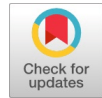


# Multimodal Analysis of Human Fear

Swagata Sarkar



**Abstract:** Human emotion detection is very much relevant in today's scenario. Human life become fast due to modernisation. People like to lead sophisticated, peaceful and healthy life. Human emotion plays a vital role in present scenario. Six basic emotions are considered for research purpose. Those are happy, sad, fear, anger, disgust and boredom. In this paper, human fear is analysed based on Electroencephalogram (EEG) signal, physical parameters and facial images. Statistical parameters both from time and frequency domain are used as feature set. Own database is used for the analysis. It is seen from the result that the efficiency is enhanced significantly after multimodal analysis of human fear. The classification results for discrete wavelet transform and logistic regression model are improved by 8.33% and 8.33% respectively.

**Keywords:** EEG signal, Emotion, Facial Images, Physical Parameters

## I. INTRODUCTION

Multimodal analysis of human emotion is an emerging field for human emotion classification. Different parameters from different domain can be used in combination for getting good results. There is a good scope to make further use of modalities. In this part of the work, EEG signal parameters are fused with physical parameters and facial image parameters using discrete wavelet transform and logistic regression model respectively. The features are classified using back propagation neural network of artificial neural network. In this part of the work, six emotions (happy, sad, surprise, angry, fear, boredom) is classified using Back Propagation Neural Network. The database used for the research purpose were collected from Sri Jeyam Clinical Laboratory, Tambaram. 50 subjects in all have contributed their valuable data for the research purpose. Thirteen features were collected from each subject. Facial images were collected as the fourteenth feature and EEG signal is also measured. Each subject was connected to various instruments measuring different parameters. Different emotions were brought about by viewing various clippings on the TV. In this part of the work emotion is analysed on the basis of facial image with physical parameters and EEG signals. Ten statistical features of facial image have been fused with thirteen physical parameters and seven statistical features of EEG signals. The thirteen physical parameters are clarified in Table 1.

**Table-I: Physical Parameters change with respect to Fear**

Serial No.	Parameters	Ranges of Parameters for Fear
1.	Electroencephalogram(A1)	Alpha(13-15 Hz) Beta(7.5-13 Hz)
2.	Heart Rate(A2)	High(84-120 BPM)
3.	Heart Rate Variability (A3)	High(0.15-0.4 Hz)
4.	Pre Ejection Period(A4)	Low(0-800 ms)
5.	Stroke Volume(A5)	High(240-400 ml)
6.	Systolic Pressure(A6)	High(120-147 Hg)
7.	Diastolic Pressure(A7)	Low(77-88 Hg)
8.	Skin Conductance Response(A8)	High(0.85-1.5ms)
9.	Tidal Volume(A9)	Rapid(100-150ml)
10.	Oscillatory Response(A10)	High(0.5-1 breaths/minute)
11.	Respiration Rate(A11)	High(15-24 breaths/minute)
12.	Non specific Skin Conductance Response(A12)	Low(0-2 /minute)
13.	Skin Conductance Level(A13)	Low(0-2 ms)

Manuscript published on 30 September 2019.

\*Correspondence Author(s)

Swagata Sarkar Assistant Professor, Sri Sairam Engineering College, Chennai, India

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>



## A. Statistical Features for Signal and Image

The statistical features are analysed for EEG signal and facial images. Both time domain and frequency domain signal analyses are done for the both EEG signal and facial images. Total seven major features are considered in the analysis. The features are Mean, Standard deviation, Skewness, Entropy, Nyquist, Second derivative sum of fast Fourier Transform. The analysis of seven basic features with respect to signal and images are analysed in the following section. There are several statistical parameters both in the time and the frequency domains, but a major seven parameters only have shown significant differences for both image and signal. The value of mode, fast fourier transform, a first order derivative of fast fourier transform and higher order derivatives also take into consideration, but those values do not have significant variation with respect to image and signal.

