SC-CVAR: Intrusion Detection Using Feature Selection and Machine Learning Techniques on UNSW-NB15 Dataset

J. Vimal Rosy[1*] and Dr. S. Britto Ramesh Kumar[2]
Research Scholar[1], Assistant Professor[2]
St.Joseph's College(Autonomous), Affiliated to Bharathidasan University, Trichy, India

Summary
This research study provides an effective mechanism in detecting and classifying the attacks taken from UNSW-NB15 dataset such as backdoor, Exploits, Shellcode, analysis, fuzzers, generic, normal, reconnaissance, DoS and Worms attacks. Specifically, to enhance the accuracy, the feature selection process is performed using SC-Sine Cosine algorithm, selected only the significant features. Finally, the classification of the intrusion is performed using the Novel CVAR- k fold Cross validated Artificial neural network weighted Random Forest classification. In the prediction phase the type of attacks are revealed. Finally, the proposed SC-CVAR model evaluated in terms of different performance metrics and compared with various existing models to prove its efficiency. The research outcome revealed that this research is highly effective in detecting and classifying the attacks in greater accuracy

Keywords: *Intrusion detection, UNSW-NB15 dataset, Random Forest classifier, Sine Cosine algorithm***.**

1. Introduction

Global Internet Statistics Report notified that the growth of internet in recent days have reached 4.66 billion active users and thus higher than 2 quintillion bytes of data have generated daily. It shows that the rate of data access has progressed from different sources are extremely faster and the development of methodologies and hacking tools also growing very fast. Hence there is a requirement for data security and pr.privacy for protecting the data from different intrusion or malicious attacks. Because of the greater volume and greater data speed, the conventional intrusion detection system couldn't detect the attacks or intrusion in efficient and faster manner. However certain computational approaches are difficult in nature to handle such kind of data and it requires advanced intelligent approaches and powerful technologies. In detecting the attacks, Intrusion detection system- IDS plays a significant role. IDS system will monitor the traffic in network in the intention of identifying threats, attacks or a suspicious activity. When such kind of activity is identified, it may issue an alert to the corresponding admin. With the purpose of handling the intrusion effectively various machine learning techniques can be employed. To handle and classify the intrusion or attacks in an efficient manner, different machine learning algorithms can be utilized[1],[3].

For higher than two decades, IDS system is generally used to improvising the security in network and information system, it has been termed as important tool[4]. In accordance with smart IoT devices protection, IDS is utilized and it has been addressed various attacks while evaluating and monitoring the suspicious traffic in networking environment[5]. Because of the specific protocol stacks, standards and architectural constraints, the traditional IDS method execution on IoT is termed as a trial. It is due to the fact that all attack types are not able to protect and thus new methods are essential which includes physical hardware application utilizing network probe which transmission the secure data to remote server and performing the malicious detection. However it needs in-depth resources[6]. The IDS plays important role in resisting hacker intrusion and developing effective IDS which is termed as major challenge. For detecting the suspicious attacks machine learning techniques can be used. For the actual detection process, machine learning algorithms are trained and applied on undetected input. In a network, there exists various classification algorithms like machine learning used for detecting the attacks. Feature reduction algorithms can be used to improvise the detection time and classifiers performance[7].

In 1980, first intrusion detection has been established and after than several mature IDS products have raised. But, several IDSs are still suffering from greater false alarm rate which generating many alerts for lower non-threatening situations, which in case raises the security analysts since serious malicious attacks can be ignored in some cases. Therefore, several researchers focusing on developing IDSs with minimized false alarm rates and greater detection rates. Another issue with the existing IDSs is they are not having the ability in detecting unknown attacks. Since the network environments rapidly changes, the new attacks and its associated variants emerges frequently. The unknown attacks is detected by enhanced IDS, which is essential. Machine learning methods are considered by the researchers in constructing IDSs. Machine learning is categorised from Artificial

Manuscript received April 5, 2022
Manuscript revised April 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.4.81

intelligence approach which can identified beneficial information from the huge datasets. When the satisfactory training data is obtained, then IDSs based on machine learning can attained reasonable detection levels and thus it attains adequate generalizability in detecting the unknown attacks and its variants. Moreover, IDS based on machine learning is not fully depends on technical knowledge and thus it is easy to construct and design[8].

Certain comparative studies have been performed but exhaustive research is not yet performed. Hence this research focus on to develop an IDS for network with better feature selection and classification approaches in studying different and effective feature selection approaches. This current study emphasizes on feature selection process combined with classification approaches to build improvised IDS accuracy. There are certain challenges in establishing the improvised IDS. Due to the architectural restrictions, protocol stacks and standards the conventional IDS system execution is considered as trial. While identifying intrusion, resource and computational constraints also another issue. All types of attacks detection is essential and it will leads to intrusion detection accuracy rate increase. To analyse all sort of performance metrics better feature dataset must be selected.

