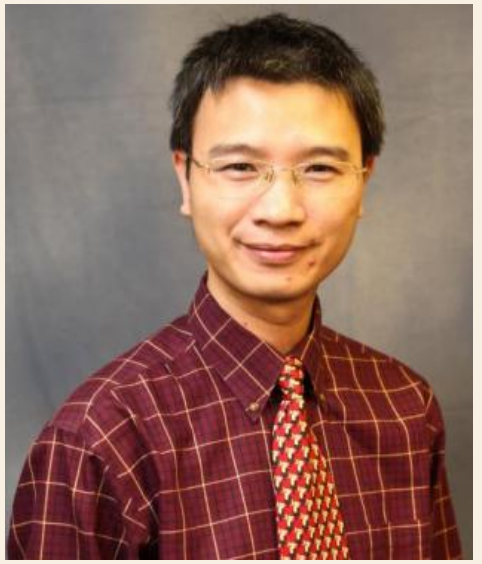
A COMPARATIVE STUDY OF MACHINE LEARNING AND QUANTUM MODELS FOR SPAM EMAIL DETECTION

Authors: Cameron Williams, Taieba Tasnim, Berkeley Wu, Mohammad Rahman, Fan Wu

Tuskegee University Department of Computer Science

PRESENTER



Dr. Fan Wu

fwu@tuskegee.edu

- Working Experience
  - ❖ Head, Computer Science Department, Tuskegee University, Tuskegee, AL
  - ❖ Professor, Computer Science Department, Tuskegee University, Tuskegee, AL
  - ❖ Director, Center of Information Assurance Education (CIAE), Tuskegee, AL
  - ❖ Director, Tuskegee University Office of Undergraduate Research (TUOUR), Tuskegee, AL
- Research Areas
  - ❖ Mobile Security, Information Assurance, Data Science, Machine Learning, Mobile Graphics,
  - ❖ Mobile Computing, Computer Graphics, Bioinformatics, Biostatistics, High Performance
  - ❖ Computing with GPGPU Technology, and Robotics

GOALS AND OBJECTIVES

Compare	Compare traditional, deep learning, and quantum models for spam detection.
Implement	Implement seven algorithms (Naive Bayes, KNN, Logistic Regression, SVM, FNN, CNN, QCNN).
Apply	Apply a consistent preprocessing and evaluation framework.
Measure	Measure performance using Accuracy, Precision, Recall, and F1-Score.
Benchmark	Benchmark QCNNs against classical models to assess real-world potential.
Provide	Provide insights to guide future quantum cybersecurity research.

AGENDA

- Introduction
- Literature Review
- Methodology
- Evaluation Metrics
- Experimental Work
- Conclusion
- Future work
- Acknowledgement
- References
- Results and Discussions

INTRODUCTION

Spam emails remain a major security and productivity concern, often used for phishing, malware distribution, and social engineering attacks. Traditional rule-based filters are no longer sufficient as these threats evolve in complexity. Machine learning offers adaptive and scalable solutions by identifying patterns in large datasets, yet few studies provide fair comparisons of different algorithms under consistent conditions. This research addresses the gap by evaluating seven models, Naive Bayes, K-Nearest Neighbors (KNN), Logistic Regression, Convolutional Neural Network (CNN), Feedforward Neural Network (FNN), Support Vector Machines (SVM), and Quantum Convolutional Neural Network (QCNN) through a standardized preprocessing, training, and evaluation pipeline.