1. **s Mean:** The mean value of the input signal is computed over a running average window of one cycle of the specified fundamental frequency. The mean for the image is the mean of all pixels without regard to color channel. The mean value identification is very important for any domain as feature because the mean value changes remarkably for different emotions. Mathematics for the image it can be represented as follows

$$\text{Mean}(X) = \frac{\sum X}{N} \quad (1)$$

Where X=Mean,  $\sum X$ = summation of image pixels and N= Number of pixels

2. **Standard Deviation:** It represents noise and other focus of interference. This gives rise to the term Signal to Noise Ratio (SNR) which is equal to the mean divided by standard deviation. It is also known as the coefficient of variation. It is a measurement of the variability of a signal about its mean value. From the image processing point of view, the unbiased estimate of the standard deviation of the brightness within a region with pixels is called the sample standard deviation. It is the estimate of the underlying brightness probability distribution. Calculation of standard deviation is important for both signal and image points of view considering the ability to its estimate SNR for the signal to identify brightness for image. For emotional facial expression brightness of the facial image changes from one emotion to another emotion. The SNR also has remarkable difference in EEG and speech signals. The example equation for standard deviation of the image is

$$(\mu) = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - x)(x_i - x)} \quad (2)$$

Where  $\mu$ =Standard Deviation, N=total number of pixels,  $x_i$ = each value of pixel,  $x$ = arithmetic means of pixels

3. **Median:** To calculate the median value of both signal and image, the input values are first sorted for calculation of the median values of both signal and image. If the number of values is odd, the median is the middle value. If the number of values is even, the median is the average of the two middle values. It is also very much significant for emotional face and speech analysis, considering the change in the median value for every emotion after sorting the data.

$$\text{Median (Mdn)} = 1 + \left( \frac{N - \sum f_0}{f_w} \right) * i \quad (3)$$

Where Mdn= Median, N=Total number of pixels,  $\sum f_0$ = summation of frequencies of pixels,  $f_w$ = frequency of pixels within the interval containing the median.

4. **Skewness:** Skewness is a measure of the asymmetrical spread of a signal or image pixel about its mean value. It is the ratio of the average cubed deviation from the mean divided by the cube of the standard deviation. If the skewness is negative, the data are spread out more to the left of the mean to the right and vice versa. Skewness is a very important feature for both signal image as it can identify the tendency of the signal or image towards negative or positive.

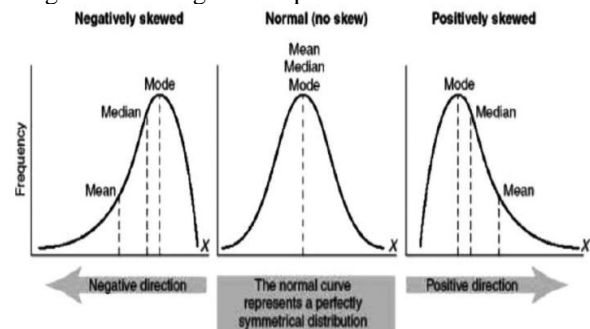


Fig. 1. Skewness Plot

The general equation for skewness is

$$\text{Skewness} = 3(\text{Mean}-\text{Median})/\text{Standard Deviation} \quad (4)$$

5. **Entropy:** The entropy or average information of the image can be obtained from the histogram of an image. It is the amount of information which must be coded by a compression algorithm. The entropy of a signal is the probability of occurrence of the signal. When the probability of occurrence is high, the information is low, otherwise, the information is high. This probability is a very important feature for both emotional facial image and signal, the value will give remarkable difference for positive and negative emotions. The equation of entropy is
 
$$\text{Entropy}(H) = -\sum_i P_i \log_2 P_i \quad (5)$$
 Where H= Entropy,  $P_i$ = Probability of instantaneous pixels
6. **Nyquist:** In this research, digital image and signal processing are used. Hence Nyquist is a very important feature. It is first the sampling of Fast Fourier transform. The Nyquist theorem states that if the sampling rate is greater than or equal to the twice of the highest frequency component present in the signal, true reproduction of the signal or image is possible. Fast Fourier transform is applicable to image from the point of view of pixel frequency. The fourier transform is a frequency domain analysis. It is known that frequency domain analysis can give better understanding and result than time domain analysis. For Fast Fourier Transform, computation time is less which can reduce the memory usage as well. The equation for Nyquist is
 
$$\text{Nyquist} = \text{Sampling of FFT} \quad (6)$$



7. **Fast Fourier Transform (FFT):** Fourier analysis can convert the time domain features to frequency domain feature. Frequency domain features can carry additional information more than a time domain feature. The FFT is the method to make Fourier transform much faster and reliable. It can improve the speed and memory utilization. The fourier transform is also a very important image processing tool. It decomposes an image to its sine and cosine components. The output of the FFT based image is in the frequency domain, whereas the input is in the time domain. The frequency domain feature has significant change for both emotional face and signal. The equation of FFT is

$$\text{Fast Fourier Transform (FFT)} X(k) = \sum_0^{N-1} x(n)e^{-j2\pi nk} \quad (7)$$

