

UNDERSTANDING AND ESTIMATION OF EMOTIONAL EXPRESSION USING ACOUSTIC ANALYSIS OF NATURAL SPEECH

Nilima Salankar Fulmare¹, Prasun Chakrabarti² and Divakar Yadav³

^{1,2}Department of Computer Science, Sir Padampat Singhanian University, Udaipur
³IET, Lucknow

ABSTRACT

Emotion expression is an essential function for daily life that can be severely affected some psychological disorders. In this paper we identified seven emotional states anger, surprise, sadness, happiness, fear, disgust and neutral. The definition of parameters is a crucial step in the development of a system for emotion analysis. The 15 explored features are energy intensity, pitch, standard deviation, jitter, shimmer, autocorrelation, noise to harmonic ratio, harmonic to noise ratio, energy entropy block, short term energy, zero crossing rate, spectral roll-off, spectral centroid and spectral flux, and formants. In this work database used is SAVEE (Surrey audio visual expressed emotion). Results by using different learning methods and estimation is done by using a confidence interval for identified parameters are compared and explained. The overall experimental results reveals that Model 2 and Model 3 give better results than Model 1 using learning methods and estimation shows that most emotions are correctly estimated by using energy intensity and pitch.

KEYWORDS

Emotion, Confidence Interval, Speech, SAVEE database, Paralinguistic information

1. INTRODUCTION

Human speech is an acoustic waveform generated by the vocal apparatus, whose parameters are modelled by the speaker to convey information. The physical characteristics and the mental state of the speaker also determine how these parameters are affected and consequently how speech conveys intended and on occasion unintended information. Knowledge about how these parameters characterise the information is not explicitly available human brain is able to decipher the information from resulting speech signal, including the emotional state of the speaker. Information about emotional state is expressed via speech through numerous cues, ranging from low level acoustic ones to high level linguistic content. Several approaches to speech based automatic emotion recognition, each taking advantage of few of these cues have been explored [1]-[9]. The most commonly used acoustic and prosodic based on cepstral coefficients, pitch, intensity and speech rate. The research of automatic speech emotion recognition enhance the efficiency of people's work and study and its helpful for solving problems more efficiently. In the present work, we report results on recognizing emotional states from a corpus of short duration spoken utterances. Specifically we aim at recognizing six emotional states fear, happiness, sadness, surprise, disgust, anger and additional neutral. The rest of the paper is as follows: In section II we describe the database that were used for an analysis. Section III describes the feature extraction process and the composition of feature vectors. Subsequently in

section IV Experiments and Evaluations are presented .Section V describes conclusion and future work.

2. WORKING DATABASE

Both extraction of features and the emotion identification experiments described in this paper were carried using the SAVEE(Surrey audio visual expressed emotion) database[10].The database consists of four actors of ages 27-31 depicting the widely used six basic emotions (fear,anger,disgust,sadness,surprise,happiness) plus the neutral state.Recording consists of 15 phonetically balanced TIMIT sentences per emotion(with additional 30 sentences for neutral state) resulting into corpus of 480 British English utterances In this study we have analysed three speakers.This database was chosen because it presents certain characteristics that were of interest.

3. FEATURE EXTRACTION

Performance of any emotion recognition strategy largely depends on how relevant features ,invariants to speaker ,language and contents could be extracted .Our approach considers a feature vector which consists of 15 basic acoustic features.Features extracted are *intensity,pitch,standard deviation,jitter,shimmer,autocorrelation,noise to harmonic ration,harmonic to noise ration,energy entropy block,short term energy,zero crossing rate,spectral roll-off,spectral centroid and spectral flux,and formants*.Previous research in the field of emotion recognition has shown that emotional reactions are strongly related to pitch and energy of the spoken message[11].For example the pitch of speech associated with anger or happiness is always higher than that of associated with sadness or fear and the energy associated with anger is greater than that associated with fear. For extraction of features software used is PRATT[12]. Some features are extracted by using this software and for the rest features matlab has been used.

A. Composition of Features

Fifteen features are grouped in a different combination for the precise understating of classification accuracy and dependency of features.

