Preserving Privacy and Optimizing Neural Network Classification by using a Mix of Soft Computing Techniques

Majid Bashir Malik
Dept. of Computer Sciences, BGSB University, Rajouri, J & K, India

M. Asger
School of Mathematical Sciences & Engg., BGSB University, Rajouri, J & K, India

Rashid Ali
Department of Computer Engineering, Aligarh Muslim University, Aligarh, India

Tasleem Arif
Department of Information Technology, BGSB University, Rajouri, J & K, India

ABSTRACT
An Artificial Neural Network (ANN) commonly called as Neural Network (NN) is a mathematical tool that is used in data mining for the analysis of data. Inherently human brain has its own limitations, the inability to handle large volumes of data being one of them. NNs with its power to handle large volumes of data along with imitation of human brain provides good solutions where the size of data is too large, beyond human interpretation capacity or beyond the capacity of conventional computational methods. NNs find applicability in healthcare among other fields like business, genetics, bioinformatics, pharmaceuticals, etc. NNs are mostly not used in data mining due to the reason that it takes lots of time for training of the networks. But as far as the utility of NNs in data mining is concerned, it is a valuable technique. Data mining in general and NNs in particular are being used in various areas where privacy of individuals is at stake. This study attempts to optimize the results of NN classification on one hand and address the problem of training time consumption on the other along with preserving privacy of the stakeholders.

Keywords
Privacy, NNs, fuzzification, rough sets, confusion matrix, fuzzy membership function

1. INTRODUCTION
Data mining is an automated knowledge discovery mechanism from huge repositories of data [1]. The focus is to gain insight into large volumes of data for a better life. The purpose of data mining is to extract or uncover previously unknown or hidden relationships and patterns from large and complex datasets using automated or semi-automated exploratory data analysis tools [3]. A large number of statistical and computational tools and techniques are used to uncover useful patterns that may be hidden within large set of databases [4-6]. Strong knowledge base with expert analytical skills and the knowledge of the domain under consideration are the key ingredients for successful and purposeful use of data mining techniques for extraction and analysis of hidden patterns and trends. The models so developed support system analyst to extract new, unknown and useful observations from the data.

Lots of data mining applications have been and are being developed in the fields like genetics, business, pharmaceuticals, research, medical diagnosis, weather forecasting, fraud detection, defense, marketing, agriculture, social networks, banking, telecommunication, anti-terrorism, education, etc. [1]. This has led to increased revenue, reduced cost, better decision support system, intrusion detection, accurate predictions and improved market place responsiveness and awareness, etc. Although data mining offers a number of benefits but there are a large number of challenges that miners face in achieving the success like non-standardization of data, distribution of data (vertical and horizontal), missing values, involvement of muti-parties/stake holders, privacy, changing data, outliers, and even handling expired data for mining purpose [2].

1.1 Soft Computing
Soft computing addresses many of such challenges [8]. Soft computing in itself is a collection of synergistic mechanism that is capable of handling real-life ambiguous problems providing subtle and flexible data processing solutions [9]. Soft computing tries to find solutions in a way the human brain does. Soft Computing while imitating human brain is capable of handling challenges that data mining task has to face in developing acceptable solutions like uncertainty in data, imprecision of the data values, missing values, ambiguity in data, and partial truth [1]. The data mining solutions developed using soft computing techniques possess high Machine Intelligence Quotient (MIQ). They are robust, tractable and are efficient as they are low at cost in terms of time and space complexity. The most widely used soft computing techniques in developing data mining solutions are Neural Networks, Rough Sets, Genetic Algorithms and Fuzzy logic [2]. Neural networks specialize in rule generation and classification. Rough sets handle indiscernibility among objects of a set because of uncertainty in the data values. Rough sets are also used in discovering dependencies and redundancies during classification of data. Genetic algorithms are robust and adaptive global search methods for very large search space and are being successfully implemented in search and optimization processes. Fuzzy logic is a natural framework for handling uncertainty and transmission at various stages even if the data is imprecise [8-9]. The whole set of soft computing techniques work in a cooperative manner rather than in a competitive manner for addressing problems in the domain of challenges of data mining.