8. **Second Derivative sum of FFT:** Second derivative sum of the FFT is the added feature over normal FFT. If the result of FFT gives minor difference between two different emotional images or signals, the second derivative sum can make a significant difference in feature. The equation for the second derivative sum of FFT is

$$\text{Second Derivative sum of FFT} = \sum \frac{d^2}{dx^2}(X(K)) \quad (8)$$

From all the seven parameters only Standard Deviation, Nyquist and Second Derivative Sum of FFT are selected as primary features as only these three are dominant different.

**B. MIN-MAX Algorithm for Image Fusion**

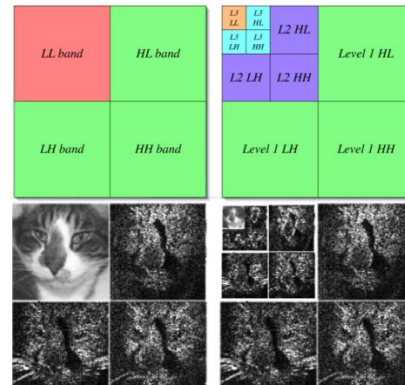
A min-max algorithm is a recursive algorithm for the choice of the next move in a n-player game, usually a two-player game. The player then makes the move that maximizes the minimum value of the position resulting from the opponent's possible following moves. In min-max algorithm depth first search technique is used. The min-max algorithm is newly used for image fusion techniques. It is found suitable because image is a combination of pixels in the two-dimensional arrangement. The min-max algorithm also can be applied efficiently in the two-dimensional case. An example can illustrate the algorithm clearly. In figure 3.3 In this work, the min-max algorithm is used to identify minimum values of each row and maximum from that. So, when an image is going to fuse, for each orientation of image row values are collected and consecutively minimum and maximum values are identified and the resultant value only gives final fused image. A unit area is identified with a fixed number of pixels. The minimum values of pixels are identified row wise and finally the maximum pixel value is identified for each set. Again, the next unit of the fixed number of pixels is identified. This way the same emotions with different face orientation of the same subjects are fused and only time and frequency domain features are identified following this. This gives much better result than single facial emotion feature extraction.

**C. Discrete Wavelet Transform**

The wavelet transform is the advanced version of Fourier Transform. The fourier transform is the frequency domain analysis of the signal. Frequency domain analysis can give a much better understanding and data than by time domain analysis. Short Time Fourier Transform (STFT) came after the Fourier transform

The Discrete Wavelet Transform (DWT) is easier to implement than CWT. The computational time of DWT is

much lesser than CWT. Mostly in DWT the position of the frequency components and the number of their repetitions can be identified, those are not possible for the fourier transform. In this work Haar wavelet is used. The following figure shows the Haar wavelet.



**Fig. 2. Three level 2D DWT**

In figure 2 low pass filter feature is further decomposed up to third level to get better results. In this research Haar decomposition is used up to fifth level to get better results. This mechanism is used for feature level fusion other than feature extraction in this work.

**D. Regression**

Regression is a statistical measure used in finance, investing and other disciplines to determine the relationship between one dependent and multiple independent variables. It is used for prediction. Regression is basically classified into seven categories. They are linear regression, logistic regression, polynomial regression, stepwise regression, ridge regression, lasso regression and elastic net regression. In this work, logistic regression model is used as it is the best fit for a non-linear model. Logistic regression is used to find the event as success and event as failure. It is widely used for classification problems. Logistic regression doesn't require linear relationship between dependent and independent variables. It can handle various types of relationships because it applies a non-linear log transformation to the predicted odds ratio.

In this work, logistic regression model is used for the feature level fusion purpose. Features from different modalities, for example, speech and physical parameters or EEG and physical parameters have been obtained and taken to single parameter value using logistic regression model.

**E. Fuzzification**

In this work, fuzzification is used for physical parameter modification. Raw data of physical parameters directly collected from experimental values are not fit for direct classification, because the same data may be closely related to two different ranges. In that scenario, the fuzzification method can give a distinct class for each value based on its maximum association to the specific class. Fuzzification is a method to convert crisp values into fuzzified values with the help of membership function. The concept of fuzzification came from real time scenario. The binary can handle only two variables, whereas fuzzy logic can handle multiple variables. Another major advantage of fuzzy logic is its ability to deal with linguistic variables.