LITERATURE REVIEW

LITERATURE REVIEW

- Spam email detection has evolved from simple rule-based systems in the mid-1990s to advanced machine learning and deep learning models today. Early methods, such as AOL's automated filters (1994) and DNS-based blacklists like MAPS's RBL, blocked known spam sources but lacked adaptability. Rule-based filters relied on keyword lists and scoring systems, which were easily bypassed by spammers. The introduction of machine learning in the 2000s, including Bayesian filters, Support Vector Machines (SVM), and Decision Trees, improved accuracy by learning from labeled datasets. Bayesian filters calculated word probabilities to classify emails, while SVMs and Decision Trees enhanced feature selection and generalization. In recent years, deep learning and NLP-based models, including transformer architectures like BERT, have enabled context-aware detection using email content, metadata, and sender behavior. Stylometric and behavioral analysis further detect phishing and impersonation attempts, while Explainable AI (XAI) improves transparency. Overall, spam detection has progressed toward adaptive, intelligent, and interpretable systems capable of countering evolving spam tactics.

METHODOLOGY

DATA ACQUISITION

- Source: Kaggle spam dataset (~5,700 emails)
- Labels: 745 spam, 4,955 non-spam (moderate imbalance)
- Format: Excel with raw text and binary labels
- Preprocessing: duplicates removed, text lowercased, punctuation & stop-words removed
- Provides a unified benchmark for fair model comparison

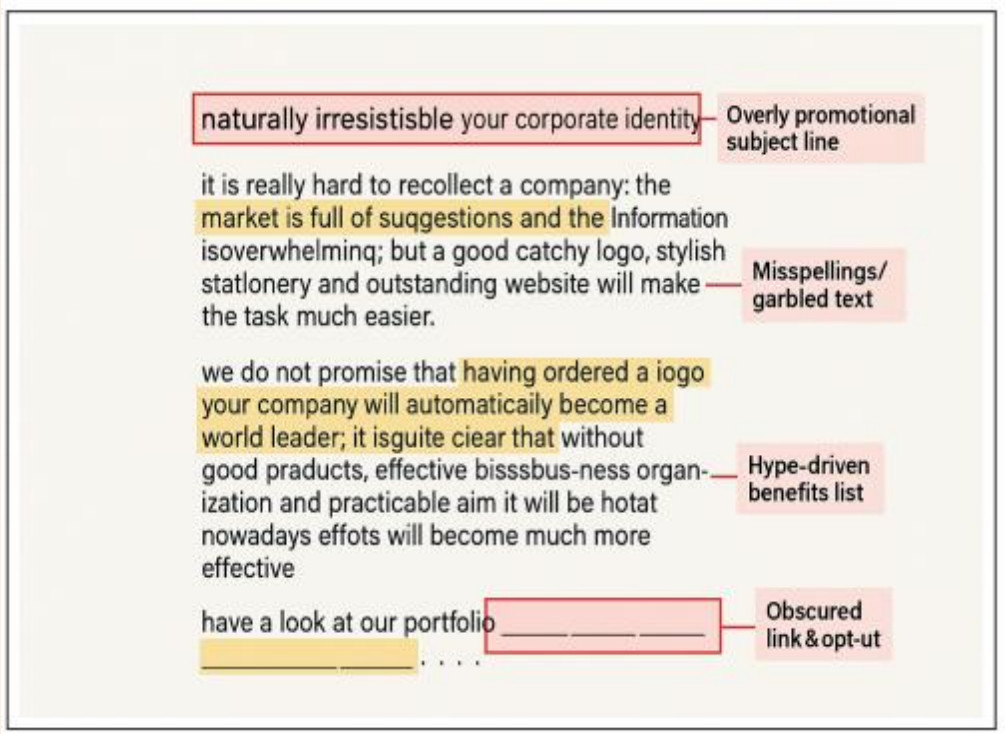


Figure 1. Sample of a potential spam email.

