ISSN 1816-6075 (Print), 1818-0523 (Online)
Journal of System and Management Sciences
Vol. 12 (2022) No. 4, pp. 398-412
DOI:10.33168/JSMS.2022.0424

Dirichlet Feature Embedding with Adaptive Long Short-Term Memory Model for Intrusion Detection System

R. Sivakami[1], G Uday Kiran[2], M. Arun[3], Leo Gertrude David[4], M. Manohar[5]

[1]Department of Computer Science and Engineering, Sona College of Technology, Salem, Tamil Nadu 636005, India

[2]Department of Computer Science and Engineering, B V Raju Institute of Technology, Narsapur - 502313, Telangana, India

[3]Department of Electronics and Communication Engineering, Panimalar Institute of Technology, Chennai

[4]Department of Visual Communication, Kumaraguru College of Liberal Arts and Science, Coimbatore, India

[5]Department of Computer Science and Engineering, CHRIST (Deemed to be University), Bangalore - 560029, Karnataka, India

arunmemba@ieee.org

Abstract. Intrusion Detection System is applied in the network to monitor the network activity and detect the intruder to protect the user data. Various existing models have been applied in the intrusion detection system and have the limitations of high False Alarm Rate (FAR), overfitting problem and data imbalance problem. In this research, Dirichlet Feature Embedding based Adaptive Long Short Term Memory (DFE-LSTM) model is proposed to improve the efficiency of the intrusion detection. The Dirichlet Feature Embedding (DFE) method is applied to effectively represent the feature to analysis the multi-variate of the input data. The enhanced Adaptive Long Short Term Memory (ALSTM) model is applied to select the optimal parameter for the LSTM model to improve the learning rate. The proposed DFE-ALSTM model is compared to three datasets such as UNSW-NB15, NSL-KDD and Kyoto 2006+ for evaluate the efficiency. The proposed DFE-ALSTM model has the accuracy of 94.32 % and existing NB-SVM has 93.75 % accuracy in intrusion detection on UNSW-NB15 dataset.

Keywords: adaptive long short-term memory, Dirichlet feature embedding, false alarm rate, intrusion detection system, and NSL-KDD.

1. Introduction

Recently, security in computer network is main concern due to development of internet services and technologies at a rapid pace (Rahman et al., 2020). Various possibilities are provided by computer technology developments including control systems and remotely manage ability, a multitude of information of online sources is opening up a gateway. Intrusion Detection System (IDS) is a tool to identify the intruder in the network and block them to increases the security of the network. IDS monitor single system or computer network to monitor the intruder and increase the security in the system. Development in network technologies and internet led to considerable increases in intrusion and attacks. Attacks detection and prevention is an important part of security that helps to protect user information in the network. In computer networks, Intrusion detection system is important way to distinguish the attacks and prevent attacks Manzoor et al., 2017). This is important to obtain a realistic performance evaluation prior to deploy IDS. Common problems face by research community in IDS are systematic metric to measure performance of IDS and realistic dataset for evaluation (Haider et al., 2017). An IDS is first line of defense in a computer network and both IDS operate using two techniques namely signature based recognition and anomaly based detection. Traditional IDSs have limitations of reduced efficiency in detecting new form of attacks and a high false alarm rate in the detection (Kasongo et al., 2020).

Machine learning techniques were commonly applied in IDS due to efficiency in pattern analysis. Some of commonly applied machine learning techniques such as Artificial Neural Networks (ANNs), Decision Tree, Support Vector Machine (SVM), and K-Nearest Neighbor (KNN) (Lv et al., 2020). Comparison to traditional machine learning algorithm, Deep learning based algorithm were applied in IDS to deal with large datasets and effectively handles the features. Some features are relevant and applied to solve a specific classification problem, others are redundant and not required (Almiani et al., 2020). Feature Extraction has gained a great deal of momentum in Deep learning and machine learning research. Feature extraction is the process of extracting or combining inputs to increases particular classifier performance (Alazzam et al., 2020). Existing IDS has the limitations of less efficiency, more false alarm rate and overfitting problem in the deep learning methods (Jin et al., 2020; Eskandari et al., 2020). The DFE-ALSTM method is applied to increases the efficiency of intrusion detection in network. The DFE-ALSTM model objectives and contribution are discussed as below.

1. The DFE is applied to analysis the multi-variate of data to effective represent the network data that increases detection accuracy and reduce the FAR. The DFE method has the capacity to analysis the multi-variate of data and existing Naïve Bayes method has capacity to analysis the uni-variate of data.

