Hybrid Intrusion Detection System using FCRM Mechanism

P. Ananthi
Assistant Professor,
Kongu Engineering College, India

P. Balasubramanie
Professor,
Kongu Engineering College, India

ABSTRACT
The necessity of efficient intrusion detection system increased recent research to be focused on hybrid techniques for better results. In recent research plenty of intrusion detection systems have been proposed with various data mining techniques, machine learning mechanisms and fuzzy logic. Existing intrusion detection systems suffered from higher false positive rate and negative rate. This paper proposes the integrated approach such as clustering with Fuzzy neural network for efficient detection rate. In this proposed approach, Fuzzy C-Regression technique is used to construct different training subsets. Then, FNN model is used to take decision making. This proposed approach significantly reduces the false positive and negative rate.

Keywords
Intrusion Detection System, Fuzzy Neural Network, Fuzzy C-Regression model, false positive

1. INTRODUCTION
Intrusion Detection Systems (IDS) used to help detect and restrain different types of attack. Nowadays the attacks in the network turn out to be unavoidable, the existing security systems cannot efficiently identify the powerful attacks such as denial of service, viruses, worms etc so that performance of the security system should be increased by using various techniques for detecting attacks earlier [3]. Many intrusion detection systems are introduced based on the statistical algorithm, heuristic algorithm and many researches has been conducting for improving the security solutions [12]. In Internet, intrusion detection system plays a vital role in detecting the network attacks, such as denial of service (DoS), viruses, worms, trojan horses, spyware, and so on. Furthermore, various kinds of attacks reduce network performance significantly and dilemma users. However, based on the high volume of data traffic involved in a network system, effects of redundant and irrelevant data should be minimized if a qualitative intrusion detection mechanism is genuinely desirous. The main goal of IDS is to prevent the happening of intrusions in the network by classifying packets into two types of attacks and normal.

IDS have been classified based on principle that intruder features which are misuse detection and anomaly detection. The differences of these two types are in their patterns. The misuse intrusion detection regularly examine the network and try according to some predefined signature patterns matches on the network by pattern matching techniques. The anomaly network intrusion based systems provide normal traffic patterns and try to find the deviation from the normal behavior. One of the most important things in the IDS is computational speed and comparison accuracy [15]. According to the tremendous features in each transaction of network a proper mechanism is required to derive an effective subset of features in order to recognize the intrusions.

A number of intrusion detection systems are developed based on many different machine learning techniques. Existing studies apply single learning techniques, such as neural networks, genetic algorithms, support vector machines, etc. Some systems are based on combining different learning techniques, such as hybrid or ensemble techniques[18]. In particular, these techniques are developed as classifiers, which are used to classify or recognize whether the incoming Internet access is the normal access or an attack. Considerably the hybrid approach provides better results than single classifier approach. Artificial Neural Network based Fuzzy c-means clustering is proposed to detect intrusion observed to provide better result and security. This method suffered from certain drawbacks such as lower detection precision for low frequency attacks [20].

The present research work develops an extension of the FC-ANN approach. In order to overcome the drawbacks of fuzzy c-means clustering, an efficient Fuzzy c-regression clustering approach is presented in this research work for clustering [13]. Additionally Fuzzy Neural Network (FNN) is used for better performance. KDD NSL data set is used for simulation result. KDD NSL is the subset of benchmark KDD 99 cup data set, which reduces the duplicate features of old data set. Section 2 describes the related work. Section 3 describes proposed methodology for detection and decision support in an intrusion detection system. Section 4 organized as conclusion and future work.

2. RELATED WORKS
In tremendous network traffic, it is tedious to maintain unbalanced distribution of data, hard to detect boundaries between normal and abnormal behaviors, and adapting to contingency environment [6]. Here we are describing some existing systems for intrusion detection and their potential shortcomings.

Jiang et al. [15] proposed serial and parallel hierarchical neural networks for IDS, which is based on radial basis function (RBF). This approach concentrates misuse and anomaly-based detection. In this approach C-mean clustering algorithm is used to group intrusions into different categories. There-fore, IDS will automatically use these groups to train a new RBF classifier to detect emergent intrusions.

Yang Li, Li Guo [17] proposed supervised intrusion detection method based on TCM-KNN algorithm and active learning method. In this approach feature selection and mapping classical attack patterns of specific application to limited points are the most important problems in real network.

