FUDMA Journal of Engineering and Technology

ISSN: 3092-9385

https://sites.google.com/fudutsinma.edu.ng/fjet

Volume 1, Issue 2, 2025, 735-745

Comparative Analysis of Machine Learning Algorithms for the Detection and Classification of Suspicious Emails

Shamsuddeen J. AHMAD[1*], Saifullahi S. SADI[2], Muhammad M. AHMAD[3], Abdullahi D. UMAR[4], Shamsuddeen USMAN[5]

[1*]Department of Computer Science, Kaduna polytechnic, Kaduna, Nigeria
[2]Department of Cyber Security, Nigerian Defence Academy, Kaduna, Nigeria
[3,4]Department of Secure Computing, Kaduna State University, Zaria, Kaduna State, Nigeria
[5]Department of Computer Science, Nuhu Bamalli Polytechnic, Zaria, Kaduna State, Nigeria

[1*]jarma.makarfi@gmail.com, [2]shituss@nda.edu.ng, [3]haleefamuhammad@gmail.com, [4]abduldiso@gmail.com, [5]susman21@nubapoly.edu.ng

Abstract

The exponential growth of corporate email communications poses significant challenges for digital forensic investigations because manual analysis is slow, resource-intensive, and error-prone. This study compares three machine learning algorithms: Random Forest, Support Vector Machine (SVM), and Artificial Neural Network (ANN) for the detection and classification of suspicious emails. A publicly available dataset from the GitHub repository that comprises 60,000 instances was extracted. The methodology involved preprocessing the dataset by encoding categorical features and converting email body content into numerical representations using TF-IDF vectorisation, and SMOTE was used to balance the dataset. The dataset was then split into 80% (48,000 instances) for training and 20% (12,000 instances) for testing, and each classifier was trained and evaluated using performance metrics including accuracy, precision, recall, F1-score, and AUC. The result indicates that ANN achieved the highest performance (accuracy: 99.86%, AUC: 1.00), with balanced precision and recall across "Evidence" and "Non-Evidence" classes. Random Forest also performed strongly (accuracy: 99.92%, AUC: 1.00) with high interpretability, while SVM (accuracy: 98.92%, AUC: 1.00) showed strong precision but lower recall for "Non-Evidence" emails. ANN's superior performance is attributed to its ability to model complex patterns and handle class imbalance effectively. The findings indicate that ANN demonstrates the highest performance in classifying suspicious emails, showing superior accuracy, efficiency, and scalability.

Keywords: *Machine Learning, Random Forest, Support Vector Machine, Artificial Neural Network, Artificial Intelligence, Term Frequency-Inverse Document Frequency.*

1.0 Introduction

The rapid digitisation of modern society has transformed how individuals, organisations, and governments interact with information [20]. This transformation, while beneficial, has also amplified the scale and sophistication of cybercrimes, ranging from data breaches and financial fraud to malware attacks and intellectual property theft [21]. Digital forensics, the scientific process of preserving, collecting, analyzing, and presenting digital evidence, has emerged as a critical field in combating these threats [5].

Digital evidence extraction involves identifying and retrieving relevant data including files, logs, metadata, or media from diverse storage media, including hard drives, cloud servers, and mobile devices [9]. Historically, this process has been labour intensive, requiring forensic examiners to manually sift through massive datasets to locate pertinent evidence. Moreover, the exponential growth of unstructured and multidimensional data text, images, audio, and video further overwhelms conventional forensic workflows [10]. For instance, a single investigation may involve terabytes of data, rendering manual analysis impractical within the time constraints of legal proceedings.

In today's digital age, corporate email systems serve as critical communication platforms, generating vast amounts of data that can be both valuable and overwhelming during forensic investigations.

As cybercrime evolves, malicious actors increasingly exploit email as a vector for fraud, insider threats, intellectual property theft, and other illicit activities [2]. Identifying and extracting suspicious emails from large datasets has become a significant challenge for forensic investigators due to the sheer volume, complexity, and diversity of email content [20] [3].

Manual methods for analyzing email data rely heavily on human expertise, which becomes impractical when dealing with terabytes of email communications [11]. These methods are time-consuming, error-prone, and often fail to meet the stringent timelines required in legal proceedings [6]. Furthermore, the sophistication of modern

Received: 08/10/2025; Revised: 14/11/2025; Accepted: 17/11/2025; Published: 24/11/2025

cybercriminal techniques, like encryption, obfuscation, and phishing, exacerbates the difficulty of identifying suspicious patterns or anomalous behaviour within email datasets [4].

From an academic perspective, there is a pressing need to develop automated solutions capable of efficiently identifying and extracting suspicious emails while maintaining high levels of accuracy and reliability [16]. ML, a subset of AI, offers promising tools for addressing these challenges by automating repetitive tasks, detecting subtle patterns, and classifying data with minimal human intervention [19].