1.2 Artificial Neural Networks
Artificial Neural Networks (ANN) also known as Neural Networks is an emulation of biological neurons. It is basically a computational or mathematical model based on biological neural network, thereby imitating human brain. Inherently humans have a limitation when it comes analyze large volumes of data. Neural networks provide solutions to handle

large volumes of data and at the same time the solutions are reached in a way human brain does. Neural networks are used in case where developing solution is either very complex or the possiblility of developing an algorithmic solution is very low. This makes Neural Networks useful in various applications like pattern recognition, prediction, classification, data compression, decision making, and optimization [7, 10].

1.3 Neural networks in medical diagnosis

Neural networks are successfully being applied in various applications of medical diagnosis like image analysis, biochemical analysis, medicines and drug development etc [10]. A wide variety of applications in medical sciences have been developed like tumor detection in ultra sonograms, classification of chest x-rays, detection of calcification in mammograms, classification of cervical cancer using pap smears, analyzing breast cancer and cardiology etc [10-20]. So it can easily be deduced that neural networks can be used in predictions of various stages of different diseases [21].

2. MOTIVATION

A lot of work has been and is being done in classifications, predictions and analysis using neural networks. And lot many successful applications have been developed in a number of critical areas where accuracy of the results is must. The accuracy of the results obviously depends on the quality of the data that is being used for training the neural networks and the quality of the algorithms of the neural networks. The quality in terms of the truthfulness of the data is dependent on the confidence that the researchers can build in the minds of the people to whom the data originally belongs i.e. data owners. The data owners mostly try to find out way to hide details or report false details as they are apprehensive of privacy breaches. As far as the quality of the algorithms of neural network is considered the constraint is the time taken to train the network. The amount of time taken to train the neural networks is too large and needs improvement.

In our study we have tried to target these two problems:

1. Privacy issue and
2. Time taken to train the Neural Networks.

Privacy in data mining has emerged as an important concern over a period of time. Stringent laws have been formulated for the same. While we try to preserve privacy, the results of the data mining get affected adversely. In order to preserve the privacy without sacrificing the results of the data mining a balance needs to be maintained where the results of the data mining are not affected and at the same time privacy is also preserved. And a lot of work has been done on the same issue. Various techniques like Anonymization, Condensation, Cryptography, Perturbation, Randomized response etc. have been proposed but they mostly suffer information loss or fail to defend against different type of privacy attacks [22-29]. Some solutions are suffer from high computational cost or they are simply application specific [22-29].

In the process of anonymization to protect privacy the identity of the individuals is kept hidden by removing identifier fields. But still the privacy is at stake when such data, particularly the quasi-identifiers are linked to publicly available data [24]. The data in anonymization is dependable but it suffers due to background knowledge and homogeneity attack. Excessive privacy preserving in anonymization process leads to heavy information loss [25].

Original values are replaced by synthetic values in case of perturbation technique to protect privacy in data mining [23]. The data so generated can defend against linkage and homogeneity attacks but reconstruction of original values is not possible and new and different set of distribution algorithms are required to de developed every time for various data mining applications like classification, clustering or association rule mining [25].

In randomized response the data is scrambled, it treats all the records equal without considering local density [10], due which the outlier records become prone to adversarial attacks [26]. Moreover adding too much of noise leads to the decrease in the utility of the data in data mining task.

Synthetic data is generated from the statistics of the clusters such that each record holds a position in the group having same anonymity level [27]. Although this approach is immune to attacks but suffers heavy information loss [25].

Cryptographic techniques are widely used for transformation of data in case where more than one party is involved in data mining task, so as to avoid disclosure of information. But the privacy of individual in this approach fails to deliver when the number of parties is more than few[25]. Moreover very little work has been done in case of malicious models [10].

A good amount of work has been done even in preserving privacy in neural network applications but the work is problem specific. Although neural networks are appropriate for wide range of data mining problems yet they are not commonly used for such tasks because of two reasons- trained NNs are usually not comprehensible and the learning method of NN is too slow [30].

3. EXPERIMENTAL RESULTS

In our study, we have tried to preserve the privacy of individual without affecting the results of the neural network predictions and at the same time, the time taken for the training of neural networks has been reduced. An application has been developed in [7] where NNs have been trained using data related to potential diabetic patients. On the basis of certain parameters like DoB, Sex, Smoking, Drinking, Thirst, Urination, Height, Weight, Fatigue etc, the application predicts the potential diabetic patient. The same data has been used here in our study.

