Journal of Data Analysis and Information Processing, 2026, 14(2), 171-188
https://www.scirp.org/journal/jdaip
ISSN Online: 2327-7203
ISSN Print: 2327-7211

Machine-Deep Learning Approaches for Efficient Persistent Threat Identification: A Performance Analysis

Vetrivelan Tamilmani[1], Venkata Deepak Namburi[2], Aniruddha Arjun Singh Singh[3], Vaibhav Maniar[4], Rami Reddy Kothamaram[5], Dinesh Rajendran[6]

[1]SAP America, Houston, TX, USA
[2]Department of Computer Science, University of Central Missouri, Warrensburg, MO, USA
[3]ADP Inc., Alpharetta, GA, USA
[4]Oklahoma City University, Oklahoma City, OK, USA
[5]California University of Management and Science, Anaheim, CA, USA
[6]Coimbatore Institute of Technology, Coimbatore, TN, India
Email: vetrivelant@gmail.com, venkatadeepak.n@gmail.com, aniruddha.singh1@gmail.com, vaibhav.maniar@gmail.com, kalyanramreddy@gmail.com, rdinesh86@gmail.com

How to cite this paper: Tamilmani, V., Namburi, V.D., Singh, A.A.S., Maniar, V., Kothamaram, R.R. and Rajendran, D. (2026) Machine-Deep Learning Approaches for Efficient Persistent Threat Identification: A Performance Analysis. *Journal of Data Analysis and Information Processing*, **14**, 171-188.
https://doi.org/10.4236/jdaip.2026.142009

Received: November 19, 2025
Accepted: March 10, 2026
Published: March 13, 2026

Abstract

The advanced persistent threat (APT) is an ever-growing issue in cybersecurity, whose emergence and evolution have been accompanied by technological advancement. The study describes an approach to identifying Advanced Persistent Threats (APTs) based on the 42 (numerical and categorical) features present in the dataset OF UNSW NB15. Exploratory Data Analysis (EDA) was done to comprehend the correlation among features and the bit of class distribution and then preprocessing of data, such as cleaning and standardization, removal and balancing of outliers and classes with SMOTE. Two models, Decision Tree (DT) and Convolutional Neural Network (CNN), were used and assessed. The results of the experiments prove that CNN model is much superior to other baseline models, including ANN, SVM, and RF, because CNN has an F1-score of 98.92, a recall of 99.14, a precision of 98.71, and an accuracy of 98.85. Conversely, DT classifier was found to have an accuracy of 95.84 with low variation in values of the F1-score at 96.12, precision, and the recall. CNN also obtained AUC of 0.9990, which indicates that it is very strong in separating normal and attack traffic. Although CNN offers better functionality due to its capability to identify intricate and nuanced patterns, the model of the DT can be interpreted and implemented easily. On the whole, the results present the superiority of CNN-based deep learning (DL) in identifying advanced and evasive APTs, as well as the supporting power of interpretable models such as DT in cybersecurity defense processes.

Keywords

Threat Detection, Advanced Persistent Threat (APT), Intrusion Detection System (IDS), Machine Learning, Cybersecurity, Deep Learning, SMOTE

1. Introduction

People all across the globe are able to access and exchange technology resources through cyberspace. A resource could be anything from an image to a tweet to an audio clip, video file, or text stored in an electronic format. When talking about cybersecurity, it refers to the tools, personnel, and procedures put in place to prevent unauthorized access to computer networks and data. An act of attempting or actually committing to steal data. Cyber threats are those that compromise system or network integrity or cause harm to computer systems or networks [1]. The private and corporate sectors are now part of the target zone for threats that previously only targeted nation-states and their related entities. Not long ago, a new type of network attack called an APT evolved. Attacks on key infrastructure, intellectual property theft, and cyber-attacks that cost millions of dollars are becoming more and more of a problem [2] [3]. These dangers, which fall into the category of APTs, are a source of concern for every country and reputable organization. The sophistication of nation-sponsored APT attacks is an undeniable characteristic [4]. Even with the rise of APTs in the business world, organizations still face significant challenges [5]. An APT is a type of multi-stage attack that is carried out with more subtlety and is intended to accomplish a certain objective, typically espionage. In order to achieve their objective, APTs employ various steps, similar to conventional multistep attacks. But advanced persistent threats (APTs) vary because they typically use what are known as "zero-day exploits" [6] and sophisticated assault methods such as social engineering [7]. Companies and governments are currently facing the greatest threat from APTs.

Particularly in the realm of smart grids, cyber dangers have grown substantially during the past decade. The sophistication of cybercriminals has grown. There are too many highly competent cybercriminals for current security procedures to protect networks [8]. Intrusion Detection and Prevention Systems (IDPS) and other cutting-edge technologies are avoided by cybercriminals. Unfortunately, these technologies are incredibly bad at detecting the Advanced Persistent Threat (APT). Data mining using ML techniques can reveal various APT attack stages, and IDPS detection rates could be enhanced with the help of AI technologies [9] [10]. Cybersecurity professionals must weigh the potential hazards and benefits of AI applications. One branch of AI, machine learning, can sift through enormous datasets in search of relevant patterns and insights automatically [11]. If there is enough training data, IDSs that rely on machine learning may detect new and different types of attacks with relative ease, and the models used by these systems are very generalizable [12]. A subfield of ML known as "DL" is capable of remarkable

results. When it comes to handling large datasets, deep learning approaches outperform more conventional machine learning methods. To top it all off, deep learning approaches are practical, end-to-end, and capable of autonomously learning feature representations from raw data before producing results.

