International Journal on
"Technical and Physical Problems of Engineering"
(IJTPE)
Published by International Organization of IOTPE

ISSN 2077-3528
IJTPE Journal
www.iotpe.com
ijtpe@iotpe.com

March 2023 | Issue 54 | Volume 15 | Number 1 | Pages 1-10

ATTRIBUTION CLASSIFICATION METHOD OF ADVANCED PERSISTENT THREAT (APT) MALWARE USING AI LEARNING

E.J. Khalefa D.A.A. Salman

Department of Computer Science, College of Science, University of Diyala, Baqubah, Iraq
scicompms2106@uodiyala.edu.iq, dhahair@sciences.uodiyala.edu.iq

Abstract- APT is a frequent and expensive system attack. Businesses, governments, and other organizations must defend against this attack. Utilizing machine learning or deep learning algorithms to scan network traffic for signals and anomalies to detect and thwart APT attacks has become commonplace. The lack of specific attack campaign data hinders APT behavior analysis and evaluation methods. Network traffic analysis can help detect APT attacks. This paper proposes two adaptable strategies. Machine learning and deep learning algorithms classify APT Malware. Binary-class classification identifies two-class APT Malware and ordinary Malware; multiclass classification identifies 15 APT malware organizations and normal Malware. Each system has two classification subsystems: machine learning based on Random Forest and LightGBM algorithms and deep learning using a hybrid CNN and long short-term memory (LSTM). EDA (exploratory data analysis) detects and removes outlier data, ETC selects essential features, and SMOTE solves unbalanced data problems. APT Malware dataset with 11,107 samples in 16 classes. Each proposed system is studied separately, and machine and deep learning accuracy results are compared. Four case studies were also conducted to evaluate machine learning algorithm performance and the impact of feature selection and SMOTE technology. Machine learning results showed the effect of feature selection and SMOTE on both proposed systems. The binary class classification system results show that machine learning outperforms deep learning, with random forest accuracy of 0.999723, Light GBM accuracy of 0.999480, and CNN-LSTM hybrid accuracy of 0.914798. The multi-class classification system showed that machine learning performs better than deep learning; Light GBM has an accuracy of 0.999727, random forest is 0.999632, and CNN-LSTM is 0.798206.

Keywords: APT, Attack, Data, Machine Learning, Deep Learning, Hybrid.

1. INTRODUCTION

The evolution of cyberattacks has paralleled that of computer technology and the Internet, from the earliest viruses and worms to today's botnets. Rootkits modify their behavior by updating their "source" in accordance with changing software and hardware [1]. Businesses and organizations consistently attempt to safeguard their data and information from cyberattacks [2]. Most enterprise threats are focused and persistent, including some APT [3]. APT is defined by UNNIST using American terminology [4]. A capable and well-resourced adversary can achieve its objectives via cyber, physical, and deception (e.g., cyber, physical, and deception). Advanced persistent threat threats concern international governments and corporations [5]. Because the attackers employ various techniques to remain undetected and evade detection, these attacks constitute an immediate, difficult-to-identify threat. APT and conventional cyberattacks are distinct. How many assets are needed for the assault? A typical cyberattack targets an individual or group of cyber criminals and organizations with inadequate or nonexistent cybersecurity [6]. Cyber espionage's "who" and "why" are determined via attribution. This technique identifies APT threat actors and infiltration targets. The security community is currently examining preliminary data. Classifying these attacks requires recognizing and overcoming organizational issues [7].

Given the importance of APT attribution and categorization to commercial security corporations and public sector organizations, both significant data processing and analysis are required. Data mining and machine learning have been proposed to tackle these challenges [8]. This research will classify APT Malware using Machine Learning and Deep Learning. Each component of the proposed system is trained, validated, and tested using the APT Malware dataset. Proposed are Binary-Class and Multiclass classifications. The deep learning branch is developed using CNN and Long Term-Short Memory [9]. The categorization algorithms of the proposed system will be compared based on accuracy metrics to evaluate their efficacy in combatting APT Malware.

2. ADVANCED PERSISTENT THREAT (APT)

APT is a new network assault that can freely employ several methodologies. APT persistently collects data from a specific target by exploiting vulnerabilities using several attack techniques [10]. As organizations and governments become more targets, APT system attacks

have become critical. These attacks target the victim's network to access sensitive information for espionage or to breach the network, compromising the victim's systems and stealing their data [11]. APT assaults, also known as targeted attacks, are always undertaken by nation-state actors. Advanced persistent threats are long-term network attacks on specific targets employing advanced attack techniques [12]. A well-organized malicious cyber-attack uses a difficult-to-detect strategy, technique, and process (TTP) and targets a particular set of enterprises for long-term network access. [13]. According to Table 1, APTs differ from typical malware assaults in attack description, perpetrator, target, objective, and life cycle.

Table 1. Distinctions between APT and Common Malware Attacks [13]

Feature	APT Attack	Common Malware Attack
Definition	Sophisticated, target very specialized groups	"Malware" attacks and disrupts digital systems (e.g., ransomware)
Attacker	Actors in the government and criminal groups	In other words, a cracker (a hacker in illegal activities)
Target	Finance and banking, Military, I.T. businesses	Business computer or target any personal
Purpose	The purpose of this assault is to harm or steal sensitive data from a specified target	This attack's goal is to financial gain
Life Cycle Attack	Use as many different approaches as possible to keep going	When the security measures (firewalls, virus scanners, etc.)

3. RELATED WORKS

Several efforts have tried identifying and categorizing the APT problem with increased cyberattacks. Will review related work in APT detection. In C.D. Xuan, et al. (2021) [14]. offer a multilayer analysis-based APT assault detection technique. By calculating and analyzing a range of Network Traffic events to find and synthesize aberrant indications and behaviors, the multilayer analysis methodology in the. proposed method APT may be recognized in the system. This method uses serial identification of adware and other potentially unwanted programs. There are three primary methods: 1) using anomalous connections to identify APT attacks, 2) using Suricata logs to detect APT attacks, and 3) using behavior profiles created from the layers mentioned above to detect APT attacks (ii). APT attack detection will use the multilayer analysis approach to help accomplish these objectives. Two The main goal is to analyze and assess network traffic components in light of odd signs or behavior. 2.) developing and classifying behavior profiles based on network data elements. The experimental portion compares and assesses how well each layer of the multilayer analysis model performs using machine learning to detect APT attacks.

