109/ACCESS.2018.2831927, 2018, pp. 26391-26403.

[5] CA. Corneanu, M. Oliu, JF. Cohn, S. Escalera, “Survey on RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Recognition: History, Trends, and Affect-related Applications”, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* doi 10.1109/TPAMI.2016.2515606, 2015.

[6] W. Wei, Q. Jia, Y. Feng, G. Chen, M. Chu, “Multi-modal facial expression feature based on deep-neural networks”, Journal on Multimodal User Interfaces. doi https://doi.org/10.1007/s12193-019-00308-9, 2020, pp. 17-23.

[7] MS. Hossain, G. Muhammad, “Emotion Recognition Using Deep Learning Approach from Audio-Visual Emotional Big Data”, *Information Fusion*, doi: https://doi.org/10.1016/j.inffus.2018.09.008, 2018.

[8] SA. Rahman, FA. AlOtaibi, WA. AlShehri, “Sentiment Analysis of Twitter Data”, *2019 International Conference on Computer and Information Sciences (ICCIS)*. Saudi Arabia. doi 10.1109/ICCISci.2019.8716464, 2019.

[9] D. Deng, L. Jing, J. Yu, S. Sun, MK. Ng, “Sentiment Lexicon Construction with Hierarchical Supervision Topic Model”, *IEEE/ACM Transactions On Audio, Speech, And Language Processing*, doi 10.1109/TASLP.2019.2892232, 2018.

[10] S. Hong, Y. Kim, “Dynamic Pose Estimation Using Multiple RGB-D Cameras”, *Sensors 2018*, 18, 3865. doi 10.3390/s18113865, 2018.

[11] N. Vryzas, L. Vrysis, R. Kotsakis, C. Dimoulas, “Speech Emotion Recognition Adapted to Multimodal Semantic Repositories”, *13th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)*, doi 10.1109/SMAP.2018.8501881, 2018, pp. 31-35.

[12] AD. Torres, H. Yan, AH. Aboutalebi, A. Das, L. Duan, P. Rad, “Patient Facial Emotion Recognition And Sentiment Analysis Using Secure Cloud With Hardware Acceleration”, *Computational Intelligence for Multimedia Big Data on the Cloud with Engineering Applications,* doi 10.1016/B978-0-12-813314-9.00003-7, 2018, pp. 61-89.

[13] A. Vij, J. Pruthi, “An automated Psychometric Analyzer based on Sentiment Analysis and Emotion Recognition for healthcare”, *Procedia Computer Science 132*, doi 10.1016/j.procs.2018.05.033, 2018, pp. 1184-1191.

[14] K. Hulliyah, NSAA. Bakar, AR. Ismail, “Emotion Recognition and Brain Mapping for Sentiment Analysis: A Review”, *2017 Second International Conference on Informatics and Computing (ICIC),* Jayapura, Indonesia. doi 10.1109/IAC.2017.8280568, 2017.

[15] SC. Yeh, WY. Hwang, “A study for the application of body sensing in assisted rehabilitation training”, *2012 International Symposium on Computer, Consumer and Control*, doi 10.1109/IS3C.2012.240, p. 92, 2012.

[16] K. Nandakumar, AK. Jain, “Soft biometrics”, *Encyclopedia of Biometrics*, doi 10.1007/978 4899-7488-4, pp. 1425-1429.

[17] JM. Girard, JF. Cohn, P. Rosenwald, “Nonverbal Social Withdrawal in Depression: Evidence from manual and automatic analysis”, *Image Vis Comput. 2014* Oct; 32(10), Doi.10.1016/j.imavis.2013.12.007, 2014, pp. 641-647.

[18] C. Gong, F. Lin, X. Zhou, X. Lü, “Amygdala-inspired affective computing: To realize personalize intracranial emotions with accurately observed external emotions”, *China Communications*, vol. 16, doi10.23919/JCC.2019.08.011, 2019, pp. 115.

[19] VK. Gunjan, PS. Prasad, S. Mukherjee, “Biometric Template Protection Scheme-Cancelable Biometrics”, Lecture Notes in Electrical Engineering,Vol 570, doi

https://doi.org/10.1007/978-981-13-87159_48, 2019, pp. 405.

[20] G. Kaur, D. Singh, S. Kaur, "Pollination Based Optimization for Feature Reduction at Feature level Fusion of Speech & Signature Biometrics", *Proceedings of 3rd International Conference on Reliability, Infocom Technologies and Optimization, IEEE,* doi 10.1109/ICRITO.2014.7014771, 2014.

[21] D. Martin, J. Croft, A. Pitt, D. Strelchuk, S. Sullivan, S. Zammit, "Systematic review and meta analysis of the relationship between genetic risk for schizophrenia and facial emotion recognition", *Schizophrenia Research,* doi https://doi.org/10.1016/j.schres.2019.12.031, 2019.

[22] A. Agarwal, A. Yadav, DK. Vishvakarma, "Multimodal sentiment via RNN variant", IEEE computer society, doi 978-1-7281-0886-5/19/, Honolulu, Hawaii, 2019, pp. 19.

[23] KJ. Oh, DK. Lee, BS. Ko, HJ. Choi, "A Chatbot for Psychiatric Counseling in Mental Healthcare Service Based on Emotional Dialogue Analysis and Sentence Generation", *Proceedings of 2017 IEEE 18th International Conference on Mobile Data Management*, doi 10.1109/MDM.2017.64, 2017.