

Fuzzy Rule Based Feature Extraction and Classification of Time Series Signal

Sandya H. B., Hemanth Kumar P., Himanshi Bhudiraja, Susham K. Rao

Abstract: - Time series signal is a continuous signal which varies continuously with respect to time. These signals involve a great deal of useful information, the information content in these signals can be used for Feature Extraction and Classification. The purpose of Feature Extraction is to reduce the dimension of feature space and achieving better performances. The Features are extracted based on the mathematical calculations like Average, Maximum, Minimum, Standard Deviation and Variance. The Classification of extracted features is carried out by Fuzzy Rule based Selection System. Fuzzy Systems (FS) are evaluated for accuracy, multiplexity, flexibility and transparency for simple and complex systems. In this paper mamdani based Fuzzy System is used to achieve accurate results. Based on feature extracted data the Fuzzy System generates a fuzzy score and the Classifier Algorithm classify the feature extracted signals as Good, Bad and Best signals.

Key words: - Fuzzy, Feature Extraction, Classification, Time series signal

I. INTRODUCTION

A Time series is a collection of observations of well-defined data items obtained through repeated measurements over time. Example, measuring the value of retail sales each month will have sales revenue of time series which are measured at every instant of time. Data which are measured randomly are not defined as time series. Therefore, a time series is a sequence number of data collected at regular intervals over a period of time. In statistics, signal processing, econometrics and mathematical finance, a time series is a sequence of data points, measured typically at successive time instants spaced at uniform time intervals. Examples of time series are the daily closing value of the Dow Jones index and the annual flow volume of the Nile River at Aswan. Time series data is analyzed to extract meaningful statistics and other characteristics of the data. The feature values are predicted based on previously observed values using Time series forecasting model. Time series can be decomposed into three components: the trend (long term direction), the seasonal (systematic, calendar related movements) and the irregular (unsystematic, short term fluctuations) [4].

This paper describes Feature Extraction and classification of Time series signals. The Feature Extraction process results in a much smaller and richer set of attributes and can greatly reduce the dimension of feature space without degrading the performance of classifier system [13].

Manuscript received May, 2013.

Sandya H.B., PG Scholar, Deptt. of ECE, AMC Engineering College, VTU, Bangalore, India.

Hemanth Kumar P., Asst. Prof., Deptt. of ECE, AMC Engineering College, VTU, Bangalore, India.

Himanshi Bhudiraja., Asst. Prof., Deptt. of ECE, AMC Engineering College, VTU, Bangalore, India.

Susham K. Rao., Asst. Prof. Deptt. of ECE, AMC Engineering College,VTU, Bangalore ,India.

II. FUZZY SYSTEM

Fuzzy is Probabilistic in nature, an Uncertain and an imprecise. Mathematical concepts are very simple in fuzzy reasoning logic system. Fuzzy logic (FL) is a convenient way to map an input space to an output space. FL is flexible, tolerant of imprecise data and is based on natural language and human communication

Fuzzy Inference System (FIS) is a system used to solve new problems. FIS maps an input features to output classes using FL. FIS can be created by hand, using graphical tools or command line functions, or automatically generated using either clustering or adaptive neuro fuzzy techniques. Fuzzy logic is easy to modify by including or excluding rules, there is no need to start a new FIS from the beginning.

FIS is used to solve “Decision Problems” and act accordingly. The structure of Fuzzy Inference System is shown in Fig 1. FIS consists of four modules, Fuzzification module, Knowledge base module, Inference engine module and Defuzzification module. Fuzzy inference methods are classified as direct methods and indirect methods. Direct methods, such as Mamdani's and Sugeno's, are the most commonly used. Indirect methods are more complex. Mamdani method is most commonly used fuzzy inference technique. Mamdani model is a knowledge driven predictive model, it works with inputs of crisp data and also with intervals and or linguistic terms. The major advantage of this model is it provides a measure of confidence for predicting future value when the actual value is unknown. The important domain of its application is WEB shopping.

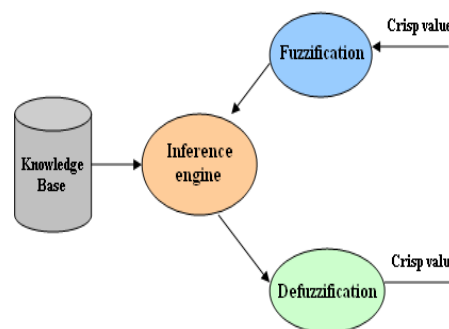


Fig.1. Structure of fuzzy inference system

A. Fuzzy Inference System modules

- 1) **Fuzzification module:** The system inputs, which are crisp numbers, are transformed into fuzzy sets. This is done by applying a fuzzification function.
- 2) **Knowledge base module:** Stores IF-THEN rules provided by experts.
- 3) **Inference engine module:** Using fuzzy inference on the inputs and IF-THEN rules simulates the human reasoning process.