According to test results utilizing the attack dataset compiled from 29 Network Traffic files in the Malware Capture CTU-13 data set, which comprises six different forms of Malware from APT assaults, the suggested technique has the best accuracy at 96.70 percent when using the Random Forest algorithm. S. Li, et al. (2021) [15]. Using machine learning, propose a categorization approach for attributing APT Malware to enterprises in the IoT. The major objective of this initiative is to identify and secure IoT devices against APT assaults. This technique employs information about APT activity collected from IoT devices to represent and choose business-specific characteristics with a high degree of variation. SMOTE-RF is used to train SMOTE-RF, a multiclass technique better equipped to manage concerns with imbalance and multi-classification. According to the experiments, the SMOTE-RF model successfully and consistently classifies APT Malware, achieving greater than 80% accuracy in general models.

W. Han, et al. (2021) [16], proposed malware detection technology called APTMalInsight will be shown. APTs (Advanced Persistent Threats) were detected and recognized using the information and ontology knowledge framework. According to the paper, many acquired feature vectors may be used to detect and cluster APT Malware. ATPs Attacks may be detected and identified using Random Forest, Decision Tree, KNN (K-Nearest Neighbors), and XGBoost machine learning algorithms. A Random Forest approach may achieve detection and clustering accuracy of 99.28 percent and 98.85 percent, respectively, based on actual APT malware samples evaluated. G. Wang, et al. in (2021) [17]. new detection approach based on belief rule basis (BRB), where expert knowledge and small samples are used to achieve interpretable detection results. Analyzing network data and building a BRB model that considers expert knowledge while effectively expressing uncertainty can help detect an APT attack. The BRB model is trained on a limited number of samples to provide more accurate results. The proposed method compares two different techniques for identifying APT in small samples. These methods support SVMs and multilayer perceptions (MLPs). BRB achieved a dependable and acceptable accuracy of 91% in the limited sample situation. Despite their great accuracy (93.23 and 95.32 percent, respectively), MLP and SVM cannot be relied upon due to their illogical nature. BRB uses extensive expertise to define the initial values of parameters, hence minimizing the number of samples necessary. Due to training, small samples are no longer relevant. The interpretation of the BR model is preferred to the fitted models since it is based on principles, making BRB outcomes easier to examine and interpret.

C.D. Xuan and M.H. Dao, (2021) [18] worked on a novel approach to detecting based on network traffic monitoring, APT attacks, and a mixed DL model. It has been shown that neural nets such as multilayer perceptron (MLP), convolutional neural nets (CNN), and long short-term memory may be used to detect APT assaults in network traffic (LSTM). The integrated models based on deep learning are used in two phases to identify APT attack signals, including extracting I.P. characteristics from the flow of analyzing network traffic by I.P. address. To extract I.P. from the integrated deep learning models, we will next use their characteristics flow; (ii) APT attack categorization attacks on I.P.s by APT I.P.s and normal I.P.s will be recognized and categorized based on I.P.

attributes retrieved in a job I in ii. This study's results demonstrate that the combined deep learning models were superior in their ability to assure accuracy on all metrics ranging from 93% to 98% which are shown by F.J. Abdullayeva (2021) [19].

Categorize APT attacks using a deep autoencoder neural network. In essence, the technique suggested in this paper uses deep learning to choose informative features as features are being selected automatically. This method uses the autoencoder to learn valuable features and then the SoftMax regression layer to categorize APT kinds. It is a machine learning dataset with three data classes: APT1, Crypto, and other assaults. There was 98.32 per cent correctness in the proposed design. This concept may identify APT threats connected to the cloud by placing them between the user and the cloud. An OTP-based two-factor authentication scheme was also presented in the article to protect Cloud Systems against APT assaults. It is nearly impossible for an intruder to get access using the suggested two-factor authentication system.

4. PROPOSED SYSTEM

The suggested method for classifying APT attack detection-based assaults comprises machine learning and deep learning techniques. It has two sub-classification systems: a Binary Class Classifier for detecting binary classes (APT malware attacks and non-APT malware attacks) and a Multi-class Class Classifier for identifying multiple categories. The proposed method included the following steps: first, read data from the dataset, clean data using Exploratory Data Analysis EDA technique, remove outlier, select important features using Extra Trees Classifier (ETC), split dataset, balance data of the training dataset using (Smote), and finally, these features are put into two models CNN-LSTM deep learning; and (LightGBM and Random Forest) Machine Learning to train a classifier for distinguishing APT samples. Figure 1 depicts the suggested processes for determining APT malware attacks from non-APT malware attacks.

