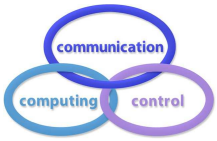
INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 17, Issue: 3, Month: June, Year: 2022
Article Number: 4481, https://doi.org/10.15837/ijccc.2022.3.4481



CCC Publications

Integration of Fuzzy with Incremental Import Vector Machine for Intrusion Detection

R. Arun Kumar, K. Karuppasamy

Arun Kumar Ramamoorthy*
Research Scholar
Faculty of Information and Communication Engineering
Anna University, Chennai, 600025, India
*Corresponding author: arunkramamoorthy@gmail.com

K. Karuppasamy
Professor and Head
Department of CSE,
RVS College of Engineering and Technology,
Coimbatore, 641402, India
karuppasamyrvs@gmail.com

Abstract

IDM design and implementation remain a difficult undertaking and an unsolved research topic. Multi-dimensional irrelevant characteristics and duplicate information are included in the network dataset. To boost the effectiveness of IDM, a novel hybrid model is developed that combines Fuzzy Genetic Algorithms with Increment Import Vector Machines (FGA-I2VM), which works with huge amounts of both normal and aberrant network data with high detecting accuracy and low false alarm rates. The algorithms chosen for IDM in this stage are machine learning algorithms, which learn, find, and adapt patterns to changing situations over time. Pre-processing is the most essential stage in any IDM, and feature selection is utilized for pre-processing, which is the act of picking a collection or subset of relevant features for the purpose of creating a solution model. Information Gain (IG) is utilized in this FGA-I2VM model to pick features from the dataset for I2VM classification. To train the I2VM classifier, FGA uses three sets of operations to produce a new set of inhabitants with distinct patterns: cross over operation, selection, and finally mutation. The new population is then put into the Import Vector Machine, a strong classifier that has been used to solve a wide range of pattern recognition issues. FGA are quick, especially considering their capacity to discover global optima. Another advantage of FGA is their naturally parallel nature of assessing the individuals within a population. As a classifier, I2VM has self-tuning properties that allow patterns to attain global optimums. The FGA-efficacy I2VM model's is complemented by information gain, which improves speed and detection accuracy while having a low computing cost

Keywords: Adversarial attacks, IDS, global optima, Network anomaly discovery and Intrusion detection.

1 Introduction

The internet has quickly become one of our society's primary means of communication. Different kinds of internet applications and use are becoming more accessible. As the number of network apps used grows, so do the security dangers to web users. We must first be able to recognize undesirable or hazardous dangers in order to avert them. As a result, building an intrusion recognition system is a difficult problem in research since network threats have diverse fingerprints and vary on a daily basis. At this time, the intrusion detection system must be capable of detecting fresh assaults.

There are dual types of intrusion recognition system algorithms: unsupervised learning and supervised learning. The supervised training technique learns patterns from supplied data to create a recognition rule/model [1]. Surveillance learning has a low computational complexity and a detection accuracy range. This method, however, only identify known assaults. As a result, it is insufficiently security because there are numerous novel and unknown threats on the network. The unsupervised learning approach is the second type of algorithm. Without any prior training, it can try different attacks. However, it has a large rate of false alerts and a low high detection.

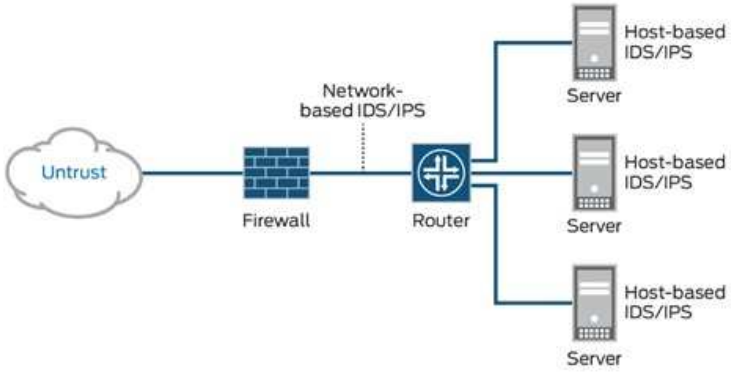


Figure 1: Basic flow of IDS

Current IDS have several flaws; there are numerous IDS programs available on the market and on the internet, but the fact remains that they are not entirely capable of detecting all types of assaults [2]. The most widely used intrusion detection tool is Snort. Snort, which is a at no cost and open-source intrusion recognition and preventive system. It may be used to needs to be transported over a computing network in real time.

Snort is a comprehensive security system that is free and open source. It may be used to transfer data in real time via a computer network. It can study protocols and identify other types of attacks in addition to packet payload analysis.

- Access control for initializers is in some way based on observed data (audit data). When an IDS is in feature extraction mode, it learns to recognize threats from the data it is watching. However, this monitoring data does not guarantee that assaults will occur; it simply indicates the likelihood of an attack [3].
- Second, the number of false positive alarms should be reduced.
- Third, even when there is no intrusion detection, IDS consumes more system resources since it must operate all of the time.
- Fourth, IDS [4] velocity.
- Fifth, precision

In conclusion, the prior study described above did not devote enough attention to the identification of unexpected intrusions. Although some of them considered the unknown assault, they did so using

the KDD99 set, which are almost twelve years of old. The out of current network and the collection lacks several recent attack types. We concentrate on intrusion recognition systems for an unknown type attack in this article, which means the technique may identify new or undiscovered forms of assaults on the internet. For example, a network ID solution should be designed to recognize common network traffic and categorize attack types. We'd want to use the Fuzzy Genetic algorithm to create an IDS method. The fuzzy inference system is a supervised learning technique that use a genetical algorithm to automatically learn new attacks. Furthermore, this method has a high detection rate and is quite reliable. As a consequence, in our online ID system, we employ the fuzzy genetic algorithm in combination with the I2VM technique, which means that data is recognized as soon as it enters the detection system. In terms of detection speed, detection rate, and false alarm rate, we grade our IDS. Section 1 of this article covers the fundamentals of the IDS. The fundamental approaches for Intrusion Detection are summarized in Section 2, Section 3 explains the recommended intrusion detection approach. Section 4 presents the results of the on empirical examination. Finally, the conclusion is reached in Section 5.