- 1.Energy Intensity+Pitch
- 2.EnergyIntensity+Pitch+Standard Deviation+Jitter+Shimmer
- 4.All extracted features

4. EXPERIMENTS AND EVALUATION

Our task was to evaluate the performance of Neural Network(NN),NaiveBayes(NB),Classification Tree(CT) and KNN algorithm for the identification of seven emotional states.For these experiments we have used Orange Canvas[13].Experimental setup used for analysis includes Neural Network with 20 hidden layers,1.0 regularization factor and maximum iterations 300. Naïve Bayes with size of LOESS window 0.5.Classification Tree with attribute selection criteria Information Gain. KNN with the metrics Euclidean .70%(252 utterances) of data has been used for training and 30% (108 utterances) data used for testing purpose. Testing method used for evaluation is CV10. The evaluation based on classification accuracy(CA),Sensitivity(SE),Specificity(SP),Brier Score(BS).An explanation of the features are CA:Proportion of the correctly classified example.

BA:Average deviation between the predicted probabilities of events and actual events.

Sensistivity and Specificity is as shown in table I.

TABLE I
SENSITIVITY AND SPECIFICITY

	<70	>=70
Clinical	True positive	False negative
NonClinical	False positive	True negative

As we are measuring human behaviour and not a physical characteristics, there is always some measurement error inherent in all clinical tests. Sometimes a test has cut-off score to determine if the client is at risk and should be referred for more in-depth testing or has a particular disorder. So in hypothetical test of language development any client with a score more than two standard deviations below the mean(i.e. 70) will be classified as having a language disorder and any subject whose scores above 70 will be classified as nonclinical[14].As shown in table II subjects in the known clinical samples with score below 70 are considered true positive because they are correctly classified as having disorder. This percentage represents an estimate of the *sensistivity* of the test. Subjects in the nonclinical samples with score of 70 or higher are considered true negatives as they are correctly classified as not having the disorder. This percentage represent an estimate of the *specificity* of the test.

B. Results

In an experimentation 3 models have been designed

Model 1:Composed of attributes Energy Intensity and Pitch.

Results of this model are summarized in Table IV and Table V.Table IV summarize the CA,BS and sensistivity and specificity for 7 identified classes.Table V summarize Confusion matrix for the correct identified class.

Model 2: Composed of attributes Energy Intensity,Pitch,Standard Deviation,Jitter and Shimmer. Results of this model are summarized in Table VI and Table VII.Table VI summarize the CA,BS and sensistivity and specificity for 7 identified classes. Table VII summarize Confusion matrix for the correct identified class.

Model 3: Composed of all extracted features mentioned in section III. Results of this model are summarized in Table VIII and Table IX. Table VIII summarized the CA,BS and sensistivity and specificity for 7 identified classes. Table IX summarize Confusion matrix for the correct identified class.

C. Comparative Results

Among three designed models highest classification accuracy achieved for **Model 2** as shown in table II.

TABLE II

CLASSIFICATION ACCURACY OF MODELS

Model	NN	NB	CT	KNN
1	70.80%	68.33%	61.53%	65.28%
2	71.39%	72.89%	73.87%	74.39%
3	59.85%	49.00%	53.80%	50.18%

As per Confusion Matrix highest accuracy achieved for the classification of seven emotions is as shown in Table III.

TABLE III

EMOTION CLASSIFICATION ACCURACY OF MODELS

Emotion	Model	Accuracy(%)	Classifier
Anger	1	87.1	NN
Disgust	2	68.2	KNN
Fear	2	66.7	NN
Happiness	2	79.3	CT
Neutral	1	91.8	NN/NB
Sadness	1	96.8	NN
Surprise	1	77.4	NN

D. Summary of Results

Model 1 and Model 2 gives the highest result. Neutral Anger,Sadness and Surprise is best classfied by Model 1 and Disgust Fear and Happiness best classified by Model 2.

TABLE IV

ENERGY INTENSITY+PITCH

Classifier	CA	BS	SE							SP						
			A	D	F	H	N	S	U	A	D	F	H	N	S	U
NB	68.33	0.44	70	51	0	51	91	93	74	94	94	99	90	94	97	91
NN	70.80	0.44	87	51	0	48	91	96	77	95	94	1	93	94	96	92
KNN	65.28	0.57	74	48	0	41	86	93	54	94	93	0	90	94	98	92
CT	61.53	0.54	74	44	0	45	85	83	54	94	91	97	88	94	96	91

