How to Cite:
Soundarraj, K., & Ravichandran, M. (2022). Enriched intuitionistic fuzzy rule pruning using butterfly optimization for rule pruning in intrusion detection. *International Journal of Health Sciences*, *6*(S6), 10574–10584. https://doi.org/10.53730/ijhs.v6nS6.9638

Enriched intuitionistic fuzzy rule pruning using butterfly optimization for rule pruning in intrusion detection

K. Soundarraj
Department of Computer Science, Sri Ramakrishna Mission Vidyalaya College of Arts and Science, Coimbatore, India
*Corresponding author email: soundarraj28@gmail.com

M. Ravichandran
Department of Computer Science, Sri Ramakrishna Mission Vidyalaya College of Arts and Science, Coimbatore, India

***Abstract*---**The process of detecting and reacting to hostile activities against the computer and network resources is known as intrusion detection. Rather of relying just on expertise and experience, a systematic and automated IDS creation process is required. This drives researchers to look into data mining-based intrusion detection frameworks. Though there are many algorithms are developed for detection intrusion, the issue of inconsistency in rule-based classification model is controlling the volume of data and removing irrelevant rules are the major problems focused in this work to improve the process of IDS. This paper proposed a framework which handles the hesitancy in classifying the unknown traffic pattern data packets by designing evolutionary algorithm optimized intuitionistic fuzzy inference system. The Intuitionistic fuzzy inference system represents each instances of KDD cup 99 dataset with the degree of membership and non-membership. With these two degrees, the hesitancy degree is computed to overcome the issue of inconsistency and uncertainty in analysing the unknown pattern of abnormal packets. In conventional Intuitionistic fuzzy System (IFS), the rules generated by the inference engine is applied for classification directly without determine the validation of it. This results in overfitting and affects the accuracy rate of the classification model. Hence, in this proposed work eliminating the redundant rules, determining the rule strength and utilizing only the best set of rules are main objective. It is accomplished by adopting Butterfly Optimization Algorithm (BOA), which balances both local and global searching capability and evaluates the fitness value of all rules generated by IFS and prunes the least fittest value rules to achieve highest accuracy rate in Intrusion detection. The simulation results explored that IFS-BOA

produced highest detection rated by controlling the volume of rule set compared to the other classification models.

***Keywords*---**intrusion detection system, inconsistency, uncertainty, intuitionistic fuzzy system, butterfly optimization, rule pruning.

Introduction

With the fast expansion of the computer network, the necessity of data security is becoming increasingly important. The term "security" refers to the level of protection provided to a network or system [1]. Intrusion refers to any collection of hostile behaviour aimed at jeopardizing the information's security aims [2]. One of the most significant issues in information security is intrusion detection. Many machine learning approaches are merged with NIDS to detect abnormal traffic patterns automatically, due to the proliferation of artificial intelligence [3]. Data mining has shown to be a viable method of actively detecting invasive attacks using rules concealed in network activity data. Intrusion detection, on the other hand, typically distinguishes between normal activity, known intrusions, and unknown intrusions, which may be viewed as a classification technique [4]. The descriptive and predictive rule learning are the two types of rule learning approaches. Predictive rule learning develops rules capable of generalizing to new data samples, whereas descriptive rule learning focuses on just discovering patterns in a current dataset without considering assessment on new data instances.

In order to identify intrusions, predictive rule learning is probably largely used. Predictive rule learning, unlike descriptive rule learning, frequently encounters two sorts of issues: several rules fire on the same new example, and no rule fires on a new example. In the first situation, several rules firing on a single example might result in a conflict, which is addressed either by prioritising rules of greater relevance or by extracting a new rule set for handling opposing predictions. Top-level control parameters are utilized to address rule conflicts, much as they are in expert systems. The second problem is solved using either a default rule that favors the majority class or more advanced algorithms that identify the appropriate rule. To overcome the issue of determining rule correctness, coverage, eliminating redundant rule and updating the rules, in this research work a novel uncertainty-based inference model is designed and its performance is enhanced by adopting evolutionary algorithm which involves in selection of best set of rules. The intuitionistic fuzzy inference model is used for generating the rules and the rule pruning is accomplished by butterfly optimization algorithm which selects the optimized rules depending on their fitness value.