Wenying Feng et.al [18] proposed IDs based on SVM with ant colony method. This IDS not consider the feature extraction of intrusion effectively. SVM classifiers are not enough for properly handling the multiclass cases.

Saurabh Mukherjee and Neelam Sharma [19] proposed Naives Bayes classifies based intrusion detection. FVBRM model for feature selection and make its comparison with three feature selectors CFS, IG and GR. Naïve bayes classifier has its own drawbacks for feature classification of attacks. This method improved the feature selection on someway but failed to produce better result for U2R attacks.

In [20], Hybrid Intelligent Intrusion Detection System is proposed based on specific AI approach for intrusion detection. The technique includes neural networks and fuzzy logic with network profiling. The system detects both anomaly and misuse attacks. Simple if – then Fuzzy rules reflect common ways of describing security attacks. There have been many techniques used for machine learning applications to tackle the problem of feature selection for intrusion detection.

Gan Xu-sheng, et al. [21] proposed anomaly detection mechanism based on PLS and CVM algorithm. the problem of feature extraction and fast modeling for large-scale sample data in anomaly intrusion detection can be solved.

3. METHODOLOGY

This research work presents an improved version of the intrusion detection system based on the Fuzzy C- regression clustering along with the Fuzzy Neural Network. The Fuzzy C-Regression Model (FCRM) of Hath-away and Bezdek [13] was introduced to classify objects into similar groups. FCRM yields simultaneous estimates of parameters for fuzzy C-regression models, while fuzzy partitioning a given dataset. It is supporting hyperplane-shaped clusters.

The proposed framework includes both training phase and testing phase. The arbitrary dataset DS is divided into training set TR and testing TS. Using the Fuzzy C-Regression clustering model the dataset is divided into different training datasets TR_1,TR_2,TR_3,…TR_k. Training subsets are applied to rule based fuzzy neural network to extract features and produce the results.

The proposed method applied the following procedure

Step 1. Apply Fuzzy C-Regression Mechanism on NSL data set

FCRM performs regression process with the number of similar clusters c(2<c<n) from training dataset. Values of cluster centers are obtained then, membership values are obtained by using these values of cluster centers.

Step 2. Form fuzzy partition matrix.

Calculate fuzzy partition matrix and obtain the representatives of each cluster from weighted member function.

Step 3. Establish the fuzzy relationship with neural network.

Attack patterns are extracted based on fuzzy classifier to discriminate normal and attack data.

Step 4. Defuzzify the results.

The FCRM clustering algorithm consists of two phases, calculating fuzzy partition matrix and obtaining the representatives of each cluster, which is carried out separately from the former phase. The solution of representatives of clusters is obtained by weighted recursive least square (WRLS), which needs iteration and this process, is embedded in the outer iterative frame [24]. The proposed system mainly used for detecting the malicious activities and it consist of the Fuzzy C- regression clustering, FNN module and fuzzy aggregation module. In intrusion detection system, detection algorithm is not only enough for detecting intrusion effectively, but also feature selection method is also an important process in IDS. The NSLKDD dataset is used for the evaluating the performance of the proposed IDS. The dataset used in this research has certain features as shown in table 1. The data type of the features is either discrete or continuous and it is labeled as either normal or an attack types.

3.1 Dataset

One of the most important deficiencies in the KDD data set is the huge number of redundant records, which leads inaccuracy in detection rate. Herewith there are two main reasons for this first one is lots of duplicates in training and testing records. The second one is the lack of difficulty measurement in records. Redundant records in training dataset prevents learning method from learning rare records such as U2R attack and R2L attack causes wrong results in testing dataset. Lack of difficulty level can wrongly increase accuracy rate. Because of the simplicity of dataset, learning methods can provide high accuracy without any trouble.

Table.1 Redundant records in KDD99 Training data set

	Original Records	Distinct Records	Reduction Rate
Normal	972,781	812,814	16.44%
Anomaly	3,925,650	262,178	93.32%
Total	4,898,431	1,074,992	78.05%

Table.2 Redundant records in KDD99 Testing data set

	Original Records	Distinct Records	Reduction Rate
Normal	60,591	47,911	20.92%
Anomaly	250,436	29,378	88.26%
Total	311,027	77,289	75.15%

NSL-KDD dataset covers four major categories of attacks such as Probing attacks (information gathering attacks), Denial-of-Service (DoS) attacks (deny legitimate requests to a system), user-to-root (U2R) attacks (unauthorized access to local super-user or root), and remote-to-local (R2L) attacks (unauthorized local access from a remote machine). NSL-KDD dataset is divided into labeled and unlabeled records which class attribute has 21 predicated labels for each record**.**

3.2 Intrusion Detection System Framework

NSL-KDD data set is classified into training and testing data set. Fuzzy c-Regression algorithm partition the training data set TR_1, TR_2,…,TR_n and forms the clusters. Output from FCRM module is passed through FNN module which has input nodes corresponding to major category of attacks such as Dos, Probe, R2L, L2R and normal. Figure.1 shows the framework of intrusion detection system.

