IEC
Institute of Electronics
and Computer

Journal of Artificial Intelligence and Systems, 2020, 2, 53-79
https://iecscience.org/journals/AIS
ISSN Online: 2642-2859

Emotion Recognition and Detection Methods: A Comprehensive Survey

Anvita Saxena[1], Ashish Khanna[1], Deepak Gupta[1, *]

[1] Computer Science and Engineering Department, Guru Gobind Singh Indraprastha University, New Delhi, India
Email: anvanvita2207@gmail.com; ashishkhanna@mait.ac.in; deepakgupta@mait.ac.in
*Corresponding Author: Deepak Gupta, Email: deepakgupta@mait.ac.in

How to cite this paper: Anvita Saxena, Ashish Khanna, Deepak Gupta (2020). Emotion Recognition and Detection Methods: A Comprehensive Survey. Journal of Artificial Intelligence and Systems, 2, 53–79.
https://doi.org/10.33969/AIS.2020.21005.

Received: December 23, 2019
Accepted: January 27, 2020
Published: February 7, 2020

Abstract

Human emotion recognition through artificial intelligence is one of the most popular research fields among researchers nowadays. The fields of Human Computer Interaction (HCI) and Affective Computing are being extensively used to sense human emotions. Humans generally use a lot of indirect and non-verbal means to convey their emotions. The presented exposition aims to provide an overall overview with the analysis of all the noteworthy emotion detection methods at a single location. To the best of our knowledge, this is the first attempt to outline all the emotion recognition models developed in the last decade. The paper is comprehended by expending more than hundred papers; a detailed analysis of the methodologies along with the datasets is carried out in the paper. The study revealed that emotion detection is predominantly carried out through four major methods, namely, facial expression recognition, physiological signals recognition, speech signals variation and text semantics on standard databases such as JAFFE, CK+, Berlin Emotional Database, SAVEE, etc. as well as self-generated databases. Generally seven basic emotions are recognized through these methods. Further, we have compared different methods employed for emotion detection in humans. The best results were obtained by using Stationary Wavelet Transform for Facial Emotion Recognition , Particle Swarm Optimization assisted Biogeography based optimization algorithms for emotion recognition through speech, Statistical features coupled with different methods for physiological signals, Rough set theory coupled with SVM for text semantics with respective accuracies of 98.83%,99.47%, 87.15%,87.02% . Overall, the method of Particle Swarm Optimization assisted Biogeography based optimization algorithms with an accuracy of 99.47% on BES dataset gave the best results.

Keywords

Emotion Recognition, Emotion Detection, Facial expressions, Speech Signals, Physiological signals (Electroencephalogram signals (EEG), Electrocardiogram signals (ECG)), Text semantics.

1. Introduction

As John McCarthy said, the science of Artificial Intelligence aims at making intelligent machines [1]. It is an interdisciplinary field [2] [3] overlapping with the fields of robotics, emotion recognition, data mining, human computer interaction to name a few. The two main fields dealing with making computers capable of sensing human emotions are Human Computer Interaction (HCI) and Affective

Computing. Affective computing [4] [5] is a science under which methods are being developed that can not only replicate but also process, identify and understand human emotions. The Association for Computing Machinery (ACM) has defined human computer interaction as a domain concerned with the development of human like interactive computing systems and the major phenomena surrounding them [6].

Emotions are a vital part of human lives which play an integral role in how humans perceive and understand things [7] [8] [9] [10]. For the last three decades, a large number of methods are continuously being devised to facilitate emotion analysis; from manual methods such as through questionnaires elaborated by psychologists to methods involving computers. Today, emotion recognition through computers has many applications. For instance, Emotion recognition through physiological signals is being utilized in the creation of smart homes, smart offices. Furthermore, Facial detection methodology is being extensively used today [11] in consumer services, education services and security related applications, to name a few. This paper aims to present the extensive and comprehensive study of significant facial, audio, physiological and textual emotion detection and recognition methods that have been proposed and developed in the last decade.

The main objective of the paper is to gather knowledge and analyze all the significant emotion recognition methods which have been developed in the last decade and determine the best suited methods for facial emotion recognition, emotion recognition through speech, physiological signals and text. The paper was comprehended using more than hundred papers including survey papers, research papers and academic articles. Analysis and comparison was carried out on the basis of features, datasets and methodologies employed for detection of emotions.

To the best of our knowledge, the presented paper is a novel approach with a detailed comparison of all the significant emotion detection and recognition methods in the mentioned four domains. Earlier works [12] [13] [14] [15] [16] [17] [18] [19] presented an explicit comparison for facial recognition, speech detection, etc., never combining all the domains together.. Moreover, the paper brings into light the limitations associated with these methods, briefly discusses the future scope and new emerging fields in this area.

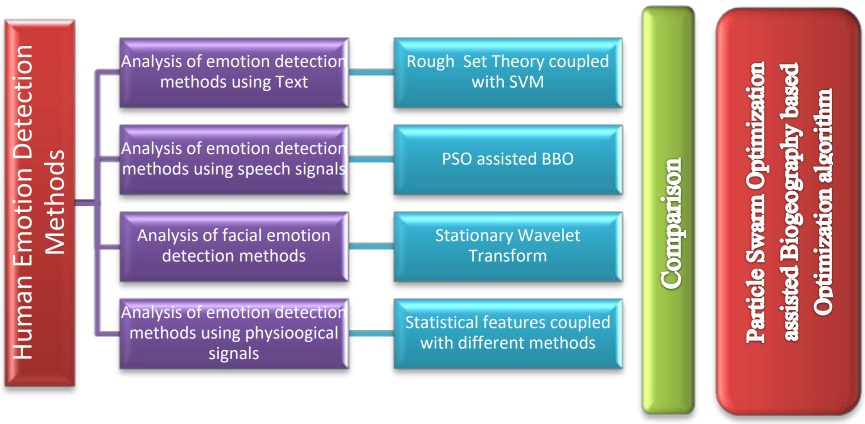

Figure 1. Graphical representation of the structure of the paper

The highlights of the paper are summarized as follows-

- The paper presents a detailed comparison and analysis of facial emotion recognition methods, models and datasets.

- The paper presents a detailed comparison and analysis of methods, models, and datasets of emotion recognition through speech and voice signals.

- The paper gives a detailed comparison and analysis of methods, models, and datasets of emotion recognition through physiological (EEG and ECG) signals.

- The paper presents a detailed comparison and analysis of methods, models, and datasets of emotion recognition through text and an inter-comparison of all the four emotion recognition methods.

- The drawbacks and future scope in the field of emotion recognition are discussed in the paper. The paper is kept as natural, comprehensible and easy as possible.

The rest of the paper is organized as follows- Second section is the Facial Emotion Detection Section which discusses the different techniques used in emotion detection through facial expressions. This section is further divided into model and feature based techniques. The third section confers about the emotion detection through speech signals. Fourth section discusses the emotion detection through physiological signals and is further segregated as EEG and ECG signals detection. The fifth section is the Textual Emotion Detection section which discusses about the detection of emotions through text semantics. The next section discusses the methodology of the best methods. The results are callibrated in the results section. The two last sections state the conclusions and future scope followed by references.

2. Facial Emotion Recognition

Neural networks are systems which are largely inspired by biological neural system. It is a framework rather than an algorithm, for many machine learning algorithms to work together. These aren't programmed with any task-specific rules. The method discussed in [20] has proposed a neural network method which evaluates seven human expressions – happy, neutral, disgust, sad, fear, surprise and anger in two steps. Multi-Layer Perceptron (MLP) and Radial Basis Function (RBF) Networks were used for classification. Another method [21] presented a system which takes the features from the outline of eyebrows, eyes and mouth using a scalable rectangle. It uses fewer features which reduces the recognition time and obtains better accuracy. Two inner canthi are used as the location points for finding the contours. For classification, Elman Neural Network of classifiers is used. Another hybrid method which uses a Wavelet Transform and Neural Network combined Ensemble [22] is discussed in which low dimension features are extracted through the image segmentation of eye and mouth regions, using the wavelet and Karhunen- Loeve transform. A bagging based neural network is used for classification.

Another Hybrid method, Convolution Recurrent Neural Network [23] system uses Convolution layers and Recurrent Neural Network (RNN). Relations within facial features are extracted by this model and then the temporal dependencies are considered during the classification by using recurrent network. The next method discussed is the Constructive Feed forward Neural Networks [24] in which feature detection is done by a 2D DCT (Discrete Cosine Transform) on the facial image and classification is done using a constructive feed forward one hidden layer neural network. Another method [25], Boosted Deep Belief Network, uses a combination of feature selector and classifier in one framework. In this model, features are jointly tuned and are selected to form a classier through a BTD-SFS process.

In a 3D meshes method [28], the uneven 3D mesh data is converted to a uniform 3D matrix by employing a unique resampling approach. The dimensionality of the features is reduced by using a Fourier spectrum of differences of the flow matrices calculated for neutral and the present expression.

A fiducial point based model is proposed in [29] in which firstly, normal localization is used for facial region detection. Then, multiple particles filter is used for the location of 26 fiducial points. According to the shift in the points, they are used as landmark points for deducing the expression for input to a basic mesh model. Then, to create a smooth wrap, elastic body spline technique is employed in the mesh. Classification is done using an Isomap-based model.

Boosting, as the name suggests is a machine learning algorithm which converts weak learning algorithms to strong learning algorithms. The method in [30] has proposed a Hybrid Algorithm in which the system is made faster by searching the prospective face regions filtered by skin color. For classification, the skin color is first scanned, and then weak and strong classifiers are applied. The next method discussed uses a popular boosting algorithm, Adaboost [31], which uses expression classifier and Haar feature based look-up table classifier.