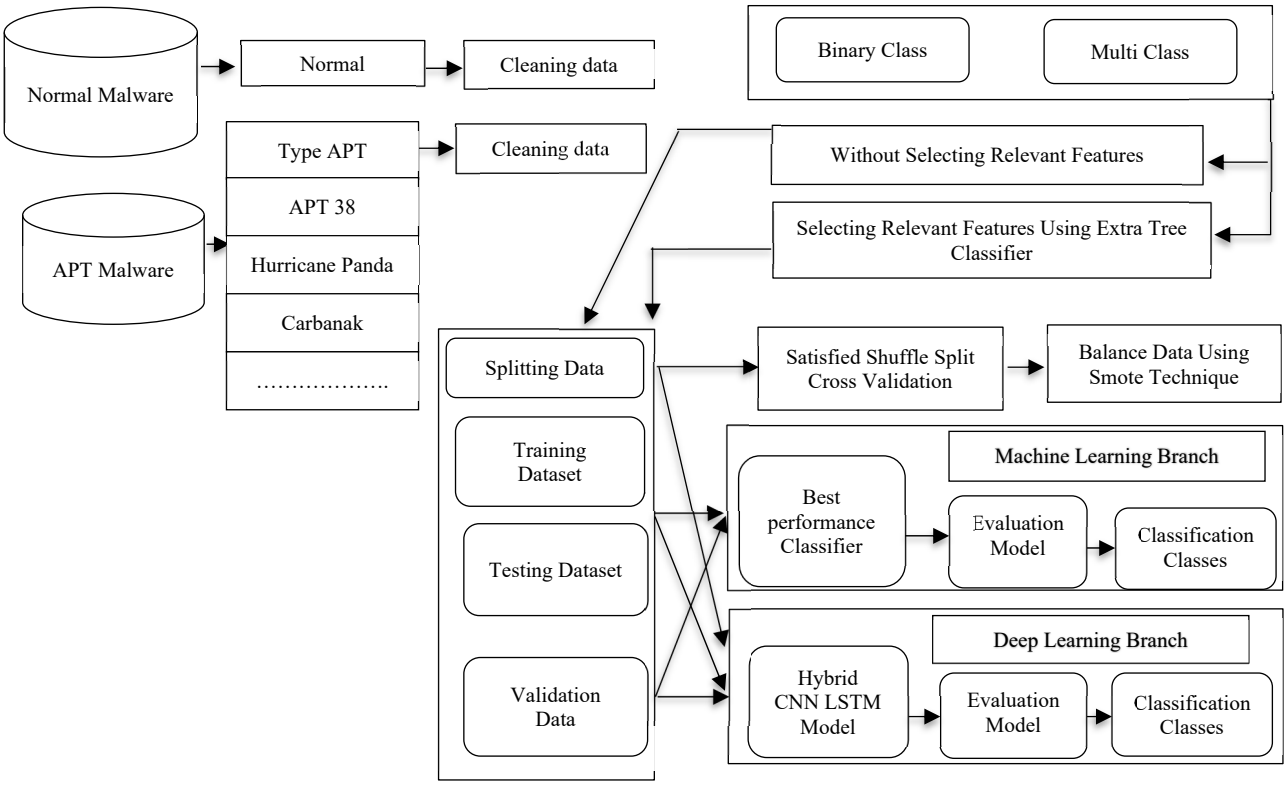

Figure 1. Block diagram of the proposed systems

4.1. Operation on Dataset

The system used APT malware data. Malware can be downloaded for free at https://github.com/cyber-research/APTmalware. APT and non-apt data. Malware apt rows and columns are (2086 and 4155) (9021 and 4154). An "Advanced Persistent Threat" (APT) malware dataset is used to train classifiers. Optional Header (30 features), MS-DOS Header (17 features), File Header (18 features), Obfuscated String Statistics (3 features), Mutex (7 features), Packer (64 features), Imported API (3917 features), and Buckets (98 features).and Methodology involves the pretreatment of input data. It involves analyzing and deleting APT malware dataset outliers to improve classification model performance. Exploratory data analysis is used to detect outliers. Unusual data are shown. Median, the 25th and 75th percentiles characterize data distribution. IQR is the difference between the lower and upper quartiles (Q3). First, the EDA technique performs statistical operations separately on each feature F in the database. When run on the dataset's 4155 columns (F) for APT Malware and 4154 columns (F) for regular Malware, it produces eight outputs: count, mean, standard deviation, minimum, maximum, and 25%, 75%, and 50% interquartile.

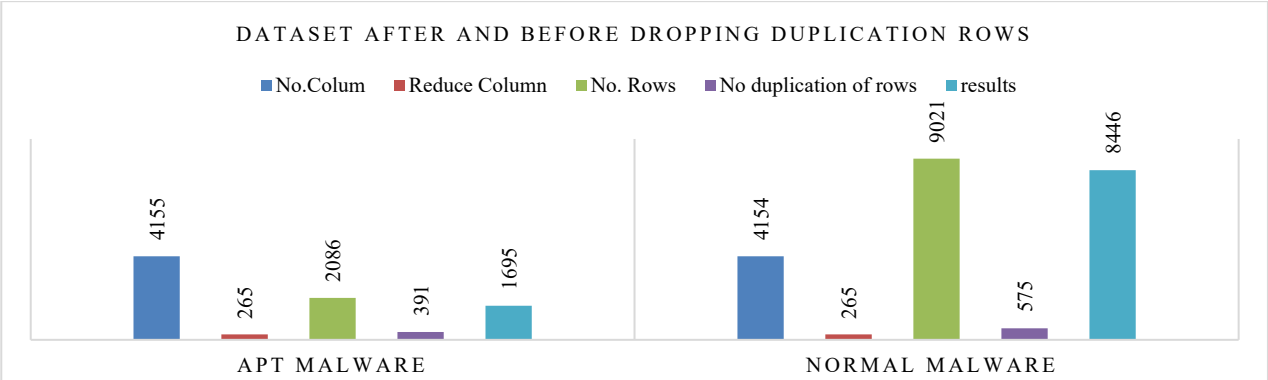

Figure 2. Finely result of the data

The mean value fills NULL values in the dataset with the feature's mean value. APT and Normal APT doesn't have null values. After removing outliers from the input dataset, this phase will yield a binary class for the Binary-Class Classifier subsystem and a multiple class for the Multiclass Classifier subsystem. To obtain correct data, duplicate and Check Missing Data have been eliminated. Figure 2 shows the finely result of the data.

4.1.1. Binary-Class Dataset

APT Malware comprises 11,107 samples from 16 malware types. The Binary Class Classifier will employ a binary class (M for APT malware attacks and N for Normal malware attacks) since APTs' behavior corresponds to the class APT malware attack. To automate the recognition and categorization of APT and Normal malware samples. It assumes that 15 APT organizations constitute a class of APT Malware. Figure 3 shows a normal malware attack.