To addressing the above complexities like focusing on all types of attacks and enhancing the accuracy of detecting intrusion rates, novel SC-CVAR algorithm is proposed. Thus, the major contribution of the stud involves,

- For fast and accurate detection of all attack types, the established IDS system should be effective. To attain this purpose the cross validated Artificial Neural Networks with Random Forest Classifier namely CVAR used for classification and to increase accuracy of classification rate efficient feature selection method used namely Sine Cosine algorithm. It further returns best solution among global optimum.
- To analyse the classification performances, the UNSW-BoT Dataset is utilized. The performance has been related with other existing models to measure the proposed IDS system effectiveness in detecting malicious attacks.

The rest of the paper is organized as, section II discussed about the related works of IDS and the various machine learning methods implementation to addressing the different attacks. Further the proposed SC-CVAR model is elaborated briefly in section III. Moreover, in section IV the results and discussion is exhibited and illustrated. Finally in section V, the paper is concluded with future works

2. Related Works

For detecting malicious behaviour in data transmission network, two machine learning approaches like feed forward neural network and Boosted Decision tree have been established. Satisfactory performance and sensitivity values are obtained. The obtained values are analysed and related with existing studies and proved its effectiveness[9]. For assuring standard security issues addressing like authenticity, trust and privacy effective deep learning based model developed in[10]. Further this[11] model used the hybrid data optimization technique comprised with feature selection and data sampling. Outliers have been eliminated by isolation forest- iForest and the sampling ratio is optimized by genetic algorithm-GA, optimal training dataset obtained by random forest classifier performing on UNSW-NB15 dataset. Rare anomaly behaviours have been detected from this model. For data optimization more time cost is required and online procedure support considered as limitations. This model can be applied to other anomaly detection area like fraud detection. Searching approaches are optimized since it takes more time for classifier training.

For all machine learning process, only individual approach is used for pre-processing the data. the random forest classifier takes more time even though it gives better performance[12]. Another ensemble learning based support vector machine, auto-encoder and random forest have been used on UNSW-NB15 and NSL-KDD dataset and network log in campus comprised with 300M daily records. The results of this model have been related with other conventional studies shows that the ensemble learning model restrict the false negative and false positive predictions[13]. IDS-CNN based on Convolutional Neural Network exhibited different open-source tools like tensor flow, traffic analysis and packet capture interface. The tensor flow is the machine learning interface focusing on neural network training and testing, intrusion response and pre-processing. Precision results shows better values compared with existing approaches[14].

Better detection rate obtained for IDS based CNN which were related with other IDS classifier system performing on KDD cup dataset. False alarm rate should be minimized[15]. Alarm filtering has concentrated to increase the accuracy of detection rate. Higher detection rate of intrusion attained using hybrid ant colony optimization with respect to unsupervised clustering resulted in false alarm rate minimization also. K-means clustering has been used for convergence of ant colony optimization algorithm[16]. For establishing RNN-IDS refer as Recurrent Neural Networks related with other machine learning models which were analysed and performance evaluated on NSL-KDD dataset. Accuracy in improvised but the time taken for training shows more and it needs to be considered[17]. By using the support vector machine and extreme learning in multilevel based hybrid IDS, the unknown attacks detection rates have been enhanced. Using the modified K-means algorithm the IDS performance and overall training time reduced shows greater accuracy[18].

Several optimization approaches have been imposed to maximize the rate of accuracy which were analysed on NSL-KDD dataset. The anomalies in network traffic has been allowed by intruders network protocol, which has further detected and evaluated based on NSL-KDD dataset[19]. In a higher dimensional network traffic, the dimensionality reduction plays significant ole since anomalies detection resulted in time consumption in higher dimensional network traffic. Bayesian network used for classification and significant features selected using firefly algorithm on KDD CUP 99 dataset. Accuracy also enhanced[20]. Moreover, in improvising the accuracy of IDS which is based on CNN and random forest classifier which extract the features based on raw network packet. Better performance resulted in case of random forest on structured data whereas in case of unstructured data, CNN has handled it. Certain attacks have been identified and however greater computational complexity has been resulted with greater length[21].

In some cases, intrusion detection is challenging since the resource constraints and computational complexity exhibited. light weight access control protocols established by K-NN and SVM to save computational resources and system lifetime in IoT devices[22]. The accuracy rate is not satisfactory in certain cases using the data mining approaches and thus new IDS based linear correlation co-efficient for feature selection with conditional random field and CNN for classification employed. Greater accuracy has attained. More efficiency has been yielded based on optimized standard approaches[23]. Using conventional IDS based advanced attacks detection shows lesser efficiency. Deep learning models can also use to address this issue and however it highly concentrated on Denial of Service- DoS attacks. For evaluation, KDD cup-99 dataset has used contains probing attacks, DoS attacks, user to root U2R attacks, remote to local attacks. Compared with RNN, CNN shows better efficiency in intrusion detection. Multi-class classification has not focused yet[24]. Different machine learning and deep learning approaches have been considered for DoS attack detection which resulted in better accuracy[25].

Different IoT malware detection method considered to perform rapidly and precisely in IoT field is another challenge. Different behaviours have been recognised using the malware detection dynamic evaluation based neural network, which were extracted in related with system call, memory and virtual process system. Further it can be transformed into malware images and by analysing these behaviour images the damages have been minimized[26]. U2R, probing attacks and R2L attacks have been focused in automated IDS system and computational overhead has resulted in detecting other attacks from NSL-KDD datasets. However stability and potentiality also recorded[27]. Extreme learning ELM based neural network has executed for handling the dimensionality reduction. Both IDS execution and detection efficiency enhanced on NSL-KDD dataset. Better performance has been observed[28].