2 Related Works

Bagging, boosting, hybrid ensemble, and other machine-learning (ML) methods for network intrusion recognition have widely explored. One of the challenges associated with network intrusion detection is the class misbalancing between regular and attack traffic. As a result, more robust techniques to network intrusion detection are required to provide constant high performance across diverse classes. Various data resampling techniques have been used to resolve class imbalance for enhancing network intrusion detection, including under sampling of the majority class [5], oversampling of the minority, and a mix of under sampling and oversampling [6]. On the other hand, the optimum class distribution is unknown, therefore the normalization ratio is usually subjective and must be adjusted to improve recognition accuracy.

The oversampling method produced over fitting for the minor classes with poor range of recall and elevated accuracy values, whereas the underneath sampling approach resulted in performance failure on the classification model, according to the authors [7]. At the data level, resampling approaches are frequently regarded as a way to minimize class imbalance. A range of techniques based on the widely used overall accuracy AdaBoost have also been designed to tackle class misbalance at the algorithmic level. Despite its prominence in machine learning, AdaBoost is widely recognized for its inability to properly solve class imbalance. AdaBoost is a classification performance-improving method that combines many fragile learners into a single well-built learner. This adjusts sample values based on classification errors, raising the weights of misclassified samples while reducing the weights of well correctly classified. As a result, classifiers that pay greater attention to misclassified data rather than minority class examples are chosen. Since AdaBoost is concerned with classification accuracy, the method is skewed toward the majority class, which contributes more to total classification accuracy [8]presented the semi-boosted layered model after demonstrating that AdaBoost alone does not perform effectively for network intrusion data. SMOTEBoost, a variant of AdaBoost oriented on combining the SMOTEBoost algorithm with the boosting approach for learning the imbalanced dataset in network intrusion (2003). The appropriate SMOTE sampling ratio, on the other hand, must be established. RareBoost is an improvement on AdaBoost in which dissimilar treatments were applied to the positive (+) and negative (-) predictions in order to better identify rare classes. If the true positive rate is more than the genuine alarm rate and the negative result rate is greater than the false negative rate, the RareBoost algorithm will collapse. In our early RareBoost tests, we observed that such a limitation is a strong requirement that our data has yet to satisfy. Another approach to class imbalance learning is cost-sensitive boosting. The authors suggested using auto-encoder to extract features from datasets, lower feature dimensions, and therefore greatly reduce memory needs, while also improving network threat detection in [9] and [10]. The automated encoder, on the other hand, was not employed for anomaly detection. Finally, to detect network risks, the authors utilized classifiers. Our solutions were unsupervised, whereas theirs were supervised, needing specialists to label the training dataset. The attack kinds of the set NSLKDD are listed in the Table 1.

Table 1: Attack types of NSLKDD set

Attack	Description
R2L	By guessing the password, a remote to local (RTL) attack gets local admittance to a system from an isolated position.
Probe	Probe attack gathers information about the system in order to exploit vulnerabilities that have been found.
U2R	A user to root (UTR) attacks gain local way into a system and develop flaws in the scheme to get right roots.
DOS	A DOS attack depletes system resources, denying or delaying legitimate users' and requests' access to resources.

The main benefit of utilizing an integrated approach in IDS is that false positive alerts are reduced. Many academics have conducted extensive research to demonstrate the benefits of combining pattern matching with machine learning. The multiple pattern matching system based on trained classifier models has emerged as a result of this combination.

3 Proposed Methodology

This article proposes a novel intrusion detection method based on FGA-I2VM to identify the DOS, Probing, R2L, and U2R types of attack classes. Using the NSLKDD dataset, the suggested method is implemented and validated.

3.1 FGA-I2VM based Model for IDS

Figure 1 depicts the suggested FGA-I2VM model's design. The architecture is divided into two phases: training phase and the testing. The NSL KDD dataset is utilized in the training phase. Data pre-processing with feature selection is handled by the Information Gain technique, fresh population creation is handled by GA, and classification is handled by the I2VM acts a classifier, which aids in the detection of DoS attacking patterns. The second stage is the analysis phase, in which the recorded traffic is analyzed, similarities are discovered and compared against storage DoS attack patterns in the database, and a decision is taken. New patterns are discovered by analyzing traffic behavior; if the behavior is illegal, the patterns are collected and updated in the database.

A. Pre-processing, B. Feature Selection, C. New population generation, and D. Classifier are the four steps of the implementation of Genetic-I2VM based IDS. The suggested FGA-I2VM IDS model is depicted in Figure 2 as a whole.