A:Anger D:Disgust F:Fear H:Happiness N:Neutral S:Sad U:Surprise

TABLE V

CONFUSION MATRIX FOR ATTRIBUTES ENERGY INTENSITY+PITCH

Algo	Anger(%)	Disgust(%)	Fear(%)	Happiness(%)	Neutral(%)	Sadness(%)	Surprise(%)
NN	87.1	51.1	0.0	45.2	91.8	96.8	77.4
KNN	74.2	48.9	23.5	41.9	86.9	93.5	54.8
NB	71.0	51.1	0.0	51.6	91.8	93.5	74
CT	74.2	44.4	0.0	45.2	85.2	83.9	54.8

TABLE VI

ENERGY INTENSITY+PITCH+STANDARD DEVIATION+JITTER+SHIMMER

Classifier	CA	BS	SE								SP							
			A	D	F	H	N	S	U	A	D	F	H	N	S	U		
CT	73.87	0.42	87	58	05	74	93	90	61	94	96	95	96	96	97	93		
NN	71.39	0.41	80	33	11	80	89	93	66	95	97	99	91	93	94	95		
KNN	74.39	0.40	77	62	50	64	89	90	61	94	96	97	93	96	98	95		
NB	72.89	0.39	80	41	05	74	97	87	77	97	94	97	91	93	97	93		

A:Anger D:Disgust F:Fear H:Happiness N:Neutral S:Sad U:Surprise

TABLE VII

CONFUSION MATRIX FOR ATTRIBUTES ENERGY INTENSITY+PITCH+STANDARD DEVIATION+JITTER+SHIMMER

Algo	Anger(%)	Disgust(%)	Fear(%)	Happiness(%)	Neutral(%)	Sadness(%)	Surprise(%)
NN	78.1	61.5	66.7	62.5	80.4	74.4	57.1
KNN	70.6	68.2	64.3	64.5	87.2	90.3	55
NB	83.3	66.7	20.0	62.2	81.8	84.4	56.0
CT	73.6	70.0	10.0	79.3	87.8	87.5	50

TABLE VIII

ALL EXTRACTED FEATURES

Classifier	CA	BS	SE							SP						
			A	D	F	H	N	S	U	A	D	F	H	N	S	U
CT	53.80	0.71	60	16	35	41	80	68	48	95	88	91	93	91	94	90
NN	59.85	0.49	78	29	38	51	88	51	51	94	92	94	94	87	96	91
KNN	50.18	0.7	43	25	38	41	67	59	58	91	94	94	90	86	92	91
NB	49.00	0.77	66	25	29	41	67	51	41	92	86	93	93	90	95	89

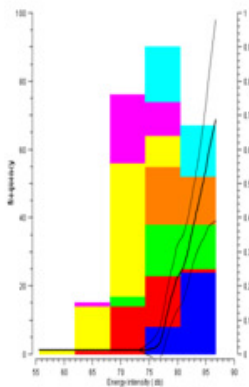
A:Anger D:Disgust F:Fear H:Happiness N:Neutral S:Sad U:Surprise

TABLE IX

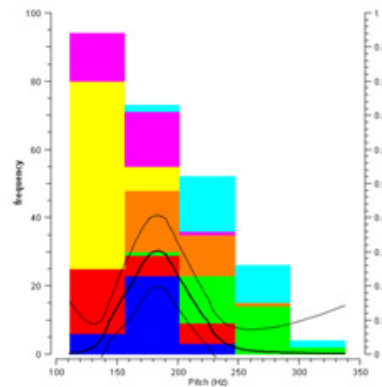
CONFUSION MATRIX FOR COMPLETE EXTRACTED FEATURES

Algo	Anger(%)	Disgust(%)	Fear(%)	Happiness(%)	Neutral(%)	Sadness(%)	Surprise(%)
NN	69.4	36.0	52.2	59.3	69.6	66.7	45.7
KNN	43.8	42.1	48.0	38.2	61.8	51.9	50.0
NB	56.8	21.1	37.5	44.4	70	61.5	36.1
CT	67.9	16.7	36.7	46.4	76.9	65.6	41.7

