Volume 2, Issue 1, January 2012 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

AMBIENT TECHNIQUES OF RANDOMNESS IN DECEPTION DETECTION UNDER FUZZIFIED APPROACH

S.RAJKUMAR[*]
Research Scholar/Bharathiar University,
AP&HOD/ CSE, NIET,Coimbatore.

V.NARAYANI
Research Scholar,
Dept. of Computer Science
St.Xavier's College, Tirunelveli

DR.S.P.VICTOR
Asso.Prof. & HOD
Dept. of Computer Science
St.Xavier's College, Tirunelveli

Abstract -**Nowadays computer mediated communication is an essential part of our efficient Information sharing system. Most of our recent communications are performed by using SMS, Email, Chatting (Text/audio/video), twitting etc among our sophisticated world of wireless communication. But criminals or antisocial elements often intentionally falsify their identity in order to escape from Identity revealing schema. This paper deals with the fuzzification features which play the major role of Deception in a Randomized environment of Email Id creation with several fields. In this paper we proposed a Research model which concedes the linkage of fuzzification and randomness in Deception Detection. We implement our proposed model with an experiment which includes the randomized probability over the fuzzification in identity detection of falsified datum. Enumerated results and discussion moulds the impact of fuzziness and randomness in Deception Detection.**

Keywords- **Deception, Ambient, Uncertainty, Fuzzy logic, Randomness**

I. INTRODUCTION

Detection of Deception is useful for managers, employers, and for anyone to use in everyday situations where telling the truth from a lie can help prevent you from being a victim of fraud/scams and other deceptions. This is just a basic run down of physical gestures and verbal cues that may indicate someone is being untruthful.

A. Identifying the Deception

The following table illustrates the identification rate for several components in deception detection system.

TABLE I
Sample Datum for Deception Detection Analysis

Component	Identification rate
Body Language of Lies	40 %
Emotional Gestures & Contradiction	60 %
Interactions and Reactions	65%
Verbal Context and Content	75%
Subject Change frequently	70%
Using humor or sarcasm to avoid a subject.	75%

B. Fuzzification

Fuzzy sets have movable boundaries, i.e., the elements of such sets not only represent true or false values but also represent the degree of truth or degree of falseness for each input.

Fuzzy logic is the part of artificial intelligence or machine learning which interprets a human's actions. Computers can interpret only true or false values but a human being can reason the degree of truth or degree of falseness. Fuzzy models interpret the human actions and are also called intelligent systems.

Fuzzy logic has mostly been applied to control systems. Fuzzy control systems interpret the expert human and replace them for performing certain tasks such as control of a power plant. Fuzzy controllers apply decision rules (if-then rules) by making use of critical variables to interpolate the output between the crisp boundaries. Some typical examples where fuzzy logic has been implemented are

1. Railway (Sendai Railways in Japan)
2. Automobile industries (transmission and braking)
3. Heating and cooling systems
4. Copy machines
5. Washing machines

Fuzzification is the process of changing a real scalar value into a fuzzy value. This is achieved with the different types of fuzzifiers.

Fuzzification of a real-valued variable is done with intuition, experience and analysis of the set of rules and conditions associated with the input data variables. There is no fixed set of procedures for the fuzzification.

C .Randomness

The Oxford English Dictionary defines 'random' as "Having no definite aim or purpose; not sent or guided in a particular direction; made, done, occurring, etc., without method or conscious choice; haphazard." This concept of randomness suggests a non-order or non-coherence in a sequence of symbols or steps, such that there is no intelligible pattern or combination.

D. Pseudorandom

A pseudorandom variable is a variable which is created by a deterministic procedure (often a computer program or subroutine) which (generally) takes random bits as input. The pseudorandom string will typically be longer than the original random string, but less random (lessen tropic, in the information theory sense). This can be useful for randomized algorithms.

E. Randomized Algorithm

A randomized algorithm is an algorithm which employs a degree of randomness as part of its logic. The algorithm typically uses uniformly random bits as an auxiliary input to guide its behavior, in the hope of achieving good performance in the "average case" over all possible choices of random bits. Formally, the algorithm's performance will be a random variable determined by the random bits; thus either the running time, or the output (or both) are random variables.

