A New Design of Custom Optimized Cnn-Lstm Assists to Detect Network Anomaly Using Categorical Data

Kanmani R (kanmanikrishna54@gmail.com)
Sri Krishna College of Technology
Dr.A.Christy Jeba Malar
Sri Krishna College of Technology
Roopa V
Sri Krishna College of Technology
Ranjani D
Sri Krishna College of Technology
Suganya R
Sri Krishna College of Technology

Research Article

Keywords: Anomaly detection, optimization, Long Short Term Memory and categorical data

Posted Date: June 3rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-490866/v1

A NEW DESIGN OF CUSTOM OPTIMIZED CNN-LSTM ASSISTS TO DETECT NETWORK ANOMALY USING CATEGORICAL DATA

Dr.R.Kanmani[1], Dr.A.Christy Jeba Malar[2], Ms.V.Roopa[3], Ms.D.Ranjani[4], Dr.R.Suganya[5]

Corresponding Author: Dr.R.Kanmani

Department of Information Technology[1-5], Sri Krishna College of Technology, Coimbatore.

Mail Id: kanmanikrishna54@gmail.com[1], a.christyjebamalar1@skct.edu.in[2], v.roopa1@skct.edu.in[3], d.ranjani1@skct.edu.in[4], r.suganya1@skct.edu.in[5]

Abstract:

For traditional intrusion detection model, the system effectiveness is fully based on training dataset and feature selection. During feature selection, it needs more labour charge and trusted mainly on expert's knowledge. Moreover, the training dataset contains more imbalanced data which in terms model tends to be biased. Here, an automatic approach is introduced to correct deficiency in the system. In this paper, the author proposes novel network anomaly detection (NID) build using categorical data. A model has to be designed with modified form of deep neural network primarily utilized for detecting anomaly within the network. Custom CNN-LSTM with Harris Hawks Optimization (named as custom optimized CNN-LSTM) is designed as a new classifier majorly used to detect the anomaly from word cloud to distinguish the data with effective performance. The experimental result shows that the proposed method achieves a promising output for network anomaly detection.

Keywords: Anomaly detection, optimization, Long Short Term Memory and categorical data.

I. Introduction:

Nowadays, an increasing availability of network technologies has been greatly influenced by spammers, cyber-attacks or intruders etc. This major thread becomes an unavoidable challenge for network security. Therefore, a defence system of network anomaly detection has become a great deal to solve this issue [1]. Anomaly is an undesirable behaviour insists to find the abnormalities within the network. For multipurpose applications, an anomaly patterns are often called by different names such as faults, noise, outliers, defects, damage, peculiarities, errors etc., [2]. Observation of anomalies detection is being considered as great advance in many different aspects. This detection process is extensively considered as an interesting active research emergence in the field of bio-surveillance, industrial process control, data entry errors, unauthorized access, fraud detection, fault diagnosis and so on [3, 4].

A major goal of network intrusion detection system (NIDS) is to distinguish the abnormal activities from the normal behaviour of the network. It is more flexible to detect a possible outcome of new emerging threads and also highly sensitive to fall in false alarm rate

[5]. Although, the other complication arises in the anomaly detection technique is that they exhibit failure in single point, speed scalability, detection rate etc.. To overcome these issues, network intrusion detection system using machine learning methodologies are introduced to enhance the system performance. Generally, three categories avail to fulfil the machine learning techniques are supervised, weakly-supervised and unsupervised. These are the predominant learning approaches helps to design several IDS system in various literatures [6-10]. A traditional model of machine learning such as Random Forest, Naïve Bayes, Support Vector Machine etc., has been well-adopted to design anomaly detection but they limited its performance evaluation while using unbalanced datasets.

In modern era, network security problem is being addressed and provide solution by using deep neural network. Multi-layer containing deep structure of convolutional neural network (CNN), deep belief network (DBN), Auto-encoders and Recurrent Neural Network (LSTM) are relatively provide promising solution in multiple disciplines [11].

Therefore, more robust deep learning techniques are required to accurately classify the network system. For detecting network anomaly, a deep architectural model of convolutional neural network (CNN) based meta-heuristic optimization is pre-trained using categorical data which are taken into practice. CNN is a well-established classifier technique withstands to conduct their experimentation in both balanced and unbalanced dataset. In data analysis applications, CNN outperform to distinct the classified data as normal and abnormal with huge training samples. Even though considering these benefits, CNN suffer from the problem of time-consuming procedure especially for patch-wise feature extraction [12]. An extended version of RNN is a longshort term memory (LSTM) that works as a time series dependencies adopted to convey their feature data into the CNN model. In this research work,a custom optimized CNN-LSTM with Harris Hawks optimization technique is employed to design the intrusion detection model. This model fit to provides effective result by discriminating the data with properly extracted features. Finally, the performance measure is employed to analyse the categorical data against several state-of-art techniques.

II. Related Works:

Some survey papers have been reviewed the machine learning methodologies to identify the anomaly (outlier). These article are discussed as follows,

Laskov et al. [13] identifies the unknown attacks by comparatively evaluate the accuracy performance of both supervised and unsupervised machine learning approaches. A survey took for network intrusion detection using supervised and unsupervised techniques by Ghorbani et al. [14]. Mahbod [15] and D. K. Bhattacharyya et al. [16] presented an anomaly based IDS which is probably compare their performance with several techniques such as SVM, Naïve Bayes etc., using NSLKDD dataset. Sandhya Peddabachigari et al. [17] presented a hybrid IDS with achieving reduced computational complexity and improved detection performance. It uses a hybrid intelligent system of combining decision tree with SVM (DT-SVM) which attains better generalization performance. As evident to shows that

the proposed technique in this paper could be performed well for identifying R2L, Probe rather than the DOS and U2R attacks.

Yasser Yasami et al. [18] proposed an unsupervised classifier to distinct normal and anomaly activities using ID3 DT and K-means clustering techniques. Training samples are partitioned into cluster and built each cluster by individual decision tree thereby enhancing the classification accuracy but this approach is restricted to specific database. Dewan et al presented a new learning algorithm based on Naïve Bayes and ID3 algorithm for NIDS [19]. A benchmark dataset of KDD99 with 5 classes are successfully investigated to improve the false positive rate using this technique. To most of the user attacks, it needs better improvement for false alarm rate. Another anomaly based IDS with hybrid approach is the enhanced data distribution presented by Roshan Chitrakar et al. [20]. For better classification, the technique established two combinations namely K-Medoids and Naïve Bayes. In this study, the aforementioned technique is well performed to improve the detection capabilities, false positive rate etc., than the K-mean clustering. The reason is that K-mean algorithm is extremely sensitive to anomaly since they loss huge value while distributing the data into each group. Instead of mean value, the centroid is considered to construct the K-Medoids which is highly dependable to group the similar behaviour as a cluster after then the clustered data is categorized into different classes of attack using Naïve Bayes classifier. The shortcoming occur in the data distribution model is that they are unable to predict the data in different environments. Moreover, it requires more time when the data size grows exponentially. In order to compensate the time complexity issues, Naïve Bayes classifier is replaced by adding support vector machine with the K-medoids clustering technique and it is proposed by Chitrakar et al. [21]. But this technique is applicable to reduce the time complexity for small-size data distribution not for large dataset.

Dino et al. [22] presented an outlier detection scheme of semi-supervised approach using categorical data. Distance learning of categorical attributes (DILCA) is a new approach performed to achieve a reasonable result. For unique framework, both ordinal and nominal data attributes are used to extent the research work with adding new active learning approach. Ashima et al. [23] designed a host based IDS incorporated with stacked CNN with Gated Recurrent Unit (GRU). It is highly feasible to minimize the training time and improves the intrusion detection system performance.

The research work is organized as follows, section III discuss the brief description of working methodology of our proposed model. Section IV and V explains the simulation work and conclusion part.

III. Methodology:

In order to design an accurate IDS system, the benchmark dataset of NSL-KDD is used to analyse the data. To build a model, the research work comprise of four stages namely pre-processing the data, feature extraction, classification by custom optimized CNN-LSTM and model evaluation. Anomaly detection system is an essential tool utilized to detect the abnormalities present within the network database. Since the database has a majority of

normal data and very low abnormal data [24]. It leads an unsafe and data integrity problem in the network, so NIDS technique is necessary to compensate the issues.

The input dataset comprises of both numerical and textual (categorical) content. Initially, the intrusion is detected in the cloud environment by loaded the test and trained data into the system. Before providing the input to the custom optimized CNN-LSTM, feature analysis is done to extract features and learn about the features. Some of the infrequent data in the dataset are removed as unused data and is sorted into dataset of reasonable size. Therefore, feature selection is an essential factor which selectively extracts most useful information by improving the learning rate accuracy and computational complexity. From the partitioned table, the text data is extracted for each of the labels by creating word cloud. To work with this extracted text data, pre-processing is done where the lowercase texts are converted and the punctuations are erased. Alternatively the words are encoding by building the document as padded and truncated to make them of the same length. Thus the concerned document is finally converted into sequences of numeric indices. At last, the converted data of important features are forward to the next stage.

