International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

Fuzzy ECOC Framework for Network Intrusion Detection System

Uma Shankar Rao Erothi, Sireesha Rodda

Abstract: ***Many aspects of our life now continually rely on computers and internet. Data sharing among networks is a major challenge in several areas, including communication, national security, medicine, marketing, finance and even education. Many small scale and large scale industries are becoming vulnerable to a variety of cyber threats due to increase in the usage of computers over network. We propose Fuzzy-ECOC frame work for network intrusion detection system, which can efficiently thwart malicious attacks. The focus of the paper is to enforce cyber security threats, generalization rules for classifying potential attacks, preserving privacy among data sharing and multi-class imbalance problem in intrusion data. The Fuzzy-ECOC framework is validated on highly imbalanced benchmark NSL_KDD intrusion dataset as well as six other UCI datasets. The experimental results show that Fuzzy-ECOC achieved best detection rate and least false alarm rate.***

Keywords: ***Network Intrusion Detection System, Fuzzy Classification, ECOC, multi class imbalance, Machine Learning.***

I. INTRODUCTION

The role of network intrusion detection system (NIDS) must effectively identify all major and minor attacks, even if they form a small fraction of network intrusion data. NIDS refers to the set of models which are used to isolate attacks against computer networks. Many Hacker/Intruders can create successful attempts to crash the networks. Finding modern technique solutions to prevent these threats are important. Cyber Security involves protecting networks, computers, and data from unintended/unauthorized access, change or destruction. Unfortunately, most cyber security defenses are reactive in nature, such that they cannot anticipate attacks and fail to keep up with new attack types.

Many researchers offer a different Machine Learning solutions for developing an effective NIDS such as Support Vector Machine [1], Decision Tree [2], K-Nearest Neighbor [3], Naïve Bayes [4], AdaBoost [5] etc. can efficiently thwart malicious attacks. Over the years, machine learning techniques have been successfully applied on many cyber security applications, such as Intrusion Detection and Misuse Detection. Machine Learning approaches are ideally suited to handle cyber-attacks but they fail to anticipate previously unknown or unseen behaviors. The process of correctly identifying unknown/unseen intrusions from the network traffic is considered as an intrusion detection classification problem. On the other hand, data sharing among inter corporation's necessitates privacy among sensitive data. A Fuzzy-ECOC framework can be built with the objective of efficiently identifying malicious activities while minimizing the false alarm rate (FAR), improving accuracy and detection rate. The main focus of Fuzzy-ECOC for NIDS System is to predict the behaviors of users over networks, and these behaviors can be analyzed/predicted as a normal or intrusion behavior. In order to identify generalization rules for classifying potential attacks. We use Fuzzy Membership function to pre-process the data to the range [0, 1] and attributes of each instance X are defined with the linguistic values (Low, Medium, High) as shown in Fig. (1). the representation of linguistic values protect sensitive data from unauthorized users, when the organization releases sensitive or confidential data to third-party for performing data mining operations. This makes fuzzy classification is a great choice for user's privacy, the boundary between benign and malicious classes are well separated for intrusion detection. Fuzzy-ECOC is trained on fuzzy space belongs to different classes at the same time. Moreover the intrusion data involves many numeric features in collected data. Building classification models on numeric type usually causes high detection errors. A malicious activities that deviate slightly from the classification model may not be detected or a small change in normal connection may increase false positive rate. With fuzzy classification, it is possible to build model for these minor deviations to maintain lower false negative / false positive rates. To preserve privacy among data, researchers in the data-mining professionals have proposed various methods such as K-anonymity [6], L-Diversity [7], perturbation based (Adding Noise [8], Randomization [9],[10], Cryptography based (Pseudonymization [11], Secure Multi Party Computation [12]) and Normalization [13],[14] techniques especially for the security related data mining applications. But these approaches often very complex and time consuming to execute and suffer from problems such as excessive generalization and suppression [15].

Many real world applications are multi class in nature, where most of the classes are very similar and very difficult to discriminate one among others. When class imbalance exists in network intrusion data, greater accuracy of the NIDS does not indicate essentially a more efficient NIDS. The NIDS also expected to identify novel connections effectively [16].

Manuscript published on 30 September 2019
* Correspondence Author
Uma Shankar Rao Erothi*, Dept of CSE, RAGHU Institute of Technology, Visakhapatnam, India. Email: umashankar.erothi3@gmail.com
Sireesha Rodda, Dept of CSE, GITAM University, Visakhapatnam, India,Email: sireesha.rodda@gitam.edu

To alleviate multi class imbalance problem, several researchers proposed different variations of ECOC frameworks, but they are mainly focused on binary class data. The multi class imbalance is more challenging than binary class imbalance. we use Error correcting output codes (ECOC) to tackle multi-class imbalance problem in intrusion data to reduce false alarm rate and to increase the overall accuracy of each attack category (Dos, Probe, U2R and R2L).