II. RESEARCH DESIGN MODEL

The proposed research design model comprises the combination of Randomness and Fuzziness as follows,

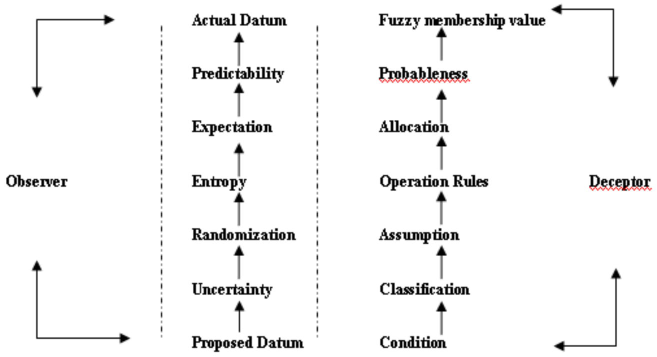


Fig-1-Proposed Model

III. RESEARCH METHODOLOGY

The incorporation of fuzzified anomaly in the field of Email identification can be analyzed as follows:

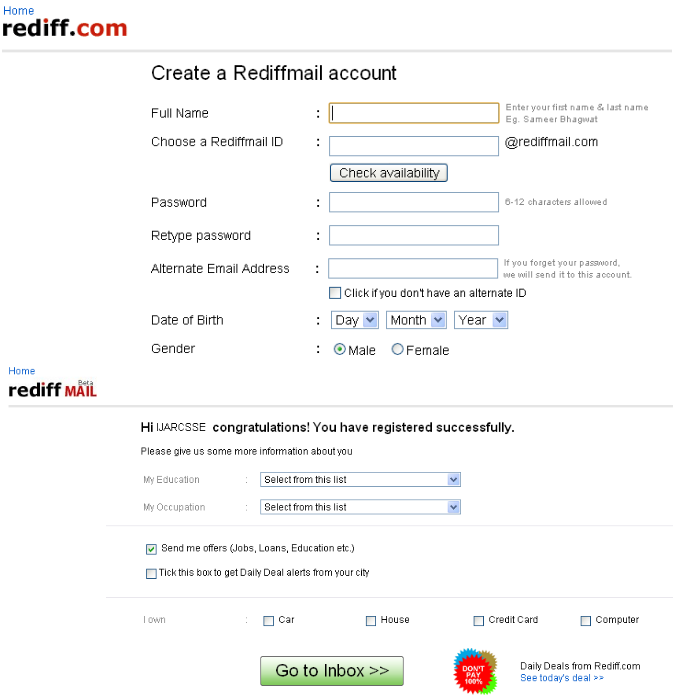
1) Specify the range of conditions

$$0 <= C_{data}(x) = \mu_t(x) <= 1$$

Candidate data at time t holds the membership function $\mu_t(x)$

2) Classification & Categorization

Let us take the most popular rediff mail account for our research analysis



Home
rediff.com

Create a Rediffmail account

Full Name : [ ] Enter your first name & last name Eg. Sameer Bhagwat

Choose a Rediffmail ID : [ ] @rediffmail.com

Check availability

Password : [ ] 6-12 characters allowed

Retype password : [ ]

Alternate Email Address : [ ] If you forget your password, we will send it to this account.

□ Click if you don't have an alternate ID

Date of Birth : Day Month Year

Gender : ● Male ○ Female

Home
rediff MAIL Beta

Hi IJARCSSE congratulations! You have registered successfully.

Please give us some more information about you

My Education : Select from this list

My Occupation : Select from this list

☑ Send me offers (Jobs, Loans, Education etc.)

□ Tick this box to get Daily Deal alerts from your city

I own : □ Car □ House □ Credit Card □ Computer

Go to Inbox >>

Daily Deals from Rediff.com
See today's deal >>

Fig-2-Rediffmail ID Creation pages

We now provide the deception labels for the mail creation as follows,

Name deception	- A
ID deception	- B
Alternate mail ID is true	- C
DOB deception	- D
Gender Deception	- E
Address Deception	- F
Education Deception	- G
Occupation Deception	- H
Send me offers – yes/no	- I
Send me updates – yes/no	- J
Owns car/house/cc/computer	- K

=> action is must

Category: 2^{12} combinations reduced as
C1={A,B,C}C2={D}C3={E,F}C4={G,H} C5= {I,J,K}
=> 2^5 combinations are possible = 32 combinations

Membership value $\mu_t(x)$		
C1	3/11	.95
C2	1/11	.80
C3	2/11	.73
C4	2/11	.3
C5	3/11	.1