4) *Defuzzification module*: The fuzzy set obtained by the inference engine transforms into a crisp value.

III. DATA DESCRIPTION

The required dataset for the extraction and classification is generated by using Matlab or Simulink with varying frequency in regular time interval. A constantly varying signal measured for uniform interval of time can be considered as a Time series signal. One best example for Time series signal is a simple sine wave whose frequency is varied with respect to time. The generated signals are similar to original real time signals (Ex: EEG or ECG) with small variations in frequency and amplitude. The data is extracted for all the time intervals for sine waves with different frequencies.

IV. METHODOLOGY

In Data Acquisition, the data samples are generated from time series signals and then Features are extracted based on parametric mathematical calculations like average, maximum, minimum, standard deviation and variance. Finally by using fuzzy Inference System, the output is classified as GOOD, BAD and BEST. The proposed methodology is shown in Fig.2.

A. Time Series Signal Generation

The Time Series Signals are generated using matlab with frequency ranging from 8Hz to 17.5Hz with total time period of 2 seconds. Generating a time series signals at frequency of every 0.5Hz. A total 20 input data signals S100 to S119 are generated at 8Hz, 8.5Hz, 9Hz, 9.5Hz, 10Hz, 10.5Hz, 11Hz, 11.5Hz, 12Hz, 12.5Hz, 13Hz, 13.5Hz, 14Hz, 14.5Hz, 15Hz, 15.5Hz, 16Hz, 16.5Hz, 17Hz and 17.5Hz. The total time period of 2 seconds is divided into 200 samples by sampling the signal at a rate of 0.01 seconds. The signals S100 to S119 generate 4000 samples. One of the generated time series signals is shown in Fig.3.



Fig.2. Processing step

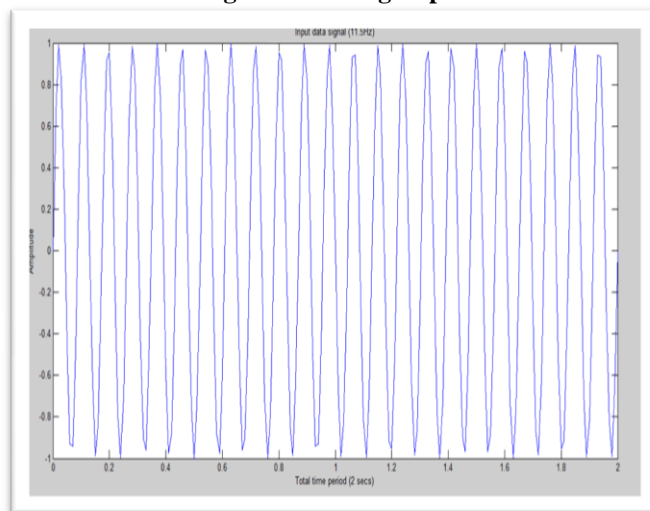


Fig.3. Generation of time series signal

B. Data Set Acquisition

Data Set is collected from the input signal as the signal is varying continuously. It has different value of amplitude at different interval of time. The data is collected for all the different input signals generated. A total of 4000 data set samples are collected. The collected data set is shown in Microsoft Excel Sheet in Fig.4.

$$\text{Total number of samples} = \frac{T}{R} \times n$$

T = Total time period = 2 seconds
 R = Sampling rate = 0.01 seconds
 n = Number of input signals = 20

$$\text{Total number of samples} = \frac{2}{0.01} \times 20$$



= 4000 samples.