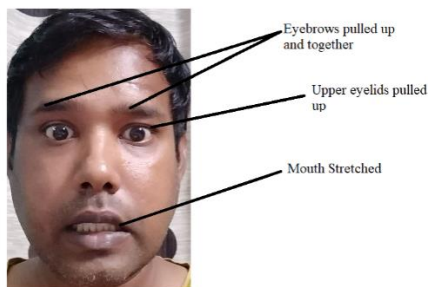


The variables like very low, low, medium, high, very high can be classified with the help of range. Membership function is related to a fixed range. There are multiple types of membership functions like triangular, Gaussian, trapezoidal, bell shaped, sigmoid etc. In this research, fuzzy logic is used for the proper data classification purpose for physical parameters. The fourteen physical parameters have different ranges and are different for different emotions. Same data belongs to different classes. So, it is difficult to identify proper emotions. The degree of association in fuzzy logic separates the data in proper class by fuzzification method and get back the original data by De-fuzzification using the centroid method. The result analysis shows that, there is a significant change in classification as fuzzified data gave more accurate output than non-fuzzified data. The sample facial images from database are shown in figure 3



**Fig. 3. Four basic emotions from single subject from Own dataset**

The fear points collected from the facial images are shown in figure 4.



**Fig. 4. Emotion Points for Fear**

In literature it is stated that multimodal analysis can furnish better result than bimodal analysis. In previous literature five parameters are used for analysis. In this work thirteen physical parameters are used for the multimodal analysis.

The arrangement of the paper is as following

- I. Introduction
- II. Proposed Work
- III. Result and Discussion

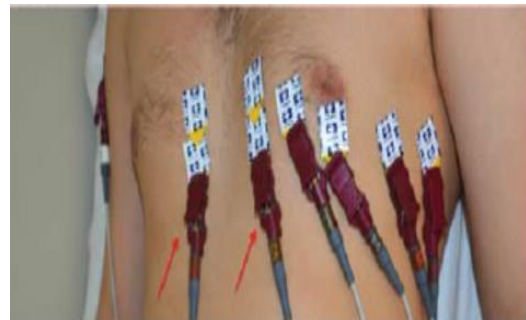
## II. PROPOSED WORK

In this paper human fear is analysed by the feature sets collected from different domains. Then the collected features are fused by discrete wavelet transform and logistic model and finally classified by back propagation algorithm of artificial neural network. The experimental setups for this work are shown in figure 5, 6 and 7. Figure 5 shows the setup for Systolic and Diastolic Blood Pressure Measurement.



**Fig. 5. Setup for Systolic and Diastolic Blood Pressure Measurement**

Figure 6 shows the setup for heart rate measurement.



**Fig. 6. Setup for Heart Rate Measurement**

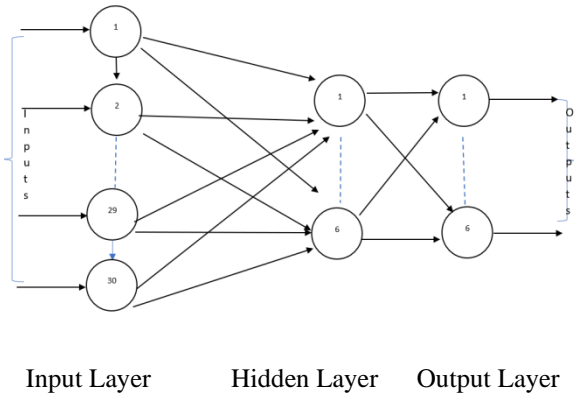
Figure 7 shows the experimental setup for EEG measurement.



**Fig. 7. Setup for EEG Measurement**

Sigmoid function is used as activation function and a threshold value for each emotion is 0.001. Seven features from facial image and thirteen features from the physical parameter domain are fused together to get 30 input feature set for neural networks. The number of input neurons is 30. Since six emotions are classified, six output layers and six hidden layer neurons have been used. 42 datasets with 30 features are taken as training data and 8 datasets with 30 features have been taken as testing data.

Since the emotions are well classified, so features were extracted directly from the dataset and experiment. Then feature level fusion was done using DWT and Logistic Regression Model. The final 30 fused features alone were classified using back propagation neural network. This classification was done for 50 subjects. 42 subjects (85%) were used as training and 8 subjects (15%) are used for testing. Centroid method was used for De-fuzzification of data. Figure 8 shows the back-propagation algorithm of Artificial Neural Network architecture.