Active Appearances Model is an algorithm used in computer vision which statistically matches the shape of the object with the appearance of an input image. A hybrid method using AAM and Manifold learning [32] is discussed, it classifies images in three distinct steps. Firstly, variant AAM features (DAFs) are calculated by using changes between AAM parameters of reference and input images. In the second step, manifold learning plants the DAFs on the feature space. And finally, recognition of expression is worked out in two steps- 1) calculation of distances using Directed Hausdroff distances and 2) selection of expression using k-NN sequences. Authors in [33] went a step further by varying the AAM and devising STAAM [33] which works in two steps. First step is performed by employing the stereo Active Appearance Model (STAAM) algorithm [33], and in the second step, a generalized discriminant analysis (GDA) classier is employed whose work is to combine 3D shape and appearance to recognize expressions.

Support Vector Machines are one of the earliest and simplest techniques of machine learning. These are supervised learning models that are used for the analysis of data models principally using regression and classification. First model discussed here, is a wavelets based method [34] in which the classification is done by using seven SVMs parallely. Each SVM classifies one expression which is later combined with others using a maximum function. Another method using Support Vector Machines (SVMs) is proposed in [35] for detection of facial expressions in live videos. Features are extracted using a facial tracker. And then are classified using a SVM. The method discussed in [36] works on the combination of features, which is done by multiple kernel learning (MKL) present in multiclass SVMs. In this method, kernel weight is calculated one at a time in the SVM, which takes into account both sparse and non-sparse kernel combinations. BBN or Bayesian Belief Networks is a probabilistic acyclic directed graph and represents the dependencies of different variables. BBN in [37] is used for development of a model which recognizes facial expressions in videos. Facial analysis is done by Kalman Filter. Feature detection process includes Principal component analysis for distinguishing facial areas. Description of expressions is done by using sets of Action Units (AUs). BBN handles the time behavior of the features. Muscle movement model [38] is a 3D/4D face emotion recognition model which doesn't require any manual work. The shape index, co-ordinate and normal are used as feature set. Optimal weights for facial regions is produced using a GA (Genetic Algorithm) and then classification is done using SVM and HMM. Further in [39], a method utilizing all facial components through spatial wavelet transform features is discussed.

Table 1. Comparison of Model Based Techniques for Facial Emotion Recognition

S. No	Method	Dataset	Recognition rate(in percentage)	Specifications	Year
1	Artificial Neural Networks	JAFFE	73	Multi-Layer Perceptron (MLP) and Radial Basis Function (RBF) Networks	2002
2	Support Vector Machines	Self-Generated	86	facial feature tracker Support Vector Machine classier	2003
3	Adaboost	JAFFE	92.4	Boosted Haar feature based weak classifiers	2003
4	Constructive feed forward neural networks	Self-generated	93.75	two-dimensional (2-D) discrete cosine transform (DCT) as feature detector, constructive one-hidden- layer feed forward neural network as classifier	2004
5	Face Profile Sequences	Self-generated	87	facial action units (AUs), a facial-action-dynamics	2006
6	Neural Networks	JAFFE	88.8	Less features extracted for better performance	2007
7	Wavelet Transform and Neural Network Combination send tabular	CMU PITTS-BURGH AU-Coded Database	98.5	Wavelet transform, Karhunen–Loe transform, neural network classifier	2008
8	3D face recognition using swarm intelligence	BU-3DFE	92.3	Based on ant colony and particle swarm optimization(ACO and PSO respectively)	2008
9	3D deformable model	CK	87	Candide-3 face model ,Tree-based classers	2009
10	MANFIS Model	JAFFE	94.29	Neuro-fuzzy based model, LBP	2009
11	AAM manifold learning model	CK	96	Active appearance model (AAM), differential facial expression probability density model, KNN classifiers	2009

12	Hybrid-Boost Learning Algorithm Model	Unknown	93.1	Skin color based segmentation, weak classifiers with the strong classifiers for classification.	2010
13	3D meshes Method	BU-3DFE	85.56	Resampling strategy, Fourier spectrum	2010
14	Wavelets based Classification using bank of SVM-send tabular	JAFFE	96	Seven SVMs working simultaneously, Distinctly classifying one expression	2011
15	Hybrid Deep Neural Network	JAFFE	77	Convolution layers, recurrent neural networks for classification	2012
16	Deformable 3-D Facial Expression Model	Self-generated	94.7	Unique proposition of locating 26 fiducial Points	2013
17	Coupled Gaussian Processes	MPFE Multiple	89 83.7	Mapping of frontal and non-frontal points followed by Coupling	2013
18	Sparse Reduced-rank Regression	BU-3DFE Multiple	64.5-87.6 80.5-92	feature extraction using grids with multi-scale sizes, expression reduction by sparse reduced-rank regression	2013
19	Boosted Deep Belief Network	JAFFE	93	Combined feature learning, feature selection, and classier construction in a single framework. Features ne-tuned jointly and selected to form a strong classier in BTD-SF process. send tabular	2014
20	lp-norm MKL multiclass-SVM	CK+ MMI GEMEP-FERA	93.6 93.6 83.6	SVM classifier with different p-norm constraints.	2015
21	Twofold random forest classier	Extended Cohn–Kanade (CK+)	96.38	Action Units were taken out from image sequences with the help of a two-step random forest classier.	2015
22	Muscle movement model	BU-3DFE BU-4DFE	83.2 87.06	Segmentation of muscle regions, Genetic Algorithm, SVM and HMM for classification.	2016
23	Stationary Wavelet Transform features	JAFFE	98.83	SWT and neural networks for classification	2017

Another system is modeled using Eigen faces approach, it uses HSV (Hue Saturation Value) for recognition of faces present in the images. After that Principal Component Analysis is applied for fitting the image into Eigen space and then Euclidean distances of the image , with the calculation of mean of Eigen. A Neuro-Fuzzy model, the Multiple Adaptive Neuro Fuzzy System (MANFIS) [41] is used which first segments the images into three regions from which feature distributions of LBP (local Binary Patterns) is extracted.

The most prevalent features employed for facial emotion recognition include Multiscale- WLD, Local Binary Patterns (LBP), Gabor Wavelets, Histograms of Oriented Gradients (HOG), Local Directional Patterns (LDP), facial landmarks. Lately, hybrid models combining two or more features are also gaining popularity.

In [42], authors have proposed a method using Gabor wavelets. Gabor filters are used to build elastic graphs for feature extraction and classification is done using elastic templates matching along with enhanced k- Nearest Neighbor Classifier. Moreover, Gabor filters in [43] are mixed with parts of frequency and orientation parameters and features are compressed with the help of Principal Component analysis (PCA) and Latent Drichilet Allocation (LDA).

In [44], Local Binary Patterns are manipulated for feature extraction. In [45], extended LBP are proposed, which enhance the distinctiveness of similar facial images. In [46] fusion of multiple features is done using spectral embedding methods, Local Binary Patterns, Multiscale-SIFT, Active Appearance Model and Gabor Magnitude features.

Another method which principally uses Local Principal Texture Pattern (LPTP) is proposed in [47], the feature computation is done by extraction of principal directions of the neighborhood, and then the differences are coded on these directions. This technique is robust against rotation changes. Another method in [48] has proposed a new PSO (Particle Swarm Optimization) algorithm which has embedded a new concept of micro Genetic Algorithm (mGA) and has used different classifiers for classification. The classification through SVM has the highest accuracy.

Table 2. Comparison of Features based techniques for FER

S.NO	Method	Database	Recognition Rate	Specifications	Year
1	Gabor wavelet transformation and elastic templates matching	CK	90.4	Segmentation of muscle regions, Genetic Algorithm, SVM and HMM classifiers.	2005
2	Spectral embedding with multiple features	JAFFE CK GWI	85.5 96 62.3	Fusion of AAM,LBP, WLD, Gabor features	2013
3	Entropy-based feature selection	BU-3DFE	90.8	Entropy based feature Selection	2013
4	Local Binary Patterns with Coarse- to-Fine Classification	JAFFE	77	LBP for feature extraction followed by two stages of classification	2004

5	Extended-LBP and Local Feature Hybrid Matching	Bosphorous	97.6	Extended LBP based on facial depth maps	2012
6	Local Principal Texture Pattern	CK	6 Class–96 7 Class- 92	Mixture of direction and contrast feature, better than LBP and LDP	2012
7	Geometric Alignment and Local Binary Patterns	JAFFE	86.1	Active mode shape algorithm, extraction of LBP, SVM classifiers	2014
8	3D Facial Expression Recognition Using Residues	CMU-Multiple	94	Spatial displacement, called residues	2009
9	Effective semantic features and SVM	CK	94.7	Active shape model, Gabor filter, Laplacian of Gaussian	2015
10	PSO	CK	98.70 10	mGA and multiple classifiers	2017

3. Speech Signals

Emotion recognition through speech is done using the variations and changes in audio signals. Speech emotion recognition has a variety of applications; it is being extensively used today in voice recognition, call centers, customer services etc. Speech emotion recognition is basically done in two steps of feature extraction and classification. In [49] a method is proposed which uses convolutional neural network (CNN) and works in two stages. First, the local invariant features (LIFs) are computed using unlabeled samples with the help of a sparse auto-encoder (SAE) with reconstruction segregation. And then in the second stage, LIF is used as the input to feature extractor. Another method is proposed in [50] which uses semi-CNN, in the first step, followed by contractive convolutional network for feature extraction. In [51] a method is proposed which uses Fourier Parameters and MFCC for improved results over MFCC. The system consists of front and back end systems, which has Gaussian Mixture Models (GMMs) for training the Multilayer Perceptron (MLP) and the Support Vector Machines (SVMs). Anchor Models are proposed in [52], which are principally based on Euclidean or Cosine distance to remove the skewed data problems. In [53] Gaussian Mixture Models are used for category-conditional distribution of speech features followed by the estimation of parameters by EM algorithm (Expectation Maximization algorithm).

