Hybrid Machine Learning Algorithm for Enhanced BGP Anomaly Detection

Nassir S. Kadhim[1,2], Nor Fadzilah Abdullah[1] and Kalaivani Chellappan[1]
p117359@siswa.ukm.edu.my fadzilah.abdullah@ukm.edu.my kckalai@ukm.edu.my nassal2005@gmail.com
[1]Department of Electrical, Electronic & Systems Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia (UKM), 43600 Bangi, Malaysia.
[2]Ministry of Communication (MOC), Iraqi Telecommunication and Post Company (ITPC), Baghdad, Iraq.

Abstract
Border Gateway Protocol (BGP) is a critical component of the Internet's infrastructure, responsible for inter-domain routing. It enables Internet Service Providers (ISPs) to manage the flow of data across the global network by announcing address prefixes and implementing routing policies. Despite its importance, BGP faces several challenges, including configuration errors and security vulnerabilities. This creates a regional or global internet service interruption. Nevertheless, the ability to detect abnormal messages transmitted via BGP enables the timely detection of such attacks. Machine learning (ML) has recently grown crucial in improving the effectiveness, efficiency, and scalability of BGP anomaly detection systems. This study evaluates the ML models for detection and identifying BGP anomalies. We applied a statistical analysis to the 24 BGP features extracted from a realistic network topology based on simulation. Three feature sets were categorized based on their significance in classifying anomalies and their potential for predicting cyberattacks. A comprehensive assessment of the performance of eight ML algorithms in detecting BGP anomalies utilizing multiple features and dataset structures has been conducted. The assessment findings revealed that the ML models exhibit consistent results with the tested dataset that containing a number of significant features data in terms of performance metrics and demonstrated that the combined dataset structure produced better results than the individual datasets. To enhance the BGP anomaly detection model and get the best results, we proposed a hybrid SGD-RF ML model, which achieved the highest accuracy by 99.3%, as well as improvement with an AUC value of 0.993 and other performance metrics as compared to the individual models.

Keywords:
Border Gateway Protocol (BGP), machine learning algorithms, Anomaly detection, cyberattacks, feature sets.

1. Introduction

Border Gateway Protocol (BGP) is an exterior gateway protocol designed to exchange reachability and routing information across Internet routers over the global network backbone [1]. BGP operates by managing a routing information table to identify the optimal path for delivering data from the source to the destination. With the rapid growth of Internet traffic, BGP has become a critical component of network infrastructure management. Due to its reliance on trust between Autonomous Systems (ASs), BGP is inherently vulnerable to various security threats that can compromise the safety and stability of the Internet. These vulnerabilities can lead to network attacks, including hijacking, denial of service (DoS), and network outages, compromising network efficiency and performance [2]. Therefore, early detection and mitigation of network anomalies are essential for maintaining Internet stability and security. Various methodologies have been proposed for detecting network anomalies, including statistical methods, data mining techniques, hegemony analysis and machine learning approaches [3-7].

Anomaly detection points to the challenge of detecting trends in data that deviate from expected behaviors. In various implementation domains, these non-conforming patterns are classified as deviations and outliers [8]. Machine learning (ML), with its inherent ability to learn patterns and make predictions, has emerged as a promising approach for BGP anomaly detection [9]. Recent BGP anomaly detection systems employ ML techniques to mine network data. ML detects anomalies by representing them as a set of features that are selected based on predicted system behavior [10]. The features selection process is used to decrease the dimensionality of a design matrix. This reduces computational complexity and memory utilization. Feature selection also helps prevent overfitting by reducing duplicated data, enhancing modeling accuracy, and decreasing training time [11]. The ML model is expected to properly categorize additional events if they exhibit similar behavior during the training phase. However, the BGP network's dynamic and complex structure poses a challenge for a single technique and classifier to detect anomalies reliably and effectively [12].

In this paper, we evaluated the performance of ML algorithms for detecting BGP anomalies using our

Manuscript received November 5, 2024
Manuscript revised November 20, 2024
https://doi.org/10.22937/IJCSNS.2024.24.11.1

simulation datasets for realistic BGP network topology. It identified 22 significant features through statistical analysis at three different levels of significance. We employed eight individual ML algorithms and assessed the performance based on accuracy, recall, precision, F1-score and ROC-AUC (receiver operating characteristic's area under curve) metrics. The algorithms implemented are Support Vector Machine (SVM), Logistic Regression (LR), Naïve-Bayes (NB), Decision Trees (DT), Quadratic Discriminant Analysis (QDA), k-nearest Neighbors (KNN), Random Forests (RF), and Stochastic Gradient Descent (SGD).

In addition to the individual evaluation of the algorithms, a hybrid approach has been developed to enhance the performance metrics by combining two classification algorithms, which are the SGD and RF. Hybrid ML approaches have proven to be useful in many situations, especially when dealing with complicated and unexpected data [13]. A hybrid ML model clusters the predictions of many separate classifiers to make a final prediction. Multiple methods may be used to combine the results of individual classifiers, such as voting, averaging, and weighting [14]. The voting method aggregates predictions from multiple classifiers, selecting the most common prediction as the final output. The averaging method produces a final prediction based on predictions from different models. Meanwhile, weighting assigns different weights to classifiers based on their performance for the final prediction. The proposed hybrid model adopted the voting method, which is effective in scenarios where classifiers have similar performance levels, as seen in network security applications.

The objective of this work is to enhance the development of sophisticated BGP anomaly detection systems through a comprehensive evaluation of ML algorithms and advancing a hybrid model. The primary contribution of our research is to select the best classifiers from the most popular models that accomplish high performance for BGP anomaly detection. By clustering the advantages of a hybrid model, the highest detection accuracies were achieved.

The outcomes of this research highlight the importance of selecting significant BGP features that reflect insights into the logical and physical BGP networks topology for developing effective detection models applicable in real-world networks. The data sets utilized in our work were created by extracting and selecting the most significant features under the BGP cyber-attack scenarios based on simulations of the actual network topology for the IRAQI Internet Gateways (IGW). This paper presents a comparative analysis and assessment of classification ML algorithms, highlighting the best performance models for BGP detecting anomalies. The importance of utilizing ML lies in improving the accuracy of BGP anomaly detection systems and relying less on human expertise in identifying unexpected traffic patterns, (i.e. automation). Consequently, the BGP infrastructure cybersecurity and data protection can evolve into proactive rather than reactive approaches.