II. RELATED WORK

Several researchers proposed different variations of ECOC techniques presented in [17],[18],[19],[20]. In general, they are categorized into problem independent and problem dependent methods. The code words created for N_c- classes in problem independent method are not always discriminative for the multi-class classification problem. Hence, problem dependent methods have been proposed such as discriminant ECOC (DECOC) [20], node embedding ECOC (ECOC-ONE) [21], deep learning ECOC (DeepECOC) [22] and imbalanced ECOC (imECOC) [19]. The basic ECOC design presented in [17] includes two major steps, coding and decoding. The popular binary ECOC coding strategies are one-vs.-all (OVA) [23] and dense random [20] strategies, they are confined to {1,-1} symbols shown in Fig. (2). In OVA, each base classifier is trained to distinguish one class from the other classes. The work presented in [24] extended the binary ECOC design by considering zero symbol, called ternary ECOC, {-1, 0, 1} shown in Fig. (3). the ternary ECOC coding strategies are one-vs.-one [25] and sparse random strategy [24]. The OVA consider all pairs of classes, hence the length of the code matrix is defined as $\frac{N_c*(N_c-1)}{2}$. [26], investigated the behavior of the ensemble ECOC framework on image vision application. The dense and sparse random ECOC methods were compared with the traditional multi class classification methods including One-Vs.-One (OVO) and One-Vs.-All (OVA). The ECOC method for logo recognition and shape categorization improved the classification performance in comparison with the traditional multiclass approaches. An ensemble model based on ECOC for the medical diagnosis application is presented in [27]. The best feature space for each dichotomizers in the ECOC design are optimized to improve their accuracy. The proposed method achieved accurate detection in classifying four Angle closure glaucoma (ACG) eye disease categories. The authors of [28] presented data-driven ECOC (DECOC) method to study real world applications such as holistic recognition (hand written numeral recognition) and for the classification of cancer tissue types. The experiments were conducted on hand written numeral pair NIST dataset and micro-array gene expression NCI cancer dataset. The performance of the DECOC is compared with the other popular decomposition methods like one-vs.-One, DAGSVM and pairwise coupling. The imECOC method in favor of minority classes presented in [19] used weighted code matrix. The weighted code matrix tackles both between class imbalance and with-in imbalance problem. The Experimental results from fourteen UCI datasets indicated that, imECOC outperforms other state of the art ECOC methods. The imECOC method is evaluated in terms F-Score, G-mean and AUC measures. To improve the process of binary coding strategies, [29] presented best decomposition methods for multi class classification problem. In order to achieve optimal results for the ECOC design model, [24] suggest that dense random and sparse random ECOC approaches requires $10\log_2(N_c)$ and $15\log_2(N_c)$ base classifier. The random ECOC coding approaches cannot guarantee that the created base code words are always discriminative for the multi-class classification task. The work proposed in [30] compared different approaches for multi-class SVM-problems, including OVO, OVA, and D-DAG. Experiments were conducted on ten benchmark datasets. The author claimed that the OVO method is better than the other approaches. OriolPujol et al [20] proposed D-ECOC method using decision tree, to find the most discriminative code words for the N_c Classes considered exactly $N_c - 1$ base classifiers, which significantly improved the performance of dichotomizers. But, the decision tree based approach has one major drawback i.e., if the parent node misclassifies an instance, the mistake will be propagated to all the subsequent child nodes. To overcome this problem [21] proposed ECOC-ONE with OVA strategy. This approach iteratively changes the dichotomy classifiers to discriminate the most confusing pairs of classes to improve the performance of ECOC ensemble model. However, this approach suffers from major drawback, if the initial coding matrix fails to perform well, the final results of ECOC-ONE are un-satisfactory.

III. METHODS

Fuzzy Logic

Fuzzy logic is a method to computing based on degree specified by its membership function (MF). MF is a curve that defines how each instance in the input space is mapped to a membership value between 0 and 1. Fuzzy logic has proved to be a powerful tool for decision making. In a fuzzy classification system, the attribute set A in X is characterized by a membership function $f_A(X)$. An instance X is classified based on the linguistic values of its attributes (LOW, MEDIUM, and HIGH) as shown in Eq. (1).The triangular membership function used in this paper is depicted in Fig1.

$$f_A(x) = \begin{cases} 0 & x < min \\ \frac{x-min}{\alpha_1-min} & min \le x \le \alpha_1 \\ \frac{x-\alpha_1}{\alpha_2-\alpha_1} & \alpha_1 < x \le \alpha_2 \\ \frac{x-\alpha_2}{max-\alpha_2} & \alpha_2 < x \le max \\ 0 & x > max \end{cases} \quad (1)$$

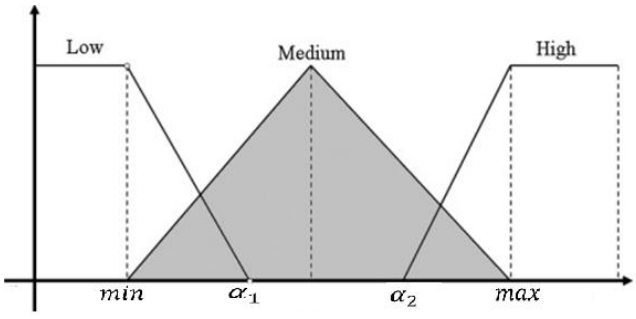

Fig. (1) Triangular Membership Function

Error Correcting Output Codes (ECOC)

The basic idea of the ECOC ensemble framework is to separate instances that belongs to multiple classes. It has three steps: coding, learning and decoding. The figures shown in 2 and 3 for N_C classes decompose the multi-class classification problem into L binary classifiers .The code matrix M with $\{1,-1\}$ or $\{1,0,-1\}$ values of size $N_C * L$ is trained on set of two-class problem consists of different combinations of the original classes. Then, these results can be combined to provide solution to the original multi-class problem. A good ECOC framework needs a proper coding and decoding strategy to enhance generalization ability, to reduce variance and bias produced by the learning algorithms. To design ECOC framework for a N_C -class problem, it should satisfy that the rows and columns are well separated in terms of Hamming Distance (HD) and Euclidian Distance (ED).

