ISSN: 2182-2069 / E-ISSN: 2182-2077

Defence Mechanism against Advanced Persistent Threat Attack Using Significant Features based Deep Learning Model

U. Sakthivelu[1], and Dr.C.N.S. Vinoth Kumar[2*]

[1]Research Scholar, Department of Networking and Communications, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chennai, India.
su4343@srmist.edu.in, https://orcid.org/0009-0004-4971-5338

[2*]Associate Professor, Department of Networking and Communications, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chennai, India.
vinothks1@srmist.edu.in, https://orcid.org/0000-0001-7622-2417

Received: July 11, 2024; Revised: August 18, 2024; Accepted: September 23, 2024; Published: November 30, 2024

Abstract

The advanced persistent threat (APT) attack is the most frequent and seriously damaging one that can be launched against the target system. This type of malicious activity is deliberate and focused, with specific objectives in mind. The rise of this attack technique poses a significant challenge to the information security systems of various businesses, governments, and organizations. The techniques of utilizing deep learning (DL) or machine learning (ML) methods to evaluate abnormal behaviours and network traffic signs for preventing and detecting APT attacks are becoming popular in recent times. Though, the APT attack detection technique utilizes behaviour evaluation and analysis methods for confronting various problems owing to the absence of distinctive data of attack drives. So, the experimental outcomes display better recognition, it does not bring higher effectiveness in practice. Therefore, this study develops a novel Defence against Advanced Persistent Threat Attack Detection Feature Selection with Deep Learning (DAPTAD-FSDL) Method. The key objective of the DAPTAD-FSDL approach lies in the robust defence against the detection of APT attacks. In the presented DAPTAD-FSDL model, Z-score normalization is performed to change the raw data into a compatible format. Then, the presented DAPTAD-FSDL technique accomplishes the feature selection process using the grey wolf optimizer (GWO) algorithm. Besides, the attention with CNN-BiLSTM model is employed for the defence against the detection of APT attacks. Finally, a multi-objective improved golden jackal optimization algorithm (MIGJOA) is exploited for the optimal hyperparameter adjustment of the A-CNN-BiLSTM model. The experimental analysis of the DAPTAD-FSDL algorithm is tested employing the APT NSL-KDD dataset. The simulation outcomes described the supremacy of the DAPTAD-FSDL technique under different measures.

Keywords: Advanced Persistent Threat, Grey Wolf Optimization, Hyperparameter Tuning, Deep Learning, Feature Selection.

Journal of Internet Services and Information Security (JISIS), volume: 14, number: 4 (November), pp. 263-277.
DOI: *10.58346/JISIS.2024.I4.016*

*Corresponding author: Associate Professor, Department of Networking and Communications, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chennai, India.

1 Introduction

Cyber defence is an important issue in the digital economy, and it present at the root, and a data synthesis issue (Singh et al., 2019). Organizations establish many sensors to support their cyber defenders by synchronized data streams broadcasting a larger and wide-ranging number of examinations about source usage, application logs, network connections, host and user behavior, threat intellect, etc (Chu et al., 2019). Defenders preserve the attention of current events, and of malicious actions especially, by creating a sense of this complex dataset. Classified among data synthesis tasks is to identify abnormalities among data streams, for example, greater event sequence, under the presumption that incipient events would include actions or behaviors detected as most exceptional (Ghafir et al., 2019). With the fast growth of information technology, the information volume in current society endures to grow, and network threats become more and more dangerous. Network attacks are developing in a unique direction, with Advanced Persistent Threat (APT) attacks slowly becoming the main techniques of cyberattacks (Alshamrani et al., 2019).

APT refers to persistent, covert, and effectual attack events directed by an organization against particular targets (Zhao et al., 2015). These attacks normally have political or commercial inspirations and contain constant monitoring of particular countries or organizations, eventually resulting in targeted disruptions or information theft, posing an important threat to nations' and organizations' information schemes and data security (Ismail, 2024). Hence, APT detection attacks became an investigation hotspot in the network security field. APT attacks show long duration, strong concealment, significant harm, and high specificity (Cho & Nam, 2019). Conventional recognition approaches like malicious code detection, log anomaly detection, and malicious traffic detection, frequently concentrate on detailed stages of APT attacks and are unable to identify unidentified attacks (Lu et al., 2019). They also fight to remove contextual correlations inside APT attacks, resulting in a higher false positive rate. Nevertheless, APT threat intellect regularly facilitates the description of the attack procedure, but the weaknesses, vulnerabilities, and attacking models used by APT organizations make systems more vulnerable to attacks (Friedberg et al., 2015). On the other hand, the analyses examined many difficulties and challenges that made the recognition of APT attacks not very effective with the scarcity of unrestricted data on the APT threat, the data inequality, applying normal rules of coding, and so on (Guo et al., 2016). To reduce this problem, new research and suggestions frequently use detection processes depending on experimenting datasets, which are made and advanced at their own (Bodström & Hämäläinen, 2018). Consequently, in experimentation and research, it can take the best outcomes, but after being used to the authentic monitor methods, it has failed to provide as predicted.

This study develops a novel Defence against Advanced Persistent Threat Attack Detection based Feature Selection with Deep Learning (DAPTAD-FSDL) Method. In the presented DAPTAD-FSDL model, Z-score normalization is performed to convert the raw data into a consistent form. Afterwards, the presented DAPTAD-FSDL technique accomplishes the feature selection process using the grey wolf optimizer (GWO) method. Besides, the attention with CNN-BiLSTM model are utilized for the defence against the detection of APT attacks. Finally, a Mult objective improved golden jackal optimization algorithm (MIGJOA) has been exploited for the optimal hyperparameter adjustment of the A-CNN-BiLSTM system. The experimental outcome analysis of the DAPTAD-FSDL method is tested employing the APT NSL-KDD dataset.