3. Methodology

In this section, the newly proposed IDS model namely SC-CVAR has been elaborated and the overall flow is shown in figure-1. Initially the UNSW-NB15 dataset is loaded and the pre-processing is performed. Further to enhance the accuracy, the feature selection process is performed using SC-Sine Cosine algorithm, selected only the significant features. Finally, the classification of the anomalies is performed using the Novel CVAR- k fold Cross validated Artificial neural network weighted Random Forest classification. In the prediction phase the type of attacks are revealed. Finally, the proposed SC-CVAR model evaluated in terms of different performance metrics and compared with various existing models to prove its efficiency.

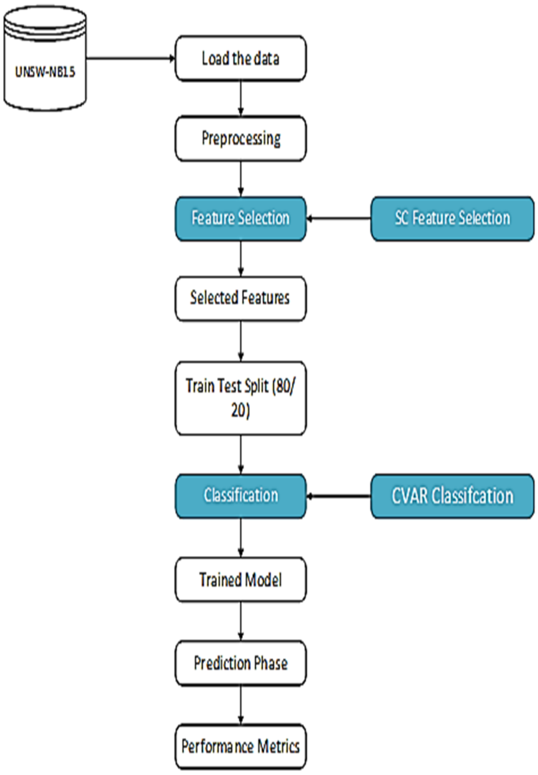

Figure-1 Proposed SC-CVAR Model flow

3.1 Feature Selection using Sine Cosine algorithm

For feature selection the proposed study utilized Sine Cosine algorithm and the flow is shown in figure2. In this procedure, the location has been initialized, objective function used to evaluate the search agents, best solution replaced and location is updated. The parameters like r1, r2, r3 and r4 are updated. Further the search agent's position is also updated. The best solution can be return as global optimum. In this way the significant feature has selected.

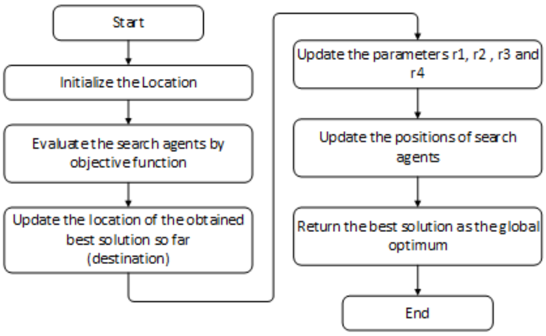
Start
Initialize the Location
Evaluate the search agents by objective function
Update the location of the obtained best solution so far (destination)
Update the parameters r1, r2 , r3 and r4
Update the positions of search agents
Return the best solution as the global optimum
End

Figure-2 Sine Cosine Algorithm for feature Selection

For solving real world issues, meta-heuristic algorithm shows greater performance. Binary search problems are exhibited from feature selection problems. The sine cosine algorithm starts with random positions in which the search agent $g_{cd} = 5$. The SC algorithm mathematical definition is,

$$g_{cd}^{v+1}\begin{cases} g_{cd}^{v} + r_1 * \sin(r_2) * |r_3(Ps_d^v) - g_{cd}^{v}| \text{if } K_1 < 0.5 \\ g_{cd}^{v} + r_1 * \cos(r_2) * |r_3(Ps_d^v) - g_{cd}^{v}| \text{if } K_1 \geq 0.5 \end{cases} \quad (1)$$

$$r_1 = a - v\frac{a}{Vmax} \quad (2)$$

If the classification performance increased over testing dataset while validating and attaining the minimal number of feature selection the sine cosine algorithm's fitness function also increased.

$$h\theta = \omega * E + (1 - \omega)\frac{\sum_c \theta c}{n} \quad (3)$$

From the Eq. (3), the fitness function $h\theta$ provides θ vector, size as n with 0/1 elements showing the selected or unselected features. The error rate of classifier is E , constant value is ω for handling the performance of classification accuracy to number of selected features. In the selected dataset, variables used are similar features which are restricted in [0,1] range, in which 1 refer as corresponding features selected. The variable is threshold in individual fitness measurements which decides the appropriate features to be evaluated shown in Eq. (4),

$$h_{cd} = 1 \text{ if } P_{cd} > 0.5, \text{otherwise } 0 \quad (4)$$

From above Eq. (4), h_{cd} is the dimension value for c search agent at d dimension. Every search agent position updated at certain dimensions, the restricted constraints [0, 1] violated by the updated value. For assuring the variable limit, standard truncation rule is followed. In one dimension, ever candidate feature is represented. The 6vectors mapping are [0,1] with respect to 0.5 threshold value shows lower bound as 0 and upper bound as 1.