An example of the ECOC coding design for a given 4-class problem is presented in Fig. 2(a) & Fig. 2(b) show class partitions in each column of the ECOC matrix in Binary Case and Ternary Case respectively. ECOC coding design for a four class, +1 indicates considered class for the dichotomy classifier H_j, -1 indicates other classes, and zero indicates classes that are not considered by the respective dichotomy classifier in the learning. Once the classifier is learned for all four dichotomizers, at the decoding step a new test sample X_i is tested by H_i classifiers. Then, the new vector (or) code word $B_i^* = \{B_1, B_2, \dots, B_n\}$ is compared with the class codeword's $B_i = \{C_1, C_2, \dots, C_n\}$, classifying the new sample by the class C_i which code word minimizes the HD.

Coding and Decoding

Several researchers proposed problem independent and problem dependent coding strategies. The code-words in ECOC techniques separate the classes between rows and columns, in order to achieve diversity among dichotomies. The popular coding strategies shown in Fig. 3(a) to Fig. 3(d) includes, OVA maintains exactly N-bit code word length for N_C number of classes, whereas OVO uses $N_C * (N_C - 1)/2$ code word length. The sparse and dense random coding design with the estimated length of $15 * \log(N_C)$ and $10 * \log(N_C)$ bits are presented in [24]. The authors [17] used well separated coding mechanisms for multi-class classification problem, where code length for N_C-classes is chosen$(2^{N_C-1} - 1)$. The exhaustive code words for 10 class and 5 class problem are presented with a row separation of hamming distance (HD) of 8.

The simplest standard decoding strategies are hamming distance $HD(x, y_i) = \sum_{j=1}^{n} |x_i - y_i^j|/2$ and Euclidean distance $ED(x, y_i) = \sqrt{\sum_{j=1}^{n} (x_i - y_i^j)^2}$, where y_i is the distance to the row for a class j, n indicates number of dichotomies, x and y indicates input vector and output vector code words respectively. If the minimum HD between any pair of code words is d, then $\frac{(d-1)}{2}$ errors in the individual binary classifier result can be corrected, since the closest code word will be the correct one.

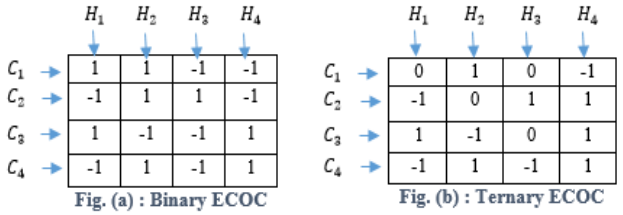

Fig.2. (a) Binary ECOC; (b) Ternary ECOC design for a four – class problem

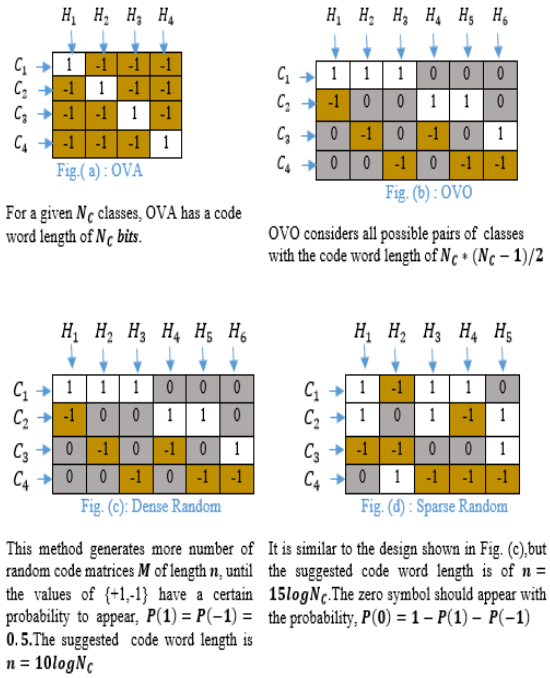

Fig. (3) Variations of ECOC Coding design for a four-class problem

Proposed Method

The aim of proposed Fuzzy-ECOC framework is to provide an efficient intrusion detection with high detection and low false alarm rate. The working of proposed framework is presented in Fig. (4) and Algorithm 1. The architecture operates in three phases: pre-processing phase, classifier building phase and validation phase.

The pre-processing phase for a given input training dataset $D = \{(x_i, H(y_i)\}_{i=1}^{n}$ with x_i in domain belongs to conditional attribute set $A_i = \{A_1, A_2, A_3, \dots, A_N\}$ and $H(y_i)$ in domain belongs to decision attribute with N_C number of classes $\{C_1, C_2, C_3, \dots, C_N\}$ is considered. In order to preserve the user's privacy, the original values in the attribute set A_i are converted to fuzzy form shown in Eq. (1). Under classification phase, the fuzzy input space is defined according to fuzzy-membership function as shown in fig (1). Once the Fuzzy-space is obtained, Fuzzy-ECOC framework with base learner C 4.5 is trained on ECOC code words as sown in fig (2) and fig (3). Then, validation phase is used to classify a test instance x^*, then we apply each of the learned function H^*_j to x^* to compute a response vector of binary decisions$B^* = \{H^*_1(x^1), H^*_2(x^2), \dots, H^*_k(x^l)\}$.

Then we find code word B_i is closest to this B^* vector using HD as shown in line 4 through 6. In coding stage (line 2), we considered different variations of coding design shown in Fig. (2) and Fig. (3), and each code-words are trained to distinguish one class from the other class. Fuzzy-ECOC (line 3) is an ensemble method which combines many binary classifiers in order to solve the multi-class classification problem. In decoding (line 5), each classifier predicts a value for an unknown instance resulting in the code-word of length l. Then a closest is assigned in the code matrix M using hamming decoding (line 6).

