Published by
NIGERIAN SOCIETY OF PHYSICAL SCIENCES
Available online @ https://journal.nsps.org.ng/index.php/jnsps

J. Nig. Soc. Phys. Sci. 6 (2024) 2093

Journal of the Nigerian Society of Physical Sciences

Analysis of support vector machine and random forest models for predicting the scalability of a broadband network

Gabriel James[a,*], Anietie Ekong[b], Etimbuk Abraham[c], Enobong Oduobuk[d], Peace Okafor[e]

[a]Department of Computing, Topfaith University, Nigeria
[b]Department of Computer Science, Akwa Ibom State University, Nigeria
[c]Department of Electrical Electronics Engineering, Topfaith University, Nigeria
[d]Department of Physics, Topfaith University, Nigeria
[e]Department of State Service, Bayelsa State Command, Bayelsa State, Nigeria

Abstract

This study proposed a machine learning approach to predict the scalability of broadband networks, which is crucial for ensuring fast and reliable internet connectivity. Scalability measures a network's ability to handle increasing users, devices, and data traffic without compromising performance. The researchers leveraged the strengths of Random Forest (RF) and Support Vector Machine (SVM) algorithms to predict scalability. A large dataset of 40,000 data points was collected, focusing on six key metrics: Response Time, Bandwidth, Latency, Error Rate, Throughput, and Number of Users Connected. The data was preprocessed and divided into training and testing sets (80:20 ratio). Both RF and SVM algorithms were trained on the dataset, and a comparative analysis was conducted to determine which algorithm performed better. The results showed that the RF model outperformed the SVM model, achieving an accuracy of 95.0% compared to 91.0%. The RF model also exhibited higher precision, recall, and AUC scores. Feature importance analysis revealed that Response Time and Throughput were the most significant factors in determining network scalability. The study demonstrated the effectiveness of the RF model in predicting broadband network scalability, with a lower loss value (0.0133 for training and 0.0160 for validation) compared to the SVM model. This approach will help network operators and administrators predict and improve network scalability, ensuring reliable and fast internet connectivity. The study contributes to the development of machine learning-based solutions for broadband network performance evaluation and optimization.

DOI:10.46481/jnsps.2024.2093

Keywords: Broadband networks, Machine learning, Network performance metrics, Random forest, Support vector machine

Article History :
Received: 21 April 2024
Received in revised form: 16 June 2024
Accepted for publication: 28 June 2024
Published: 27 July 2024

© 2024 The Author(s). Published by the Nigerian Society of Physical Sciences under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

Communicated by: O. Akande

1. Introduction

The word "broadband" refers to a broad category of technologies that provide families, businesses, and other institutions with high-speed internet access. Broadband networks are high-speed, high-capacity transmission networks that can handle multiple types of data, traffic, and signals at once; this makes it play a crucial role in modern communication, supporting various applications and services that require high-speed internet connectivity. These networks are designed to provide fast and reliable internet connectivity, enabling the transfer of large amounts of data at high speeds. The key characteristics of

*Corresponding author: Tel.: +234-810-738-1867.
Email address: g.james@topfaith.edu.ng (Gabriel James)

broadband networks are their performance abilities which are determined by some key metrics such as; capacity, speed, multiple signal capacity, bandwidth, and scalability, with scalability as the most important metric. Scalability here means the ability of a broadband network to efficiently handle and accommodate growth in user demand, data traffic, and connected devices while maintaining optimal performance. Scalability is a critical aspect of network design, ensuring that the network can expand or contract seamlessly to meet changing requirements without significant degradation in speed, capacity, or service quality. Broadband network scalability is particularly crucial in the context of rapidly evolving technologies such as the proliferation of connected devices, and the increasing demand for high-speed internet access. As user requirements change and data traffic grows, a scalable network ensures that service providers can meet these demands without significant disruptions or the need for extensive redesigns. Scalability contributes to the long-term viability and competitiveness of broadband networks in meeting the needs of users and supporting emerging applications and services. Poor scalability in broadband networks can have various detrimental effects on service providers and end-users. Thus, to effectively address these challenges, broadband networks must be designed and maintained with scalability in mind.

According to Ruan *et al.* [1], Network Scalability refers to the capability of a network to handle increased demands and grow its capacity as the need arises. It involves designing and implementing a network infrastructure that can accommodate rising traffic, data volume, and user requirements without compromising performance or reliability [2]. So, evaluating the scalability of a broadband network with emphasis on some key factors is crucial to ensuring the ability to accommodate the growing demands of users and applications over time. Network planners and administrators could benefit from a methodical analysis of these aspects as it could shed light on how scalable broadband networks are both presently and in the future, which could improve proactive planning. Proactive planning and routine evaluations are necessary to guarantee that the network can adapt and satisfy the needs of a constantly shifting digital environment [3, 4]. Many technologies have been employed to predict the scalability and performance of broadband networks which has further improved the performance of broadband networks. For example, Umoren *et al.* [3] employed a fuzzy knowledge-based method and a computational intelligence framework with numerous criteria. VHD methods use the Adaptive Intelligence Multi-Factored Algorithm (AIMFA) and the Multi-Criteria Algorithm to anticipate the performance of a mobile broadband communication network. A highly skilled optimal-based handoff decision algorithm was produced as a result of the work, and it helped to reduce the ping-pong effect and call drops that network operators face to provide more effective service. Also in 2020, Ituma *et al.* [5] used a Fuzzy type 1 to provide a precise and accurate solution to the packet switching problem, as opposed to non-linguistic specifications which produce unreliable solutions. With the use of their underlying notion of linguistic variable definitions for variables with uncertainties that have gone largely unacknowledged, the technique was able to ensure increased quality of service; this is related to the idea of accuracy. Ituma *et al.* [6] work used a fuzzy logic-based methodology to optimize packet switching in wireless communication systems. To do this, their primary packet-switching factors were determined and examined. The factors investigated were: transmitted packet length (TPL), packet loss (PL), packet arrival rate (PAR), traffic intensity (TI), latency (L), or delay. Data from an already-existing organization—the Akwa Ibom Broadcasting Corporation (AKBC), a Third Generation (3G) government-owned business based in Uyo was used. It was noted that research was also done on how some of these elements interacted with the total network throughput. They were shown to exhibit comparable patterns as well, demonstrating the significant dependency or correlation between the various factors and the total network throughput [5, 6]. The majority of research went so far as to use machine learning models to assess, forecast, and figure out broadband networks' performance capabilities. The creation and study of statistical algorithms that can learn from data and generalize to new data, allowing them to carry out tasks without explicit instructions, is the focus of the artificial intelligence discipline [7, 8]. By applying statistical models and algorithms to analyze and deduce patterns in data, it is used to create computer systems that can learn and adapt without having to be explicitly instructed [9, 10]. Machine learning is a key technology for making predictions and achieving accuracy when working with datasets [11], By applying machine learning to a dataset, it is possible to uncover insights, identify trends, and make informed decisions with greater accuracy [12].