Another method discussed in [54] combines the Gaussian Mixture Models and the CHMM, and tested on English and German language datasets. Another method [54] employing fuzzy interface system, using all the prosodic features of speech, i.e. , pitch, duration and energy features is tested on a self-generated dataset which comprises of an user talking to a call center machine has shown significant results. Another method discussed in [54] adopts Linear Predictive Cepstral Coefficient

(LPCC) features for the detection of dysfluencies in the audio signals. Classification is done using k-NN and Linear discriminant analysis (LDA). A method discussed in [54] presents Teager Energy Operator (TEO) features as one of the best features for speech recognition, the method is tested on SUSAS database and Bayesian Hypothesis is used for classification. Another method discussed in [54] uses Low Frequency components of speech using DWT, this is tested on a Malayalam language database which is self-generated, using three layer MLP classifier. The method described in[55] has proposed a new method for feature selection for reducing the dimensionality of the features. Different simulations on different datasets are carried out using a Particle Swarm Optimization assisted biogeography based algorithm giving accuracy as high as 99.47%.

Table 3. Comparison of Emotion recognition techniques using speech signals

S.No	Method	Database	Recognition Rate	Specifications	Year
1	Artificial Neural Networks tested on Gender-Dependent Databases	Self-generated	Male 72.055 Female 65.5	Discrete Wavelet Transform (DWT), Artificial neural network.	2009
2	Anchor Models	FAU-AIBO	44.19	Anchor Models based on cosine distances	2013
3	ASR	FAU-AIBO	67.7	Vector space modeling vs. string kernels	2009
4	CNN	SAVEE Emotional Database DES MES	71.8 57.2 60.4 57.8	LIF, SAE, SDFA	2014
5	Semi-CNN	SAVEE Emotional Database DES MES	73.6 85.2 79.9 78.3	contractive convolutional neural network to learncandidate features, a novel function for classification	2014
6	Hybrid Deep Neural Network Hidden Markov Model (DNN-HMM)	Berlin emotional database	77.92	Restricted Boltzmann Machine(RBM) based unsupervised learning, and DNN-HMMs with discriminative learning	2015
7	GMM-CHMM	Self-generated	GMM-86.8 CHMM- 77.8	Uses Pitch and energy as features	2004
8	Fuzzy C means	Self-generated	Male-63 Female 73.7	Pitch, Interface duration, System energy	2003
9	Bayesian hypothesis testing	Self-generated	TEO-based features are the best for stress classification.	HMM	2003

10	HMM N-D HMM	SUSAS	94.41	Intensity, pitch, duration	2000
11	PSO and BBO	BES Database	99.47	PSO and BBO based algorithm	2017

4. Physiological signals

Physiological signals are biochemical signals which are generated as a response to stimuli. Physiological signals are hard to extract and process, thus require an extensive preprocessing.

ECG signals or Electrocardiographic signals are the electric signals recorded for tracing the activity of the human heart. Some very promising techniques have been devised recently to detect human emotions from cardiac activity. A method proposed in [56], preprocesses the signal using Empirical Mode Decomposition (EMD) into smaller constituents called the Intrinsic Mode Functions (IMFs). Feature vectors are computed using Hilbert transform and classification is done using Multi-Class Support Vector Machines (SVMs). In the next method [57], the signal decomposition is done with the help of digital filters and feature extraction methods such as EMD integrated either with Hilbert Transform or Discrete wavelet Transform (DWT). These are used for five emotions namely (Happiness, Surprise, Disgust, Fear and Sadness). Another method [58] developed by researchers at MIT, is a signal emitting wireless device called EQ radio, in which the signals get reflected by one's body. And are finally again fed into the device for interpretation. The method presented in the paper[59] explored the method of supervised dimensionality reduction, LDA (Linear Discriminant Analysis), NCA (Neighborhood Components Analysis), and MCML (Maximally Collapsing Metric Learning), based on a three class valence arousal problem. The method improved the accuracy for valence from 55.8% to 64.1%, and for arousal from 59.7 % to 66.1% with the NCA method.

EEG signals or Electroencephalographic signals are electric signals recorded to monitor brain activity. These signals are recorded through different channels or points on the brain, and then are decomposed. The method discussed in [60] uses Support Vector Machines, k- Nearest Neighbor algorithm and Multi-layer Perceptron (MLP) as classifiers. Feature extraction is carried out using Minimum redundancy and maximum relevance method. Next method [61], along with Minimum Redundancy and Maximum Relevance method uses Principal Component Analysis (PCA) to process features and then classification is done using a combination of k-Nearest Neighbor, Support Vector Machines (SVMs) and least square distance classifiers. Method discussed in [62] uses Kernel Eigen Emotion Pattern (KEEP) for extracting features and adaptive SVM classifier for managing the problem of imbalanced EEG datasets. Another method discussed in [63] uses combination of time frequency analysis of wavelet transform, Surface Laplacian (SL) filtering and Linear Classifiers for classification. Signal decomposition is done using wavelet transform, which uses statistical features for extraction. Method discussed in [64] uses a new feature extraction technique called the Hjorth parameters which extracts features from the preprocessed EEG signals and classification is done using SVM.

A completely new method, called the Mirror Neuron System is discussed in [65] which used the process of imitation for emotion investiture. Feature extraction is done using Higher- Order Crossings analysis and a classifier combined of four different classifiers (Mahanobolis distance, SVM, QDA and k-NN) for classification. The paper presented in [66] has given a comparison of different feature selection techniques and different classification techniques; it has acquired the data using MUSE headband and has extracted features using a unique time windowing technique.

Table 4. Comparison of Emotion Recognition techniques using ECG signals

S.No	Name	Dataset	Recognition Rate	Specifications	Year
1	Human Emotion Recognition using Electro-cardiogram Signals	University of Augsberg generated dataset	57.5	Signal decomposition by Empirical mode decomposition, Feature vector composed using Hilbert- Huang transform, classification by multi-class SVM.	2014
2	Empirical mode decomposition (EMD) and discrete Fourier transform	Self-Generated	52	FFT based feature extraction	2013
3	Emotion recognition using wireless signals	Self-Generated	87	Emission and Detection of Wireless signals	2016
4	Supervised dimensionality reduction	Mahnob-HCI database	66.1(arousal) 64.1(valence)	LDA, NCA and MCML	2017

Table 5. Comparison of Emotion Recognition techniques using EEG signals

S.No	Name of the method	Dataset	Recognition Rate	Specifications	Year
1	Frequency Domain Features and Support Vector Machines	Self-Generated	66.51	KNN,SVM, multilayer perceptron as classifiers, minimum redundancy maximum relevance for feature extraction	2011
2	Support Vector Machine and Linear Dynamic System	Self-Generated	83.01	Music used as stimuli To evoke emotions, Minimum redundancy, principal component analysis, KNN, SVM and least square classifiers	2012
3	Kernel Eigen-Emotion Pattern and Adaptive Support Vector Machine	Self-generated	73.42-80	Kernel Eigen-emotion pattern for extracting features, adaptive SVM classifier	2013
4	Combination of Spatial Filtering and Wavelet	Self-generated	62 channels- 83.04 24 channels- 79.17	Surface Laplacian filtering, time-frequency analysis of wavelet transform, linear classifiers	2010

5	Higher Order Crossings method	Self-generated 1	83.3	Higher Order 5Crossings feature extraction, HOC classifier	2010

5. Textual Emotion Recognition

Emotion recognition from text is an extensively researched field with Natural Language Processing (NLP) continuously being an advancing field. Online sentiment analysis is one of the most conventional and popular ways for interpreting user's state of mind through her written text and activity on the web. Traditionally, emotion recognition through text is done by selecting emotional keywords, bag of words and N-grams. But keywords might or might not be present in a given sentence. Thus, to overcome this problem, an alternate method called the knowledge based-ANN is introduced [67] in which meaning of words in ontology is used as features. Another method [68] to recognize emotions in Chinese language using Chinese NLP is proposed in which first training and testing tables are prepared to sample the dataset and then rough set theory coupled with SVM are applied for classification. Another method [69] using semantic labels (SLs) and attributes (ATTs) is used , which is usually inferred with the help of psychological analysis. And a Separable Mixture Model (SMM) is used for identifying the correspondence between the input sentences and labels. A deep learning based personality detection method is discussed in [70] which has used common classifiers for classification and has an accuracy of 62% for identifying a personality type.

Table 6. Comparison of Emotion Recognition Techniques using text

S.No	Name of the method	Dataset	Recognition Rate	Specifications	Year
1	Knowledge Based ANN	Unknown	65	Meaning of words in ontology used	2008
2	Rough Set Theory and SVM	Self-prepared from center for Chinese linguistics PKU	87.02	Testing tables and Rough set theory and SVM	2007
3	Semantic Labels and ATTs, Separable Mixture Model	Dialogue system created from students' expressions	83.94	Semantic Labels and ATTs, Separable Mixture Model	2006
4	Deep learning based personality detection	James Pennebaker and Laura King's stream-of- consciousness essay dataset	62.68	Numerous steps feature selection and extraction, classification using two layer perceptron	2017

6. Methodology

This section discusses the methodology of the methods with the best accuracies. A detailed comparison and discussion of the features, input parameters and experimental setup is expanded in this section.

6.1. Hybrid PSO assisted Biogeography-based Optimization

The proposed algorithm used three databases to test its accuracy. The databases used are BES, SAVEE and SUSAS. BES database has samples of Anxiety, Angry, Happiness, Disgust, Sad, Boredom and Neutral emotions from ten German speakers. SAVEE database has samples of the states- Happiness, Fear, Neutral, Surprise, Sadness, Anxiety and Disgust of four English speakers. SUSAS database has samples of simulated stressful and multi- style dialogues, in which six words are vocalized in four different emotional manners (Angry, Loud, Lombard and Neutral). All speech signals were reduced to 8 kHz using sampling and divided into 256 non overlying frame samples (32 ms). The silent portions were eliminated before the feature extraction process by establishing a threshold value where each database had its separate threshold value. Linear predictive analysis method along with inverse filtering was applied to extract glottal waveforms. To attend the glottal and speech waveforms spectrally, first order pre-emphasis filter was employed. The resulting waveforms were divided into frames with a 50 percent overlap. Finally, windowing of each frame was done by hamming window method, which helped in reducing the discontinuity and distortion in the signal. The bispectral and bicoherenece features were computed and averaged for the frames. A block representation of the process is presented in figure 2.