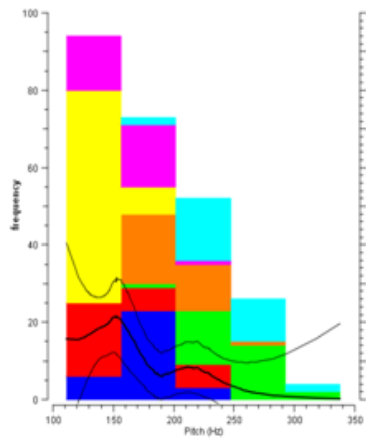
We have done estimation of emotions on the basis of two features Energy Intensity and Pitch. Graphs 1a and 1b, shows that a confidence interval is more wider for emotion anger by using pitch as an attribute as compare to energy intensity. Graphs 2a and 2b shows that confidence interval is more wider for emotion disgust by using energy intensity attribute as compare to pitch. Graphs 3a and 3b shows that confidence interval is more wider for emotion fear by using pitch attribute as compare to Energy Intensity. Graphs 4a and 4b shows that confidence interval is almost same for emotion happiness by using energy intensity and pitch. Graphs 5a and 5b shows that estimation is not possible for an emotion sadness by using an attribute energy intensity and pitch we need to provide some more information for its estimation. Graphs 6a and 6b shows that confidence interval is more wider for emotion surprise by using an attribute pitch as compare to energy intensity.



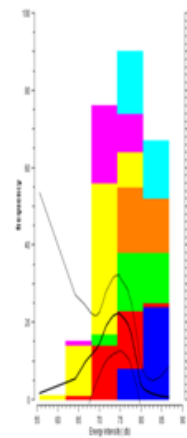
Graph 1 a (Anger)



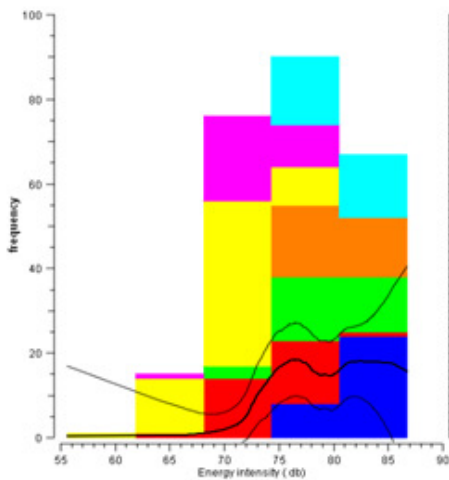
Graph 1b(Anger)



Graph 2 a (Disgust)

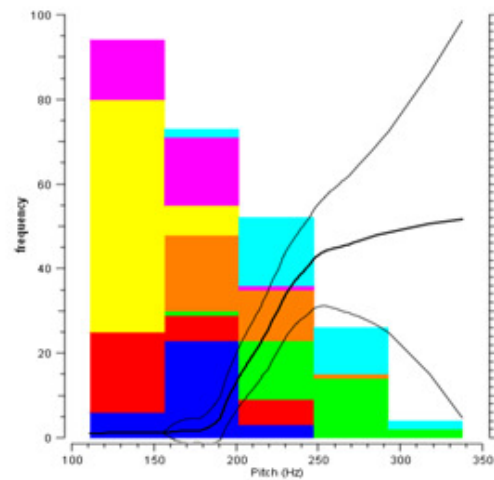


Graph 2b(Disgust)



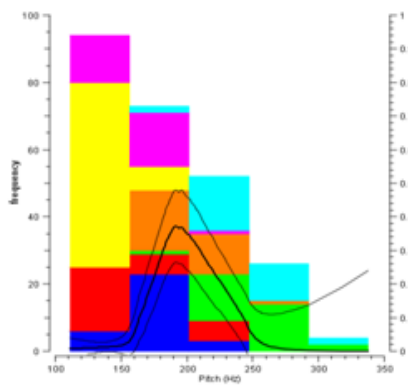
Graph 3 a (Fear)

Graph 3 a (Fear)

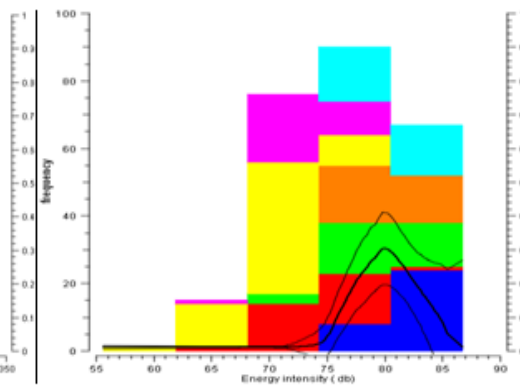


Graph 3b(Fear)