Deep learning (DL) is a class of machine learning algorithms that uses multiple layers to progressively extract higher-level features from the raw data (input) [3–14]. The adjective "DEEP" refers to the use of multiple layers in the network in either supervised, semi-supervised, or unsupervised methods. It is more of a mathematical distribution for complex behavior than traditional machine learning. One such DL model is the Convolutional Neural Network (CNN) which is most commonly employed in computer vision [11]. Given a series of images or videos from the real world, with the utilization of a CNN, the AI system learns to automatically extract the features of these inputs to complete a specific task, e.g., image classification, face authentication, and image semantic segmentation. Different from fully connected layers in MLPs, in CNN models, one or multiple convolution layers extract the simple features from input by executing convolution operations [15]. Each layer is a set of nonlinear functions of weighted sums at different coordinates of spatially nearby subsets of outputs from the prior layer, which allows the weights to be reused [16]. Indeed, selecting the right machine learning algorithm depends on several factors, including, but not limited to: data size, quality, and diversity, as well as what answers businesses intended derivable from that data. Additional considerations include accuracy, training time, parameters, data points, and much more [17]. Therefore, choosing the right algorithm is a combination of business needs, specifications, experimentation, and time available [18].

According to Eyceyurt *et al.* [19], applying machine learn-

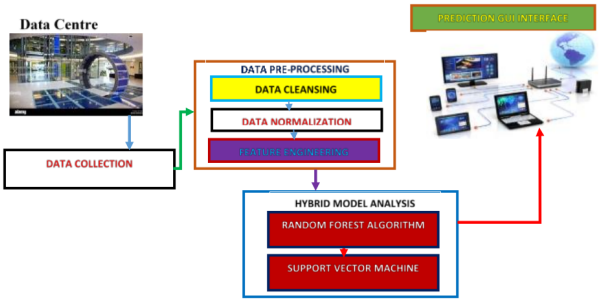


Figure 1: Conceptual framework of the proposed model.

ing to assess a broadband network's scalability entails employing models and algorithms to analyze data, spot trends, and forecast how the network will function in different scenarios. It is important to approach machine learning with a full grasp of the network's architecture, domain expertise, and the unique scalability difficulties faced by the broadband network [19]. Machine learning can offer useful insights into network scalability. The quality of the data, feature selection, and ongoing model improvement over time are critical to the machine learning approach's effectiveness [20, 21].

However, it has been noticed that all these works are network-centered as all gear towards enhancing the functional ability of the network alone. But our concept shall be user-com user-networks centered as it will help enhance the performances of the broadband network and as well accepting more users with user-friendly experiences. This work, Comparative Analysis of Support Vector Machine and Random Forest Models for Predicting the Scalability of Network Broadband. This approach leverages the strengths of the Machine Learning Model based on RF and SVM Algorithms to enhance accuracy and robustness for the specific problem of the prediction of broadband network scalability. This approach is motivated by the complementary nature of RF's ensemble learning and SVM's ability to handle non-linear decision boundaries, which enhances the prediction accuracy and robustness for the specific problem of broadband network scalability.

2. Methodology

Computational intelligence (CI) otherwise known as machine learning (ML), is an offshoot of the field of data analytics and pattern analysis which are both domains of artificial intelligence. It is made up of several arithmetic expressions that enhance its capability to carry out reliable predictions for a given task. Because of its vast application areas, ML has become a very unique tool for researchers, engineers as well as data scientists to build detection and prediction systems [22, 23]. In this unit, we shall examine in detail, the machine learning technique used to enhance the prediction of the scalability of broadband networks.

Figure 1 shows the framework for the hybrid intelligent models for the evaluation of the scalability of broadband networks.

2.1. Overview of the major components of the framework

2.1.1. Data pre-processing

Data pre-processing is a crucial step in the machine learning pipeline that involves cleaning, transforming, and organizing raw data into a format suitable for training and evaluating models. The ultimate goal is to ensure that the data is clean, relevant, and well-suited for training a model that can generalize well to new, unseen data. In this work, we identified and dealt with missing values by directly removing all instances with missing values, duplicates, and outliers using the RF built-in missing value handler. The MinMaxScaller techniques were used to perform the Data Transformation process where features were scaled, encoded, and dimensionally reduced. For encoding the OHE methods were used. The Data was normalized to the common range using Normalizer and the relevant features were selected with the Random Forest features selection. The dataset was then divided into training and testing sets in the ratio of 80:20 to assess/validate the models' performance on unseen data.

2.1.2. Data Collection

The broadband networks dataset which consists of about 40,000 data points was collected from Coquina Software Company Limited, F3 Ewet Housing Extension, Uyo, Akwa Ibom State, Nigeria; a well-known software company in Akwa Ibom State that offers broadband internet services. The data contains historical data on network performance, including bandwidth usage, latency, error rates, and other relevant metrics. The NS-3 Network Simulator was used to simulate network scenarios and the Prometheus software platform was used to collect and store network data; this data was then passed through the various pre-processing stages of preparations and usage in the machine learning model training. Relevant features were identified which may impact network scalability. This includes the number of users, response time, Bandwidth, Latency, error rate, Throughput, and the class representing the scalability as the target features. The raw data was transformed into a format suitable for machine learning, ensuring that it reflects the key aspects of network behavior, and split into training and testing sets to train and evaluate the model's performance with the percentage of 80:20; where 80% representing 32,000 was used as training data sets and 20% representing 8,000 was used as testing data sets and Table 1 contains a segment of the dataset.