High Order Spectra is the spectral representation of higher order cumulants of a random process. Bispectrum and Bispectral are the third order cumulant spectra. The 2D Fourier Transform is called Biospetrum Fourier. Different from the power spectrum, the bispectrum is a function of two frequencies. The normalized bispectrum is termed as bicoherence of the signal.

$$B(f_1, f_2) = E\{X(f_1)X(f_2)X * (f_1 + f_2)\} \quad (1)$$

Where B(f1, f2) is the bispectrum of bi-frequency (f1, f2), E[.] denotes the expectation operation , * signifies the complex conjugate, X(f)represents the Fourier transform of the signal .The non-linearity of the signals leads to the production of phase at frequency f1+f2. Varied voice portions were obtained for the signal as the recording duration of speech signals varied. The OpenSmile Toolbox was used to compute interspeech 2010 feature set. The features obtained are combined with these features. Feature selection was done using PSO, genetic and Tabu search algorithms. For the worst half, a modified PSO velocity and position update was applied. The classification was done by the ELM Classifier. Extreme Learning Machines are less complex computation machines which can be utilized for regression, multi-classs classification and feature mapping. The operation of PSOBBO is shown by Algorithm 1. In the algorithm, P signifies population, Pnew signifies the habitat after migration process, V signifies the velocity of the particle. The values c1 = 0.5 and c2 = 2 are employed as constant weight factors.

6.2. Stationary Wavelet Transform

There are abundant signal transform methods that can be used to convert the signal into fundamental sinusoids of altered frequencies. Wavelet transform helps in conserving the time and frequency by disintegrating the signal in a order of increasing resolve. Discrete wavelet transform (DWT) is implemented either through filter bank approach or lifting scheme. Filter bank approach is a series of filtration in which the signal is sequentially passed first through low l[m], then high h[m]. It is then reduced by a factor of 2 for the computations of the coefficients. For preprocessing an image, DWT is applied to every dimension separately. In terms of shift invariance and decimation, Stationary Wavelet Transform (SWT) is better than DWT for pattern detection, feature extraction and change detection. Conventionally, in DWT every single level of transform input signal is convoluted with high and low pass filters. Afterwards, they are reduced by a factor of two for the procurement of wavelet coefficients. Differently in SWT, after the convolution with low l[m] and high h[m] no reduction or decimation is done.

The first step was performed using the Viola Jones algorithm where an integral Iint represents an input image-

$$I_i(x,y) = \sum_{x_i,y_i}^{x,y} I(x_{i,}, y_i) \tag{2}$$

The calculations and computations were carried out on the image using the following equations:

$$(x, y) = (x, y - 1) + I(x, y) \tag{3}$$

$$I_i(x, y) = I_i(x - 1, y) + s(x, y) \tag{4}$$

where(x,y) represents the row sum, s(x,–1),and $I_i(-1, y)$equals zero.

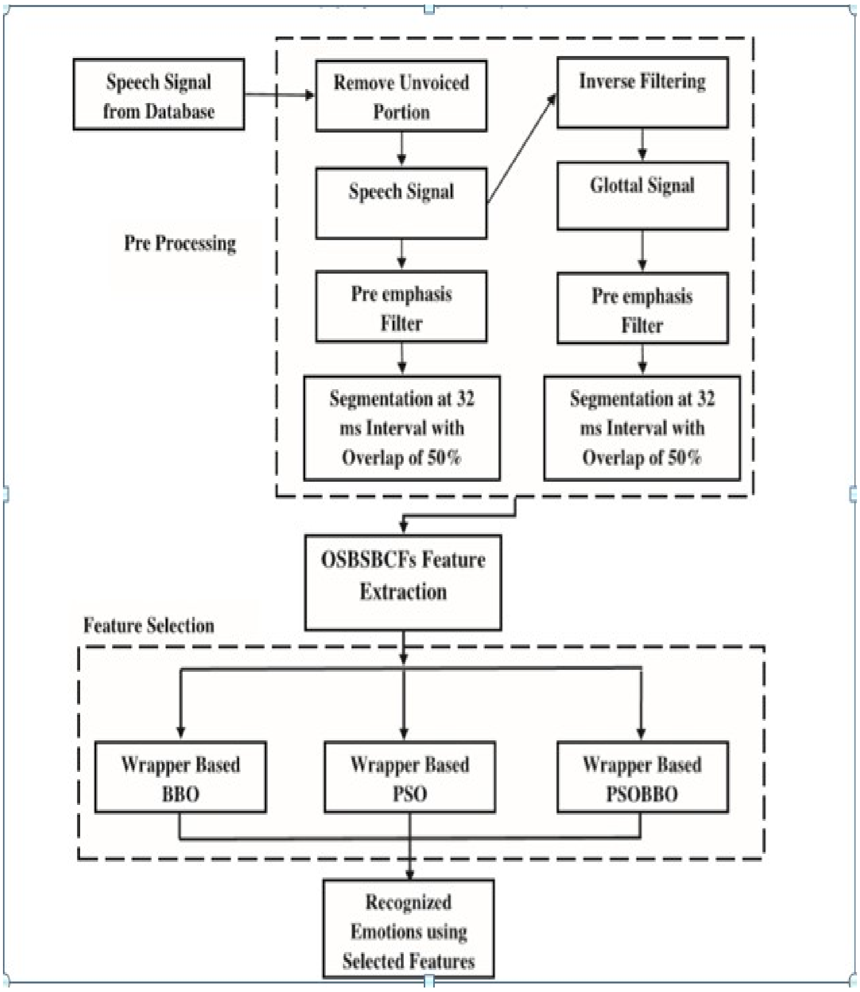

Figure 2. Hybrid PSO assisted Biogeography-based Optimization[55]

Algorithm 1: The PSOBBO framework

```
Randomly initialize the population of P habitats
Calculate the tness for each habitat
Sort the habitats in descending order based on the tness
Update gBest
for (m = 1 to Maximum Iteration)
—— for (i = 1 to P)
—— Update a and b
—— end
Perform Migration operations
—— for (p = 1 to P)
——— for j = 1 to Number Of features
—————-if rand () ¡ i
—————-Select a habitat Pp with probability i
—————-Pnewp Pp
—————-end if
——— end for
——end for
——for p = round (length (P/2)) to P
———for j = 1 to Number Of features
———V(p, j) = rand V(p, j) + c1 rand P(p, j) + c2 rand gBest( j) P(p, j)
———S = abs((2/pi) a tan((pi/2) V(p, j)))
—————-if rand () ¡ S
—————-Pnew(p, j) = 1
—————-else
—————-Pnew(p, j) = 0
—————-end if
———end for
——end for
P = Pnew
Calculate the tness for each habitat
Sort the habitats in descending order based on the tness
Update gBest
end for
```

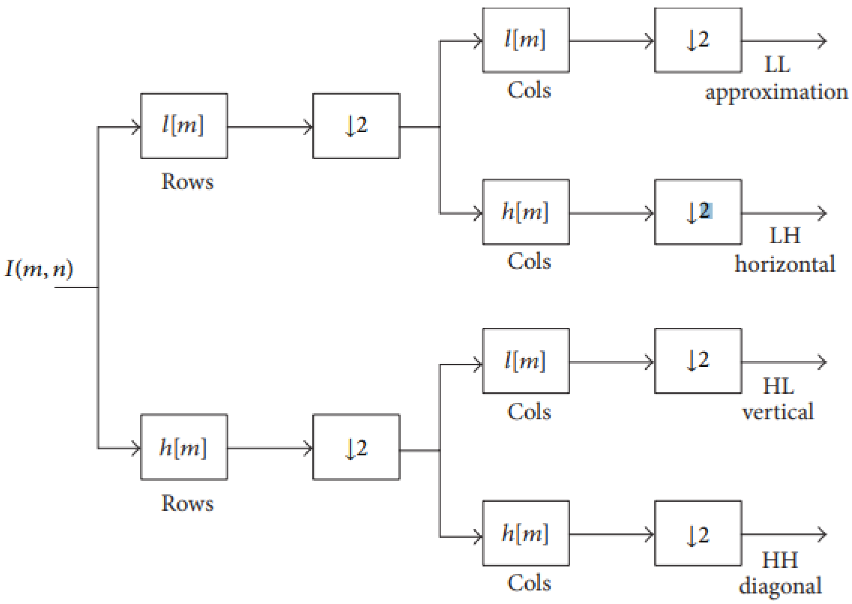

Figure 3. Discrete Wavelet Transform [39]

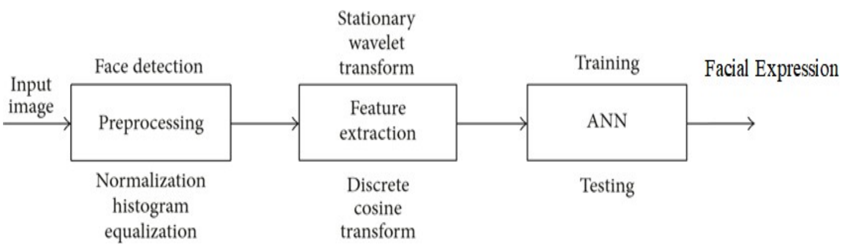

Figure 4. Face detection Process [39]

