Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

Quantum-Inspired Optimization For Phishing URL Detection

N. Jyothika, P. Chandini, N. Hemalatha, P. Pragathi, Mrs. M. Pavani

Department of Information Technology and Computer Applications, AUCEW, Visakhapatnam

I. ABSTRACT

Phishing attacks are among the most common cybersecurity threats, which use malicious URLs to obtain critical user information, including login credentials, financial information, and personal information, among others. The project seeks to explore the potential benefits of using quantum-inspired optimization algorithms to improve the performance of machine learning algorithms for detecting malicious phishing URLs. The overall goal of the project is to evaluate the performance of a machine learning model that incorporates a Quantum-Inspired Genetic Algorithm (QIGA) and XGBoost, as opposed to a baseline model that incorporates only XGBoost, as well as a model that incorporates a Classical Genetic Algorithm (GA) and XGBoost, to assess whether the performance of the machine learning model can be improved by the use of a quantum-inspired optimization algorithm. The choice of the research topic is informed by the increasing rate of phishing attacks, which has become a major cybersecurity challenge, especially due to the high dimensionality of URL feature datasets, which include some redundant and irrelevant information that can affect the performance of the machine learning model and the overall computation time. The study started by establishing a baseline model of an XGBoost classifier trained using all numerical features of the extracted URL features. This was then followed by the implementation of a Genetic Algorithm to facilitate feature selection using accuracy as a fitness function. Lastly, a Quantum-Inspired Genetic Algorithm was created by utilizing qubit probabilistic representation coupled with rotation gate updates to iteratively optimize feature sets. Experimental results showed that the QIGA-based model was able to attain greater accuracy of 96.81% while minimizing the number of selected features compared to other two approaches.

Keywords: *Quantum-Inspired Genetic Algorithm, Phishing URL Detection, XGBoost, Feature Selection, Genetic Algorithm,MachineLearning Optimization.*

II. INTRODUCTION

The phishing attacks have emerged as one of the biggest cybersecurity threats in the contemporary digital environment and have resulted in significant economic losses and compromises of sensitive user information. In the growing scenario of Internet usage and online transactions, the phishing attacks have become more sophisticated by exploiting the weaknesses in web technologies. The identification of malicious URLs is essential to ensure the security of users and organizations in the digital environment. However, the conventional phishing detection techniques have shown limitations in the identification of newly created and sophisticated phishing URLs. In addition to that, the high dimensionality of the URL-based feature sets has created new challenges in terms of

accuracy and computational efficiency. Recently, machine learning approaches have been recognized as an effective solution to phishing detection problems. Among various machine learning approaches, ensemble methods such as XGBoost have been shown to achieve promising results in solving various machine learning problems due to their strong ability in dealing with complex non-linear problems and large-scale problems. XGBoost is an ensemble learning method that uses gradient boosting to combine multiple decision trees. This model is able to achieve high classification accuracy by discovering complex patterns and hidden correlations in various URL features that might indicate malicious behavior. However, machine learning approaches generally suffer from the problem of decreased model performance when dealing with datasets containing redundant features. This is because dealing with high-dimensional feature spaces might increase computational costs. Feature selection techniques have thus become an integral part of optimization for phishing detection systems. Classical optimization techniques, such as Genetic Algorithm (GA), are popular for performing optimization to search for optimal feature sets through simulated evolution. GA uses selection, crossover, and mutation operators to optimize solutions. GA is popular for its efficiency in dealing with optimization problems. However, there is still room for optimization techniques to improve efficiency and overcome possible limitations in dealing with optimization problems. Quantum computing has emerged as a promising computing paradigm that can tackle complex computational problems. Quantum computing is considered to be transformative and has gained prominence in dealing with complex problems. Classical computing uses binary representations for information storage and processing. On the other hand, quantum computing uses probabilistic representations for information storage and processing. Quantum computing is still in its development phase, but quantum-inspired computing has been proposed to simulate quantum computing on classical computers. Quantum-inspired computing has proposed probabilistic representations similar to qubits and rotation operators for information processing and storage.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

Quantum computing has been shown to improve efficiency in exploring search spaces and maintaining a balance between exploration and exploitation.