2.1.3. Random forest algorithm (RFA)

The Random Forest algorithm is an ensemble learning method used for both classification and regression tasks. It operates by constructing a multitude of decision trees during training and outputs the mode of the classes (for classification) or the mean prediction (for regression) of the individual trees [24–26]. It builds multiple decision trees by sampling the training data with replacement. This process, known as bootstrapped sampling, ensures that each tree is trained on a slightly different subset of the data. In addition to sampling data points, Random Forest introduces randomness by considering only a

Table 1: Sample dataset.

Response time (ms)	Band width (Mbps)	Latency (ms)	Error rate (%)	Through put (Mbps)	Users connected	Class
109.83	40	22.48	1.7	100.59	51	1
97.23	78	19.31	1.4	110.33	51	1
112.95	112	23.24	1.8	80.37	58	1
130.46	130	27.62	2.2	71.90	42	1
95.32	50	18.83	1.3	102.14	33	1
72.45	100	51.41	1.2	82.47	30	1
118.95	100	51.40	3.9	78.93	56	1
56.74	100	51.41	2.8	75.40	60	0
100.21	100	51.41	2.3	71.87	33	1
56.76	100	51.40	1.2	68.34	60	0
73.55	100	51.41	1.9	64.80	50	1
39.41	100	51.40	0.8	61.27	40	0
54.23	100	51.40	1.5	57.74	50	0
34.56	100	51.10	2.7	54.20	50	0

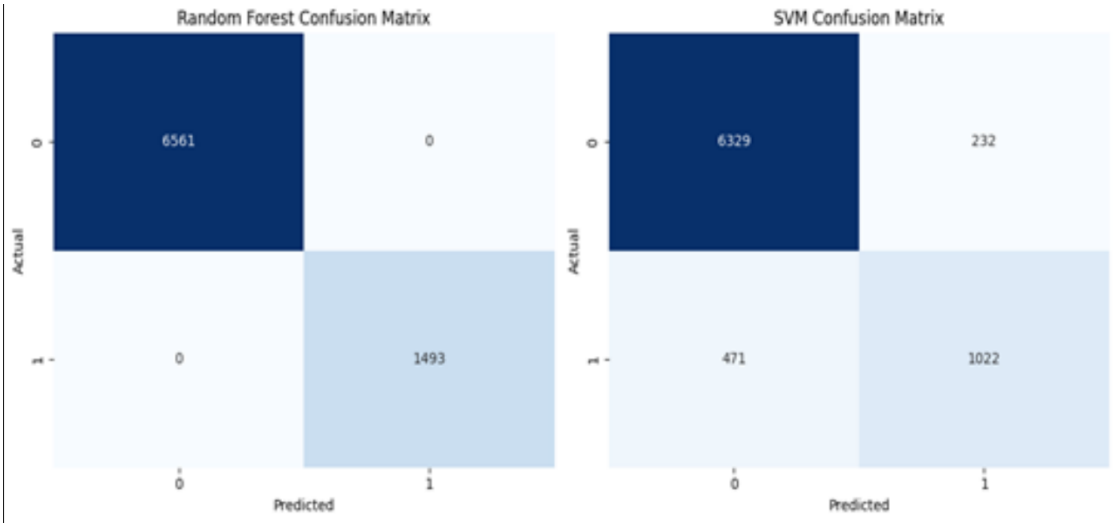


Figure 2: Confusion metric plot for the hybrid model.

Table 2: Sample dataset.

	Model	Accuracy	Precision	Recall	F1-score	AUC
1	RF	1.0000	1.0000	1.0000	1.0000	1.0000
2	SVM	0.9082	0.8231	0.6563	0.7303	0.8117

Table 3: Summary of the confusion matrices.

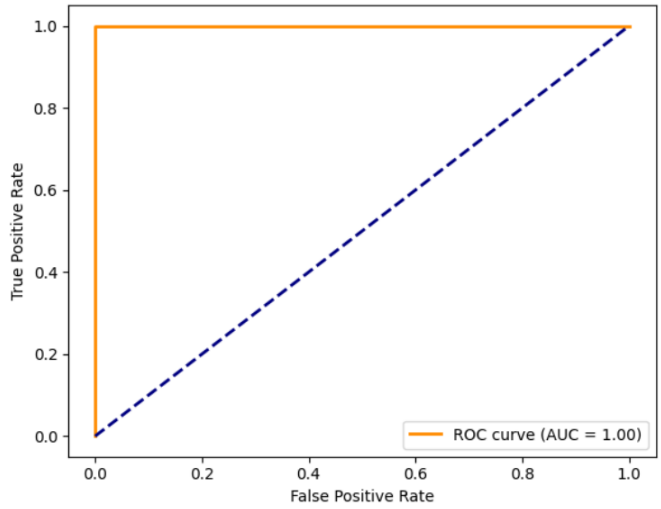
Metrics	Random forest classifier	Support vector machine
True Positive (TP)	1525	1001
True Negative (TN)	6314	6314
False Positive (FP)	0	215
False Negative (FN)	0	524
Accuracy	0.95	0.91
Precision	0.94	0.82
Recall (Sensitivity)	0.95	0.66
Specificity	0.93	0.73
F1 Score	1.00	0.81

random subset of features at each split in the decision tree. This helps to decorate the trees, making the ensemble more robust and reducing the risk of overfitting. The mathematical model for the random forest algorithm involves the combination of decision trees. For classification tasks, the output of each decision tree is a class label $C_i(x)$, where x is the input data. The final prediction of the Random Forest is obtained through majority voting:

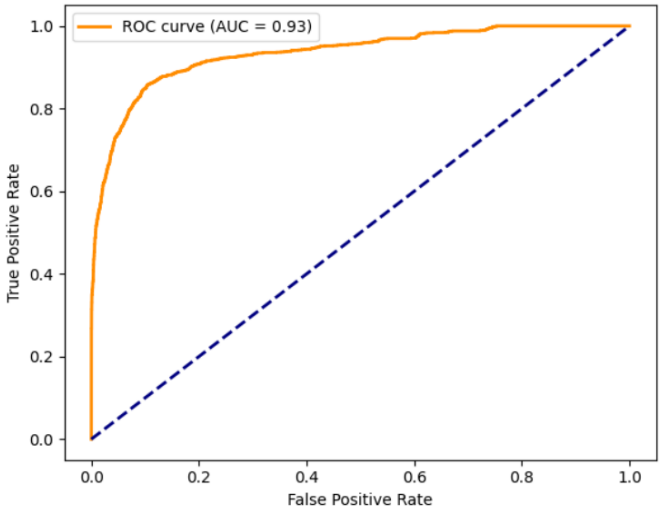
$$\sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j K\left(x_i, x_j\right), \tag{1}$$

Table 4: Summary of the model's classification reports.

Model	Characteristics	AUC	Precision	Recall	F1-score	Support
RF		0.0	0.96	0.93	0.94	6561
		1.0	0.97	0.91	0.93	1493
	accuracy				0.95	8054
	macro avg		0.94	0.96	0.94	8054
	weighted avg		1.00	0.94	0.95	8054
SVM		0.0	0.93	0.96	0.95	6561
		1.0	0.81	0.68	0.74	1493
	accuracy				0.91	8054
	macro avg		0.87	0.82	0.85	8054
	weighted avg		0.91	0.91	0.91	8054

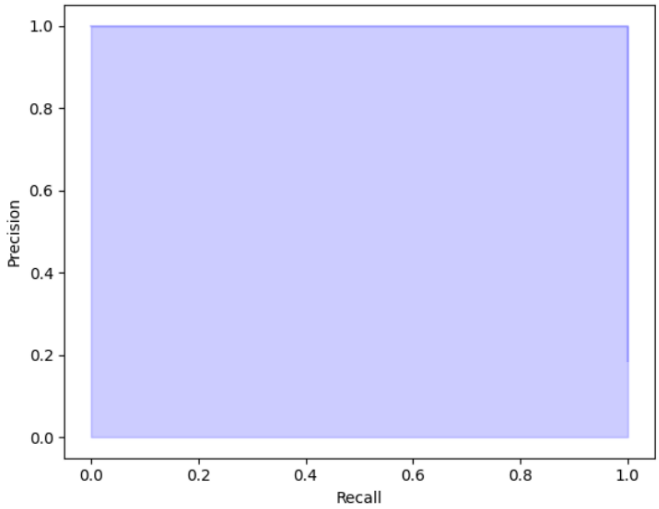


(a) Random forest ROC curve.

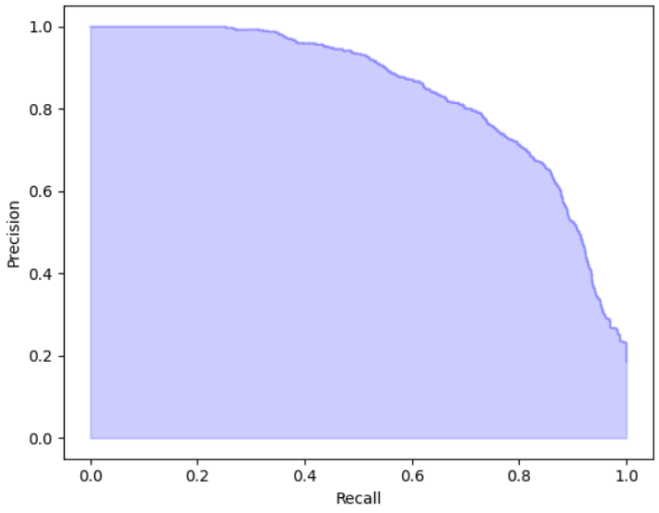


(b) SVM ROC curve.

Figure 3: The ROC curve.



(a) Random forest precision recall curve (AP= 1.00).



(b) SVM precision recall curve (AP= 0.84).

Figure 4: Precision recall curves.

subject to:

$$\sum_{i=1}^{n} \alpha_i y_i = 0, \tag{2}$$

where α_i is the weight assigned to the training sample x_1. If $\alpha_i > 0$, x_1 is called a support vector C is a "regulation pa-

Table 5: Summary of feature importance.

Feature description	Percentage of importance
response time	0.603
Throughput	0.388
Latency	0.004
Error Rate	0.002
Users Connected	0.001
Bandwidth	0.001

rameter" used to trade off the training accuracy and the model complexity so that a superior generalization capability can be achieved. K is a kernel function, which is used to measure the similarity between two samples. A popular radial basis function (RBF) kernel function, as shown in [25]. The following Algorithm [26] can be used in the classification process.

```
Input: sample x to classify Training set
T = {(x1,y1),(x2,y2), ... (xn,yn)}
Number of nearest neighbors k.
Output: decision yp in {-1,1}
Find k sample (xi, yi) with minimal values
of K (xi, xi) - 2 * K(xi, x)
Train an SVM model on the k-selected samples
Classify x using this model, get the result yp
Return yp
```

2.1.4. Support vector machine (SVM)

A Support Vector Machine (SVM) is a machine learning algorithm that looks at data and sorts it into one of two categories [27]. It is a supervised and linear Machine Learning algorithm most commonly used for solving classification problems and is also referred to as Support Vector Classification. It is characterized by the equation of the main separator line is called a hyperplane equation. The equation becomes $mx + c = 0$, where the hyperplane equation dividing the points (for classifying) can now easily be written as:

$$H: \; wT(x) + b = 0. \tag{3}$$

Here, b is intercept and bias term of the hyperplane equation.