The algorithmic solution/model that has been proposed in the study for the same application where in we have tried to preserve the privacy and optimize the NN training is as:

1. The data regarding potential diabetic patients has been requested from various diagnostic labs.
2. First level privacy preserving:

   Identifier fields, like name, address, which can be used to identify the individual have been removed.
3. Optimizing data for NN training:

   Johnson's reducer algorithm has been used for optimizing data by reducing the data. The non participating fields in NN training have been removed.
4. Second level privacy preserving step:
   a. Some fields have been generalized like DoB.
   b. For the purpose of privacy preservation fuzzification of the all fields has been done (except those which take boolean values) using S-shaped fuzzification membership function.
5. Train the Neural networks using multiple back-

propagation.

6. Test the results.
7. In case the training is not satisfactory the NNs are further re-trained.

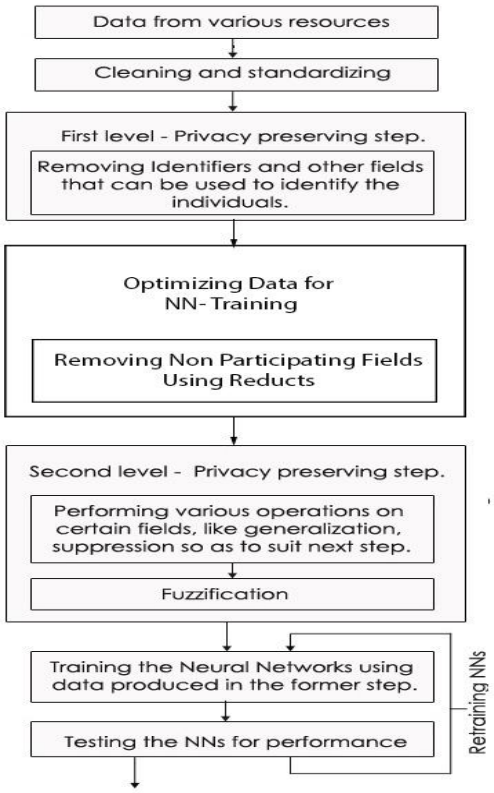


Fig. 1: Algorithm/Model for preserving privacy and optimizing NN Classification

For the purpose of study we have trained NNs using three sets of data:

1. Normal data: The data as received from various sources with some changes like cleaning and standardizing.
2. Fuzzified data: All the fields, except those, having boolean data have been fuzzified using S-shaped fuzzy membership function for privacy preservation. The equation representing S- shaped fuzzy membership function is as:

$$f(x;a,b)\begin{cases} 0, & x \le a \\ 2\left(\frac{x-a}{b-a}\right)^2, & a \le x \le \frac{a+b}{2} \\ 1-2\left(\frac{x-b}{b-a}\right)^2, & \frac{a+b}{2} \le x \le b \\ 1, & x \ge b \end{cases}$$

Eq. 1: S-shaped fuzzy membership function

Where

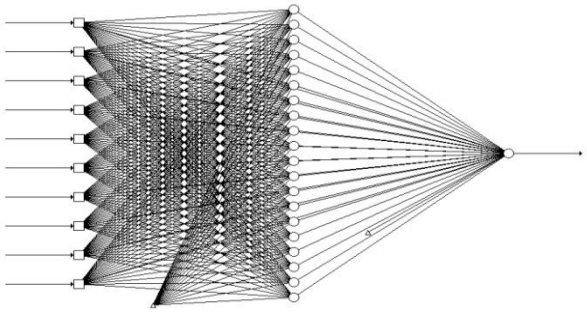
'x' is the value to be fuzzified

'a' and 'b' are lower and upper limits of data.

3. Optimized data: Using Johnson's reducer algorithm non-participating fields have been removed from the dataset. Rosetta, a roughest based toolkit has been used for the same. The number of fields, 10, as in case of Normal and fuzzified data have been reduced to 4 along with one decision field. The data so produced has been generalized for some fields and then fuzzified for privacy preserving using S-shaped fuzzy membership function as represented in Eq. 1.

Now we have three sets of data: Normal/Original data, Fuzzified data and Optimized data. Moreover the data in all the cases has been divided into four sets for the sake of cross validation. The neural networks have been trained by 4 sets of cross validated in all of three types of cases i.e. Normal/Original data, Fuzzified data and Optimized data, using same set of parameters, NN architecture and topology as represented in the below diagram. The data in all the cases has been divided into four sets for 4-level cross validation to establish the results as true as possible.