Algorithm: Algorithm for proposed IDS model
Input : intrusion dataset
Output : Classification of intrusion
1. Begin
2. Load dataset
3. Partition the dataset for training and testing
4. Extract the text data into a table
5. Create wordcloud chart from a table
6. Pre-processing the training and testing data
7. Create a word embedding
8. Convert the documents into image indices
9. Create optimized custom layer
10. Set
11. Initialization parameters
12. Describe layers
13. Set an optimizer threshold value as 1
14. Set Initial learn rate as 1
15. Test the custom optimized CNN-LSTM network
16. Convert the text documents to sequences
17. Classify the documents
18. End

Custom optimized CNN-LSTM:

A hybrid approach of combining convolutional neural network with Long Short Term Memory (LSTM) that should be optimized by using Harris hawks algorithm. This new approach is proposed for NIDS and it is named as custom optimized CNN-LSTM.

The paper starts to describe the general view of optimization technique, Convolutional neural network and long-short term memory followed by the proposed description. The proposed block diagram is shown in Figure 1.

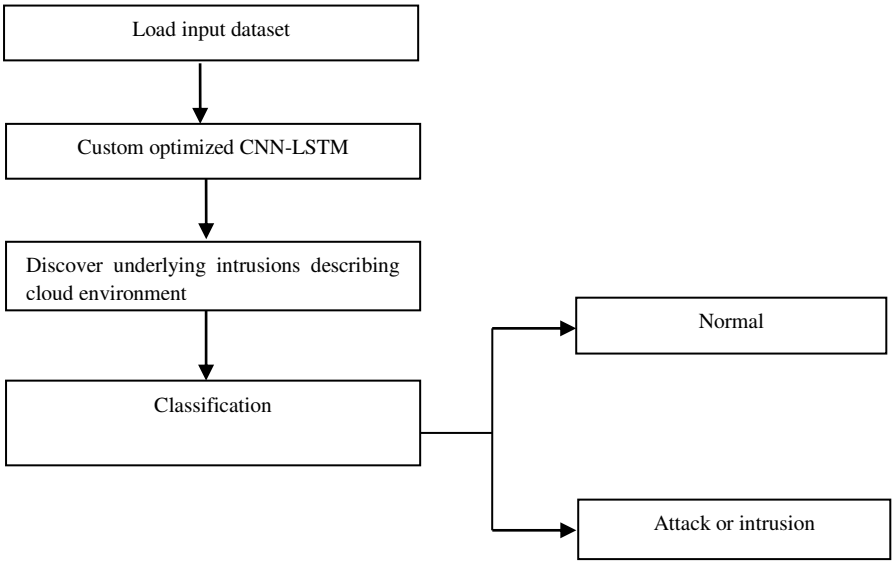

Figure 1. Block diagram for proposed methodology

Harris hawks optimization (HHO):

It is a meta-heuristic technique utilized to track the prey by both exploration and exploitation phases. Hawk is a most intelligent bird which perch on any location to search their prey (rabbit) with its powerful eyes. Number of hawks is the individuals required to use their strategy to find the optimal solution. During optimization, HHO uses three major stages to perform the operation namely perching on tall place, tracing the prey, attack them on different sorts [25, 26]. At exploration phase, hawk intended to monitor the prey by waiting for a long time and it is represented by,

$$Z(iter+1) = \begin{cases} Z_{rand}(iter) - r1[Z_{rand}(iter) - 2Z(iter)r2] & p \geq 0.5 \\ \left(Z_{prey}(iter) - Z_{mean}(iter)\right) - r3(lb + r4(ub - lb)) & p < 0.5 \end{cases} \quad (1)$$

Where, $Z_{mean}(iter) = \frac{1}{n}\sum_{j=1}^{n} Z_j(iter)$, Z_{mean} specifies the mean location of every hawk, $n\ and\ j$ are size and position of each hawk respectively.

Where $Z(iter)\ and\ Z_{prey}(iter)$ represents the present position of hawks and rabbit,$Z(iter+1)$is the next iteration of each hawk's position, r, q are random variables specified in [0, 1] interval and $lb\ and\ ub$ represents lower and upper bonds variables. At transforming from exploration to exploitation phase, it represents the external strength of the rabbit (S) and it is attained using equation (2)

$$S = 2S_0(1 - {iter}/{t_{max}}) \quad (2)$$

Where, S_0represents the physical strength of rabbit changes randomly over (-1, 1) interval. If the prey is actually energetic, then S_0 maximize its value from 0 to 1 or else minimize from 0 to -1. At exploitation phase, the hawks attempt to attack their prey in four possible strategies. Here r is the chance of rabbit when it successfully escape before pounce or not. During

chasing, the prey tries to escape from danger which in terms it significantly reduces its energy. At that moment, the hawks effortlessly capture the prey using the besiege process. The four possible pouncing process in exploitation phase is hard besiege, soft besiege, rapid dives while hard and soft besiege [31]. In this concern, the soft besiege occur when $|F| \geq 0.5$ and hard besiege occur when$|F| < 0.5$. The following condition from $r\ and\ |F|$ reveals the four strategies, whether the prey is escape from hawks or not.

Soft besiege & hard besiege: the condition is $r \geq 0.5\ and\ |F| \geq 0.5 \& r \geq 0.5\ and\ |F| < 0.5$

During this attempt, the prey tries to escape but can't. The hawk softly and hardly encircle the prey and make a surprise pounce. The mathematical expression for both besiege are modeled as,

$$Z(iter+1) = \Delta Z(iter) - F\left(IZ_{prey}(iter) - Z(iter)\right)\ ; for\ soft\ besiege \quad (3)$$

$$Z(iter+1) = Z_{prey}(iter) - F\left(\Delta Z(iter)\right)\ \ ; for\ hard\ besiege \quad (4)$$

Where, ΔZ indicates the position between rabbit and hawk in current iteration, $I = 2(1 - r5)$ which means the rabbit strength of randomly changing movement in each iteration.

Soft besiege with advanced rapid dives:

When $r < 0.5\ and\ |F| \geq 0.5$, the prey has a chance to escape since it contains enough energy but soft besiege is still there. The prey uses zigzag deceptive motion to change their direction in irregular form. This mechanism is named as levy flight (LF) model being utilized in HHO. The hawk observes the prey's strategies and progressively correct their direction based on previous move. The updated position of hawks using this mechanism is given by,

$$Z(iter+1) = \begin{cases} v = Z_{prey}(iter) - F\left(IZ_{prey}(iter) - Z(iter)\right)\ ; if\ v < f(Z(iter)) \\ u = Z_{prey}(iter) - F\left(IZ_{prey}(iter) - Z(iter)\right) + M \times LF(x)\ ;\ if\ u < f(Z(iter)) \end{cases} \quad (5)$$

where, M represents the random value of 1xd size, v indicates the soft besiege by hawks evaluating the next move, u represents irregular movement of rabbit based on LF approach.

Hard besiege with advanced rapid dives:

When$r < 0.5\ and\ |F| < 0.5$, the prey cannot escape from hawks since it doesn't have enough energy and also hard besiege make a shot of surface pounce on it. This strategy is similar to soft besiege with advanced rapid dives but the only difference is that hawk tries to minimize the average distance. The following equation is formulated below,

$$Z(iter+1) = \begin{cases} v = Z_{prey}(iter) - F\left(IZ_{prey}(iter) - Z_{mean}(iter)\right)\ ; if\ v < f(Z(iter)) \\ u = Z_{prey}(iter) - F\left(IZ_{prey}(iter) - Z(iter)\right) + M \times LF(x)\ ;\ if\ u < f(Z(iter)) \end{cases} \quad (6)$$

Convolutional Neural Network:

It is a neural architecture comprise of multi-hidden layers to extract the features from input to output layer. For deep image extraction, we have to use more hidden layer which in terms an improved result is generated at the end. Following stages arise within the network layer are convolutional layer, activation layer, pooling layer followed by fully-connected layer. Convolutional layer uses certain parameters such as kernel, padding, stride etc. to create a feature map. The corresponding output is generated based on convolving the input with the kernel. The next stage of ReLU activation layer is to improve the nonlinearity in feature map. After then, the pooling layer downsample the input dimensionality to minimize the parameters. Finally, FC layer collect each feature maps from previous layer to effectively classify the data.

Long-short term memory:

It is an extended version of RNN that works as a chain like structure with repeating module. LSTM has a capability to learn long term dependencies using feedback mechanism. A major purpose of this architecture is to forget or add information to the cell state. With regard to this information, it has both short and long term memory. To construct a LSTM layer, it consists of one memory state and three main gates of input, forget and output as shown in Figure 2. In order to sustain long terms dependency, the gate mechanism in LSTM network can decide which information to hold or leave [27].