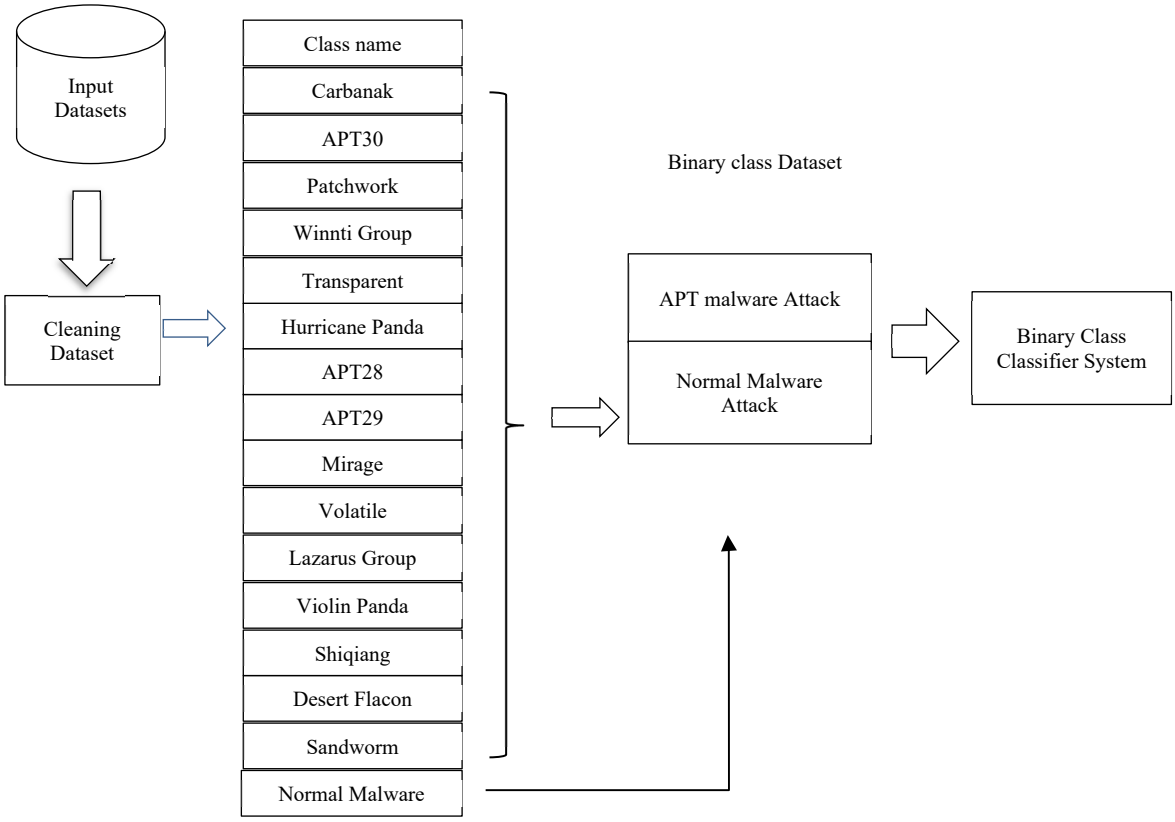

Figure 3. Constructs classes in the binary-class system

4.1.2. Multi-Classes Dataset

Multiclass Classifier handles APT dataset multiclassification. Some firm APTs share behavioral tendencies. A multiclass–classifier system identifies and categorizes samples from the same firm based on APT malware data behavior. Figure 4 demonstrates that every APT organization considers multiclass an independent class. This approach is based on APT malware dataset behavior data.

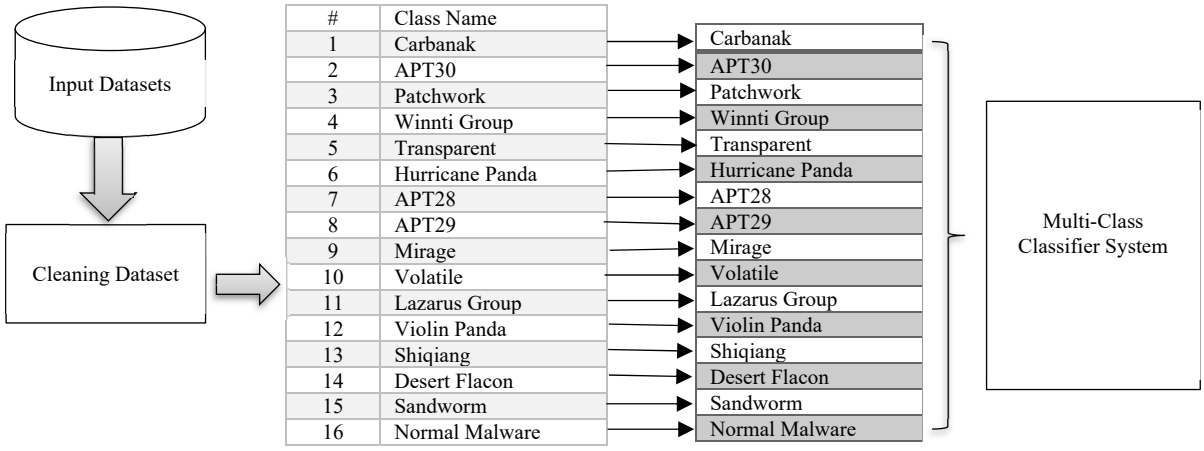

Figure 4. Constructs classes in the multi-class system

Feature selection chooses the categorization model's most important characteristics for classes. The features selection technique improves APT detection throughput and efficacy. This system uses Extra Trees Classifier to select features. The different tree classifier determines attribute importance. It's an ensemble strategy for choosing the best dataset features based on decision trees and random forests. The Gini index serves as the decision tree's attribute split criterion to calculate the significance of dataset features.

4.1.3. Balance Training Dataset Using SMOTE Technique

Over-sampling and under-sampling are frequent dataset parity strategies. Under-sampling is performed by decreasing the higher attribute to match the lower attribute. Oversampling is used to compare the values of lower and higher qualities. This study examined under- and over-sampling findings. SMOTE creates fake samples for underrepresented groups. This method fixes the dataset's unequal data distribution and eliminates random oversampling's overfitting. Minority classifications in the APT malware collection include violin panda (23 samples), Shiqiang (23 samples), volatile rice (34 samples), etc. The majority class has the most samples, such as 507. Thus, data distribution amongst classes will affect the classification model's performance.

5. CLASSIFICATION

Two classifier model scenarios in binary-class systems. Normal/non-APTMalware (N) and APTMalware (M) are two kinds. (N) is the normal/non-APT malware attack class in a Multiclass Classifier system, and (APT organization name) is the APT malware attack organization.