Related Work

Aburomman et al [5] in their survey state about the importance of machine learning algorithms, like classification algorithms, hybrid methods and ensemble algorithms. The voting methods are simpler to implement and it produce more favorable results during intrusion detection. Kabir et al [6] devised an optimal allocation least square SVM for detecting intrusion in two different stages. The

first step involves in partitioning the dataset into subgroups depending on the variance of observation. To extract samples to identify intrusion least square SVM. Nannan et al [7] constructed an evolving technique to excerpt rules for intrusion detection system. The qualified rule sets are extracted to control the quantity and diversified rules. Using the distance measure the rules set of the same class and different classes are examined using the genetic algorithm. Mehr et al [8] in their work developed a DDOS attack detection model using support vector machine. They used flows in switches and focused on detecting ddos with specific interval of time. The SVM often faces overfitting problem when the traffic flow is high and in this work Ryu controller is used in tree network topology.

Elhag et al [9] constructed a multi-objective fuzzy system which uses various metrics for training process. While enlarging the search space the ids process can be optimized. The fuzzy system infers the pattern of KDD cup 99 dataset in terms of degree of truthiness alone. The C4.5 decision tree is also involved in detection of IDS. Bridges et al [10] performed a detail survey on leverage host-based data sources for identifying attacks on commercial network. In this work a targeted sub survey of host-based intrusion detection and using publicly available system calls are considered in this study. Alzahrani et al [11] in their work explained about importance of machine learning in monitoring of network traffic to analyze the malicious behavior inside the network in Software designed network controller. Both classical and advanced tree-based classifiers are deployed for IDS.

Varanasi and Razia [12] designed an ensemble model which combines many week learning algorithms to enhance the intrusion detection model. When the volume of dataset grows, the popularity and need of deep learning techniques are used in real time applications. For limited dataset, transfer learning is well suited in intrusion detection system. Latif et al [13] introduced an innovative light weight random neural network is used for predicting cyber security attacks such as probing, denial of service, spying and malicious control. The neural network is enriched with the training dataset and detects the unknown patterns of abnormal packets. Khan et al [14] constructed an ensemble-oriented voting model which integrates different classifiers to predict the IDS. The voting model improves the effectiveness of the testing phase for both binary and multi-class attack detection.

Proposed Methodology

Intuitionistic Fuzzy Rule Pruning using Butterfly Optimization

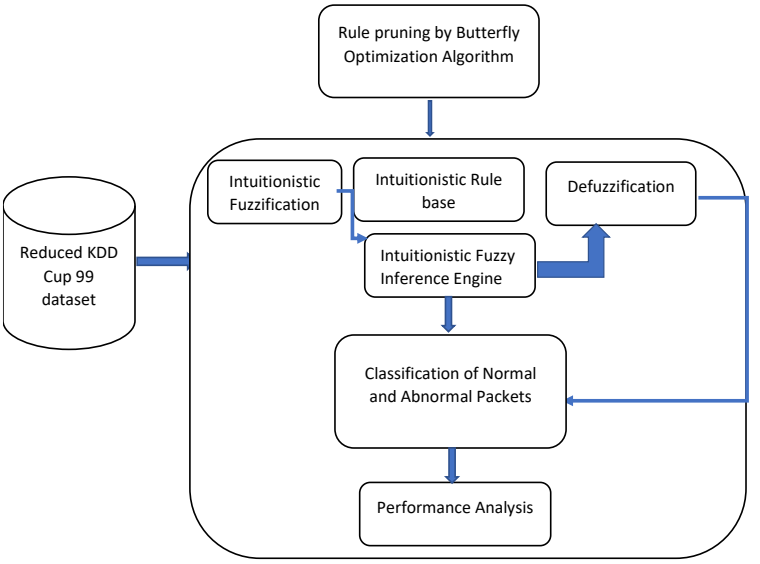

Figure Workflow of Intuitionistic fuzzy Inference System optimized by Butterfly Optimization Algorithm