The rest of the paper is organized as follows. Section 2 presents the related work and background, followed by the methodology in Section 3. The experimental results and discussions are presented in Section 4. Finally, Section 5 concludes the paper.

2. Background and Related Work

The exponential growth of internet traffic and the increasing complexity of network infrastructures have led to the vulnerability of BGP routing systems to performance degradation, security breaches, and malicious attacks. BGP plays a crucial role in facilitating communication between autonomous systems in the internet, making it a prime target for various security threats. BGP anomalies, such as hijacks, route leaks, and DOS, can disrupt normal traffic patterns and pose significant security risks. ML-based approaches have shown great potential in effectively detecting BGP anomalies, enabling timely response and mitigation [15].

The adoption of ML-based anomaly detection in BGP routing offers several advantages for network security enhancement. First, ML models can adapt to changing network conditions and identify new types of anomalies, providing proactive protection against emerging threats. Second, ML approaches can reduce false positives and negatives, improving the accuracy of anomaly detection and reducing the burden on network administrators. Third, ML systems can continuously learn and update their knowledge, ensuring the detection models remain effective over time [16].

Table 1. A summary of BGP anomalies detection based on various machine learning techniques.

Author-Year	Type of anomaly	ML algorithm
Hoarau et al. 2022 [9]	Forged AS path	LSTM-RNN
Verma et al. 2023[12]	Indirect anomaly	KNN, NB, and ELM
Sanchez et al. 2019 [18]	RTL	NB, DT, SVM, RF & MLP
Bhatnagar et al. 2019 [21]	Indirect, hardware failures, prefix hijacking	DT, RF, AdaBoost & Gradient Boosting
Kalra et al. 2021[22]	Direct unintended anomaly, system	SVM & MLP
Thales et al. 2021 [23]	Indirect, direct or link failure of BGP anomalies	LSTM
Al-Rousan & Trajkovi´c 2012 [24]	Indirect anomaly	SVM & HMMs
Lutu et al. 2014 [25]	Direct unintended anomaly	Winnowing
Allahdadi et al. 2017 [26]	DDoS attacks, power outage	SVM
Cheng et al. 2018 [27]	DDoS attacks	Multi-scale, LSTM, SVM, NB, AdaBoost
Dai et al. 2019 [28]	DDoS attacks	SVM
Cosovic et al. 2017 [28]	RTL, DDoS attacks, power outage	ANN
Li et al. 2020 [29]	DDoS attacks, WannaCrypt, power outage	LSTM
Hoarau et al. 2021 [30]	Google leak	RNN
Park et al. 2023 [31]	Indirect Anomaly	AE, One-class SVM, RF, CNN-LSTM

Table 1 presents a summary of the research articles identified in the related works. The utilized ML algorithms include Decision Tree (DT), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest (RF), Extreme Learning Machine (ELM), AdaBoost, Gradient Boosting, Naive Bayes (NB), Multi-Layer Perception (MLP), Long Short-Term Memory (LSTM), Artificial Neural Network (ANN), Convolutional Neural Network (CNN) Recurrent Neural Network (RNN), Hidden Markov Models (HMM), Autoencoder (AE) and Winnowing.

3. Methodology

The BGP serves as the backbone for routing information exchange across the internet. However, anomalies within BGP routing, stemming from misconfigurations, hardware failures, or malicious activities like DOS, outage and route hijacking, pose significant threats to network stability, security, and reliability [26]. Given the critical importance of BGP routing, there is a pressing need to develop effective anomaly detection mechanisms to identify deviations from expected routing behaviors [25]. This study offers a comparative analysis and assessment of classification ML algorithms for BGP detecting anomalies. BGP anomalies are detected by employing binary classification techniques and distinguishing between normal and abnormal classes. Based on the performance evaluation of eight ML algorithms, we proposed an effective and accurate hybrid model clustered the strengths of two tested algorithms, capable of promptly detecting anomaly BGP behaviors and mitigating potential network disruptions produced by cyber-attacks.

Our research process is segmented into six primary stages, as seen in Figure 1. Stage 1 involved BGP data collection, which consists of extracting and verifying 24 BGP features, and visualizing features most affected under various cyber-attack scenarios. In Stage 2, own datasets were created based on extracted BGP features [38]. During Stage 3, data preprocessing on the 24 feature datasets was conducted, and statistical analysis using the ANOVA test and p-value to determine significant features was performed in Stage 4. In Stage 5, eight ML algorithms were evaluated based on performance metrics such as accuracy, recall, precision, and F1-score. Finally, in stage six, further assessments were performed and a hybrid model combining two algorithms, SGD and RF was proposed. The proposed hybrid ML model improved the detection accuracy and other performance metrics. Therefore, this model has proven to be very effective in detecting anomalies in the BGP network.

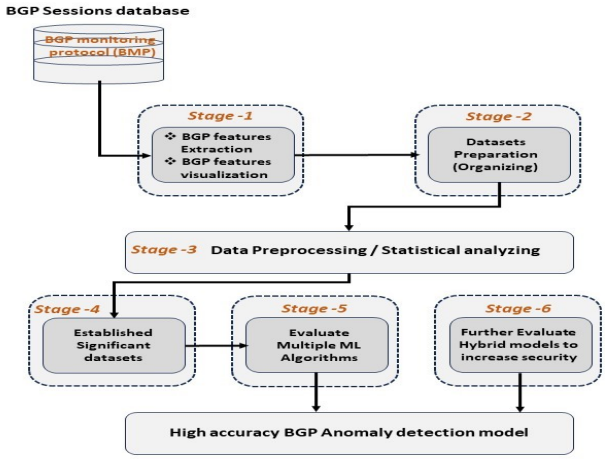


Fig. 1. Research process stages

3.1 Feature Selection

The BGP traffic behavior is grounded on realistic network data and extensive network testing conducted through simulation rather than depending on historical BGP security incidents and a publicly available data set. BGP traffic is analyzed in the context of three BGP cyberattack scenarios: hijacking, DOS, and outages. BGP features were extracted from the simulator's BGP database using PostgreSQL scripts to extract 14 features from the BGP AS-PATH attribute and ten (10) features from the number of BGP UPDATE messages (volume), and then 24 datasets were created accordingly. To assess the significance of features in a binary classification context and predict a binary outcome (Normal, Anomaly) based on extracted features, ANOVA statistical analysis is adopted to generate the F-Statistic and the p-values and identify the most impactful features by the cyber-attacks.