For evaluating the feature subset in SC algorithm search space, the fitness function is employed based on CVAR classifier. The proposed fitness function is measured and shown in Eq. (3). For providing new set of individuals mutation operator can be used with different features which are not in ancestor. Also, it can be applied for binary, real or integer representation and comprised with several types. Mutations has been established by selecting one or higher bits randomly and flipping the value based on specific probability. In SC algorithm, mutation is performed as internal function which further establishes new solution and enhances the exploration ability after crossover operator employed. The mutation operator is expressed as,

$$P_c^{v+1} = \text{Mutation}(P_c^v) \quad (5)$$

SC Algorithm:

```
INPUT: Search agent number, dataset, dimension,
        max_iter, upper bound, lower bound
                max_iter, upper bound, lower bound
OUTPUT: Vector with 20 best solutions
 Initialize the SCA population yc, $(c = 1,2 \cdots n)$
FS = the best search agent by equation(5)
u = 1
While $(u < max\_iter + 1)$
For c = 1: N
Calculate $r_1, r_2, r_3$ and $r_4$
Identify best solution Ps
If $(K_1 < 0.5)$
Update the position of the SCA by sin equation(1);
Else
Update the position of the SCA by cos equation(2);
end if
        Improvement the current position by proposed
CVAR algorithm
End for
Amend the SCA Agent based on the upper
and lower bound of variables;
FS = the best search agent by equation (5);
End for
u = u + 1;
End while
End for
```

3.2 Classification of identified intrusion using CVAR algorithm

Before establishing the classifier model, the quality should be assured and the candidate issue compatibility is essential. Data standardization is considered as next step. In different formats, the standardization can present in which the Artificial Intelligence approaches are assume the data insights. For feature column measurement, the standardized value is,

$$z_{cd} = \frac{g_{cd}-\bar{x}}{s_c} \qquad (6)$$

In the following step, the significant feature set is deduced using random forest classifier used as wrangler model. Chi-square feature score measurement has been performed by using,

$$g_c^2 = \sum \frac{(o_c-E_c)^2}{E_c} \qquad (7)$$

The activation node function is the same as set of node outputs, utilized as next node input in ANN network. Based on desired range of 0 and 1 the output units have been generated. The ReLU activation function is expressed in,

$$h(g) = g^+ = \max(0,g) \qquad (8) \qquad \backslash$$

$$h_{a-b} = \varphi(Z_{a-b}) \qquad (9)$$

$$Z_{a-b} = P_{ua-b}^{v}H = P_{un1}H_1 + P_{un2}H_2 + P_{U0} \qquad (10)$$

$$H \Longrightarrow \varphi(\mathrm{H}) \Longrightarrow \mathrm{P}^{\mathrm{v}}H \qquad (11)$$

$$\hat{Y} = \sigma(\varphi) \qquad (12)$$

$$\hat{Y} = \frac{e^{\varphi_k(H)}}{\sum_{k=1}^{k} e^{\varphi_k(H)}} \qquad (13)$$

The hidden layer output function is calculated by using Eq. (9), in which contribution of hidden layer is explained in Eq.(10). Using Eq. (11) to (13), feature vector of input transformation to output classes is performed. Hidden layer activation function is φ and sigmoid activation function is σ. This ANN model is combined with k fold cross validation and random forest classifier. This proposed CVAR model is removes the hyper parameter tuning and thus appropriate for huge data volume and reduced memory requirements.

Algorithm: K- fold Cross validation

```
1. Input Selected Feature
      a) The dataset was selected randomly
      b) The dataset was split into K groups
      c) For each K groups:
           i.   The groups are input mode,
                Hidden mode, Output mode.
           ii.  The one group is selected as a test
                set.
           iii. The random forest method was
                fitted in the group.
           iv.  Creating a training set.
      d) Random forest was evaluated on the test
         set.
      e) The evaluation weight was retained.
```

The cross-validation method has utilized to evaluate the skill of random forest classifier. Resampling process is considered for evaluating the model skill. K parameter shows the no. of groups based on splitting of dataset. Furthermore, for accurate classification of intrusion the CVAR algorithm is utilized and this algorithm is more robust. The following algorithm is the CVAR pseudo code. The random forest classifier comprised with several decision trees and it create a decision tree based on data sample which predict everyone and best solution is selected finally through voting. Compared with decision tree it performed better

4. Results and Discussions

The proposed SC-CVAR method used to detect the intrusion is performing on UNSW-NB15 dataset. The dataset is obtained from https://research.unsw.edu.au/projects/unsw-nb15-dataset. The UNSW-NB15 dataset raw network packets have been generated by UNSW Canberra cyber range lab for synthetic contemporary behaviour attack and generating activities of real modern normal.