The distance of any line, $ax + by + c = 0$ from a given point say, (x_0, y_0) is given by d. Similarly, the distance of a hyperplane equation: $wT\Phi(x) + b = 0$ from a given point vector $\Phi(x_0)$ can be easily written as:

$$d_H(\varphi(x_0)) = \frac{\left|w^T\left(\varphi\left(w^T\right)\right) + b\right|}{\|w\|_2}, \tag{4}$$

where $\|w\|_2$ is the Euclidean norm for the length of w given by

$$\|w\|_2 =: \sqrt{w_1^2 + w_2^2 + w_3^2 + \ldots w_n^2}. \tag{5}$$

In this work, the SVM shows effectiveness in the handling of classification and hyperplane optimization.

3. Discussion of results

The evaluation of the result is an organized and unbiased arrangement geared towards explaining the output. This enables the researcher to determine if the goal(s) are being met [28–31]. In this work, a comparative analysis of the performance of machine learning models (i.e. RF and SVM) is used to predict the scalability of a broadband network. Binary classifications were carried out using the major metrics in the dataset. It was observed that the throughput and latency depend largely on the network bandwidth whose satisfaction is based on the number of users connected to the network at any given time. The dataset was supplied into the RF and the SVM separately. In each of these instances, the model results, the confusion matrix, as well as the comparative analysis, were extracted and presented in this report. Confusion matrix is a representation of classification prediction outcomes, where the count values are used to total and breakdown down the number of true and false predictions by class. Table 2 depicts the confusion matrix for the RF and SVM classification models used in this work respectively. The table provided the summary of the performance metrics of two models, random forest and support vector machine (SVM), for the binary classification task.

Random Forest model achieved a perfect accuracy of 1.0, indicating that it made correct predictions for all instances in the test set whilst SVM achieved an accuracy of 0.908244 or 90.82%, indicating that it correctly predicted the class for approximately 90.82% of instances. Random Forest Recall is 1.0, suggesting that the Random Forest captured all positive instances in the dataset whilst SVM Recall is 0.656393 or 65.64%, indicating that the SVM captured approximately 65.64% of the actual positive instances. In all, the Random Forest model appears to perform exceptionally well, achieving perfect scores across all metrics. The SVM model also performs well but has slightly lower accuracy and recall compared to the Random Forest. Choosing between the two models depends on specific requirements and considerations of the application. With the values in Table 2, Table 3 is completed using the formulas as follows:

i Accuracy (AC):

$$AC = \frac{(TP + TN)}{(TP + TN + FP + FN)}. \tag{6}$$

ii Precision (P):

$$P = \frac{\mathrm{TP}}{(TP + FP)}. \tag{7}$$

iii Recall (Sensitivity) (RS):

$$RS = \frac{\mathrm{TP}}{(TP + FN)}. \tag{8}$$

iv Specificity (S):

$$S = \frac{\mathrm{TN}}{(TN + FP)}. \tag{9}$$

v F1 Score (F1-Score):

$$F1 - \mathrm{Score} = 2 \times \frac{(\mathrm{Precision} \times \mathrm{Recall})}{(\mathrm{Precision} + \mathrm{Recall})}. \tag{10}$$

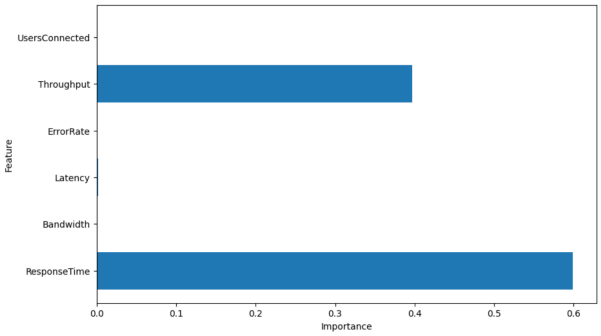
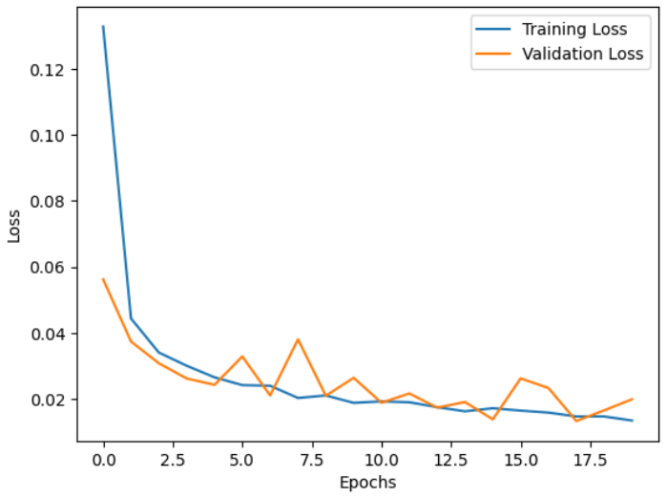


Figure 5: The feature importance curve.

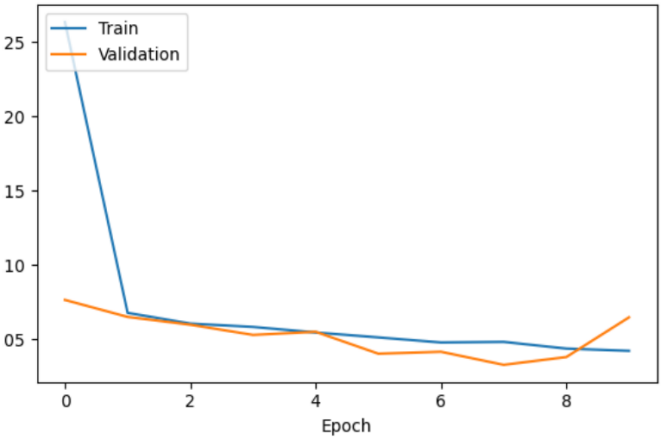
However, the training process was divided into epochs, where one epoch is a complete pass through the entire training dataset and according to [32], in Machine Learning, the test dataset always follows the pattern of the training set. At epoch 1, the hybrid model starts with an accuracy of 95.10% on the training set and 97.93% on the validation set. The loss is 0.1388 for training and 0.0575 for validation. After 20 epochs, the model achieves an accuracy of 99.54% on the training set and 99.50% on the validation set. The output of the loss function decreases to 0.0133 for training and 0.0160 for validation. It was noticed that the model shows a consistent improvement in both accuracy and loss over epochs, indicating effective learning. The decreasing loss values suggest that the model is converging towards an optimal state and the accuracy on the validation set is also consistently high, indicating good generalization. The training process appears successful, with the model learning from the data and generalizing well to unseen samples. A slight increase was noticed in the validation loss after epoch 15, which might be a sign of overfitting. In general, the training log indicates a well-behaved training process with improvements in accuracy and decreasing loss over epochs. Table 4 shows the summary of the classification report.