A. Pre-processing Phase

All of the recorded data must be arranged in a specific structure or pattern for categorization purposes, and this entire process is known as pre-processing. The I2VM classification system cannot handle the NSL KDD DS in its current set-up. As a result, pre-processing is necessary before constructing an I2VM classification system. Pre-processing includes the following steps:

- Mapping emblematic features to some numerical value.
- Converting our feature values to 0 and 1 using min-max normalization.
- The titles of attacks were classified into one of two categories: regular or DoS. (Denial - of - Service).

B. Normalization

Various characteristics are on different scales, which is an issue with normal data. This results in a preference for certain traits over others. To overcome this challenge, we use the training dataset's distribution to transform the data instances to a standard format. That is, we assume that the training dataset properly represents the range and variance of feature values over the whole distribution.

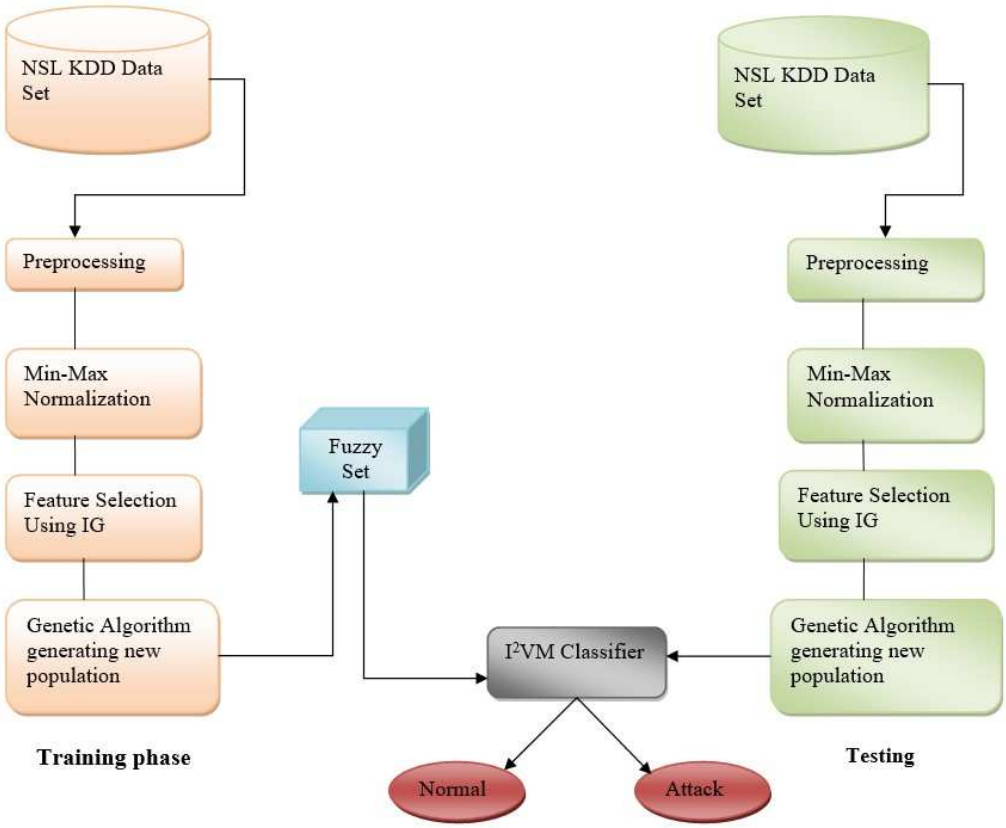


Figure 2: The FGA-I2VMmodel

The data in the range of 0 and 1 is likewise converted during normalization. In normalization, we first choose the maximum and minimum values in a column, then conduct the Normalization procedure.

C. Feature Range using IG

If all of the dataset's accessible attributes are picked, intrusion detection will be ineffective due to the extended training and testing time and low detection accuracy. The classifier's accuracy is determined by the selection of the best feature subset. As a result, only relevant qualities must be extracted for training and testing. The feature collection method is mostly used to choose a compartment of features from a larger dataset. Filter and wrapper techniques are two often used feature selection strategies. The filter approach relied mostly on generic characteristics of data features rather than machine language. These characteristics are ranking according to a set of criteria, with the highest-ranking traits being chosen as the best. The primary benefits of the filter technique are its cheap computing cost and the fact that it does not require any machine language algorithms for future range. The Information gain method, which is used as the feature selection mechanism in the proposed FGA-I2VM model, is a frequently used filter method.

Information gain (IG) measures the volume of information in bits[11] about the model training when the only data provided is the existence of a characteristic and the corresponding class distribution.

The decrease in uncertainty regarding the value of A after seeing values of B is the information gain (IG) of a given quality B with respect to the class quality A. The Equation determines the amount of information gained in Equation (1).

$$IG = A|B \tag{1}$$

When A and B are distinct variables with ranges in a1....ak and b1....bk, the uncertainty about A's value is quantified by its entropy in Equation (2).

$$H(A) = -\sum_{i-1}^{k} p(a_i) log_2(p(a_i)) \tag{2}$$

Where (ai) is the prior probability for all A values. The conditional entropy of A given B expresses the degree of uncertainty regarding A's value after witnessing B's values in Equation (3).

$$(A|b) - \sum_{j-1}^{n_1} p(b_j) \sum_{i-1}^{k} p(a_i|b_j) log_{n(p(a_i|b_j))} \tag{3}$$

Where (ai|bj) denotes the posterior probability of A based on the values of B. As a result, the information gain is provided by Equation (4).

$$IG(A|x) = H(A) - H(a|b) \tag{4}$$

If IG (A | B) > IG (A | Z), an attribute B is considered to be more linked to class A than an attribute Z. We may rank the correlations of each feature to the class and select key qualities based on the expected information gain.