- Random Forest (ML): R.F. employs decision trees and bagging (sometimes referred to as Bootstrap aggregation) approaches. Bagging requires each decision tree to be trained on a subset of the dataset. Finally, each tree is classified based on the decision tree's outcome of a majority vote. The random forest comprises two crucial parameters: *n* estimators and training data. This forest contained 150 trees. Two phases are required to construct a random forest. The algorithm first selects k features from m. Utilizing the learned random forest approach and the test characteristics and rules of each randomly generated decision tree, make predictions. After recording the estimated result, compute votes for each expected objective. The final prediction of R.F. should receive the most votes.
- LightGBM (ML): Light Gradient Boosting Machine enhances model performance while requiring less memory. This method combines Gradient-based One-Sided Sampling with Exclusive Feature Bundling (EFB) to solve the drawbacks of the histogram-based approach standard to all GBDT designs. LightGBM detects and classifies APT Malware attacks.
- CNN-LSTM: analyze geographical and temporal data. This work uses deep hybrid learning. CNN with LSTM for APT analysis. CNN-LSTM cannot handle the class symbol; hence it must be translated to a number or code. In a binary–class classifier based on a machine learning algorithm, they can detect if an input test sample is a normal APT attack or APTMalware. In CNN-LSTM, it may be identified if the input test sample belongs to (0) regular APT attack or (1) APT malware attack. The same example applies to the multiclass Classifier based on the machine learning method, where each APT organization must have a unique code. To code classes in On-class and multiclass classifiers using training SMOTE data, first label data, then use One-hot encoding to encode classes, and then normalize data using the scaling equation in Figure 5.

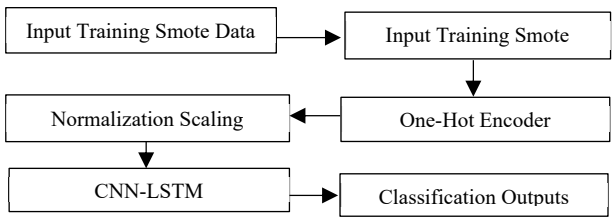

Figure 5. Block diagram of the coding classes in the training apt malware attack

Where data labelling for each class in the training APT dataset, the proposed system used the Python Label encoder library to convert labels into numeric and CNN-LSTM-readable forms. CNN-LSTM can better decide how to use labels.

All 15 APT organizations will assign code 1 (Figure 6). The standard attack class has a binary-class classifier code of 0. Each APT organization gets a multiclass classifier. Carbank=0, APT30=1, ..., and Normal malware=15 are examples of regular classes.

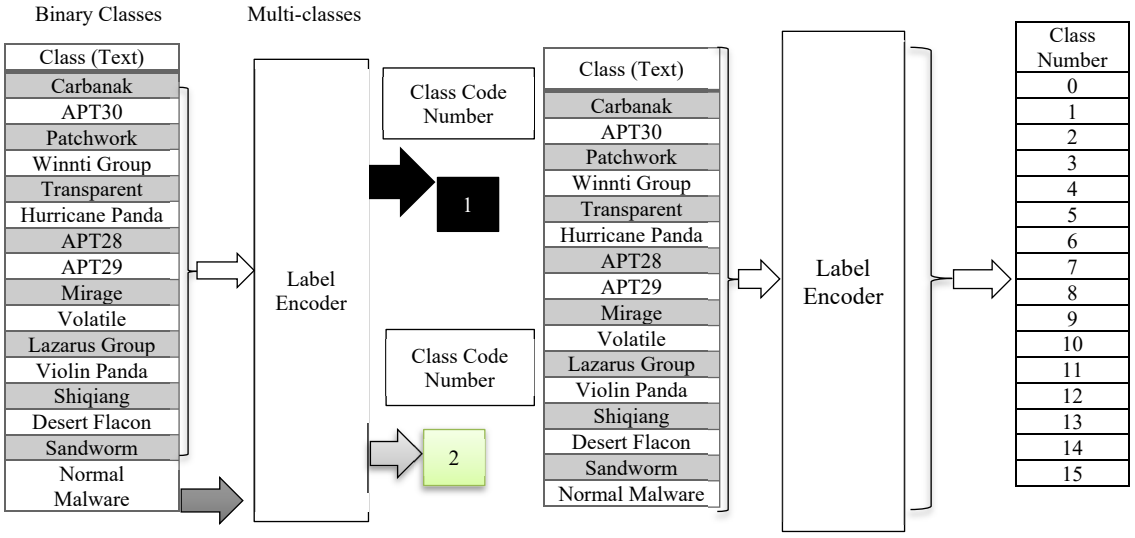

Figure 6. Label encode for binary-class and multi-class classification systems

The proposed method uses One-Hot Encoding to accommodate the CNN algorithm's SOFTMAX function. This method will only be utilized on a multiclass classifier with 16 classes, as opposed to a two-class classifier (0 for normal Malware and 1 for APT malware attack). This technique enhanced CNN precision. LSTM's One-Hot Encoder works on multiclass classifiers. Each dataset class is converted into a vertical and horizontal vector where the Table 2 is an example; it will take the first horizontal vector (carbanak) with a vertical vector (carbanak) and then apply the one-hot encoder condition, which is true or one of the horizontal vectors (carbanak) appears in the vertical vector (carbanak) else false or 0. Then combine the horizontal and vertical vectors (carbanak) (APT 30). The encoder value is 0 since the carbanak class does not present in the APT30 class, but it does in the horizontal vector (APT30). One-Hot Encoder creates a 16×16 array. These dimensions represent dataset classes. The diagonal line has one symbol since it connects horizontal and vertical classes of the same category.

When APT malware assault dataset properties are similar and close to a normal distribution, CNN-LSTM performs well. Standardization is a scaling procedure with a mean and unit standard deviation.

Table 2. Example of the One-Hot Encoding Technique

Label Encoder		One-Hot Encoding																
Class	Code	Class	Carbanak	APT30	Patchwork	Winnti Group	Transparent	Hurricane Panda	APT28	APT29	Mirage	Volatile	Lazarus Group	Violin Panda	Shiqiang	Desert Flacon	Sandworm	Normal Malware
Carbanak	0	Carbanak	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
APT30	1	APT30	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Patchwork	2	Patchwork	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Winnti Group	3	Winnti Group	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Transparent	4	Transparent	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Hurricane Panda	5	Hurricane Panda	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
APT28	6	APT28	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
APT29	7	APT29	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Mirage	8	Mirage	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Volatile	9	Volatile	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Lazarus Group	10	Lazarus Group	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Violin Panda	11	Violin Panda	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Shiqiang	12	Shiqiang	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Desert Flacon	13	Desert Flacon	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Sandworm	14	Sandworm	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Normal Malware	15	Normal Malware	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

The attribute's mean becomes zero, and its distribution has a unit standard deviation. Normalize APT data by using Equation (1). A scaling algorithm may convert APT data to 0-100 or 0-1.

$$s = \sum_{(i=0)}^{n} ((xi - \mu)^2) / (n-1) \quad (1)$$

where, μ is mean, $\sum xi$ is sum of data values, n is number of values in the sample dataset, and s is sample standard deviation.