In this proposed methodology to handle uncertainty in rule-based model that involves in prediction of normal and abnormal packets, an enriched intuitionistic fuzzy rule classifier is developed and its work flow is shown in the figure. The KDD cup 99 [15] dataset is used in this work for intrusion detection. With the significant subset of features obtained using neutrosophic correlation, the reduced feature set is passed as input to the Intuitionistic fuzzy Inference system. The crisp value of input is converted into intuitionistic fuzzy values by representing them as grade of membership, non-membership and hesitation. The Intuitionistic knowledge base obtains the pattern of instances and produces the set of rules to classify the incoming packets as normal or malicious. The generated rules are pruned by applying the Butterfly optimization algorithm which searches for best rules that improves the accuracy rate of IDS and minimize the false detection.

Intuitionistic Fuzzy Inference System (IFS)

The Intuitionistic Fuzzy [16] is the generalization of fuzzy logic that assess the elements by the two function namely membership and non-membership, which belongs to the interval value [0,1]. The Intuitionistic Fuzzy incorporates hesitation margin which is represented as hesitation degree which is to be considered in real time applications due to uncertainty and vagueness. The hesitation degree is defined as one minus sum of membership and non-membership degrees respectively. In IFS, the given input is transformed into intuitionistic domain which defines each input into corresponding memberhsip

and non membership functionality. After receiving the membership and hesitacy or non-membership functions, their updation is carried out depending on the desired inference operation [17]. The overall framework of intuitonistic fuzzy inferesence system as shown in the figure.

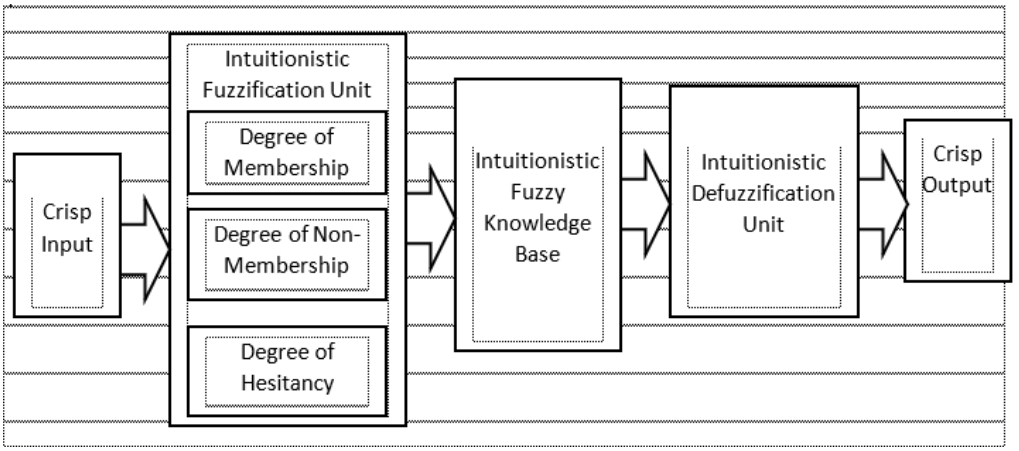

Figure Overall Workflow of Intuitionistic Fuzzy Inference System

Steps involved in Intuitionistic Fuzzy Inference System

- Fed the KDD cup 99 dataset as input value which is in the crisp form into the IFS system
- The intuitionistic fuzzification process transforms the crisp value into membership and non-membership representation grades.
- To infer information about input values and construct intuitionistic fuzzy rules, an intuitionistic fuzzy knowledge base is used.
- The output value is determined by the rules generated. Converting intuitionistic value to crisp value requires defuzzification.