Table 2: Selected Features from NSL KDD using IG

Attribute	**Description**	**Type**
Protocol	Protocol type (e.g., TCP, UDP, etc.)	separate
Source bytes	The amount of data bytes sent from the source to the end node	nonstop
Service	The number of times the same service has been requested from the client to the server	nonstop
Host Count	In the last two seconds, the number of links to the same host as the current link has increased.	nonstop
Flag	Set the Flagvalue=1 or Set_Flagvalue=0.	nonstop

The 41 properties of the NSL KDD dataset are reduced to 5 traits that are related to DoS attack features, saving 87.8% of feature space. Information is gathered from a selection of data sets, and Table 2 displays the selected characteristics and their descriptions from NSL KDD.

D. Generating new population with FGA

The characteristics chosen by the Information Gain technique are sent into the genetic algorithm as input. To increase classification accuracy, GA produces a fresh population set for training the I2VM classifier. Normal data (A) and assault data (B) make up the population (B). The issue domain is transformed into sets of chromosomes like data structure since GA exclusively works with chromosome like data structure. GA uses an objective function to calculate population fitness. If the training recording matches each gene on the chromosome in the predefined normal chromosomal set, the normal variable 'Y' is increased by one. If the result of 'Y' was initially '0,' the value of 'Y' becomes '1' after matching the training record with the normal population, and if matches. If the training plan don't match the regular population, they're compared to the attack population, and parameter "X" is raised by one. Equation (5) is used to determine the consolidated fitness value.

$$F = \frac{(Y|N + X|N_{range})}{2} \tag{5}$$

The total number of records in the data collection is N.

If the fitness threshold falls below the 0.5 threshold, three activities were carried out: selection, cross-over, and mutations. The crossovers stage follows, in which the second season of the first chromosome is joined to the first half of the second chromosomes. The one-point cross over concept is the foundation of the cross over. Mutation is the last stage, which includes randomly altering the bit position of the training set. The fitness function is reassessed, and if the objective is met, the operation is halted, followed by mutation and crossover. The algorithm for generating a new population via GA is found in following Algorithm.

Algorithm 1: Generating new population

1. Create a population from pre-processed data.
2. For each unique pre-processed data, calculate an objective function based on determined DoS attack principles.
3. Individualized solution selection.
4. Mating of a couple of populaces.
5. Execute mutation operation
6. Determine the objective function for the newly generated population.
7. If it is satisfied, discontinue the operation.
8 If not, replicate step 3.
9. return the best features from the NSL KDD set that indicate DoS attack characteristics.

3.2 Fuzzy Ruleset: Derivative of NSL KDD

Following selection, cross-over, and mutations, GA produce a new set of population, and the fresh batch of patterning or rule set created for six types of DoS assaults in the NSL KDD dataset is obtainable in Table 3.

- The attack anticipated is a Smurf type if the protocol type is ICMP, the service type is ecr_i, the source byte is 1032, Set_Flag is 1, and the Host Count is 512.
- It's a Neptune attack if the protocol type is TCP, the service type is private, the source byte is 0, the Set_Flag is 1, and the host count is 160.
- A Back type attack is identified if the protocol type is TCP, the service type is http, the Set_Flag is 1, the source byte is 54540, and the host count is 160.
- A Teardrop assault is one in which the protocol type is UDP, the service type is UDP, the Set Flag is 1, F, the source byte is 28, and the host count is 255.
- The predicted attack type is Land if the protocol type is TCP, the service type is finger, the set flag is 1, the source byte is 0, and the host count is 211.
- The attack type is considered aPod if the protocol type is ICMP, the service type is secr_i, the Set_Flag is 1, the source byte is 1480, and the host count is 255.
- If conditions other than those listed above occur, the data is classified as abnormal.

To construct a detection rule in our system, we employ a trapezoidal shape, as shown in Figure 3. The a, b, c, and d are four parameters that make up the trapezoidal form. The method uses the parameters listed below to determine the likelihood of being attacked.

```
Fuzzy rule set

If (range>a) && (range<b) {
probability = range – a/(b-a)
}
else if (range>=b and range<=c) {
probability = 1;
}
else if (range>c and range<d) {
probability = d – range/(d-c)
}
else {
probability =0
}
```

$$\mu_A(x) = \begin{cases} 0, & x \le a \\ \dfrac{x-a}{b-a}, & a < x \le b \\ 1, & b \le x \le c \\ \dfrac{c-x}{d-c}, & c < x \le d \\ 0, & x \ge d \end{cases}$$

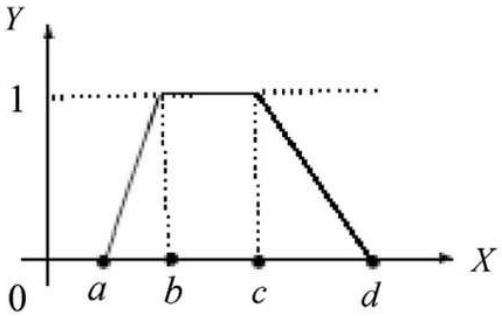


Figure 3: Fuzzy set based on trapezoidal form

Table 3: Fuzzy Rule Structure of NSL KDD DS

S. No	Attack Description	Attack Type
1	Pprotocol=ICMP, flag=SF, host_count=512, Service=ecr_i, src_byte=1032.	Smurf
2	protocol=TCP, src_byte= 0, service=private flag=SF, host_count=160	Neptune
3	Protocol=TCP, flag=SF, src_byte=54540, service=http, host_count=160	Back
4	protocol=UDP, service=UDP, flag=SF, src_byte=28, host_count=255	Teardrop
5	protocol=TCP, service=finger, flag=SF, src_byte=0, host_count=211	Land
6	Protocol=ICMP, flag=SF, src_byte=1480, service=ecr_i, host_count=255	Pod