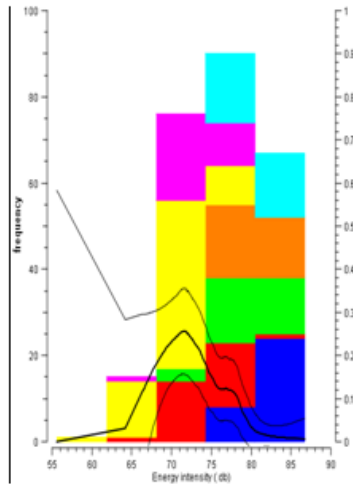
Graph 3b(Fear)



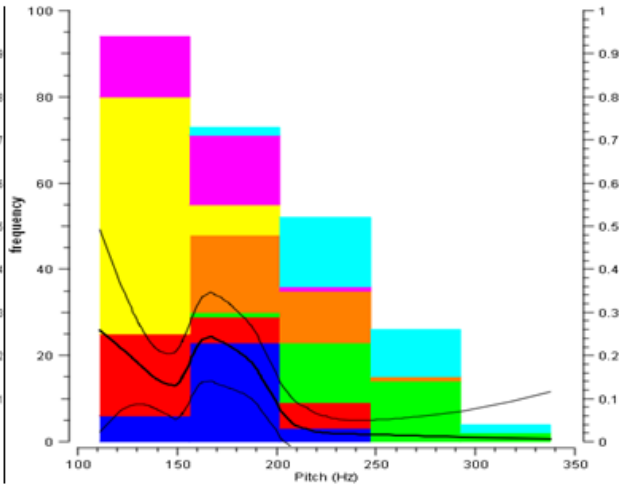
Graph 4 a (Happiness)



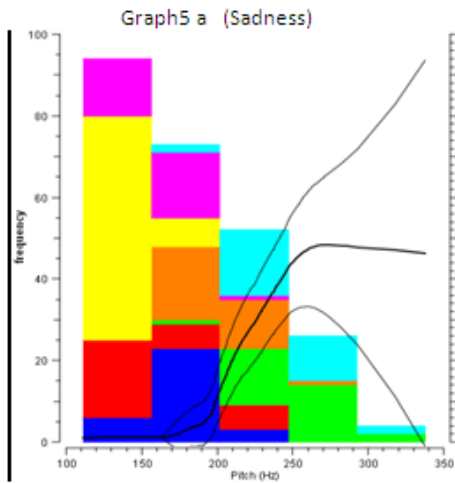
Graph 4b(Happiness)



Graph5 a (Sadness)

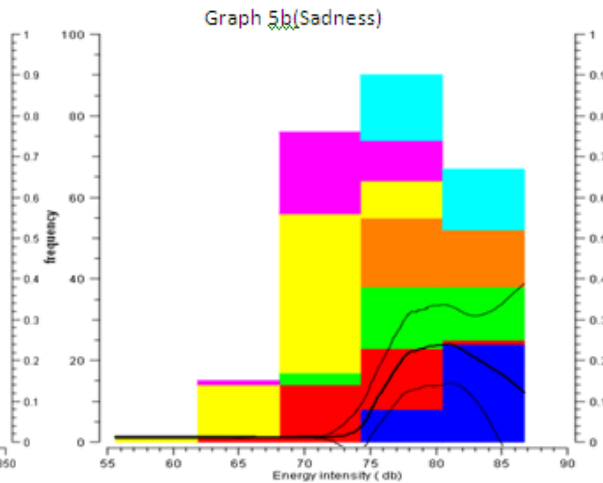


Graph 5b(Sadness)



Graph6 a (Surprise)

Graph a (Surprise)



Graph 6b(Surprise)

Graph 6b(Surprise)

5. CONCLUSION and FUTURE WORK

This paper presents a comparative study of different classifiers based on different combination of features. Feature Extraction was performed by using PRATT software. This paper analyses which combination of features provide better results. By using confidence interval estimation

we can estimate most of the basic emotions with less wider range either by using an energy intensity or pitch.

Future work includes expanding of the study based on larger extracted features and more number of database. And also modelling the understanding of the classification based on different feature combination by using an Event-B approach.