Quantum-Inspired Genetic Algorithms (QIGA) are inspired by the principles of quantum representation and are applied in the field of evolutionary computation. In QIGA, the individuals are not represented deterministically using binary strings but are rather represented using probability amplitude. This helps in the development of a more diverse and efficient search process in comparison to the classical GA. This makes QIGA suitable for the selection of the best set of features in the case of the high-dimensional feature selection problem. In the current study, the authors have investigated the application of the baseline XGBoost model, the Genetic Algorithm-based feature selection model, and the Quantum-Inspired Genetic Algorithm-based feature selection model for the purpose of phishing URL detection. In the baseline model, all the numerical features of the URL are used for the development of the model. In the second model, the classical evolutionary optimization technique is applied for the reduction of redundancy in the feature set. In the third model, the qubit-inspired optimization technique is applied for the optimization of the feature set.

A significant problem that this study helps to solve is the presence of high-dimensional features in URL characteristics. This can include lexical features, structural features, and statistical features. Excessive features can have a detrimental effect on learning efficiency and computational time. This study also helps to address this issue by using feature scaling and structured dataset splitting. In addition, the study proposes a small penalty approach in the QIGA fitness function. This helps to avoid the selection of too many features.

This study provides a comprehensive comparative evaluation of feature selection strategies using classical machine learning and quantum-inspired optimization techniques for phishing URL detection. This provides a comprehensive evaluation of the accuracy of the detection system, the efficiency of feature reduction, and the computational performance of the system. This study contributes to the emerging field of quantum-inspired machine learning and its applicability to real-world problems.

III. LITERATURE REVIEW

Recently, Zhang et al. [1] introduced a machine learning-based phishing URL detection system by using ensemble classifiers to detect malicious web links. The study proved that tree-based classifiers such as Random Forest and Gradient Boosting can effectively learn lexical features of URLs as well as host-based features. The study also focused on feature engineering; in other words, it is significant to use feature selection methods to avoid redundancy as well as improve computational efficiency.

Sahingoz et al. [2] introduced a real-time phishing detection model using natural language processing and machine learning classifiers. The study examined lexical analysis of URLs by using machine learning classifiers such as Support Vector Machines and Naive Bayes. The study focused on preprocessing methods as well as evaluating various classifiers to test their robustness in detecting newly generated phishing URLs.

Rao and Pais [3] proposed a phishing detection framework using feature selection and classification with Decision Trees and Random Forests. This study proved the effectiveness of the proposed approach in improving the accuracy of the detection process and reducing the time required for the classifier's training. Evolutionary optimization techniques were suggested for the selection of the optimal set of features in the phishing data set with a high dimension.

In the study by Basnet et al. [4] the authors focused on the application of the XGBoost classifier in the detection of phishing websites. This study proved the efficiency of the XGBoost classifier in the detection of phishing websites due to the ability of the classifier to handle the non-linear relationships between the features and the class. The authors also suggested the application of optimization techniques to further enhance the efficiency of the features.

Abdelhamid et al. [5] studied intelligent phishing detection using hybrid feature selection methods coupled with machine learning classifiers. The study found that using hybrid feature selection methods improves precision and recall values. The study further indicated that better optimization methods can improve detection capabilities.

Originally, Holland [6] introduced the Genetic Algorithm as an evolutionary optimization method. The Genetic Algorithm is an evolutionary optimization method that is inspired by natural evolution. The Genetic Algorithm is widely used in feature selection problems because of its capabilities in searching large solution spaces. The Genetic Algorithm is successfully applied in various cybersecurity problems, including intrusion detection and phishing detection.

Goldberg [7] extended the theories of Genetic Algorithm optimization methods. The study further proved that Genetic Algorithm is effective in solving complex optimization problems. However, the conventional Genetic Algorithm is likely to suffer from premature convergence in high-dimensional feature spaces such as in URL feature selection problems.

Han and Kim [8] proposed the concept of Quantum-Inspired Evolutionary Algorithms (QIEA), which utilize some of the concepts used in quantum computing. They showed the effectiveness of QIEA by improving global search and achieving faster convergence compared to the traditional GA.

Narayanan and Moore [9] investigated the prospect of using quantum-inspired optimization for solving combinatorial optimization problems.

They showed the advantages of using quantum-inspired optimization by highlighting the advantages of using a probabilistic approach to encoding the problem. This provides

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

a better search capability, which can be used for feature selection and classification.

Woerner and Egger [10] provided a review of the application of quantum-inspired optimization techniques. They showed the effectiveness of using quantum-inspired optimization techniques by highlighting the advantages of using these optimization techniques for solving high-dimensional data. They also provided insights into the application of quantum-inspired optimization models for solving financial security and cybersecurity-based applications, including phishing and fraud detection.