4.1 Performance analysis

The outcomes attained for K-Fold cross validation of the proposed system are shown in table-1. Different accuracy rate has been attained for various kinds of cross validation. Accordingly, 1-fold cross validation showed 68.11%, 2-fold cross validation showed 99.84%, 3-fold cross validation showed 99.85% etc. However, the 5-fold cross validation showed maximum accuracy rate of 99.87%.

Table 1*: K-Fold cross validation*

Method	***Accuracy***
1-fold cross validation	68.11
2-fold cross validation	99.84
3-fold cross validation	99.85
4-fold cross validation	99.86
5-fold cross validation	99.87

In addition, the overall count of columns and the selected columns of the introduced system are shown in table-2. From the outcomes, it is found that nearly 18 columns have been selected from 40 columns. The 18 selected columns have found to be relevant.

Table 2*: Columns selected from the dataset*

Total number of columns	***40***
Number of Selected Columns	2,5,6,8,9,10,12,16,18,20, 24,29,31,33,34,36,40

Moreover, the confusion matrix is generally utilised to assess the classifier model's performance. Obtained results are shown in figure3. From the results, it is found that, 11098 have been correctly identified as attacks, while, 23857 have been correctly detected to be normal. On contrary, 91 attacks have been misinterpreted as 91, whereas, 23 normal have been misclassified as attacks. Though the correct classification rate is maximum in comparison to the misclassified rate, the proposed classifier is found to be effective.

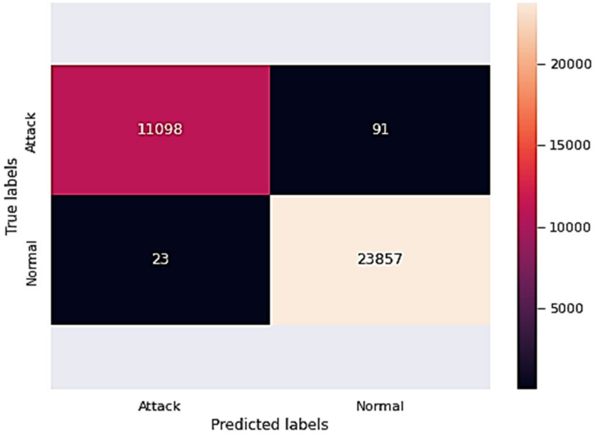

Figure-3 Confusion Matrix

4.2 Comparative Analysis

Performance of the proposed system is comparatively analysed with respect to precision, recall, F-measure, FPR (False Positive Rate) and accuracy. Obtained results are presented in table-3. Abnormal, Normal and Weighted average are the existing methods considered for analysis. The graphical results are also presented in figure-4 and figure-5.

Table 1: Comparative analysis in terms of performance metrics[29]

Method	***Precision***	Recall	F-measure	FPR	Accuracy
Abnormal	0.9482	0.9647	0.9564	0.0646	0.9515
Normal	0.9558	0.9354	0.9455	0.0353	0.9515
Weighted average	0.9516	0.9515	0.9515	0.0540	0.9515
SC-CVAR	0.9915	0.9915	0.9915	0.0256	0.9987

From the analysis, it is revealed that normal method showed maximum precision rate of 0.9558%, however, the proposed SC-CVAR (k-fold Cross Validated Artificial Neural Network Weighted Random Forest Classification) expose maximum precision than normal methodology with 0.9915% as precision rate. Similarly, abnormal technique showed maximum recall rate with 0.9647%. But, the introduced system is higher than the traditional methods with 0.915%. Likewise, the F—measure and accuracy rate of the considered methods varied in each cases. However, the proposed work exposed better outcomes than conventional techniques. On contrary, minimum FPR rate confirms a system to be effective. Accordingly, proposed work showed less FPR than existing methodologies as shown in figure4.

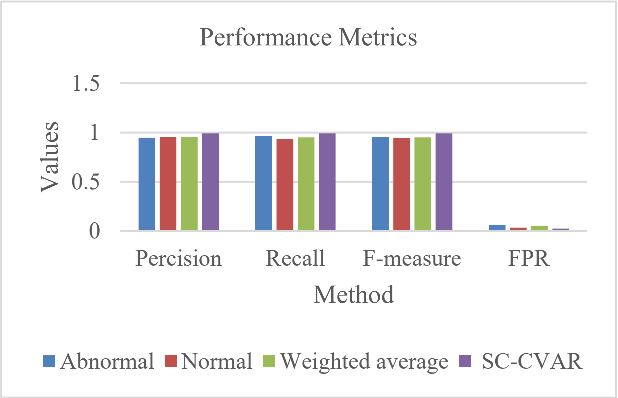

Figure-4 Comparative analysis in terms of performance metrics[29]

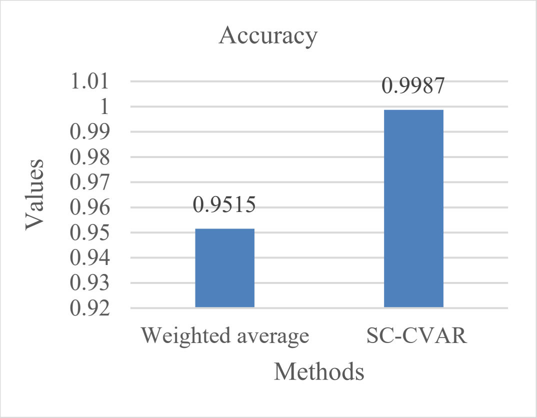

Figure-5 Comparative analysis in terms of Accuracy metrics[29]