It was observed in Figure 2 that the precision, recall, and F1-score vary between the two classes. Class 0 has a higher precision, recall, and F1 score compared to Class 1. Weighted averages reflect a balance between the classes based on the number of instances.

These visualizations help in the assessment and comparison of the performance of the Random Forest and SVM classifiers in terms of classification and probability prediction [33, 34]. They provide insights into the models' ability to discriminate between positive and negative classes and their overall effectiveness. The ROC curve in Figure 3 shows the trade-off between a true positive rate (sensitivity) and a false positive rate. A steeper curve and higher AUC indicate better model performance. The Precision-Recall curve in Figure 4 evaluates the trade-off between precision and recall. Higher precision and recall values lead to a higher area under the curve (AP). Table 5 shows the feature importance of the dataset whilst Figure 5 presents the feature importance curve. The table presents details representing the feature importance scores for different features in the hybrid model. In this case, the feature importance is presented as a series of values ranging from 0 to 1,



(a) Training and validation loss.

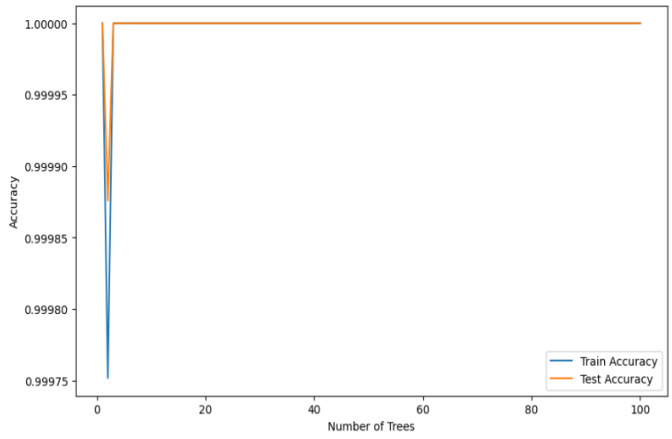


(b) Model loss.

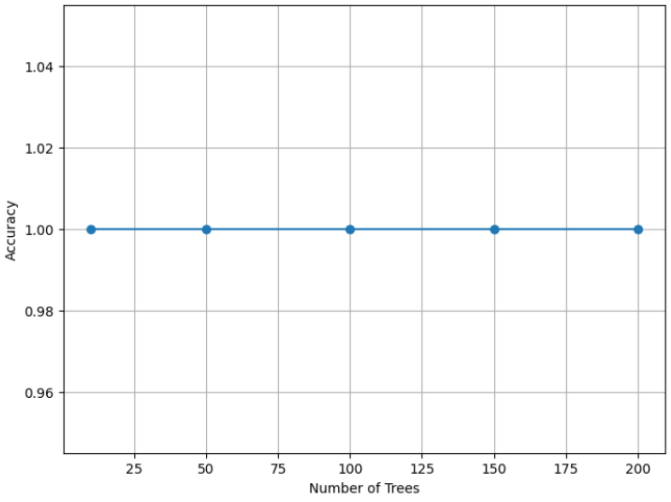
Figure 6: Precision recall curve.

where a higher value indicates a more important feature. Feature importance scores indicate the relative contribution of each feature to the model's predictive performance. Features with higher importance scores are more influential in determining the model's output. The sum of all importance scores is equal to 1. Figures 6-8 show the various curves by the proposed model.

From the tables and the ROC plots, it is observed that RF was able to obtain better classification results in terms of ROC, compared to the SVM classifier. The two models performed reasonably well on precision and recall. Although RF performed better than SVM in simulations, in the real dataset SVM achieved slightly better testing results. The hybrid model, by learning sparse representation using RF as a feature detector, improved over the SVM model in terms of testing data classification. The large sample size may be the reason why the two methods performed better. Even though, the RF achieved slightly better testing data classification error rate, indicating its applicability on scalability prediction in broadband networks.



(a) Accuracy over number of trees.



(b) Random forest accuracy versus number of trees.

Figure 7: Accuracy versus number of trees.

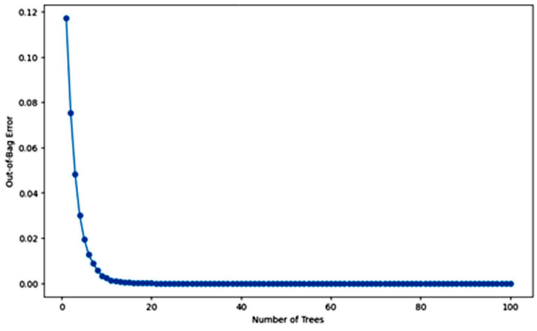


Figure 8: Out-of-bags over several trees.