1.1. Motivation and Contribution

The rapid evolution of cyberspace has introduced unprecedented opportunities for information sharing, but it has also created an environment highly vulnerable to cyber threats, particularly APTs. Unlike traditional cyberattacks, APTs are stealthy, multi-stage, and often exploit zero-day vulnerabilities that traditional Intrusion Detection and Prevention Systems (IDPS) have a hard time picking up on. The increasing sophistication of cybercriminals, coupled with the expansion of targets from nation-states to corporate sectors and critical infrastructures such as smart grids, highlights the inadequacy of existing security mechanisms. AI and ML, especially DL approaches, provide a promising direction by enabling automated feature extraction, pattern recognition, and enhanced detection of novel attack variants. Motivated by these challenges, this study aims to explore and advance AI/ML-based solutions for strengthening cyber threat detection, thereby addressing the limitations of current systems and contributing to more resilient cybersecurity infrastructures. This research offers several key contributions as listed below:

- Utilized the dataset of UNSW_NB15 from Kaggle, which is a comprehensive and widely recognized benchmark for evaluating intrusion detection systems.
- Implemented systematic preprocessing steps, including data cleaning, standardization of numerical features using StandardScaler, removal of outliers, and application of SMOTE to address class imbalance, ensuring high-quality and balanced input data.
- Proposed and implemented two different approaches: Convolutional Neural Network (CNN) to exploit deep feature learning capabilities, and Decision Tree (DT) to ensure simplicity and interpretability.
- Conducted a comprehensive performance assessment using multiple evaluation metrics, including precision, accuracy, F1-score, recall, loss, and ROC, ensuring a well-rounded comparison of models.

1.2. Novelty and Justification of the Study

The novelty of the research lies in combining advanced preprocessing methods like outlier removal, feature standardization, and SMOTE-based balancing with the evaluation of both CNN and DT models on dataset of UNSW_NB15, which is a modern and complex benchmark for network intrusion detection. Unlike many prior works that focus on either deep learning or classical models in isolation, this study justifies its approach by aiming to capture the strengths of both paradigms, leveraging CNN's capability for hierarchical feature extraction and DT's interpretability for decision-making. By considering the problem from two angles, it can

learn more about how the models perform on actual cybersecurity data, which in turn can help us design better intrusion detection systems.

1.3. Structure of the Paper

The article is structured like this: Section II presents a review of related work on advanced persistent threat (APT) detection. Section III details the dataset, preprocessing techniques, and the implementation of the proposed model. The experimental results and a comparison to previous methods are detailed in Section IV. The study is concluded and future research directions are outlined in Section V.

2. Literature Review

The research is based on an extensive review and critical analysis of the literature on APT detection, which influenced the scope of the research and the general direction of the work.

Yan and Xiong (2020) proposed Web-APT-Detect (WAD), an unsupervised anomaly detector that builds a self-translation machine using an attention-based encoder-decoder. To better identify harmful patterns in HTTP requests, their attention techniques can enhance the quality of self-translation machines. Their technique achieves an F1-Score of 0.9844 in experiments on CSIC 2010 dataset, matching or exceeding that of the state-of-the-art supervised algorithm and the known unsupervised algorithm [13].

Joloudari *et al*. (2020) stated that NSL-KDD was input into the machine learning frameworks of Bayesian networks, quickly detect and classify APT-attacks using DL and C5.0 decision trees. When testing these models experimentally, a 10-fold cross-validation method is also employed. Results show that the C5.0 Bayesian network model, decision tree model, and 6-layer DL model all have respectable ACCs: 95.64, 88.37, and 98.85, respectively. In addition, the 6-layer DL model had FPR of 1.13, the Bayesian network model of 10.47 and the decision tree model of C5.0 model had FPR of 2.56, which is an essential threshold [14].

Ayoade *et al*. (2020) stated that attack campaigns are typically prolonged in an effort to avoid being discovered. In order to identify unusual activity, they utilize a method that generates execution traces of the host nodes by utilizing a provenance graph. When compared to baseline models, their methodology outperforms them by a wide margin, with an average increase of 11.3% in detection rate and 18.3% in true positive rate (TPR) [15].

Ghafir *et al*. (2019) stated that APT campaign elementary alerts are correlated in the phase by a correlation structure. Rudimentary notifications were generated with the help of an adjustable time window, and its properties were compared to form a correlation. Attack decoding is the second step of the suggested system. The proposed approach yields better predictability of the order of APT stages with at least 91.80 accuracy. Moreover, with two, three and four subsequent alerts, it predicts the subsequent phase of APT campaign with 66.50 percent, 92.70 percent

and 100 percent, respectively [16].

Prakash, Sankaran and Jithish (2019) proposed a new method of detecting threats of malware on smartphones based on elements of information theory and statistical techniques. They use the Kullback-Leibler divergence measure to identify unusual smartphone use patterns by parameterizing CPU utilisation, RAM utilisation, and network data. The findings of the experiment conducted to test the proposed approach indicate that the proposed method can detect dangerous activity of Android cellphones with an accuracy of 86.24 [17] [18].

Khosravi-Farmad, Ramaki and Bafghi (2018) stated that this was one of the frequently presented IKCs that were selected to be examined further on presentation and discussion. Further, the mapping study is conducted between models of IKC and the current methods of MTD following a novel and comprehensive taxonomy of MTD methods at different levels. Finally, and definitely not the last, the effect of MTD is evaluated in the case of a case study (IP Randomisation in particular). To ensure that they are not infiltrated by attacks based on IKC, the results of the testing indicate that the MTD approaches are the best [19] [20].

The current body of work on advanced persistent threat (APT) detections has looked at a range of solutions, with encoder-decoder models with attention mechanisms, ML and DL-based classification, provenance graph-based analysis of behaviors, correlation frameworks based on probabilistic models, lightweight statistical solutions in mobile settings, and moving target defense (MTD) solutions. Although these approaches show considerable improvement, a number of issues are yet to be addressed. Most methods are very dependent on certain datasets which restrict their applicability in varied settings. Others concentrate on one of the individual components of APT detection, including anomaly detection or stage prediction, but do not offer an integrated framework capable of addressing the stealthy, multi-stage and evolving aspects of APTs. Also, model-based approaches that are lightweight tend to lack accuracy, and all models based on deep learning are computationally intensive, which makes them hard to deploy in real-time. Besides, the flexibility of the existing methods to new types of attacks and zero-day attacks remains low. These loopholes point to the necessity of stronger, scaled, and smarter detection paradigms capable of compromising efficiency against quality, adjusting to any changing threats, and offering all-encompassing protection in various areas of application.

Table 1 below presents a consolidated overview of recent studies on Advanced Persistent Threat Detection, outlining the models applied, key findings, datasets utilized, existing challenges, and potential directions for future research.