In the next step, AdaBoost algorithm was employed for selecting features. A waterfall combination of classifiers was used for fast elimination of background regions, thus, allocating more time for the evaluation of face like regions. These classifiers used features computed on the area of rectangular neighborhood of pixels. Next, the processes of image normalization along with histogram equalization were carried out for the removal of unassociated and undesirable parts. The normalized image norm was given by,

$$\mathrm{I_{norm}(x,y)} = \mathrm{I}_d\mathrm{(x,y)min(I}_d\mathrm{(x,y))/max(I}_d\mathrm{(x,y))min(I}_d\mathrm{(x,y))} \quad (5)$$

where $I_d(x,y)$ is the sub image identified as face region and min() and max() are functions used to identify the minimum and maximum pixel values. Normalization was accomplished to alter the intensity of the images into a new range [0-1]. The detected and preprocessed face from the image was firstly disintegrated into various sub bands with the help of Stationary Wavelet Transform. In SWT, input image was convoluted with low pass and high pass filter to obtain estimated and comprehensive coefficients without decimation. For the detected face image of size M N, the SWT at th level is given as,

$$\mathrm{LL_{j+1}(a,b)} = \sum y \ \sum_x I_{\mathrm{xy}} I_{jy} LL_y(\mathrm{a} + \mathrm{x}, \mathrm{b} + \mathrm{y}) \quad (6)$$

$$\mathrm{LH_{j+1}(a,b)} = \sum y \ \sum_x h_{\mathrm{xy}} I_{jy} LL_y(\mathrm{a} + \mathrm{x}, \mathrm{b} + \mathrm{y}) \quad (7)$$

$$\mathrm{HL_{j+1}(a,b)} = \sum y \ \sum_x I_{\mathrm{xy}} h_{jy} LL_y(\mathrm{a} + \mathrm{x}, \mathrm{b} + \mathrm{y}) \quad (8)$$

$$HH_{j+1}(a,b) = \sum y \sum_x h_{xy} h_{jy} LL_y(a + x, b + y) \quad (9)$$

Where a=1,2,...,M, b=1,2,...,N and h and l represent the low and high pass filters. LL, HL, HH and LH are the approximate, vertical, diagonal and horizontal sub bands. Different sub bands have different information in each SWT. The overall image estimate was given by the LL sub band and other sub bands (LH, HH and HL) have the horizontal, diagonal and vertical information. The resulting SWT disintegration was similar to the input image, having four times the number of coefficients than the original image, as the data was two-dimensional. To reduce the vector length of features, 88 block DCT was employed to three bands (LH, HL, and HH) of SWT. The DCT was calculated as,

$$X(u,V) = C(u)C(V)/4 \sum_{m=0} \sum_{n=0} \cos((2m+1)u\pi/16).\cos((2n+1)V\pi/16) \quad (10)$$

Where,

$$(u) = \frac{1}{\sqrt{2}}, u = 0,1,1 \le u \le 7$$

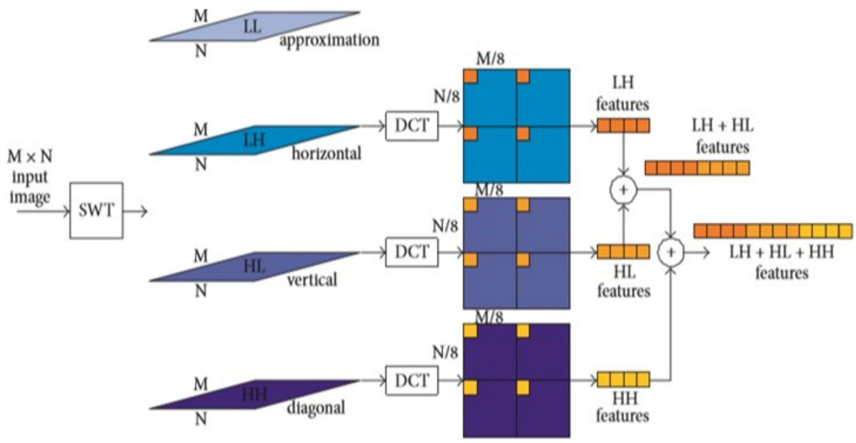

Figure 5. Feature Extraction [39]

Table 7. Databases in SWT method

Properties	No. of subjects	No. of images	Static/Video	Single/Multiple Faces	Expressions
CK+	210	N/A	N/A	Single	23
JAFFE	5749	13,233	Static	Single	Various

As DC denotes common of the energy of that sub band, it was selected from each block as features. The features were combined from different sub bands, that is, LH, HL, LH + HL, and LH + HL + HH, to achieve better features in terms of diagonal , vertical and horizontal directions. Finally, the features were fed to the Artificial Neural Network.

The Neural Network was trained to classify seven emotions. It had fully connected layers. It had k inputs (f_1–f_k), which denoted the seven outputs and feature vector length (1–7)that related to the emotions being classified. The training data was ordered in pairs (F_i,Y_i),where F is input vector and Y is target output. Feed Forward Networks were used for training and Optimization was done using back propagation. The output was taken as O rather than Y. This design was verified for the classification of facial expressions of JAFFE and CK+ database. The details of the databases are given in table 7.

6. Results

In the previous sections, we compared and discussed all the methods and models of the same field as well as intercompared the methods and models of the four different emotion recognition methods. For facial emotion recognition, we compared the model and feature based methods. The results are shown in the figure 5. It can be inferred that model using Static Wavelet Transforms has the highest efficiency of 98.83% followed by two fold random forest classifiers. Figure 6 shows the accuracy comparison of feature based techniques.

For emotion recognition through speech signals, we found that Anchor models have the least accuracy and HMM, N-D HMM has the highest accuracy of 94.41%. As observed, the accuracy for detection of emotion using speech signals has significantly grown over the years. For physiological signals, we analyzed the two domains of emotion recognition through EEG and ECG signals. These signals need extensive preprocessing and thus feature extraction is difficult. Consequently, these are the least employed methods for emotion recognition. As we can observe from figure 8 that EQ radio has the highest accuracy whereas most of the other methods have their accuracies lying between 50-60 %.

EEG signals are physiological signals recording the brain activity.Just like ECG signals, extensive preprocessing is required for EEG signals before feature extraction and selection. The graph for ER through EEG signals is given which shows that Hybrid Filtering and Higher Order Crossings is the most efficient method from the given methods, whereas Frequency domain features with SVM classifiers has the least accuracy.