Based on the findings of statistical analysis p-value (p=0.005), it has been observed that six (6) of the BGP features are most significant by BGP cyber-attack effect with (F≥90), known henceforth as features Set A, as shown in Table 2. Next, nine (9) features with moderate significance (30< F<90) were clustered as feature Set B, as shown in Table 3. Finally, seven (7) features Set C with (F≤30), with the least significance were clustered in Table 4. Note that two (2) of the 24 extracted features are unaffected by cyber-attacks, with (p>0.005), and thus removed from further analysis.

Table 2. BGP's most significant features (Set A)

Feature 1	Average/ Maximum edit distance
Feature 2	Prefix origin change
Feature 3	Number of ORIGIN changes
Feature 4	Number of implicit withdrawals with same/different path
Feature 5	Maximum unique AS path length
Feature 6	Number of rare ASs

Table 3. BGP's moderate significant features (Set B)

Feature 1	Number of plain new announcements
Feature 2	Number of new paths announced after withdrawing an old path
Feature 3	Maximum/average announcements per AS
Feature 4	Announcements to longer paths
Feature 5	Number of announcements/withdrawals
Feature 6	Maximum/average announcements per prefix
Feature 7	Number of IGP/EGP/ INCOMPLETE messages
Feature 8	Number of announced prefixes
Feature 9	Average number of rare ASs

Table 4. BGP's least significant features (Set C)

Feature 1	Maximum number of rare ASs
Feature 2	Announcements to Shorter paths
Feature 3	Average AS Path Length
Feature 4	Average unique AS path length
Feature 5	Number of new-path announcements
Feature 6	AS-Path changes according to geographic location
Feature 7	Maximum AS Path Length

3.2 Dataset Modelling

The dataset modelling of this study builds on the statistical analysis outcome of 22 data sets classified into three main sets (A, B, C). Furthermore, by combining two of the main datasets, three more combinations of feature sets and an additional fourth set containing all three main feature sets were created, as shown in Table 5. Thus, seven (7) datasets is generated to be tested with the selected ML models.

Table 5. Summary of BGP feature combinations for simulation

Seq.	Combining sets	No. of significant features
1	Set (A) + (B)	15
2	Set (A) + (C)	13
3	Set (B) + (C)	16
4	Set (A) + (B) + (C)	22

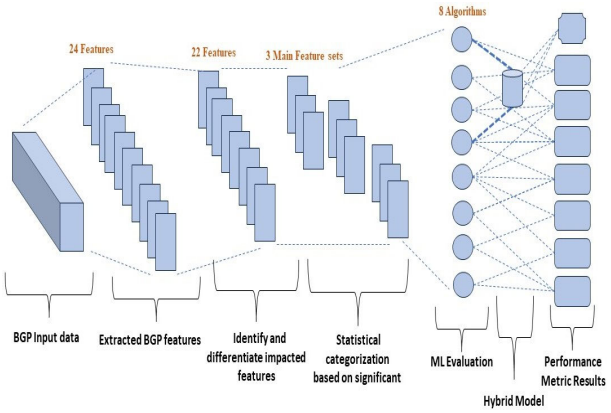


Fig. 2. Data processing and modeling

The goal of utilizing multi-feature sets to produce the datasets is to expand the input space of ML models, enabling them to learn more effectively from the data and produce accurate predictions. Figure 2 illustrates the data processing and modeling steps to provide a better understanding of the model's performance.

3.3 ML Binary classification algorithms

An evaluation was conducted to assess the effectiveness of eight ML techniques in detecting BGP anomaly behaviors and identifying potential cyber-attacks such as Hijacking, DOS, and Outage. The study focuses on binary classification and utilizes datasets containing the most important BGP features.

Table 6. Pros and Cons of evaluated ML classification algorithms

ML Algorithm	Pros	Cons
SVM (Support Vector Machine)	Effective in high-dimensional spaces	Can be sensitive to kernel choice; Computationally expensive for large datasets
LR (Logistic Regression)	Simple and interpretable, works well with linearly separable data	The performance is poor when the feature space is large
NV (Naïve Bayes)	Fast and simple- Performs well on high-dimensional data	Assumes independence between features
DT (Decision Tree)	Easy to understand and interpret- Can model complex relationships	Prone to overfitting, especially with deep trees
QDA (Quadratic Discriminant Analysis)	Preserves class separability	Assumes normal distribution of features
KNN (K-Nearest Neighbors)	Simple, no training phase, Works well with local patterns	Computationally expensive for large datasets, Sensitive to irrelevant features
RF (Random Forests)	Reduces overfitting through ensemble - Handles high-dimensional data	Complexity increases with the number of trees
SDG (Stochastic Gradient Descent)	Efficient and scalable for large datasets	Requires careful tuning of hyperparameters, Sensitive to feature scaling

Table 6 illustrates the strengths and weaknesses of eight ML classification algorithms that are evaluated. Despite ML algorithms shared aims, each technique attempts to classify data in different circumstances. K-nearest neighbors (KNN) [27], for instance, is a non-parametric method used to categorize data. KNN is a form of instance-based learning algorithm. This classification approach utilizes a distance function to calculate the estimated distances between objects. It then assigns all unlabeled items to the most common category among their K-nearest neighbors. The value of K is always a positive integer. Naive Bayes is a classifier that utilizes Bayes' theorem of prior probability to classify data examples into a certain class. Logistic Regression (LR) [28] is a classifier that utilizes a logistic regression function, which is also known as a sigmoid function. The decision tree (DT) approach [11] categorizes data into branch segments consisting of a root node, internal nodes, and leaf nodes. Random Forests (RF) [29] classification is a technique that effectively tackles the problem of over-fitting in decision trees by aggregating many decision trees.

3.4 Hybrid ML model

Hybrid machine-learning models combine the characteristics of multiple models to create an effective tool for detecting BGP anomalies. The ML algorithm combinations provide better accuracy, robustness, scalability, and resilience, making them optimal for real-world network monitoring and security protection.

To improve the model detection efficiency, a hybrid SGD-RF ML model has several advantages as follows [30]:

(i) **Improved Accuracy and Performance**: By leveraging the strengths of both algorithms, the hybrid model achieved better overall accuracy compared to using each algorithm individually. The combination allowed the model to address the limitations of one algorithm with the strengths of the other.

(ii) **Flexibility in Hyperparameter Tuning**: The hybrid model provides more flexibility in hyperparameter tuning. Experimentation of different hyperparameters for both SGD and RF is conducted to find the optimal configuration and improve the performance of the model.