Algorithm 1 : Fuzzy- ECOC

Input : Training set $D = \{(x_i, H(y_i)\}_{i=1}^{n}$;
$H(y_i) = \{1,2,..,N_C\}$; binary classifier l

//Classification phase

1 Generate $N_c * l$ distinct binary or ternary code word matrix$M$, where $M \in \{-1,+1\}^{Nc*l}$ in binary case or $M \in \{-1,0,+1\}^{Nc*l}$ in ternary case.
2 Each class is assigned one row from a code matrix M
for i = 1 to n
$C_i \leftarrow M_i$
end for
3 Train the base classifier to learn the $H(y_i)$ binary functions (one for each column)
for i = 1 to l
$H_i \leftarrow l(x_i, H(y_i))$
end for

//Validation phase

4 Apply each of the H_i learned classifiers to the test example
For each test instance x^* in dataset D
for i = 1 to l
$B(x_i, H(y_i)), \forall x_i \in D, \forall H(y_i) = 1,2,\dots,N_c$
5 Combine the predictions to form new output vector B^* of length l.
6 Classify to the class with the nearest code word
$B_i = argmin\, B^*(x_i, H(y_i))$
7 evaluate measures
8 If measures not found satisfactory
9 repeat process till results are satisfactory

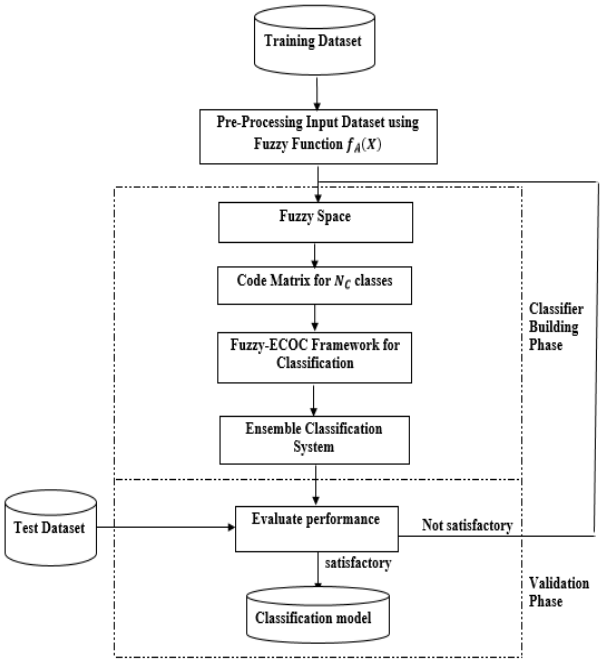

Fig. (4) Fuzzy ECOC framework for Network Intrusion Detection System

IV. DATASET DESCRIPTION

The performance of the proposed Fuzzy-ECOC framework for network intrusion detection system is evaluated on benchmark NSL_KDD [31] NIDS dataset. Since 1999 many researchers utilized KDDCUP'99 dataset for anomaly-detection approach. It has approximately four million records collected from DARPA'98[32] dataset, which consists of 3GB tcp_dump connection data. NSL-KDD is another improved well known utilized version of KDDCUP'99 intrusion dataset; it eliminates redundant records from the training and test dataset. NSL_KDD contains 41 conditional features and a decision feature, and these connections are labelled as either benign (normal) or a malicious (attack). All conditional features includes basic categories such as Traffic (TF), Host (HF), Basic (BF) and Content (CF) features. The number of training and test samples in NSL_KDD for different types of attacks are presented in [33].The dataset includes 22 attacks and they are classified into four categories: Probe, Dos, U2R, and R2L shown in Fig. (5).The complete details of different attacks in the dataset are clearly outlined in [34]. The summary of other benchmark datasets from University of California Irvine Machine Learning Repository (UCI_ML) [35] are presented in Table 1. The summary of class distribution is shown in Fig. (5).

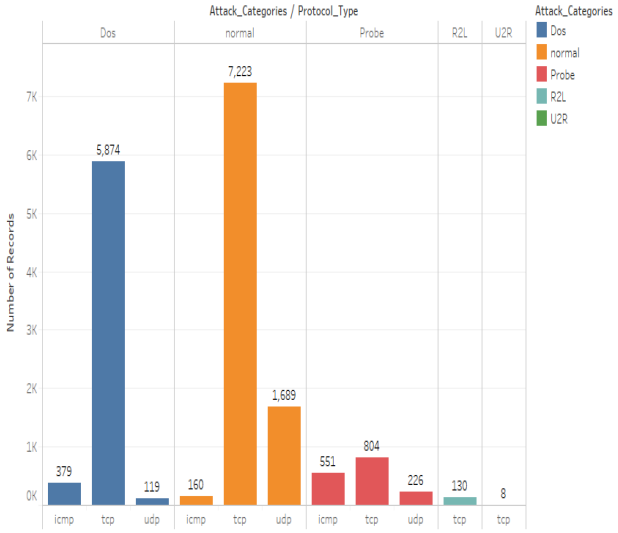

Fig. (5) Summary of Class Distribution in NSL_KDD dataset

Table 1: Summary of UCI datasets

Sno	Dataset	#Train	#Attribute	#Class
1	Iris	150	4	3
2	Wine	178	13	3
3	Balance Scale	625	4	3
4	Zoo	101	17	7
5	Car	1728	6	4
6	Glass	214	10	7

V. EXPERIMENTAL RESULTS AND ANALYSIS

All the experiments were conducted on Intel® Core™ i3-5005U CPU @ 2.00GHz PC with 4GB RAM running on 64-bit OS and x64-based processor. The implementation is done using Java programming language with the help of java weka library tool [36]. Experiments were conducted on NSL_KDD and other UCI datasets to evaluate the performance of proposed method. The Fuzzy-ECOC method presented in this paper are evaluated with four different coding strategies shown in Fig. (2) and Fig. (3). Dichotomizers in ECOC design are trained using C 4.5 as base learner. The decoding process is chosen based on minimum HD. The performance of the proposed Fuzzy-ECOC is evaluated by considering all the measures Accuracy, Detection Rate (DR), Precision, False Alarm Rate (FAR), G-mean, F-Score, and Area under Curve (AUC). As accuracy is not an adequate evaluation measure in presence of class imbalance. F-Score, G-mean and AUC are also evaluated to study the multiclass imbalance problem. The measures are provided in Eq. (2) through (8) in terms of confusion matrix metrics True Positive (TP), False Negative (FN), False Positive (FP) and True Negative (TN) shown in Table 2.