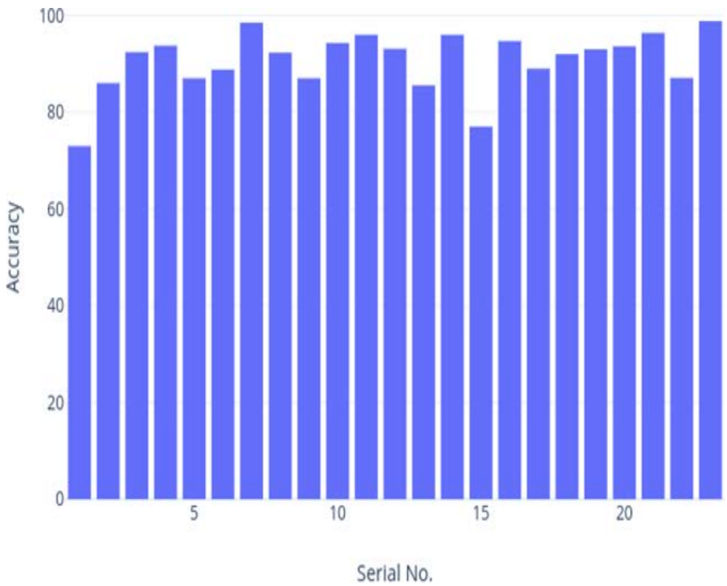

Figure 6. Accuracy rate depiction of model based approaches for FER

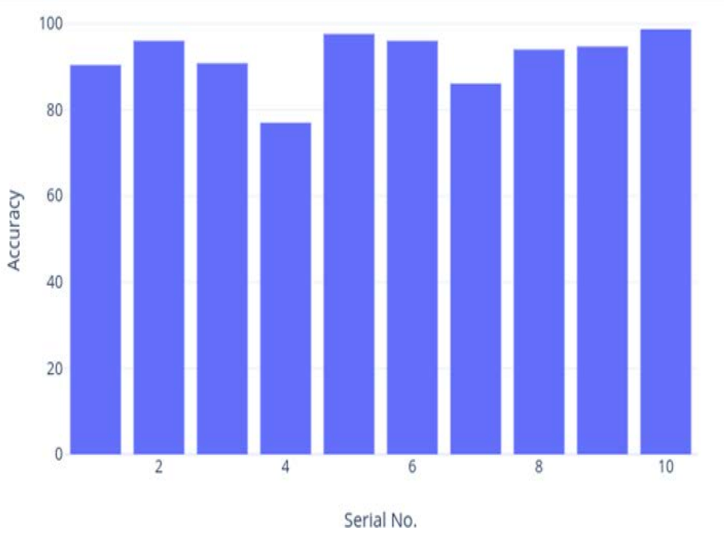

Figure 7. Accuracy rate depiction of feature based techniques used for FER

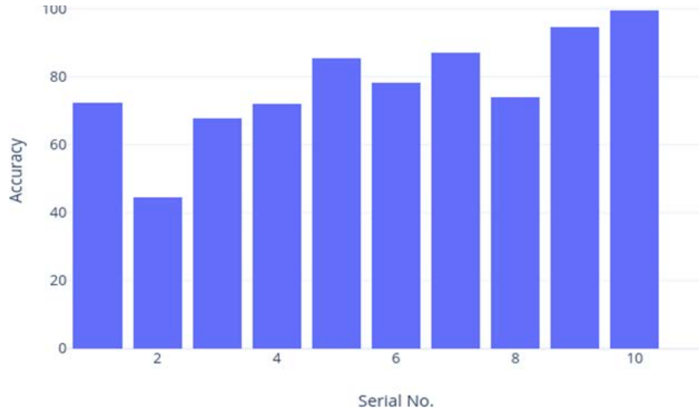

Figure 8. Recognition rate depiction of Speech based ER

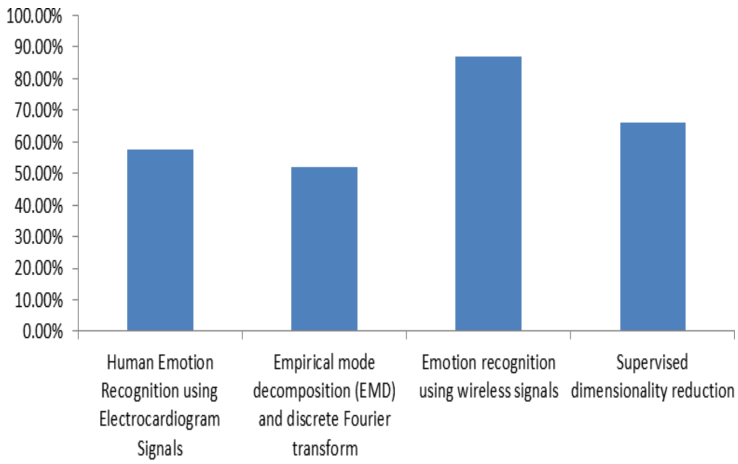

Figure 9. Recognition rate depiction of ER through ECG signals

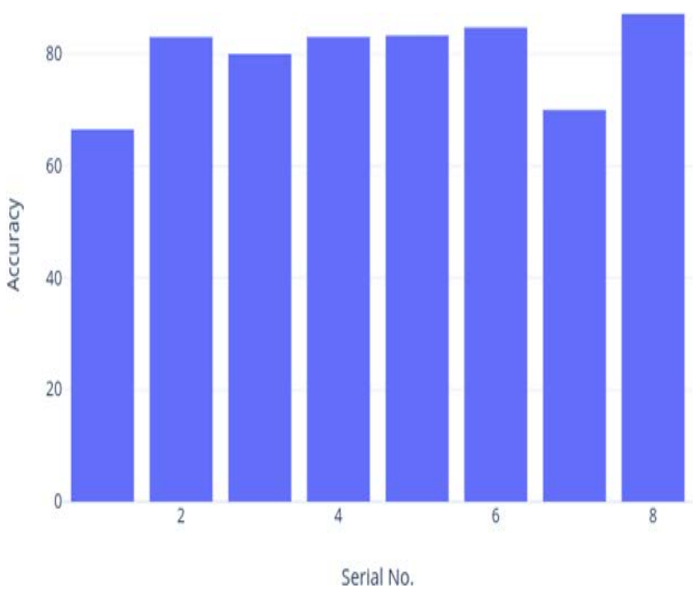

Figure 10. Recognition rate depiction through EEG signals

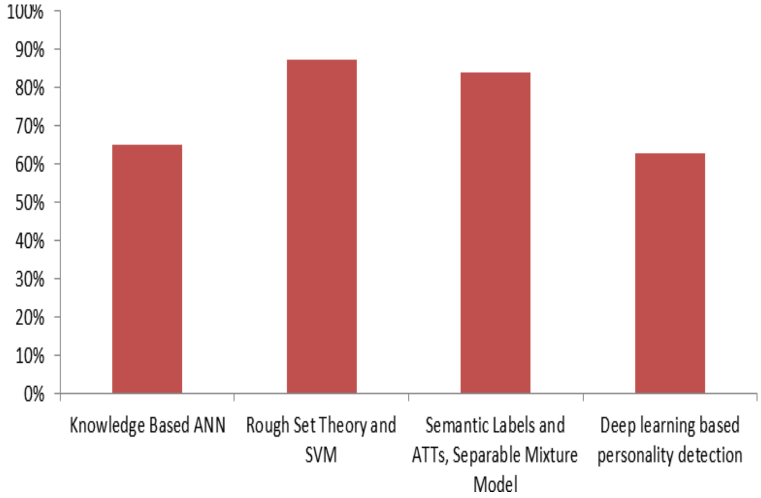

Figure 11: Recognition rate depiction of Textual ER

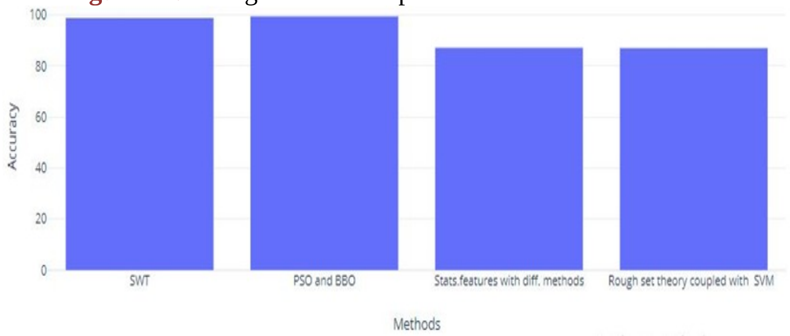

Figure 12. Comparison the best method of the four domain

7. Conclusion

Emotions play an important role in human sphere of life [79]. This paper was comprehended to assess and gather all the significant and efficient emotion recognition techniques developed in the last decade. Today, we have a wide range of methods, from being based on a single mathematical or neural model to a combination of multiple features, models and classifiers [80] [81] [82] [83]. A considerable amount of work has been done in the fields of facial emotion recognition, emotion recognition through speech signals. The six elementary emotions which humans display are sadness, surprise, disgust, happiness, fear and anger. Facial Emotion Recognition (FER) is largely done through two categories of methods, namely, feature and model based techniques. Feature extraction and selection approaches such as Gabor wavelets, Facial landmarks,), Local binary patterns (LBP),Weber Local Descriptors (WLD, Active units (AUs), Histogram of Oriented Gradients (HOG), Geodesic path difference, Local Directional Pattern (LDP) etc. are extensively used. Model based techniques including Neural Network Models, 3D Face Recognition Models, Multi View Models, Models principally based on Support Vector Machines (SVMs) Classifiers, Bayesian Belief Networks, etc. are popular for Facial Emotion Recognition. Speech emotion recognition is mainly done through four types of methods, namely prosodic features, phonetic features, mathematical models and neural models. Some of the popular methods include Convolutional and Artificial Neural Networks, Discrete Wavelet Transform (DWT) based models, Anchor Models, Vector Space Modeling, Gaussian Mixture Models and Hybrid Models.

In emotion recognition through physiological signals, signals are first decomposed to smaller activities or features. Thus, they require preprocessing and then feature extraction and classification is done. In ECG signals, decomposition is done through methods like Empirical Mode Decomposition (EMD). Further, EMD is also used in hybrid feature extraction algorithms and parameters. Similarly, EEG signals are preprocessed first and then feature extraction step takes place followed by classification. Though, there has been a considerable amount of work done in the field of emotion recognition through EEG signals, but there are not many methods for emotion detection through ECG signals. In emotion recognition through text, a varied range of techniques are used, including manual methods. Furthermore, there is a large number of public datasets available for emotion recognition and detection methods.

The best results were obtained by using Stationary Wavelet Transform for Facial Emotion Recognition , Particle Swarm Optimization assisted Biogeography based optimization algorithms for emotion recognition through speech, Statistical features coupled with different methods for physiological signals, Rough set theory coupled with SVM for text semantics with respective accuracies of 98.83 % ,99.47 %, 87.15 % ,87.02% . Overall, the method of Particle Swarm Optimization assisted Biogeography based optimization algorithms with an accuracy of 99.47% on BES dataset gives the best results.

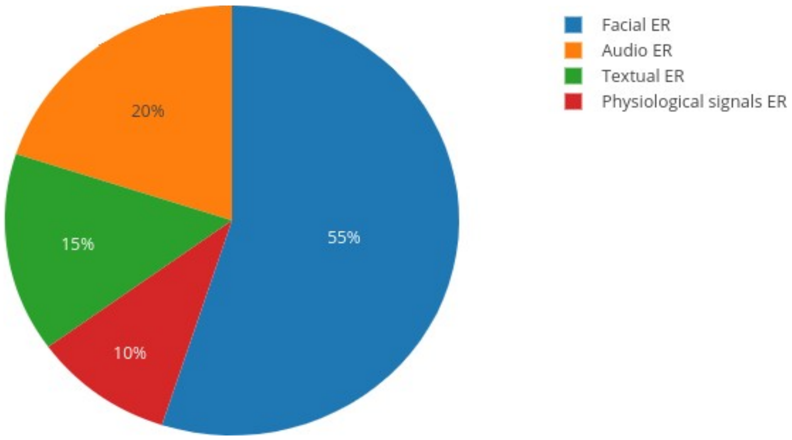

Figure 13. Domain Classification of Emotion Recognition Techniques

8. Future Scope

There are numerous limitations and a wide scope of improvements in this field. The complexity of preprocessing the physiological signals is a big challenge for emotion detection through physiological signals. This is a competent area of research. Till now, only seven basic emotions have been successfully identified. Research should be carried out for identifying more than seven emotions. Emotion detection through ECG signals and features such as skin temperatures, Electromyography signals (EMG) which use muscle movement signals [84][85] are still emerging . A detailed research can be carried out to check the potency of these methods. There is still a dearth of accurate combined and hybrid models. More effective hybrid and combined methods can be developed for better estimation of human emotions.

Acknowledgement

We are very grateful to the editors and reviewers for their suggestions on our paper.

Conflicts of Interest

There is no conflict of interest.

References

[1] McCarthy, John. What is artificial intelligence? URL: http://wwwformal. stanford.edu!jmciwhatisai.html, 1998.

[2] L. Steels,The artificial life roots of artificial intelligence , Artificial Life. 1(1994). 75-110.

[3] Michael Brady, Artificial intelligence and robotics , Artificial Intelligence. 26(1985). 79-121.https://.org/10.1016/0004-3702(85)90013-X.

[4] E. Cambria, Affective Computing and Sentiment Analysis, IEEE Intelligent Systems.31(2016)102-107.

[5] Tao J., Tan T., Affective Computing: A Review. In: Tao J., Tan T., Picard R.W. (eds) Affective Computing and Intelligent Interaction. ACII 2005, Lecture Notes in Computer Science,3784.Springer, Berlin, Heidelberg.

[6] Yashaswi Alva M, Nachamai M and J. Paulose, A comprehensive survey on features and methods for speech emotion detection, 2015 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore.(2015)1-6. 10.1109/ICECCT.2015.722604.

[7] Vandervoort, D. J.,The importance of emotional intelligence in higher education. Current Psychology: Development, Learning, Personality, Social, 25(1) (2006)4–7.

[8] Bagozzi, R. P., Gopinath, M., Nyer, P. U. The Role of Emotions in Marketing. Journal of the Academy of Marketing Science. 27(2) (1999)184–206. 10.1177/0092070399272005

[9] Scotty Craig, Arthur Graesser, Jeremiah Sullins Barry Gholson , Affect and learning: An exploratory look into the role of affect in learning with AutoTutor , Journal of Educational Media.29:3(2004) . 241-250. 10.1080/1358165042000283101

[10] Nussbaum M., Emotions as Judgments of Value and Importance. In R. C. Solomon (Ed.), Series in affective science. Thinking about feeling: Contemporary philosophers on emotions. (2004). 183-199. New York, NY, US: Oxford University Press.