Sample rules generated by Intuitionistic fuzzy Rule generator is as follows:

- IF src_bytes is low and dst_bytes is low and srv_count is low diff_srv_rate is low and dst_host_srv_count is low then packet is normal
- IF src_bytes is low and dst_bytes is medium and srv_count is low diff_srv_rate is medium and dst_host_srv_count is low then packet is normal
- IF src_bytes is medium and dst_bytes is medium and srv_count is medium diff_srv_rate is medium and dst_host_srv_count is medium then packet is uncertainty
- IF src_bytes is high and dst_bytes is high and srv_count is high diff_srv_rate is high and dst_host_srv_count is high then packet is abnormal.

Likewise. all the possible rules are generated by the Intuitionistic fuzzy System Classifier, the main issue in rule generation is its volume. All the rules generated by the IFS classifier has to be validated and irrelevant and redundant rules must be eliminated to improvise the accuracy of the IDS. Hence, in this work to perform

rule pruning process butterfly optimization algorithm is adopted to search the most fittest rules. Its detailed working process is explained the next section.

Butterfly Optimization Algorithm (BOA)

In this study, an evolutionary approach based on the Butterfly Optimization Algorithm (BOA) [18] is used to improve and enhance the productivity of the traditional Intuitionistic Fuzzy Inference Model. The BOA concept is used to prune the rules generated by IFS. Unlike traditional IFS, which develops the rule base only by experts, the butterflies in this study assess the best fittest rules based on their food source finding behaviour, and only those fittest rules are employed for intrusion detection. BOA is based on the metaheuristic sensory aspect of the butterfly's foraging feeding behaviour like sight, hear, touch, smell and taste which are the major factors involved in detecting food sources. Its sensing behaviour assists in the placement of eggs in a safe location away from predators. Even if it is from a great distance, the sense of smell is highly crucial in assisting the butterfly in determining its nectar (meal). It has sensory receptors all throughout its body, including palps, antennae, legs, and so forth. Butterflies' receptors are called chemoreceptors, and they are nerve cells. They can detect differences in smells and intensities because it has a more exact sense of place of the aroma source. It functions as a search agent in BOA, generating fragrances of appropriate intensity that are linked to its fitness.

In this research work, the BOA is integrated to search the best fittest rules which influence the accuracy of intrusion detection by well balancing both global and local searching strategy. During global searching, the butterfly takes a step toward the fittest butterfly (or) solution s^* is mathematically formulated as

$$y_i^{t+1} = y_i^t + (rd^2 + s^* - y_i^t) * fn_i$$

Where y_i^t refers to solution vector y_i for the i^{th} butterfly at time period t, s* indicates the present best solution (rule) determined among all the solutions (set of rules) in current iteration. Fragrance of i^{th} butterfly is denoted by fn_i and rd is the random number lies between [0.1]. The local search of butterfly is signified as

$$y_i^{t+1} = y_i^t + \left(rd^2 + y_j^t - y_k^t\right) * fn_i$$

Where y_j^t and y_k^t are the solution space of j^{th} and k^{th} butterflies. If both belongs to same swarm then rd is a random number lies between [0.1] and exhibits local random walk.

Algorithm: Rule Pruning usign Butterfly Optimization Algorithm

Input : Set of Intuitionistic Fuzzy Rules Z

Set $of(Z), Z = (z_1, z_2, \dots, z_d), d = no\ of\ directions$
Produce initial population of butterflies Z_i=(i=1, 2, …, n)
Intensity of Stimulation SI_i at Z_i is computed by $fn(z_i)$
Determine probability to switch sp, modality sensor m, exponent power e
while termination criteria are not met
{

for each *bf* in population
Estimate fragrance for *bf* using equation below
fn = m* SI^e
discover best butterfly bf
for every *bf* in population
{
Generate random value rd from [0,1]
if rd<*ps* then
move to the best fittest butterfly/solution as computed below

$$y_i^{t+1} = y_i^t + (rd^2 + s^* - y_i^t) * fn_i$$

else
move butterflies randomly using equation

$$y_i^{t+1} = y_i^t + \left(rd^2 + y_j^t - y_k^t\right) * fn_i$$

end if }

Update power exponent value e

End while
Output: Best fittest rule selection for IDS

Experimental Results

In this section the discussion about the experimental results conducted on the proposed model IFS-BOA for intrusion detection in network is explained in detail. The dataset used for IDS is collected from KDD Cup 99 Dataset [6] which is managed by MIT Lincoln labs. In this work 60,000 records of IDs dataset which comprised of four different attacks. The IFS-BOA algorithm is simulated using Matlab code for detecting attacking instances. The performance of the proposed IFS-BOA is assessed with the existing classification models namely Rule Induction (RI), Support Vector Machine (SVM) and Fuzzy Inference System (FIS).