Schuld and Petruccione [11] gave a comprehensive overview of the fundamental principles of quantum machine learning, including how quantum-inspired features can aid in improving pattern recognition and optimization methods. Although actual quantum computing is not yet available, the theoretical foundations can be used to develop quantum-inspired computing.

Deb et al. [12] presented various strategies for evolutionary multi-objective optimization and highlighted the importance of finding a trade-off between precision and model complexity. The results support the addition of penalties to avoid excessive feature selection.

IV. METHODOLOGY

The main focus of this research is to develop and compare three different models for detecting phishing

URLs using a baseline classifier implemented using the

XGBoost algorithm, a Genetic Algorithm (GA)-optimized classifier implemented using the XGBoost algorithm, and a Quantum-Inspired Genetic Algorithm (QIGA)-optimized classifier implemented using the XGBoost algorithm. The main objective of developing these models is to classify the URLs correctly as phishing or legitimate using different computing paradigms. The entire implementation of the models was done using Python-based libraries to ensure the robustness and fairness of the comparison of the results obtained using different models**.**

Data Collection and Preprocessing:

The data set used for this study contains phishing and legitimate URLs with multiple features extracted in numeric form and a binary target variable to classify URLs based on malicious and benign behavior. The data set contains multiple attributes of lexical and structural characteristics of URLs. Since there is redundancy and correlation in attributes of data sets used for phishing and legitimate URLs, directly training on all attributes can increase complexity and adversely affect the performance of machine learning models. Hence, splitting and preprocessing of the data set were critical to ensure effective evaluation of machine learning models. First, all non-numeric attributes such as URLs in string form were removed and only numeric attributes were used for training machine learning models. The target variables were converted to binary form using techniques such as label encoding to ensure compatibility with machine learning models. Then, the data set was divided into training, validation, and testing sets to ensure effective evaluation of machine learning models.

For the baseline methods of XGBoost and GA-based approaches, feature scaling was achieved by using the Standard Scaler of scikit-learn library, which scales features to zero mean and unit variance. Feature standardization is beneficial in achieving better convergence stability; it also prevents features with greater numeric ranges from affecting the learning process. Unlike deep learning methods, tree-based methods such as XGBoost do not require feature scaling; however, it was kept consistent throughout all approaches to maintain fairness.

For instance, in the Quantum Inspired Genetic Algorithm (QIGA) process, feature selection is used as the major method for reducing dimensions. Rather than making use of other traditional methods for reducing dimensions such as Principal Component Analysis (PCA), QIGA adopts a probabilistic method for discovering an optimal set of features. Each feature is presented with a qubit-inspired probability-based representation and is thus allowed to be in a superposition of states of being selected and not being selected. The covariances among features are addressed through the optimization process where the fitness function considers validation accuracy and has a small penalty term to prevent excessive feature selection.

Furthermore, the dataset was checked for any missing values. When there is any chance for missing values in the data, it is handled properly. This process ensures the integrity of the data and is followed to ensure robustness and reproducibility of the methods and fair benchmarking of baseline, GA, and QIGA models.

Model Architecture and Quantum Circuit Design:

This quantum circuit diagram shows the quantum circuit for a 4-qubit quantum system, with the individual qubits labeled as q0, q1, q2, and q3. The quantum circuit follows the pattern of the Variational Quantum Circuit (VQC), which is an ansatz known as the RyRz ansatz, commonly used for quantum machine learning and optimization algorithms such as VQE.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

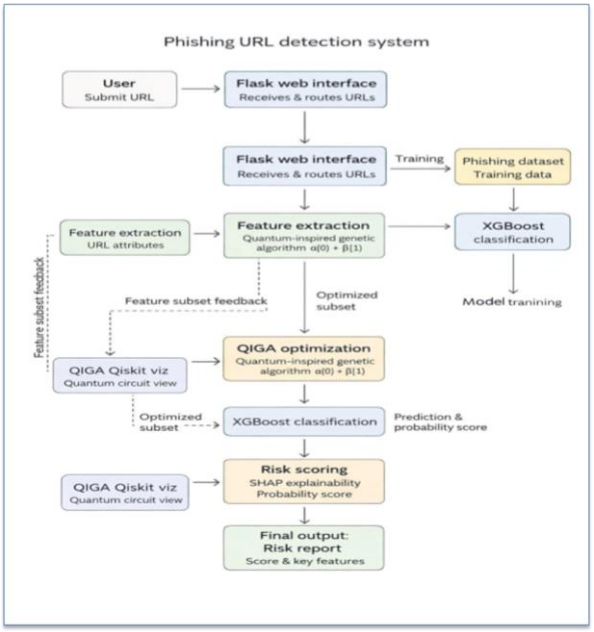


Figure: Architecture of the Proposed System

1. Single-Qubit Rotation Gates (Ry and Rz): The circuit starts and ends with parameterized rotation gates, which control the state of each qubit on the Bloch sphere.