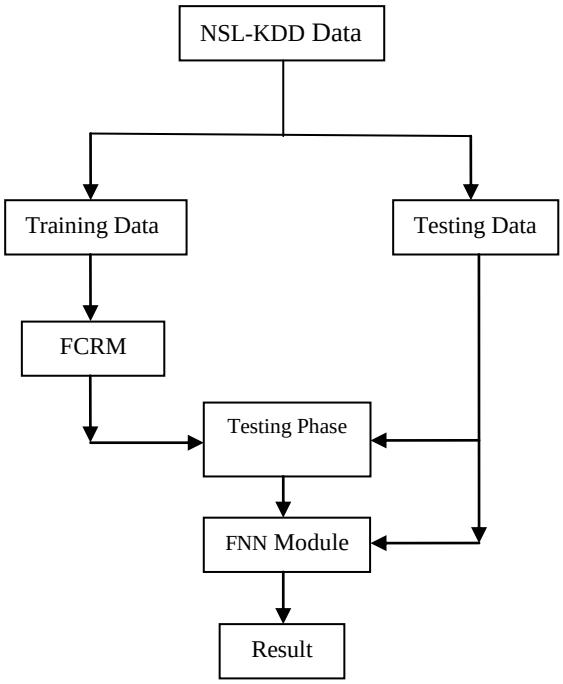

Fig.1 Hybrid Intrusion Detection System

3.3 Fuzzy C-Regression Clustering

3.3.1 Objective Function

Let $S = \{(\mathbf{x}_1,y_1),..,(\mathbf{x}_N,y_N)\} = \{(\mathbf{x}_k,y_k),\ k =1,...,N\}$ be a set of input–output sample data pairs. Assume that the data pairs in S are drawn from c different fuzzy regression models. The hyper-plane of the i-th cluster representative is expressed as follows:

$$y_k = f_i(\mathbf{x}_k,\boldsymbol{\theta}_i) + E_{ik}(\boldsymbol{\theta}_i),= a_{i1}x_{k1} + a_{i2}x_{k2} +\dots + a_{iM}x_{kM}+ b_{i0} + E_{ik}(\boldsymbol{\theta}_i),$$

$$= [\mathbf{x}_k\ 1].\boldsymbol{\theta}_i^T + E_{ik}(\boldsymbol{\theta}_i), i= 1,2,\dots,c \qquad (1)$$

where $\mathbf{x}_k = [x_{k1}, \dots, x_{kM}] \in \mathbb{R}M$ is the input vector, $y_k \in \mathbb{R}$ is the output and $\boldsymbol{\theta}_i = [a_{i1}, \dots, a_{iM}, b_{i0}] \in \mathbb{R}^{M+1}$ is the parameter vector of the corresponding local linear model.

The distance (error measure) between the value predicted by the model $f_i(\mathbf{x}_k, \boldsymbol{\theta}_i)$ and the output y_k is defined by

$$E_{ik}(\boldsymbol{\theta}_i) = |y_k - [\mathbf{x}_k\ 1].\ \boldsymbol{\theta}_i^T|. \qquad (2)$$

The distances $(E_{ik}(\boldsymbol{\theta}_i))$ are weighted with the membership values μ_{ik} in the objective function that is minimized by the clustering algorithm and is given as

$$J(S; U, \theta) = \sum_{k=1}^{N}\sum_{i=1}^{c}(\mu_{ik}^m)E_{ik}^2(\theta_i), \qquad (3)$$

where m is the weighting exponent and μ_{ik} is the membership degree of $\mathbf{x}_k$ to the i-th cluster. The membership values μ_{ik} have to satisfy the following conditions:

$$\mu_{ik} \in [0\ 1],\ i = 1, 2, \dots, c, k = 1, 2, \dots, N, \qquad (4)$$

$$0 < \sum_{k=1}^{N}\mu_{ik} < N, \quad i = 1, 2, \dots, c, \qquad (5)$$

$$\sum_{i=1}^{c}\mu_{ik} = 1, \quad k = 1, 2, \dots, N. \qquad (6)$$

The identification procedure of the FCRM algorithm is summarized as follows[35]. Given data S, set m>1and specify regression model (eqn.1) and choose error measure (eqn.2). Select termination threshold $\varepsilon > 0$ and initialize $U^{(0)}$.