4. Conclusion

As the number of users or devices increases, a poorly scalable network may become congested, leading to slower data transfer rates and degraded performance for all users which invariably results in a poor user experience, especially during peak usage hours [35–37]. Insufficient scalability may lead to network overloads, causing service interruptions or downtime. This can impact critical services, disrupt business operations, and inconvenience users who rely on a stable internet connection. Similarly, poor scalability limits the ability of the network to accommodate the growing demand for bandwidth and services. This lack of capacity can hinder the expansion of services and hinder the adoption of new technologies that require higher data transfer rates [37]. To address these challenges, broadband networks must be designed and maintained with scalability in mind. Regular assessments, upgrades, and investments in infrastructure are necessary to ensure that networks can handle increasing demands and provide a satisfactory user experience. This study proposed a novel approach to broadband network scalability prediction, leveraging the strengths of Machine Learning Models based on the Random Forest Model and Support Vector Machine to predict the scalability of a broadband network and a comparative analysis carried out to evaluate the best model in terms of accuracy and robustness for the specific problem of the prediction of broadband network scalability. The Random Forest Algorithm (RF) compared to the Support Vector Machine (SVM), appears to be more promising in predicting the scalability of broadband networks. After model training and testing, the RF model achieves an accuracy of 95.0%; while the SVM model achieves an accuracy of 91.0%. This means that the RF model exhibited stronger predictive performance than the SVM model, as evidenced by high accuracy, precision, recall, and AUC scores. Feature importance analysis revealed the significance of ResponseTime and Throughput in determining network scalability. With the RF model, the loss decreases to 0.0133 for training and 0.0160 for validation which proves that the RF model is more effective in predicting the scalability of broadband networks than the SVM model.

Further research should consider the following:

(i) Monitor the model's performance on an independent test set to ensure it generalizes well to new, unseen data.

(ii) Consider techniques like regularization or dropout to address overfitting if needed.

(iii) Evaluate the model's performance on real-world scenarios and domain-specific metrics.

(iv) Explore hyperparameter tuning for both models to optimize performance.

(v) Finally, investigate any class imbalances and consider strategies like oversampling, undersampling, or using class weights.

5. Acknowledgment

I would like to thank God almighty the Giver of Wisdom who helped the team to complete the first and second faces of this project. Immediately after the article is published, the second article from this project will be forwarded for review. I would also like to sincerely appreciate the Editorial team and reviewers of this article for shaping the work with a very insightful review. I want to say I am more satisfied with the current state of the work than when we submitted it at the first

instant. I shall continue to publish with this journal and as well introduce other researchers to the journal.

References

[1] L. Ruan, M. P. I. Dias & E. Wong, "Machine learning-based bandwidth prediction for low-latency h2m applications," IEEE Internet Things J. **6** (2019) 3743. https://doi.org/10.1109/JIOT.2018.2890563.

[2] L. Huang, X. Dong & T. Edward, "A scalable deep learning platform for identifying geological features from seismic attributes", The Leading Edge **36** (2017) 249. https://doi.org/10.1190/tle36030249.1.

[3] I. Umoren, S. Inyang & A. Ekong, "A fuzzy knowledge-based system for modeling handoff prediction in mobile communication networks", Journal of mobile communication **15** (2021) 1. https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://docsdrive.com/%3Fpdf%3Dmedwelljournals/jmcomm/2021/1-19.pdf.

[4] R. M. Morais & J. Pedro, "Machine learning models for estimating quality of transmission in DWDM networks", J. Opt. Commun. Netw **10** (2018) 84. https://doi.org/10.1364/JOCN.10.000D84.

[5] C. I. Ituma, S. O. Iwok & G. G. James, "Implementation of an optimized packet switching parameters in wireless communication networks", Int. J. Sci. Eng. Res. **11** (2020) 350. https://www.ijser.org/researchpaper/IMPLEMENTATION-OF-AN-OPTIMIZED-PACKET-SWITCHING-PARAMETERS-IN-WIRELESS-COMMUNICATION-NETWORKS.pdf.

[6] C. Ituma, S. O. Iwok and G. G. James, "A model of intelligent packet switching in wireles communication networks," Int. J. Sci. Eng. Res. **11** (2020) 2341. http://www.ijser.org.

[7] Y. Wang & M. Kosinski, "Deep neural networks can detect sexual orientation from faces", Graduate School of Business, Stanford University, Stanford, CA94305, USA, 2020. https://osf.io/zn79k/.

[8] G. G. James, P. C. Okafor, E. G. Chukwu, N. A. Michael & O. A. Ebong, "Predictions of criminal tendency through facial expression using convolutional neural network", J. Inf. Syst. Inform. **6** (2024) 635. https://journal-isi.org/index.php/isi/article/view/635.

[9] I. J. Goodfellow et al., "Challenges in representation learning: a report on three machine learning contests". arXiv, 2013. [online]. http://arxiv.org/abs/1307.0414.

[10] G. G. James, W. F. Ekpo, E. G. Chukwu, N. A. Michael, O. A. Ebong & P. C. Okafor, "Optimizing business intelligence system using big data and machine learning", J. Inf. Syst. Inform. **6** (2024) 1. http://journal-isi.org/index.php/isi.

[11] H. Jha & V. Vijayarajan, "Mobile internet throughput prediction using machine learning techniques", in 2020 International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India **8** (2020) 253. https://doi.org/10.1109/ICOSEC49089.2020.9215436.

[12] B. S. Sahay and J. Ranjan, "Real-time business intelligence in supply chain analytics", Inf. Manag. Comput. Secure. **16** (2008) 28. https://doi.org/10.1108/09685220810862733.

[13] T. Subramanya, D. Harutyunyan & R. Riggio, "Machine learning-driven service function chain placement and scaling in MEC-enabled 5G networks", Comput. Netw. **166** (2020) 106980. https://doi.org/10.1016/j.comnet.2019.106980.

[14] A. Ekong, I. Attih, G. James & U. Edet, "Effective classification of diabetes mellitus using support vector machine algorithm", Res. J. Sci. Technol. **4** (2024) 18. https://rejost.com.ng/index.php/home.

[15] N. Kundariya et al., "A review on integrated approaches for municipal solid waste for environmental and economical relevance: Monitoring tools, technologies, and strategic innovations", Bioresour. Technol. **342** (2021) 25982. https://doi.org/10.1016/j.biortech.2021.125982.