In addition, analysis has been undertaken by considering various algorithms like LOGNN (Logarithmic Neural Network), RF (Random Forest), KNN (K Nearest Neighbour), CNN (Convolutional Neural Network), RNN (Recurrent Neural Network), DNN (Deep Neural Network), LSTM (Long Short-Term Memory), SVM_rbf, ConvLSTM_SAE_NN, SDAE_ELM1, SDAE_ELM2, SDAE_ELM3, stacked_CNN_LSTM_SAE_NN and DBN_EGWO_KELM. Obtained results are shown in table-4. Accuracy rate of traditional stacked_CNN_LSTM_SAE_NN and ConvLSTM_SAE_NN showed 0.954% which has been better than other existing works. But proposed work showed high accuracy than the considered methods showing its efficiency. Similarly, the precision, recall and F1-score of introduced system exposed better results in comparison to conventional researches. The attained outcomes are presented in figure5.

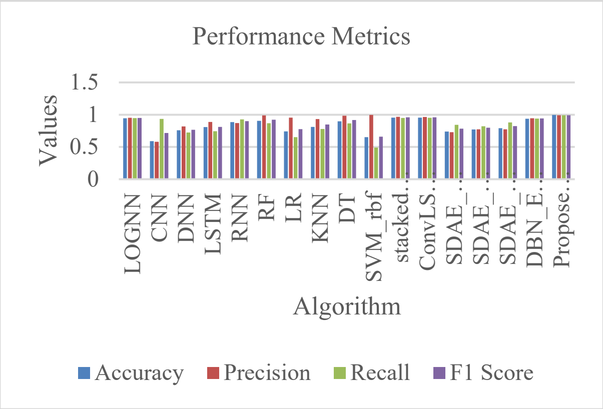

Figure-6 . Comparative analysis with respect to performance metrics [30]

Additionally, performance of proposed methodology has been comparatively analysed with conventional research[31]. Accuracy, F1-score, recall, time and precision has been considered for analysis. Though the existing techniques showed better outcomes, proposed system showed effective performance with respect to all the metrics considered. Moreover, execution time of the introduced work has also been found to low that explored its efficacy than traditional methods. The graphical representation is shown in figure-7 and figure-8.

Table 4*: Analysis with respect to performance metrics[30]*

Algorithm	***Accuracy***	Precision	Recall	F1-Score
LOGNN	0.944	0.951	0.947	0.949
CNN	0.590	0.579	0.936	0.716
DNN	0.758	0.817	0.723	0.767
LSTM	0.807	0.887	0.744	0.810
RNN	0.884	0.870	0.928	0.898
RF	0.903	0.988	0.867	0.924
LR	0.743	0.955	0.653	0.775
KNN	0.810	0.932	0.778	0.848
DT	0.897	0.982	0.864	0.919
SVM_rbf	0.653	0.998	0.492	0.659
stacked_CNN_LSTM_SAE_NN	0.954	0.967	0.950	0.958
ConvLSTM_SAE_NN	0.954	0.964	0.952	0.958
SDAE_ELM1	0.740	0.728	0.843	0.782
SDAE_ELM2	0.770	0.774	0.821	0.797
SDAE_ELM3	0.791	0.773	0.879	0.823
DBN_EGWO_KELM	0.937	0.945	0.940	0.942
Proposed SC-CVAR	0.9987	0.9915	0.9915	0.9915

Table-5. Analysis with respect to various performance metrics [31]

Model	***Metrics***				
	Accuracy	Precision	Recall	F1 Score	T+ime(s)
MLP-1	98.98	93.365	98.46	95.85	430.95
MLP-2	98.99	93.61	98.38	95.94	595.03
MLP-3	98.96	93.20	98.55	95.80	606.21
MLP-4	98.91	92.66	98.67	95.57	617.15
CNN-1	99.10	95.50	97.74	96.38	423.58
CNN-2	99.10	94.71	98.08	96.38	892.19
RNN-1	98.91	93.62	97.68	95.61	128.94

RNN-2	98.80	92.20	98.29	95.15	273.47
RNN-LSTM-1	98.94	93.95	97.57	95.73	301.56
RNN-LSTM-2	98.88	92.63	98.11	95.29	388.29
RNN-GRU-1	98.88	93.21	97.87	95.48	317.04
RNN-GRU-2	98.81	92.10	98.38	95.14	412.73
CNN+LSTM-1	99.13	95.93	97.127	96.53	1355.8
CNN+LSTM-2	99.12	95.97	97.08	96.52	1334.1
Proposed SC-CVAR	**99.87**	**99.15**	**99.15**	**99.15**	**126.30**