(iii) **Feature Importance Analysis**: RF is known for its ability to assess feature importance. By combining with SGD, the hybrid model offered a more comprehensive analysis of feature importance.

(iv) **Complementary Strengths**: SGD and RF have distinct qualities that complement one another. SGD is effective in training on huge datasets, while RF is robust to overfitting. Therefore, the combination performs well with high-dimensional data.

3.5 Performance Metrics

Binary classification is the process of identifying a single class, specifically an anomaly. A classifier categories situation as either anomalies or regular (non-anomalies). Table 7 confusion matrix demonstrates the classifier's decision [31].

Table 7. ML Confusion Matrix

		Predicted Class	
		Anomaly	Regular
Actual Class	Anomaly (True)	TP	FN
	Regular (False)	FP	TN

The factors of the classification are termed as: TP (true positive), which is the number of correctly classified attacks; TN (true negative), the number of normal flows correctly classified; FP (false positive), the number of normal instances misclassified as attacks; and FN (false negative), the number of attack instances misclassified as normal [32]. Based on these elements, the performance metrics can be defined as the following [33]:

(i) **Accuracy**: measures the classification model's ability to correctly categorize data examples, regardless of whether they are positive or negative. It is determined as the ratio of the total number of correctly categorized records in a dataset to the total number of rows in the dataset:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

(ii) **Precision**: a metric that quantifies the correctness of positive predictions provided by a model. The calculation involves determining the proportion of accurate positive predictions produced by the model, relative to the total number of true positives:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

(iii) **Recall**: a metric that quantifies the proportion of positive instances correctly detected by the model. It is computed by dividing the number of true positive predictions by the total number of real positive occurrences in the dataset:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

(iv) **F1-score**: commonly referred to as the F-score or F-measure, is a ML evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The formulation of the harmonic mean of precision and recall is given by:

$$\text{F1-score} = 2 \times \left(\frac{Precision \times Recall}{Precision + Recall}\right) \quad (4)$$

The effectiveness of a classification model relies on its capacity to correctly predict classes [34]. Anomaly detection models were trained on a limited data set and evaluated on a separate dataset (train/test split). An assessment of implemented models was conducted to verify their capacity to achieve high levels of detection accuracy. The evaluation of all models was evaluated using five performance metrics: accuracy, precision, True Positive Rate (TPR), also known as Recall, F1-Score and ROC-AUC.

4. Results and Discussions

The ML methods used in this work were chosen based on their ability to detect anomalies, with a focus on models designed for binary classification. These algorithms were tested to analyses their performance and suitability for tasks requiring a clear distinction between two classes. The datasets were divided into two parts, where a significant amount of the data was chosen as a training dataset, while the remaining data was designated as the test dataset to assess the developed ML models [31].

The datasets are evaluated using binary classification methods, where each label represented one of two possible outcomes: normal or anomaly. This comprehensive evaluation was specifically conducted to identify and analysis three distinct types of BGP attacks: hijacking, denial of service (DoS), and outage. By focusing on these attack types, the

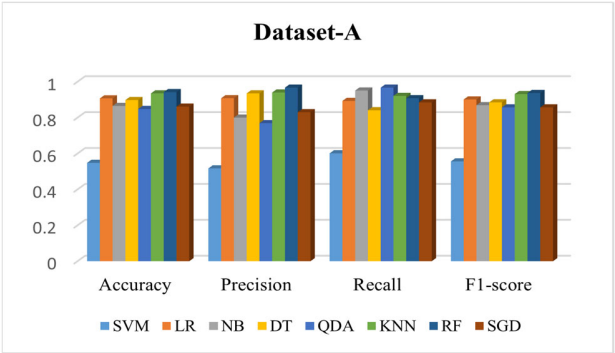
study aimed to assess the effectiveness of binary classification in accurately distinguishing between normal operations and various anomaly activities. The results showed that the KNN, RF, SGD, and LR algorithms outperformed the QDA, NB, and DT algorithms. Notably, the SVM algorithm exhibited subpar performance across all statistical feature sets.

For further explanation, the accuracy values of the Random Forest (RF) algorithm across the three main datasets (A, B, and C) were 94%, 88%, and 75%, respectively. In contrast, the SVM algorithm yielded unsatisfactory results, with accuracy values of only 54%, 59%, and 52% for the same feature sets. These results demonstrate the wide range of performance levels among the tested algorithms. They show that KNN, RF, SGD, and LR are all very good at binary classification tasks for tested datasets, while SVM performs poorly.

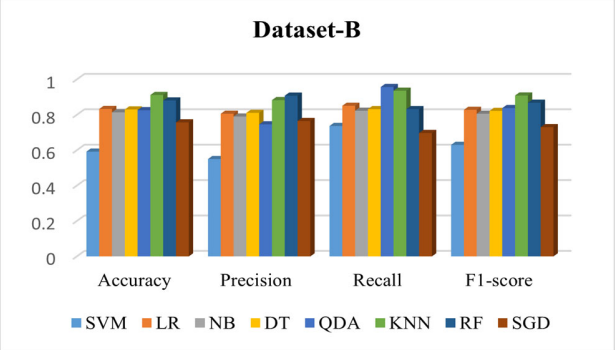
Table 8 displays the accuracy, precision, recall, and F-score metrics for all algorithms for the main datasets (A, B, and C). The results clearly show that the obtained precision and accuracy are proportional to the significance level of the feature data sets that are used for evaluation. It also becomes evident that both the KNN and RF accuracy results are better than the other methods for all the tested datasets. This indicates that the KNN and RF classifiers can individually detect BGP attacks effectively. A comparison of pattern detection results in terms of performance metrics was applied, as shown in Figure 3, where a slight disparity is clearly visible among high-performance algorithms and high contrast with low-performance algorithms like SVM.

Table 8. Performance results for three main datasets (A, B, C).