[11] M. S. Bartlett, G. Littlewort, I. Fasel and J. R. Movellan, Real Time Face Detection and Facial Expres- sion Recognition: Development and A lications to Human Computer Interaction., 2003 Conference on Computer Vision and Pattern Recognition Workshop, Madison, Wisconsin, USA(2003). 53-53. 10.1007/BF02884429.

[12] Moataz El Ayadi, Mohamed S. Kamel, Fakhri Karray, Survey On Speech Emotion Recognition: Features, Classification Schemes, And Databases , Pattern Recognition, 44(2011). 572-587. 10.1016/J.Patcog.2010.09.020.

[13] Nicu Sebe, Ira Cohen, Theo Gevers, and Thomas S. Huang Multimodal approaches for emotion recog- nition: a survey , Internet Imaging VI.(2005) .https:// .org/10.1117/12.600746.

[14] Anagnostopoulos, CN., Iliou, T. Giannoukos, I., Artifical Intelligence Review . ACM. 2()155-177. 10.1007/s10462-012-9368-5.

[15] Z. Zeng, M. Pantic, G. I. Roisman and T. S. Huang, A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions, in IEEE Transactions on Pattern Analysis and Machine Intelligence. 31(2009). 39-58 . 10.1109/TPAMI.2008.52

[16] H. Gunes, B. Schuller, M. Pantic and R. Cowie, Emotion representation, analysis and synthe- sis in continuous space: A survey, Face and Gesture 2011, Santa Barbara, CA(2011). 827-834. 10.1109/FG.2011.5771357

[17] B. Fasel, Juergen Luettin, Automatic facial expression analysis: a survey , Pattern Recognition,36(2003). 259-275.10.1016/S003.

[18] S. Mitra and T. Acharya, Gesture Recognition: A Survey, in IEEE Transactions on Systems, Man, and Cybernetics, Part C (A lications and Reviews). 37(2007). 311-324. 10.1109/TSMCC.2007.893280.

[19] Chan HL, Kuo PC, Cheng CY, Chen YS. Challenges and Future Perspectives on Electroencephalogram- Based Biometrics in Person Recognition Front Neuroinform .(2018).12:66. 10.3389/fninf.2018.00066.

[20] M. Gargesha, P. Kuchi, Facial expression recognition using a neural network, Artificial Neural Compu- tation Systems, 31 (2002). 709-724.

[21] S. C. Tai and K. C. Chung, Automatic facial expression recognition system using Neural Networks, TENCON 2007 - 2007 IEEE Region 10 Conference, Taipei.(2007). 1-4. 10.1109/TENCON.2007.4429124.

[22] F. Chen, Z. Wang, Z. Xu, J. Xiao and G. Wang, Facial Expression Recognition Using Wavelet Trans- form and Neural Network Ensemble, 2008 Second International Symposium on Intelligent Information Technology A lication, Shanghai.(2008) 871-875.10.1109/IITA.2008.

[23] Neha Jain, Shishir Kumar, Amit Kumar, Pourya Shamsolmoali, Masoumeh Zareapoor, Hybrid deep neural networks for face emotion recognition, Pattern Recognition Letters. 115(2018). 101-106. https://.org/10.1016/j.patrec.2018.04.010.

[24] L. Ma and K. Khorasani, Facial expression recognition using constructive feedforward neural networks, in IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 34(2004).1588-1595.10.1109/TSMCB.2004.825930.

[25] P. Liu, S. Han, Z. Meng and Y. Tong, Facial Expression Recognition via a Boosted Deep Belief Net- work, 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus. 2014.1805-1812 . 10.1109/CVPR.2014.233.

[26] I. Mpiperis, S. Malassiotis, V. Petridis and M. G. Strintzis, 3D facial expression recognition using swarm intelligence, 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV. (2008). 2133-2136 . 10.1109/ICASSP.2008.4518064.

[27] C. Mayer, M. Wimmer, M. Eggers and B. Radig, Facial Expression Recognition with 3D Deformable Models, 2009 Second International Conferences on Advances in Computer-Human Interactions, Cancun. (2009). 26-31 . 10.1109/ACHI.2009.33.

[28] Y. V. Venkatesh, A. K. Kassim and O. V. R. Murthy, Resampling A roach to Facial Expression Recogni- tion Using 3D Meshes, 2010 20th International Conference on Pattern Recognition, Istanbul.(2010)3772- 3775. 10.1109/ICPR.2010.91.

[29] Y. Tie and L. Guan, A Deformable 3-D Facial Expression Model for Dynamic Human Emotional State Recognition, in IEEE Transactions on Circuits and Systems for Video Technology, 23(2013). 142-157. 10.1109/TCSVT.2012.2203210

[30] H. Chen, C. Huang and C. Fu, Hybrid-Boost Learning for Multi-Pose Face Detection and Facial Expres- sion Recognition, 2007 IEEE International Conference on Multimedia and Expo, Beijing.(2007). 671-674.10.1109/ICME.2007.4284739

[31] Yubo Wang, Haizhou Ai, Bo Wu and Chang Huang, Real time facial expression recognition with Ad- aBoost, Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004., Cambridge. 3(2004). 926-929. 10.1109/ICPR.2004.1334680.

[32] Y. Cheon and D. Kim, A Natural Facial Expression Recognition Using Differential-AAM and k-NNS, 2008 Tenth IEEE International Symposium on Multimedia, Berkeley, CA. (2008).220- 227.10.1109/ISM.2008.121

[33] L. Zhang and D. Tjondronegoro, Facial Expression Recognition Using Facial Movement Features, in IEEE Transactions on Affective Computing. 2(2011) .219-229. 10.1109/T-AFFC.2011.13

[34] Kazmi, S.B., Qurat-ul-Ain Arfan Jaffar, M. Soft Comput .(2012) 16: 369. https:// .org/10.1007/s00500- 011-0721-4

[35] Michel, Phili El Kaliouby, Rana. Real time facial expression recognition in video using su ort vector machines. ICMI ’03 Proceedings of the 5th international conference on Multimodal interfaces . Vancouver, British Columbia, Canada. (2003). 258-264 .10.1145/958432.958479.

[36] Zhang, Xiao Mahoor, Mohammad Mavadati, Seyedmohammad. Facial expression recognition using lp-norm MKL multiclass-SVM. Machine Vision and A lications. 26 (2015). 467-483. 10.1007/s00138-015- 0677-y.

[37] D. Datcu and L. J. M. Rothkrantz, Automatic recognition of facial expressions using Bayesian belief networks, 2004 IEEE International Conference on Systems, Man and Cybernetics . The Hague 3(2004). 2209-2214. : 10.1109/ICSMC.2004.1400656

[38] Q. Zhen, D. Huang, Y. Wang and L. Chen, Muscular Movement Model-Based Automatic 3D/4D Facial Expression Recognition, in IEEE Transactions on Multimedia, 18(2016) .1438-1450. 10.1109/TMM.2016.2557063

[39] Qayyum, Huma Majid, Muhammad Anwar, Syed Khan, Bilal, Facial Expression Recognition Using Stationary Wavelet Transform Features . Mathematical Problems in Engineering. (2017).1-9. : 10.1155/2017/9854050.

[40] Pu, Xiaorong Fan, Ke Chen, Xiong Ji, Luping Zhou, Zhihu. Facial expression recognition from image sequences using twofold random forest classifier. Neurocomputing. 168(2015).1173-1180.10.1016/j.neucom.2015.05.005.

[41] V. Gomathi, K. Ramar, A. S. Jeevakumar, Human Facial Expression Recognition using MANFIS Model , International Journal of Computer Science and Engineering, 3(2009) .93-97.

[42] Zhan Yong-zhao, Ye Jing-fu, Niu De-jiao and Cao Peng, Facial expression recognition based on Gabor wavelet transformation and elastic templates matching, Third International Conference on Image and Graphics (ICIG’04), Hong Kong, China.(2004) .254-257. 10.1109/ICIG.2004.63

[43] Yurtkan, Kamil Demirel, Hasan, Feature selection for improved 3D facial expression recognition . Pattern Recognition Letters.38. (2013).26-33. 10.1016/j.patrec.2013.10.026.

[44] Xiaoyi Feng, Facial expression recognition based on local binary patterns and coarse-to-fine classification, The Fourth International Conference onComputer and Information Technology, 2004. CIT ’04., Wuhan, China(2004) . 178-183. 10.1109/CIT.2004.1357193.

[45] D. Huang, M. Ardabilian, Y. Wang and L. Chen, 3-D Face Recognition Using eLBP-Based Facial Description and Local Feature Hybrid Matching, in IEEE Transactions on Information Forensics and Security . 7(2012). 1551-1565.

10.1109/TIFS.2012.2206807

[46] Yu, Kaimin Wang, Zhiyong Hagenbuchner, Markus Feng, David Dagan Feng, Spectral embed- ding based facial expression recognition with multiple features, Neurocomputing. 129(2014). 136–145. 10.1016/j.neucom.2013.09.046.

[47] A. Ramirez Rivera, J. A. Rojas Castillo and O. Chae, Recognition of face expressions using Local Principal Texture Pattern, 2012 19th IEEE International Conference on Image Processing, Orlando, FL. (2012). 2609-2612. 10.1109/ICIP.2012.6467433.

[48] K. Mistry, L. Zhang, S. C. Neoh, C. P. Lim and B. Fielding, A Micro-GA Embedded PSO Feature Selec- tion A roach to Intelligent Facial Emotion Recognition, in IEEE Transactions on Cybernetics .47(2017). 1496-1509.