Acquisition of Data Set		0	0.01	0.02	0.03	0.04	0.05	0.08	0.09	0.1	0.11	0.12	1.46	1.47	1.93	1.97	1.98	1.99	2
8	(0)	0.481754	0.844320	0.998027	0.904827	0.587785	-0.77051	-0.98229	-0.95106	-0.68455	-0.24869	-0.90483	-0.99803	0.368125	-0.99803	-0.84433	-0.48175	-3.9515	
8.5	(0)	0.589041	0.876007	0.999037	0.844328	0.45399	-0.90403	-0.99956	-0.80902	-0.39715	0.125313	0.535827	0.034211	0.562083	-0.99951	-0.87611	-0.58904	2.94515	
9	(0)	0.535827	0.904827	0.992153	0.770513	0.389017	-0.98229	-0.93970	-0.58779	-0.06279	0.481754	0.770513	0.992153	0.720969	-0.99211	-0.90483	-0.53583	-4.4515	
9.5	(0)	0.562083	0.929776	0.979151	0.684547	0.156434	-0.99803	-0.79016	-0.30902	0.278991	0.770513	-0.72897	-0.22834	0.866742	-0.97915	-0.92978	-0.56208	-1.2514	
10	(0)	0.587785	0.951057	0.951057	0.587785	1.22516	-0.95106	-0.58779	-2.4516	0.587785	0.951057	0.951057	-0.58779	-0.95106	-0.95106	-0.95106	-0.58779	-4.9515	
10.5	(0)	0.622907	0.968083	0.917753	0.481754	-0.15643	-0.84433	-0.33874	0.389017	0.827881	0.998027	0.676307	0.397940	0.999502	-0.91775	-0.96808	-0.62291	1.96515	
13.5	(0)	0.750111	0.992115	0.562083	-0.34869	-0.89101	0.481754	0.979157	0.809017	0.040308	-0.68455	-0.90836	-0.82788	0.387938	-0.56208	-0.99211	-0.75011	-1.4514	
14	(0)	0.770513	0.982387	0.481754	-0.38912	-0.95106	0.684547	0.998027	0.587785	-0.24869	-0.90483	0.368125	-0.48175	0.125313	-0.48175	-0.98229	-0.77051	-6.9515	
14.5	(0)	0.790155	0.968083	0.397348	-0.48175	-0.58779	0.844328	0.948081	0.389017	-0.56208	-0.99803	0.676307	0.917753	-0.84433	-0.39735	-0.96808	-0.79016	2.48515	
15	(0)	0.809017	0.951057	0.389017	-0.58779	-1	0.951057	0.809017	3.67516	-0.80902	-0.95106	-0.58779	0.389017	-0.38902	-0.38902	-0.95106	-0.80902	-2.9514	
15.5	(0)	0.827881	0.929776	0.22834	-0.68455	-0.58779	0.998027	0.612907	-0.30902	-0.94808	-0.77051	-0.72897	-0.97915	-0.58904	-0.22834	-0.92978	-0.82788	-1.5514	
16	(0)	0.844328	0.904827	0.125313	-0.77051	-0.95106	0.982387	0.368125	-0.58779	-0.99803	-0.48175	0.770513	-0.125313	-0.68455	-0.125313	-0.90483	-0.84433	-7.8515	
16.5	(0)	0.880784	0.876307	0.034211	-0.84433	-0.89101	0.948081	0.040308	-0.80902	-0.91775	-0.125313	0.535827	0.999501	-0.82788	-0.03421	-0.87611	-0.88078	-8.8515	
17	(0)	0.876307	0.844328	-0.68279	-0.90483	-0.80902	0.770513	-0.18738	-0.95106	-0.72897	0.24869	-0.90483	-0.68279	-0.92978	0.062793	-0.84433	-0.87611	5.88515	
17.5	(0)	0.891017	0.809017	-0.15643	-0.95106	-0.70711	0.587785	-0.45399	-1	-0.45399	0.587785	-0.38902	-0.80789	-0.80789	0.156434	-0.80902	-0.89101	-1.6514	

Fig.4. Data Set Acquisition

C. Feature Extraction

The purpose of Feature extraction is to reduce the dimension of feature space and achieving better performances. The features are extracted based on the parameters like Average, Maximum, Minimum, Standard Deviation and Variance methods. Total 200 features are extracted. The Feature Extraction results as shown in Microsoft Excel Sheet in Fig.5.

1. Average Method

Average is defined as the sum of 'x' data sets divided by 'N' total number of data sets. It is called the mean average.

$$\text{Average} = \frac{\sum x}{N}$$

Where, x is data

N is total number of datasets

2. Maximum Method

The maximum is defined as the greatest quantity or the greatest value attainable in a given data sets.

$$\text{Max} = \text{maximum}(x_1, x_2, x_3, x_4, \dots, x_N)$$

3. Minimum Method

The Minimum is defined as the lowest quantity or the lowest value attainable in a given data sets.