REFERENCES

- [1] R Barra .JM Montero,J Macias-Guarasa,LF D'Haro,R san-Segundo,R Cordoba,prosodic and segmental rubrics in emotion identification,in Proceedings of the 2006 IEEE International Conference on Acoustic,Speech and Signal processing ,vol 1(IEEE Piscataway,2006),pp 1
- [2] M Borchert .A Dusterhoft,Emotions in speech-experiments with prosody and quality features in speech for use in categorical and dimensional emotion recognition environments in proceedings of 2005 International Conference on natural Languages Processing and Knowledge Engineering(IEEE NLP-KE '05),(IEEE ,Piscataway,2005),pp. 147-151
- [3] M Lugger,B Yang,An incremental analysis of different feature groups in speaker independent emotion recognition,in Proceedings of the 16th International Congress of Phonetic Sciences,Saarbruecken,August 2007(IEEE,Washington,2007) pp. 2149-2152
- [4] M pantic, LJM Rothkrantz,Toward an affect-sensitive multimodal human computer interaction,proc IEEE 91,1370-1390(2003)
- [5] D Ververidis,C Kotropoulos,Emotional speech recognition :resources,features,and methods,Speech Communication 48,1162-1181(2006)
- [6] L Vidrascu,L. Devillers,Five emotion classes detection in real world call center data:the use of various types of paralinguistic features,in Proceedings of International Workshop on Paralinguistic Speech-2007,Saarbrucken,3 August 2007,pp.11-16
- [7] S Yacoub,S Simske,X Lin,J Burns,Recognition of emotions in interactive voice response systems,in Proceedings of the 8th European Conference on Speech Communication and Technology,EUROSPEECH 2003-INTERSPEECH 2003,Geneva,1-4 September 2003,pp.729-732
- [8] D Bitouk,R verma,A Nenkova,Class-level spectral features for emotion recognition,Speech Communicatin 52,613-625(2010)
- [9] M ElAyadi,MS kamel,F Karray,Survey on speech emotion recognition:features,classification schemes, and databases,pattern Recognition 44,572-587(2011)
- [10] S.Haq,P.Jackson,and J.Edge,Audio-Visual fature selection and reduction for emotion classification .In proc,AUSP,pages 185-190,2008
- [11] Tatham M.,Morton K.,Expressions in Speech:Analysis & Synthesis,Oxford Lingiistics,2004
- [12] P.Boersma,"Accurate short term analysis of the fundamental frequency and the harmonics-to-noise ration of a sampled sound".Procs. institute of Phonetic Sciences 17.Univ. Amsterdam. Pp. 97-110.1993.
- [13] Demsar J,Zupan B(2004) Orange:From Experimental machine Learninfo to Interactive Data Mining ,White Paper(www.ailab.si/orange),Faculty of Computer and Information Science University of Ljubjana.
- [14] Lawrence G.Weiss,Patrica Zureich,A primer on Psychometrics,The important points for Speech language-Pathologists.

Authors

Nilima Salankar Fulmare is serving as Assistant Professor Engineering of Sir Padampat Singhania University ,Udaipur,India. She has done her B.E. in computer science and engineering from Shri Guru Govind Singhji college of Science and Technology.Mtech.Tech in Computer Engineering and pursuing PhD(Engg) in the area of



Human Computer Interaction

Dr.Prasun Chakrabarti (09/03/1981) is serving as Associate Professor and Head of the department of Computer Science and Engineering of Sir Padampat Singhania University ,Udaipur,India. He has done his B.Tech in Computer Engineering in 2003 , M.E. inComputer Science and Engineering in 2005 and PhD(Engg) in the area of Information Security in 2009. Dr.Chakrabarti has filed 8 Indian Patents and has about 113 publications in his credit. He visited Waseda University, Japan from 16th to 25th May 2012 as Honorary Visiting Professor under INSA-CICS Travel Fellowship Award. He is a senior member of IACSIT(Singapore) ; life-member of Indian Science Congress Association , Calcutta Mathematical Society , Calcutta Statistical Association , Indian Society for Technical Education , Cryptology Research Society of India , IAENG(Hong Kong), Computer Science Teachers' Association(USA), International Institute for Software Testing(USA) .He is Reviewer of International journal of Information Processing and Management (Elsevier) , IEEE Transactions on Industrial Informatics , Journal of Medical Systems , Springer , etc.



Divakar Singh Yadav,working as Professor in Department of Computer Science & Engineering, Head, Department of Computer Science and Engineering, Head, Training and Placement Cell, Institute of Engineering and Technology, Lucknow.