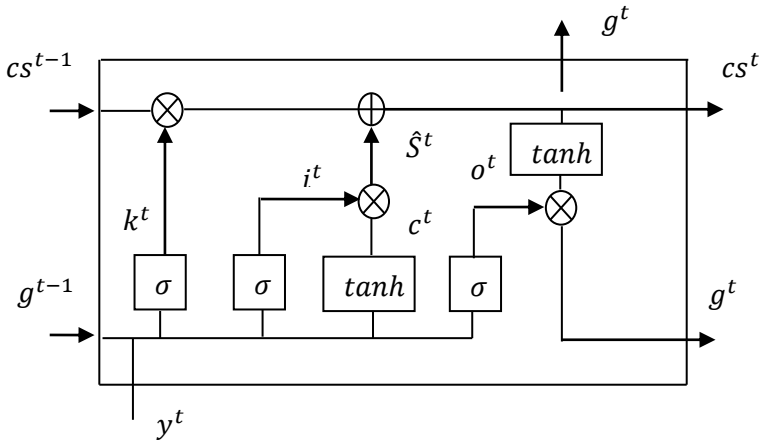

Figure 2. Internal architecture of LSTM

The below expression derives the individual stages of forget gate layer, input gate layer and output gate layer which computes to solve the problem of long term dependencies,

$$k^t = \sigma[w_k.(g^{t-1}, y^t) + b_k] \tag{7}$$

$$i^t = \sigma[w_i.(g^{t-1}, y^t) + b_i] \tag{8}$$

$$\hat{S}^t = \tanh[w_c.(g^{t-1}, y^t) + b_c] \tag{9}$$

$$o^t = \sigma[w_o.(g^{t-1}, y^t) + b_o] \tag{10}$$

$$cs^t = k^t * cs^{t-1} + i^t * \hat{S}^t \tag{11}$$

$$g^t = o^t * \tanh(c^t) \tag{12}$$

Where, $k^t, i^t \; and \; o^t$ are forget gate, input gate and output gate layer respectively, σ represents the sigmoid function, $w \; and \; b$ are the weight and bias representing the gate, $g^t \; and \; g^{t-1}$ specifies the hidden node at time $t \; and \; t-1$, cs^trepresents the memory state in which they hold new information as$\hat{S}^t$. The output of hidden and memory state are turns as an input to the next LSTM layer.

Proposed custom optimized CNN-LSTM:

A layer description of custom optimized CNN-LSTM model should be clearly explained in detail. It is a deep learning network composed of sequence input layer, convolution2dlayer, batch normalization, relu layer, 2 layer of lstm and dropout layer, fully connected layer, softmax and classification layer. It is a sequential layer operation carried out to process the previous layer output to the next layer input. Firstly, the extracted text document is a sequence index that act as an input to the convolutional layer. Each layer contains several neurons that constitute in the form of convolving the input with its filter size. This convolution process moves over the region to create several feature maps. For each region, the convolution layer size works with insist of parameters such as filter, padding, stride etc. Each region subset has been adjusting its parameters vertically and horizontally over the network to generate number of feature maps. To improve the convergence speed, the batch normalization is utilized to normalize each training samples of batch data and is place in between the convolutional and relu activation function. In optimized layer, the Harris hawks optimization is deployed as an activation function by even replacing relu. An optimization technique is defined within the predict function which in terms a proper functioning is activated via each layer. The predict function is given by,

$$[y_1, y_2, y_3, \ldots y_n] = predict(layer, x_1, x_2, x_3, .. x_m) \tag{13}$$

Where, $x_1 \ldots x_m$ is the mthneuron output from the previous layer which propagates to the next layer of nthneuron $y_1 \ldots y_n$. Dropout is placed in between 2 LSTM layer followed by connecting another dropout layer. With its special memory state, LSTM has been able to hold long temporal data for better extraction process. It uses 125 hidden_units for LSTM1, 100 hidden_units for LSTM2 and 0.2% for dropout layer. By adding dropout next to LSTM layer may have a possibility to memorising something and also reduce overfitting within the network. Finally, the output layer is trained by connecting each neuron output to the single flatten-wise layer called fully connected layer. It adjusted its weight and bias value to train the categorical samples and consequently, softmaxlayer of activation function assign its real vector in between 0 and 1, so that they easily interpret its value as a probability function. The formula for soft-max is given by,

$$Softmax = \frac{e^{fi}}{\sum_{n=1}^{k} e^{fn}} \tag{14}$$

Where, k represents each classes within the network, f_i specifies the ith output vector of dense layer. Finally, the classification layer mutually gathers data from softmax to establish the resultant output.

During training, the system train the network effectively to computes the validation accuracy and loss value at regular time interval so that the performance is evaluated properly.

IV. Experimental Result and Discussion:

Network anomaly detection is classified using custom optimized CNN-LSTM model. In this experimentation, NSL-KDD dataset of categorical text document is used to distinct the normal and abnormal data. Due to this context, the proposed technique started to train the document of sequence indices so that the accuracy and loss functions are validated. For training process, the categorical data is partitioned into training, testing and validation data and also its protocol type, number of attacks are tabulated in Table 1,

Table 1. NSL-KDD dataset defined for analysis purpose

	Protocol	Attack Count	Training Set	Testing Set	Validation	Total
Categorical data	3	38	54,078	11, 587	11,588	77,253

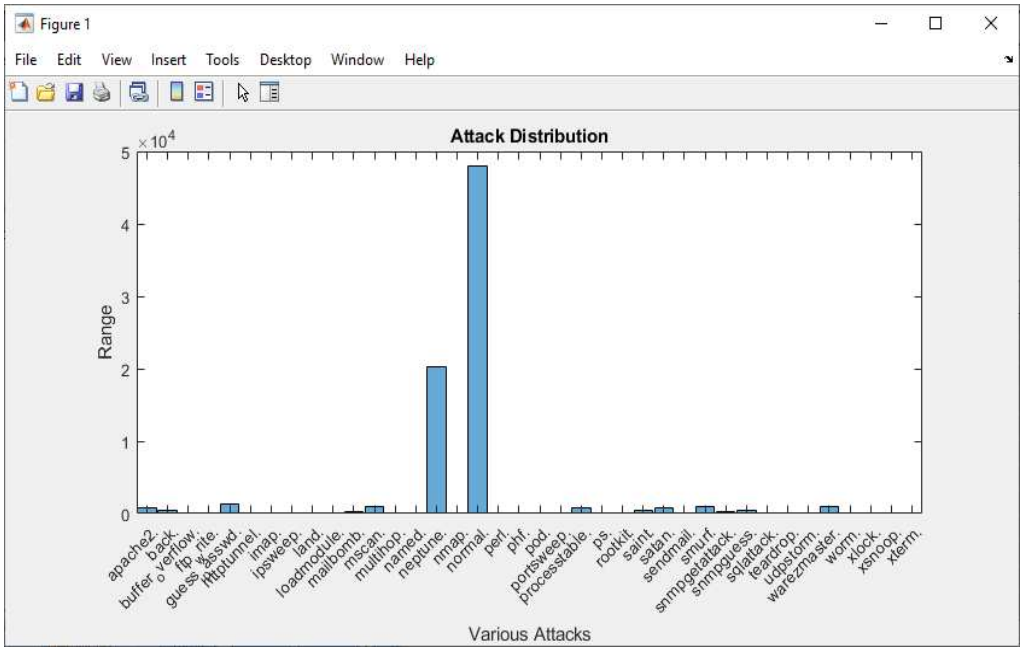

Figure 3. Data distribution from NSL-KDD dataset

To detect intrusion from cloud environment, the NSL-KDD data contains several attacks with maximum of normal data. Various attacks are positioned from the categorical dataset as classified events and are depicted in Figure 3.

Before initiate the process, the dataset is partitioned into set of training and testing samples. Then frequent data of important features are only extracted and is stored in word cloud. To visualize the text data by word cloud and the chart is depicted in Figure 4. It uses protocol types such as tcp, udp and icmp standards to hold the data within the cloud. Additionally, the service provided for the corresponding protocol are also created in word cloud as shown in Figure 5.Pre-process the text is the next phase which erase the unwanted distortion in the document.