TOOLS & ENVIRONMENT

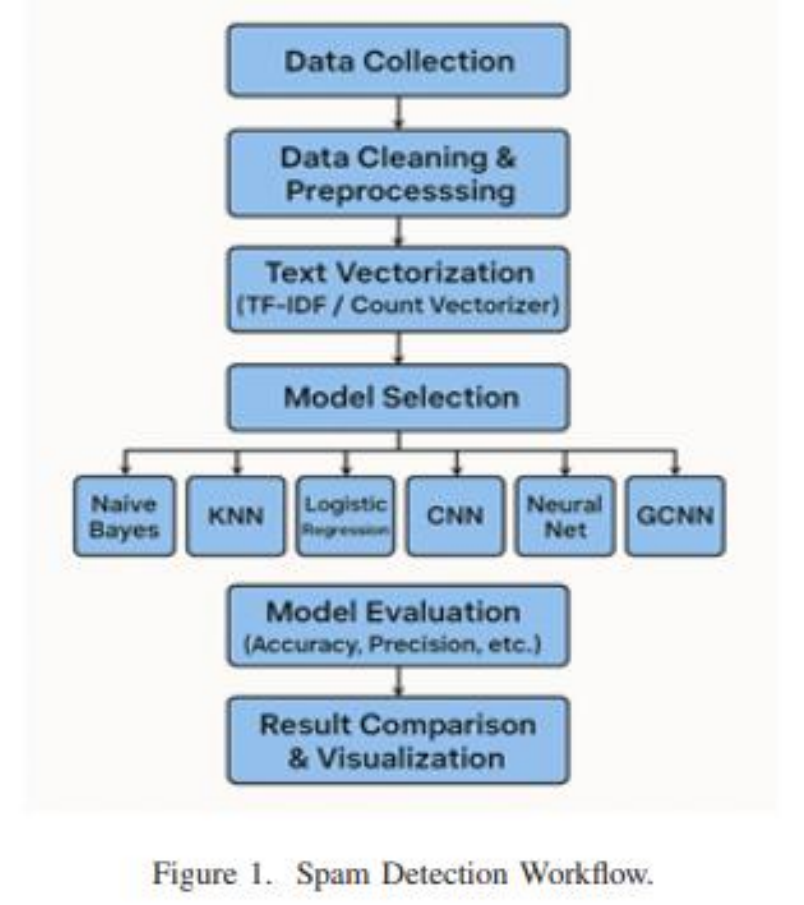


Figure 1. Spam Detection Workflow.

- Google Colab (GPU-enabled)
- Integration with Google Drive for datasets/scripts
- Consistent preprocessing, training & evaluation pipelines
- Workflow:

1. Data collection
2. Preprocessing
3. Vectorization (TF-IDF/Count Vectorizer)
4. Model training (Naive Bayes, KNN, Logistic Regression, SVM, FNN, CNN, QCNN)
5. Performance evaluation & visualization

EVALUATION METRICS

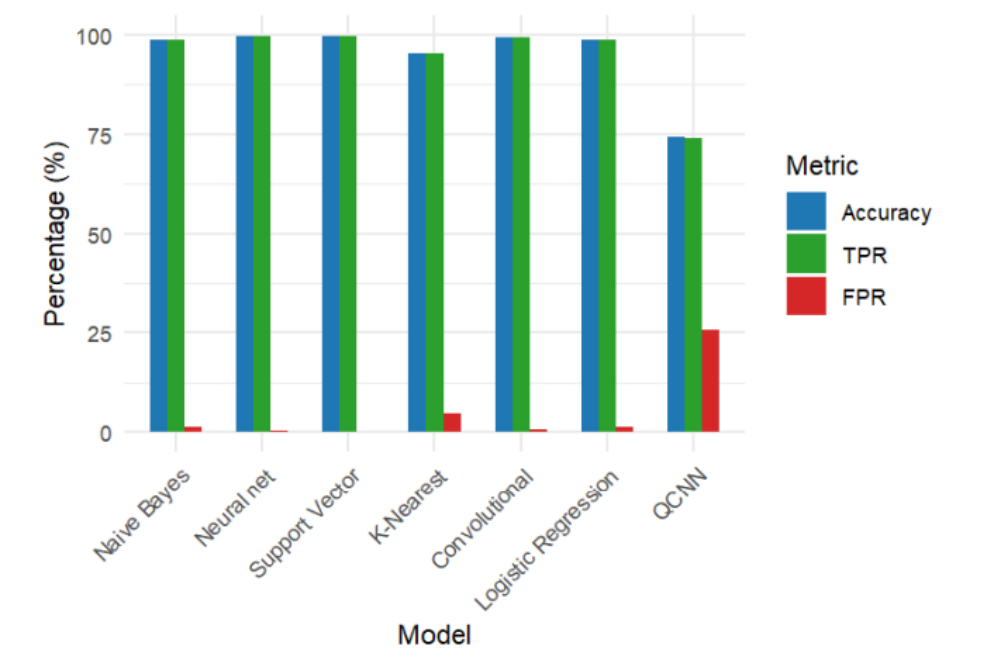
•**Accuracy:**
•Used for overall comparison between models.
•Showed which algorithms classified emails (spam vs. non-spam) most correctly.
•Example: SVM achieved the highest accuracy (99.70%), QCNN the lowest (74.17%).