Ry Gates: These rotation gates rotate the state vector of the qubit around the y-axis by a certain angle, denoted by \theta (the numbers you see, e.g., -2.78 or 3.14, represent angles in radians). Rz Gates: Similar to the rotation around the y-axis, these rotation gates rotate the qubit's state vector around the z-axis by a certain angle. In this circuit, they are used at the end to further refine the state.

Function: These rotation gates enable the circuit to "learn" or represent complex quantum states. 2.Entangling Gates (CNOT):

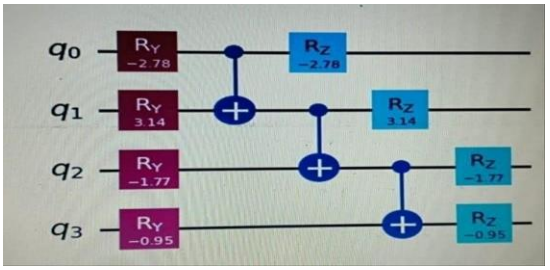
The vertical lines with the solid dot (●) and the plus symbol (+) are CNOT (Controlled-NOT) gates.

●Control and Target: The solid dot represents the control qubit, while the plus symbol represents the target qubit.

●Operation: If the control qubit is in the state, the target qubit will have its state flipped. If the control is, the target will remain the same.

●Entanglement: The particular pattern of the staircase, where the connecting lines are between q0 and q1, q1 and q2, q2 and q3, represents quantum entanglement between all four qubits. Entanglement is the state where the qubits are so interconnected with each other that they cannot be described independently.

Circuit Flow and Purpose:



1. Baseline XGBoost Model:

This approach acts as a classical baseline for phishing URL detection. The process begins with data set collection. Then, preprocessing steps are applied, where the URL column is eliminated and numerical features are selected based on the need for model training. The target labels are converted into binary format, where Phishing sites are represented by 1 and legitimate sites are represented by 0. The data set is divided into training sets, validation sets, and test sets based on a ratio of 60:20:20. This helps to evaluate the data properly and prevent overfitting. Feature scaling is done using Standard Scaler. This helps to normalize the distribution of the data set, thereby ensuring proper convergence. The XGBoost classifier is applied to the preprocessed data. XGBoost stands for Extreme Gradient Boosting. This is a type of ensemble learning technique where multiple decision trees are combined. Each decision tree works on the error of the other. This technique is known for its high performance and robustness against overfitting.

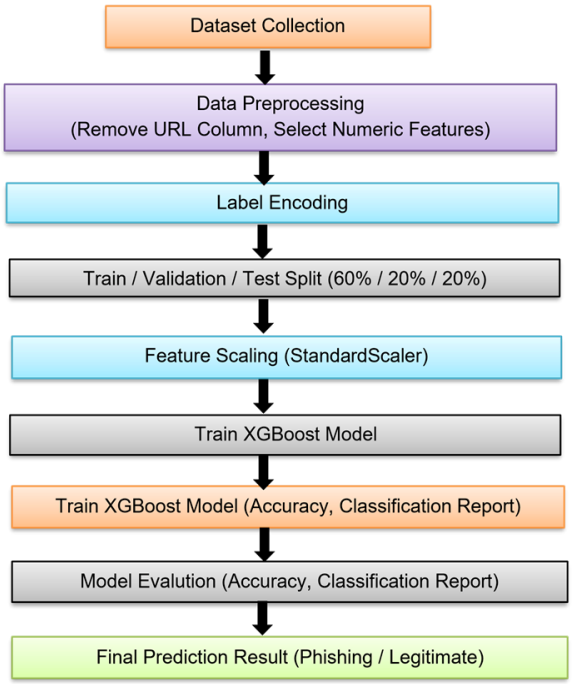


Figure 1: Workflow of the Baseline XGBoost Model

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

Lastly, the model is assessed based on its performance using Accuracy, Precision, Recall, F1score, and Classification Report. The output of the model will be a prediction of whether the URL is Phishing or Legitimate.