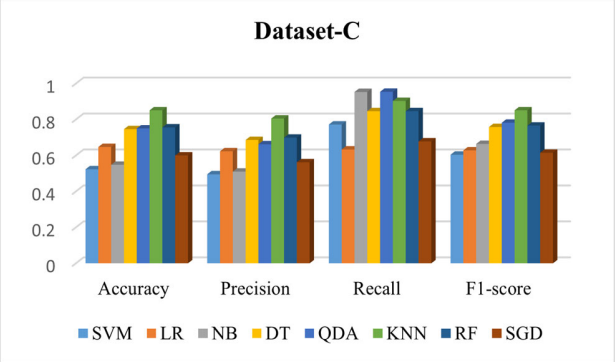
Dataset A				
ML	Accuracy	Precision	Recall	F1-score
SVM	0.548	0.517	0.601	0.556
LR	0.907	0.908	0.893	0.901
NB	0.865	0.800	0.951	0.869
DT	0.898	0.935	0.841	0.885
QDA	0.848	0.769	0.967	0.857
KNN	0.935	0.940	0.921	0.931
RF	0.942	0.967	0.908	0.937
SGD	0.861	0.830	0.885	0.857
Dataset B				
SVM	0.592	0.550	0.737	0.630
LR	0.833	0.806	0.851	0.828
NB	0.814	0.790	0.823	0.806
DT	0.830	0.811	0.832	0.822
QDA	0.826	0.746	0.957	0.838
KNN	0.912	0.883	0.936	0.909
RF	0.881	0.908	0.832	0.868
SGD	0.757	0.765	0.697	0.730
Dataset C				
SVM	0.523	0.495	0.772	0.604
LR	0.647	0.623	0.633	0.628
NB	0.548	0.510	0.953	0.664
DT	0.746	0.686	0.846	0.758
QDA	0.750	0.662	0.954	0.782
KNN	0.851	0.805	0.903	0.851
RF	0.756	0.699	0.846	0.7658
SGD	0.600	0.562	0.678	0.615



(a)



(b)



(c)

Fig. 3. Performance comparisons for three main feature sets

In addition to the comparison of the main feature sets, Table 9 examined the additional combination dataset performance. When comparing the combined feature sets' performance to the individual dataset results for most evaluated ML algorithms,

improvements were particularly found in the accuracy metric. In particular, the (A+B+C) dataset, which included all the data on significant BGP features, demonstrated the highest improvements in performance metrics.

Table 9. Performance results for combination datasets.

Dataset (A+B)				
ML	Accuracy	Precision	Recall	F1-score
SVM	0.585	0.550	0.654	0.597
LR	0.933	0.915	0.945	0.930
NB	0.887	0.832	0.952	0.888
DT	0.826	0.758	0.925	0.833
QDA	0.924	0.870	0.985	0.924
KNN	0.926	0.909	0.936	0.922
RF	0.945	0.979	0.902	0.939
SGD	0.881	0.906	0.834	0.868
Dataset (A+C)				
SVM	0.595	0.573	0.545	0.559
LR	0.912	0.914	0.898	0.906
NB	0.799	0.712	0.960	0.818
DT	0.938	0.899	0.978	0.937
QDA	0.801	0.772	0.819	0.795
KNN	0.934	0.946	0.911	0.928
RF	0.946	0.974	0.909	0.940
SGD	0.901	0.905	0.882	0.893
Dataset (B+C)				
SVM	0.568	0.542	0.525	0.533
LR	0.872	0.845	0.891	0.867
NB	0.696	0.615	0.943	0.745
DT	0.885	0.917	0.832	0.872
QDA	0.889	0.832	0.957	0.890
KNN	0.919	0.893	0.941	0.916
RF	0.842	0.833	0.832	0.832
SGD	0.826	0.797	0.846	0.821
Dataset (A+B+C)				
SVM	0.588	0.541	0.821	0.652
LR	0.938	0.928	0.942	0.935
NB	0.847	0.775	0.951	0.854
DT	0.909	0.919	0.886	0.902
QDA	0.862	0.940	0.756	0.838
KNN	0.923	0.898	0.943	0.920
RF	0.946	0.981	0.902	0.940
SGD	0.930	0.905	0.933	0.919

Figure 4 presents a comparative of the performance metric patterns for different combinations datasets (A + B), (A + C), (B + C), and (A + B + C). This figure provides a detailed visual representation and highlights differences between the metrics when considered in different groupings, offering insight into which combinations yield the most effective or optimal performance.

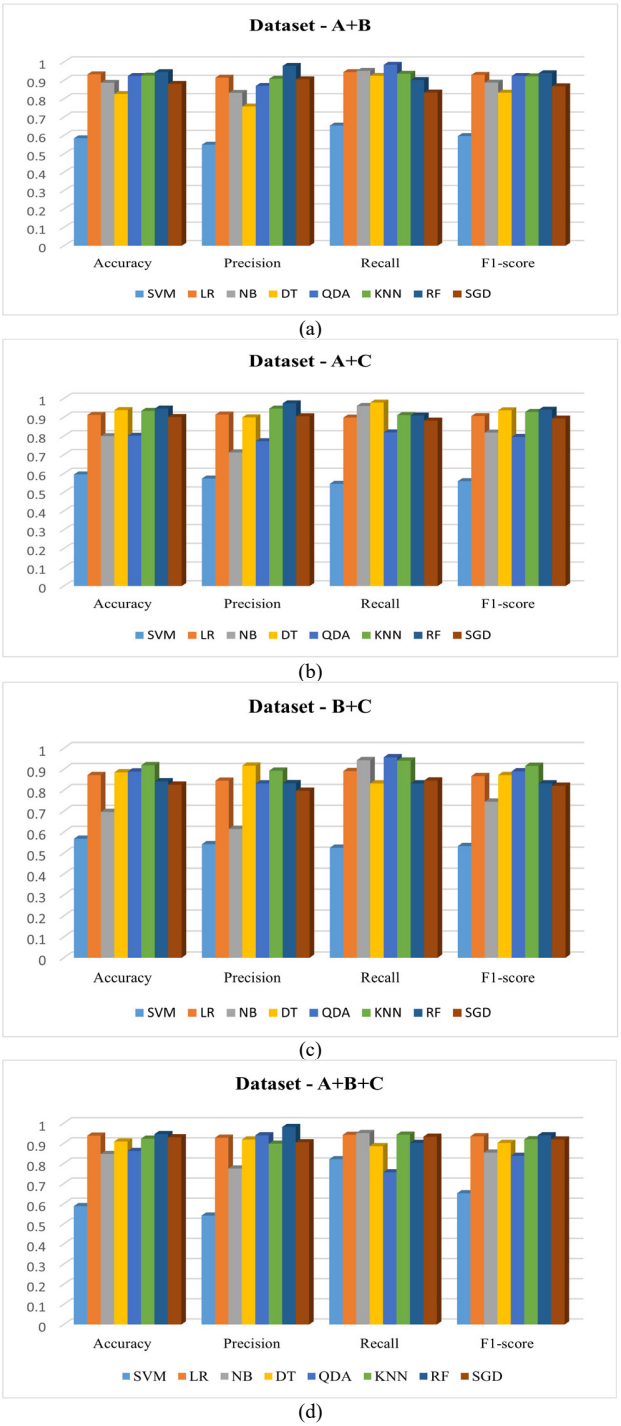


Fig. 4. Performance comparisons for combined datasets

Next, the hybrid SGD-RF model with the was proposed to further improve the detection performance. Table 10 displays the performance metrics results of testing all feature sets under the hybrid model. The table shows that throughout the