3.3 I2VM Classification

I2VM area category of supervised learning methods based on statistical learning methods that can be used for classification or regression. I2VM has proven to be effective in a variety of sectors. Because of its adaptability and capacity to operate effectively with high-dimensional data while avoiding the dimensionality problem, I2VM has gained a lot of interest as a classification approach from a variety of research fields [12]. I2VM was created for binary classification with the goal of constructing an ideal or maximum hyper plane that maximize the margin of division between the negative and positive data sets.

To categorize the data, a variety of hyper planes can be employed. The optimal hyperplane is the one that shows the largest separation, or margin, between the two classes. So, we picked the hyperplanes so that the distance between it and the nearest data point on either side is as little as possible. An I2VM is a program that converts linear algorithms into non-linear space. A feature known as the kernel function is used to accomplish this mapping. Kernel functions such as polynomial and RBF are used to partition the feature space by constructing a hyper plane. The kernel functions may be used to pick support vectors along the surface of this function during the training of classifiers.

I2VM categorizes data using these support vectors, which define the hyperplane in the feature space. This technique will utilize a quadratic programming problem, with the result being a global optimal solution. Assume that N training datapoints (b1, a1), (b2, a2), (b3,a3)...(bN,aN) exist, where xi€Rd and yi€+1,1. Consider the plane (w, b), where w - weight vector and b is the bias. Categorization of a fresh item x has been carried by using Equation (6),

$$f(x) = signal(w.b + b) = signal(\sum_{i}^{N} \alpha 1_i a_i (b_i b1) + b) \tag{6}$$

Only a dot product may be used to represent the training vectors. There is a Lagrangian theory for every training end. The significance of each data piece is represented by the Only the places closest to the plane will have I 0, and these points will be remembered as support vectors after the maximum margin hyper-plane has been found i=0 will be used for all other points. That is, only the points closest to the hyperplane are used to represent the hypothesis/classifier.

A. The main steps of the I2VM Classifier

Step 1: Establish a new enhanced chromosomal population, which is a group of humans with different chromosomes produced by GA. Each person's chromosome is made up of a collection of pre portioned parameters. The user must correctly calculate the pattern weight of the starting population in order to include as many alternative solutions as feasible.

Step 2: Using Equation (7), calculate the fitness range of each member in the initial population and rank them accordingly.

$$F(x) = \sum_{i=1}^{n_1} q_i.x_i + b \tag{7}$$

Where 'Q' is weight vector, 'b' is bias range, and 'n' is features count. They assigned a weight value to each attribute and rated them accordingly. The most essential qualities are those with the greatest weight values, and they are used in the detection method.

To determine the fitness value of an individual or chromosome, the learning record is matched to each gene of the chromosome in the regular range. Different pattern values for different feature values are generated by each and every record. Similarly, each gene on the chromosome in the assault population is compared to the training record. As a result, each record creates a unique pattern value for each feature value. Finally, using pattern weight for normal and attack populations, we will compute support vector values for both normal and attack patterns. We'll now obtain two I2VM values: 1 and 0.F>=1 denotes a normal record, whereas F1 denotes an attack record. Our dataset is classified by I2VM using freshly created hyperplane values. Now each and every testing record is compared with each and every gene of the normal and attack population. This will provide a pattern value of 0,1,2,3,4,5, based on which it will be determined if our testing data corresponds to the normal or attack model.

4 Simulations and Outcome Discussions with NSL KDD Datasets

We utilized an AMD Athlon TM64X2 Dual Core Processor 6000+2.59 GMHz computer with 4GB RAM, running on Windows Operating System.MatLab7.2 was used to test the proposed FGA-I2VM model. During the testing, 10% of the NSL KDD labelled data was utilized to train the suggested FGA-I2VM model. The NSL KDD dataset comprises three categories of traffic and six types of DoS attacks, totaling roughly four terabytes, with each traffic record containing 41 characteristics that help distinguish the category as normal or assault. In the 10% tagged NSL KDD dataset, there are 97,277 usual and 3, 91,458 Denial assaults traffic records. There are 2,80,790 Smurf records, 1,07,201 of Neptune recordings, 2,203 of Back records, 979 of teardrop, 21 land, and 264 attacks on pod among the 3,91,458 of DoS attack records. After eliminating repeated occurrences, the traffic reports examined for training the suggested IDs are 97277 of usual, 641 of Smurf, 51820 of Neptune, 994 of return, 19

of land, 918 of teardrop, and 206 of pod. The traffic reports chosen for testing are 60255 of usual, 400 of Smurf, 20500 of Neptune, 714 of back, 300 of land, 7 of teardrop, and 101 of pod.

4.1 Performance evaluation with NSL KDD datasets

The precision, recall, and F-measure of the proposed FGA- I2VM model are computed using the confusion matrix in Table 4 to determine its detection accuracy. The requirements outlined below are valid for both the NSL and KDD datasets. True_negatives values (TNV), True_positives values (TPV), False_positives values (FPV), and False_negatives Values (FNV) are the four potential prediction results shown in the confusion matrix (FN).