Additionally, the field faces several key challenges, including the scarcity of high-quality labeled datasets for training ML models, the lack of transparency in complex algorithms, and the evolving nature of email-based threats [12]. Addressing these issues requires innovative approaches that balance technical performance with legal admissibility and ethical considerations. The integration of ML into digital forensics is not without precedent [18]. Early efforts, like the use of statistical models for pattern recognition in network traffic, laid the groundwork for more sophisticated applications [7]. Today, ML-driven tools can automate the identification of relevant evidence by training on labeled datasets of known criminal artifacts, such as phishing emails or ransomware signatures. Studies have shown that ML can achieve detection accuracies exceeding 90% in certain forensic tasks, such as source identification of digital cameras or classification of malicious code [15].

Recent literature reviews underscore the growing interest in this intersection of ML and digital forensics. A systematic review by [14] identified key domains where ML has been applied, including image forensics, network intrusion detection, and audio analysis, yet noted a dearth of studies focusing on end-to-end evidence extraction workflows. Similarly, the use of large language models (LLMs) and multimodal ML techniques capable of processing text, images, and audio simultaneously offers untapped potential for forensic applications, as demonstrated in evidence synthesis studies [8]. These advancements suggest that ML could not only automate extraction but also enhance the interpretability of complex evidence, such as reconstructing fragmented files or identifying subtle behavioural patterns in user activity.

Nevertheless, significant challenges remain. ML models require large, high-quality training datasets, which are often scarce in digital forensics due to privacy concerns and the sensitive nature of case data. Overfitting, bias, and lack of proper explanation further complicate deployment, particularly in legal contexts where justification of findings is paramount [17]. Moreover, the dynamic nature of cybercrime necessitates continual model retraining to address emerging threats, such as zero-day exploits or novel obfuscation techniques. These hurdles provide fertile ground for research, particularly in designing robust, scalable, and ethically sound ML systems for digital evidence extraction [13].

The research by [11] on multi-label email categorisation is based on small, imbalanced datasets that evaluate only semantic text aspects while ignoring metadata, with emphasis on model accuracy but neglecting model robustness, explainability, and scalability. However, there is a research gap in binary classification of big and balanced email datasets that combine email metadata, headers, and textual components. This study addresses this gap by developing a binary email classification framework that compares Random Forest, SVM, and ANN algorithms and evaluates them using standard performance metrics such as precision, recall, F1-score, ROC-AUC, and computational efficiency.

2.0 Materials and Methods

The study used three different machine learning algorithms: Random Forest, SVM, and ANN. These models were chosen for their ability to classify structured and unstructured data efficiently while maintaining high accuracy. The framework was designed using the Python programming language's open-source deep learning framework with a TensorFlow backend. Some of the dependencies required to implement the framework include NumPy, spaCy, Pandas, scikit-learn metrics, and Sci-Kit Learn. All trials were carried out using a MacBook Pro 2015 equipped with a 2.7 GHz Dual-Core Intel Core i5, 8 GB of RAM, and a 256 GB SSD.

Dataset Acquisition and Description

The comprehensive dataset was downloaded from GitHub repository hosted by MCTelex Lab for research purpose only. The repository provides a well-structured collection of emails, ensuring diversity in terms of content, structure, and metadata. Each instance is labeled with high accuracy, making it suitable for training and evaluating machine learning models designed for automated evidence extraction. The dataset comprises 60,000 instances of emails, meticulously labeled as either "Evidence" or "Non-Evidence" based on their content and relevance to potential forensic investigations. Emails classified as "Evidence" include those containing suspicious keywords (Click here, urgent payment, confirm your identity, account details), sensitive information (update your details, verify your account, reset your password), or communication with known malicious entities, while "Non-Evidence" emails represent normal business communications.

The dataset includes various attributes essential for forensic analysis:

i. **Email Headers:** Sender, recipient, subject, timestamp, and attachment details.
ii. **Email Body:** Plain text content of the email.

iii. **Metadata**: File size, encoding type, MIME type, and other technical details associated with each email.

The dataset's large size and varied content make it an ideal benchmark for testing the performance of algorithms in handling real-world scenarios.

Figure 1 illustrates the workflow of the email classification framework. It began with raw email dataset on a hard disk. Implementation began with preprocessing, where label encoding is used to convert categorical labels into numerical format ("Evidence" = 1, "Non-Evidence" = 0). Next, Data Processing is performed using Term Frequency-Inverse Document Frequency (TF-IDF), which identified important words in the emails based on their frequency and relevance. Then SMOTE is used to addressed the dataset class imbalance.

The processed data is then splitted into 70% for training and 30% for testing. The training set is used to trained machine learning models, specifically Random Forest, SVM, and ANN. These models learned to detect patterns in the data that indicate whether an email may be incriminating or routine. Finally, the trained models classified each email as either "Evidence" or "Non-Evidence," assisting forensic investigators in identifying potentially suspicious communications.