2. The LSTM model is used in this method as classifier and these store relevant features for long term in the network. The LSTM model also has the capacity to analysis the input data in the backward manner.
3. The enhanced Average Stochastic Gradient Descent (ASGD) is used to estimate optimal parameter settings for LSTM model. The enhanced ASGD method improves model learning rate and also helps to increases detection performance.
4. Three datasets such as UNSW-NB15, NSL-KDD and Kyoto 2006+ were applied to evaluate the model performance. The proposed DFE-ALSTM model is compared with existing Naïve Bayes-SVM model to evaluate the efficiency.

This paper is formulated as literature survey of recent methods in intrusion detection system is discussed in section 2, the explanation of proposed DFE-ALSTM model is given in section 3, the results of the intrusion detection model is given in section 4, and conclusion is given in section 5.

2. Literature Survey

Intrusion Detection System examines the features of the network such as network traffic, data rate etc., to identify the attacker in the network. Recent methodologies in the intrusion detection system were reviewed in this section for better understanding of the existing models.

Guo, et al. (2016) applied hybrid technique to reduce false alarm rate and improve detection rate of misuse and anomaly detection component. The low computing complexity of anomaly detection is developed and k-Nearest Neighbor (kNN) algorithm for misuse detection model. The misuse detection method coordinates with intrusion detection method to improve learning performance and reduce false alarm rate. The Kyoto University benchmark dataset and KDD'99 datasets are applied to examine hybrid model with a low positive rate. The hybrid method has lower efficiency in handling large dataset and irrelevant features selected by the model affects the efficiency.

Wang, et al. (2017) proposed Support Vector Machine (SVM) with augmented features to increases intrusion detection efficiency. The density ratios of logarithm marginal are applied to transform the new and better-quality features to increases the intrusion detection method. Two models are developed named as LMDRT-SVM and LMDRT-SVM2 models are developed for intrusion detection system. Intrusion detection performance is evaluated using NSL-KDD dataset in intrusion detection. The SVM model with transformed features has the higher performance in NSL-KDD dataset. The LMDRT-SVM model has the higher performance in binary classification and multi-class classification performance is needed to be carried out to evaluate its efficiency.

Ahsan, et al. (2019) proposed Kernel Density Estimation (KDE) and fast Minimum Covariance Determinant (MCD) for the multivariate control chart to increases outlier detection and reduce the false alarm rate. The Fast-MCD method improves the capacity of the control chart model to quickly and accurately detect the outliers. The developed KDE-CL method in intrusion detection performance is evaluated in three datasets such as UNSW-NB 15, NSL-KDD, and KDD'99 datasets. The developed model has higher performance in three datasets and also has higher efficiency than existing model. The developed model eliminates the relevant features in the classification that degrades model efficiency.

Zhang, et al. (2019) proposed unified model based on LSTM-CNN to improve the performance of intrusion detection system. The features of spatial temporal were applied for the classification in the intrusion system. The benchmark dataset of UNSW-NB15 was applied to evaluate model efficiency. The model effectively analysis the features and improve the efficiency of intrusion detection. The CNN model is compared with LSTM-CNN and analysis shows that developed MSCNN-LSTM model has higher performance. The model effectively handles imbalance dataset and improves the performance. The model creates the overfitting in the classification and overfitting problem is needed to be solved for effective performance.

Gu, et al. (2020) applied Naive Bayes feature embedding technique and SVM model is applied to improve efficiency of model. Original data is applied with Naïve Bayes feature embedding and transform into high quality data. Transformed features are applied in SVM classifier in the model. The NSL-KDD, CICIDS2017, and UNSW-NB15 benchmark datasets were applied in model for evaluation. The SVM with Naïve Bayes model has the higher efficiency in terms of false alarm rate, accuracy and detection rate. The relevant features are eliminated in the analysis that affects the performance of the SVM with Naïve Bayes model in the intrusion detection.

3. Method

Intrusion Detection System is required in the network to detect the intruder in the network and protect the user data. Various existing models have been applied for intrusion detection system and existing models have limitations of data imbalance, overfitting, and higher FAR. The DFE-ALSTM model is proposed to increases the detection performance in intrusion detection system. The DFE method is applied to analysis the multi-variate of input data and provides effective representation of input data. The enhanced ASGD method is applied in LSTM to optimal parameter to improve the learning rate and detection performance. The LSTM model is applied as classifier based on the DFE method to evaluate detection performance in intrusion detection system. The DFE-ALSTM model flow diagram is shown in the Figure 1.