•**True Positive Rate (TPR) & False Positive Rate (FPR):**
•Helped reveal how models performed on imbalanced data (745 spam vs. 4,955 non-spam).
•TPR showed how well spam was caught, while FPR showed how many legitimate emails were mistakenly flagged.

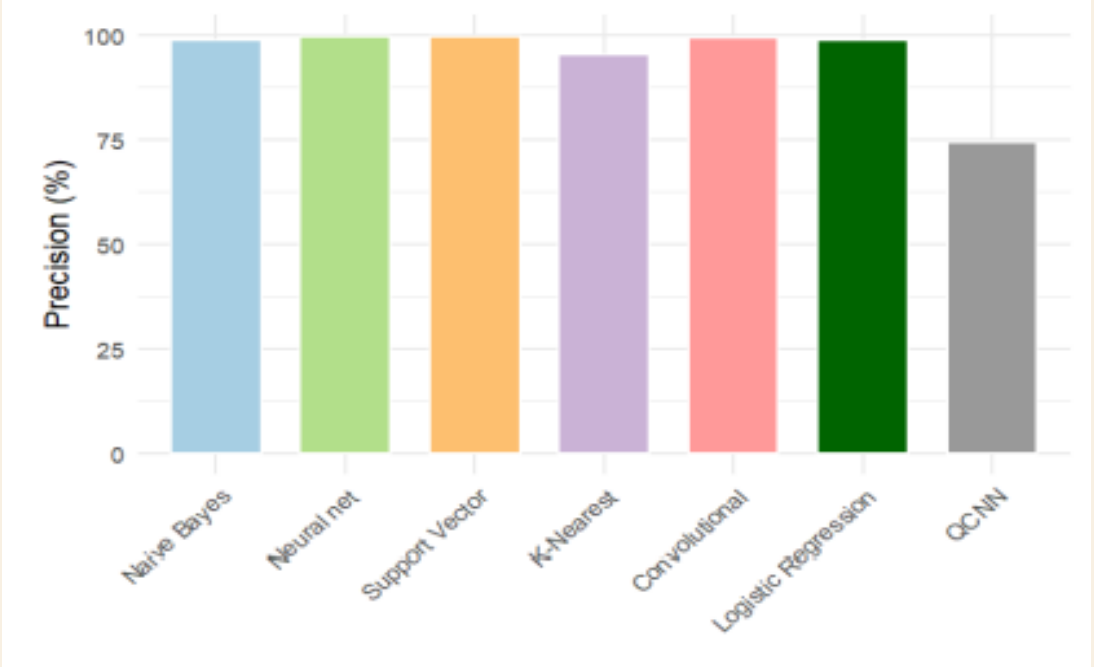
•**Precision & Recall:**
•Important to balance catching spam (Recall) with avoiding false alarms (Precision).
•Naive Bayes and Logistic Regression both scored ~98.67% Precision/Recall, meaning they rarely misclassified.
•CNN, FNN, and SVM performed even better, maintaining near-perfect balance.