$$\text{Min} = \text{minimum}(x_1, x_2, x_3, x_4, \dots, x_N)$$

4. Standard Deviation (σ)

Standard Deviation is defined as the how much variation or "dispersion" exists from the average (mean, or expected value). A low standard deviation indicates that the data points tend to be very close to the mean; high standard deviation indicates that the data points are spread out over a large range of values.

$$SD(\sigma) = \sqrt{\frac{\sum(x - \mu)^2}{N}}$$

Where, σ = Symbol of Standard Deviation
μ = mean of all the values in the data set
N = Total Number of values in data set
x = each value in the data set

5. Variance (σ²)

Variance is defined as the square of standard deviation or the variance of a random variable or distribution is the expectation, or mean, of the squared deviation of that variable from its expected value or mean. Thus the variance is a measure of the amount of variation of the values of that variable, taking account of all possible values and their probabilities.

$$\text{variance}(\sigma^2) = SD^2 = \frac{\sum(x - \mu)^2}{N}$$

D. Fuzzy Rule Based Model

The Extracted parameters Average, Max, Min, Standard Deviation and Variance is considered as input variables to the Fuzzy rule based selection process block. Inference System (FIS) maps an input features to output classes using FL. Fuzzy logic are easy to modify a FIS just by including or excluding rules. The fuzzy rules have written for Extracting Features to get results as Good, Bad and Best data sample values. Fuzzy Rule Based selection for five inputs and one output model is shown in Fig.6.

1. Fuzzy Rules

Fuzzy rules are linguistic IF-THEN- constructions that have the general form "IF A THEN B" where A and B are propositions containing linguistic variables. A is called the premise and B is the consequence of the rule.

Feature Extracted Data						
Signals	Frequency	Avg	Max	Min	SD	Var
S100	8	0.190836	0.99803	-0.9823	0.03377	0.02281
S101	8.5	0.1434	0.99951	-0.9956	0.0342	0.02339
S102	9	0.083236	0.99803	-0.9823	0.03562	0.02537
S103	9.5	0.030215	0.99556	-0.998	0.03656	0.02673
S104	10	0	0.95106	-0.9511	0.03627	0.02632
S105	10.5	-0.00075	0.99803	-0.9956	0.03539	0.02505
S106	11	0.024206	0.98229	-0.998	0.03509	0.02463
S107	11.5	0.062829	0.99556	-0.9877	0.03561	0.02536
S108	12	0.099532	0.99803	-0.9823	0.03606	0.02601
S109	12.5	0.120711	1	-1	0.03574	0.02555
S110	13	0.1192044	0.998027	-0.98229	0.035018	0.024526
S111	13.5	0.0962552	0.992115	-0.99951	0.034855	0.024297
S112	14	0.0604868	0.998027	-0.95106	0.035534	0.025253
S113	14.5	0.0244408	0.968583	-0.99951	0.036273	0.026315
S114	15	1.055E-16	1	-1	0.036274	0.026316
S115	15.5	-0.005677	0.998027	-0.98769	0.035661	0.025434
S116	16	0.0076459	0.998027	-0.99803	0.035268	0.024877
S117	16.5	0.0335974	0.929776	-0.94088	0.035563	0.025295
S118	17	0.0617774	0.992115	-0.98229	0.036059	0.026004
S119	17.5	0.0815926	0.987688	-1	0.036032	0.025549

Fig.5. Feature Extraction Result

The Fuzzy Classification result is shown in Fig.10. The interval between 0 to 0.38 is classified as BAD, interval between 0.39 to 0.47 is classified as GOOD and the interval between 0.48 to 0.5 is classified as BEST. These results have plotted in Membership Function as shown in Fig.9.

V. RESULT

The Fuzzy rule based Classification was evaluated for the Feature Extracted data Set. FIS generates score for each input signal based on the fuzzy Constraints and Fuzzy Rules. Finally, the obtained Fuzzy Score is classified as GOOD, BAD and BEST by using algorithm.

if (Fuzzy Score \leq 0.38)

Result = BAD

else if (Fuzzy Score \leq 0.47)

Result = GOOD

else

Result = BEST

The Final Classification of Extracted Features is shown in Fig 10.