Table 4: Confusion Matrix (Predictor Class)

	Usual	**Attacked**
Usual	True_Positivesvalues (TPV)	False_Positivesvalues (FPV)
Attacked	False_Negativesvalues (FNV)	True_Negativesvalues (TNV)

where,

- True_positives values (TPV): the number of usual occurrences that have been successfully labelled as normal.
- False_positives values (FPV) are the count of regular occurrences that are mistakenly identified as assaults.
- False_negatives values (FNV): The count of attack occurrences is projected wrongly as usual.
- True_negatives values (TNV): The attack's number of occurrences is accurately anticipated.

The following performance metrics were calculated as a result of these:

$$Recallvalue = \frac{TPV}{TPV + FNV}$$

$$Precisionvalue = \frac{TPV}{TPV + FPV}$$

IDS should attain a high recall value without sacrificing accuracy, where F-measure values is a weighted mean that evaluates the swapping.

$$Recallvalue = \frac{2 + Recallvalue + Precisionvalue}{Recallvalue + precisionvalue}$$

$$OnthewholeAccurateness = \frac{TPV + TNV}{TPV + TNV + FNV + FPV}$$

Table 5: Accuracy comparison

Iterations	**5**	**10**	**15**	**20**	**30**
FGA-I2VM	81.88	86.97	90.78	92.57	95.78
Single SVM	80.32	76.74	84.96	81.4	76.761
STL-IDS	65.24	69.65	71.23	76.87	80.48

Table 6: Precision comparison

Iterations	**5**	**10**	**15**	**20**	**30**
FGA-I2VM	75.24	86.97	92.56	95.66	100
Single SVM	71.80	78.12	86.55	90.47	92.98
STL-IDS	76.8	80.47	87.41	92.58	93.92

Table 5 depicts the accuracy anticipated FGA-I2VM with prevailing models like single SVM and STL- IDS. The experimentation of the anticipated FGA-I2VM model is done with 5 to 30 iterations.

Table 7: Recall comparison

Iterations	5	10	15	20	30
FGA-I2VM	65.98	76.11	81.52	86.47	89
Single SVM	50.78	55.68	58.12	60.25	61.84
STL-IDS	51.02	56.77	60.85	63.99	68.28

Table 8: F-measure comparison

Iterations	5	10	15	20	30
FGA-I2VM	72.58	79.56	82.98	89.92	94.17
Single SVM	55.58	60.04	62.89	69.87	74.28
STL-IDS	59.22	63.22	67.58	72.45	79.07