Table 2:Confusion Matrix

		Actual	
		Intrusion	Normal
Predicted	Intrusion	TP	FP
	Normal	FN	TN

Performance Measures

$$Accuracy = \left(\frac{TP+TN}{TP+TN+FP+FN}\right) \quad (2)$$

$$Recall\ or\ DR = \left(\frac{TP}{TP+FN}\right) \quad (3)$$

$$Precision = \left(\frac{TP}{TP+FP}\right) \quad (4)$$

$$FAR = \left(\frac{FP}{FP+TN}\right) \quad (5)$$

$$F-Score = \left(\frac{2*Precision*Recall}{Precision+Recall}\right) \quad (6)$$

$$G-Mean = \sqrt{Precision * Recall} \quad (7)$$

$$AUC = (1+TP-FP)/2 \quad (8)$$

First, we evaluate Fuzzy-ECOC method on six UCI machine learning repositories shown in Table 2. The performance of the proposed method is evaluated against various state-of-the-art ECOC encoding strategies OVO, OVA, Dense Random and Sparse Random methods. The classification results produced by these methods are evaluated from the confusion matrix, which gives the actual vs. predicted results. The results indicated in Fig. (6). for the iris dataset are best for the Fuzzy-ECOC method for all the considered methods. It is clearly observed that OVO method shows slight improvement in all the measures over OVA, Dense Random and Sparse Random methods. The OVO method outperforms in all the measures. The best results indicated by all the methods may be the class distribution of iris dataset is balanced. The Sparse and Dense Random methods achieved highest measures in Fig. (7), highest FAR and lower DR is indicated by OVO and OVA methods.

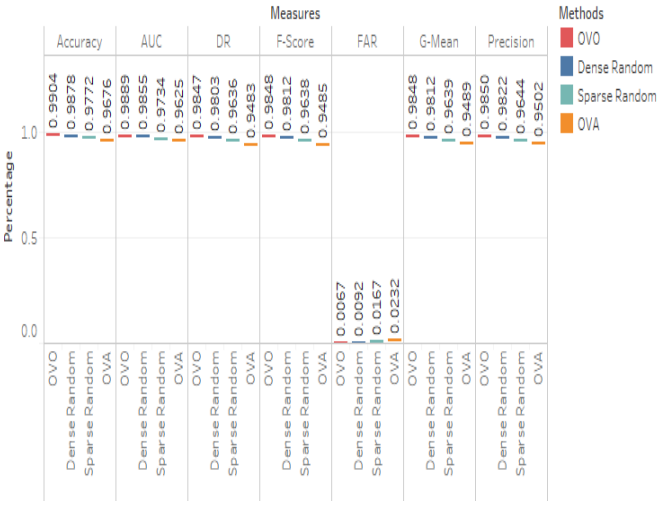

Fig. (6): Performance Measures of Iris Dataset

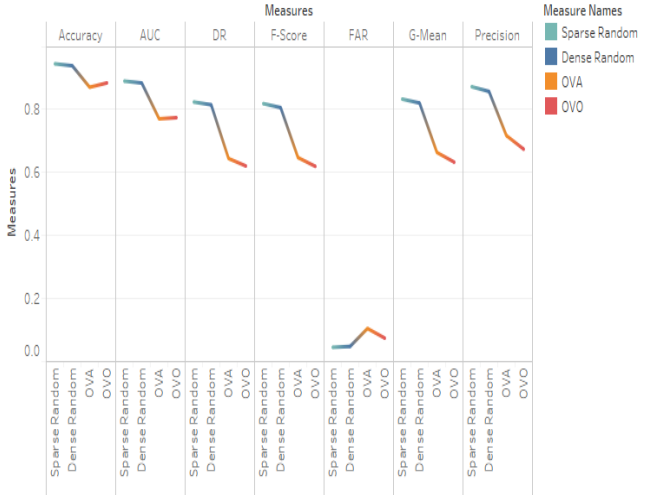

Fig. (7): Performance Measures of Glass Dataset

Fig. (7). shows the performance measures obtained on the Glass dataset. Sparse Random method outperforms in all the measures with a slight improvement over dense random method. The improvement of F-Score, G-mean and AUC values for sparse random method are best in detecting minority classes. Dense random method also perform very similar to sparse random method.

The results indicated in Fig. (8), OVA method achieved lowest values when compared to other methods. The highest accuracy, DR and FAR is obtained by dense random method. On the other hand, the higher AUC and lower FAR is indicated by sparse random method. It is clearly observed that OVA method increasing the FAR. The good ECOC architecture for Network Intrusion Detection System should always need lowest FAR and higher DR. The dense and sparse methods performs similar with slight variations in all the measures.