A hybrid CNN/LSTM extracts spatial and temporal data from network traffic to improve intrusion detection. For practical model training, we recommend deep learning with category weights. This strategy reduces the influence of imbalanced samples during model training, enhancing training and prediction. Finally, we tested our network-labelling approach. CNN-LSTM uses two convolutional layers with one dimension and two functions (Rectified Linear Unit or ReLU and Max_pooling 1D). Its simplicity makes CNN-LSTM a cheap computing function, and the model trains quickly. A-Max pooling one dimension reduces the input's dimensionality to reduce overfitting. Convolutional layers feed LSTM. Using Activation Function [ReLU], batch-normalization to scale input data between [0,1], flatten to turn input data into a vector, and the F.C. layer to integrate and flatten all most profound convolutional layer characteristics. Softmax classifier categorizes input and flattens output for the output layer. Malware classification. Figure 7 shows the intended (CNN-LSTM).

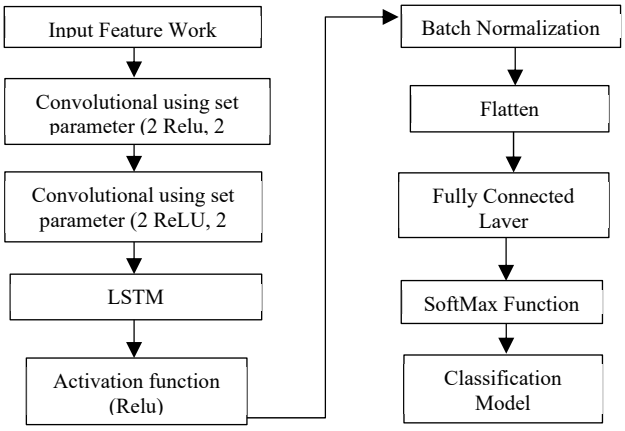

Figure 7. The proposed hybrid CNN-LSTM network algorithm

6. EVOLUTION OF THE RESULT

The recommended evaluation strategy will be implemented. A confusion matrix. Will be completed in error rate and provides a summary of the number of instances correctly or incorrectly predicted by a classification model. Counts calculated in a confusion matrix are typically referred to as follows [20]:

- TP: True Positive, TN: True Negative, FP: False Positive, and FN: False Negative. Based on the below-mentioned equations.

Tables 3, 4, 5, and 6 show the best result of an algorithm for two models. Tables 7 and 8 show how the system in CNN-LSTM work in layer and information of work.

Table 3. Values of random forest (binary)

The scale of the Binary class classifier system (random forest)	Value
TP	2503
TN	10822
FP	0
FN	4

Table 4. Values of LightGBM (binary)

The scale of the proposed multiclass classifier system (Light BGM)	Value
T.P	2501
TN	10824
FP	2
FN	2

Table 5. Values of random forest (multiple)

The scale of the Binary class classifier system (random forest)	Value
TP	2497
TN	10826
FP	0
FN	0

Table 6. Values of LightGBM (multiple)

The scale of the proposed multiclass classifier system (LightGBM)	Value
TP	2499
TN	10824
FP	4
FN	2

Table 7. Model: "Sequential" Binary classification

Layer (type)	Output Shape	Param #
conv1d (Conv1D)	(None, 264, 512)	1024
activation (Activation)	(None, 264, 512)	0
max_pooling1d	(None, 132, 512)	0
conv1d_1 (Conv1D	(None, 132, 256)	131328
activation_1 (Activation)	(None, 132, 256)	0
max_pooling1d (MaxPooling1D)	(None, 66, 256)	0
LSTM (LSTM)	(None, 512)	1574912
activation_2 (Activation)	(None, 512	0
batch normalization	(None, 512)	2048
flatten (Flatten)	(None, 512)	0
dropout (Dropout)	(None, 512)	0
dense (Dense)	(None, 2048)	1050624
dense_1 (Dense)	(None, 1024)	2098176
dense_2 (Dense)	(None, 2)	2050

Table 8. Model "Sequential" Multi classification

Layer (type)	Output Shape	Param #
conv1d (Conv1D)	(None, 264, 512)	1024
activation (Activation)	(None, 264, 512)	1024
max_pooling1d	(None, 264, 512)	0
conv1d_1 (Conv1D	MaxPooling1D (None, 132, 512)	0
activation_1 (Activation)	(None, 132, 256)	131328
max_pooling1d (MaxPooling1D)	(None, 132, 256	0
LSTM (LSTM)	(None, 66, 256)	0
activation_2 (Activation)	(None, 512)	1574912
batch normalization	(None, 512)	0
flatten (Flatten)	(None, 512)	2048
dropout (Dropout)	(None, 512)	0
dense (Dense)	(None, 512)	0
dense_1 (Dense)	(None, 2048)	1050624
dense_2 (Dense)	(None, 1024)	2098176

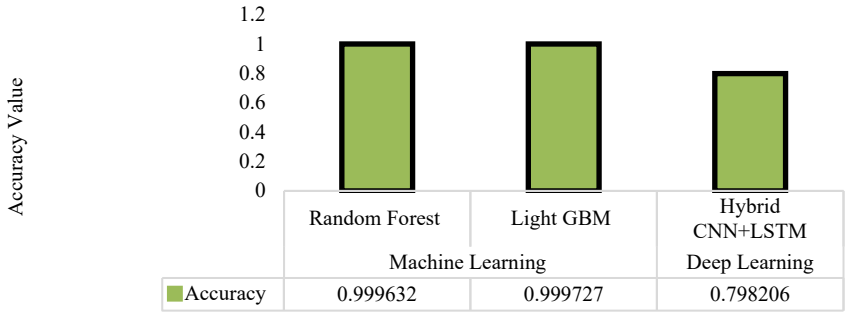

Figure 8. Compared to the performance of algorithms multiclass classification system

• Accuracy: The classification accuracy will be tested on the search data collection (Figure 8). It is assumed that each category of membership is described.

$$accuracy\ (\mathrm{AC}) = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

7. COMPARISON OF PREVIOUS RELATED STUDIES

Several studies have been concerned with APT malware attack classification using different methods and techniques adopted in previous years. Table 9 shows a comparison between proposed classification systems of APTs Malware and the same importin studies based on a set of qualitative metrics, such as the classification methods, dataset name used, and accuracy value. Figure 9 shows the result in a summary way.