3. Research Methodology

The proposed methodology begins with collecting the UNSW_NB15 dataset from Kaggle, followed by a comprehensive data preprocessing stage that includes data cleaning, standardizing numerical features using StandardScaler, removing outliers, and applying SMOTE for class balancing. After preprocessing, dataset is split

Table 1. Machine-deep learning approaches for efficient persistent threat identification.

Author & Year	Proposed Work	Dataset	Key Findings	Challenges & Future Work
Yan & Xiong (2020)	Web-APT-Detect (WAD), an unsupervised anomaly detection algorithm using encoder-decoder with attention mechanism	CSIC 2010	Achieved high F1-score, comparable to state-of-the-art supervised models	Limited to HTTP request patterns; needs evaluation on diverse datasets and real-world traffic
Joloudari *et al.* (2020)	Implemented a Bayesian network, DL, and C5.0 decision tree system for APT classification and detection.	NSL-KDD	DL model achieved the highest accuracy with the lowest false positive rate	Dependence on benchmark dataset; real-time adaptability and handling novel attacks remain challenging
Ayoade *et al.* (2020)	Provenance graph-based approach with online adaptive metric learning to detect anomalous behavior in prolonged APT campaigns	Execution traces of host nodes	Outperformed baseline models, improving detection accuracy and true positive rate	Computational overhead of provenance graph analysis; scalability and deployment in large systems need further study
Ghafir *et al.* (2019)	Correlation framework and Hidden Markov Model (HMM) for APT stage prediction	Correlated alerts from security systems	Accurately estimated APT stages and predicted next attacker steps with high precision	Relies on accurate correlation of alerts; effectiveness may decrease with noisy or incomplete data
Prakash, Sankaran & Jithish (2019)	Statistical and information-theoretic approach (Kullback-Leibler divergence) to detect malware-based attacks on smartphones	Smartphone system data (CPU, RAM, network usage)	Lightweight method, less computationally intensive, effective for low-power devices	Lower accuracy compared to ML/DL models; needs robustness against complex and evolving mobile malware
Khosravi-Farmad, Ramaki & Bafghi (2018)	Comprehensive taxonomy of MTD techniques; case study on IP randomization against intrusion activities	Case study (IKC models & MTD techniques)	MTD techniques improved defense against intrusion activities	Practical implementation challenges; performance overhead and integration with existing security systems require attention

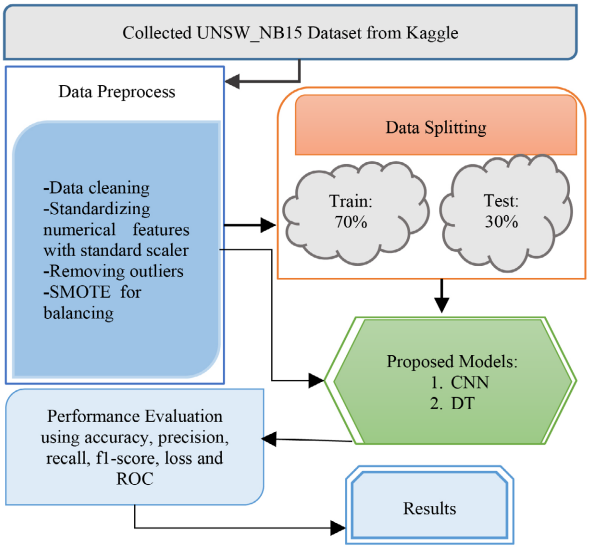

Figure 1. Proposed flowchart for Advanced persistent threat detection.

into testing (30%) and training (70%) subsets. Two ML models, CNN and DT, are then proposed for experimentation. Important measures are used to assess performance of these models like F1-score, precision, recall, accuracy, ROC, and loss, enabling a robust comparison and analysis of their effectiveness for the given task. Figure 1 shows the suggested methodology's total workflow [21].

The following section details of the steps outlined of flowchart are explained below.

3.1. Data Collection and Visualization

The UNSW_NB15 dataset, which was obtained from Kaggle, is used in experimental procedures. The UNSW_NB15 has 42 features in its clean format. Of the 42 features, 39 are numerical in nature and 3 are non-numeric (categorical) qualities. It has conducted EDA on the dataset. Here's the EDA plan used and analyzed for research work:

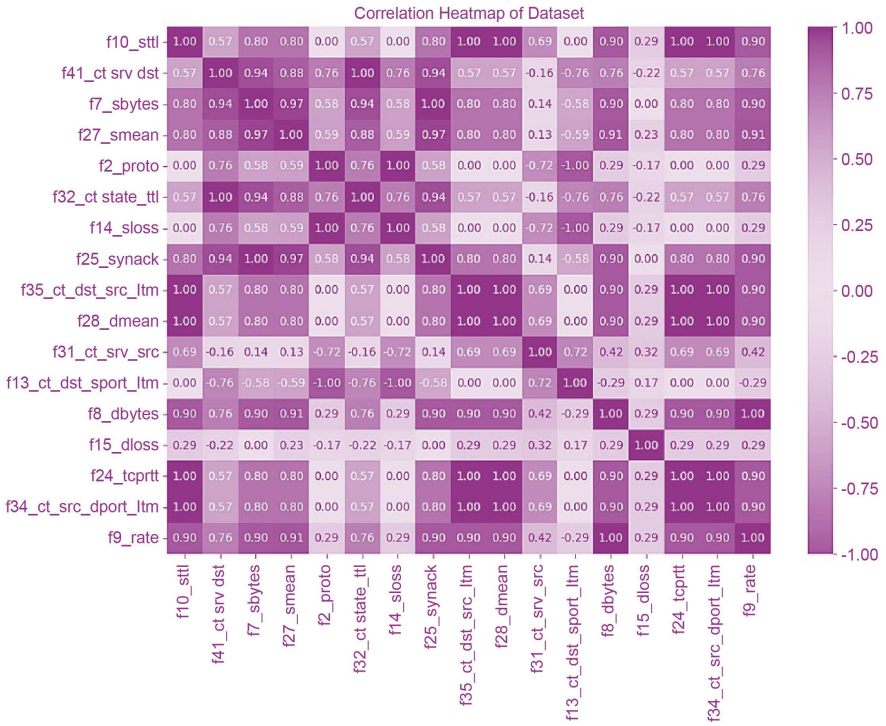

Figure 2. Correlation analysis.