3.3.2 Algorithm

Repeat for *l=1,2,3….,*

Step 1. Calculate values for c model parameters $\theta_i^{(l)}$ in eqn.1 and that globally minimize the restricted function in Eqn.3.

Step 2. *Update* $U^{(l)}$ *with* $E_{ik}(\theta_i^{(l)})$ *to satisfy*

$$U_{ik}^{(l)} = \begin{cases} \left[\sum_{j=1}^{c}\left(\frac{E_{ik}}{E_{jk}}\right)^{\frac{2}{m-1}}\right]^{-1} & \text{if } E_{ik} > 0 \text{ for } 1 \le i \le c \\ 0 \quad otherwise \end{cases}$$

Until $\left\|U^{(l)} - U^{(l-1)}\right\| \le \varepsilon$ then stop. Otherwise, set $l=l+1$ and return to step 1.

3.4 Fuzzy Neural Network

The combination of the fuzzy logic and artificial neural network is used in the neural network is explained in [24, 25]. The FNN is one of the important topics in the research field because it is used in various applications.

Fuzzy neural network is used to learn parameters of the fuzzy sets, fuzzy rules and weights of the rules of a fuzzy system in an iterative way. A neuro-fuzzy system can be interpreted as a set of fuzzy rules. This system can be total created from input output data or initialised with the a priori knowledge in the same way of fuzzy rules. The resultant system by fusing fuzzy systems and neural networks has as advantages of learning through patterns and the easy interpretation of its functionality. Fuzzy systems are suitable for uncertain or approximate reasoning, especially for the system with a mathematical model that is difficult to derive. Fuzzy logic allows decision making with estimated values under incomplete or un-certain information. Neural networks are used to tune membership functions of fuzzy systems that are employed as decision-making systems for controlling equipment. This system applies fuzzy IF-THEN rules in a constructive way.

The basic functions of neural network is training the input data, output data, parameter connection between the neurons which is adjusted through the repeated error corrections and these functions are mainly used to achieve the purpose of learning. The normal if-then rules in the network cannot be encoded directly. The only method is giving a large number of training data to the system. When the fuzzy system is compared with the neural network, the input values can be directly encoded in the fuzzy systems and the tolerance level is also high in fuzzy based system than the neural network [26].

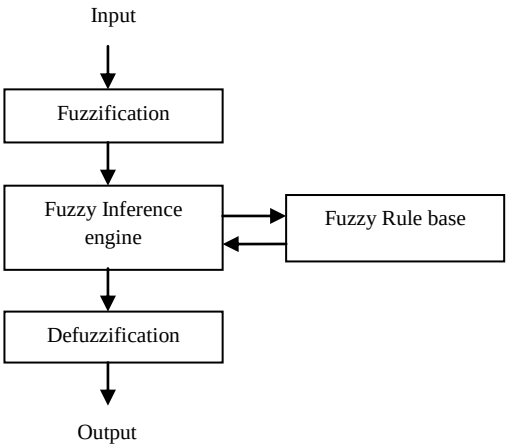

Fig 2. Fuzzy System

4. EXPERIMENTAL RESULTS AND DISCUSSION

This section depicts the experimental results and performance evaluation of the proposed system. Today network traffic data is increasing rapidly. In order to detect intrusion from large traffic data, detection algorithm, and feature selection method have to more efficient. NSL KDD data set is used for evaluating intrusion detection system. The proposed system can easily filters records to improve detection accuracy. The data in NSL-KDD dataset is either labeled as normal or the 24 different kinds of attack. These attacks can be grouped into four major types Probe, DoS, R2L, and U2R. In this proposed mechanism, FCRM algorithm clustering data by various parameters and FNN used to classify network traffic as normal and attack behavior. The effectiveness of the algorithm is identified from high detection ratio and accuracy. The results of the proposed method are presented in table 3.