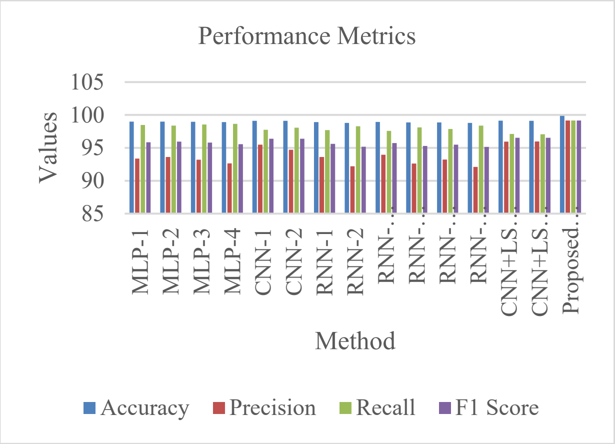

Figure-7 Comparative analysis with respect to performance metrics [31]

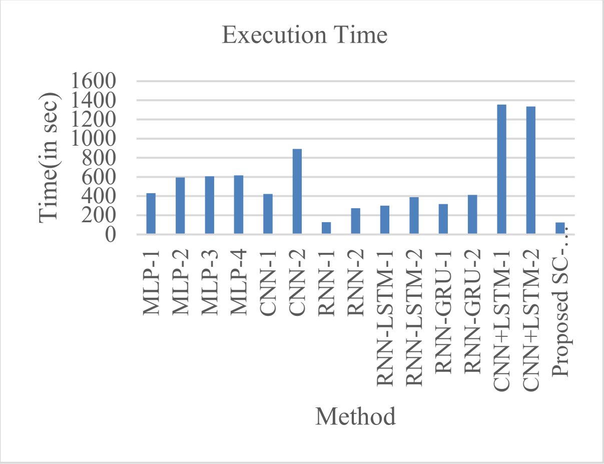

Figure-8 Comparative analysis with respect to execution time [31]

The analytical outcomes revealed that the proposed system is effective and efficient than conventional methods with respect to the considered metrics. The misinterpretation rate and execution time of the introduced work is minimum that exhibited its efficacy. These results make is suitable for intrusion detection.

Conclusion

Therefore, this research using the Sine Cosine algorithm for feature selection used for selecting the UNSW-NB15 dataset's optimal features. Once the detection of attack is made, the appropriate classification is made using the proposed Novel CVAR- k fold Cross validated Artificial neural network weighted Random Forest classification. The proposed system SC-CVAR attained the highest accuracy rate of 0.9987 for UNSW-NB15 Dataset and detection rate proved that the proposed method outperforms the existing techniques to classify and detect various attacks Thus, the system is restored effectively in minimal time.

References

[1] H. Alqahtani, I. H. Sarker, A. Kalim, S. M. M. Hossain, S. Ikhlaq, and S. Hossain, "Cyber intrusion detection using machine learning classification techniques," in International Conference on Computing Science, Communication and Security, 2020, pp. 121-131.

[2] Z. Ahmad, A. Shahid Khan, C. Wai Shiang, J. Abdullah, and F. Ahmad, *Network intrusion detection system: A systematic study of machine learning and deep learning approaches*, Transactions on Emerging Telecommunications Technologies, vol. 32, p. e4150, 2021.

[3] T. Saranya, S. Sridevi, C. Deisy, T. D. Chung, and M. A. Khan, *Performance analysis of machine learning algorithms in intrusion detection system: A review*, Procedia Computer Science, vol. 171, pp. 1251-1260, 2020.

[4] B. B. Zarpelão, R. S. Miani, C. T. Kawakani, and S. C. de Alvarenga, *A survey of intrusion detection in Internet of Things*,Journal of Network and Computer Applications, vol. 84, pp. 25-37, 2017.

[5] S. Pundir, M. Wazid, D. P. Singh, A. K. Das, J. J. Rodrigues, and Y. Park, *Intrusion detection protocols in wireless sensor networks integrated to Internet of Things deployment: Survey and future challenges*, IEEE Access, vol. 8, pp. 3343-3363, 2019.

[6] E. Benkhelifa, T. Welsh, and W. Hamouda, *A critical review of practices and challenges in intrusion detection systems for IoT: Toward universal and resilient systems*, IEEE Communications Surveys & Tutorials, vol. 20, pp. 3496-3509, 2018.

[7] S. K. Biswas, *Intrusion detection using machine learning: A comparison study*, International Journal of pure and applied mathematics, vol. 118, pp. 101-114, 2018.

[8] Gao, C., Zhu, H., & Guo, Y. (2012), "*Analysis and H. Liu and B. Lang, Machine learning and deep learning methods for intrusion detection systems: A survey*, applied sciences, vol. 9, p. 4396, 2019.

[9] J. Zhang, R. Gardner, and I. Vukotic, *Anomaly detection in wide area network meshes using two machine learning algorithms*, Future Generation Computer Systems, vol. 93, pp. 418-426, 2019.

[10] Y. He, S. Nazir, B. Nie, S. Khan, and J. Zhang, *Developing an efficient deep learning-based trusted model for pervasive computing using an LSTM-based classification model*, Complexity, vol. 2020, 2020.

[11] J. Ren, J. Guo, W. Qian, H. Yuan, X. Hao, and H. Jingjing, *Building an effective intrusion detection system by using hybrid data optimization based on machine learning algorithms*, Security and Communication Networks, vol. 2019, 2019.