[16] N. G. Resmi, A. Shajan, J. Jose, J. P. George & M. Harikrishnan, "Solid waste tracking and route optimization using geotagging and k-means clustering", Int. J. Appl. Eng. Res. **16** (2021) 633. https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://www.ripublication.com/ijaer21/ijaerv16n8_01.pdf.

[17] J. Wang, X. Wu & C. Zhang, "Support vector machines based on K-means clustering for real-time business intelligence systems", Int. J. Bus. Intell **1** (2005) 54. https://doi.org/10.1504/IJBIDM.2005.007318.

[18] S. Biswas & J. Sen, "A proposed architecture for big data driven supply chain analytics", Innovation Practice eJournal **14** (2016) 23414. https://doi.org/10.2139/ssrn.2795884.

[19] E. Eyceyurt, Y. Egi & J. Zec, "Machine-learning-based uplink throughput prediction from physical layer measurements", Electronics, **11** (2022) 1227. https://doi.org/10.3390/electronics11081227.

[20] Y. Nkansah-Gyekye, *An intelligent vertical handoff decision algorithm in next-generation wireless networks*, Ph.D. dissertation, Department of Computer Science, University of the Western Cape, South Africa, 2010. https://core.ac.uk/download/pdf/58913821.pdf.

[21] T. A. Olukan, Y.C. Chiou, C. H. Chiu, C.Y. Lai, S. Santos & M. Chiesa, "Predicting the suitability of lateritic soil type for low-cost sustainable housing with image recognition and machine learning techniques", J. Build. Eng. **29** (2020) 101175. https://doi.org/10.1016/j.jobe.2020.101175.

[22] W. A. Awad & S. M. Elseuofi, "Machine learning methods for e-mail classification", Int. J. Comput. Appl. **16** (2011) 39. https://doi.org/10.5120/1974-2646.

[23] T. S. Guzella & W. M. Caminhas, "A review of machine learning approaches to spam filtering", Expert Syst. Appl. **36** (2009) 10206. doi: https://doi.org/10.1016/j.eswa.2009.02.037.

[24] O. A. S. Carpinteiro, I. Lima, J. M. C. Assis, A. C. Z. De Souza, E. M. Moreira & C. A. M. Pinheiro, "A neural model in anti-spam systems", in artificial neural networks – ICANN **4132** (2006) 847. https://doi.org/10.1007/11840930_88.

[25] H. Zhang, A. C. Berg, M. Maire & J. Malik, "SVM-KNN: Discriminative nearest neighbor classification for visual category recognition", in 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition **2** (2006) 2126. https://doi.org/10.1109/CVPR.2006.301.

[26] A. Ekong, B. Ekong & A. Eden. Supervised machine learning model for effective classification of patients with COVID-19 symptoms based on Bayesian belief network", Researchers Journal of Science and Technology **2** (2022) 27. https://www.rejost.com.ng/index.php/home/article/view/14.

[27] G. G. James, E. G. Chukwu & P. O. Ekwe, "Design of an intelligent based system for the diagnosis of lung cancer", Int. J. Innov. Sci. Res. Technol. **8** (2023) 791. https://www.ijisrt.com/volume-8-2023_issue-6-june.

[28] G. G. James, A. E. Okpako, C. Ituma & J. E. Asuquo, "Development of hybrid intelligent based information retreival technique", Int. J. Comput. Appl. **184** (2022) 13. https://doi.org/10.5120/ijca2022922401.

[29] C. Ituma, G. G. James & F. U. Onu, "A neuro-fuzzy based document tracking & classification system", Int. J. Eng. Appl. Sci. Technol. **4** (2020) 414. https://doi.org/10.33564/IJEAST.2020.v04-i10.075.

[30] G. G. James, U. A. Umoh, U. G. Inyang & O. M. Ben, "File allocation in a distributed processing environment using gabriel's allocation models", Int. J. Eng. Tech. Math **5** (2012) 56. https://www.researchgate.net/publication/378372640_File_Allocation_in_a_Distributed_Processing_Environment_Using_Gabriel%27s_Allocation_Model.

[31] A. P. Ekong, G. G. James & I. Ohaeri, "Oil and gas pipeline leakage detection using iot and deep learning algorithm" **6** (2024) 421. https://doi.org/10.51519/journalisi.v6i1.652.

[32] G. James, A. Ekong & H. Odikwa, "Intelligent model for the early detection of breast cancer using fine needle aspiration of breast mass.", Int. J. Res. Innov. Appl. Sci. **4** (2024) 348. https://doi.org/10.51584/IJRIAS.2024.90332.

[33] C. Ituma, G. G. James & F. U. Onu, "Implementation of intelligent document retrieval model using neuro-fuzzy technology", Int. J. Eng. Appl. Sci. Technol. **4** (2020) 65. https://doi.org/10.33564/IJEAST.2020.v04i10.013.

[34] G. G. James, G. J. Ekanem, E. A. Okon & O. M. Ben, "The design of e-cash transfer system for modern bank using generic algorithm", International journal of science and technology research **9** (2012) 47. https://www.ajol.info/index.php/jcsia/article/view/15%203911.

[35] G. G. James, A.E. Okpako & J.N. Ndunagu, "Fuzzy cluster means algorithm for the diagnosis of confusable disease", Journal of Computer Science and Its Application **23** (2017) 234. https://www.ajol.info/index.php/jcsia/article/view/153911.

[36] F. U. Onu, P. U. Osisikankwu, C. E. Madubuike & G. G. James, "Impacts of object oriented programming on web application development", Int. J. Comput. Appl. Technol. Res. **4** (2015) 706. https://ijirt.org/Issue?volume=4&issue=9&month=February%202018.

[37] U. A. Umoh, A. A. Umoh, G. G. James, U. U. Oton & J. J. Udoudo, B.Eng., "Design of pattern recognition system for the diagnosis of gonorrhea disease", International Journal of Scientific & Technology Research **1** (2012) 74. http://www.ijstr.org/archive.php.