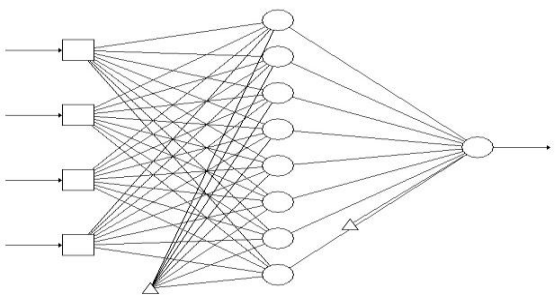
Here in the study standard back propagation based multilayer perceptron (MLP) architecture of ANN has been used for training all the sets of data so that the results after training the NN by all the forms of data can be compared. This architecture followed here is often used for ANN in medical research [7]. It's a directed graph of multiple layers of artificial neurons where each layer is connected to next layer. Back propagation is a gradient descent technique that minimizes the error criteria and it is a simple way to determine the error criteria and to determine the error values in hidden layers. The hidden layer allows ANNs to develop its representation of input–output mapping. In case of back propagation, the error data at the output layer is propagated back to earlier layers, thereby allowing incoming weights to these layers be updated and adjusted such that the error between the input and desired output is as least as possible [7]. The topology employed is 10 inputs, one hidden layer and one output in case of Normal and Fuzzified data and is represented as:



Input Layer Hidden Layer Output layer

Fig. 2: NN Topology followed in case of NNs trained by Normal and Fuzzified data

The topology employed is 4 inputs, one hidden layer and one output in case of optimized data as this data has only four fields as input attributes. It is represented as:



Input Layer Hidden Layer Output layer

Fig. 3: NN Topology followed in case of NNs trained by optimized data

4. PERFORMANCE OF ANNs

The NNs have been trained by all the sets of cross validated data for 1000000 epochs using multiple back propagation (MBP) tool version 2.2.4. The number of records for training purpose is 291 and for testing purpose is 97, for all sets of data. After training and testing the neural networks the results were drawn and compared for all the sets of data i.e. normal, fuzzified and optimized data. Average of all sets or cross validated results in all of the three cases is as:

Table 1: Average results of cross validated data

	Average results of Cross Validated data		
	Normal data	**Fuzzified data**	**Optimized data**
RMS Error Training	0.08186559	0.05442622	0.06861812
RMS Error Testing	0.20095253	0.19822708	0.18473622
Wrong Predictions-Training	8.25	4.25	6.25
Wrong Predictions-Testing	18	16.50	13.75
Time taken for training NNs	03:08:07	02:06:18	00:34:46

The Root Mean Square (RMS) error for training, in case of Fuzzified data seems to be better as compared to Normal and optimized data as far as the figures generated during the experimentation are concerned. But when we see the results of RMS Error for training and testing, it is clearly visible that in case of testing the RMS Error is least for optimized data as compared to normal and fuzzified data. Thereby it can be deduced that the results in case of optimized data do not suffer from overfitting problem. Basically overfitting is problem where the results of training are good but in case of testing, results are not good. The accuracy rate of the ANNs', root mean square error for testing and the time taken to train the NNs are better in case of the NNs when trained by fuzzified data as compared to NNs trained by normal data. The results are even better when the NNs are trained by optimized data.
During testing, out of 97, only 13.75 instances were predicted wrongly in case of optimized data where as in case of normal and fuzzified data it is 18 and 16.50 respectively. The accuracy rate of NNs trained by optimized data is highest i.e. 85.82% and that of NNs trained by normal data and fuzzified data is 81.44% and 82.98%. The average time taken to train the NNs is least in case of NNs trained by optimized data i.e. 34 minutes and 46 seconds whereas in case of NNs trained by normal data it is 3 hours, 8 minutes and 7 seconds and in case of NNs trained by fuzzified data it is 2 hours, 6 minutes and 18 seconds.

5. CONFUSION MATRIX

A confusion Matrix (proposed by Kohavi and Provost in 1998) represents information in a tabular form regarding predicted and actual classification calculated by the system. This information in the matrix is used to evaluate the performance of systems. The following table shows the confusion matrix for a two class classifier.

Table 2: Confusion Matrix for a two classifier

		Predicted	
		Negative	Positive
Actual	Negative	a	b
	Positive	c	d

a: number of correct predictions that an instance is negative,

b: number of incorrect predictions that an instance is positive,

c:number of incorrect of predictions that an instance negative, and

d: number of correct predictions that an instance is positive.