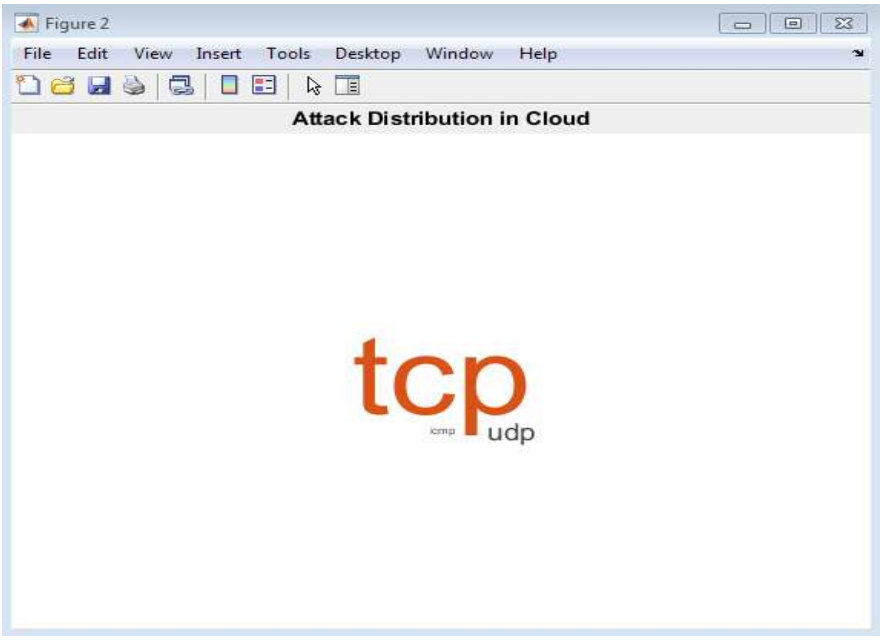

Figure 4. Create wordcloud chart from a table

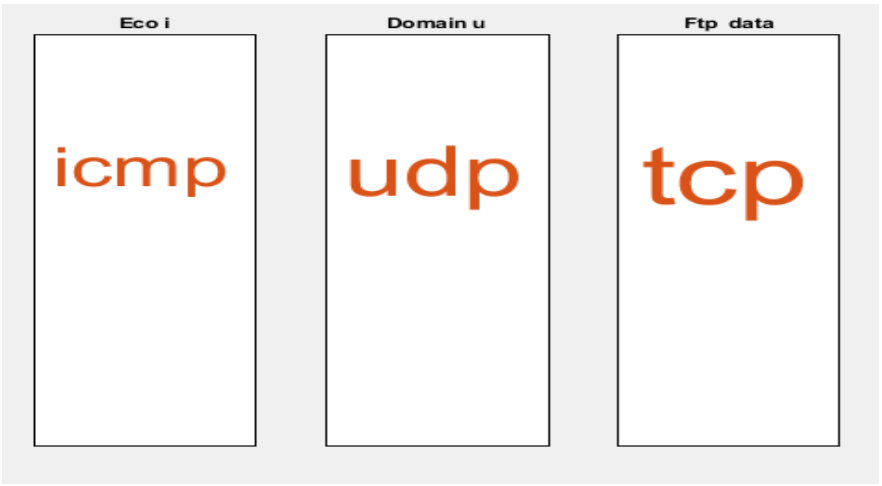

Figure 5. Services for each of the protocol type in wordcloud

After embedding the document, the training progress is created using custom optimized CNN-LSTM. It is a deep learning network composed of sequence input layer, convolution2dlayer, batch normalization, relu layer, 2 layer of lstm and dropout layer, fully connected layer, softmax and classification layer to train the model. The training options involved within the network is prior to use 1024 mini-batch size, 0.0005 L2 regularization and maxEpochs of 3. To create and train lstm network, they set input size as 1, embedded dimension=100, hidden_units1= 125 and hidden_units2= 100. Set the base learning rate as 1 and lesser it after every two epochs by a drop factor of 0.2% learning rate. During training, the system continuously monitors the accuracy by specifying validation frequency and data.

The validation accuracy for proposed technique achieves 98.24% at epoch 3 of 150th iteration. When training the samples, the proposed technique improves its performance level of accuracy at every iteration whereas loss factor gradually reduces. As compared to other state-of-art techniques, the proposed approach obtains greater performance. The training progress of NSL-KDD dataset using custom optimized CNN-LSTM is depicted in Figure 6.

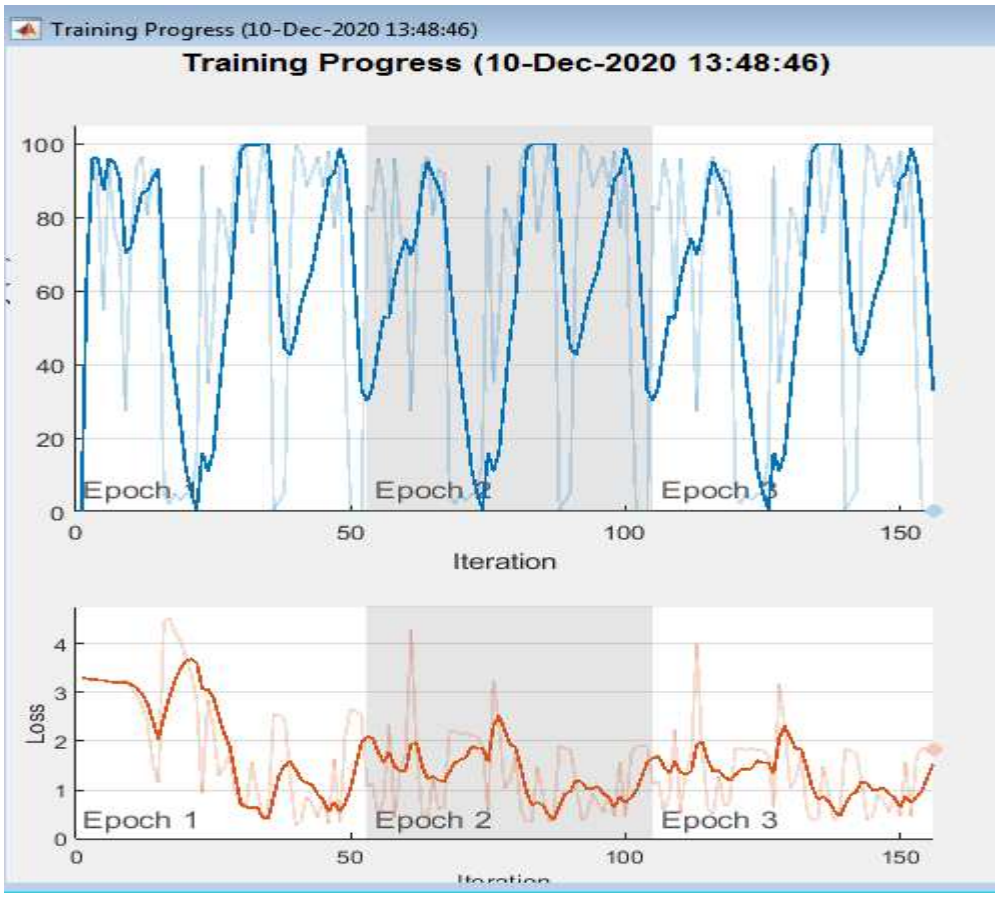

Figure 6. Training progress of proposed technique

Validation loss is another factor which shows 3.2982% at 1st iteration, 2.2795% at 50th iteration, 0.4176% at 100th iteration and 0.4639% at 150th iteration for the proposed technique. Comparatively analysing the loss function, the proposed observed an improved result than the existing techniques of CNN [28], LSTM [29] and CNN+LSTM [30]. The

validation accuracy, mini-batch loss and validation loss of proposed and existing techniques are comparatively simulated and plotted its resultant value in table 2.

Table 2.Evaluating values of accuracy and error rate for different techniques

Techniques	Epoch	Iteration	Accuracy (%)	Mini Batch loss (Error rate)	Validation Loss (%)
Proposed	1	1	0.39	0.0058	3.2982
	1	50	0	0.0056	2.2795
	2	100	94.82	0.0926	0.4176
	3	150	98.24	0.0015	0.4639
CNN	1	1	2.54	11.8538	5.9526
	1	50	54.49	1.9389	-
	1	52	83.89	0.5267	2.4418
	2	100	7.32	9.9479	-
	2	104	96.29	0.1371	3.5605
	3	150	96.29	0.3438	-
	3	156	99.32	0.0213	4.3303
CNN +LSTM	1	1	0.00	3.4161	2.0354
	1	50	0.00	2.5859	-
	1	52	0.00	2.4993	1.3062
	2	100	93.36	0.5469	-
	2	104	0.00	1.7340	1.1662
	3	150	98.05	0.5196	-
	3	156	0.00	1.4776	1.1289
LSTM	1	1	56.05	3.2897	3.2759
	1	50	60.74	1.3337	1.2545
	2	100	61.62	1.2052	1.2169
	3	150	61.23	1.2223	1.2099

Furthermore, the performance is evaluated for proposed and exiting techniques that should be tells the effectiveness of the system. It is measured by means of using parameters such as accuracy, sensitivity, precision, specificity and F-score. Precision is more evident to correctly predict normal data among total predicted document and it is expressed in eqn (16). Sensitivity or recall is formulated as a ratio of correctly predicted normal data to total normal data within the dataset as in eqn (17). Specificity actually corresponds to correctly predict the abnormal data among the abnormal categories and is formulated in eqn (18). Accuracy is defined as a ratio of correctly predict both normal and abnormal data among the total data as in eqn (15). F-score is defined as a ratio of weighing both recall and precision into single measure and is expressed in eqn (19).

$$Accuracy = {}^{True\ Positive + True\ Negative}/_{Total\ Sum} \tag{15}$$

$$Precision = {}^{True\ Positive}/_{True\ Positive + False\ Positive} \tag{16}$$

$$Sensitivity = {True\ Negative}/{False\ Positive + True\ Negative} \quad (17)$$

$$Specificity = {True\ Positive}/{True\ Positive + False\ Negative} \quad (18)$$

$$F\_score = 2 * \frac{Precision * Sensitivity}{Precision + Sensitivity} \quad (19)$$

Figure 7, shows that the performance measure for proposed and existing techniques with reasonable values. When comparing the result, the proposed technique provides better performance than the other approaches. Precision is assessed for the overall techniques and it shows 0.6201 for LSTM, 0.4522 for CNN, 0.0581 for CNN+LSTM and 0.6202 for proposed method. Accuracy is evaluated for the proposed technique and it improves its effectiveness by 36%, 40.64% and 35.89% than the LSTM, CNN and CNN+LSTM respectively.