2 Literature Review

Amaru et al., (2024) present RAPID and new DL-based techniques for powerful APT detection and investigation, levering context aware alert tracing and anomaly recognition. By using self-supervised structure learning and constantly learned embedding, our technique efficiently adjusts the behaviour of the dynamic system. The usage of provenience findings either enhances the detection or enriches the alert abilities of our method. Sakthivelu & Vinoth Kumar, (2022) presented several learning-based methods to assess sensory measures and network traffic in present to recognize and find cyberattacks. To attain this, various learning-based methods have been proposed, like scalable DL and classification model and self-tuning for cyberattack spot detection and an ensemble DL-based cyberattack recognition technique for unstable APT datasets. Kumar et al., (2024) propose an effective and new APT attack detection method, specifically SFCA-DeepCNN. At this time, the detection of APT attack is made by the DeepCNN, in which the DeepCNN weight is upgraded by the presented SFCA. The SFCA can be formed by combining the SFOA and SCA. Furthermore, the augmentation of data has been executed to upsurge the data dimension by executing the oversampling, which evades the problem of overfitting. Also, the presented enhanced DL system identifies the precise APT detection result.

Mei et al., (2024) propose a DL based network forensics method for automatically tracking and identifying network attacks, offering a network forensics process complete review. In particular, the author extracted network traffic and utilized encryption to assure the security and data reliability (Kalinin et al., 2024; Hlushenkova et al., 2024). Afterward, feature filtering techniques are applied to maintain vital traceable information, and DL method parameters have been spontaneously enhanced by utilizing the hyperparameter optimization methods. Finally, an MLP DNN method is developed with a perceptive ability for identifying anomalous actions in the network. Wang & Fu, (2024) introduces an enhanced Transformer-based techniques for APT malware detection and attribution. Relating to recognition, dynamic APT malware behaviours were extracted, and a data filtering gate method was used to decrease the redundant feature noise in the original Transformer method. A contrastive learning restrained method can be utilized for self-training, optimization, and data filtering. For attribution, APT malware samples static features have been extracted, sequence data global features were recognized by utilizing the Transformer method, local features were created by utilizing an Incremental Dilated CNN, and features have been united by utilizing an attention mechanism. This technique surpasses the primary technique.

Nallapaneni et al., (2024) present APT malware as the recognition object and gather 6777 malware from 20 kinds of APT attack agents. The enhanced CNN method integrated with an attention mechanisms for completing the weight of static malware and automated extraction features that could discover and detect the APT malware. Simultaneously, the identification outcome for APT malware is superior to other present techniques, which recognize the citation of APT attack groups (Atapour et al., 2018). Sakthivelu & Vinoth Kumar, (2024) introduce DeepTaskAPT, a heterogeneous task-tree based DL technique to create a standard method depends upon the sequence of tasks by utilizing an LSTM-NN that could be utilized over various users to classify anomalous behavior. Instead of utilizing the method to sequential log entries directly, since many existing methods can achieve this, DeepTaskAPT utilizes a method for creating a task-tree method to create sequential log entries for the DL method.

3 Materials and Methods

In this research, we have developed a novel DAPTAD-FSDL method. The major intension of the DAPTAD-FSDL method exists in the robust defense against the detection of APT attacks. It involves four various methods specifically data normalization, GWO based feature selection, APT attack detection using A-CNN-BiLSTM, and parameter optimizer. Figure 1 illustrates the entire flow of the DAPTAD-FSDL method.

3.1. Z-score Normalization

Initially, the presented DAPTAD-FSDL model takes place Z-score normalization is performed to transform the raw data into a consistent format. Z-score normalization can be a statistical method, which converts data into a normal scale by deducting the mean and separating by the standard deviation (Geem et al., 2024). This technique is utilized in cybersecurity, especially in the protection against APT attacks. By normalizing the data of network traffic, anomalies signifying possible APT activities are identified more effectually. Z-score normalization improves the precision of intrusion detection systems (IDS) by emphasizing abnormalities from distinctive patterns, helping in the initial recognition of complicated threats. This technique supports the entire security position against APTs.