[49] Q. Mao, M. Dong, Z. Huang and Y. Zhan, Learning Salient Features for Speech Emotion Recogni- tion Using Con utional Neural Networks, in IEEE Transactions on Multimedia. 16(2014). 2203-2213. 10.1109/TMM.2014.2360798

[50] Huang, Zhengwei Dong, Ming Mao, Qirong Zhan, Yongzhao, Speech Emotion Recognition Us- ing CNN. MM 2014 - Proceedings of the 2014 ACM Conference on Multimedia . (2014). 801-804 . 10.1145/2647868.2654984.

[51] E. H. Kim, K. H. Hyun, S. H. Kim and Y. K. Kwak, Speech Emotion Recognition Using Eigen-FFT in Clean and Noisy Environments, RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication, Jeju. (2007) . 689-694. 10.1109/ROMAN.2007.4415174.

[52] Y. Attabi and P. Dumouchel, Anchor Models for Emotion Recognition from Speech, in IEEE Transactions on Affective Computing. 4(2013). 280-290 . 10.1109/T-AFFC.2013.17.

[53] J. Liu, C. Chen, J. Bu, M. You and J. Tao, Speech Emotion Recognition using an Enhanced Co-Training Algorithm, 2007 IEEE International Conference on Multimedia and Expo, Beijing, (2007) .999-1002 . 10.1109/ICME.2007.4284821.

[54] Yashaswi Alva M, Nachamai M and J. Paulose, A comprehensive survey on features and methods for speech emotion detection, 2015 IEEE International Conference on Electrical, Computer and Communi- cation Technologies (ICECCT), Coimbatore.(2015).1-6. 10.1109/ICECCT.2015.7226047.

[55] Yogesh C.K., M. Hariharan, Ruzelita Ngadiran, Abdul Hamid Adom, Sazali Yaacob, Chawki Berkai, Kemal Polat, A new hybrid PSO assisted biogeography-based optimization for emotion and stress recog- nition from speech signal, Expert Systems with A lications, 69(2017). 149-158.

[56] Paithane, A. N.. Human Emotion Recognition using Electrocardiogram Signals. , International Confer- ence on Pervasive Computing (ICPC). (2014). 10.1109/PERVASIVE.2015.7087042.

[57] M, Muruga an Khairunizam, Wan Yaacob, Sazali Selvaraj, Jerritta, Electrocardiogram-based emotion recognition system using empirical mode decomposition and discrete Fourier transform. Expert Systems. 31 (2013). 10.1111/exsy.12014.

[58] By Mingmin Zhao, Fadel Adib, Dina Katabi , Communications of the ACM,61 No. 9(2018) . 91-100. :10.1145/3236621.

[59] By Mingmin Zhao, Fadel Adib, Dina Katabi , Communications of the ACM,61 No. 9(2018) . 91-100. :10.1145/3236621 59. Ferdinando, Hany Se ¨anen, Tapio Alasaarela, Esko, 2017, Enhanc- ing Emotion Recognition from ECG Signals using Supervised Dimensionality Reduction. 112-118 . 10.5220/0006147801120118.

[60] Ferdinando, Hany Se ¨anen, Tapio Alasaarela, Esko, 2017, Enhanc- ing Emotion Recognition from ECG Signals using Supervised Dimensionality Reduction. 112-118 . 10.5220/0006147801120118.

[61] Duan RN., Wang XW., Lu BL, EEG-Based Emotion Recognition in Listening Music by Using Su ort Vector Machine and Linear Dynamic System. In: Huang T., Zeng Z., Li C., Leung C.S. (eds) Neural Information Processing,Lecture Notes in Computer Science. (2012). 7666. Springer, Berlin, Heidelberg.

[62] Y. Liu, C. Wu, Y. Kao and Y. Chen, Single-trial EEG-based emotion recognition using kernel Eigen-emotion pattern and adaptive su ort vector machine, 2013 35th Annual International Confer- ence of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka. (2013). 4306-4309. 10.1109/EMBC.2013.6610498.

[63] Suwicha Jirayucharoensak, Setha Pan-Ngum, and Pasin Israsena 2014. EEG-Based Emotion Recognition Using Deep Learning Network with Principal Component Based Covariate Shift Adaptation, The Scientific World Journal,627892. https:// .org/10.1155/2014/627892.

[64] Mehmood, Raja Majid Lee, Hyo Jong., EEG based Emotion Recognition from Human Brain using Hjorth Parameters and SVM. International Journal of Bio-Science and Bio-Technology, 7(2015). 23-32. 10.14257/ijbsbt.2015.7.3.03.

[65] Tauseef Sohaib, Ahmad Qureshi, Shahnawaz Hagelb¨ack, Johan Hilborn, Olle Jerˇci´c, Petar. Evaluating Classifiers for Emotion Recognition Using EEG . Found. Augment. Cognit. Lecture Notes Comput. Sci.. 8027(2013). 492-501. 10.1007/978-3-642-39454-6 53.

[66] Gao, Yongbin et al. Deep learninig of EEG signals for emotion recognition. 2015 IEEE International Conference on Multimedia Expo Workshops (ICMEW) . (2015).1-5.

[67] J. J. Bird, L. J. Manso, E. P. Ribiero, A. Ekart, and D. R. Faria, A study on mental state classifi- cation using eeg-based brain-machine interface, in 9th International Conference on Intelligent Systems, IEEE(2018).

[68] Y.-S. Seol, D.-J. Kim, and H.-W. Kim. — Emotion recognition from text using knowledge- based ANN, in proceedings of IC-ISCC. (2008) 1569-1572.

[69] Z. Teng, F. Ren and S. Kuroiwa, Emotion Recognition from Text based on the Rough Set Theory and the Su ort Vector Machines, 2007 International Conference on Natural Language Processing and Knowledge Engineering, Beijing. (2007). 36-41. 10.1109/NLPKE.2007.4368008.

[70] Wu, Chung-Hsien et al. Emotion recognition from text using semantic labels and separable mixture models. ACM Trans. Asian Lang. Inf. Process. 5 (2006).165-183. 10.1145/1165255.1165259.

[71] N. Majumder, S. Poria, A. Gelbukh and E. Cambria, Deep Learning-Based Document Modeling for Personality Detection from Text, in IEEE Intelligent Systems, 32(2017) .74-79. 10.1109/MIS.2017.23

[72] M. Pantic and I. Patras, Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences, in IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 36(2006).433-449 . 10.1109/TSMCB.2005.859075.

[73] I. Mpiperis, S. Malassiotis, V. Petridis and M. G. Strintzis, 3D facial expression recognition using swarm intelligence, 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV(2008). .2133-2136. 10.1109/ICASSP.2008.4518064.

[74] O. Rudovic, M. Pantic and I. Patras, Coupled Gaussian processes for pose-invariant facial expression recognition, in IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(2013) .1357-1369. 10.1109/TPAMI.2012.233.

[75] W. Zheng, Multi-View Facial Expression Recognition Based on Group Sparse Reduced-Rank Regression, in IEEE Transactions on Affective Computing, 5(2014) . 71-85. 10.1109/TAFFC.2014.2304712.

[76] W. Liu, S. Li and Y. Wang, Automatic Facial Expression Recognition Based on Local Binary Patterns of Local Areas, 2009 WASE International Conference on Information Engineering, Taiyuan, Chanx. (2009). 197-200. 10.1109/ICIE.2009.36.

[77] Li, Huibin Ding, Huaxiong Huang, di Wang, Yunhong Zhao, xi Morvan, J. - M. Chen, Liming, 2015, An Efficient Multimodal 2D + 3D Feature-based A roach to Automatic Facial Expression Recognition. Computer Vision and Image Understanding. 140. 10.1016/j.cviu.2015.07.005.

[78] Wang, Xun et al. A New Facial Expression Recognition Method Based on Geometric Alignment and LBP Features. 2014 IEEE 17th International Conference on Computational Science and Engineering. (2014). 1734-1737.

[79] R. Srivastava and S. Roy, 3D facial expression recognition using residues, TENCON 2009,2009 IEEE Region 10 Conference, Singapore. (2009).1-5. 10.1109/TENCON.2009.5395856.

[80] Antoine Bechara, The role of emotion in decision-making: Evidence from neurological patients with orbitofrontal damage ,Brain and Cognition, 55(2004). 30-40. https:// .org/10.1016/j.bandc.2003.04.001.

[81] Lakshmanaprabu SK, Shankar K, Deepak Gupta, Ashish Khanna, Joel J. P. C. Rodrigues, Plácido R. Pinheiro, Victor Hugo C. de Albuquerque, “Ranking Analysis for Online Customer Reviews of Products Using Opinion Mining with Clustering”, Complexity, June 2018, Article ID 3569351. https://doi.org/10.1155/2018/356935.

[82] Aditya Khamparia, Deepak Gupta, Nguyen Gia Nhu, Ashish Khanna, Babita Shukla, Prayag Tiwari, “Sound Classification Using Convolutional Neural Network and Tensor Deep Stacking Network”, IEEE Access, doi: 10.1109/ACCESS.2018.2888882.

[83] Anvita Saxena, Kaustubh Tripathi, Ashish Khanna, Deepak Gupta, Shirish Sundaram, “Emotion Detection through EEG Signals using FFT and Machine learning techniques”. International Conference on Innovative Computing and Communications (ICICC2019). Advances in Intelligent Systems and Computing, Springer. [Accepted]

[84] Deepak Gupta, Nimish Verma, Mayank Sehgal, Nitesh, “Feature selection using multiobjective grey wolf optimization algorithm”, 2019, *Innovative Computing and Communication*. Vol.1 (July 2019) 15–18.

[85] Haag A., Goronzy S., Schaich P., Williams J. (2004) Emotion Recognition Using Bio-sensors: First Steps towards an Automatic System. In: Andr´e E., Dybkjær L., Minker W., Heisterkamp P. (eds) Affective Dialogue Systems. Lecture Notes in Computer Science. ADS 2004. 3068. Springer, Berlin, Heidelberg, : https:// .org/10.1007/978-3-540-24842-2 4.

[86] K. Takahashi, Remarks on emotion recognition from bio-potential signals, Proc. 2nd Int. Conf. Auton. Robots Agents, 1315 (2004). 186-191. 10.1.1.125.2544.