The evaluating parameters such as precision recall F-value are used in this study to evaluate the proposed system. The formula are defined as

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{F} - \text{value} = \frac{(1 + \beta^2) * \text{recall} * \text{precision}}{\beta^2 * (\text{recall} + \text{precision})}$$

where TP, FP, and FN represents the number of true positives, false positives, and false negatives, respectively, and β represents to the relative importance of precision versus recall and is usually set to 1. The percentage of the evaluation criteria is also measured by using the parameters ie., successful training is measured by using the detection stability.

$$\text{percentage of training successfully} = \frac{\text{the number of training successfully}}{\text{the number of training}}$$

The experiment is conducted and the simulated results are compared for 21 features of NSL-KDD dataset for 5000 records with the other types of intrusion detection system. The parameters are taken here are measurement of precision, recall and f-value of the systems.

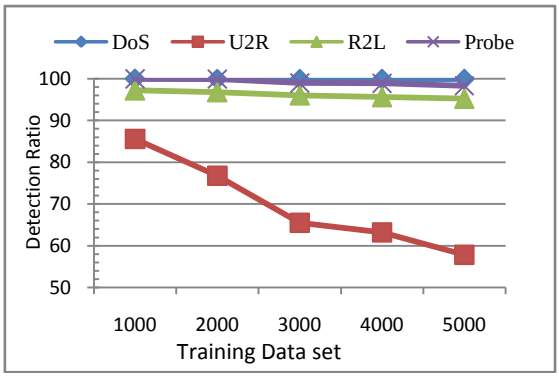

Fig 3. Attack Detection of FCRM-FNN methods

Table 3. Comparison of Detection Ratio with 21 features of NSL-KDD data set

Class Name	SVM system	FC-ANN	FCRM-FNN
Normal	99.5%	99.6	99.89
DoS	99.2%	99.91	99.76
U2R	81.2%	83.33	57.72
R2L	54.6%	93.18	95.2
Probe	95.3%	48.12	95.5

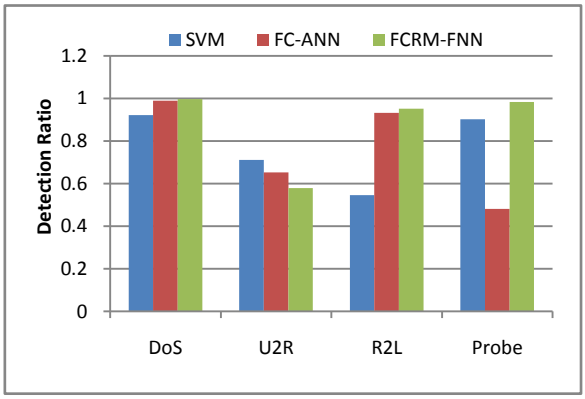

Fig 4: Comparison of Detection Ratio

The comparison of detection ratio between the SVM, FC-ANN and FCRM-FNN is represented in the graph. The detection ratio of the proposed method is compared with the existing system detection ratio. The system performance is measured based on the detection ratio and the proposed approach outperforms the existing system.

5. CONCLUSION

For efficient intrusion detection system approach the fuzzy clustering and fuzzy rule based neural network mechanisms have been employed in order to obtain accurate results [23]. The existing mechanism FCM does not consider the functional relationship of clustering variables. The proposed FCRM algorithm improves the modeling accuracy by forming fuzzy partition matrix of data and parameters which represents cluster centers. This is forming similar clusters with function relationship of input and output variable in first phase. Then second phase fuzzy rule discriminates normal and attach behavior of data to achieve desired results efficacy. The evaluation of the algorithm prominently measured in precision, recall and f-value analysis.

6. REFERENCES

[1] S.-X. Wu and W. Banzhaf, 2010. The Use of Computational Intelligence in Intrusion Detection Systems: A Review. *Elsevier Applied Soft Computing*, vol. 10(1), pp. 1–35.

[2] H. T. Elshoush and I. M. Osman, 2000. Reducing False Positives through Fuzzy Alert Correlation in Collaborative Intelligent Intrusion Detection Systems — A Review. *IEEE Int'l. Conf. Fuzzy Systems*, pp. 1–8.

[3] Patcha, A., & Park, J. M. 2007. An overview of anomaly detection techniques: Existing solutions and latest technological trends. Computer Networks, 51(12), 3448–3470.

[4] Manikopoulos, C., & Papavassiliou, S. 2002. Network intrusion and fault detection: A statistical anomaly approach. IEEE Communications Magazine, 40(10), 76–82.