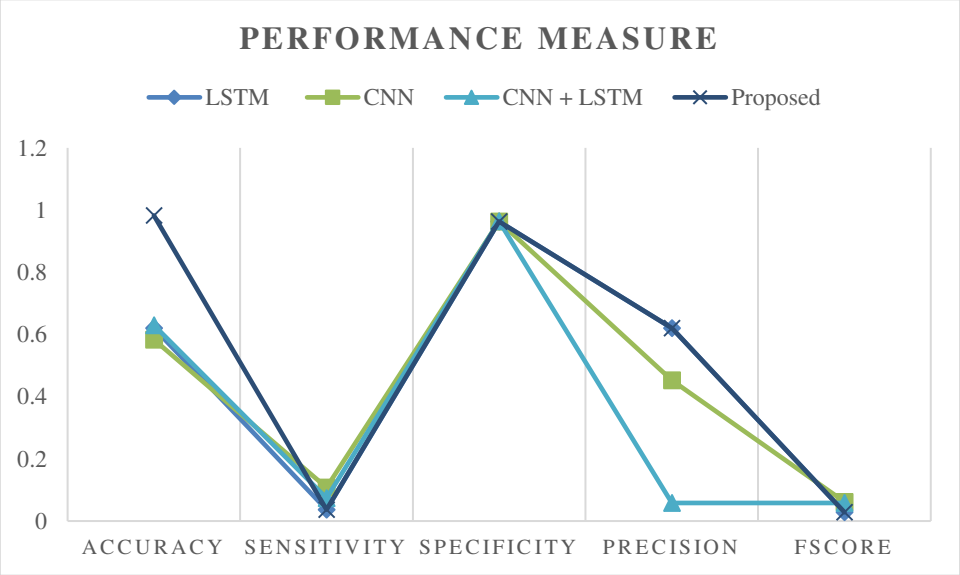

Figure 7. Performance measure using NSL-KDD dataset

V. Conclusion:

This study introduced a new deep network of custom optimized CNN-LSTM effectively used for detecting anomaly within the network. It is designed in the form of connecting LSTM and Harris hawks optimization within the convolutional neural network, which is capable of providing learned temporal parameters and proper functioning made to improve the system performance. In order to enhance the generalization capability of the proposed system, NSL-KDD dataset is used for investigation. From the simulated result, the proposed model of custom CNN-LSTM with Harris hawks optimization classified the categorical data as a validation accuracy of 98.24%. The system effectiveness is proved by measuring the performance metrics and it shows that the proposed technique improves its accuracy by 36%, 40.64% and 35.89% than the existing techniques of LSTM, CNN and CNN+LSTM respectively.

Funding: There is no funding source.

Conflict of Interest: The authors declare that they have no conflict of interest

References:

[1] Monowar H. Bhuyan, Dhruba K. Bhattacharyya, and Jugal K. Kalita, "Survey on incremental approaches for network anomaly detection." *International Journal of Communication Networks and Information Security (IJCNIS)*, Vol. 3, No. 3, 2011

[2]PrasantaGogoi, D. K. Bhattacharyya, Bhogeswar Borah, and Jugal K. Kalita, "A survey of outlier detection methods in network anomaly identification" *The Computer Journal* 54, no. 4, pp. 570-588, 2011

[3] Dino Ienco, Ruggero G. Pensa, and Rosa Meo, "A semi-supervised approach to the detection and characterization of outliers in categorical data", *IEEE transactions on neural networks and learning systems* 28, no. 5, pp.1017-1029, 2016

[4] U. Porwal and S. Mukund, ``Credit card fraud detection in e-commerce: An outlier detection approach,"*arXiv:1811.02196*, 2018

[5] Songlin Dai, Jubin Yan, Xiaoming Wang, and Lin Zhang, "A Deep One-class Model for Network Anomaly Detection." In *IOP Conference Series: Materials Science and Engineering*, vol. 563, no. 4, p. 042007, IOP Publishing, 2019.

[6] Taeshik Shon, and Jongsub Moon. "A hybrid machine learning approach to network anomaly detection." *Information Sciences* 177, no. 18, pp. 3799-3821, 2007

[7] Ying Gao, Yu Liu, Yaqia Jin, Juequan Chen, and Hongrui Wu, "A novel semi-supervised learning approach for network intrusion detection on cloud-based robotic system." *IEEE Access* 6, pp. 50927-50938, 2018

[8] Hyunseung Choi, Mintae Kim, Gyubok Lee, and Wooju Kim, "Unsupervised learning approach for network intrusion detection system using auto-encoders", *The Journal of Supercomputing* 75, no. 9, pp. 5597-5621, 2019

[9] FrancescoPalmieri, Ugo Fiore, and Aniello Castiglione. "A distributed approach to network anomaly detection based on independent component analysis." *Concurrency and Computation: Practice and Experience* 26, no. 5, pp.1113-1129, 2014

[10] Hamed Haddad Pajouh, GholamHosseinDastghaibyfard, and SattarHashemi. "Two-tier network anomaly detection model: a machine learning approach." *Journal of Intelligent Information Systems* 48, no. 1, pp. 61-74, 2017

[11] SherazNaseer, YasirSaleem, Shehzad Khalid, Muhammad Khawar Bashir, Jihun Han, Muhammad Munwar Iqbal, and Kijun Han. "Enhanced network anomaly detection based on deep neural networks." *IEEE Access* 6, pp. 48231-48246, 2018

[12] Mohammad Sabokrou, Mohsen Fayyaz, MahmoodFathy, Zahra Moayed, and ReinhardKlette, "Deep-anomaly: Fully convolutional neural network for fast anomaly

detection in crowded scenes." *Computer Vision and Image Understanding* 172, pp. 88-97, 2018

[13]PavelLaskov, Patrick Düssel, Christin Schäfer, and KonradRieck' "Learning intrusion detection: supervised or unsupervised?", In *International Conference on Image Analysis and Processing*, pp. 50-57,Springer, Berlin, Heidelberg, 2005.

[14] A. A. Ghorbani, W. Lu, and M. Tavallaee, "Network Intrusion Detection and Prevention", *Advances in Information Security*, vol. 47. Boston, MA, USA: Springer, 2010.

[15] MahbodTavallaee, "An adaptive hybrid intrusion detection system", PhD diss., University of New Brunswick, Faculty of Computer Science, 2011.

[16] D. K. Bhattacharyya and J. K. Kalita, Network Anomaly Detection: A Machine Learning Perspective. Boca Raton, FL, USA: CRC Press, 2013.

[17] SandhyaPeddabachigari, Ajith Abraham, CrinaGrosan, and Johnson Thomas, "Modeling intrusion detection system using hybrid intelligent systems", *Journal of network and computer applications* 30, no. 1, pp. 114-132, 2007

[18] Yasser Yasami, and Saadat Pour Mozaffari, "A novel unsupervised classification approach for network anomaly detection by k-Means clustering and ID3 decision tree learning methods", *The Journal of Supercomputing* 53, no. 1, pp. 231-245, 2010.

[19] DewanMdFarid, NouriaHarbi, and Mohammad Zahidur Rahman. "Combining naive bayes and decision tree for adaptive intrusion detection." *arXiv preprint arXiv:1005.4496*, 2010.

[20] RoshanChitrakar, and Chuanhe Huang. "Anomaly based intrusion detection using hybrid learning approach of combining k-medoids clustering and naive bayes classification", In *2012 8th International Conference on Wireless Communications, Networking and Mobile Computing*, pp. 1-5. IEEE, 2012.

[21] Chitrakar. R, .Chuanhe,H., "Anomaly detection using Support Vector Machine classification with k-Medoids clustering", *In Proceedings of IEEE Third Asian Himalayas International Conference on Internet (AH-ICI*), p. 1-5, 2012

[22] Dino Ienco, Ruggero G. Pensa, and Rosa Meo, "A semi-supervised approach to the detection and characterization of outliers in categorical data" *IEEE transactions on neural networks and learning systems* 28, no. 5, pp. 1017-1029, 2016

[23] Ashima Chawla, Brian Lee, Sheila Fallon, and Paul Jacob."Host based intrusion detection system with combined CNN/RNN model", *In Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 149-158. Springer, Cham, 2018.

[24] ChiiranjiiLalChowdhary, AbhiishekRanjan, and D. S. Jat, "Categorical Database Information--Theoretic Approach of Outlier Detection Model", *Annals. Computer Science Series* 14, no. 2, 2016.

[25] Ali AsgharHeidari, SeyedaliMirjalili, HossamFaris, Ibrahim Aljarah, MajdiMafarja, and Huiling Chen, "Harris hawks optimization: Algorithm and applications." *Future Generation Computer Systems* 97, pp. 849-872, 2019

[26] HosseinMoayedi, AbdolrezaOsouli, Hoang Nguyen, and Ahmad Safuan A. Rashid, "A novel Harris hawks' optimization and k-fold cross-validation predicting slope stability." *Engineering with Computers*, pp. 1-11, 2019

[27] Prerna Singh, and PritiSehgal, "GV Black dental caries classification and preparation technique using optimal CNN-LSTM classifier", *Multimedia Tools and Applications*, pp. 1-18, 2020

[28] Taejoon Kim, Sang C. Suh, Hyunjoo Kim, Jonghyun Kim, and Jinoh Kim, "An encoding technique for CNN-based network anomaly detection" In *2018 IEEE International Conference on Big Data (Big Data)*, pp. 2960-2965, IEEE, 2018.