The parameters calculated on the basis of the results in the above table:

1. Accuracy (AC) is the proportion of the total number of predictions that are correct and is calculated by the equation: AC= (a + d) / (a + b + c + d).
2. True positive rate (TP) is the proportion of positive cases and is calculated by the equation: TP= d / (c + d).
3. False positive rate (FP) is the proportion of negative cases that are incorrectly classified as positive: FP= b / (a + b).
4. True negative rate (TN) is the proportion of negative cases that are classified correctly and is represented by: TN= a / (a + b).
5. False negative rate (FN) is the proportion of positive cases that are incorrectly classified as negative and is represented by: FN= c / (c + d).
6. Precision (P) is the proportion of predicted positive cases that are correct and is represented as: P= d / (b + d).

Table 3: Confusion matrix Training: Normal data (Average values from four sets of data)

		Predicted	
		Negative	**Positive**
Actual	**Negative**	**139**	**5.75**
	Positive	**2.5**	**143.75**

AC= 0.9716495

TP= 0.9829060

FP= 0.0397237

TN= 0.9602764

FN= 0.0170940

P= 0.9615385

Table 4: Confusion matrix Testing: Normal data (Average values from four sets of data)

		Predicted	
		Negative	**Positive**
Actual	**Negative**	**39.25**	**9.25**
	Positive	**8.25**	**40.25**

AC= 0.8195876

TP= 0.8298969

FP= 0.1907216

TN= 0.8092784

FN= 0.1701031

P= 0.8131313

Table 5: Confusion matrix Training: Fuzzified data (Average values from four sets of data)

		Predicted	
		Negative	**Positive**
Actual	**Negative**	**142**	**3**
	Positive	**1.25**	**144.75**

AC= 0.9853952

TP= 0.9914384

FP= 0.0206897

TN= 0.9793103

FN= 0.0085616

P= 0.9796954

Table 6: Confusion matrix Testing: Fuzzified data (Average values from four sets of data)

		Predicted	
		Negative	**Positive**
Actual	**Negative**	**39.5**	**8.75**
	Positive	**7.75**	**41**

AC= 0.8298969

TP= 0.8410256

FP= 0.1086960

TN= 0.8186529

FN= 0.1589744

P= 0.8241206

Table 7: Confusion matrix Training: Optimized data (Average values from four sets of data)

		Predicted	
		Negative	**Positive**
Actual	**Negative**	**140.25**	**4.5**
	Positive	**1.75**	**144.5**

AC= 0.9785223

TP= 0.9880342

FP= 0.0310881

TN= 0.9689119

FN= 0.0119658

P= 0.9697987

Table 8: Confusion matrix Testing: Optimized data (Average values from four sets of data)

		Predicted	
		Negative	**Positive**
Actual	**Negative**	**41**	**7.25**
	Positive	**6.75**	**42**

AC= 0.8556701

TP= 0.8615385

FP= 0.1502591

TN= 0.8497409

FN= 0.1384615

P= 0.8527919

Table 9: Comparative results of normal, fuzzified and optimized training and normal, fuzzified and optimized testing on the basis of various parameters calculated through Confusion Matrix

	Data Training			**Data Testing**		
	Normal	**Fuzzified**	**Optimized**	**Normal**	**Fuzzified**	**Optimized**
AC	0.9716495	0.9853952	0.9785223	0.8195876	0.8298969	0.8556701
TP	0.9829060	0.9914384	0.9880342	0.8298969	0.8410256	0.8615385
FP	0.0397237	0.0206897	0.0310881	0.1907216	0.1813472	0.1502591
TN	0.9602764	0.9792103	0.9689119	0.8092784	0.8186529	0.8410257
FN	0.0170940	0.0085616	0.0119658	0.1701031	0.1589744	0.1384615
P	0.9615385	0.9796954	0.9697987	0.8131313	0.8241206	0.8527919

The results of the confusion matrices shown in the table 9 are calculated as average of 4-stage cross validation for normal/original, fuzzified and optimized data. Table 9 shows that the values of AC, TP, FP, TN, FN and P are better in case of NNs trained by optimized data.