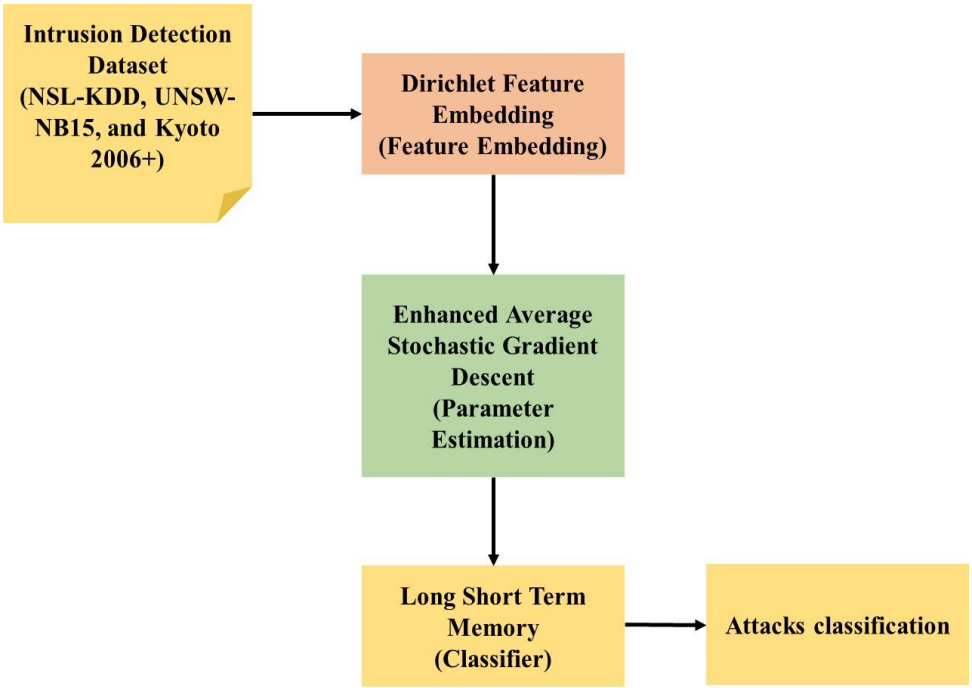

Fig. 1: The flow diagram of DFE-ALSTM.

3.1. Dirichlet feature embedding

Scalar data is used as feature embedding technique and applied Dirichlet feature embedding to compute the data. For classification in univariate problem, density ratio of marginal is considered as power classifier. The original classification problem of decision boundary x is find using a threshold in $\frac{f(x)}{g(x)}$. Naïve Bayes independence assumption, generalization is applied to convert univariate to multivariate. The density ration of corresponding marginal of each feature is transformed independently and newly obtained features are trained together for classifier.

This process is not suitable for compositional data and this is applied to deal with scaler features. Multiple components of compositional feature and each composition carry relative information and has no real meaning. The algorithm for the Dirichlet Feature embedding is shown below.

Algorithm: Dirichlet Feature Embedding

Input: $S = (X_{n\times p}, Y)$ with n samples; p compositional data feature and class label $Y \in \{0,1\}$

Output: $S_2^{new} = (X^{2'}, Y^2)$

1. Randomly split S into two mutually exclusive subsets with equal size and same class distribution: $S_1(X_{\frac{n}{2}}^1, Y^1)$, $S_2 = (X_{\frac{n}{2}}^2, Y^2)$
2. Split X^1 into X^{1+} and X^{1-} according to Y^1
3. For $j = 1,2,\dots,p$ do
    a. Estimate class conditional Dirichlet densities $\hat{f}_j(x_j)$ and $\hat{g}_j(x_j)$ using MLE on x_j^{1+} and x_j^{1-}, respectively
4. End for
5. For $j = 1,2,\dots,p$ do
    a. Feature transformation on X_j^2 in S_2 using $\log \hat{f}_j(x_j^2) - log\hat{g}_j(x_j^2)$ and denote the new feature by $x_j^{2'}$
6. End for
7. Merge $x_1^{2'}, \dots, x_p^{2'}, Y^2$ by column to obtain the transformed data

3.2. Long short term memory

Network intrusion of one-step ahead prediction requires latest and historical data. LSTM hidden layer has self-feedback technique and this has handles dependence problem in long term (Sherstinsky 2020). Three gates in LSTM handles storing of information in LSTM memory cell (Chimmula et al., 2020; Kim et al., 2019). The LSTM unit structure is shown in Figure 2.