A correlation heatmap of the dataset is displayed in Figure 2, which displays pairwise correlation coefficients between characteristics along the x and y axes. Correlation strength and direction are shown by color intensity; An extremely negative correlation is shown by dark blue (−1.00), a perfectly positive connection by dark red (1.00), and no correlation by white (0.00). The diagonal line of dark red squares reflects each feature's self-correlation. The heatmap reveals strong correlations among features such as f10_sttl, f41_ct_srv_dst, and f27_smean, while others, like f13_ct_dst_sport_ltm and f8_dbytes, exhibit weaker or negative

relationships.

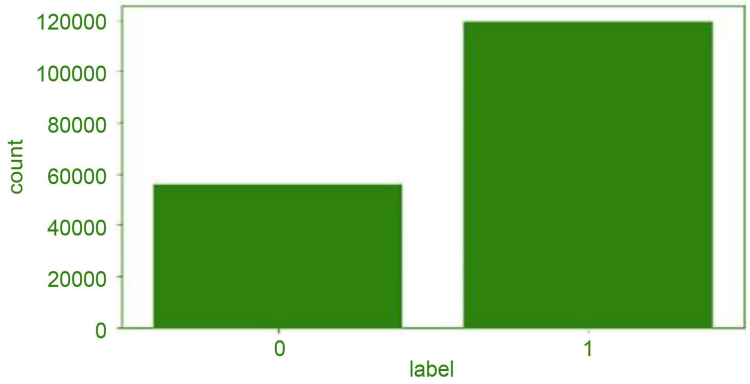

Figure 3. Number of normal and attack samples.

Figure 3 illustrates the class distribution within the dataset, with category 0 having just over 50,000 instances and category 1 nearly 120,000. The taller bar for category 1 highlights a clear class imbalance, where category 1 is significantly more prevalent than category 0.

3.2. Data Pre-Processing

Data preprocessing is initial and most crucial step in ML model creation [22]. In this study, several preprocessing techniques were applied, including data cleaning, outlier removal, standardization and SMOTE. The complete sequence of preprocessing steps is explained in detail below to ensure a clear understanding of how the data was prepared for modeling and visualization.

- **Data Cleaning:** The dataset had all the necessary data points. Most of the features are continuous in nature and are considered useful for the classification process.
- **Removing Outliers:** Outliers were removed from the dataset prior to the data visualization step to ensure a clearer and more accurate understanding of feature distributions.
- **Standardizing Numerical Features:** To guarantee uniform scaling across features, numerical columns were standardised using the StandardScaler. After being trained on training data, the "scaler" object was applied to both testing and training sets. It can find the formula in Equation (1):

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

in which z stands for the regularized or scaled value, the feature's starting value is denoted by x, its mean by μ, and its standard deviation by σ.

- **Data Balancing:** The UNSW_NB15 dataset is imbalanced, with one class smaller than the other. Imbalance can be addressed via sampling-based or cost function-based approaches. This paper adopts sampling, which includes under-sampling and over-sampling. By eliminating samples at random, under-sampling

lowers the majority class, but this may discard important data and reduce accuracy [23]. Over-sampling duplicates minority-class examples, avoiding data loss but risking overfitting. To address this, this paper employs SMOTE, which interpolates between a sample and its closest neighbors to create synthetic minority-class samples, improving CNN and DT model training [24] [25]. If the data point (x_1, x_2) belongs to a minority class and its nearest neighbor is chosen as (x_1', x_2'), as shown in Equation (2), then the data point is synthesized.

$$(X_1, X_2) = (x_1, x_2) + \text{random}(0,1) \times \Delta \tag{2}$$

where, $\Delta = \{(x_1' - x_1), (x_2' - x_2)\}$ and $\text{random}(0,1)$ represent a random value between 0 and 1.

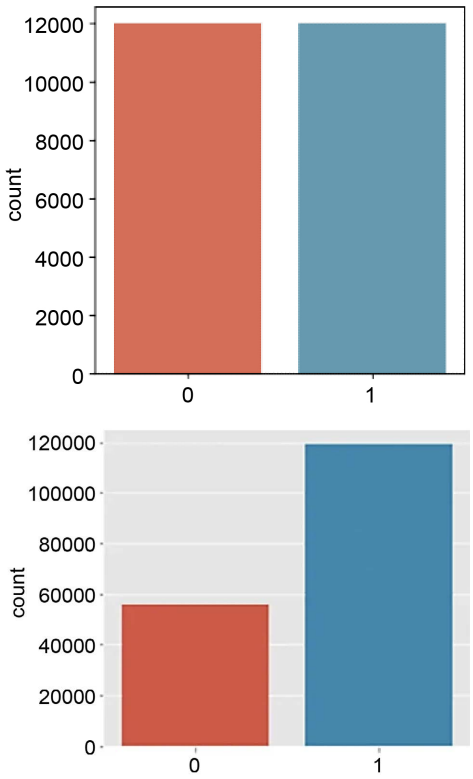

Figure 4. Imbalanced and Balanced graph of the dataset.

Figure 4 shows the dataset's class distribution both before and after the SMOTE was applied. The unbalanced dataset is displayed in the graph on the left, where the majority class significantly outnumbers the minority class. The graph on the right depicts the balanced dataset obtained after applying SMOTE, where both classes have an equal number of samples, ensuring that model training process is not biased toward the majority class [26].

3.3. Data Splitting

A 70-30 split ratio was employed to separate dataset into testing and training sub-

sets, with the majority of the data being used for model training. and the remaining 30% being set aside for performance evaluation.

3.4. Proposed ML and DL Models

In this section discuss the CNN and DT models.