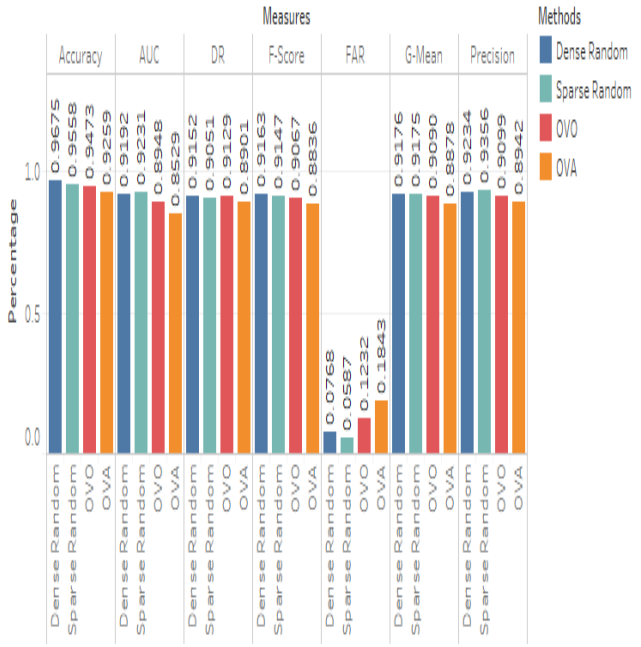

Fig. (8): Performance Measures of Car Dataset

Similar performance may be observed in Fig. (10). for Balance Scale dataset, However, the OVO method performs better than dense and sparse random method for the Zoo dataset shown in Fig (9). The dense random method indicated highest G-mean that indicates it is very good at detecting minority classes. The highest FAR is indicated by both OVA and OVO methods in Fig (10). However, the OVA method is poor choice for ECOC learning, but still used in ECOC framework, because it takes less time to build ECOC-Classifier as it contains less number of dichotomizers when compared to other methods. The improvement in sparse and dense random coding strategies for minority classes is due to the optimal dichotomizers involved at each step of the method when there is a difficult cases for classifying majority and minority classes. The performance of OVA and OVO methods are unstable on the datasets Iris, Glass, Wine and Balance Scale. The performance of OVA and OVO methods depends on the initial code matrix.

By examining the Fig. (9) to Fig. (12), we can see that the Fuzzy-ECOC framework with C 4.5 as base learner achieved best results for all the considered measures. We can clearly observed that the performance of OVO, Sparse and Dense ECOC results with a slight variation when the number of dichotomizers increases. However for a large number of classes, the sub-problems of sparse and dense random designs is considerably less when compared to OVA and OVO methods. The results shown in Fig. (9). for the Zoo Dataset, OVO method achieved highest accuracy, DR, F-Score, AUC and Precision values. The dense random method increased FAR.

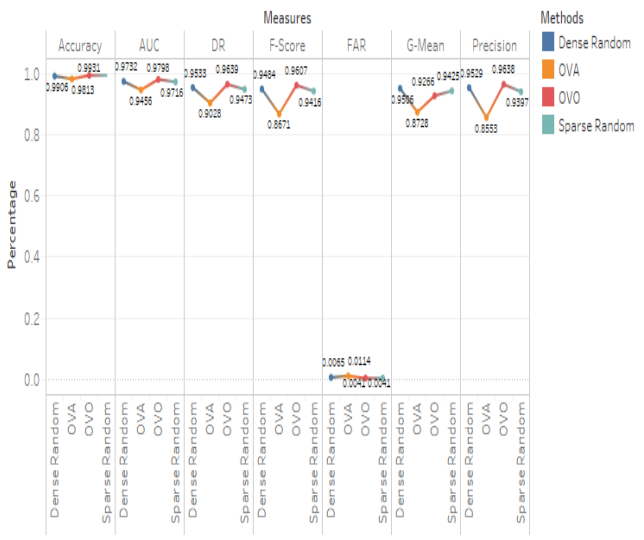

Fig. (9): Performance Measures of Zoo Dataset

Results presented in Fig. (12). show that the proposed Fuzzy-ECOC framework achieved better accuracy, DR, precision, G-Mean and F-score values when we use OVO, Dense and Sparse Random method when compared to OVA method. C 4.5 has given good measures for all the coding methods with the exceptions of OVA. It is clearly observed that, the majority class detection i.e., normal and dos classes given good results for all the coding strategies. On the other hand, the lowest FAR is obtained by the sparse random approach. The best F-Score, G-Mean and AUC values for network intrusion dataset i.e., NSL_KDD, improved the performance of minority class detection. It is observed that the dense random method achieved good results in detecting all types of attacks without any exceptions. It can be concluded that Fuzzy-ECOC ensemble framework with C 4.5 and dense random coding strategy achieved best values for detecting all major attacks viz., Dos, Probe, U2R and R2L.