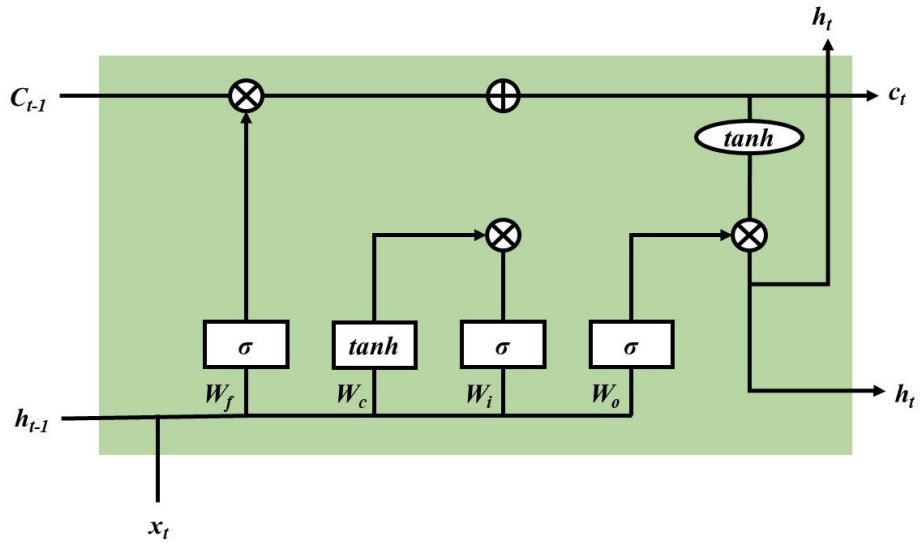

Fig. 2: The LSTM unit structure

The input data is x_t at time t, previous moment LSTM cell output is h_{t-1}, memory cell value is c_t, , and the output LSTM cell is h_t. The LSTM calculation process is divided into steps as:

The LSTM output unit h_t is calculated, as in equation (1)

$$h_t = o_t * \tanh(c_t) \quad (1)$$

Calculate output gate o_t, output gate bias is b_0, output gate weight matrix is W_0, state value of memory cell is control by output gate, as in equation (2).

$$o_t = \sigma(W_0.[h_{t-1}, x_t] + b_0) \quad (2)$$

Calculate the Memory cell c_t of current moment value, and state value of last LSTM unit is c_{t-1}, as in equation (3).

$$c_t = f_t \times c_{t-1} + i_t * \widetilde{c_t} \quad (3)$$

Forget gate f_t value is measured, forget gate bias is b_f, forget gate weight matrix is W_f, forget gate controls state value memory cell of historical data, as in equation (4).

$$f_t = \sigma(W_i.[h_{t-1}, x_t] + b_f) \quad (4)$$

Input gate i_t is calculated, input gate bias is b_i, input gate weight matrix is W_i, sigmoid function is σ, as in equation (5). The input gate handles state value of memory cell in current input data.

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \quad (5)$$

Candidate memory cell $\widetilde{c_t}$ is calculated, candidate memory cell bias is b_c, candidate memory cell weight matrix is W_c, as in equation (6).

$$\widetilde{c_t} = \tanh\left(W_c.[h_{t-1}, x_t] + b_c\right) \tag{6}$$

LSTM handles sequence of data based on memory cell and control gate to decide for read, update and reset timestamp information. LSTM internal parameters output dimensions are controlled by weight matrix dimensions and sharing parameters.

3.3. Parameter estimation – optimization

Deep learning models training using Stochastic Gradient Descent across various modalities including reinforcement learning, natural language processing, and computer vision (Merity et al., 2017). Non-convex optimization problems are applied to train deep networks, as in equation (7).

$$\min_{w} \frac{1}{N} \sum_{i=1}^{N} f_i(w) \tag{7}$$

Where expectation is taken over the data, network weights is denoted as w, ith data point of loss function is f_i. SGD iteratively takes form steps and learning rates sequence is denoted as γ_k, as in equation (8).

$$w_{k+1} = w_k - \gamma_k \widehat{\nabla} f(w_k) \tag{8}$$

Where stochastic gradient is denoted $\widehat{\nabla}$ and iteration number is denoted in subscript. Mini-batch of data points are computed using SGD. SGD method perform well in practice and provide several properties such as better generalization performance, saddle point avoidance, and linear convergence. Tradition SGD without momentum in neural language modeling of specific task outperforms other algorithms such as RMSProp, Adagrad, Adam, and SGD in margin of statistical significance. The Averaged SGD (ASGD) further improves training process and ASGD method provides surprising results in asymptotic second-order convergence. ASGD consider identical in equation (1) instead of last iterate as solution, returns $\frac{1}{K-T+1} \sum_{i=T}^{K} w_i$, where user-specified averaging trigger is $T < K$, total number of iterations is K. The algorithm for enhanced ASGD is shown below.