**Fig. 8. ANN Architecture for Multimodal Human Emotion Recognition (Facial Image, EEG Signal & Physical Parameters) Multimodal human emotion analysis, shows that the confusion matrix giving very good results for most of the emotions.**

### III. RESULTS AND DISCUSSIONS

The multimodal result analysis is done by confusion matrix. This matrix can identify that how emotions are exactly detected by classification. The confusion matrices for two different fusion methods are shown in Table 2 and Table 3.

**Table-II: confusion matrix of the emotion recognition system based on discrete wavelet transform based multimodal feature level fusion.**

Types of Emotions	a (%)	b (%)	c (%)	d (%)	e (%)	f (%)
a	100	0	0	0	0	0
b	0	100	0	0	0	0
c	0	0	100	0	0	0
d	0	0	0	100	0	0
e	0	10	0	0	90	0
f	0	0	0	0	0	100

Notes: a=Happy, b=Surprise, c=Neutral, d= Anger, e= Sad, f=Fear

**Table-III: Confusion matrix of the emotion recognition system based on logistic regression based multimodal feature level fusion.**

Types of Emotions	a (%)	b (%)	c (%)	d (%)	e (%)	f (%)
a	100	0	0	0	0	0
b	0	100	0	0	0	0
c	0	5	85	0	10	0
d	0	0	0	90	0	10
e	0	0	0	0	100	0
f	0	0	0	0	0	100

**Notes: a=Happy, b=Surprise, c=Neutral, d= Anger, e= Sad, f=Fear**

The classification result of discrete wavelet transform can give better result than logistic regression model. The average classification percentage using discrete wavelet transform is 98.33%. The average classification percentage using logistic regression is 95.83%. The bimodal classification result is 90% for discrete wavelet transform. The average classification percentage using logistic regression is 87.5%. Both the results are improved by 8.33% and 8.33% respectively.

### IV. CONCLUSIONS

This work analyses the multimodal aspect of human emotion recognition. The literature shows the possibility of the multimodal analysis to give better results compared to unimodal and bimodal analysis. The confusion matrix shows almost all the emotions having 100% accuracy for discrete wavelet based feature level fusion compared to logistic regression based feature level fusion. Only sad emotion has been misclassified for discrete wavelet based classification, whereas neutral and anger have been misclassified using logistic regression based feature level fusion. . The average classification percentage using discrete wavelet transform is 98.33%. The average classification percentage using logistic regression is 95.83%. The bimodal classification result is 90% for discrete wavelet transform. The average classification percentage using logistic regression is 87.5%. Both the results are improved by 8.33% and 8.33% respectively.

### ACKNOWLEDGEMENT

The authors are very much grateful to Sree Jeyam Laboratory to render experimental facilities. They are also thankful to the university to provide adequate research facilities.

### REFERENCES

1. Andrea Kliensmith and Nadia, (2013), "Affective body expression perception and recognition: A Survey," IEEE Transactions on Affecting Computing, Vol. 4, pp. 15-33.
2. Ayadi M E, Kamel M S, Karray F, (2011), "Survey on Speech Emotion Recognition: Features, Classification Schemes and Databases," Image and Vision Computing, Vol. 31, pp. 175-185, DOI: 10.1109/TMM.2010.2051872.
3. Baveye Y, Dellandrea E, Chamaret C, Chen L, (2015), "LIRIS-ACCEDE: A Video Database for Affective Current Analysis," IEEE Transactions on Affective Computing, Vol. 4, pp. 43-55.
4. Busso C, Bulut M, Lee C, Kazemzadeh A, Mower E, Kim S, Chang J, Lee S, Narayanan S, (2008), "IEMOCAP: Interactive Emotional Dyadic Motion Capture Database," Journal of Language Resources and Evaluation, Vol. 42, pp. 335-359.
5. Caridakis G, (2007), "Multimodal emotion recognition from expressive faces, body gestures and speech," In Proceedings of International Conference. Artificial Intelligence Applications and Innovations, pp. 375-388, DOI: 10.1007/978-0-387-74161-1-41.