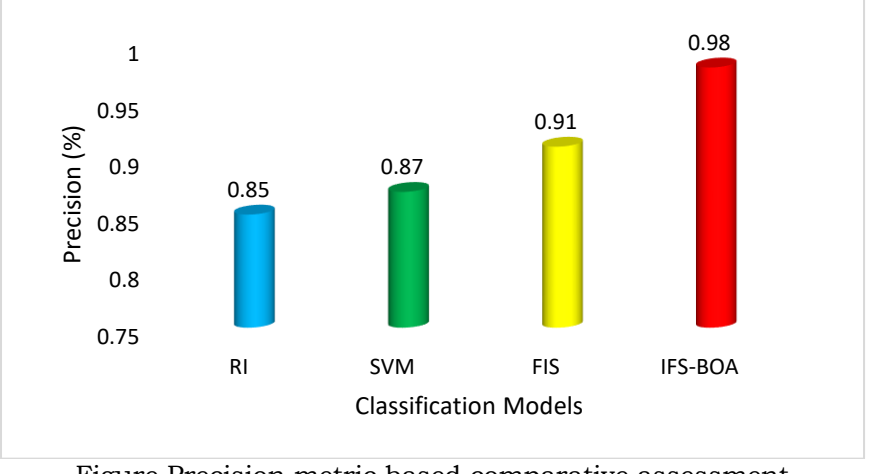

Figure Precision metric based comparative assessment

Figure 6 depicts the performance of proposed intuitionistic fuzzy inference system enriched with butterfly optimization algorithm in intrusion detection produced better precision value compared to Rule Induction (RI), Support Vector Machine (SVM) and Fuzzy Inference System (FIS). The IFS-BOA involves in pruning of irrelevant rules generated by the Intuitionistic Fuzzy Inference system by inheriting the knowledge of butterfly food searching strategy the least fit rules are eliminated for better classification results

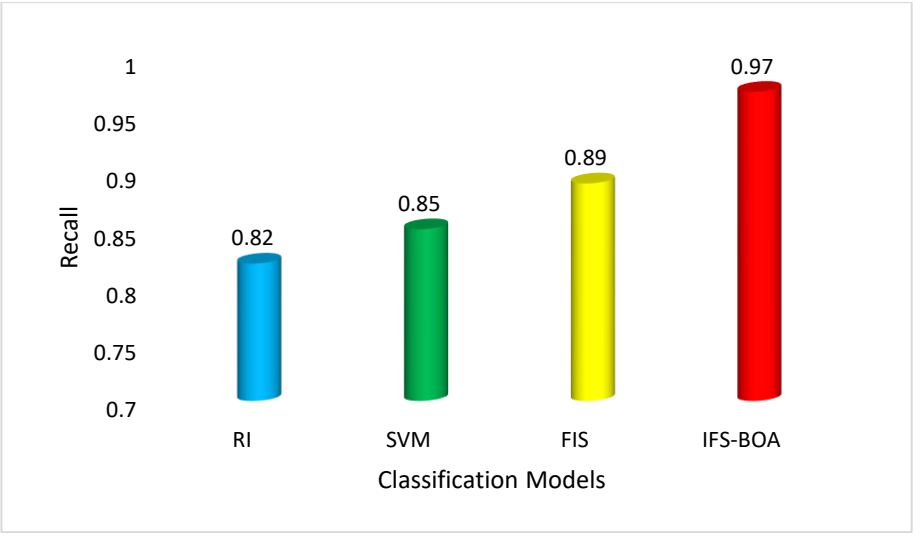

Figure illustrates the performance of four different classification models based on the recall rate obtained by each of them against intrusion detection. The intuitionistic fuzzy inference system defines each data packets characteristics in terms of membership degree, non-membership and hesitancy degree. The irrelevant and redundant rules are identified by the butterfly optimization algorithm, which evaluates the fitness value of each rule. The highest fitness value gaining rules are retained. Thus the proposed IFS-BOA produced highest recall rate compared to other classification models FIS, SVM and RI in normal and abnormal packets detection.