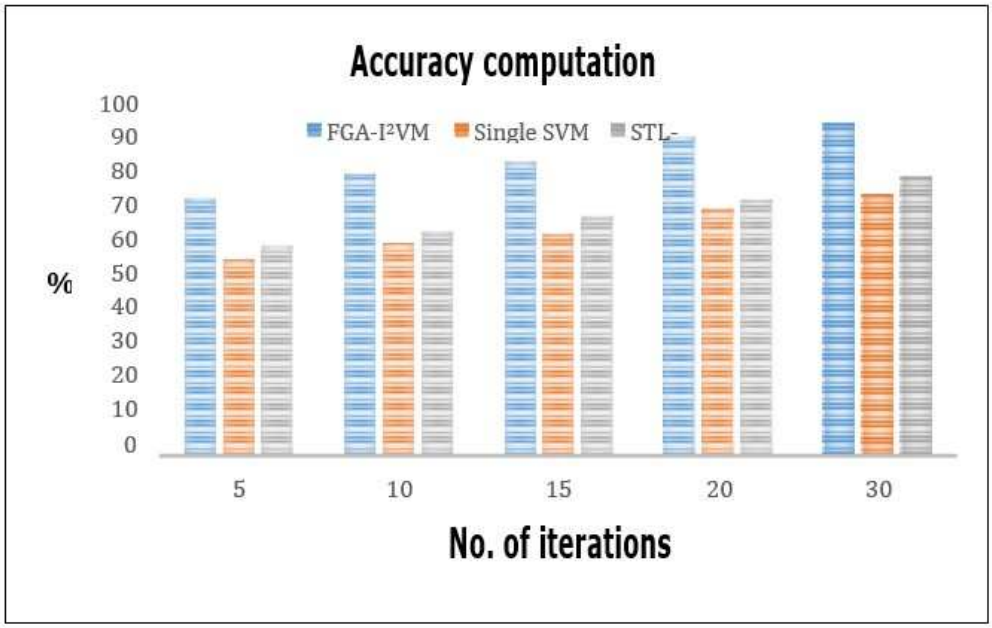


Figure 4: Accuracy computation

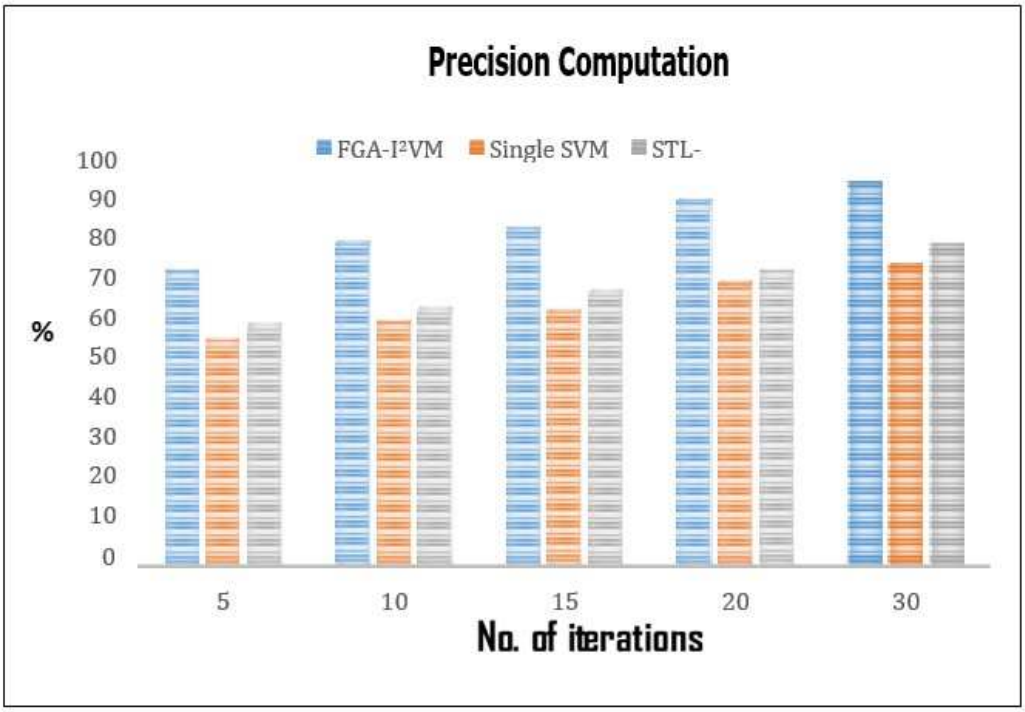


Figure 5: Precision computation

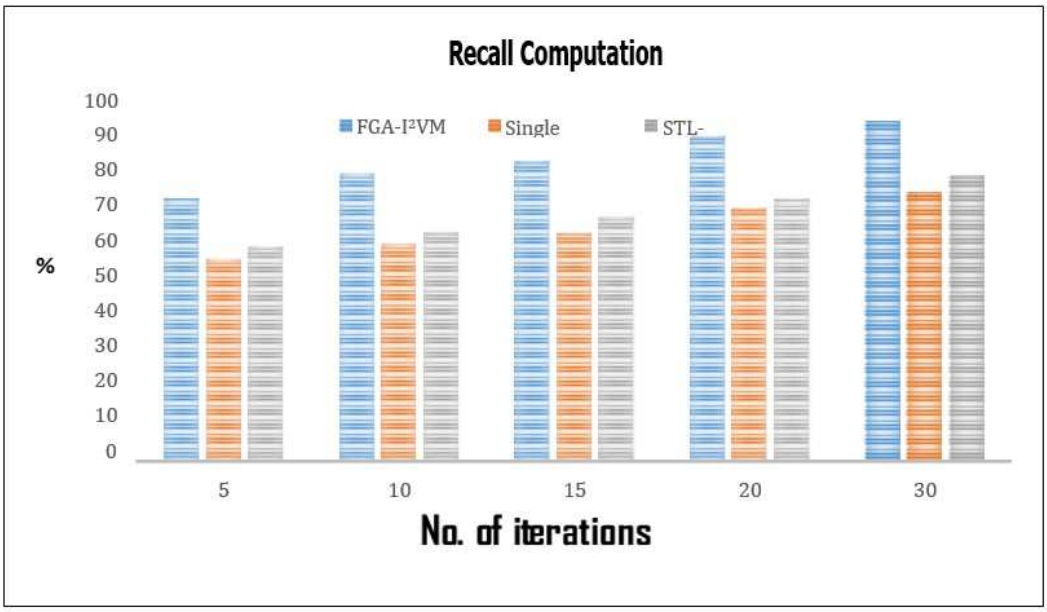


Figure 6: Recall computation

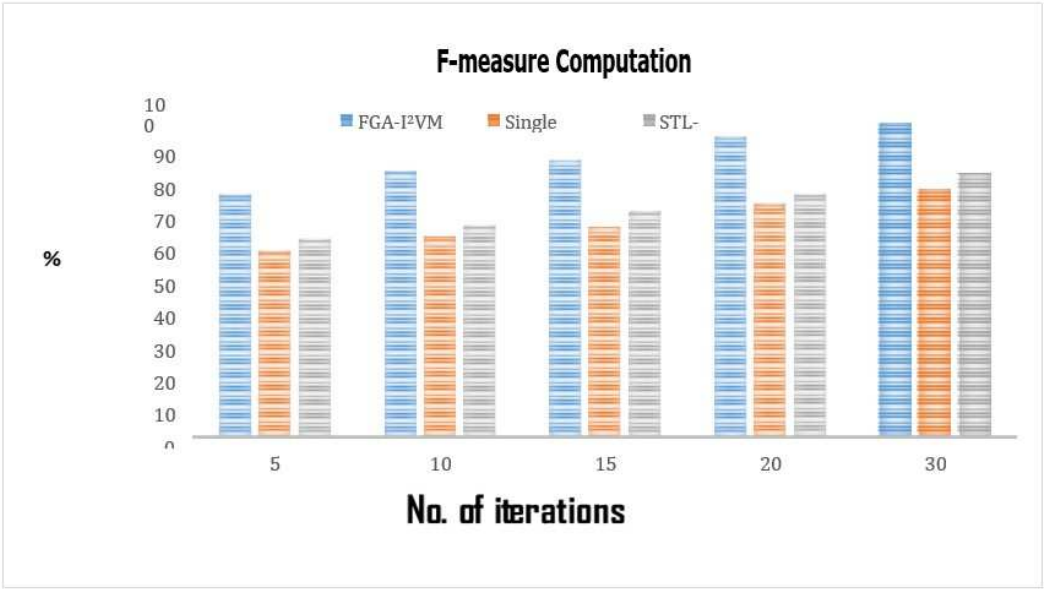


Figure 7: F-measure computation

The accuracy of FGA-I2VM over 30th iteration is 95.78% which is higher than other approaches (See Figure 4). Table 6 depicts the comparison of precision where the FGA-I2VM shows 100%, which is 7.02% and 6.08% higher than single SVM and STL-IDS, respectively (See Figure 5). Table 7 depicts recall where the FGA-I2VM model shows 89%, 27.66% and 20.72% higher than SVM and STL- IDS, respectively (See Figure 6). Table 8 depicts the comparison of F-measure where the FGA-I2VM model shows 94.17%, which is 17.77%, 19.89%, and 15.1% higher than single SVM, and STL- IDS, respectively (See Figure 7). The above analysis shows that the anticipated model works in predicting the intrusion over the network model using learning and IVM models.

5 Conclusion

Genetic algorithms are quick due to their capacity to discover global optima, and another benefit is their intrinsically parallel nature of assessing individuals within a population. The self-tuning properties of the I2VM as a classifier allow the patterns to achieve global optimum. The FGS-I2VM model's efficacy is complemented by information gain, which improves speed and detection accuracy while having a low computational cost. This method is extremely effectual for identification of various attacks and showed superior detection accuracy when compared to other models. The testing findings show that the FGA-I2VM model has a very excellent detection accuracy of 99.49% and a 0.51 percent false alarm rate against NSL KDD, making it one of the good detection approaches in IDS. Similarly, zero-day attacks should also be analyzed as it shows huge impact towards real-time applications. In future, deep learning classifier or ensemble classifier is used for enhancing prediction accuracy.

Funding

No funding is involved in this work.

Author contributions

The authors contributed equally to this work.

Conflict of interest

The authors declare no conflict of interest.

References

[1] Al Shahrani, B. M. M. (2021). Classification of Cyber-Attack using Adaboost Regression Classifier and Securing the Network, *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(10), 1215-1223.

[2] Sallam, A. A., Kabir, M. N., Alginahi, Y. M., Jamal, A., and Esmeel, T. K. (2020). IDS for Improving DDoS Attack Recognition Based on Attack Profiles and Network Traffic Features, In *16th IEEE International Colloquium on Signal Processing and Its Applications (CSPA)*, IEEE, (pp. 255-260).