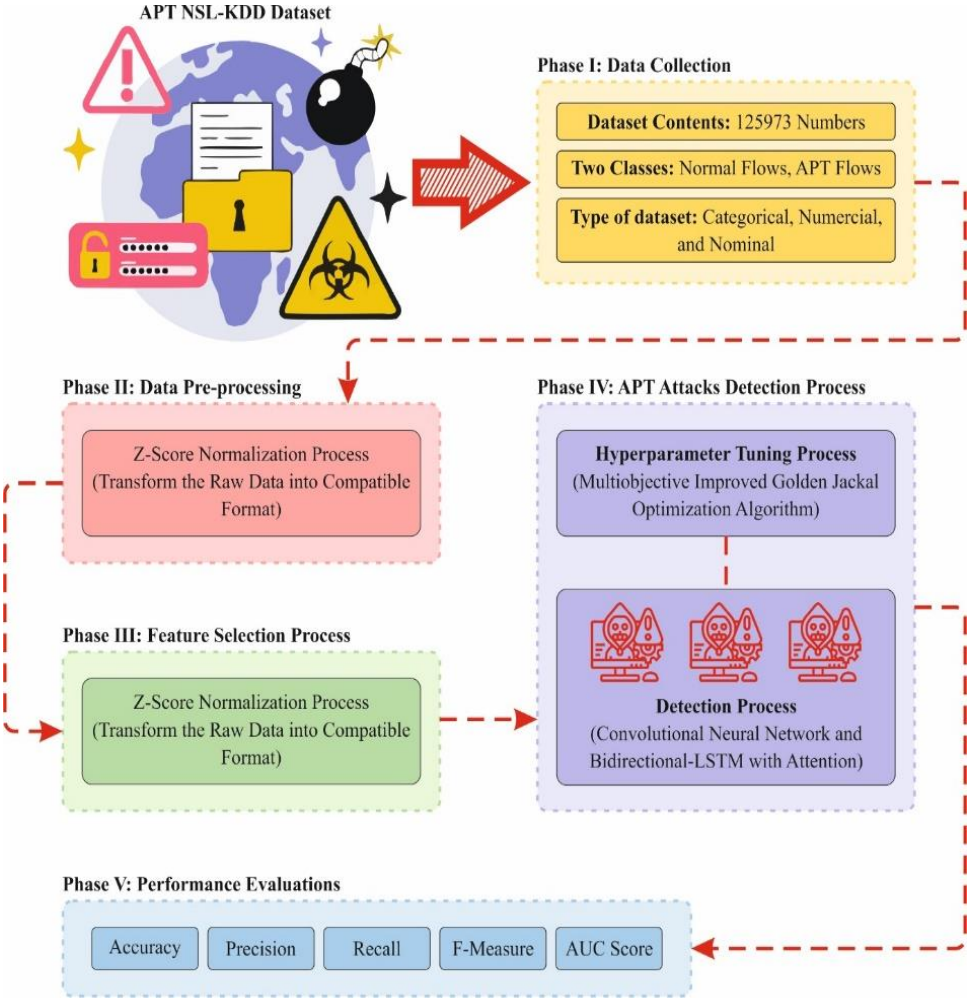

Figure 1: Overall Flow of DAPTAD-FSDL Method

3.2. Feature Selection using GWO

Then, the presented DAPTAD-FSDL technique accomplishes the feature selection (FS) method using the GWO method. GWO has been projected to depend upon the Grey wolf (Canis lupus), which relates to a family of Canidae (Tang & Krezger, 2024). They commonly used to survive in a pack. In an algorithm of GWO, the optimizer procedure was directed by the beta, alpha, and delta wolves, reflecting the behavior of hunting detected in real wolf packs. The 3 wolves take the guide while discovering the search space and defining the optimizer procedure route. The omega wolves, signifying sub-ordinate members, then the beta, delta, and alpha, adjusting their choices and movement with those leaders. By imitating the hierarchical form of wolf packs, the optimizer procedure is inclined by the main wolves, whereas the sub-ordinate wolves adjust and absorb their actions. This hierarchical structure permits the GWO model to attack a balance between exploitation and exploration, using the guidance of beta, delta, and alpha wolves to direct the optimizer near a superior solution.

To integrate the position of wolves, a mathematical representation is employed. In the population, the appropriate solution is labeled as alpha (α), signifying the maximum-ranking wolf. Likewise, the 2nd and 3rd finest solutions are mentioned as (β) and delta (δ), correspondingly. The residual candidate solution is reflected as omega (ω), which signifies lower-ranking wolves.

Encircling Prey

The numerical method of grey wolves enclosing victims throughout the search is signified below Eq. (1):

$$D = \left|C \cdot X_p(t) - X(t)\right| \quad (1)$$

$$X(t+1) = X_p(t) - A \cdot D \quad (2)$$

Whereas, X_p and X states the position of prey and grey wolf, respectively; t represents the present iteration; C and A represent the vectors of coefficient in Eq. (2).

$$A = 2a \cdot rand(\,) - a \quad (3)$$

$$C = 2 \cdot rand(\,) \quad (4)$$

The Eq. (3) – (4) attained an outcome by decreasing the a value. It is noticeable that the variation range of A is similarly diminished by a. While, A denotes a randomly produced integer in the interval of $[-a, a]$. While a is reduced from *2* to *0* during iteration.

Wolf Hunting

To arithmetically pretend the behavior of hunting, it is expected that the alpha, delta, and beta have superior awareness regarding the latent site of the target. So, the initial 3 finest solutions attained until now are kept and the further search agents were driven to upgrade their locations as per the location of the finest search agent that was exposed in Eqs. (5)- (7):

$$D = \begin{cases} D_\alpha |C_1 \cdot X_\alpha - X| \\ D_\beta \left|C_1 \cdot X_\beta - X\right| \\ D_\delta |C_1 \cdot X_\delta - X| \end{cases} \quad (5)$$

$$X = \begin{cases} X_1 = X_\alpha - A_1 \cdot (D_\alpha) \\ X_2 = X_\beta - A_2 \cdot (D_\beta) \\ X_3 = X_\delta - A_3 \cdot (D_\delta) \end{cases} \quad (6)$$

$$X(t+1) = \frac{X_1 + X_2 + X_3}{3} \quad (7)$$

The fitness function (FF) used in the GWO method is considered to have a balance between the value of chosen features in every solution (minimum) and the classification precision (maximum) obtained by employing these chosen features, Eq. (8) denotes the FF to assess solutions.

$$Fitness = \alpha \gamma_R(D) + \beta \frac{|R|}{|C|} \quad (8)$$

whereas $\gamma_R(D)$ denotes the classification rate of error of a given classification. $|R|$ represents selected subset cardinality and $|C|$ signifies the overall quantity of features in the dataset, α and β are two parameters equal to the importance of classification quality and subset length. $\in$ [1,0] and $\beta = 1 - \alpha$.

3.3. APT Attack Detection using A-CNN-BiLSTM

Besides, the A-CNN-BiLSTM model is utilized for the defense against the detection of APT attacks. CNNs are also broadly employed in time sequences. This network contains a pooling, an input, a convolutional, an output, and a fully connected (FC) layer (Muralidharan, 2024). Every layer contains its part, where the convolutional layer is capable of input data extracting features. The layer of pooling executes a dimension-reducing process by the data to constrict the extracted features in the convolutional layer, and the FC layer could weighting the compact features. Still, as the input data and model bound the convolution kernel dimension, CNN can't effectively remove features from longer temporal serious data. Nevertheless, new investigation demonstrates that standard neural networks can attain high-quality outcomes over particular combination structures.

According to the gradient difficulty by RNN, LSTM can deal with the long-term dependencies of time-based data. LSTM can be separated into 3 gates, like the output gate, input gate, and oblivion gate for controlling the states and adding the process of upgrading the past data. The oblivion gate observes $ht\text{-}1$ and Xt over a sigmoid unit to output of zero to one vector (keep is 1, discard 0). The gate of input required to select the efficient data and get the novel information via $tanh$. Then, the updated LSTM operation. These processes will utilize the old information $Ct\text{-}1$, remove a portion of the data over the gate of forgetting and later add the candidate data using the inputs to transform into novel data Ct. Computing other LSTM layers above the LSTM to obtain reversal data processing decreases the limitations of a solitary LSTM and accomplishes bi-directional information understanding.