Algorithm: Enhanced ASGD

Inputs: Initial point w_0, learning rate γ, logging interval L, non-monotone interval n.

1. Initialize $k \leftarrow 0, t \leftarrow 0, T \leftarrow 0, logs \leftarrow []$
2. While stopping criterion not met do
3. Compute stochastic gradient $\widehat{\nabla} f(w_k)$ and take SGD
4. If $mod(k, L) = 0$ and $T = 0$ then
5. Compute Validation perplexity v
6. If $t > n$ and $v > \min_{l \in \{t-n, \ldots, t\}} logs[1]$ then
7. Set $T \leftarrow k$

8. End if
9. Append v to logs
10. $t \leftarrow t + 1$
11. End if
12. End while
13. Return $\frac{\sum_{i=T} w_i}{k-T+1}$

4. Results

Intrusion Detection System is required in the network to monitor the network and identify the intruder in the network to protect the user data. Various deep learning and machine learning based models were applied in intrusion detection system for effective detection of the model. Existing intrusion detection techniques have the limitation of data imbalance problem, higher FAR, and overfitting. The DFE-ALSTM model is proposed in intrusion detection system to enhance efficiency.

Datasets: The UNSW-NB15, NSL-KDD and Kyoto 2006+ datasets were applied to examine the DFE-ALSTM model efficiency. The three datasets details were given in the Table 1.

Table 1: Data instances on three datasets.

	Number of Instances	
Datasets	Normal	Attacks
UNSW-NB15 dataset	56,000	119,341
NSL-KDD	67343	58630
Kyoto 2006+	113120	118191

Metrics: The three metrics are need to be measured such as False Alarm Rate (FAR), Detection Rate (DR), and Accuracy, from the DFE-ALSTM model. The equations of FAR, DR, and Accuracy are given in Equation below.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{9}$$

$$DR = \frac{TP}{TP+FN} \tag{10}$$

$$FAR = \frac{FP}{FP+TN} \tag{11}$$

System Configuration: The DFE-ALSTM model is executed in system of 22 GB Graphics card, 128 GB of RAM, Intel i9 processor, and 1 TB hard disk. The DFE-ALSTM model is implemented using the Python 3.7 tool.

Parameter Settings: The number of iteration is set as 100 and number of epoch is set as 50 in the DFE-ALSTM model. The learning rate of DFE-ALSTM model is set as 0.001 and decay rate is set as 0.0001.

Table 2: The performance analysis on UNSW-NB15 dataset.

Methods	Accuracy (%)	DR (%)	FAR (%)
LMDRT - SVM [12]	93.3	94.98	8.55
LMDRT - SVM2 [12]	92.85	95.96	10.58
KDE - CL [13]	91.01	99.69	27.48
MSCNN [14]	91.4	92.3	15.5
EnSVM [15]	91.19	96	14.13
DT - EnSVM [15]	93.35	93.65	6.99
DT - EnSVM2 [15]	93.25	93.56	7.1
NB - SVM [15]	93.75	94.73	7.33
NB - SVM2 [15]	93.35	95.27	8.78
DFE - ALSTM	94.32	97.23	6.64

Table 2 evaluate existing and DFE-ALSTM model on UNSW-NB15 dataset and compared. The DFE-ALSTM model has the higher performance in terms of accuracy, DR and FAR compared to existing model in the intrusion detection. The proposed DFE-ALSTM model has the advantage of effectively represent the network data based on the Dirichlet Feature Embedding. The LSTM in the proposed model effectively analysis the input data, relevant features are store in long term and also analyze in the backward manner. The LSTM model has the effective performance in large datasets of the intrusion detection and existing SVM has lower efficiency. The enhanced ASGD selects optimal parameter for LSTM to increase the learning rate. The existing NB - SVM2 model has the second higher performance in the intrusion detection due to its effective data representation based on Naïve Bayes. The proposed DFE – ALSTM model has the effectively represent the multi-variate data and existing Naïve Bayes has lower efficiency in representing the multi-variate data. The proposed DFE – ALSTM model has the higher DR of 97.23 % and existing NB-SVM2 has 95.27 % in the intrusion detection.