[3] Mazini, M., Shirazi, B., and Mahdavi, I. (2019). Anomaly network-based intrusion detection system using a reliable hybrid artificial bee colony and AdaBoost algorithms, *Journal of King Saud University-Computer and Information Sciences*, 31(4), 541-553.

[4] Yulianto, A., Sukarno, P., and Suwastika, N. A. (2019). Improving adaboost-based intrusion detection system (IDS) performance on CIC IDS 2017 dataset, In *Journal of Physics: Conference Series*,IOP Publishing, Vol. 1192, No. 1, p. 01201.

[5] Mikhail, J. W., Fossaceca, J. M., and Iammartino, R. (2019). A semi-boosted nested model with sensitivity-based weighted binarization for multidomain network intrusion detection, *ACM Transactions on Intelligent Systems and Technology*, 10(3), https://doi.org/10.1145/3313778.

[6] Akashdeep, I., Manzoor I., and Kumar, N. (2017). A feature reduced intrusion detection system using ANN classifier, *Expert Systems with Applications*, 88, 249–257, https://doi.org/10.1016/j.eswa.2017.07.005.

[7] Li, K., Zhou, G., Zhai, J., Li, F., and Shao, M. (2019). Improved PSO_AdaBoost ensemble algorithm for imbalanced data, *Sensors (Switzerland)*, 19(6), https://doi.org/10.3390/s19061476.

[8] Shahraki, A., Abbasi, M., and Haugen, Ø. (2020). Boosting algorithms for network intrusion detection: A comparative evaluation of Real AdaBoost, Gentle AdaBoost and Modest AdaBoost, *Engineering Applications of Artificial Intelligence*, 94, 103770.

[9] Han, K., Wang, Y., Zhang, C., Li, C. and Xu, C. , (2018). Autoencoder inspired unsupervised feature selection, In *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*,IEEE, pp. 2941–2945.

[10] Wang, C., Zheng, J., and Li, X., (2017). Research on DDOS attacks detection based on RDF-SVM, In *Proceedings of the 2017 10th International Conference on Intelligent Computation Technology and Automation (ICICTA)*,IEEE, pp. 161–165.

[11] Revathy, G., Kumar, P. S., and Rajendran, V. (2021). Development of IDS using mining and machine learning techniques to estimate DoS malware, In *International Journal of Computational Science and Engineering*, 24(3), 259-275.

[12] Jianjian, D., Yang, T., and Feiyue, Y. (2018). A novel intrusion detection system based on IABRBFSVM for wireless sensor networks, In *Procedia computer science*, 131, 1113-1121.

Journal's webpage: http://univagora.ro/jour/index.php/ijccc/

This journal is a member of, and subscribes to the principles of,
the Committee on Publication Ethics (COPE).
https://publicationethics.org/members/international-journal-computers-communications-and-control

Cite this paper as:

Arun Kumar, R.; Karuppasamy, K. (2022). Integration of Fuzzy with Incremental Import Vector Machine for Intrusion Detection, *International Journal of Computers Communications & Control*, 17(3), 4481, 2022.
https://doi.org/10.15837/ijccc.2022.3.4481