6. BAR CHART REPRESENTATIONS OF THE RESULTS SHOWN IN THE TABLE 8

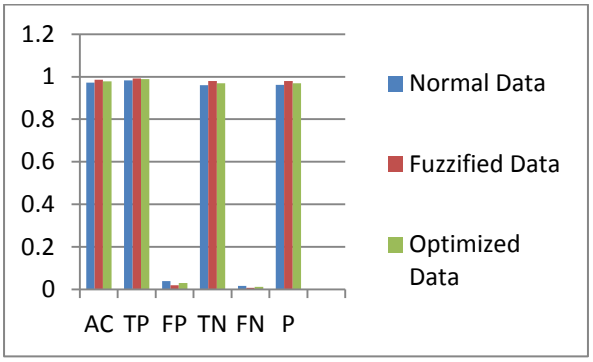


Fig.: 4: Bar Chart representation of different parameters of Normal Data Training, Fuzzified data training and optimized data training.

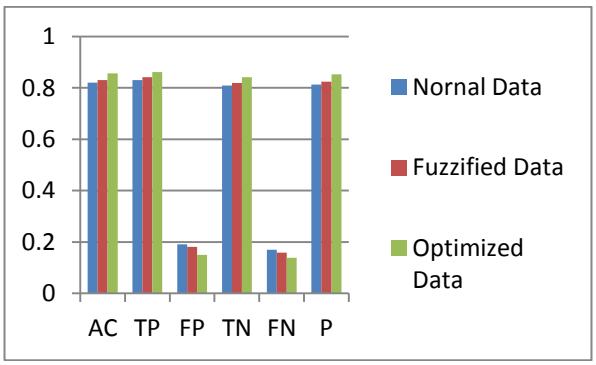


Fig.: 5: Bar Chart representation of different parameters of Normal Data Testing, Fuzzified data testing and optimized data testing

7. CONCLUSION

Privacy preserving in data mining has emerged as an important research field and of great interest over a period of time. Lots of research is being targeted towards preserving the privacy of the stake holders. And at the same time the results of the mining need not to be compromised. But most of the solutions suffer information loss or complexity. As far as Neural networks is concerned it has established itself as one of the successful techniques for predictions, but the time it takes for training has also emerged as a problem. In our study we have preserved the privacy by generalizing and fuzzifying the data and at the same time the results of the NNs trained by optimized data are better as compared to NNs trained by normal data. Moreover the time taken to train the NNs has been reduced drastically by using rough sets as shown in the table 1.

8. REFERENCES

[1] M. B. Malik, M. A. Ghazi, R. Ali, "*Privacy preserving data mining techniques: Current Scenario and future prospects"*, Third International Conference on Computer and Communication Technology, 2012.

[2] M. B. Malik, M. A. Ghazi, R. Ali, T. Arif, "*Privacy Preserving Data Mining Using Fuzzy Based Approach"*, COMMUNE, International Conference on Advances in Computers, Communication and Electronic Engineering, March, 2015.

[3] Zhihua, X., "Statistics and Data Mining", Department of Information System and Computer Science, National University of Singapore, 1998.

[4] Tsantis L & Castellani J, "Enhancing Learning Environment Solution-based knowledge Discovery Tools: Forecasting for Self-perpetuating Systematic Reform", JSET Journal, 2001.

[5] Luan, J, "Data Mining Application in Higher education", SPSS Executive Report. Retrieved from http://www.crisp-dm.org/CRISPWP.pdf, 2002.

[6] A. Ahmad, and L. Dey, "A k-mean clustering algorithm for mixed numeric and categorical data", Data & Knowledge Engineering, vol. 63, no. 2, pp. 503-527, 2007.

[7] Abid Sarvar, Vinod Sharma, "Comparative analysis of machine learning techniques in prognosis of type II diabetes", AI and Society, Journal of Knowledge, Culture and Communication, Vol. 29, No. I, 2014.

[8] Sushmita Mitra, Sankar K. Pal and Pabitra Mitra, 2002, "Data Mining in Soft Computing Framework: A Survey", *IEEE Transactions on Neural Networks*, Vol 13, No. 1, 2002.

[9] L. A. Zadeh, "Fuzzy Logic, Neural Networks, and Soft Computing", *Communications of the ACM*, vol. 37, no. 3, pp: 77-84, 1994.

[10] M B Malik, M Asger, R Ali, A Sarvar, *"A Model for Privacy Preserving in Data Mining using Soft Computing Techniques",* INDIACom 2015, International Conference on Computing for Sustainable Global Development, PP 181-186, March 2015.