[5] Ryan, J., Lin, M., & Miikkulainen, R. 1998. Intrusion detection with neural networks. Advances in neural information processing systems (Vol. 10). Cambridge, MA: Springer.

[6] P. Spathoulas and S. K. Katsikas, 2009. Using a Fuzzy Inference System to Reduce False Positives in Intrusion Detection. *Proc. 16th Int'l. Conf. Systems, Signals and Image Processing.*

[7] Kosko, Bart, 1992. Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence. Englewood Cliffs, NJ: Prentice Hall. ISBN 0-13-611435-0.

[8] Lin, W. J. Hwang, and R. J. Wai, 1999. A supervisory fuzzy neural network control system for tracking periodic inputs. IEEE Trans. Fuzzy Systems, Volume 7, No.1, pp. 41-52.

[9] Y. C. Chen and C. C. Teng, 1995. A model reference control structure using a fuzzy neural network. Fuzzy Sets and Systems, Volume 73, pp.291-312.

[10] Bezdek, J. 1974. Fuzzy mathematics in pattern classification. Ph.D. thesis. Ithaca, NY: Cornell University.

[11] Beghdad, R. 2008. Critical study of neural networks in detecting intrusions. Computers and Security, 27(5-6), 168–175.

[12] Axelsson, S. 2003. The base-rate fallacy and the difficulty of intrusion detection. ACM Transaction on Information and System Security, 3, 186–205.

[13] Hathaway, R.J., Bezdek, J.C., 1993. Switching regression models and fuzzy clustering. IEEE Trans. Fuzzy Syst. 1 (3), 195–204.

[14] Moez Solutani, Abdelkadar Chaary, Faycal Benhimda, 2012. A Novel Fuzzy C-regression Model using a new error measure and particle swarm optimization. International Journal of applied Mathematics and computer science, 22(3), 617-628.

[15] Jiang J, Zhang C, Kame M. 2003. RBF-based real-time hierarchical intrusion detection systems. In Pro-ceedings of the International Joint Conference on Neural Networks (IJCNN'03), vol. 2, pp. 1512–1516.

[16] M. Tavallaee, E. Bagheri, W. Lu, and A. Ghorbani.2009. A Detailed Analysis of the KDD CUP 99 Data Set. *Second IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA)*.

[17] Yang Li, Li Guo, 2007. An active learning based TCM-KNN algorithm for supervised network intrusion detection, Computers & Security 26 , 459–467

[18] Wenying Feng, Qinglei Zhang, Gongzhu Hu, Jimmy Xiangji Huang, Mining network data for intrusion detection through combiningSVMs with ant colony networks,Future Generation Computer Systems 37,127–140

[19] Dr. Saurabh Mukherjee,, Neelam Sharma, 2014. Intrusion Detection using Naive Bayes Classifier with Feature Reduction Procedia Technology 4, 119 – 128, Elsevier

[20] Norbik Bashah, Idris Bharanidharan Shanmugam, and Abdul Manan Ahmed,2005. Hybrid Intelligent Intrusion Detection System. World Academy of Science, Engineering and Technology.

[21] Gan Xu-sheng, Duanmu Jing-shun, Wang Jia-fu, Cong Wei, 2013. Anomaly intrusion detection based on PLS feature extraction and core vector machine, Knowledge-Based Systems 40, 1–6, Elsevier.

[22] Hsu-Kun Wu, Jer-Guang Hsieh, Yih-Lon Lin, Jyh-Horng Jeng, 2010. On maximum likelihood fuzzy neural networks, Fuzzy Sets and Systems 161, 2795 – 2807, Science Direct.

[23] Michel Menard, 2001. Fuzzy clustering and switching regression models using ambiguity, and distance rejects, Fuzzy Sets and Systems 122, 363–399, Elsevier

[24] I.B. Türkşen, 2011. A review of developments in fuzzy system models: Fuzzy rule bases to fuzzy functions Scientia Iranica D 18 (3), 522–527.

[25] Heba F. Eid, Ashraf Darwish, Aboul Ella Hassanien, and Ajith Abraham, 2010. Principle Components Analysis and Support Vector Machine based Intrusion Detection System 978-1-4244-8136, IEEE.

[26] Chaoshun Li, Jianzhong Zhou , Xiuqiao Xiang, Qingqing Li, Xueli An, 2009. T–S fuzzy model identification based on a novel fuzzy c-regression model clustering algorithm, Engineering Applications of Artificial Intelligence, 646–653.