[29] R. Vinayakumar, K. P. Soman, and PrabaharanPoornachandran, "Long short-term memory based operation log anomaly detection", In *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 236-242, IEEE, 2017.

[30] Mostofa Ahsan, and Kendall E. Nygard, "Convolutional Neural Networks with LSTM for Intrusion Detection" In *CATA*, pp. 69-79, 2020.

[31] Christy Jeba A. Malar, Deva M. Priya, and Sengathir Janakiraman, "Harris hawk optimization algorithm-based effective localization of non-line-of-sight nodes for reliable data dissemination in vehicular ad hoc networks" *International Journal of Communication Systems* 34, no. 1, 2021.

Figures

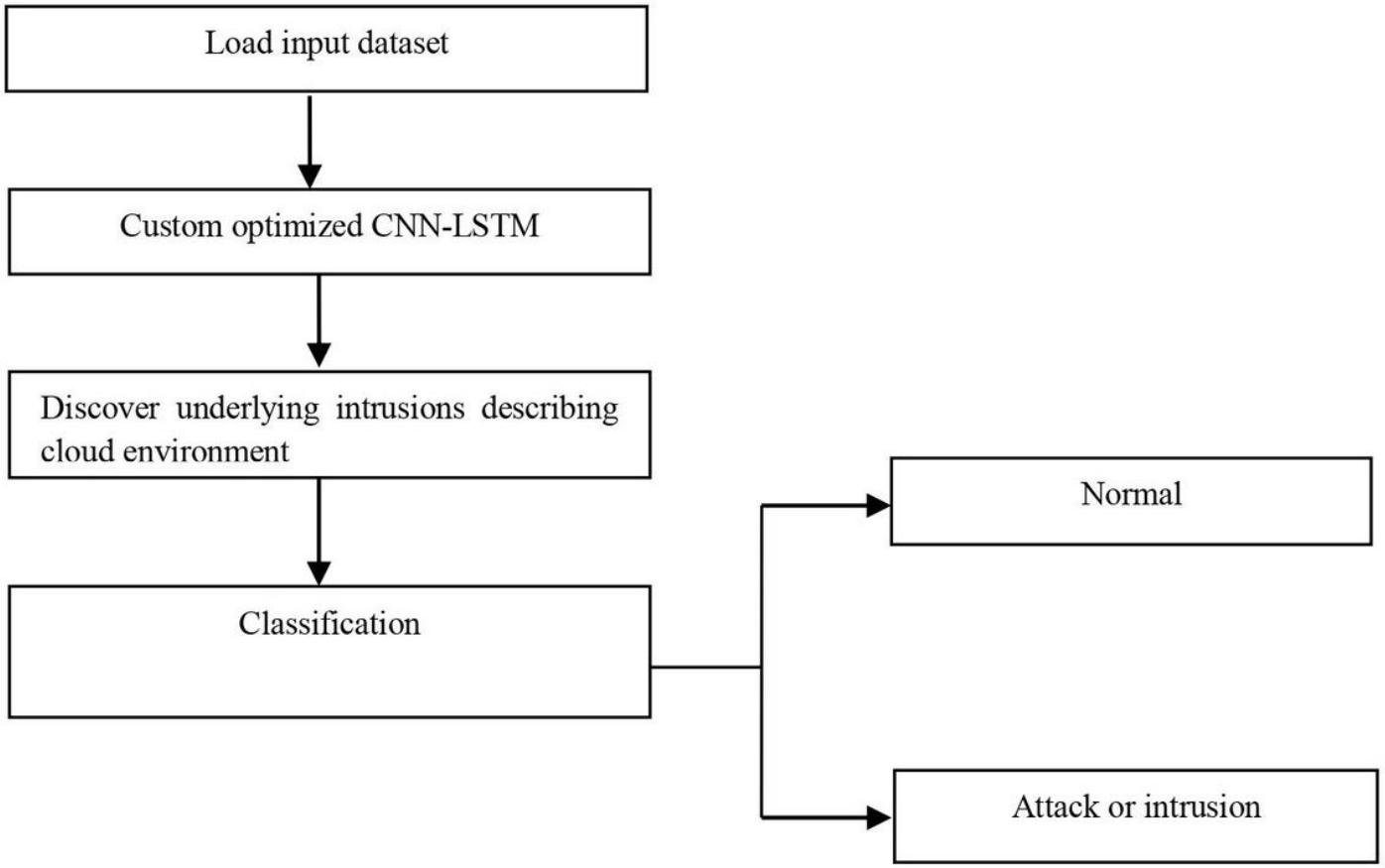

Figure 1

Block diagram for proposed methodology

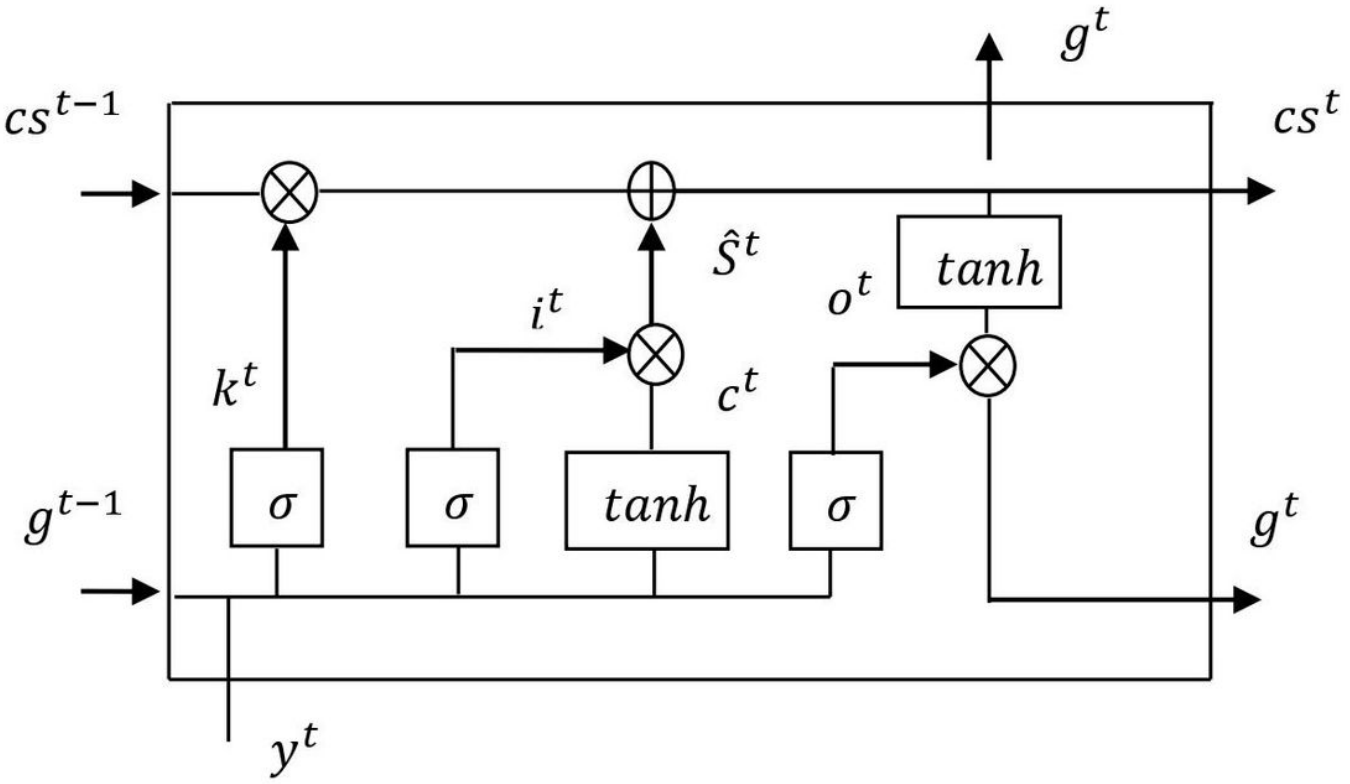

Figure 2

Internal architecture of LSTM

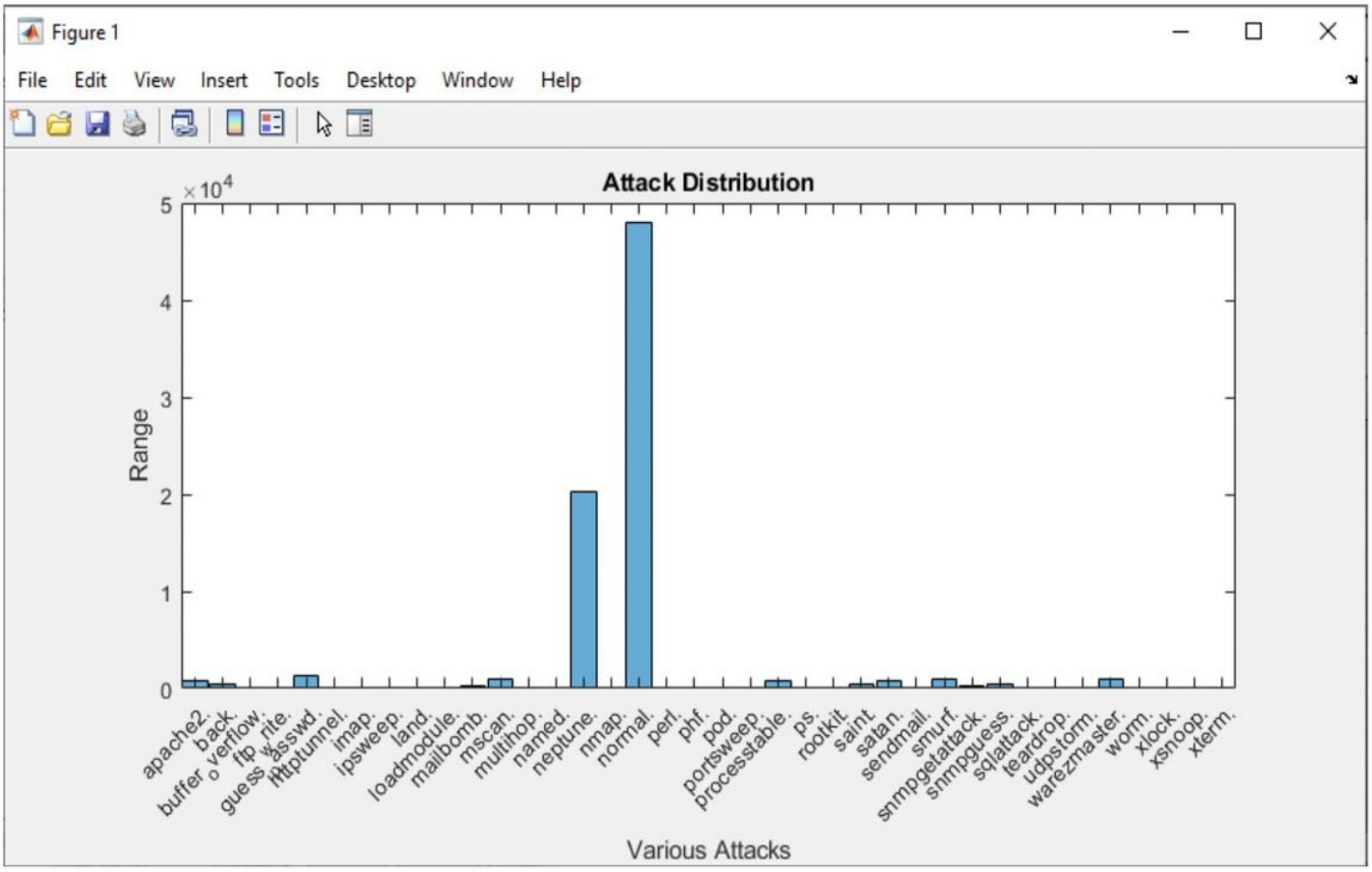