8. CONCLUSION

Since APT attempts to steal sensitive information or conduct harmful network espionage, research on the identification and prevention of APT is urgently required. There are several alternative names for the APT. The great majority of these organizations are classified as terrorist groups. This paper proposes binary and multiclass classification based on machine learning and deep learning approaches to categorizing advanced persistent threats (APT) and conventional malware infections. Based on the study of the data set, they have developed a CNN-LSTM-based deep learning model that is both practical and adaptable for spotting Advanced persistent threats (APT) and conventional Malware.

Table 9. Comparison Between the Proposed Systems and Related Methods

Ref.	Classification Approach	Algorithms	Dataset Name	Accuracy
C.D. Xuan, et al., (2021) [13]	Machine learning	Random forest	CTU-13 data set	96.70%
S. Li, et al. (2021) [14]	Machine learning	Proposed SMOTE-RF	APT data set	80%
W. Han, et al., (2021) [15]	Machine learning	R.F., Decision Tree, KNN, and XGBoos	APT malware dataset	99.28%
G. Wang, et al., (2021) [16]	Machine learning	Method of detection based on belief rule (BRB)	APT and normal dataset	95.32%
C.D. Xuan and M.H. Dao, (2021) [17]	Deep learning	Hybrid CNN-LSTM	XSS global dataset	98%
F.J. Abdullayeva, (2021) [18]	Machine learning	Autoencoder and softmax regression algorithm	machine learning dataset	98.32%.
Our proposed (R.F.)	Machine and deep learning	Random forest, LightBGM and Hybrid CNN-LSTM	APT and Normal malware dataset	random forest multiclass Classifier The binary class classifier system
Our proposed (LightBGM)				LightBGM multiclass Classifier The binary class classifier system
Our proposed Hybrid CNN-LSTM				The binary class classifier system 91.4% multiclass classifier system 80%

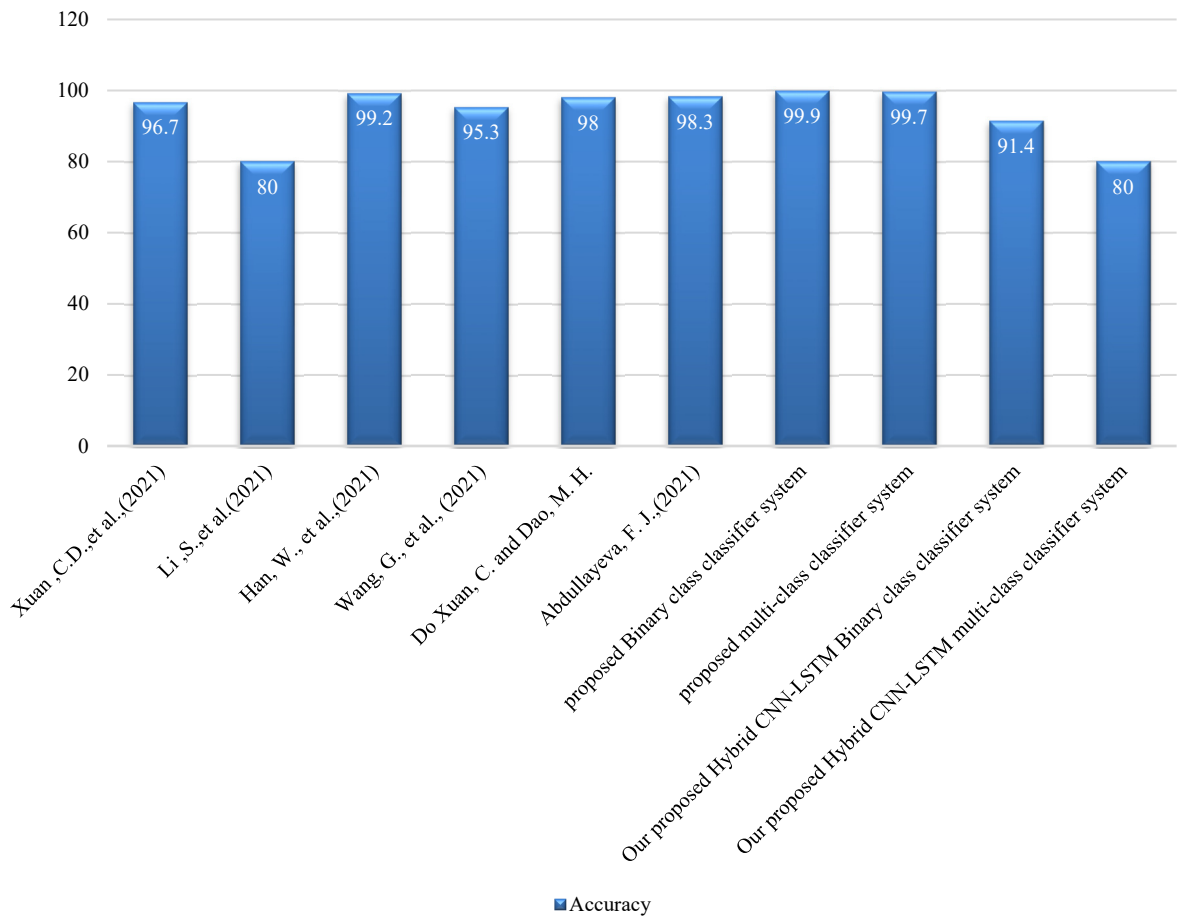

Figure 9. compatriot with privies studies

REFERENCES

[1] G. Laurenza, R. Lazzeretti, L. Mazzotti, "Malware Triage for Early Identification of Advanced Persistent Threat Activities", Digit. Threat. Res. Pract., Vol. 1, No. 3, pp. 1-17, September 2020.