feature sets, the proposed hybrid model produced significant improvements in overall performance metrics in varying percentages. This improvement confirms the benefits of the hybrid SGD-RF ML model, we were able to get more comprehensive information using two of the best-performing algorithms. Thus, the model enabled us to understand the underlying patterns in the data better. Furthermore, the selection of significant BGP features resulted in the model being equipped with a greater amount of information, particularly as these features are the most impacted by cyberattacks. Consequently, there was an improvement in the accuracy of predictions and the overall performance metrics of the hybrid model.

Table 10: Performance matric results for Hybrid SGD-RF ML Model

Feature set (A)			
Accuracy	Precision	Recall	F1-score
0.968	0.960	0.980	0.970
Feature set (B)			
0.973	0.969	0.979	0.974
Feature set (C)			
0.877	0.913	0.849	0.880
Feature set (A+B)			
0.990	0.984	0.998	0.991
Feature set (A+C)			
0.981	0.972	0.992	0.982
Feature set (B+C)			
0.978	0.968	0.991	0.979
Feature set (A+B+C)			
0.993	0.989	0.998	0.994

Table 11: ROC-AUC values for Hybrid SGD-RF Model

Dataset - **A**	ROC-AUC = 0.963
Dataset - **B**	ROC-AUC = 0.965
Dataset - **C**	ROC-AUC = 0.879
Dataset - (**A+B**)	ROC-AUC = 0.986
Dataset - (**A+C**)	ROC-AUC = 0.981
Dataset - (**B+C**)	ROC-AUC = 0.977
Dataset - (**A+B+C**)	ROC-AUC = 0.993

Additionally, we analyzed the ROC-AUC metric, which is a method for organizing, visualizing the trade-off between TPR and false positive rate as shown in Figure 5, and selecting classification models based on their performance [37]. The ROC-AUC values for the proposed hybrid model are presented in Table 11. It can be seen that the SGD-RF model outperforms the rest of the individual methods in detection anomalies, with an ROC-AUC value of 0.993 for the dataset (A+B+C).

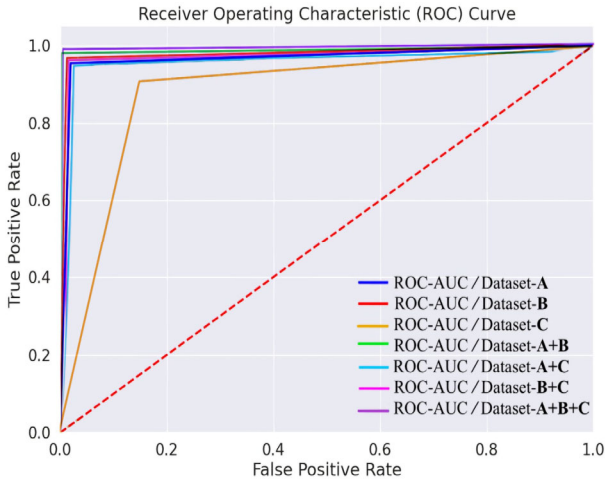


Fig. 5. ROC-AUC for Hybrid SGD-RF ML model

The findings of this research can provide valuable insights into the strengths and limitations of various ML algorithms when applied to BGP anomaly detection. By comparing the performance of different algorithms, we can identify the most suitable ML approach for BGP anomaly detection in real-world scenarios. Our results show that four of eight tested ML algorithms performed remarkably well in detecting BGP anomalies for the three feature sets generated based on statistical analysis. These algorithms achieved high accuracy, precision, and recall scores, demonstrating their potential for effective anomaly detection, particularly within feature sets or datasets most affected by cyber-attacks. Optimizing the model's performance and increasing its flexibility to obtain a wide range of better performance metrics results. Additional combinations of datasets were formed by merging two of the primary feature sets, and an extra dataset was established, including all the significant features that were found to be significant in the detection of cyber-attacks. The testing of these datasets resulted in various enhancements to the performance metrics. This indicates that our statistical analysis selected the most significant features in binary classification that capture essential BGP traffic patterns or behaviors

within our data sets. Thus, it combines more produced accuracy enhancement by giving the models additional information for making predictions [35].

Additionally, we implemented a hybrid model that clustered the strengths of two of the best-performing algorithms for improved anomaly detection. The hybrid model of SGD and RF holds great potential in BGP anomaly detection by overcoming the limitations of individual models. Clustering RF and SGD models substantially improved performance metrics on the hybrid model for all tested datasets, achieving the highest accuracy, surpassing the individual performance of Random Forest (RF) and Stochastic Gradient Descent (SGD). For instance, in the dataset that provided the best results (A+B+C), the accuracy measure showed an increase of 4.6% and 7.6%. Also, the hybrid model demonstrated enhancements in Recall since increased by 10.1% and 7% compared to the models RF and SGD, respectively.

The ROC-AUC improved in the hybrid model also, the AUC is essentially the integral of the ROC curve and influences the Recall. High accuracy and Recall are essential for guaranteeing that most anomalies are detected and are deeply relevant to BGP feature selection, particularly AS-PATH attribute features, which are critical in identifying anomalies in routing behavior. Our study features were extracted and evaluated under three common BGP cyber-attack conditions. The statistical analysis employed the ANOVA test, an effective tool for selecting the most significant features and determining the importance of the extracted feature in BGP anomaly detection. Based on statistical analysis, all 14 features from AS-PATH attributes were selected, with none excluded. AS-path features reflect insights into the logical and physical topology of BGP networks, as well as the logical relationships and interconnections between different ASs. This outcome strongly supported our work objectives by involving and considering the entire BGP network's topological criteria when extracting relevant features and datasets for BGP detection models. Such a methodology has a substantial contribution to creating and developing effective detection models with high accuracy that are applicable in real-world networks.