•**F1 Score:**
•Used as a single summary metric to evaluate trade-offs between Precision and Recall.
•FNN and SVM had the highest F1-scores (~99.65%–99.69%), showing both completeness and correctness.

RESULTS

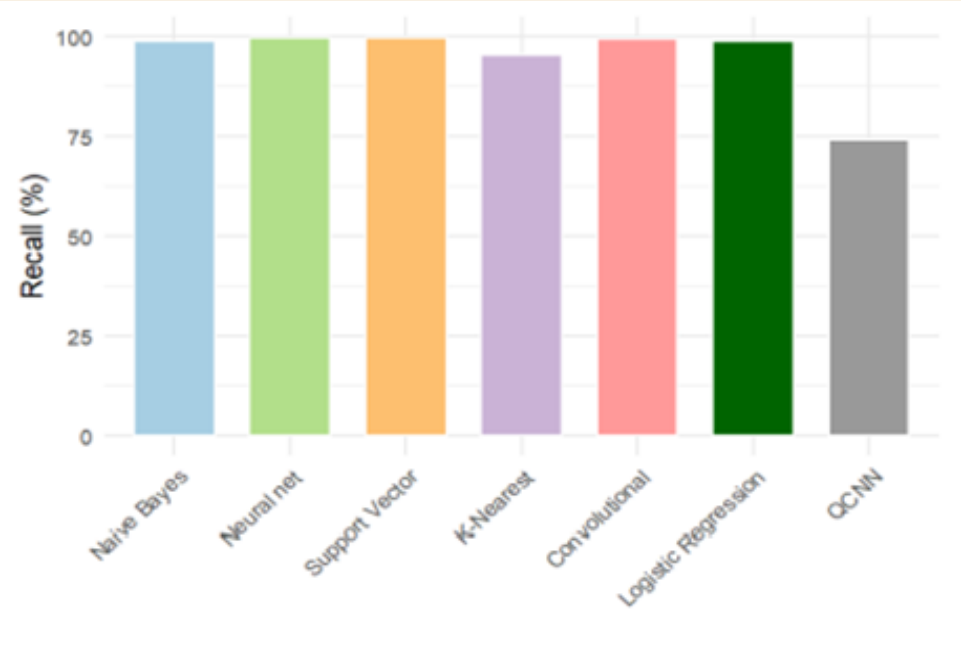


Model performance comparison by accuracy

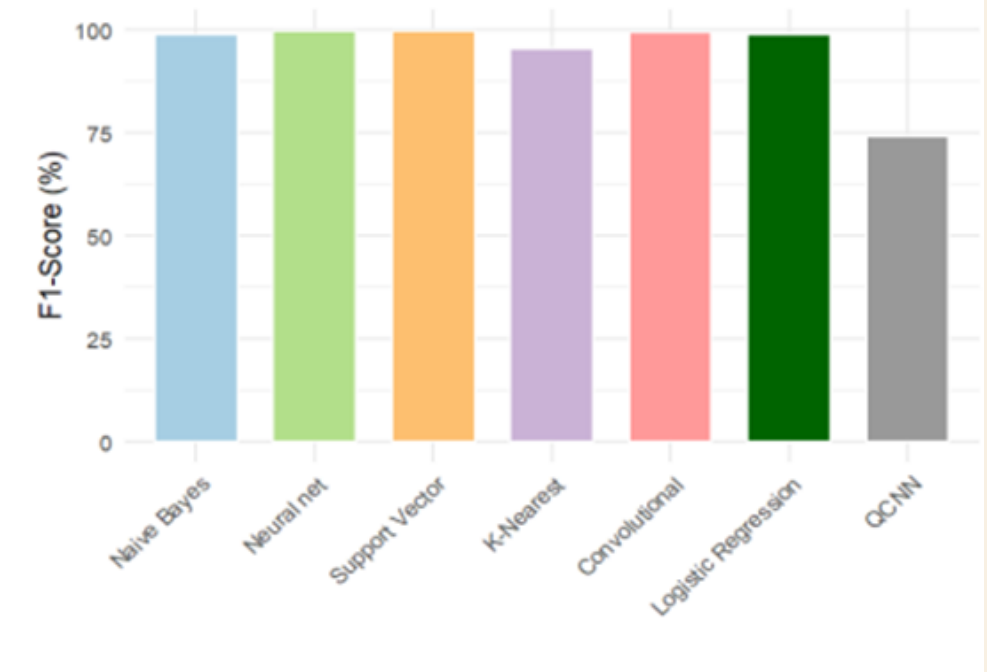


Precision comparison of classification models

RESULTS



Recall comparison of classification models



F1-Score comparison of classification models

EXPERIMENTAL WORK

Experimental Work – Setup

- Preprocessing steps:
  - Duplicate removal, lowercasing, punctuation/special-character removal
  - Stop-word filtering, tokenization
  - TF-IDF vectorization (unigrams & bigrams)
- Algorithms tested:
  - Classical: Naive Bayes, KNN, Logistic Regression, SVM
  - Deep Learning: CNN, FNN
  - Quantum-inspired: QCNN
- Consistent experimental pipeline ensured fair evaluation

Experimental Work – Observations