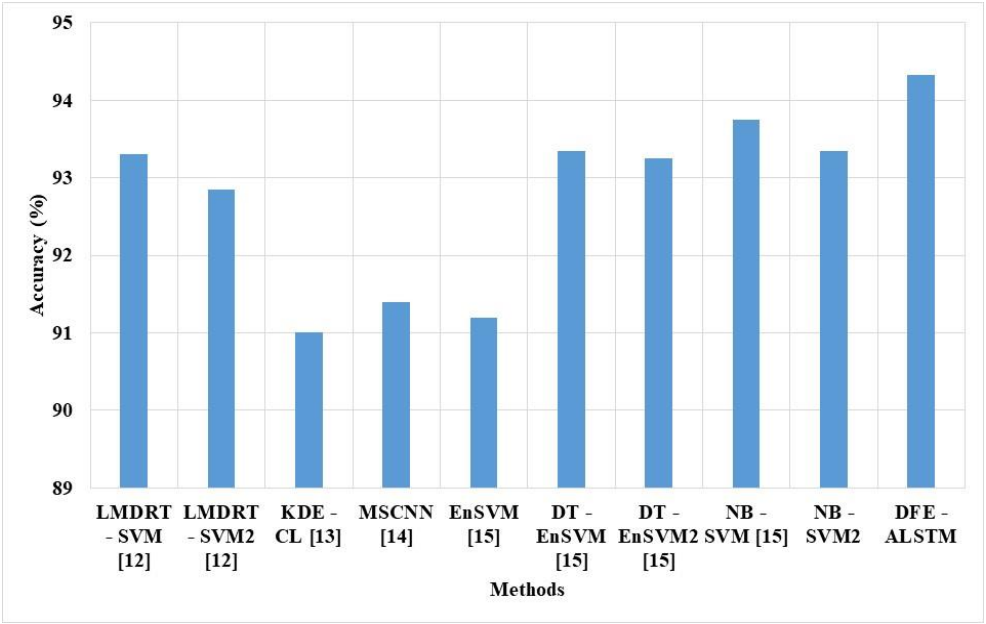

Fig. 3: Accuracy of the proposed and existing models in on dataset of UNSW-NB15.

The accuracy of the DFE – ALSTM and existing models on dataset of UNSW-NB15 is shown in Figure 3. The DFE – ALSTM has the higher accuracy compared to existing intrusion detection model compared. The Dirichlet Feature Embedding method in the proposed DFE – ALSTM model has the effective representation of multi-variate of the input data. The enhanced ASGD method and Dirichlet Feature Embedding method improves the accuracy of the model. The proposed DFE – ALSTM model has the 94.32 % and existing NB-SVM has 93.75 % accuracy in intrusion detection.

Table 3: Performance analysis on NSL-KDD dataset.

Methods	Accuracy (%)	DR (%)	FAR (%)
LMDRT - SVM2 [12]	99.28	99.16	0.61
LMDRT - SVM [12]	99.31	99.2	0.6
DT - EnSVM2 [15]	99.41	99.09	0.31
NB - SVM2 [15]	99.36	99.25	0.54
NB - SVM [15]	99.35	99.24	0.56
DT - EnSVM [15]	99.36	99.07	0.38
EnSVM [15]	97.88	96.93	1.29
DFE - ALSTM	99.52	99.31	0.3

The proposed DFE-ALSTM model is evaluated on dataset of NSL-KDD and compared with existing models in the intrusion detection, as show in Table 3. The DFE-ALSTM has the higher performance compared to other models in the NSL-

KDD dataset in terms of FAR, DR and accuracy. The proposed DFE-ALSTM model has the advantages of effective feature representation using Dirichlet Feature Embedding method. The proposed DFE-ALSTM model has the advantage of optimal parameter settings based on the enhanced ASGD method. The proposed DFE-ALSTM model has the higher DR of 99.31 % and existing NB-SVM model has 99.24 % DR.