1) Convolutional Neural Network (CNN)

The area of computer vision has been greatly improved by use of Convolutional Neural Networks (CNNs) [27] [28]. An ordinary convolutional neural network (CNN) has three layers: a pooling layer, a convolutional layer, and a fully connected prediction layer [21]. One extremely beneficial property in the context of classification of images is the fact that the pooling and convolutional layer ensures that features extracting the inputs are positional and rotational invariant features. Having tens of convolutional and pooling layers, the modern CNNs are incredibly deep and can demonstrate promising results in the field of image categorization. As an example, Equation (3) shows the flow of data of the simple CNN. The equation can be shown as follows: "FC" stands for "fully connected line". "Pool" denotes the stratum that collects excess water; The activation function is denoted as ReLU. The meaning of "conv" is "convolutional layer." The CNN takes $XR(N\times T\times F)$ as input. N means nodes; T means length of time series; and F means features.

$$Y = \mathrm{FC}\left(\mathrm{Pool}\left(\mathrm{ReLU}\left(\mathrm{Conv}\,\mathrm{X}\right)\right)\right) \tag{3}$$

The simple CNN can be used as an example of data flow as depicted in Equation (3). Here is the equation: Y symbolizes the prediction outcome; CNN input is N; T stands for time series length; A fully connected layer is represented by FC, and F stands for the number of features. The activation function is denoted as ReLU, the pooling layer is known as Pool, and the convolutional layer is called Conv [29].

2) Decision Tree (DT)

One kind of machine-learning technology that is frequently used for both regression and classification applications is decision trees. Using the data from the training set, the algorithm constructs a decision-and-result tree model [30] [31]. The decision-making process is easier to see and understand because of the tree layout. Decision trees offer insights about feature relevance in an easily interpreted tree-like form. Non-linear relationships can be captured using the tree-based method. Several threat detection designs in IDS have included decision trees, especially when it comes to identifying pertinent aspects for classification [32].

3.5. Evaluation Metrics

In this work, a confusion matrix was used to assess the models created with classification algorithms. F-score, recall, precision, and accuracy were the four statistical metrics used to assess performance. Performance metrics listed below are as follows:

- **True Positives (TP):** indicates that an APT assault happens when an alert is generated [33].
- **True Negatives (TN):** indicates that an APT assault does not take place when no alarm is raised.
- **False Positives (FP):** represents the situation in which an APT attack does not take place despite a warning being generated.
- **False Negatives (FN):** represents the situation where an APT attack takes place without an alert being triggered.

1) Accuracy

The percentage of data that are correctly classified relative to the total amount of data is called accuracy [34]. The percentage of correctly categorized outcomes can be expressed using Equation (4) as follows.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{4}$$

2) Precision

Precision measures the proportion of accurately classified attack data to total attack data, indicating the number of expected attacks that really occur. This can be illustrated as shown in Equation (5).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5}$$

3) Recall

Recall displays the ratio of anticipated attacks to total attacks. Equation (6) provides a description of this.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{6}$$

4) F1-Score

It represents balanced performance in both precision and recall and as weighted is calculated average of PR and RE. The weighted average of all values is one way to conceptualize it, as illustrated in Equation (7).

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{7}$$

Loss: It refers to the numerical value that measures how far a model's predicted outputs deviate from the actual target values [35] [36].

ROC: A graphical depiction of a classification model's performance across all categorization criteria is called Receiver Operating Characteristic Curve (ROC).

4. Results and Discussion

A system with 16 GB of RAM and an Intel Core i5 CPU running at 2.30 GHz is used for the experiment. Spyder and Jupyter Notebook, which are installed within the Anaconda Python 3.7 environment, are the IDEs utilized. Table 2 shows a comparison of the performance of the proposed model of APT detection, which uses CNN and DT classifiers. The CNN model had the best results in all the measures of evaluation with a recall of 99.14, a precision of 98.71, an F1-score of

98.92, and an accuracy of 98.85, indicating its high ability to identify and categorize the persistent threats with a low number of FNs and FPs. Conversely, the DT classifier, although having a relatively good performance, had relatively lower scores in terms of precision of 96.12, F1-score of 96.12, recall 96.12, and accuracy of 95.64. The obtained results are a clear indication that CNN-based method is better than the DT model and is better fitted to the complexity of APT detection and its elusive nature.

Table 2. The Recommended Model's Execution on the UNSW NB15 Dataset.

Performance Matrix	CNN	DT
Accuracy	98.85	95.64
Precision	98.71	96.12
Recall	99.14	96.12
F1-Score	98.92	96.12

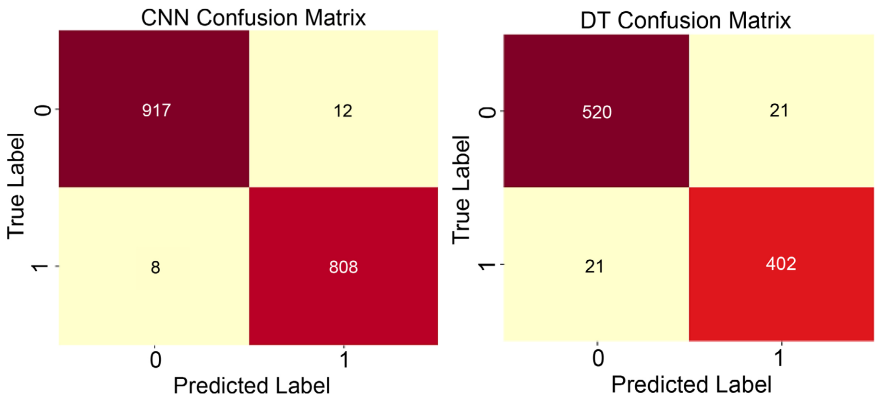

Figure 5. Confusion matrices of the CNN & DT model.

A CNN model's and a DT model's confusion matrices are shown in Figure 5. The model produced 808 True Positives (identifying 808 cases of class 1) and 8 False Negatives (erroneously classifying 8 examples of class 1 as class 0), according to the CNN Confusion Matrix. The model produced 21 False Positives (erroneously categorizing 21 cases of class 0 as class 1) and 520 True Negatives (rightly classifying 520 instances of class 0 as class 0), according to the DT Confusion Matrix.

The CNN and DT models' ROC curves are shown in Figure 6. The CNN achieved an AUC score of 0.9990, displaying nearly faultless categorization capabilities, with an exceptionally high rate of correct classifications and an extremely low rate of incorrect ones. In comparison, the DT model obtained an AUC of 0.9775, also indicating strong performance but slightly lower than the CNN. The results highlight the superior predictive capability of the CNN model over the DT model.