5. Conclusions

This study compares and evaluates the performance metrics of eight ML models (SVM, LR, NB, DT, QDA, KNN, RF, SGD) in the context of BGP anomaly detection. Based on statistical analysis, we assessed BGP behaviors by utilizing three primary feature sets across three levels of feature significance and generated four additional combined sets. Performance metrics were evaluated using data sets generated from simulated cyber-attack scenarios applied on realistic internet gateway networks. The best-performing algorithms were KNN, LR, RF, and SGD, which gained the highest metrics results. To further develop complementary strengths and robustness, we proposed a hybrid ML model that combines SGD and RF algorithms. This combination highly improved performance metrics for all tested datasets, achieving maximum accuracy of 99.3% for the (A+B+C) dataset, additionally the hybrid model exhibited a performance in Area Under Curve (AUC) of 0.993, exceeding this metric of the two combining models independently.

ML shows immense promise for the detection of anomalies in BGP. Through improved accuracy, real-time detection, and development and implementation of ML hybrid models based on realistic network datasets, exhibit potential in addressing the ever-increasing challenges in secure BGP networks and ensuring their reliable and efficient operation. We believe that our research will contribute to the existing literature on BGP anomaly detection and provide a foundation for future studies in this area.

Acknowledgment

Part of this work is funded by Universiti Kebangsaan Malaysia (ref: DPK-2023-009). The authors would like to thank the Ministry of Communication (MOC), Iraqi Telecommunication and Post Company (ITPC) for the shared data.

References

[1] Alotaibi, H. S., Gregory, M. A., Li, S. & Ali, I., 2022. Multidomain SDN-based gateways and Border Gateway Protocol. Journal of Computer Networks and Communications, 2022(1): 1-23.

[2] Mala, S. & Mallapur, S. V., 2022. A brief analysis of Border Gateway Protocol for Internet controlling and malicious attacks. International Conference on Computing, Communication, Electrical and Biomedical, pp. 561-572.

[3] Liao, H., Murah, M.Z., Hasan, M.K., Aman, A.H.M., Fang, J., Hu, X. and Khan, A.U.R., 2024. A survey of deep Learning technologies for intrusion detection in Internet of Things. IEEE Access.

[4] Mousa'B, M.S., Hasan, M.K., Sulaiman, R., Islam, S. and Khan, A.U.R., 2023. An explainable ensemble deep learning approach for intrusion detection in industrial Internet of Things. IEEE Access, 11 :115047-115061.

[5] Mohamed, M. and Alosman, K., 2024. A comprehensive machine learning framework for robust security management in cloud-based Internet of Things systems. Jurnal Kejuruteraan, 36(3), pp.1055-1065.

[6] Ahmad, Z., Shahid Khan, A., Nisar, K., Haider, I., Hassan, R., Haque, M.R., Tarmizi, S. and Rodrigues, J.J., 2021. Anomaly detection using deep neural network for IoT architecture. Applied Sciences, 11(15), p.7050.

[7] Al-Daweri, M.S., Abdullah, S. and Ariffin, K.A.Z., 2021. An adaptive method and a new dataset, UKM-IDS20, for the network intrusion detection system. Computer Communications, 180, pp.57-76.

[8] Vafaei Sadr, A., Bassett, B. A. & Kunz, M. 2023. A flexible framework for anomaly detection via dimensionality reduction. Neural Computing and Applications 35(2): 1157-1167.

[9] Hoarau, K., Tournoux, P. U. & Razafindralambo, T. 2022. Detecting forged AS paths from BGP graph features using Recurrent Neural Networks. IEEE 19th Annual Consumer Communications & Networking Conference (CCNC), pp. 735-736.

[10] Chliah, H., Battou, A., Hadj, M. a. E. & Laoufi, A. 2023. Hybrid machine learning-based approach for anomaly detection using Apache Spark. International Journal of Advanced Computer Science and Applications 14(4).

[11] Ding, Q., Li, Z., Haeri, S. and Trajković, L., 2018. Application of machine learning techniques to detecting anomalies in communication networks: Datasets and feature selection algorithms, pp. 47-70. Springer International Publishing.

[12] Verma, R.D., Govil, M.C. and Keserwani, P.K., 2023. ELM based ensemble of classifiers for BGP security against network anomalies. 11th International Symposium on Electronic Systems Devices and Computing (ESDC), pp. 1-6.

[13] Dietterich, T., 2000. Ensemble Methods in Machine Learning. Multiple Classifier Systems, pp.1-15.

[14] Banda, M., Ngassam, E.K. And Mnkandla, E., 2024. Enhancing Classification and Prediction through the Application of Hybrid Machine Learning Models. IEEE IST-Africa Conference (IST-Africa), pp. 1-12.

[15] Barbir, A., Murphy, S. and Yang, Y., 2006. Generic threats to routing protocols. http://www.ietf.org/rfc/rfc4593.txt

[16] Ibidunmoye, O., Rezaie, A.R. and Elmroth, E., 2017. Adaptive anomaly detection in performance metric streams. IEEE Transactions on Network and Service Management, 15(1): 217-231.

[17] Poupart, P., Chen, Z., Jaini, P., Fung, F., Susanto, H., Geng, Y., Chen, L., Chen, K. and Jin, H., 2016. Online flow size prediction for improved network routing. IEEE 24th International Conference on Network Protocols (ICNP), pp. 1-6.

[18] Sanchez, O.R., Ferlin, S., Pelsser, C. and Bush, R., 2019. Comparing machine learning algorithms for BGP anomaly detection using graph features. 3rd ACM CoNEXT Workshop on Big DAta, Machine Learning and Artificial Intelligence for Data Communication Networks, pp. 35-41.

[19] Karimi, M., Jahanshahi, A., Mazloumi, A. and Sabzi, H.Z., 2019. Border gateway protocol anomaly detection using neural network. IEEE International Conference on Big Data, pp. 6092-6094.

[20] Al-Musawi, B., Branch, P. & Armitage, G. 2016. BGP anomaly detection techniques: A survey. IEEE Communications Surveys & Tutorials, 19(1): 377-396.

[21] Bhatnagar, A., Majumdar, N. and Shukla, S., 2019. BGP anomaly detection using decision tree-based machine learning classifiers. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8: 4015-4020.

[22] Kalra, H., Singh, A.P. and Sadhya, D., 2021. Anomaly detection in Border Gateway Protocol using supervised machine learning. IEEE Bombay Section Signature Conference (IBSSC), pp. 1-6.