•Naive Bayes: efficient baseline, 98.67% accuracy
•KNN: effective but computationally intensive, 95.20% accuracy
•Logistic Regression: interpretable, 98.67% accuracy
•CNN & FNN: solid results (99.39% & 99.65%), but gains limited by dataset size
•SVM: best performance at 99.70%
•QCNN: underperformed (74.17%), limited by classical simulation & immaturity
•Classical models remain strongest; QCNN serves as future-oriented benchmark

CONCLUSION

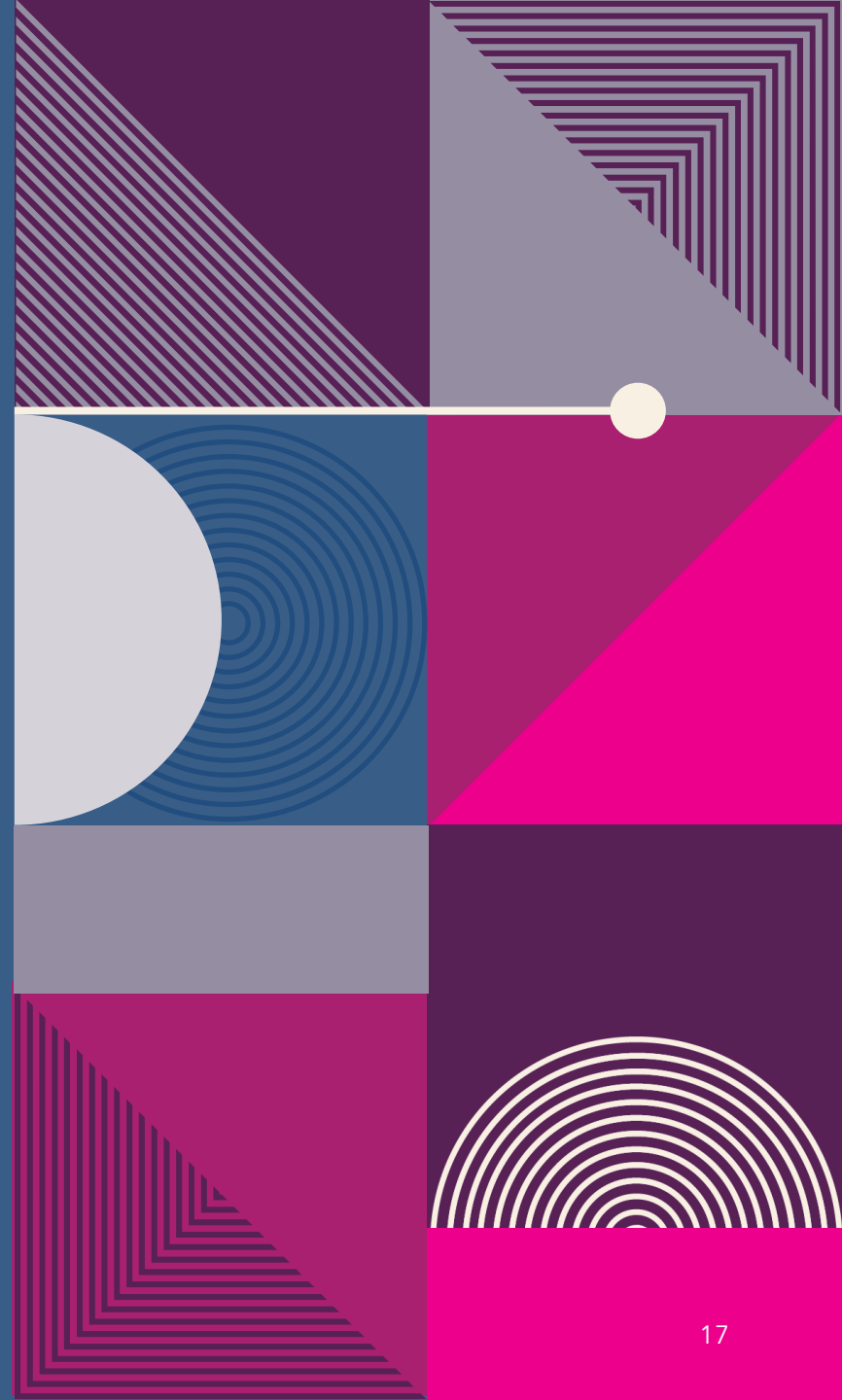
- SVM delivered the best overall performance with 99.69% accuracy.
- Naive Bayes and Logistic Regression remain strong, efficient options for lightweight filtering.
- Deep learning models (CNN, FNN) showed solid but limited gains at current data scale.
- QCNN results highlight potential for future quantum-based text classification.
- Ongoing work will explore larger datasets, ensemble approaches, and native quantum hardware to boost accuracy and scalability.