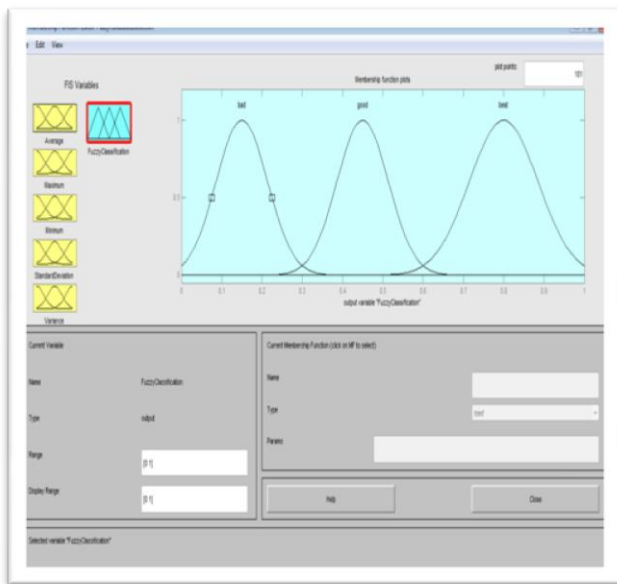


Fig.9. Membership Function for Output Classifier

VI. CONCLUSION

The Time series signals of different frequency ranging from 8Hz to 17.5Hz, 20 input data signals S100 to S119 are generated with time period of 2 seconds and 0.01 sampling rate by using Matlab. The generation of input data signal is shown in Fig.3. Total 4000 samples are acquired in data set acquisition table is shown in Fig.4. The Features are extracted from data set acquisition table, based on the Mathematical calculations like Average Maximum, Minimum, Standard Deviation and Variance methods. Total 200 features are extracted as shown in Fig.5.

These Feature extracted five parameters are considered as input to the Fuzzy Rule Based Selection Block in Fuzzy System. The Fuzzy Rules have designed in Fuzzy Inference System as if sample value is less than or equal to 0.38 considered as bad, if it is less than or equal to 0.47 considered as good and if it is less than or equal to 0.5 considered as best as shown in Fig.7. The Fuzzy Classifier Algorithm assigns classes to the Feature extracted outputs as GOOD, BAD and BEST. Finally, 7 BEST, 4 GOOD and 9 BAD Fuzzy results have obtained, as shown in Fig.10.

Classification of Extracted Features								
Signals	Frequency (Hz)	Avg	Max	Min	SD	Var	Fuzzy_Result	Classification
S100	8	0.19084	0.99803	-0.9823	0.03377	0.02281	0.50	BEST
S101	8.5	0.1434	0.99951	-0.9956	0.0342	0.02339	0.50	BEST
S102	9	0.08324	0.99803	-0.9823	0.03562	0.02537	0.47	GOOD
S103	9.5	0.03021	0.99556	-0.998	0.03656	0.02673	0.47	GOOD
S104	10	0	0.95106	-0.9511	0.03627	0.02632	0.37	BAD
S105	10.5	-0.0007	0.99803	-0.9956	0.03539	0.02505	0.38	BAD
S106	11	0.02421	0.98229	-0.998	0.03509	0.02463	0.50	BEST
S107	11.5	0.06283	0.99556	-0.9877	0.03561	0.02536	0.47	GOOD
S108	12	0.09953	0.99803	-0.9823	0.03606	0.02601	0.47	GOOD
S109	12.5	0.12071	1	-1	0.03574	0.02555	0.50	BEST
S110	13	0.119204	0.998027	-0.98229	0.035018	0.024526	0.5	BEST
S111	13.5	0.096255	0.992115	-0.99951	0.034855	0.024297	0.47	GOOD
S112	14	0.060487	0.998027	-0.95106	0.035534	0.025253	0.47	GOOD
S113	14.5	0.024441	0.968583	-0.99951	0.036273	0.026315	0.5	BEST
S114	15	1.05E-16	1	-1	0.036274	0.026316	0.38	BAD
S115	15.5	-0.00568	0.998027	-0.98769	0.035661	0.025434	0.38	BAD
S116	16	0.007646	0.998027	-0.99803	0.035268	0.024877	0.5	BEST
S117	16.5	0.033597	0.929776	-0.94088	0.035563	0.025295	0.47	GOOD
S118	17	0.061777	0.992115	-0.98229	0.036059	0.026004	0.47	GOOD
S119	17.5	0.081593	0.987688	-1	0.036032	0.025549	0.47	GOOD

Fig.10. Feature Extraction and Classification Result

REFERENCES

[1] Ayan Banerjee, Kanad Basu and Aruna Chakraborty, "Prediction of EEG Signal by Digital Filtering", University of Florida, branch of Computer & Information Science and engineering.