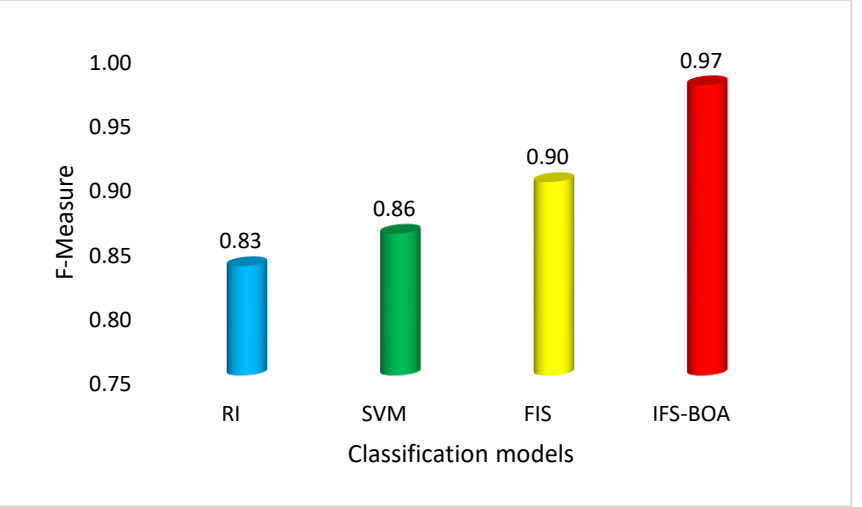

Figure explores the performance based on the F-Measure value generated by four different classification models involved in IDS. The inconsistency in representing unknown traffic patterns is defined precisely by introducing degree of hesitancy in intuitionistic fuzzy improves IDS process more prominently. While using fuzzy inference system it defines only based on degree of membership. The SVM and RI is capable of handling small set of data, when volume of data increases they results in overfitting. Thus, IFS-BOA produced better F-measure rate compared to other state of arts classification algorithms.

Conclusion

In this proposed work the major issue of inconsistency in rule-based classifier is overcome by applying the rule pruning process which controls the volume of rules by eliminating redundant and irrelevant rules generated by Intuitionistic fuzzy System. The inconsistency and unceratainty in detecting new unknwon pattern of abnormal packets is effectively determined by the IFS model. The rules generated by IFS is effectively pruned by adopting butterfly optimization algorithm which evaluates the fitness value of each rule in terms of rule coverage and rule strength. The best fittest rules are maintained and utilized for intrusion detection. The local and global search strategy of butterfly algorithm well balances the optimized searching of optimized rule sets by alleviating irrelevant and redundant rules. The simulation result exposes that the proposed IFS-BOA prevails accuracy in detection of intrusion compared to the existing classification models.

References

1. Li J, Qu Y, Chao F, Shum HP, Ho ES, Yang L. Machine learning algorithms for network intrusion detection. AI in Cybersecurity. New York, NY: Springer; 2019: 151- 179
2. Khraisat A, Gondal I, Vamplew P, Kamruzzaman J. Survey of intrusion detection systems: techniques, datasets and challenges. Cybersecurity. 2019; 2(1): 20.