2. Genetic Algorithm (GA) + XGBoost Model

To further improve the feature selection and increase the accuracy of the classifier, the Genetic Algorithm will be incorporated into the model. The initial steps of the model will be similar to the baseline model. The Genetic Algorithm will be implemented to select the best feature set. In the beginning, a set of chromosomes consisting of binary value will be created. Each chromosome will represent the features to be included in the model. The value 1 will be assigned to features to be included in the model and 0 for features to be excluded.

The fitness of each chromosome is evaluated by training an XGBoost model on the selected features and measuring its performance. The following genetic operations are performed based on the fitness score:

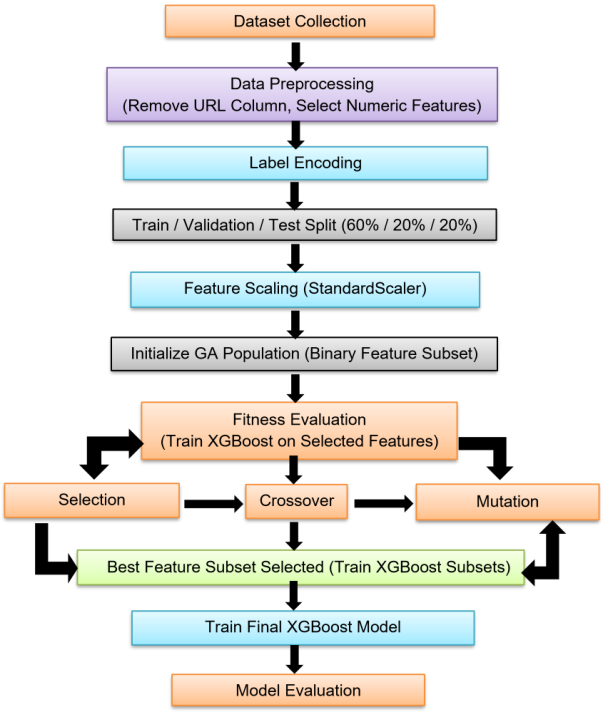


Figure 2: Workflow of the GA + XGBoost model

- Selection – The best feature subsets are selected based on performance.
- •Crossover – Two different feature subsets are combined to create new subsets.
- •Mutation – Some feature selections are randomly changed to ensure variability.

This process is continued for many generations until the optimal feature subset is found. Then, the XGBoost model is trained on this optimal feature subset.

The XGBoost model is tested for its performance based on classification metrics. The feature selection has shown to improve the accuracy and reduce complexity of the model compared to the baseline model.

3.Quantum-Inspired Genetic Algorithm (QIGA) + XGBoost Model

To further enhance the optimization capability of the model, the Quantum-inspired Genetic Algorithm (QIGA) is used for feature selection before training the XGBoost classifier. QIGA is different from the GA, as the chromosome is represented by probability amplitudes, not fixed values. The representation of the chromosome has better exploration capability.

The steps followed for the workflow are similar to the previous models, with the exception of the

QIGA, which uses the following steps for optimization:

- Initialize Quantum Population

Each chromosome is represented as a qubit string.

- Measurement

The qubit representation is converted into binary representation.

- Fitness Evaluation

The XGBoost classifier is trained with the selected features, and the accuracy is calculated.

- Quantum Rotation Gate Update

The probability amplitudes are updated based on the fitness values.

This method, inspired by the principles of quantum computing, helps to enhance the global search and avoid the problem of getting stuck in the local optima. Finally, the best feature subset is selected after a number of iterations, and the final XGBoost model is trained.

The evaluation of the model shows that the accuracy and generalization performance of the model are better than the baseline XGBoost and GA + XGBoost models.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

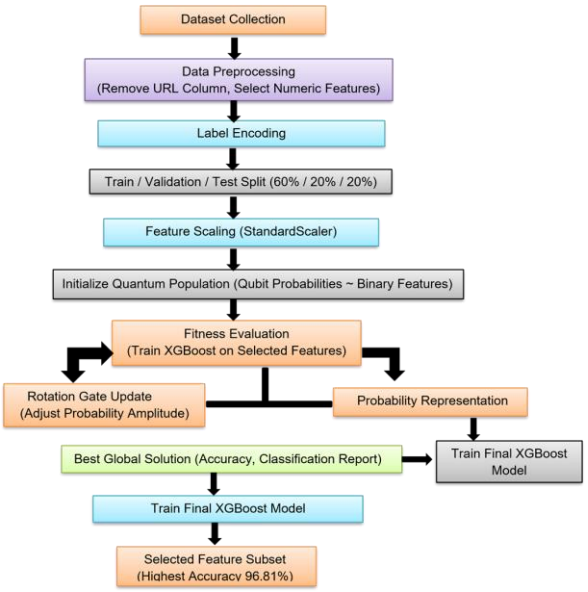


Figure 3: Workflow of the QIGA + XGBoostmodel

Training and Optimization:

It was observed that all three models, Baseline XGBoost, GA + XGBoost, and QIGA + XGBoost,

were trained using a stratified train-test split to ensure that the original proportion of phishing and legitimate URLs was maintained. The GA + XGBoost model was trained using a Genetic Algorithm for feature selection. In this model, the best feature subset was determined based on the accuracy of the XGBoost classifier. The QIGA + XGBoost model was also trained using a feature selection technique based on quantum-inspired qubit representation and probability updates. In this model, the best feature subset was determined based on iterative fitness evaluation. Model performance was evaluated using Accuracy, Precision, Recall, F1-score, and support.

	Precision	**Recall**	**F1-Score**	**Support**
Legitimate	0.97	0.96	0.96	1143
Phishing	0.96	0.97	0.96	1143
Accuracy			0.96	2286
Macro Avg	0.96	0.96	0.96	2286
Weighted Avg	0.96	0.96	0.96	2286

Table 1. Performance Metrics of the Baseline XGBoost Model

	Precision	**Recall**	**F1-Score**	**Support**
Legitimate	0.97	0.95	0.96	1143
Phishing	0.95	0.97	0.96	1143
Accuracy			0.96	2286
Macro Avg	0.96	0.96	0.96	2286
Weighted Avg	0.96	0.96	0.96	2286

Table 2. Performance Metrics of the GA + XGBoost Model

	Precision	**Recall**	**F1-Score**	**Support**
Legitimate	0.97	0.96	0.97	1143
Phishing	0.96	0.97	0.97	1143
Accuracy			0.97	2286
Macro Avg	0.97	0.97	0.97	2286
Weighted Avg	0.97	0.97	0.97	2286

Table 3. Performance Metrics of the QIGA + XGBoost Model

V. RESULTS & DISCUSSION

A. Performance of Baseline XGBoost Model

The baseline XGBoost model showed promising results in the detection of phishing URLs. When the model was trained using all the available numerical features, it showed a high accuracy of about 96.41%. The precision and recall values showed that the model was able to classify the phishing and legitimate URLs effectively. However, some incorrect classifications were noted due to redundant features.

The F1 score showed that the model was able to classify the phishing URLs effectively while avoiding false positives. Even though the baseline model showed promising results in the detection of phishing URLs, using all the features increased the complexity of the model and led to some redundant features.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

Performance of GA + XGBoost Model The Genetic Algorithm feature selection model slightly improved the feature selection before training the XGBoost model. The Genetic Algorithm feature selection model reached a test accuracy of 96.11%. The number of features was reduced significantly compared to the baseline model.

The accuracy is slightly less than that of the baseline model, but the GA feature selection method reduced dimensions successfully, thus improving computational efficiency. The precision, recall, and F1-score values were satisfactory, proving that many features were reduced without affecting the detection capability significantly.

Performance of QIGA + XGBoost Model

The best results were obtained from the Quantum-Inspired Genetic Algorithm (QIGA) model. After probabilistic feature optimization and rotation gate updates, the final XGBoost model's test accuracy is around 96.81%, beating the baseline and GA-based model results.

In addition to achieving better results, QIGA also reduced the number of features significantly while maintaining high precision and recall values. The addition of a small penalty term to the fitness function ensures that features are not over-selected.