Figure 3

Data distribution from NSL-KDD dataset

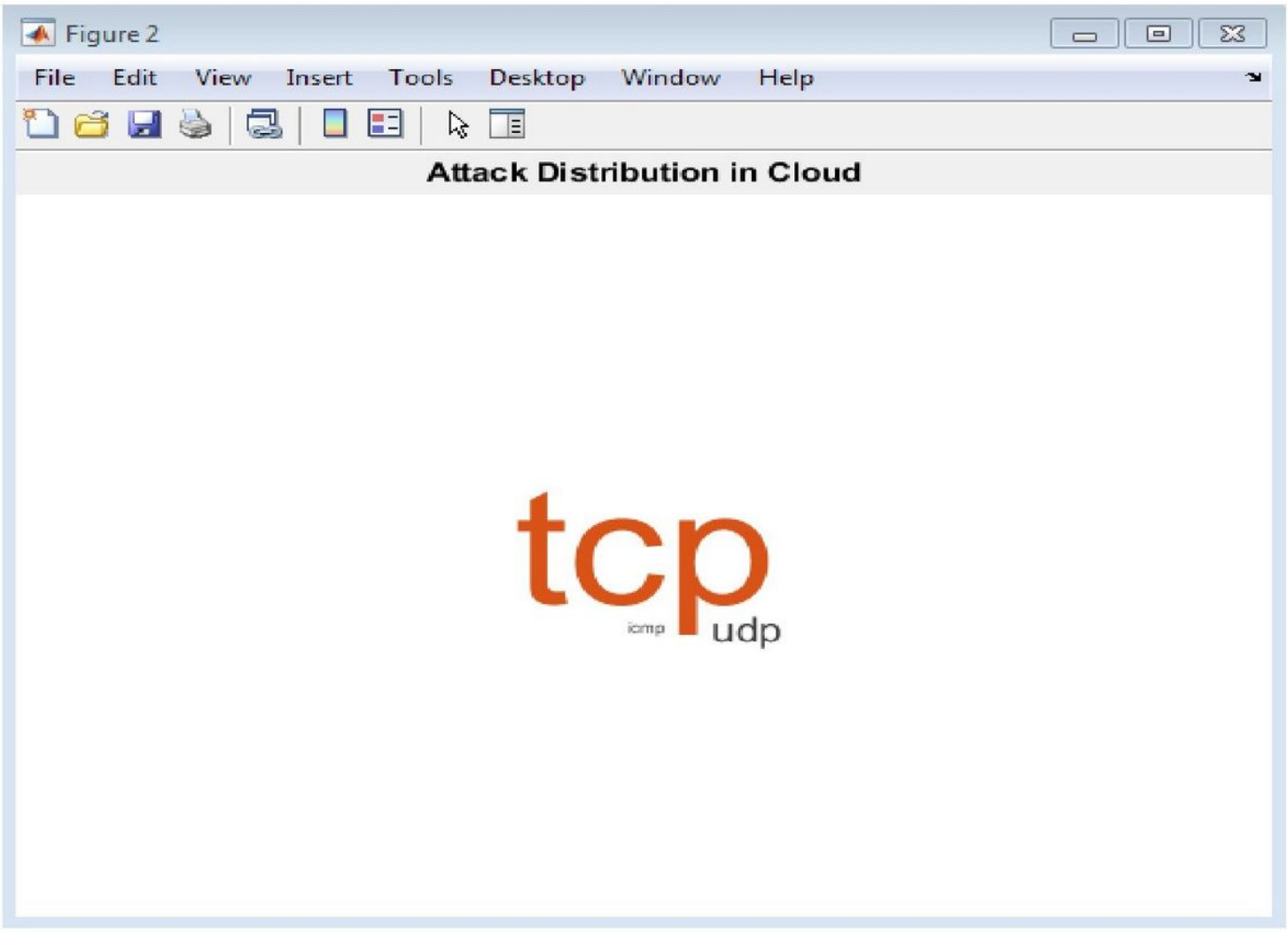

Figure 4

Create wordcloud chart from a table

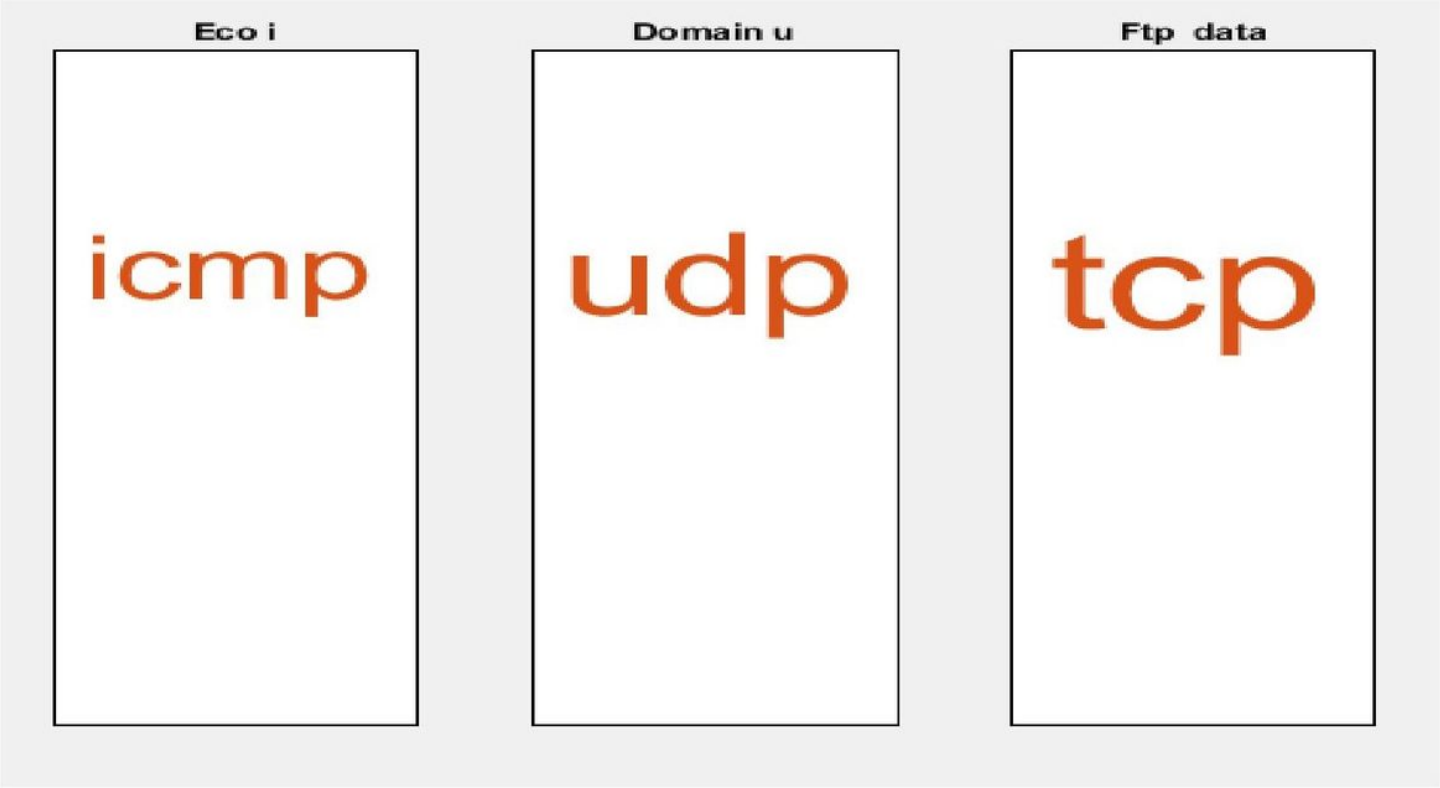

Figure 5

Services for each of the protocol type in wordcloud

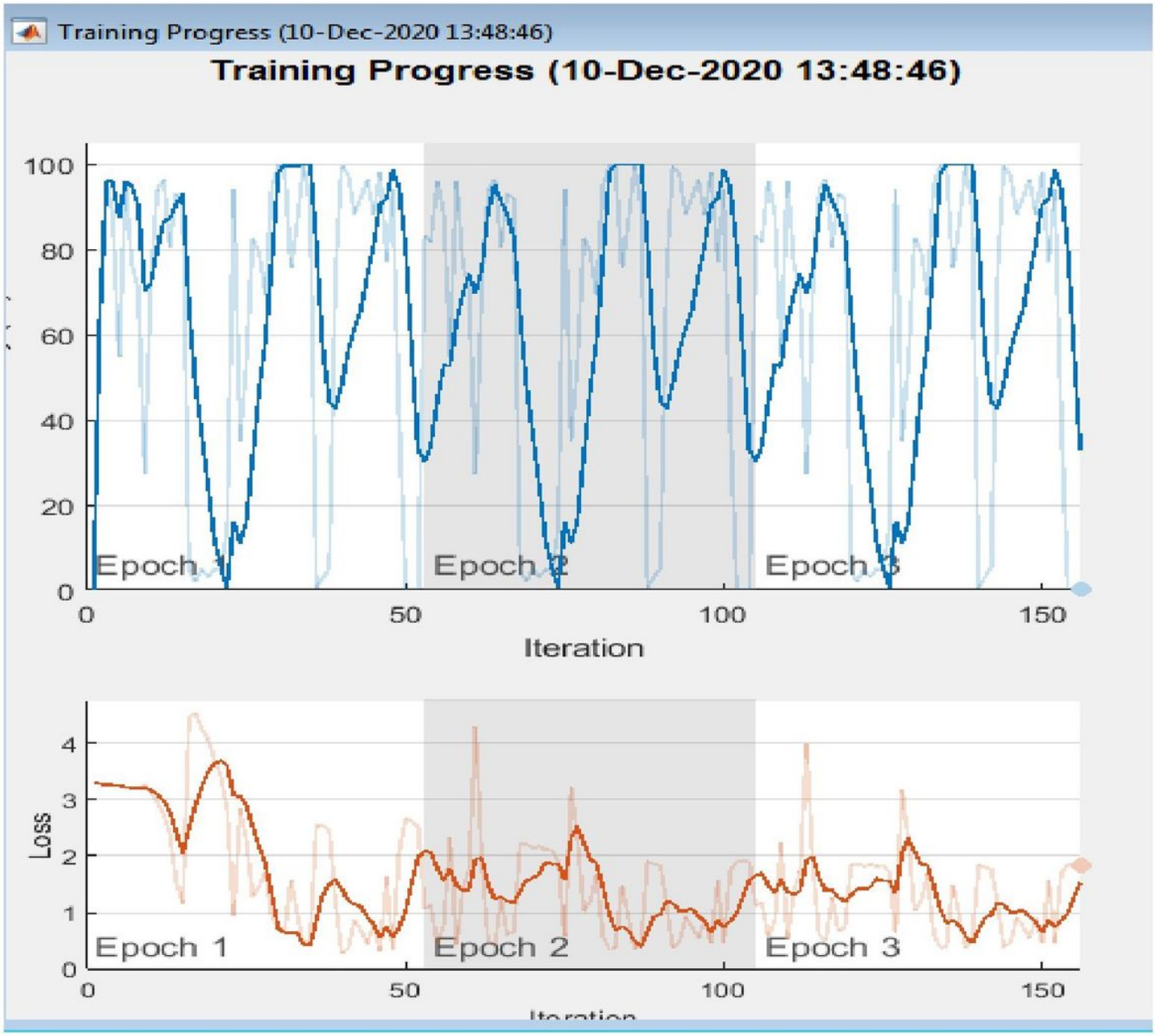

Figure 6

Training progress of proposed technique

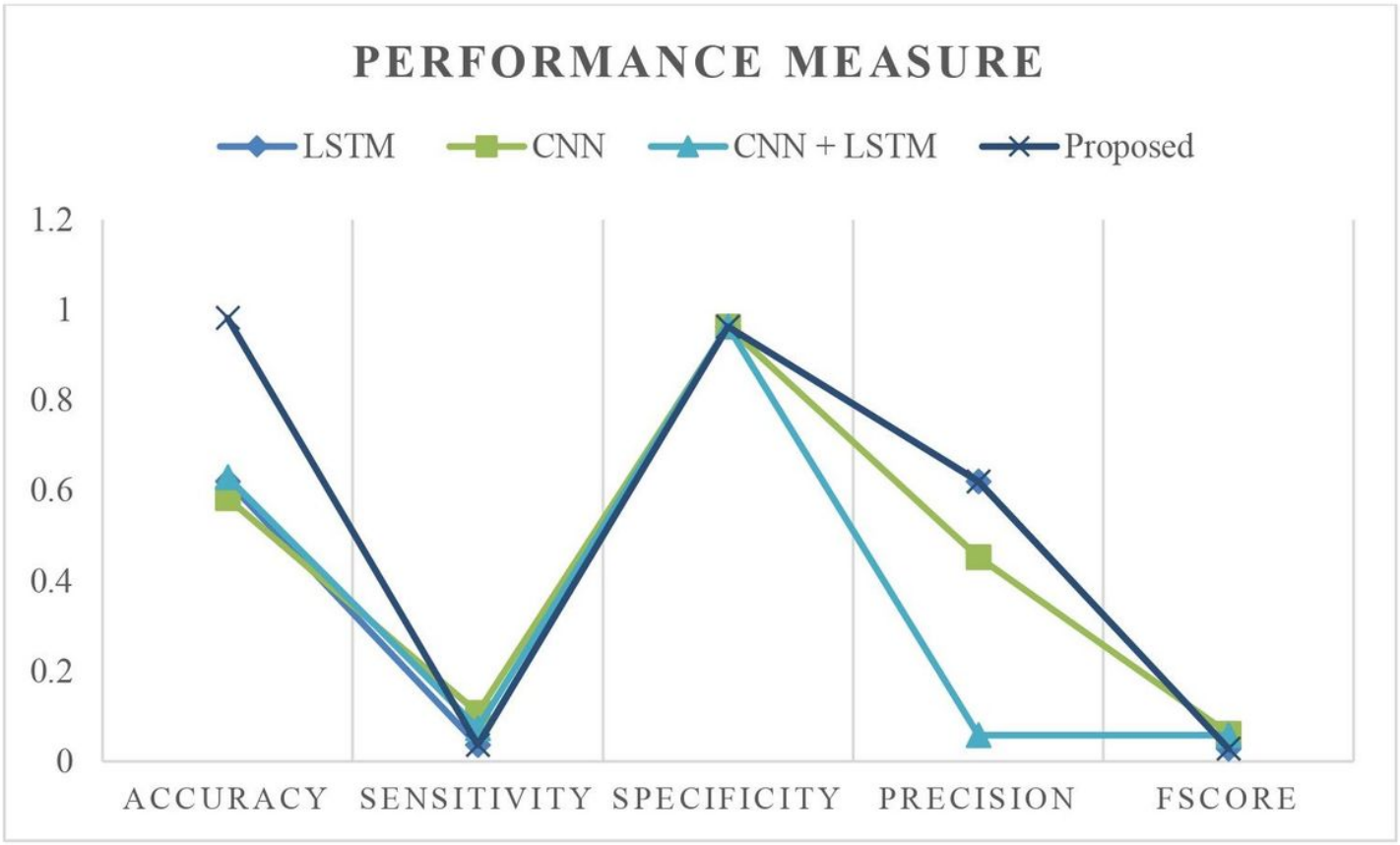

Figure 7

Performance measure using NSL-KDD dataset