3) Probing the assumptions

Based on the associated
Probing the assumptions
We have C1, C2 &C3
Number of combinations = 2^3 = 8
= {ø, C1, C2, C3, C1C2, C1C3, C2C3, C1C2C3}

4) Operational Rules

```
If ( DD(C1) = true) then
  If (DD(C2) = true) then
    If (DD(C3) = true) then
If (DD(C1&C2) = true) then
        C1& C3
        C2&C3
        C1&C2&C3 = true
    DD =true
        Action of deceiver = crime possible
        $\mu_t(x)$ = 0.75
Else If (DD(C4&C5) = true) then
    Action of deceiver = Expectation & Exaggeration
Else if (DD(Page 1) = true & Page 2 =Null &
    TimeDiif(Pageloadtime–Time(clickgotoinbox)<=3
        DD=true
        $\mu_t(x)$ = 0.9
```

IV.RANDOMNESS-ENTROPY

Shannon denoted the entropy H of a discrete random variable X with possible values {x_1, ..., x_n} as,

$$H(X) = \mathrm{E}(I(X)).$$

Here E is the expected value, and I is the Information content of X. I(X) is itself a random variable. If p denotes the probability mass function of X then the entropy can explicitly be written as

$$H(X) = \sum_{i=1}^{n} p(x_i)\, I(x_i) = \sum_{i=1}^{n} p(x_i) \log_b \frac{1}{p(x_i)} = -\sum_{i=1}^{n} p(x_i) \log_b p(x_i),$$

where b is the base of the logarithm used. Common values of b are 2, Euler's number e, and 10, and the unit of entropy is bit for b = 2, nat for b = e, and dit (or digit) for b = 10.[3]

In the case of $p_i = 0$ for some i, the value of the corresponding summand 0 $\log_b$ 0 is taken to be 0, which is consistent with the limit:

$$\lim_{p \to 0+} p \log p = 0 .$$

The proof of this limit can be quickly obtained applying l'Hôpital's rule.

$$H(x) = - \sum_{i=1}^{n} P(x_i)\log_b(P(x_i))$$

V. EXPERIMENT

The collection of 30 students from final year CSE and ask them to sent the feedback of staff members to ertrewtewrt@rediffmail.com (which initiates Deception) which will not be explored to anybody in a lab.

All the subjects are given 20 minutes for simultaneous rank providing strategy towards staff members. After that we open the mail-ID ertrewtewrt@rediffmail.com in front of all the students and then delete all the 30 information mails without opening it. (Such instructions each are mail using a new ID). We then explain the research activity about the Deception Detection, so we requested them to open their mail settings with their permission. We perform the entire analysis with the following stages.

Name deception:

In Stage-1-Evaluating the datum based on Fuzzy conditions we obtain the following,

Original name: Sundar

Possible alterations = {Chandar, Sandar, Sundari, Sundarrajan, Sndar, Sndr, Sunder1980, Sunder2011, Sund1980, Sund_2011}

Name Alterations	= 04/40	- A
Celebrity names	=08/40	- B
Nick names	=08/40	- C
Irrelevant names	= 05/40	- D
Family name mixing	= 02/40	- E
Friends name mixing	=02/40	- F
Non related (Family object, Household, Natural things etc)	=11/40	-G

Now calculate the values using the Random Entropy we obtain the following,

TABLE II
Entropy Values for Name Deception

X_i	$P(X_i)$	$LogP(X_i)$	$P(X_i) * Log(P(X_i))$
Name Alterations	0.1	-1	-0.1
Celebrity names	0.2	-0.69897	-0.13979
Nick names	0.2	-0.69897	-0.13979
Irrelevant names	0.16	-0.79588	-0.12734
Family name mixing	0.05	-1.30103	-0.06505
Friends name mixing	0.05	-1.30103	-0.06505
Non related (Family object, Household, Natural things etc)	0.275	-0.56067	-0.15418

$H(x) = - \sum P(X_i) LogP(X_i) =0.79122$

0.79122 is the probability for Name deception

Stage-1 Result: The Randomness Entropy based on Fuzzified Datum towards Name deception has its value as 0.79122 (HIGH IMPACT)

ID Deception

In Stage-2-While we analyse the datum based on fuzzy methodologies ,we categorize the subjects into the following criterias,

ID related with name	= 5/40
ID associated with year (DOB/ mo)	= 3/40
ID associated with current year	= 1/40
ID with special symbol	= 7/40
ID with combination of (A to G)	= 12/40
ID based on server recommendation (non availability)	= 3/40
ID with numbers (Age/carno/fancy)	= 4/40
Irrelevant ID	= 5/40