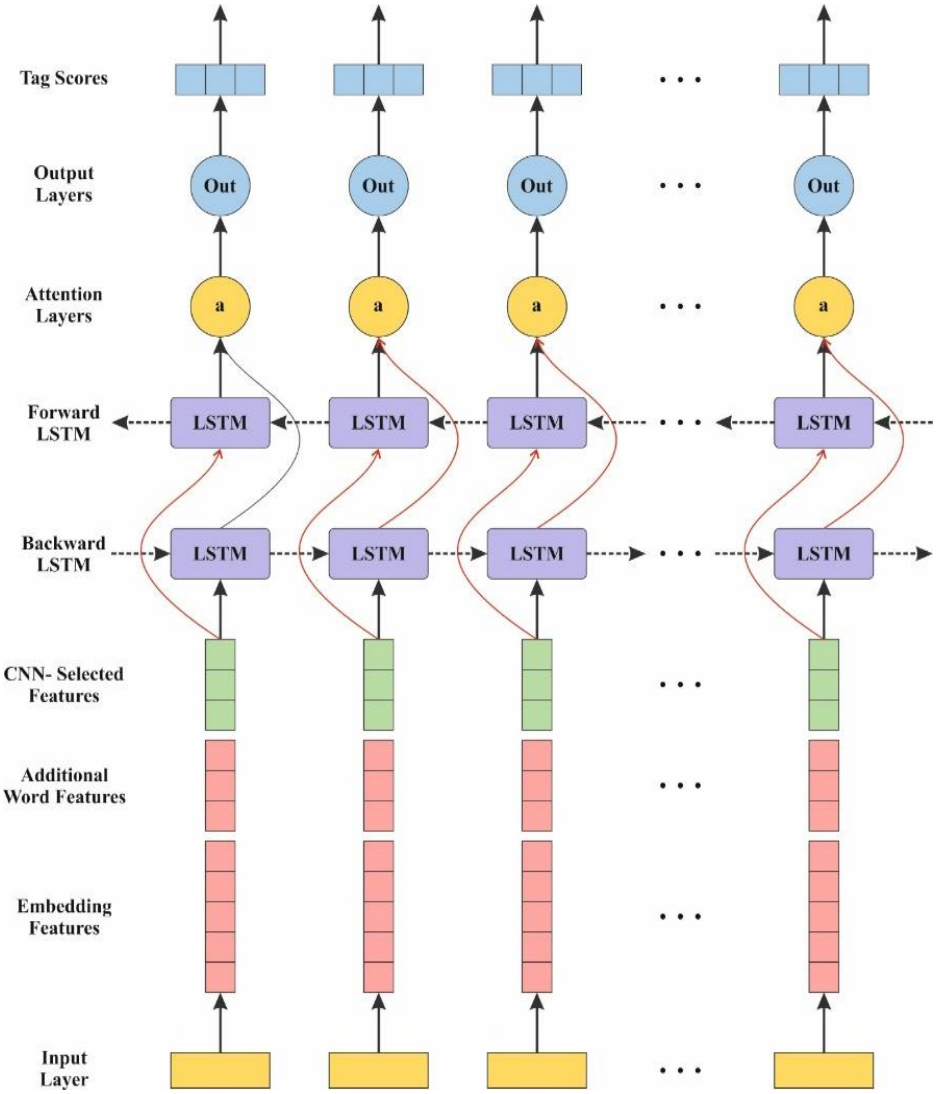

Figure 2: Structure of A-CNN-BiLSTM

It has reached excellent results in ML fields namely prediction, detection, and classification. The SE-attention mechanism is capable of handling the features of a time-based sequence, which the recent technique extracts and ignores the prominent features based on them. Primarily, the FC layer squeezes the channels and decreases the computation amount. Next, over the RELU non-linear activation layer, the size of the channel can be restored by the FC layer. This Sigmoid function stimulates the channel weights. This Scale performs the last weighting process for every channel. Figure 2 represents the infrastructure of A-CNN-BiLSTM.

3.4. Parameter Tuning Process

In conclusion, the MIGJOA is deployed for the optimal hyperparameter adjustment of the A-CNN-BiLSTM system. The present GJOA is dependent on hunting behavior and search capability but its challenged issues are linked to the amount of adverse features such as poor conjunction, stuck in the local solution to reduce the exploration capability (Sakthivelu & Kumar, 2023).

In MIGJOA, the iterative process is dependent upon exploitation and exploration besides opposition based learning. The exploration stage is dependent upon the jackal search capability the exploitation phase is reliant on the behavior of hunting. The procedure contained in phases of exploitation and exploration besides opposition-based learning was defined below.

Exploration phase: Jackals recognize and follow the victim but they run away from being captured, so they need sufficient time and hunt for their novel victim. In the hunting method, the male jackals execute the main role, and female jackals will follow them. The mathematical computation of this action is signified in Eqs. (9) and (10).

$$Z_1(t) = Z_M(t) - E|Z_M(t) - rl \times prey(t)| \quad (9)$$

$$Z_2(t) = Z_M(t) - E|Z_F(t) - rl \times prey\ (t)| \quad (10)$$

Whereas, r represents iteration in the present state, and the location vector is signified as $prey\ (t)$. A female and male jackal location is represented as $Z_F(t)$ and $Z_M(t)$ correspondingly. E is the energy shown by the victim is signified in Eq. (11):

$$E = E_1 \times E_0 \quad (11)$$

Where the decreasing phase and initial energy are represented as E_1 and E_0, which is exemplified in Eq. (12) and (13) as below:

$$E_0 = 2 \times r - 1 \quad (12)$$

$$E_1 = c_1 \times \left(1 - \left(\frac{\mathrm{t}}{\mathrm{T}}\right)\right) \quad (13)$$

Where r signifies a randomly produced value that lies amongst the interval of [0, 1] and c_1 denotes a value of constant. The maximum iteration and the present state iteration are signified as t and T respectively.