[11] A. Kusiak, K.H. Kernstine, J.A. Kern, K A. McLaughlin and T.L. Tseng, (2000) "Data mining: Medical and Engineering Case Studies", Proceedings of the Industrial Engineering Research Conference, Cleveland, Ohio, May21-23, pp.1-7.

[12] H. B. Burke, (1994) "Artificial neural networks for cancer research: Outcome prediction," *Sem.Surg. Oncol.*, vol. 10, pp. 73–79.

[13] Siri Krishan Wasan1, Vasudha Bhatnagar2 and Harleen Kaur, (2006)" The impact of Data Mining Techniques on Medical Diagnostics", Data Science Journal, Volume 5, 119-126.

[14] Scales, R., & Embrechts, M., (2002) "Computational Intelligence Techniques for Medical Diagnostic", Proceedings of Walter Lincoln Hawkins, Graduate Research Conference from the World Wide Web: http://www.cs.rpi.edu/~bivenj/MRC/proceedings/papers/ researchpaper.pdf

[15] S. M. Kamruzzaman, Md. Monirul Islam, (2006) "An Algorithm to Extract Rules from Artificial Neural Networks for Medical Diagnosis Problems", International Journal of Information Technology, Vol. 12 No. 8.

[16] Hasan Temurtas, Nejat Yumusak, Feyzullah Temurtas, (2009)" A comparative study on diabetes disease diagnosis using neural networks", Expert Systems with Applications: An International Journal , Volume 36 Issue 4.

[17] D Gil, M Johnsson, JM Garcia Chamizo, (2009) , "Application of artificial neural networks in the diagnosis

of urological dysfunctions", Expert Systems with Applications Volume 36, Issue 3, Part 2, Pages 5754-5760, Elsevier

[18] R. Dybowski and V. Gant, (2007), "Clinical Applications of Artificial Neural Networks", Cambridge University Press.

[19] O. Er, N. Yumusak and F. Temurtas, (2010) "Chest disease diagnosis using artificial neural networks", Expert Systems with Applications, Vol.37, No.12, pp. 7648-7655.

[20] S. Moein, S. A. Monadjemi and P. Moallem, (2009) "A Novel Fuzzy-Neural Based Medical Diagnosis System", International Journal of Biological & Medical Sciences, Vol.4, No.3, pp. 146-150.

[21] Dr. K. Usha Rani, **"**Analysis of heart diseases dataset using Neural Network", International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.1, No.5, September 2011 DOI: 10.5121/ijdkp.2011.1501 1.

[22] Benjamin C M Fung, Ke Wang, Rui Chen, Philip S Yu, "Privacy Preserving Data Publishing: A Survey of recent developments", ACM Computing Surveys, Vol. 42, No. 4, Article 14, June 2010.

[23] B. Karthikeyan, G Manikandan, V. Vathiyanathan, "A fuzzy based approach for privacy preserving clustering", Journal of Theoretical and Applied Information technology", Vol. 32, No. 2, Oct. 2011

[24] Sweeney L, "Achieving k-Anonymity privacy protection using generalization and suppression" International journal of Uncertainty, Fuzziness and Knowledge based systems, 10(5), 571-588, 2002.

[25] Gayatri Nayak, Swagatika Devi, "A survey on Privacy Preserving Data Mining: Approaches and Techniques", International Journal of Engineering Science and Technology, Vol. 3 No. 3, 2127-2133, 2011.

[26] Y. Lindell and B. Pinkas, "Privacy Preserving Data Mining", Journal of Cryptology, 15(3), pp.36-54, 2000.

[27] Aggarwal C, Philip S Yu, "A condensation approach to privacy preserving data mining", EDBT, 183-199, 2004.

[28] Aggarwal C, Philip S Yu, "A General Survey of Privacy-Preserving Data Mining Models and Algorithms", Springer Magazine, XXII, 11-52, 2008.

[29] Clifton C, Kantarcioglu M, Vaidya J, Xiaodong L, Michael Y, "Tools for Privacy Preserving Distributed Data mining", SIGKDD Explorations letters Vol. 4, Issue 2, December 2002.

[30] Mark W. Craven, Jude W. Shavlik, "Using Neural Networks in Data Mining", Future Generation Computer Systems, special issue on Data Mining, 2001.