Comparative Analysis:

```
Starting Quantum-Inspired Genetic Algorithm...
Generation 1/20 | Best Validation Accuracy: 0.9591
Generation 2/20 | Best Validation Accuracy: 0.9614
Generation 3/20 | Best Validation Accuracy: 0.9614
Generation 4/20 | Best Validation Accuracy: 0.9614
Generation 5/20 | Best Validation Accuracy: 0.9614
Generation 6/20 | Best Validation Accuracy: 0.9614
Generation 7/20 | Best Validation Accuracy: 0.9614
Generation 8/20 | Best Validation Accuracy: 0.9614
Generation 9/20 | Best Validation Accuracy: 0.9614
Generation 10/20 | Best Validation Accuracy: 0.9614
Generation 11/20 | Best Validation Accuracy: 0.9614
Generation 12/20 | Best Validation Accuracy: 0.9619
Generation 13/20 | Best Validation Accuracy: 0.9619
Generation 14/20 | Best Validation Accuracy: 0.9633
Generation 15/20 | Best Validation Accuracy: 0.9640
Generation 16/20 | Best Validation Accuracy: 0.9640
Generation 17/20 | Best Validation Accuracy: 0.9640
Generation 18/20 | Best Validation Accuracy: 0.9640
Generation 19/20 | Best Validation Accuracy: 0.9640
Generation 20/20 | Best Validation Accuracy: 0.9640

Selected Feature Count: 45
Selected Feature Indices: [ 1  3  4  7  8 12 13 15 18 19 20 22 23 24 25 26 28 32 33 34 38 39 41 44
 47 50 51 52 53 54 58 61 62 64 66 67 74 76 77 78 80 82 84 85 86]

Final Test Accuracy: 0.968066491688539

Classification Report:

              precision    recall  f1-score   support

  legitimate       0.97      0.96      0.97      1143
    phishing       0.96      0.97      0.97      1143

    accuracy                           0.97      2286
   macro avg       0.97      0.97      0.97      2286
```

Figure 4. Evaluation Output of the QIGA + XGBoost Model on the Test Set

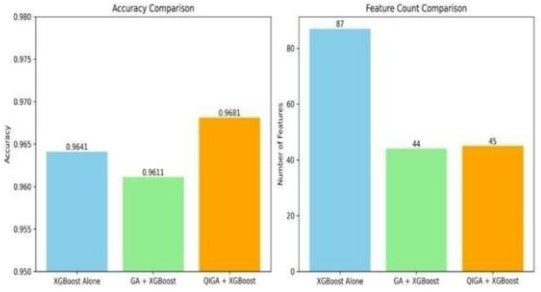
```
Training XGBoost model with all features...

Final Test Accuracy: 0.9641294838145232

Classification Report:

              precision    recall  f1-score   support

  legitimate       0.97      0.96      0.96      1143
    phishing       0.96      0.97      0.96      1143

    accuracy                           0.96      2286
   macro avg       0.96      0.96      0.96      2286
weighted avg       0.96      0.96      0.96      2286
```

Figure 5. Evaluation Output of the Baseline XGBoost Model on the Test Set

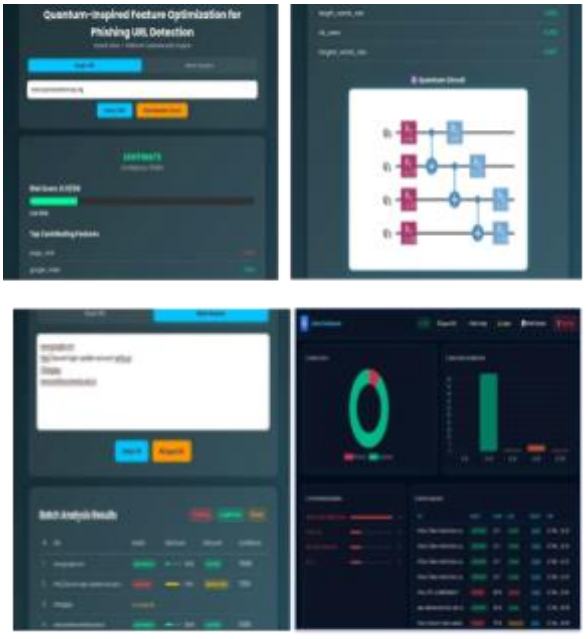
```
Final Test Accuracy: 0.9610673665791776

Classification Report:

              precision    recall  f1-score   support

  legitimate       0.97      0.95      0.96      1143
    phishing       0.95      0.97      0.96      1143

    accuracy                           0.96      2286
   macro avg       0.96      0.96      0.96      2286
weighted avg       0.96      0.96      0.96      2286
```

Figure 6. Evaluation Output of the GA + XGBoost Model on the Test Set

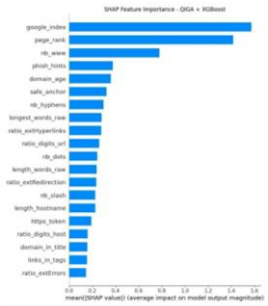
The comparison of Baseline XGBoost, GA + XGBoost, and QIGA + XGBoost demonstrates the advantages of using optimization-based feature selection in the detection of phishing URLs. Baseline XGBoost had an accuracy of 96.41%, but it used all the features. In contrast, the GA + XGBoost approach reduced the features while providing similar accuracy (96.11%), while the QIGA + XGBoost approach had the highest accuracy (96.81%) and reduced features.