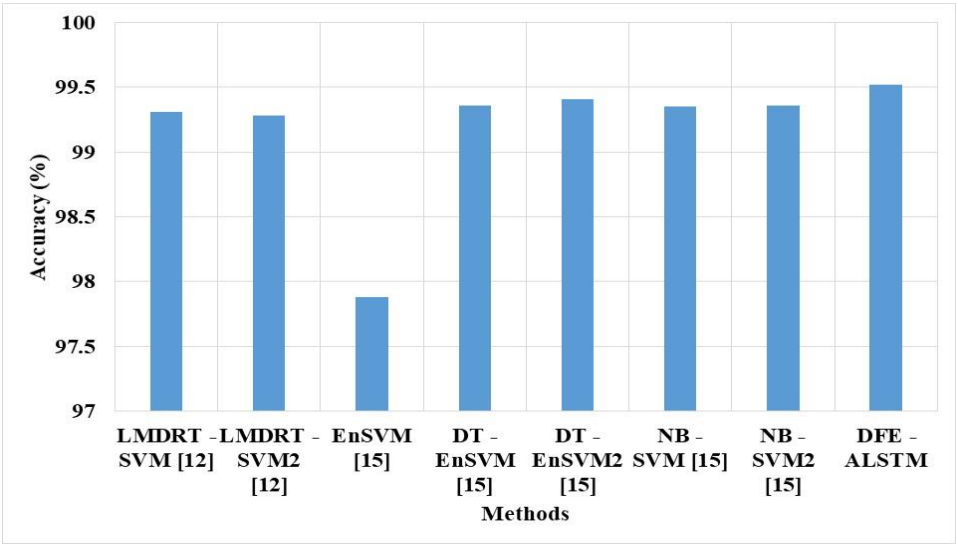

Figure 4: Accuracy of the proposed DFE-ALSTM and existing models in intrusion detection on NSL-KDD dataset.

The DFE-ALSTM model accuracy is compared with existing technique on dataset of NSL-KDD, as shown in Figure 4. The DFE-ALSTM method has advantage of effective feature representation and optimal parameter settings for the LSTM. The proposed DFE-ALSTM model has the accuracy of 99.52 % and existing NB-SVM model has 99.35 % accuracy.

Table 4: Performance analysis on Kyoto 2006+ dataset.

Methods	Accuracy (%)	DR (%)	FAR (%)
LMDRT - SVM [12]	98.33	99.85	3.25
LMDRT - SVM2 [12]	98.47	99.85	2.96
EnSVM [15]	97.83	99.54	3.96
DT - EnSVM [15]	98.34	99.82	3.21
DT - EnSVM2 [15]	98.48	99.81	2.92
Single - SVM [15]	97.51	99.6	4.67
NB - SVM [15]	98.58	99.73	2.62
NB - SVM2 [15]	98.55	99.84	2.79
DFE - ALSTM	98.76	99.87	2.6

The performance of the proposed DFE-ALSTM model is evaluated and compared with existing models, as shown in Table 4. The DFE-ALSTM has the advantage of effective feature representation using Dirichlet Feature Embedding. The enhanced ASGD method in the proposed DFE-ALSTM model provides the optimal parameter settings for the LSTM. The proposed DFE-ALSTM model has the accuracy of 98.76 % and existing NB-SVM model has 99.58 % accuracy.

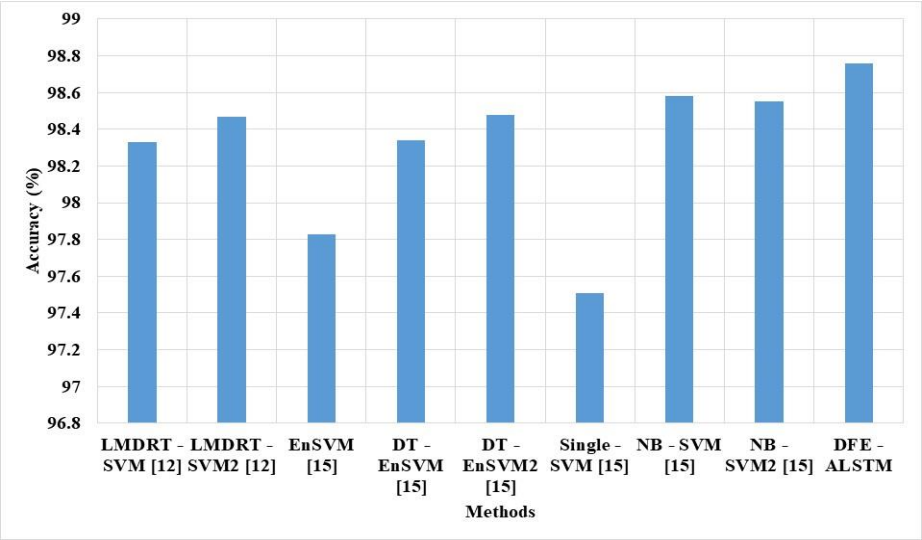

Fig. 5: Accuracy of the proposed DFE-ALSTM model and existing model in intrusion detection on Kyoto 2006+ dataset.

The accuracy of the proposed DFE-ALSTM model is evaluated and compared with existing models in Figure 5. The DFE-ALSTM model has the advantage of optimal parameter settings based on the enhanced ASGD method. The proposed DFE-ALSTM model has the effective feature representation based on the Dirichlet Feature Embedding. The proposed DFE-ALSTM model has the accuracy of 98.76 % and existing NB-SVM model has 98.58 % on Kyoto 2006+ dataset.