Exploitation phase: In this phase, the prey that can go away from the jackal will be reduced. The set of jackals surround the victim and their behavior of hunting with female and male jackals was signified in Eq. (14) and (15) as below:

$$Z_1(t) = Z_M(t) - E.|rl.Z_M(t) - prey(t)| \quad (14)$$

$$Z_2(t) = Z_F(t) - E.|rl.Z_F(t) - prey(t)| \quad (15)$$

Here, factor rl is answerable for making the behaviour of random in the exploitation phase that highlights search and evades local goals.

In this part, a brief explanation of IGJOA the mixture of GJOA with OBL has been clarified for enhancing the search capability and improve the precision in perceiving the optimum solution. It incorporates the value of opposite that protects searching region depending upon possible solutions. IGJOA depends upon dual stages like initialization and upgrading the novel generation.

Initialization: In this stage, the novel population $Z_0 = \{z_{i1}, z_{i2}, \ldots tz_{ij}, \ldots z_{iD}\}$ is formed randomly in a search area, whereas $= 1,2,3, \ldots, NP; j = 1,2, \ldots D$, the size is signified as D and the population size is denoted as NP. The opposite-based learning approach is reflected and the opposite jackal was formed depending upon Eq. (16):

$$\overline{Z} = L + U - Z \tag{16}$$

Whereas, the opposite and real location vector is signified as$\overline{Z}$ and Z. Lower and Upper bound values are signified L and U. Next, the actual and original location is combined. Then, pick the finest solution from the size of population NP in the interval of $\{Z_i, \overline{Z_i}\}$.

Update stage: The search for an optimum result utilizing female and male jackals was recognized and these dual jackals were measured as the appropriate one. The GJO convergence passes rapidly but it effortlessly drops into the local goals owing to reduced search capability. Therefore, opposition-based learning is used to generate novel jackals with a likelihood of P_γ and the randomly generated valuation between *0* and 1 was made. If it is < P_r, then the opposition-based learning method has been employed for acquiring novel jackals depending upon the dominant population. The MIGJOA method advances an FF to obtain enhanced classification performance. It identifies a positive integer to signify the greater performances of the candidate outcomes. In this research, the reduction of the classification error value can be examined as the FF, as specified in Eq. (17).

$$fitness(x_i) = ClassifierErrorRate(x_i)$$
$$= \frac{No.of\ misclassified\ samples}{Total\ no.of\ samples} * 100 \tag{17}$$

4 Performance Validation

The experimental outcome study of the DAPTAD-FSDL methodology can be tested using the APT NSL-KDD dataset. The dataset comprises 125973 samples under 2 classes defined in Table 1. The number of features is 41 and the selected features are 22.

Table 1: Details on Dataset

Class	**values**
Normal	67343
Attack (Dos, R2L, U2R, Probe)	58630
Sum	**125973**

Figure 3 represents the classification outcomes of the DAPTAD-FSDL method on the test database. Figures 3a-3b illustrates the confusion matrices with precise identification and classification of all 2 class labels under 70%TRAP and 30%TESP. Figures 3c defines the PR analysis, indicating maximum performance across all classes. At last, Figure 3d represents the ROC analysis, representing efficient results with high ROC values for 2 classes.

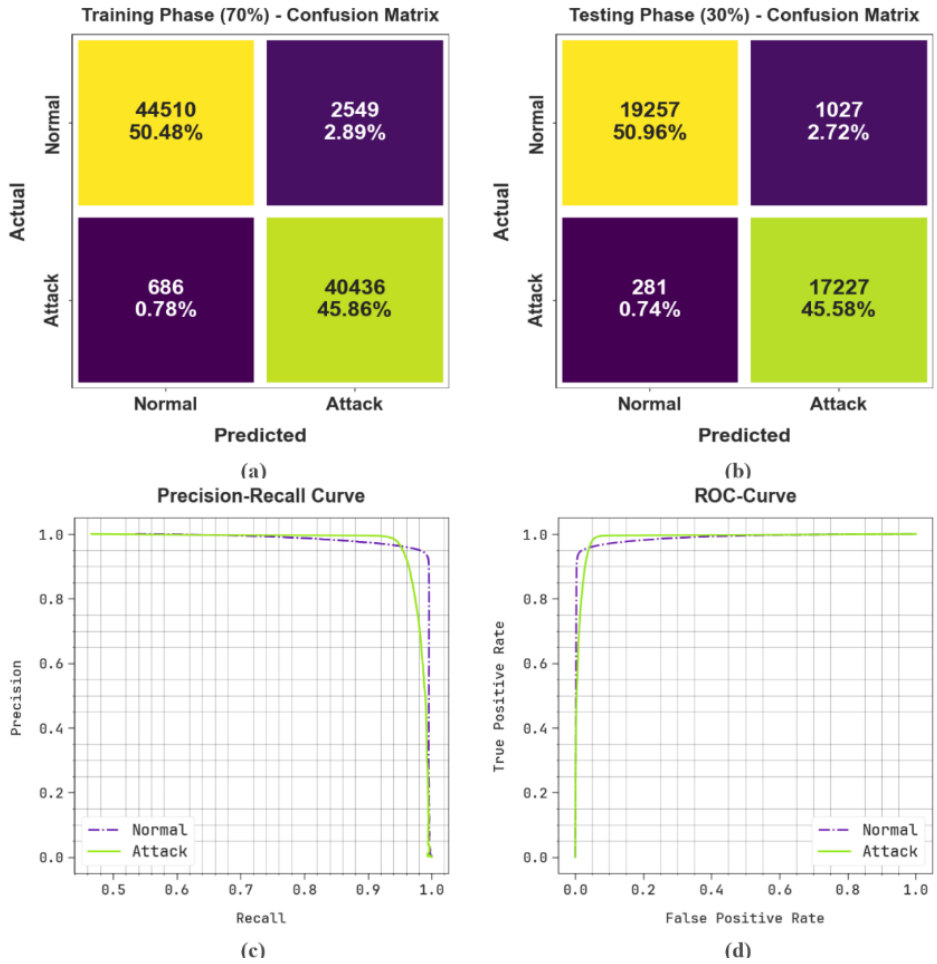

Figure 3: Classifier Outcome of (a-b) Confusion Matrices and (c-d) Curves of PR and ROC

In Table 2 and Figure 4, the APT attack detection outcome of the DAPTAD-FSDL process is clearly depicted under 70%TRAP and 30%TESP. The results signified that the DAPTAD-FSDL technique properly identified the normal and attack samples. With 70%TRAP, the DAPTAD-FSDL technique gains average $accu_y$ of 96.46%, $prec_n$ of 96.28%, $reca_l$ of 96.46%, $F_{measure}$ of 96.32%, and AUC_{score} of 96.46%. In addition, with 30%TESP, the DAPTAD-FSDL system attains average $accu_y$ of 96.67%, $prec_n$ of 96.47%, $reca_l$ of 96.67%, $F_{measure}$ of 96.53%, and AUC_{score} of 96.67%.