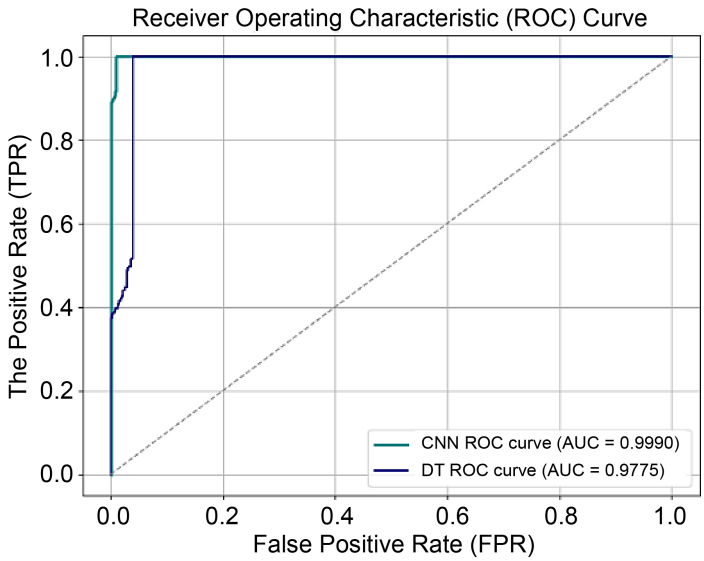

Figure 6. ROC analysis of the CNN and DT models.

The CNN model's validation and training results across 30 epochs are shown in Figure 7. The loss curves are displayed on the left graph, where training and validation losses gradually decline and converge to zero, indicating effective error minimization and absence of overfitting. The right graph presents the accuracy curves, with both training and validation accuracy rising rapidly and stabilizing above 0.9, demonstrating that the model generalizes well and achieves consistently high performance on both validation and training datasets.

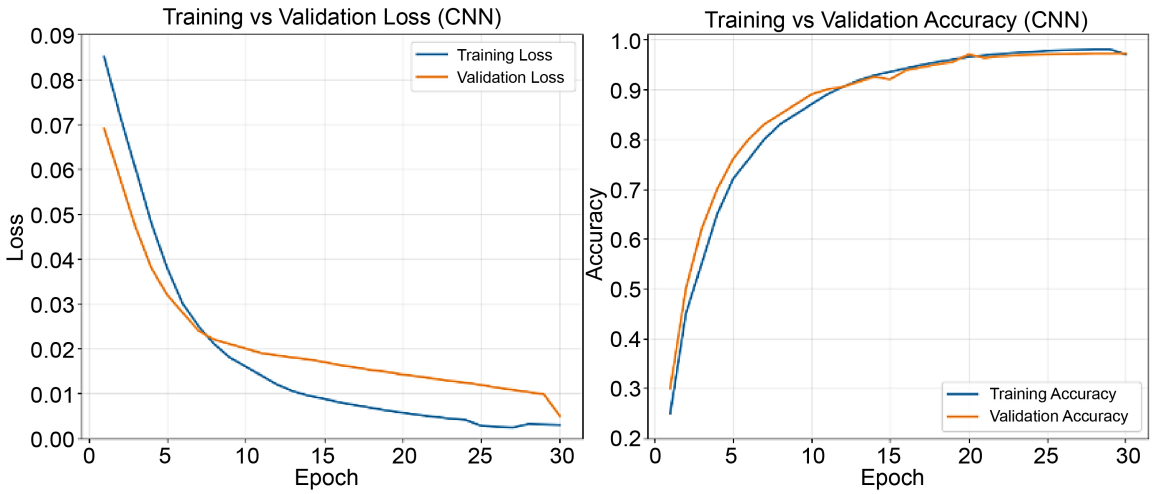

Figure 7. Validation and Training Performance Loss and Accuracy Graphs of the CNN Model.

Comparative Analysis

Table 3 presents the performance comparison of various models applied to Advanced Persistent Threat (APT) detection. Among the baseline models, the ANN achieved a recall of 98.06%, a precision of 81.54%, an accuracy of 86.71%, and an

F1-score of 89.04%, while the SVM performed similarly with 87.05% accuracy, 82.95% precision, an F1-score of 89.11%, and 96.27% recall. The RF yielded balanced results with 86.5% accuracy, 87.6% recall, 85.9% precision, and an F1-score of 86.7%. With the best recall (99.14%), precision (98.71%), accuracy (98.85%), and F1-score (98.92%) of any model tested, the suggested CNN clearly came out on top. The DT also showed competitive performance, with 95.84% accuracy and consistent F1-score, recall, and precision values of 96.12%. The results show that the CNN model is the best at detecting APTs.

Table 3. Performance comparison of different models for advanced persistent threat detection.

Models	Accuracy	Precision	Recall	F1-Score
ANN [23]	86.71	81.54	98.06	89.04
SVM [24]	87.05	82.95	96.27	89.11
RF [25]	86.5	85.9	87.6	86.7
CNN	98.85	98.71	99.14	98.92
DT	95.84	96.12	96.12	96.12

The suggested CNN and DT models provide different detection of APT. Its CNN model uses its deep learning structure to automatically identify the more complicated patterns and subtle features of the data, and is able to be effective in capturing the nature of persistent threats, which is both stealthy and evolving [37]. The fact that it can generalize well on both training and validation datasets makes it very applicable in real-life, where the behavior of the attacks can be different. Conversely, the DT model is easy to understand, interpret, and use, which is valuable in situations where transparency and speed of decisions are paramount. The combination of these models exhibits a tradeoff between high feature learning and interpretability, which reinforces their use in cybersecurity defense systems [38] [39].

5. Conclusions and Future Study

Another type of cyberattack that is terrifying, APTs are characterized by their specificity, advanced level, and long-term course. These information attacks that are usually executed in silence and without notice are a great danger to individuals and organizations and are designed by an extremely motivated enemy. This paper used the dataset of UNSW_NB15 as its conclusion and showed the efficiency of ML and DL methods in identifying APT. Models were evaluated and trained through extensive EDA, data preprocessing and class balancing, and results indicated that CNN was much better than baseline models and the DT classifier, with better recall, precision, accuracy, and F1-score as well as an unparalleled AUC score. The CNN, where complex patterns are automatically learnt is most appropriate to identify stealth and changing cyber-attacks, and the DT offers interpret-

ability and ease in the required decision-making procedures.