Now calculate the values using the Random Entropy we obtain the following,

TABLE III
Entropy Values for ID Deception

X_i	$P(X_i)$	$LogP(X_i)$	$P(X_i) * Log(P(X_i))$
ID related with name	.125	-0.90309	-0.11289
ID associated with year (DOB/ mo)	.075	-1.12494	-0.08437
ID associated with current year	.025	-1.60206	-0.04005
ID with special symbol	.175	-0.75696	-0.13247
ID with combination of (A to G)	.3	-0.52288	-0.15686
ID based on server recommendation (non availability)	.025	-1.60206	-0.04005
ID with numbers (Age/carno/fancy)	.025	-1.60206	-0.04005
Irrelevant ID	.625	-0.20412	-0.12757

$$H(x) = - \sum P(X_i)\, LogP(X_i)= 0.73432$$

Stage-2- Result: The Randomness Entropy based on Fuzzified Datum towards ID deception has its value as 0.73432 (HIGH IMPACT)

Alternate Email Id

In Stage-3The fuzzy classification based on probable ness we obtain the following sets of variations,

Provided with true mail	- 10/40
Provided with frequent usage	- 3/40
Provided with rare usage (used for deception)	- 7/40
Provided with false mail	- 5/40
Provided with irrelevant	- 3/40
Provided with known members in the circle	- 1/40
Provided with familiar mail –IDs	- 1/40
Leave the option	- 25/40

Now calculate the values using the Random Entropy we obtain the following

TABLE IV
Entropy Values for Alt-Email ID Deception

X_i	$P(X_i)$	$LogP(X_i)$	$P(X_i) * Log(P(X_i))$
Provided with true mail	0.25	-0.60206	-0.15051
Provided with frequent usage	0.075	-1.12494	-0.08437
Provided with rare usage (used for deception)	0.175	-0.75696	-0.13247
Provided with false mail	0.125	-0.90309	-0.11289
Provided with irrelevant	0.075	-1.12494	-0.08437
Provided with known members in the circle	0.025	-1.60206	-0.04005
Provided with familiar mail –IDs	0.025	-1.60206	-0.04005
Leave the option	0.625	-0.20412	-0.12757

$$H(x) = - \sum P(X_i)\, LogP(X_i)= 0.77229$$

Stage-3 Result: The Randomness Entropy based on Fuzzified Datum towards Alternate ID deception has its value as 0.77229(HIGH IMPACT)

DOB

In Stage-4 we perform the fuzzy allocation towards the datum of the subjects 8-Byte value of Date of Birth we segmented the following sectors,

Original	- 2/40
Day change	- 11/40
Month change	- 8/40
Year change	- 3/40
Day & Month change	- 2/40
Day & year change	- 2/40
Month & year change	- 3/40
Day, month & year change	- 9/40
Trivial	- 3
Porpose for 18+	- 6

Now calculate the values using the Random Entropy we obtain the following,

TABLE V
Entropy Values for DOB Deception

X_i	$P(X_i)$	$LogP(X_i)$	$P(X_i)* Log(P(X_i))$
Day change	0.275	-0.56067	-0.15418
Month change	0.2	-0.69897	-0.13979
Year change	0.075	-1.12494	-0.08437
Day & Month change	0.05	-1.30103	-0.06505
Day & year change	0.05	-1.30103	-0.06505
Month & year change	0.075	-1.12494	-0.08437
Day, month & year change	0.225	-0.64782	-0.14576
Trivial	0.075	-1.12494	-0.08437
Porpose for 18+	0.15	-0.82391	-0.12359

$$H(x) = - \sum P(X_i)\, LogP(X_i)= 0.94654$$

Stage-4 Result: The Randomness Entropy based on Fuzzified Datum towards Date Of Birth deception has its value as 0.94654 (VERY HIGH IMPACT)

Gender

In Stage-5 the fuzzy classification based on probable ness we obtain the following sets of variations

Original	- 25/40
Falsified suspection	- 15/40
Related to Name	- 10/40
No Relation	- 5/40