Table 2: APT Attack Detection of DAPTAD-FSDL Technique Under 70%TRAP and 30%TESP

Class	$Accu_y$	$Prec_n$	$Reca_l$	$F_{Measure}$	AUC_{Score}
TRAP (70%)					
Normal	94.58	98.48	94.58	96.49	96.46
Attack	98.33	94.07	98.33	96.15	96.46
Average	**96.46**	**96.28**	**96.46**	**96.32**	**96.46**
TESP (30%)					
Normal	94.94	98.56	94.94	96.72	96.67
Attack	98.40	94.37	98.40	96.34	96.67
Average	**96.67**	**96.47**	**96.67**	**96.53**	**96.67**

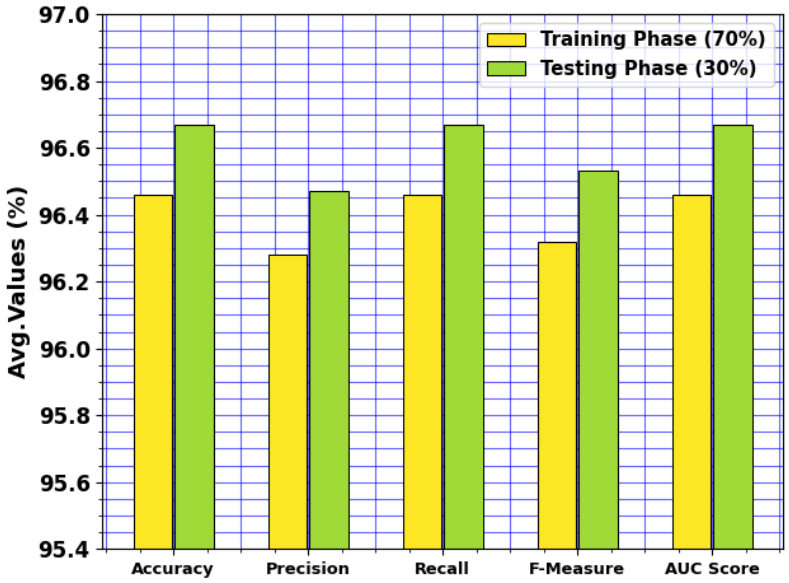

Figure 4: Average of DAPTAD-FSDL Technique Under 70%TRAP and 30%TESP

In Figure 5, the training (TRA) and validation (VLA) accuracy outcomes of the DAPTAD-FSDL system is displayed. The accuracy values are computed over an interval of 0-25 epochs. The figure emphasized that the TRA and VLA accuracy values show a rising trend that notified the capability of the DAPTAD-FSDL method with improved performance over several iterations. Additionally, the TRA and VLA accuracy remains closer over the epochs, which indicates low minimal overfitting and demonstrations the improved performance of the DAPTAD-FSDL model, guaranteeing consistent prediction on unseen samples.

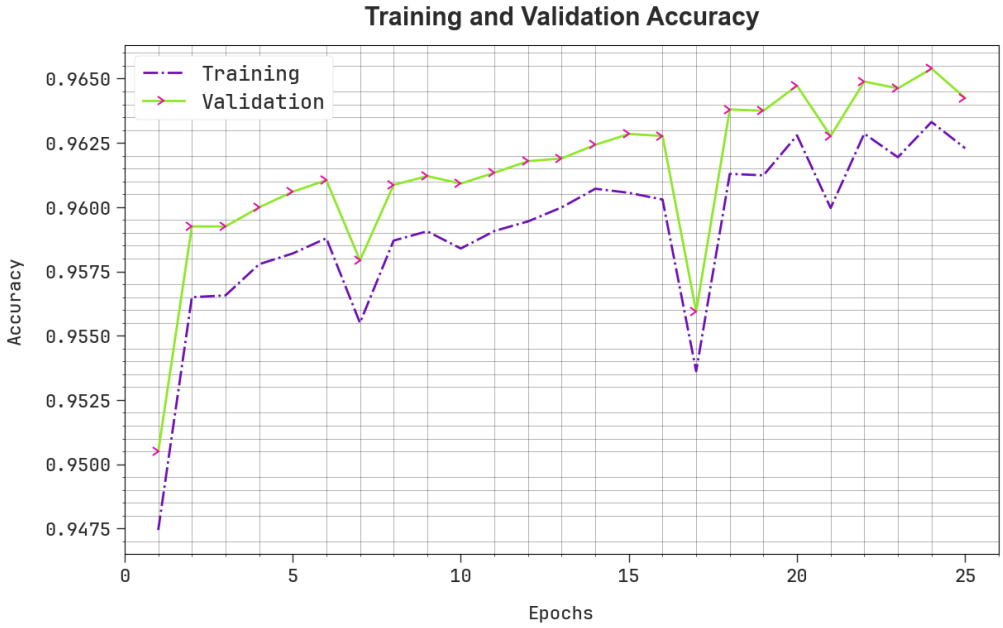

Figure 5: $Accu_y$ Curve of the DAPTAD-FSDL Model

Figure 6: Loss Curve of the DAPTAD-FSDL Technique

In Figure 6, the TRA and VLA loss graph of the DAPTAD-FSDL model is displayed. The loss values are computed over an interval of 0-25 epochs. It is denoted that the TRA and VLA accuracy values show a reducing trend, which notified the ability of the DAPTAD-FSDL method in balancing a trade-off between data fitting and generalization. The continual reduction in loss values additionally guarantees the improved performance of the DAPTAD-FSDL process and tune the prediction results over time.