FUTURE WORK

- Test models on larger, real-world datasets
- Apply ensemble and hybrid techniques for improved accuracy
- Enhance feature engineering and tuning methods
- Explore native quantum hardware for QCNN optimization
- Investigate transfer learning and transformer-based models
- Evaluate model robustness against adversarial and obfuscated spam

ACKNOWLEDGEMENT

- The work is supported in part by the National Science Foundation under NSF grants #2417608, and #2234911, #2209637, #2131228, and #2100134, Any opinions, findings and recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.



REFERENCES

- Zaragoza, H., Gallinari, P., & Rajman, M. (2000). Machine learning and textual information access. PKDD Workshop, Lyon, France.
- Eskin, E., Arnold, A., Prerau, M., Portnoy, L., & Stolfo, S. (2002). A geometric framework for unsupervised anomaly detection. Springer.
- Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. Statistical Science, 17(3), 235–255.
- Singh, M., Pamula, R., & Shekhar, S. K. (2018). Email spam classification by support vector machine. GUCON, 878–882.
- Jeong, Y. S., Woo, J., Lee, S., & Kang, A. R. (2020). Malware detection using spatial pyramid average pooling. Sensors, 20(18), 5265.
- Jantan, A. B., Ghanem, W. A. H. M., & Ghaleb, S. A. A. (2017). Modified bat algorithm for neural network spam detection. J. Theor. Appl. Inf. Technol., 95(24), 6788–6799.
- Tasnim, T., Rahman, M., & Wu, F. (2024). CNN vs QCNN performance in binary classification. IEEE Big Data, 3770–3777.
- Cong, I., Choi, S., & Lukin, M. D. (2019). Quantum convolutional neural networks. Nature Physics, 15, 1273–1278.
- Kaggle Spam Email Dataset. (2021). https://www.kaggle.com/datasets/jackksoncsie/spam-email-dataset
- Google Colaboratory. (2024). https://colab.research.google.com

THANK YOU