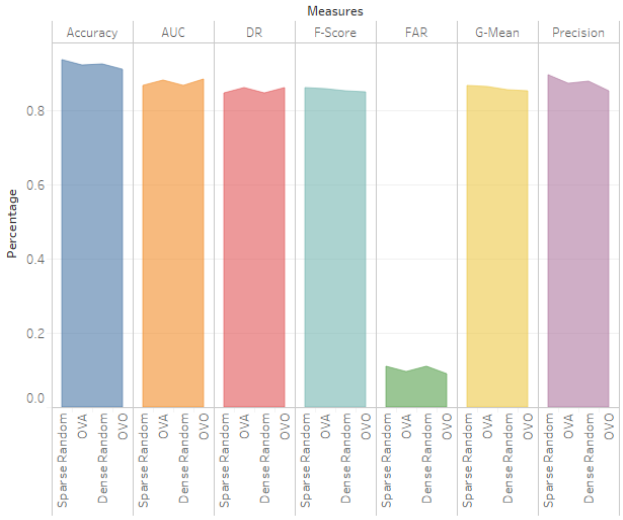

Fig. (10): Performance Measures of Balance Scale Dataset

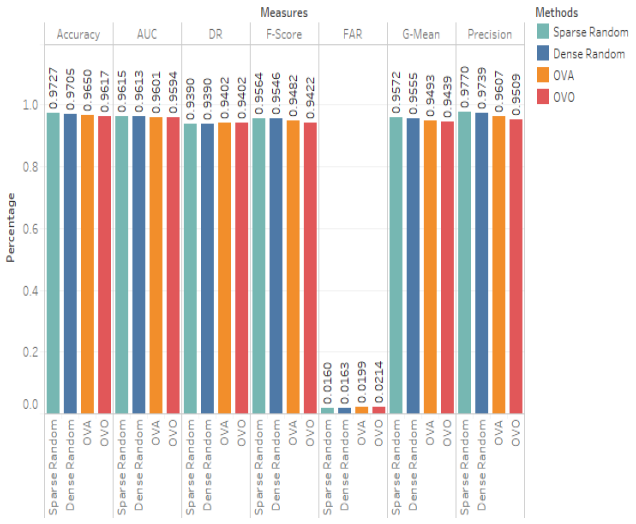

Fig. (11): Performance Measures of Wine Dataset

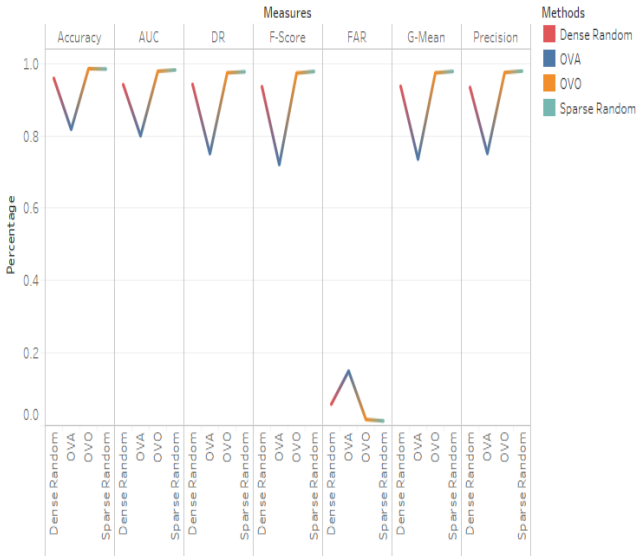

Fig. (12): Performance Measures of NSL_KDD Dataset

VI. CONCLUSION

In this paper, we presented a novel Fuzzy-ECOC framework to deal with multi-class classification problem. The proposed ECOC technique learns on Fuzzy-subspaces. Each dichotomizer is trained on popular ECOC coding strategies OVO, OVA, dense random and sparse random. The proposed framework shows improved performance for imbalanced datasets, the independent classifiers increased the overall accuracy of the ensemble. Experimental results shown that the proposed approach provides best detection rate and least false alarm rate for most of the standard datasets considered. The Fuzzy-ECOC significantly improved the performance of evaluation measures in detecting both known and unknown attacks. The conventional machine algorithms alone is not enough to classify potential attacks when there is a class imbalance problem. Hence, we designed a specialized ECOC framework by giving due importance to the minority class detection.

ACKNOWLEDGMENT

The authors would like to express their gratitude to the Science and Engineering Research Board (SERB), Ministry of Science & Technology, Govt. of India under grant No. SB/FTP/ETA -0180/2014 for providing partial funding support.

REFERENCES

1. Snehal A.Mulay, P.R.Devale,G.V.Garje.”Intrusion Detection System using Support Vector Machine and Decision Tree.”, International Journal of Computer Applications,3(3): 40-43,2010.
2. Vaishali Kosamkar, Sangita S. Chaudhuri.“Improved Intrusion Detection System using C 4.5 Decision Tree and Support Vector Machine”, International Journal of Computer Science and Information Technologies,5(2):1463-1467,2013.
3. Wenchao Li, Ping Yi, Yue Wu, Li Pan, and Jianhua Li.”A new Intrusion Detection System based on KNN classification algorithm in wireless sensor network.”, Journal of Electrical and Computer Engineering, http://dx.doi.org/10.1155/2014/240217. pp 1-7, 2014.
4. Mrutyunjaya Panda, Manas Ranjan Patra.”Network Intrusion Detection using Naïve Bayes, International Journal of Computer Science and Network Security.”, 7(12): 258-263, 2007.
5. Weiming, Hu, Wei, Hu, & Maybank, S.”AdaBoost-based Algorithm for Network Intrusion Detection.”, IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics, 38:577–583, 2008.
6. Ninghui Li, Tiancheng Li and Suresh Venkatasubramanian.”t-closeness: Privacy Beyond k-anonymity and l-diversity.”, IEEE 23rd International Conference on Data Engineering, IEEE, 1-10, 2007.
7. Ashwin Machanavajjhala, Johannes Gehrke, Daniel Kifer,Muthuramakrishnan Venkitasubramaniam.”l-diversity: Privacy beyond k-anonymity.”, ACM Transactions on Knowledge Discovery from Data (TKDD), 1(1):1-12, 2007.
8. Ashish. E. Mane and Sushma Gunjal.”Privacy preserving using additive perturbation based on multilevel trust in relational streaming data.”, Multidisciplinary Journal of Research in Engineering and Technology (MJRET), 2(2): 392-397, 2015.
9. Swapnil Kadam and Navnath Pokale.”Preserving Data Mining through Data Perturbation.”, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 4(11):4128-4131, 2015.
10. Yu Zhu and Lei Liu.”Optimal randomization for privacy preserving data mining.”, Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 761-766, 2004.
11. Wenliang Du and Mikhail J.Atallah.”Secure multy-party computation problems and their applications: a review and open problems.”, Proceedings of the 2001 workshop on new security paradigms, ACM, 13-22, 2001
12. Benny Pinkas.”Cryptographic techniques for privacy-preserving data mining.”, ACM Sigkdd Explorations Newsletter, 4(2):12-19, 2002.
13. Syed Md. Tarique Ahmad, Shameemul Haque and Prince Shoeb Khan.” Privacy Preserving in Data Mining by Normalization.”, International Journal of Computer Applications, 96(4):14-18, 2014.
14. C.Saranya and G.Manikandan.”A Study on normalization techniques for privacy preserving data mining.”, International Journal of Engineering and Technology (IJET),5(3): 2701-2704, 2013.
15. Uma Shankar Rao Erothi, Sireesha Rodda.”Data Transformation Technique for Preserving Privacy in Data.”, International Journal of Computer Sciences and Engineering, 6(5):42-50, 2018.
16. Rodda, S., & Erothi, U. S.”A Roughset Based Ensemble Framework for Network Intrusion Detection System.”, International Journal of Rough Sets and Data Analysis (IJRSDA), 5(3):71-88, 2018.
17. T. G. Dietterich and G. Bakiri.”Solving multiclass learning problems via error-correcting output codes.”, Journal of artificial intelligence research, 263–286, 2012.
18. Gholam Ali Montazer, Sergio Escalera, et al.”Error correcting output codes for multiclass classification: Application to two image vision problems.”, In AISP, 508–513, 2012.
19. Xu-Ying Liu, Qian-Qian Li, and Zhi-Hua Zhou.”Learning imbalanced multi-class data with optimal dichotomy weights.”, In Data Mining (ICDM), 2013 IEEE 13th International Conference on, 478–487, 2013.
20. OriolPujol, PetiaRadeva, Jordi Vitria.”Discriminant ECOC: a heuristic method for application dependent design of error correcting output codes.”, In IEEE Transaction on Pattern Analysis and Machine Intelligence, 1007–1012, 2016.
21. Sergio Escalera ,OriolPujol.”Ecoc-one: A novel coding and decoding strategy.”, In ICPR, 578–581, 2006.