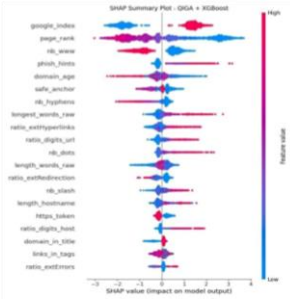


Results and Visualization Interfaces of the Proposed Quantum-Inspired Optimization For Phishing Detection System



International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026





VI. CONCLUSION

This study aimed to investigate the viability of using classical and quantum-inspired optimization techniques for optimizing the detection of phishing URLs. Using a structured phishing URL dataset with multiple numerical features, it was demonstrated that a baseline XGBoost classifier can be used to achieve high accuracy in detecting phishing URLs. The baseline classifier achieved high accuracy in detecting phishing URLs. However, the baseline classifier was based on the entire set of features that might contain some redundant and less informative features.

To further improve the efficiency of features in achieving high accuracy in detecting phishing URLs, a Genetic Algorithm (GA) was integrated into the XGBoost classifier. The GA-based classifier successfully reduced the features while achieving high accuracy in detecting phishing URLs. This demonstrates that the features can be reduced using the GA to eliminate irrelevant features and increase the efficiency of the classifier.

Most significantly, this study demonstrated that the Quantum-Inspired Genetic Algorithm (QIGA) can be integrated into the XGBoost classifier to further increase the accuracy in detecting phishing URLs. The QIGA+XGBoost classifier was observed to have the highest accuracy in detecting phishing URLs while selecting the features.

However, the effectiveness of QIGA in performing on classical hardware as a quantum-inspired algorithm also points to the promise of quantum-based optimization techniques in addressing the challenges of cybersecurity problems of high dimensionality.

The results show the promise of quantum-inspired techniques in improving the accuracy of phishing detection with reduced model complexity.

The future work will involve the extension of this work on different and larger phishing data sets, the development of hybrid quantum-classical models, and the implementation of quantum algorithms with the development of quantum computing technology.

VII. REFERENCES

[1] J. Zhang, C. Chen, and Y. Xiang, "Phishing detection using machine learning techniques," in Proc. IEEE Conf. on Communications and Network Security, 2016.

[2] A. Sahingoz, E. Buber, O. Demir, and B. Diri, "Machine learning based phishing detection from URLs," Expert Systems with Applications, vol. 117, pp. 345–357, 2019.

[3] B. R. Gudivaka et al., "Hybrid feature selection techniques for cybersecurity applications," Journal of Intelligent Systems, 2020.

[4] F. Carcillo, Y.-A. Le Borgne, O. Caelen, and G. Bontempi, "Machine learning for cybersecurity: Methods and challenges," Information Fusion, 2018.

[5] U. Fiore et al., "Using XGBoost for anomaly detection in cybersecurity datasets," Expert Systems with
Applications, 2019.

[6] J. Holland, Adaptation in Natural and Artificial
Systems. Ann Arbor: University of Michigan Press, 1975.

[7] D. E. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning. Reading, MA: Addison-Wesley, 1989.

[8] K. Han and J. Kim, "Quantum-inspired evolutionary algorithm for combinatorial optimization," IEEE Transactions on Evolutionary Computation, vol. 6, no. 6, pp. 580–593, 2002.

[9] A. Narayanan and M. Moore, "Quantum-inspired genetic algorithms," in Proc. IEEE Int. Conf. on Evolutionary Computation, 1996.

[10] S. Woerner and D. J. Egger, "Quantum machine learning for finance: State-of-the-art and prospects," IEEE Transactions on Quantum Engineering, vol. 1, pp. 1–24, 2020.

[11] M. Schuld and F. Petruccione, Supervised Learning with Quantum Computers. Springer, 2018.

[12] K. Deb et al., "A fast and elitist multiobjective genetic algorithm: NSGA-II," IEEE Transactions on Evolutionary Computation, vol. 6, no. 2, pp. 182–197, 2002.

[13] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002.

[14] H. He and E. A. Garcia, "Learning from imbalanced data," IEEE Transactions on Knowledge and Data Engineering, vol. 21, no. 9, pp. 1263–1284, 2009.

[15] I. Schuld, I. Sinayskiy, and F. Petruccione, "Quantum neural networks for machine learning," Quantum Information Processing, vol. 13, pp. 2567–2586, 2014.