In future research, the research may be conducted by examining hybrid deep learning models, ensemble models, and attention-based mechanisms to achieve better detection. Furthermore, real-time APT detection systems, scaling tests in large-scale network traffic and implementation in dynamic cybersecurity systems need to be considered to guarantee the resilience and flexibility to the emerging and advanced threats.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Shaukat, K., Luo, S., Chen, S. and Liu, D. (2020) Cyber Threat Detection Using Machine Learning Techniques: A Performance Evaluation Perspective. 2020 *International Conference on Cyber Warfare and Security* (*ICCWS*), Islamabad, 20-21 October 2020, 1-6. https://doi.org/10.1109/iccws48432.2020.9292388

[2] Berrada, G., Cheney, J., Benabderrahmane, S., Maxwell, W., Mookherjee, H., Theriault, A., *et al.* (2020) A Baseline for Unsupervised Advanced Persistent Threat Detection in System-Level Provenance. *Future Generation Computer Systems*, **108**, 401-413. https://doi.org/10.1016/j.future.2020.02.015

[3] Thapliyal, A., Bhagavathi, P.S., Arunan, T. and Rao, D.D. (2009) Realizing Zones Using UPnP. 2009 6*th IEEE Consumer Communications and Networking Conference*, Las Vegas, 10-13 January 2009, 1-5. https://doi.org/10.1109/ccnc.2009.4784867

[4] Alshamrani, A., Myneni, S., Chowdhary, A. and Huang, D. (2019) A Survey on Advanced Persistent Threats: Techniques, Solutions, Challenges, and Research Opportunities. *IEEE Communications Surveys & Tutorials*, **21**, 1851-1877. https://doi.org/10.1109/comst.2019.2891891

[5] Rao, D.D. (2009) Multimedia Based Intelligent Content Networking for Future Internet. 2009 *Third UKSim European Symposium on Computer Modeling and Simulation*, Athens, 25-27 November 2009, 55-59. https://doi.org/10.1109/ems.2009.108

[6] Ghafir, I. and Prenosil, V. (2014) Advanced Persistent Threat Attack Detection: An Overview. *International Journal of Advancements in Computer Networks and Its Security*, **4**, 50-54.

[7] Kapadia, H.P. (2020) Cross-Platform UI/UX Adaptions Engine for Hybrid Mobile Apps. *International Journal of Novel Research and Development*, **5**, 30-37.

[8] Neeli, S.S.S. (2020) Real-Time Data Management with In-Memory Databases: A Performance-Centric Approach. *Journal of Advances in Developmental Research*, **11**, 1-8.

[9] Hasan, K., Shetty, S. and Ullah, S. (2019) Artificial Intelligence Empowered Cyber Threat Detection and Protection for Power Utilities. 2019 *IEEE* 5*th International Conference on Collaboration and Internet Computing* (*CIC*), Los Angeles, 12-14 December 2019, 354-359. https://doi.org/10.1109/cic48465.2019.00049

[10] Neeli, S.S.S. (2019) The Significance of NoSQL Databases: Strategic Business Approaches and Management Techniques. *Journal of Advances in Developmental Research*, **10**, 1-11.

[11] Liu, H. and Lang, B. (2019) Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey. *Applied Sciences*, **9**, Article 4396.

https://doi.org/10.3390/app9204396

[12] Kushwaha, A., Pathak, P. and Gupta, S. (2016) Review of Optimize Load Balancing Algorithms in Cloud. *International Journal of Distributed and Cloud Computing*, **4**, 1-9.

[13] Yan, L. and Xiong, J. (2020) Web-Apt-Detect: A Framework for Web-Based Advanced Persistent Threat Detection Using Self-Translation Machine with Attention. *IEEE Letters of the Computer Society*, **3**, 66-69. https://doi.org/10.1109/locs.2020.2998185

[14] Hassannataj Joloudari, J., Haderbadi, M., Mashmool, A., Ghasemigol, M., Band, S.S. and Mosavi, A. (2020) Early Detection of the Advanced Persistent Threat Attack Using Performance Analysis of Deep Learning. *IEEE Access*, **8**, 186125-186137. https://doi.org/10.1109/access.2020.3029202

[15] Ayoade, G., Akbar, K.A., Sahoo, P., Gao, Y., Agarwal, A., Jee, K., *et al.* (2020) Evolving Advanced Persistent Threat Detection Using Provenance Graph and Metric Learning. 2020 *IEEE Conference on Communications and Network Security* (*CNS*), Avignon, 29 June-1 July 2020, 1-9. https://doi.org/10.1109/cns48642.2020.9162264

[16] Ghafir, I., Kyriakopoulos, K.G., Lambotharan, S., Aparicio-Navarro, F.J., Assadhan, B., Binsalleeh, H., *et al.* (2019) Hidden Markov Models and Alert Correlations for the Prediction of Advanced Persistent Threats. *IEEE Access*, **7**, 99508-99520. https://doi.org/10.1109/access.2019.2930200

[17] Prakash, J., Sankaran, S. and J., J. (2019) Attack Detection Based on Statistical Analysis of Smartphone Resource Utilization. 2019 *IEEE* 16*th India Council International Conference* (*INDICON*), Rajkot, 13-15 December 2019, 1-4. https://doi.org/10.1109/indicon47234.2019.9030310

[18] Khosravi-Farmad, M., Ramaki, A.A. and Bafghi, A.G. (2018) Moving Target Defense against Advanced Persistent Threats for Cybersecurity Enhancement. 2018 8*th International Conference on Computer and Knowledge Engineering* (*ICCKE*), Mashhad, 25-26 October 2018, 280-285. https://doi.org/10.1109/iccke.2018.8566531

[19] Fernandez, A., Garcia, S., Herrera, F. and Chawla, N.V. (2018) SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-Year Anniversary. *Journal of Artificial Intelligence Research*, **61**, 863-905. https://doi.org/10.1613/jair.1.11192

[20] Yang, R., Zhang, C., Gao, R. and Zhang, L. (2016) A Novel Feature Extraction Method with Feature Selection to Identify Golgi-Resident Protein Types from Imbalanced Data. *International Journal of Molecular Sciences*, **17**, Article 218. https://doi.org/10.3390/ijms17020218

[21] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/cvpr.2016.90

[22] Mutyala, N.K., Koushik, K.V.S. and Deepak, K. (2018) Prediction of Heart Diseases Using Data Mining and Machine Learning Algorithms and Tools. *International Journal of Scientific Research in Computer Science*, *Engineering and Information Technology*, **3**, 262-269.