Now calculate the values using the Random Entropy we obtain the following,

TABLE VI
Entropy Values for Gender Deception

X_i	$P(X_i)$	$LogP(X_i)$	$P(X_i)*Log(PXi))$
Original	0.625	-0.20412	-0.12757
Related to Name	0.25	-0.60206	-0.15051
No Relation	0.125	-0.90309	-0.11289

$$H(x) = - \sum P(X_i)\, LogP(X_i)= 0.2634$$

Stage-5 Result: The Randomness Entropy based on Fuzzified Datum towards Gender deception has its value as 0.2634(BELOW AVERAGE IMPACT)

Education

In Stage-6 we perform the fuzzy allocation towards the datum of the subjects Education or Qualification Datum we obtain the following,

a) Original	- 5/40
b) Falsified suspection	- 25/40
Related to DOB	- 15/40
Non related to DOB	- 10/40
c) Leave the field	- 10/40

Now calculate the values using the Random Entropy we obtain the following,

TABLE VII
Entropy Values for Education Deception

X_i	$P(X_i)$	$LogP(X_i)$	$P(X_i)$* $Log(P(X_i))$
Related to DOB	0.375	-0.42597	-0.15974
Non related to DOB	0.25	-0.60206	-0.15051
Leave the field	0.25	-0.60206	-0.15051

$H(x) = - \sum P(X_i) LogP(X_i)= 0.4676$

Stage-6 Result: The Randomness Entropy based on Fuzzified Datum towards Education value deception has its value as 0.4676(AVERAGE IMPACT)

Occupation

In Stage-7 we perform the fuzzy allocation towards the datum of the subjects job descriptions we segmented the following sectors,

a) Original	- 1/40
b) Falsified	- 29/40
Related to course (may be infuture)	- 20/40
Non related to course	- 9 /40
c) Leave the field	-11/40

Now calculate the values using the Random Entropy we obtain the following,

TABLE VIII
Entropy Values for Occupation Deception

X_i	$P(X_i)$	$LogP(X_i)$	$P(X_i)$* $Log(P(X_i))$
Related to course	0.5	-0.30103	-0.15051
Nonrelated to course	0.225	-0.64782	-0.14576
Leave the field	0.275	-0.56067	-0.15418

$H(x) = - \sum P(X_i) LogP(X_i)= 0.449$

Stage-7 Result: The Randomness Entropy based on Fuzzified Datum towards Occupation value deception has its value as 0.449(AVERAGE IMPACT)

Choosing the options

In Stage-8 we perform the fuzzy allocation towards the datum of the subjects Options selecting strategies we segmented the following sectors,

Send me offers & Send me updates – Default	- 28/40
Check point removed	- 12 /40
For 1	- 8/40
For 2	- 4/40

Now calculate the values using the Random Entropy we obtain the following,

TABLE IX
Entropy Values for Offers Deception

X_i	$P(X_i)$	$LogP(X_i)$	$P(X_i)$ * $Log(P(X_i))$
Send me offers & Send me updates - Default	0.7	-0.1549	-0.10843
Check point removed for first	0.2	-0.69897	-0.13979
Check point removed for second	0.1	-1	-0.1

$H(x) = - \sum P(X_i) LogP(X_i)= 0.347$

Stage-8 Result: The Randomness Entropy based on Fuzzified Datum towards Choosing the Options value deception has its value as 0.347(BELOW AVERAGE).

I own

In Stage-9 we perform the fuzzy allocation towards the datum of the subjects Component ownership we segmented the following sectors,

C H CC C

Null	- 20/40
Any one	- 3/40
Any two	- 6/40
Any three	- 1/40
All	- 10/40

Now calculate the values using the Random Entropy we obtain the following,

TABLE X
Entropy Values for Ownership Deception

X_i	$P(X_i)$	$LogP(X_i)$	$P(X_i)$ * $LogP(X_i)$
Null	0.5	-0.30103	-0.15051
Any one	0.075	-1.12494	-0.08437
Any two	0.15	-0.82391	-0.12359
Any three	0.025	-1.60206	-0.04005
All	0.25	-0.60206	-0.15051

$H(x) = - \sum P(X_i) LogP(X_i)= 0.54904$

Stage-9 Result: The Randomness Entropy based on Fuzzified Datum towards I Own Value deception has its value as 0.54904(AVERAGE)

We performed some exterior characteristics for accessing the Mail server is as Follows,

Verifying with the ownership 30/40 is falsified response
=> 75%

Directly click goto inbox
(Time gap is low /access fastly) => 12/40

VI. RESULT & DISCUSSION

Comparing the result by calling another set of 40 students with the same approach, the results are shown below:

TABLE XI
Entropy Comparison Values for Deception

Factors	Set1- $H(x_a)$	Set2- $H(x_a)$	Diff.	Diff. %
Name	0.79	0.68	0.11	11
ID	0.73	0.72	0.01	01

Alternate	0.77	0.81	0.04	04
DOB	0.94	0.91	0.03	03
Gender	0.26	0.25	0.01	01
Education	0.46	0.52	0.06	06
Occupation	0.44	0.42	0.02	02
Option	0.34	0.32	0.02	02
Own’s	0.55	0.61	0.06	06

So our proposed research model insists the selected fields for suspecting and capture the Deception Detection as early as possible.

The Random values with Fuzzy descriptions for concentrating on the components of Mail-ID-creation are evaluated in our proposed model. The Impact calculations are made for our self references which points as follows,

Table XII
Impact Assignment

Slno	**Range of evaluated values**	**Impact nature**
1	0.95 to 1.00	Peak
2	0.81 to0.949	Very high
3	0.71 to 0.800	High
4	0.61 to 0.709	Above average
5	0.41 to 0.609	Average
6	0.25 to 0.409	Below average
7	0.1 to 0.249	Low
8	0.0 to 0.099	Least

VII. CONCLUSION

We achieve somewhat better results when combine more than on of individual component implementation of Fuzzy, Randomness and Uncertainty rather than with its individuality.

Detecting the deception in an online computer mediated communication in a tedious process. Here the fuzziness in identifying false information can be reduced to focusing on specific fields using fuzzy classifications and categorization. The randomness entropy strategy helps us to detect the variation of deception in information theory to detect the probable fields with underlying uncertainty. In near future we will combine fuzzy logic, uncertainty and randomization in the area of deception detection.

Applying the Randomization techniques and Fuzzylogic towards identifying the deception is a critical process of complexity,but the results are more effective when compare it with implementing each phase individually.

Many automated systems are now required for deception detection handling in an optimized manner,we moreover try to implement the concept of artificial intelligence,neurofuzzy and Genitic algorithm combinations for detecting Deceptions in our recent data communication strategies and components.

REFERENCES

[1] Steve Woznaik, Kevin D.Mitnick, Willaim L.Simon,2002. “*The art of deception: controlling the human element of security”*. Wiley; 1 edition.

[2] Zuckerman, M.,DePaulo, B.M. and Rosenthal, R.”*Verbal and Nonverbal Communication of Deception*”.In L.Berkowitz(Ed)(1981)

[3] Burgoon, J.K., and Qin,T. “*The Dynamic Nature of Deceptive Verbal Communication*”. Journal of Language and Social Psychology, 2006, vol25(1), 1-22.

[4] Bond,c.,F. “*A world of lies: the global deception research team*”, Journal of Cross-culture Psychology, 2006, Vol.37(1), 60-74.

[5] Pennebaker,J.W,Mehl,M.R.&Niederhoffer,K. ”*Psychological aspects of natural language use: our words, ourselves*”. Annual Review of Psychology, 2003, 54,547-577

[6] Whissell,C., Fournier,M.,Pelland,R., Weir, D.,& Makaree,K. ”*A comparison of Classfiifcation methods for predicting deception in computer-mediated communication*”. Journal of Management Information systems, 2004,20(4),139-165.

[7] http://register.rediff.com/register/register.php?FormName=user_details

Authors Profile

Mr.S.Rajkumar completed his M.E–CSE at Sathyabama University, Chennai and currently doing his Ph.D in the area of Computational Science. He is a Research Scholar of Bharathiar University and working as a HOD/CSE at NIET Coimbatore.

Ms.V.Narayani completed her M.C.A in M.S University, Tirunelveli and M.Phil in Mother Teresa University, Kodaikanal. She submitted her Ph.D thesis in the area of Data Mining.

Dr. S. P. Victor earned his M.C.A. degree from Bharathidasan University, Tiruchirappalli. The M. S. University, Tirunelveli, awarded him Ph.D. degree in Computer Science for his research in Parallel Algorithms. He is the Head of the Department of Computer Science, and the Director of the Computer Science Research centre, St. Xavier’s college (Autonomous), Palayamkottai, Tirunelveli. The M.S. University, Tirunelveli and Bharathiar University, Coimbatore has recognized him as a research guide. He has published research papers in international, national journals and conference proceedings. He has organized Conferences and Seminars at national and state level.