6. Carlos B, Mariooryad S, Metallinou A and Narayanan. S, (2013), "Iterative feature normalization scheme for automatic emotion detection from speech," IEEE Transactions on Affective Computing, Vol. 4, pp. 386-397, DOI: 10.1109/T-AFFC.2013.26.
7. Castellano G, Villalba S D and Camuri A, (2007), "Recognizing Human Emotions from Body Movement and Gesture Dynamics," In Proceedings of 2nd International Conference on Affective Computing and Intelligent Interaction, pp. 71-82, DOI: 10.1007/978-3-540-74889-2\_7.
8. Chanel G, Rebetez C, Betrancourt M, Pun T, (2011), "Emotion assessment from physiological signals for adaptation of game difficulty", IEEE Transactions on Systems, Man and Cybernetics, Vol. 4, pp. 1052-1063.
9. Chen S, Tian Y L, Liu Q and Metaxas D N, (2013), "Recognizing expressions from face and body gesture by temporal normalized motion and appearance features," Image and Vision Computing, Vol. 31, pp. 175-18, DOI: 10.1109/CVPRW.2011.5981880.
10. Douglas Cowie E, Campbell N, Cowie R P, (2003), "Emotional Speech: Towards a new generation of databases," Speech Communication, Vol. 40, pp. 33-60.
11. Dornaika F, Moujahid A, Raducanu B, (2013), "Facial expression recognition using tracked facial actions: classifier performance analysis," Engineering Applications of Artificial Intelligence, Vol. 26, pp. 467-477, DOI: 10.1016/j.engappai.2012.09.002.
12. Dornaika F, Davinie F, (2008), "Simultaneous facial action tracking and expression recognition in the presence of head motion," International Journal of Computer Vision, Vol. 76, pp. 257-281.
13. Gabriele Fanelli, Juergen Gall, Harald Romsdorfer, Thibaut Weise, Luc Van Gool, (2010), "A 3-D Audio-Visual Corpus of Affective Communication," IEEE Transactions on Multimedia, Vol. 12, pp. 591-598.
14. Gangeh M J, Fewzee P, Mohamed G A, Kamel S and. Karray F, (2014), "Multiview supervised dictionary learning in speech emotion recognition," IEEE Transactions on Audio, Video, and Language Processing, Vol. 22, pp. 1056-1068. DOI: 10.1109/TASLP.2014.2319157.
15. Gunes H and Piccardi M, (2005), "Affect Recognition from Face and Body: Early Fusion vs. Late Fusion," In Proceedings of IEEE International Conference on Systems, Man and Cybernetics, Vol. 4, DOI: 10.1109/ICSMC.2005.1571679.
16. Gunes H and Piccardi M, (2009), "Automatic temporal segment detection and affect recognition from face and body display," IEEE Transactions on Systems, Man and Cybernetics, Vol. 39, DOI: 10.1109/TSMCB.2008.927269.
17. Hayat M, Bennamoun M, (2014), "An automatic framework for textured 3D video-based facial expression recognition," IEEE Transactions on Affective Computing, Vol. 5, pp. 301-313, DOI: 10.1109/TAFFC.2014.2330580.
18. Hernandez J, McDuff D, Fletcher R and Picard R W, (2013), "Inside-Out: reflecting on your inner state," In Proceedings of IEEE International Conference on Pervasive Computing and Communications Workshops, pp. 324-327, DOI: 10.1109/PerComW.2013.6529507.
19. Jagdish Lal Raheja, Umesh Kumar, (2010), "Human facial expression detection from detected in captured image using back propagation neural network," International Journal of Computer Science and Information Technology, Vol. 2, pp. 116-123.
20. Karla Conn Welch, (2012), "Physiological signals of autistic children can be useful," IEEE Instrumentation and Measurement Magazine, pp. 28-32.

## AUTHORS PROFILE



**Swagata Sarkar** obtained his Bachelor's degree in Electronics and Communication Engineering from Institution of Engineers, India with all India first rank in ECE. Then she obtained two consecutive Master's degrees in Mobile Communication and Network Technology and Instrumentation and Control Engineering from West Bengal University of Technology and Birla Institute of Technology (BIT, Mesra) respectively.

She has more than 7 years of Industrial and 10 years of teaching experience. She has guided 40 projects in the undergraduate level and 5 projects at the post graduate level.

She has completed DST funded project of Rs.1,00,000/-. She also filed patent for the same project. She has completed project funded by TNSCST of Rs. 10,000/-. She has published 50 research papers in the referred journals and conferences. Her current research interest is Digital Image and Signal Processing and Electromagnetics.