[23] Kasongo, S.M. and Sun, Y. (2020) Performance Analysis of Intrusion Detection Systems Using a Feature Selection Method on the UNSW-NB15 Dataset. *Journal of Big Data*, **7**, Article No. 105. https://doi.org/10.1186/s40537-020-00379-6

[24] Li, Y., Xu, W. and Ruan, Q. (2020) Research on the Performance of Machine Learning Algorithms for Intrusion Detection System. *Journal of Physics*: *Conference Series*, **1693**, Article ID: 012109. https://doi.org/10.1088/1742-6596/1693/1/012109

[25] Ren, J., Guo, J., Qian, W., Yuan, H., Hao, X. and Jingjing, H. (2019) Building an Effective Intrusion Detection System by Using Hybrid Data Optimization Based on Machine Learning Algorithms. *Security and Communication Networks*, **2019**, Article ID: 7130868. https://doi.org/10.1155/2019/7130868

[26] Gangineni, V.N., Tyagadurgam, M.S.V., Pabbineedi, S., Penmetsa, M., Bhumireddy, J.R. and Chalasani, R. (2024) AI-Powered Cybersecurity Risk Scoring for Financial Institutions Using Machine Learning Techniques. *Journal of Artificial Intelligence & Cloud Computing*, **3**, 1-9.

[27] Vangala, S.R., Polam, R.M., Kamarthapu, B., Kakani, A.B., Nandiraju, S.K.K. and Chundru, S.K. (2025) A Machine Learning-Based Framework for Predicting and Improving Student Outcomes Using Big Educational Data (approved by ICITET 2024) *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.5515379

[28] Gangineni, V.N., Pabbineedi, S., Kakani, A.B., Nandiraju, S.K.K., Chundru, S.K. and Tyagadurgam, M.S.V. (2023) AI-Enabled Big Data Analytics for Climate Change Prediction and Environmental Monitoring. *International Journal of Emerging Trends in Computer Science and Information Technology*, **4**, 71-79.

[29] Pabbineedi, S., Kakani, A.B., Nandiraju, S.K.K., Chundru, S.K., Tyagadurgam, M.S.V. and Gangineni, V.N. (2023) Scalable Deep Learning Algorithms with Big Data for Predictive Maintenance in Industrial IoT. *International Journal of AI, BigData, Computational and Management Studies*, **4**, 88-97.

[30] Bhumireddy, J.R., Chalasani, R., Tyagadurgam, M.S.V., Gangineni, V.N., Pabbineedi, S. and Penmetsa, M. (2022) Predictive Models for Early Detection of Chronic Diseases in Elderly Populations: A Machine Learning Perspective. *International Journal of Computing and Artificial Intelligence*, **3**, 71-79. https://doi.org/10.33545/27076571.2022.v3.i1a.169

[31] Polam, R.M. (2023) Predictive Machine Learning Strategies and Clinical Diagnosis for Prognosis in Healthcare: Insights from MIMIC-III Dataset. *International Journal of Computing and Artificial Intelligence*, **4**, 80-88.

[32] Gangineni, V.N., Pabbineedi, S., Penmetsa, M., Bhumireddy, J.R., Chalasani, R. and Tyagadurgam, M.S.V. (2022) Efficient Framework for Forecasting Auto Insurance Claims Utilizing Machine Learning Based Data-Driven Methodologies. *International Research Journal of Economics and Management Studies*, **1**, 45-53.

[33] Vikram Tyagadurgam, M.S., Gangineni, V.N., Pabbineedi, S., Penmetsa, M., Bhumireddy, J.R. and Chalasani, R. (2022) Designing an Intelligent Cybersecurity Intrusion Identify Framework Using Advanced Machine Learning Models in Cloud Computing. *Universal Library of Engineering Technology*, 18-26. https://doi.org/10.70315/uloap.ulete.2022.003

[34] Chalasani, R., Tyagadurgam, M.S.V., Gangineni, V.N., Pabbineedi, S., Penmetsa, M. and Bhumireddy, J.R. (2022). Leveraging Big Datasets for Machine Learning-Based Anomaly Detection in Cybersecurity Network Traffic.

[35] Bhumireddy, J.R., Chalasani, R., Tyagadurgam, M.S.V., Gangineni, V.N., Pabbineedi, S. and Penmetsa, M. (2022) Big Data-Driven Time Series Forecasting for Financial Market Prediction: Deep Learning Models. *Journal of Artificial Intelligence and Big Data*, **2**, 153-164. https://doi.org/10.31586/jaibd.2022.1341

[36] Vangala, S.R., Polam, R.M., Kamarthapu, B., Kakani, A.B., Kireeti Nandiraju, S.K. and Chundru, S.K. (2022) Leveraging Artificial Intelligence Algorithms for Risk Prediction in Life Insurance Service Industry. *Universal Library of Engineering Technology*, 27-34. https://doi.org/10.70315/uloap.ulete.2022.004

[37] Sandeep Kumar, C., Srikanth Reddy, V., Ram Mohan, P., Bhavana, K. and Ajay Babu,

K. (2022) Efficient Machine Learning Approaches for Intrusion Identification of DDoS Attacks in Cloud Networks.

[38] Polu, A.R., Narra, B., Buddula, D.V.K.R., Patchipulusu, H.H.S., Vattikonda, N. and Gupta, A.K. (2022) Blockchain Technology as a Tool for Cybersecurity: Strengths, Weaknesses, and Potential Applications. *Accent Journal of Economics Ecology & Engineering*, **7**, 167-174.

[39] Nandiraju, S.K.K., Chundru, S.K., Vangala, S.R., Polam, R.M., Kamarthapu, B. and Kakani, A.B. (2022) Advance of AI-Based Predictive Models for Diagnosis of Alzheimer's Disease (AD) in Healthcare. *Journal of Artificial Intelligence and Big Data*, **2**, 141-152. https://doi.org/10.31586/jaibd.2022.1340