22. GuoqiangZhong,YuchenZheng,Peng Zhang,Mengqi Li,JunyuDong." DEEP Error Correcting Output Codes.", Under review as a conference paper at ICLR, 1-11, 2017.
23. N.J. Nilsson, "Learning Machines." New York: McGraw-Hill, 1965.
24. Allwein, E. L., Schapire, R. E., & Singer, Y."Reducing multiclass to binary: A unifying approach for margin classifiers, Journal of machine learning research.", 1:113-141, 2001.
25. T. Hastie and R. Tibshirani."Classification by Pairwise Grouping.", Proc. Conf. Neural Information Processing Systems, 26:201-233, 2002.
26. Bagheri, Mohammad Ali, Qigang Gao, and Sergio Escalera."Rough set subspace error-correcting output codes.", ICDM, IEEE 12th International Conference on Data Mining, 2012.
27. Bai, Xiaolong, Swamidoss Issac Niwas, Weisi Lin, Bing-Feng Ju, Chee Keong Kwoh, Lipo Wang, Chelvin C. Sng, Maria C. Aquino, and Paul TK Chew."Learning ECOC code matrix for multiclass classification with application to glaucoma diagnosis.", Journal of medical systems 40, 4:78, 2016.
28. Zhou, Jie, Hanchuan Peng, and Ching Y. Suen."Data-driven decomposition for multi-class classification.", Pattern Recognition 41, 1: 67-76, 2008.
29. Utschick, Wolfgang, and Werner Weichselberger."Stochastic organization of output codes in multiclass learning problems.", Neural Computation, 13(5):1065-1102, 2001.
30. Hsu, Chih-Wei, and Chih-Jen Lin."A comparison of methods for multiclass support vector machines.", IEEE transactions on Neural Networks, 13(2):415-425, 2002.
31. Tavallaee, Mahbod, Ebrahim Bagheri, Wei Lu, and Ali A. Ghorbani."A detailed analysis of the KDD CUP 99 data set.", In Computational Intelligence for Security and Defense Applications, 2009. CISDA 2009. IEEE Symposium on, 1-6, 2009.
32. Lippmann, R. P., Fried, D. J., Graf, I., Haines, J. W., Kendall, K. R., McClung, D.,& Zissman, M. A."Evaluating intrusion detection systems: The 1998 DARPA off-line intrusion detection evaluation.", DARPA Information Survivability Conference and Exposition, IEEE, 2:12-26, 2000.
33. Rodda, S., & Erothi, U. S. R."Network Intrusion Detection System to Preserve User Privacy.", In Proceedings of International Conference on Computational Intelligence and Data Engineering. Springer, Singapore, 85-94, 2018.
34. Rodda, S., & Erothi, U. S. R."Class Imbalance Problem in the Network Intrusion Detection Systems.", In Electrical, Electronics, and Optimization Techniques (ICEEOT), International Conference on IEEE, 2685-2688, 2016.
35. Blake, Catherine, and Christopher J. Merz."{UCI} Repository of machine learning databases.", 1998.
36. Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard P fahringer, Peter Reutemann, Ian H. Witten."The WEKA Data Mining Software: An Update."; SIGKDD Explorations, 11(1), 2009.

AUTHORS PROFILE

Mr. E. Uma Shankar Rao presently working as Assistant Professor in RAGHU Institute of Technology, Visakhapatnam, Andhra Pradesh, India. He received his M.Tech (CSE) degree in the year 2013.He received B.Tech (CSE) degree in the year 2007.His research interest Data mining, Image Processing, Data Structures.

Dr. Sireesha Rodda is a Professor in Department of Computer Science and Engineering at GITAM University, Visakhapatnam, India. She has published more than 25 papers in refereed National and International Journals. Her research interests include machine learning and big data analytics.