[23] Paiva, T.B., Siqueira, Y., Batista, D.M., Hirata, R. and Terada, R., 2021. BGP anomalies classification using features based on as relationship graphs. IEEE Latin-American Conference on Communications (LATINCOM), pp. 1-6.

[24] Al-Rousan, N.M. and Trajković, L., 2012. Machine learning models for classification of BGP anomalies. IEEE 13th International Conference on High Performance Switching and Routing, pp. 103-108.

[25] Lutu, A., Bagnulo, M., Cid-Sueiro, J. and Maennel, O., 2014. Separating wheat from chaff: Winnowing unintended prefixes using machine learning. IEEE Conference on Computer Communications (INFOCOM), pp. 943-951.

[26] Allahdadi, A., Morla, R. and Prior, R., 2017. A framework for BGP abnormal events detection. arXiv preprint: 1708.03453.

[27] Cheng, M., Li, Q., Lv, J., Liu, W. & Wang, J., 2021. Multi-scale LSTM model for BGP anomaly classification. IEEE Transactions on Services Computing 14(3): 765-778.

[28] Dai, X., Wang, N. and Wang, W., 2019, March. Application of machine learning in BGP anomaly detection. Journal of Physics: Conference Series, 1176(3): 032015.

[29] Cosovic, M., Obradovic, S. and Junuz, E., 2018. Deep learning for detection of BGP anomalies. Time Series Analysis and Forecasting: Selected Contributions from ITISE 2017, pp. 95-113. Springer International Publishing.

[30] Li, Z., Rios, A.L.G. and Trajković, L., 2020. Detecting internet worms, ransomware, and blackouts using recurrent neural networks. IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 2165-2172.

[31] Hoarau, K., Tournoux, P.U. and Razafindralambo, T., 2021, June. BML: an efficient and versatile tool for BGP dataset collection. IEEE International Conference on Communications (ICC) Workshops, pp. 1-6.

[32] Park, H., Kim, K., Shin, D. and Shin, D., 2023. BGP dataset-based malicious user activity detection using machine learning. Information, 14(9): 501.

[33] Butler, K., Farley, T. R., McDaniel, P. & Rexford, J., 2009. A survey of BGP security issues and solutions. Proceedings of the IEEE, 98(1): 100-122.

[34] Hoarau, K., Tournoux, P. U. & Razafindralambo, T., 2021. Suitability of graph representation for BGP anomaly

detection. IEEE 46th Conference on Local Computer Networks (LCN), pp. 305-310.
[35] Edgar, T.W. and Manz, D.O., 2017. Research methods for cyber security. Syngress.
[36] Ali, J., Khan, R., Ahmad, N. and Maqsood, I., 2012. Random forests and decision trees. International Journal of Computer Science Issues (IJCSI), 9(5): 272.
[37] Muneer, A., Mohd Taib, S., Mohamed Fati, S., O. Balogun, A. & Abdul Aziz, I., 2022. A hybrid deep learning-based unsupervised anomaly detection in high dimensional data. Computers, Materials & Continua 70(3): 5363-5381.
[38] Bernieri, G., Conti, M. and Turrin, F., 2019. Evaluation of machine learning algorithms for anomaly detection in industrial networks. IEEE International Symposium on Measurements & Networking (M&N), pp. 1-6.
[39] Rodríguez, M., Alesanco, Á., Mehavilla, L. and García, J., 2022. Evaluation of machine learning techniques for traffic flow-based intrusion detection. Sensors, 22(23): 9326.
[40] Kubat, M., 2017. An introduction to machine learning. Springer, Cham.
[41] Bognár, L. & Fauszt, T., 2022. Factors and conditions that affect the goodness of machine learning models for predicting the success of learning. Computers and Education: Artificial Intelligence 3: 100100.
[42] Ray, S,. 2024. 8 Ways to Improve Accuracy of Machine Learning Models. Available at: https://www.analyticsvidhya.com/blog/2015/12/improve-machine-learning-results/ (Accessed: 25 October 2024).
[43] Ridwan, M.A., Radzi, N.A.M., Azmi, K.H.M., Abdullah, F. & Ahmad, W.S.H.M.W. 2023. A new machine learning-based hybrid intrusion detection system and intelligent routing algorithm for MPLS network. International Journal of Advanced Computer Science and Applications 14(4): 94-107.
[44] Yu, Y., Lv, P., Tong, X. and Dong, J., 2020. Anomaly detection in high-dimensional data based on autoregressive flow. Database Systems for Advanced Applications: 25th International Conference (DASFAA), pp. 125-140.
[45] Kadhim, N.S., Chellappan, K. & Abdullah, N.F., 2024. BGP security analysis using network simulation: An impact study of cyber attacks. Jurnal Kejuruteraan.

NASSIR SALLOM KADHIM received B.Sc. in Communication and Electronic Engineering from University of Technology, Baghdad, Iraq in 1998. The Postgraduate Diploma in Computer Science from the Informatics Institute of Postgraduate Studies in Baghdad, Iraq in 1999. The M.Sc. in Computer Networking, Wireless ad hoc network (VANET) from University Malaysia Pahang (UMP), Malaysia in 2016.He is Currently, Ph.D. candidate specializing in Computer Engineering in University of Kebangsaan, Malaysia (UKM), with a research focus on BGP network security analysis and impact of cyber-attacks.

NOR FADZILAH ABDULLAH received the B.Sc. degree in Electrical and Electronics engineering from Universiti Teknologi Malaysia, in 2001, the M.Sc. degree (Hons.) in Communications Engineering from the University of Manchester, U.K., in 2003, and the Ph.D. degree in Electrical and Electronic Engineering from the University of Bristol, U.K., in 2012. She is currently an Associate Professor with Universiti Kebangsaan Malaysia, Selangor, Malaysia. Her research interests include 5G, millimeter wave, LTE-A, vehicular networks, massive MIMO, space-time coding, fountain codes, and channel propagation modeling and estimation.

KALAIVANI CHELLAPAN is an Associate Professor in Faculty of Engineering & Built Environment, Universiti Kebangsaan Malaysia since Nov 2011. Her research interests include Cardiovascular Engineering Modeling, IoT in Healthcare and Data Modeling & Analytics. She obtained the BEng (Electrical & Electronics) in 1995 from Universiti Sains Malaysia, MSc (CompSc) and PhD (Engineering) from Universiti Kebangsaan Malaysia in 2001 and 2009, respectively.