[2] M. Alrehaili, A. Alshamrani, A. Eshmawi, "A Hybrid Deep Learning Approach for Advanced Persistent Threat Attack Detection", The 5th Int. Conference on Future Networks and Distributed Systems, pp. 78-86, 2021.

[3] I. Ghafir, V. Prenosil, "Advanced Persistent Threat Attack Detection: An Overview", International Journal of Advancements in Computer Networks and Its Security - IJCNS, Vol. 4, No. 4. pp. 50-54, 2014.

[4] P. Chen, L. Desmet, C. Huygens, "A Study on Advanced Persistent Threats", Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 8735 LNCS, pp. 63-72, 2014.

[5] A. Zimba, H. Chen, Z. Wang, M. Chishimba, "Modeling and Detection of the Multi-Stages of Advanced Persistent Threats attacks based on Semi-Supervised Learning and Complex Networks Characteristics", Futur. Gener. Comput. Syst., Vol. 106, pp. 501-517, 2020.

[6] S. Quintero Bonilla, A. Martin del Rey, "A New Proposal on the Advanced Persistent Threat: A Survey", Appl. Sci., Vol. 10, No. 11, p. 3874, 2020.

[7] Y. Haddi, A. Moumen, A. Kharchaf, "Study of a Mobile Robot's Obstacle Avoidance Behavior in a Radioactive Environment with a High Level of Autonomy", International Journal on Technical and Physical Problems on Engineering (IJTPE), Issue 50, Vol. 14, No. 1, pp. 34-41, March 2022.

[8] P.V.S. Charan, P.M. Anand, S.K. Shukla, "DMAPT: Study of Data Mining and Machine Learning Techniques in Advanced Persistent Threat Attribution and Detection", Data Mining-Concepts & Applications, IntechOpen, 2021.

[9] V. Jain, Y. Jain, H. Dhingra, D. Saini, M.C. Taplamacioglu, M. Saka, "A Systematic Literature Review on QR Code Detection and Pre-Processing", International Journal on Technical and Physical Problems on Engineering (IJTPE), Issue 46, Vol. 13, No. 1, pp. 111-119, March 2021.

[10] G. Yan, Q. Li, D. Guo, X. Meng, "Discovering Suspicious APT Behaviors by Analyzing DNS Activities", Sensors, Vol. 20, No. 3, p. 731, 2020.

[11] G. Ayoade, et al., "Evolving Advanced Persistent Threat Detection using Provenance Graph and Metric Learning", The IEEE Conference on Communications and Network Security (CNS), pp. 1-9, 2020

[12] Q. Wu, Q. Li, D. Guo, X. Meng, "Exploring the Vulnerability in the Inference Phase of Advanced Persistent Threats", Int. J. Distrib. Sens. Networks, Vol. 18, No. 3, p. 155013292210804, 2022.

[13] T. Jabar, M. Mahinderjit Singh, "Exploration of Mobile Device Behavior for Mitigating Advanced Persistent Threats (APT): A Systematic Literature Review and Conceptual Framework", Sensors, Vol. 22, No. 13, p. 4662, 2022.
[14] C. Do Xuan, D. Duong, H.X. Dau, "A multilayer Approach for Advanced Persistent Threat Detection using Machine Learning Based on Network Traffic", J. Intell. Fuzzy Syst., Vol. 40, No. 6, pp. 11311-11329, 2021.
[15] S. Li, Q. Zhang, X. Wu, W. Han, Z. Tian, "Attribution Classification Method of APT Malware in IoT Using Machine Learning Techniques", Secur. Commun. Networks, Vol. 2021, pp. 1-12, September 2021.
[16] W. Han, J. Xue, Y. Wang, F. Zhang, X. Gao, "APTMalInsight: Identify and Cognize APT Malware Based on System Call Information and Ontology Knowledge Framework", Inf. Sci. (Ny)., Vol. 546, pp. 633-664, February 2021.
[17] G. Wang, Y. Cui, J. Wang, L. Wu, G. Hu, "A Novel Method for Detecting Advanced Persistent Threat Attack Based on Belief Rule Base", Appl. Sci., Vol. 11, No. 21, p. 9899, October 2021.
[18] C.D. Xuan, M.H. Dao, "A Novel Approach for APT Attack Detection Based on Combined Deep Learning Model", Neural Comput. Appl., Vol. 33, No. 20, pp. 13251-13264, October 2021.
[19] F.J. Abdullayeva, "Advanced Persistent Threat attack Detection Method in Cloud Computing Based on Autoencoder and Softmax Regression Algorithm", Array, Vol. 10, p. 100067, July 2021.
[20] J. Xu, Y. Zhang, D. Miao, "Three-Way Confusion Matrix for Classification: A Measure Driven View", Inf. Sci. (Ny)., Vol. 507, pp. 772-794, January 2020

BIOGRAPHIES

Name: **Eman**
Middle Name: **Jalal**
Surname: **Khalefa**
Birthday: 12.10.1984
Birth Place: Baghdad, Iraq
Bachelor: Computer Science, University of Diyala, Baqubah, Iraq, 2006
Master: Student, Computer Science, University of Diyala, Baqubah, Iraq, 2020
Research Interests: Deep Learning Technologies, Branches of Artificial Intelligence
The Last Scientific Position: Senior Programmer, Department Accounting, Ministry of Finance, Baghdad, Iraq, Since 2020

Name: **Dhahir**
Middle Name: **A. Abdullah**
Surname: **Salman**
Birthday: 26.01.1961
Birth Place: Iraq
Bachelor: Electrical and Electronics Engineering, Al-Rasheed College of Science and Engineering, University of Technology, Baghdad, Iraq, 1982
Master: Research Decision Making Operations, Al-Rasheed College of Science and Engineering, University of Technology, Baghdad, Iraq, 1989
Doctorate: Research Simulations and Artificial Intelligence, Al-Rasheed College of Science and Engineering, University of Technology, Baghdad, Iraq, 2002
The Last Scientific Position: Prof., Department of Computer Science, College of Science, University of Diyala, Baqubah, Iraq
Research Interests: Deep Learning Technologies, Branches of Artificial Intelligence