5. Conclusion

Intrusion Detection System is required in network to monitor the network and detect the intruder in the network to protect the user data. Data imbalance, overfitting, and high FAR are major limitations in existing methods. The DFE-ALSTM model is applied to increases efficiency of intrusion detection system. The DFE method provides the effective representation of the multi-variate of intrusion data to increases the performance. The ALSTM model is based on the enhanced ASGD method for optimal selection of the parameters for the model. The three datasets such as UNSW-NB15, NSL-KDD and Kyoto 2006+ were applied to examine the efficiency of DFE-

ALSTM model. The future direction of this model involves in applying the feature selection model to increases model detection efficiency.

References

Ahsan, M., Mashuri, M., Lee, M.H., Kuswanto, H., & Prastyo, D. D. (2020). Robust adaptive multivariate Hotelling's T2 control chart based on kernel density estimation for intrusion detection system. *Expert Systems with Applications*, 145, 113105.

Alazzam, H., Sharieh, A., & Sabri, K. E. (2020). A feature selection algorithm for intrusion detection system based on pigeon inspired optimizer. *Expert systems with applications*, 148, 113249.

Ali, M. H., Al Mohammed, B. A. D., Ismail, A., & Zolkipli, M. F., (2018). A new intrusion detection system based on fast learning network and particle swarm optimization. *IEEE Access*, 6, 20255-20261.

Almiani, M., AbuGhazleh, A., Al-Rahayfeh, A., Atiewi, S., & Razaque, A. (2020). Deep recurrent neural network for IoT intrusion detection system. *Simulation Modelling Practice and Theory*, 101, 102031.

Chimmula, V. K. R. & Zhang, L. (2020). Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons & Fractals*, 135, 109864.

Eskandari, M., Janjua, Z. H., Vecchio, M., & Antonelli, F. (2020). Passban IDS: An intelligent anomaly-based intrusion detection system for IoT edge devices. *IEEE Internet of Things Journal*, 7(8), 6882-6897.

Gu, J. & Lu, S. (2021). An effective intrusion detection approach using SVM with naïve Bayes feature embedding. *Computers & Security*, 103, 102158.

Guo, C., Ping, Y., Liu, N., & Luo, S. S. (2016). A two-level hybrid approach for intrusion detection. *Neurocomputing*, 214, 391-400.

Haider, W., Hu, J., Slay, J., Turnbull, B. P., & Xie, Y., (2017). Generating realistic intrusion detection system dataset based on fuzzy qualitative modeling. *Journal of Network and Computer Applications*, 87, 185-192.

Jin, D., Lu, Y., Qin, J., Cheng, Z., & Mao, Z., (2020). SwiftIDS: Real-time intrusion detection system based on LightGBM and parallel intrusion detection mechanism. *Computers & Security*, 97, 101984.

Kasongo, S. M. & Sun, Y. (2020). A deep learning method with wrapper based feature extraction for wireless intrusion detection system. *Computers & Security*, 92, 101752.

Kim, T. Y. & Cho, S. B. (2019). Predicting residential energy consumption using CNN-LSTM neural networks. *Energy*, 182, 72-81.

Kingma, D. P. & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv*:1412.6980.

Lv, L., Wang, W., Zhang, Z., & Liu, X., (2020). A novel intrusion detection system based on an optimal hybrid kernel extreme learning machine. Knowledge-based systems, 195, 105648.

Manzoor, I. & Kumar, N., (2017). A feature reduced intrusion detection system using ANN classifier. *Expert Systems with Applications*, 88, 249-257.

Merity, S., Keskar, N. S., & Socher, R. (2017). Regularizing and optimizing LSTM language models. *arXiv preprint arXiv*:1708.02182.

Rahman, M.A., Asyhari, A.T., Leong, L.S., Satrya, G.B., Tao, M.H., & Zolkipli, M. F. (2020). Scalable machine learning-based intrusion detection system for iot-enabled smart cities. *Sustainable Cities and Society*, 61, 102324.

Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena*, 404, 132306.

Wang, H., Gu, J., & Wang, S. (2017). An effective intrusion detection framework based on SVM with feature augmentation. *Knowledge-Based Systems*, 136, 130-139.

Zhang, J., Ling, Y., Fu, X., Yang, X., Xiong, G., 7 Zhang, R. (2020). Model of the intrusion detection system based on the integration of spatial-temporal features. *Computers & Security*, 89, 101681.