[12] N. Bindra and M. Sood, *Detecting DDoS attacks using machine learning techniques and contemporary intrusion detection dataset,*

Automatic Control and Computer Sciences, vol. 53, pp. 419-428, 2019.

[13] Y.-F. Hsu, Z. He, Y. Tarutani, and M. Matsuoka, *Toward an online network intrusion detection system based on ensemble learning*, in 2019 IEEE 12th international conference on cloud computing (CLOUD), 2019, pp. 174-178.

[14] Hafiza Huma Taha, Syed Sufyan Ahmed, Haroon Rasheed, *Tumor Detection through Image Processing Using MRI*, International Journal of Scientific & Engineering Research, Volume 6, Issue 2, Feb.2015, pp.10-15

[15] H. Wang, Z. Cao, and B. Hong, *A network intrusion detection system based on convolutional neural network*, Journal of Intelligent & Fuzzy Systems, pp. 1-15, 2019.

[16] Y. Liu, S. Liu, and X. Zhao, *Intrusion detection algorithm based on convolutional neural network*, DEStech Transactions on Engineering and Technology Research, 2017.

[17] X. Yang and Z. Hui, *Intrusion detection alarm filtering technology based on ant colony clustering algorithm*, in 2015 Sixth International Conference on Intelligent Systems Design and Engineering Applications (ISDEA), 2015, pp. 470-473.

[18] W. L. Al-Yaseen, Z. A. Othman, and M. Z. A. Nazri, *Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system*, Expert Systems with Applications, vol. 67, pp. 296-303, 2017.

[19] H. Ji, D. Kim, D. Shin, and D. Shin, *A Study on comparison of KDD CUP 99 and NSL-KDD using artificial neural network*, in Advances in computer science and ubiquitous computing, ed: Springer, 2017, pp. 452-457.

[20] B. Selvakumar and K. Muneeswaran, *Firefly algorithm based feature selection for network intrusion detection*, Computers & Security, vol. 81, pp. 148-155, 2019.

[21] E. Min, J. Long, Q. Liu, J. Cui, and W. Chen, *TR-IDS: Anomaly-based intrusion detection through text-convolutional neural network and random forest*, Security and Communication Networks, vol. 2018, 2018.

[22] L. Xiao, X. Wan, X. Lu, Y. Zhang, and D. Wu, *IoT security techniques based on machine learning: How do IoT devices use AI to enhance security?*, IEEE Signal Processing Magazine, vol. 35, pp. 41-49, 2018.

[23] B. Riyaz and S. Ganapathy, *A deep learning approach for effective intrusion detection in wireless networks using CNN*, Soft Computing, pp. 1-14, 2020.

[24] J. Kim, J. Kim, H. Kim, M. Shim, and E. Choi, *CNN-Based Network Intrusion Detection against Denial-of-Service Attacks*, Electronics, vol. 9, p. 916, 2020

[25] B. Susilo and R. F. Sari, *Intrusion Detection in IoT Networks Using Deep Learning Algorithm*, Information, vol. 11, p. 279, 2020.

[26] J. Jeon, J. H. Park, and Y.-S. Jeong, *Dynamic Analysis for IoT Malware Detection with Convolution Neural Network model*, IEEE Access, 2020.

[27] M. Almiani, A. AbuGhazleh, A. Al-Rahayfeh, S. Atiewi, and A. Razaque, *Deep recurrent neural network for IoT intrusion detection system*, Simulation Modelling Practice and Theory, vol. 101, p. 102031, 2020.

[28] D. Zheng, Z. Hong, N. Wang, and P. Chen, *An improved LDA-based ELM classification for intrusion detection algorithm in IoT application*, Sensors, vol. 20, p. 1706, 2020.

[29] Q. Tian, D. Han, M.-Y. Hsieh, K.-C. Li, and A. Castiglione, *A two-stage intrusion detection approach for software-defined IoT networks*, Soft Computing, pp. 1-17, 2021.

[30] Z. Wang, Z. Xu, D. He, and S. Chan, *Deep logarithmic neural network for Internet intrusion detection*, Soft Computing, vol. 25, pp. 10129-10152, 2021.\

[31] . Gaifulina and I. Kotenko, *Selection of Deep Neural Network Models for IoT Anomaly Detection Experiments*, in 2021 29th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP), 2021, pp. 260-265.

[32]

J.Vimal Rosy has completed her Masters in Computer Science, Masters in Philosophy Computer Science, and is currently pursuing her Ph.D in Computer Science in the Field of Cloud Computing. She currently serves as an Head & Assistant Professor in the Department of Computer Science, Soka Ikeda College of Arts and Science for Women, Chennai, Tamil Nadu, India

Dr. S. Britto Ramesh Kumar is an Assistant Professor of Computer Science at St. Joseph's College (Autonomous), Tiruchirappalli. His research interests include software architecture, wireless and mobile technologies, information security and Web Services. He has published many journal articles and book chapters on the topics of Mobile payment and Data structure and algorithms. His work has been published in the International journals and conference proceedings, like JNIT, IJIPM, IEEE, ACM, Springer and Journal of Algorithms and Computational Technology, UK. He was awarded as the best researcher for the year 2008 in Bishop Heber College, Tiruchirappalli. He has completed a minor research project. He has visited countries like China, South Korea and Singapore.