In Table 3 and Figure 7, an extensive comparison study of the DAPTAD-FSDL system is clearly displayed (Joloudari et al., 2020; Nadim et al., 2024). The outcomes show that the ANN process has shown ineffective performance with $accu_y$ of 81.20%, $prec_n$ of 84.13%, $reca_l$ of 85.61%, and $F_{measure}$ of 83.18%. Along with that, the J48-AdaBoost technique has displayed slightly boosted results with $accu_y$ of 90.56%, $prec_n$ of 95.19%, $reca_l$ of 95.42%, and $F_{measure}$ of 90.41%. Meanwhile, the DBN-SVM, LSTM, NB, HMM, and MLP approaches have exhibited moderately closer results. However, the DAPTAD-FSDL system outperforms the other models with increased $accu_y$ of 96.67%, $prec_n$ of 96.47%, $reca_l$ of 96.67%, and $F_{measure}$ of 96.53%.

Table 3: Comparative Analysis of DAPTAD-FSDL Method with Recent Approaches

Techniques	$\boldsymbol{Accu_y}$	$\boldsymbol{Prec_n}$	$\boldsymbol{Reca_l}$	$\boldsymbol{F_{Measure}}$
DBN-SVM Model	92.84	91.67	93.49	89.81
Naïve Bayes	93.87	93.95	95.33	90.87
ANN Algorithm	81.20	84.13	85.61	83.18
J48-AdaBoost	90.56	95.19	95.42	90.41
Hidden Markov Model	95.14	90.11	91.94	93.47
MLP Algorithm	96.00	94.23	93.00	90.75
LSTM Classifier	93.13	93.81	89.98	94.14
DAPTAD-FSDL	96.67	96.47	96.67	96.53

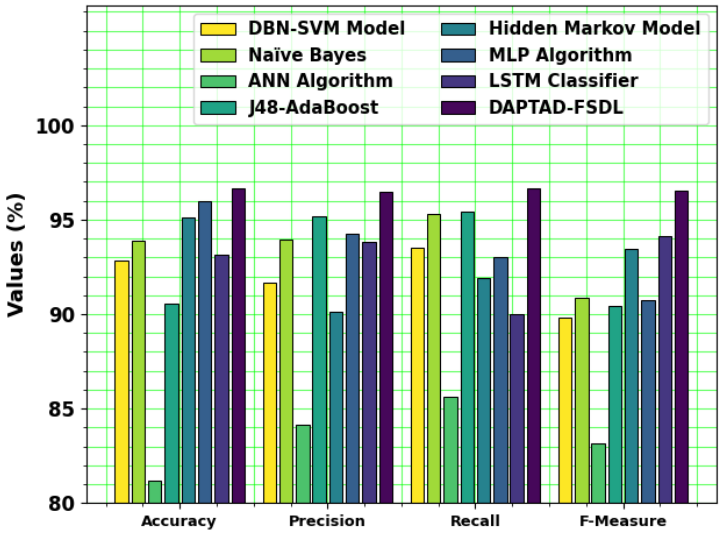

Figure 7: Comparative Analysis of DAPTAD-FSDL Approach with other Approaches

5 Conclusion

In this study, we have developed a DAPTAD-FSDL approach. The major intension of the DAPTAD-FSDL methodology lies in the robust defense against the detection of APT attacks. Initially, the presented DAPTAD-FSDL model takes place Z-score normalization is performed to change the raw data into a compatible format. Then, the presented DAPTAD-FSDL technique accomplishes the feature selection process using the GWO algorithm. Besides, the A-CNN-BiLSTM model is utilized for the defense against the detection of APT attacks. Eventually, the MIGJOA has been employed for the better hyperparameter adjustment of the A-CNN-BiLSTM algorithm. The experimental outcome analysis of the DAPTAD-FSDL approach is tested using the APT NSL-KDD database. The simulation outcomes defined that the higher of the DAPTAD-FSDL technique under different measures.

References

[1] Alshamrani, A., Myneni, S., Chowdhary, A., & Huang, D. (2019). A survey on advanced persistent threats: Techniques, solutions, challenges, and research opportunities. *IEEE Communications Surveys & Tutorials*, *21*(2), 1851-1877.

[2] Amaru, Y., Wudali, P., Elovici, Y., & Shabtai, A. (2024). RAPID: Robust APT Detection and Investigation Using Context-Aware Deep Learning. https://doi.org/10.48550/arXiv.2406.05362

[3] Atapour, C., Agrafiotis, I., & Creese, S. (2018). Modeling Advanced Persistent Threats to enhance anomaly detection techniques. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, *9*(4), 71-102.

[4] Bodström, T., & Hämäläinen, T. (2018). A novel method for detecting APT attacks by using OODA loop and black swan theory. In *Computational Data and Social Networks: 7th International Conference, CSoNet 2018, Shanghai, China, December 18–20, 2018, Proceedings 7* (pp. 498-509). Springer International Publishing.

[5] Cho, D. X., & Nam, H. H. (2019). A method of monitoring and detecting APT attacks based on unknown domains. *Procedia Computer Science*, *150*, 316-323.

[6] Chu, W. L., Lin, C. J., & Chang, K. N. (2019). Detection and classification of advanced persistent threats and attacks using the support vector machine. *Applied Sciences*, *9*(21), 4579. https://doi.org/10.3390/app9214579

[7] Friedberg, I., Skopik, F., Settanni, G., & Fiedler, R. (2015). Combating advanced persistent threats: From network event correlation to incident detection. *Computers & Security*, *48*, 35-57.

[8] Geem, D., Hercules, D., Pelia, R. S., Venkateswaran, S., Griffiths, A., Noe, J. D., ... & Kugathasan, S. (2024). Progression of Pediatric Crohn's Disease Is Associated with Anti–Tumor Necrosis Factor Timing and Body Mass Index Z-Score Normalization. *Clinical Gastroenterology and Hepatology*, *22*(2), 368-376.