3. Z. Malek and B. Trivedi, “The Rule Based Intrusion Detection Model For User Behavior”, International Journal of Advanced Research in Computer Science and Software Engineering, vol. 5, no. 12, (2015).
4. Ö. Cepheli, S. Büyükçorak and G. K. Kurt, “Hybrid Intrusion Detection System for DDoS Attacks”, Journal of Electrical and Computer Engineering, vol. 2016, (2016).
A. Aburomman, M. B. I. Reaz, A survey of intrusion detection systems based on ensemble and hybrid classifiers,” Computer Security, vol. 65, pp. 135–152, Mar. 2017.
5. E. Kabir, J. Hu, H. Wang, and G. Zhuo, “A novel statistical technique for intrusion detection systems,” Future Generation Computer Systems, vol. 79, pp. 303–318, 2018.
6. Nannan Lu, Yanjing Sun, Hui Liu, Song Li, "Intrusion Detection System Based on Evolving Rules for Wireless Sensor Networks", Journal of Sensors, vol. 2018, Article ID 5948146, 8 pages, 2018
7. Mehr, S.Y.; Ramamurthy, B. An SVM Based DDoS Attack Detection Method for Ryu SDN Controller. In Proceedings of the 15th International Conference on Emerging Networking Experiments and Technologies, Orlando, FL, USA, 9–12 December 2019, pp. 72–73.
8. S. Elhag, A. Fernández, A. Altalhi, S. Alshomrani, F. Herrera, A multi-objective evolutionary fuzzy system to obtain a broad and accurate set of solutions in intrusion detection systems, Soft Computing, vol. 23, no. 4, pp. 1321–1336, Feb. 2019
9. R. A. Bridges, T. R. Glass-Vanderlan, M. D. Iannacone, M. S. Vincent, and Q. Chen, “A survey of intrusion detection systems leveraging host data,” ACM Comput. Surv., vol. 52, no. 6, pp. 1–35, Jan. 2020
10. Alzahrani, A.O, Alenazi, M.J.F. Designing a Network Intrusion Detection System Based on Machine Learning for Software Defined Networks. Future Internet 2021, 13, 111.
11. V. Varanasi and S. Razia, "Network Intrusion Detection using Machine Learning, Deep Learning - A Review," *2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2022, pp. 1618-1624,
12. S. Latif, Z. Zou, Z. Idrees, and J. Ahmad, “A novel attack detection scheme for the industrial internet of things using a lightweight random neural network,” IEEE Access, vol. 8, pp. 89 337–89 350, 2020
13. M. A. Khan, S. Latif, A. A. Shah, M. U. Rehman, W. Boulila, M. Driss, and J. Ahmad, “Voting classifier-based intrusion detection for IoT networks,” in 2021 2nd International Conference of Advance Computing and Informatics (ICACIN). Springer, 2021 http://kdd.ics.uci.edu/databases/kddcup99/
14. K. Atanassov, P. Vassilev, R. Tcvetkov, "Intuitionistic fuzzy sets measures and integrals" in Sofia:” Prof M. Drinov, AcademicPublishing House, 2013.
A. Hernandez-Aguila, M. Garcia-Valdez and O. Castillo, "A proposal for an intuitionistic fuzzy inference system," 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2016, pp. 1294-1300, doi: 10.1109/FUZZ-IEEE.2016.7737838.
15. Xingyue Cui , Zhe Chen , Fuliang Yin, Differential Evolution and Local Search based Monarch Butterfly Optimization Algorithm with Applications, International Journal of Computational Intelligence Systems, Vo l. 12 (2018) 149-163.Gede Budasi, I. & Wayan Suryasa, I. (2021). The cultural view of

North Bali community towards Ngidih marriage reflected from its lexicons. *Journal of Language and Linguistic Studies*, *17*(3), 1484–1497

16. Kustina, K.T., Dewi, G.A.A.O., Prena, G.D., Suryasa, W. (2019). Branchless banking, third-party funds, and profitability evidence reference to banking sector in indonesia. *Journal of Advanced Research in Dynamical and Control Systems, 11*(2), 290-299.
17. Nyandra, M., Kartiko, B.H., Susanto, P.C., Supriyati, A., Suryasa, W. (2018). Education and training improve quality of life and decrease depression score in elderly population. *Eurasian Journal of Analytical Chemistry*, *13*(2), 371-377.