[2] Ramasubramanian V., Time series analysis. I.A.S.R.I., Library Avenue, New Delhi-110 012

[3] Nandish.M, Stafford Michahial, Hemanth Kumar P, Faizan Ahmed, "Feature Extraction and Classification of EEG Signal Using Neural Network Based Techniques", ISSN: 2277-3754 ISO 9001:2008 Certified International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 4, October 2012

[4] Regina Kaiser and Agustín Maravall, Notes on Time Series Analysis Arima Models and Signal Extraction.

[5] Arshdeep Kaur, Amrit Kaur, "Comparison of Mamdani-Type and Sugeno-Type Fuzzy Inference Systems for Air Conditioning System", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-2, May 2012

[6] HU Jian-feng, "Multifeature analysis in motor imagery EEG classification", Proc. IEEE, 2010 Third International Symposium on Electronic Commerce and Security, pp.114-117, 2010.

[7] Abdulhamit Subasia, M. Kemal Kiyimika, Ahmet Alkana, Etem Koklukayab, "Neural Network Classification of EEG signals by using AR with MLE preprocessing for epileptic seizure detection", Mathematical and Computational Applications, Vol. 10, No. 1, pp. 57-70, 2005.

[8] Boyu Wang, Chi Man Wong, Feng Wan, Peng Un Mak, Pui In Mak, and Mang I Vai, "Comparison of Different methods for EEG signal classification", IEEE Trans. Neural Syst. Rehabil. Eng., vol.11, pp. 141-144, June 2003.

[9] D. Garrett, D. A. Peterson, C. W. Anderson, and M. H. Thaut, "Comparison of linear, nonlinear, and feature selection methods for EEG signal classification", IEEE Trans. Neural Syst. Rehabil. Eng., vol.11, pp. 141-144, June 2003.

[10] Deepak R. Keshwani, David D. Jones, George E. Meyer, Rhonda M. Brand, "Rule based Mamdani type fuzzy modelling of skin

permeability", Published in Applied Soft Computing 8 (2008), pp. 285-294; doi: 10.1016/j.asoc.2007.01.007

[11] J.R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interface for communication and control," Clin. Neurophysiol., vol. 133, pp. 767-791, June 2002.

[12] Jang, J.-S. R., "ANFIS: Adaptive-Network-based Fuzzy Inference Systems", IEEE Transactions on Systems, Man, and Cybernetics, Vol. 23, No. 3, pp. 665-685, May 1993.

[13] J. Yen and R. Langari, Fuzzy Logic. Pearson Education, 2004.

[14] K.P. Mohandas and S. Karimulla, "Fuzzy and Neuro-fuzzy modeling and control of non linear systems", Second International Conference on Electrical and Electronics, 2001.

[15] T. J. Ross, Fuzzy Logic with Engineering Applications. John Wiley and sons, 2010.



Sandya H.B is currently pursuing M.Tech, final year in the stream of Digital Electronics and Communication from AMC Engineering College, Bangalore, VTU, Belgaum. She had received her Bachelor's of degree in Electronics and Communication Engineering from V.V.I.E.T College, Mysore, Karnataka, India. Her area of interest includes Digital and Analog Communication, Fuzzy logic, Digital Signal Processing etc.



Hemanth Kumar P. currently pursuing Ph.D. in the stream of Computer Science Engineering from VTU Resource Research Centre, Belgaum, Karnataka had received his Master's degree in Digital Communication from National Institute of Technology (MANIT) Bhopal, Madhya Pradesh, India in the year 2010 and Bachelor's Degree from Global Academy of Technology, Bangalore in 2008. Currently working as Assistant Professor in AMC Engineering College, Bangalore in Department of Electronics & Communication Engineering. His area of interest includes Soft Computing Techniques, Data Mining, Pattern Classification, Communication etc. He has have three years of Teaching Experience and guided six M.Tech projects successfully. He has published more than 10 International Journals.



Himanshi Budhiraja is currently working as Assistant Professor in AMC Engineering College, Bangalore in Department of Electronics & Communication Engineering. She had received Master's degree and Bachelor's Degree in branch ECE from Kurukshetra University, Kurukshetra, Haryana, India. She has Six years of Teaching Experience. Her specialization is in field of Optical

Communication



Susham K Rao is currently working as Assistant Professor in AMC Engineering College, Bangalore in Department of Electronics & Communication Engineering. She had received Master's degree in branch VLSI and Embedded Systems from Reva Institute of Technology (RITM) Bangalore, Karnataka, India and Bachelor's Degree from KVG Engineering College, Sullia. She has seven years

of Experience which includes Industry and Teaching. Her area of interests is VLSI, Signal Processing, Embedded systems etc.