[9] Ghafir, I., Kyriakopoulos, K. G., Lambotharan, S., Aparicio-Navarro, F. J., Assadhan, B., Binsalleeh, H., & Diab, D. M. (2019). Hidden Markov models and alert correlations for the prediction of advanced persistent threats. *IEEE Access*, *7*, 99508-99520.

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

[11] Hlushenkova, A., Kalinin, O., Navrozova, Y., Navolokina, A., Shcherbyna, V., & Doroshenko, T. (2024). Management of Strategies for Shaping the Innovative and Investment Potential of Enterprises as a Factor Ensuring Their Economic Security. *Indian Journal of Information Sources and Services*, *14*(3), 16–22. https://doi.org/10.51983/ijiss-2024.14.3.03

[12] Ismail, W. S. (2024). Threat Detection and Response Using AI and NLP in Cybersecurity. *Journal of Internet Services and Information Security*, *14*(1), 195-205.

[13] Joloudari, J. H., Haderbadi, M., Mashmool, A., GhasemiGol, M., Band, S. S., & Mosavi, A. (2020). Early detection of the advanced persistent threat attack using performance analysis of deep learning. *IEEE Access*, *8*, 186125-186137.

[14] Kalinin, O., Gonchar, V., Abliazova, N., Filipishyna, L., Onofriichuk, O., & Maltsev, M. (2024). Enhancing Economic Security through Digital Transformation in Investment Processes: Theoretical Perspectives and Methodological Approaches Integrating Environmental Sustainability. *Natural and Engineering Sciences*, *9*(1), 26-45.

[15] Kumar, A., Noliya, A., & Makani, R. (2024). Fuzzy inference based feature selection and optimized deep learning for Advanced Persistent Threat attack detection. *International Journal of Adaptive Control and Signal Processing*, *38*(2), 604-620.

[16] Lu, J., Chen, K., Zhuo, Z., & Zhang, X. (2019). A temporal correlation and traffic analysis approach for APT attacks detection. *Cluster computing*, *22*, 7347-7358.

[17] Mei, Y., Han, W., Li, S., Lin, K., Tian, Z., & Li, S. (2024). A Novel Network Forensic Framework for Advanced Persistent Threat Attack Attribution Through Deep Learning. *IEEE Transactions on Intelligent Transportation Systems*. https://doi.org/10.1109/TITS.2024.3360260

[18] Muralidharan, J. (2024). Advancements in 5G Technology: Challenges and Opportunities in Communication Networks. *Progress in Electronics and Communication Engineering*, *1*(1), 1-6.

[19] Nadim, I., Rajalakshmi, N. R., & Hammadeh, K. (2024). A Novel Machine Learning Model for Early Detection of Advanced Persistent Threats Utilizing Semi-Synthetic Network Traffic Data. *Journal of VLSI Circuits and Systems*, *6*(2), 31-39.

[20] Nallapaneni, S., Suneel, E., Dileep, J., Kumar, G. D., Shankar, B. S. S. M., & Sai, E. V. (2024). Implementation of Adaptive Motion Controlled Wheel Chair. *International Journal of communication and computer Technologies*, *12*(1), 45-50.

[21] Sakthivelu, U., & Kumar, C. V. (2023, December). An Adaptive Defensive Mechanism for DQN Storage Resources Allocation from Advanced Persistent Threats. In *2023 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)* (pp. 1-8). IEEE. https://doi.org/10.1109/ICSES60034.2023.10465545

[22] Sakthivelu, U., & Vinoth Kumar, C. N. S. (2022). An approach on cyber threat intelligence using recurrent neural network. In *ICT Infrastructure and Computing: Proceedings of ICT4SD 2022* (pp. 429-439). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-19-5331-6_44

[23] Sakthivelu, U., & Vinoth Kumar, C. N. S. (2024). A multi-step APT attack detection using hidden Markov models by molecular magnetic sensors. *Optical and Quantum Electronics*, *56*(3), 282. https://doi.org/10.1007/s11082-023-05905-3

[24] Singh, S., Sharma, P. K., Moon, S. Y., Moon, D., & Park, J. H. (2019). A comprehensive study on APT attacks and countermeasures for future networks and communications: challenges and solutions. *The Journal of Supercomputing*, *75*, 4543-4574.

[25] Tang, U., & Krezger, H. (2024). Design and validation of 6G Antenna for Mobile communication. *National Journal of Antennas and Propagation*, *6*(1), 6-12.

[26] Wang, N., & Fu, H. (2024, May). Research on deep-learning-based techniques for advanced persistent threat malware detection and attribution. In *Fourth International Conference on Sensors and Information Technology (ICSI 2024)* (Vol. 13107, pp. 596-603). SPIE.

[27] Zhao, G., Xu, K., Xu, L., & Wu, B. (2015). Detecting APT malware infections based on malicious DNS and traffic analysis. *IEEE access*, *3*, 1132-1142.

Authors Biography

U. Sakthivelu, received his B.E. degree in the discipline of Electronics and Communication Engineering from Annamalai University, India, M.E Degree in the discipline of Computer Science Engineering from Annamalai University, India and Pursuing PhD in the Department of Networking and Communications from SRM Institute of Science and Technology. His area of interest is Cyber Security, Wireless Sensor Networks, & Network Security. He is having two years of teaching experience. He is currently pursuing his PhD as a Full-Time Research Scholar in the Department of Networking and Communications in SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu, India.

Dr.C.N.S. Vinoth Kumar, received his B.E. degree in the discipline of Information Technology from Annamalai University, India, M.E Degree in the discipline of Computer Science Engineering from Annamalai University, India and Ph.D. degree in the discipline of Computer Science and engineering from Annamalai University. His area of interest is Cyber Security, Wireless Sensor Networks, Cryptography & Network Security, Artificial Intelligence, and Graphics. He is the Life member of computer society of India and International associations of engineers. He is having totally of 13 years of experience in both teaching and research. He has published over 50+ articles in reputed Journals and International conference proceedings. He has filed & published 6 national patents and one international patent was granted. Currently He is working as Associate Professor, Department of Networking and Communications in SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu, India.