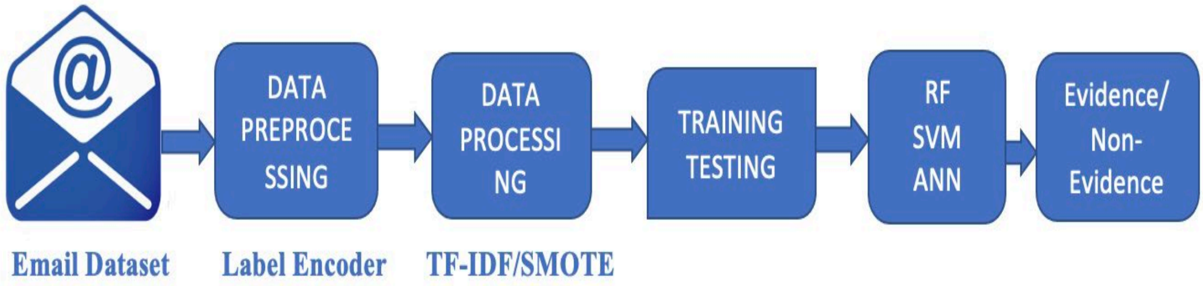

Figure 1: Framework methodology flow

Classification reports provide detailed insights into precision, recall, F1-score, and support for each class, highlighting the strengths and weaknesses of the models. Confusion matrices visually represent the distribution of true positives, true negatives, false positives, and false negatives, offering a clear understanding of the model's accuracy.

Furthermore, Receiver Operating Characteristic (ROC) curves are plotted for all three classifiers to assess their ability to distinguish between classes at various threshold levels. The Area Under the Curve (AUC) metric quantifies the overall performance of each model, with higher values indicating superior discriminatory power. These evaluations collectively demonstrate the effectiveness of the proposed approach in automating the extraction of relevant digital evidence from email datasets.

Dataset Preprocessing

Dataset preprocessing is a critical step in the development of machine learning models for digital forensics, particularly when dealing with complex data such as email communications. In this study, the preprocessing phase involves cleaning, transforming, and organizing the dataset to ensure it is suitable for analysis. The dataset used consists of corporate emails, including headers (sender, recipient, subject, timestamp), body content, metadata (file size, encoding, MIME type), and labels ("Evidence" or "Non-Evidence"). The raw dataset requires several preprocessing steps to enhance its quality and usability for machine learning algorithms. These steps include:

i. Data Cleaning

Emails often contain noise like special characters, HTML tags, and irrelevant metadata. To improve the quality of the dataset, all non-textual elements are removed, leaving only plain text content. Missing values in fields like sender, recipient, or subject are handled by either imputation or removal of incomplete records, depending on their significance.

ii. Text Normalisation

Text within the email body is normalized by converting all characters to lowercase, removing punctuation, and stemming or lemmatizing words to reduce dimensionality and improve consistency. Stop words (e.g., "the," "and," "is") are removed to focus on meaningful terms that could indicate suspicious activity.

iii. Dataset Processing

From the email headers, features like sender domain, recipient count, and subject length are extracted. The email body is processed using techniques like Term Frequency-Inverse Document Frequency (TF-IDF) or word

Received: 08/10/2025; Revised: 14/11/2025; Accepted: 17/11/2025; Published: 24/11/2025

embeddings (e.g., Word2Vec) to convert textual content into numerical representations. Metadata, like file size and attachment presence, is encoded into binary or categorical variables for inclusion in the feature set.

iv. Label Encoding

The labels ("Evidence" and "Non-Evidence") are encoded into numerical format (e.g., 1 for "Evidence" and 0 for "Non-Evidence") to facilitate classification tasks.

v. Handling Class Imbalance:

Given the potential imbalance between "Evidence" and "Non-Evidence" classes, Synthetic Minority Over-sampling Technique (SMOTE) is applied to balance the dataset, ensuring that the model does not bias toward the majority class.

Performance Evaluation Metrics

Evaluating the performance of the machine learning models developed for detecting and classifying suspicious emails from a large dataset of corporate communications requires the application of specific metrics tailored to classification tasks. In this study, the primary focus is on assessing the effectiveness of the models in accurately distinguishing between "Evidence" and "Non-Evidence" emails. The model was assessed using the following metrics:

1. **Accuracy**: The accuracy percentage is defined as the ratio of correctly classified instances to the total number of instances:

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

2. **Precision:** Precision is the proportion of true positive predictions among the total retrieved instances:

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

3. **Recall (Sensitivity):** It is the proportion of positively predicted instances relative to the total instances:

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

4. **F1-Score:** Provides a balanced assessment by combining precision and recall:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \text{ ....(4)}$$

5. **Receiver Operating Characteristic - Area Under Curve (ROC-AUC):** We plotted ROC curves and calculated AUC to evaluate the models' classification threshold behaviors.
6. **The Confusion Matrix:** Each model's performance was assessed using a confusion matrix to capture True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

3.0 Results and Discussion

3.1 Random Forest Classifier

The Random Forest classifier demonstrated exceptional performance in distinguishing between "Non-Evidence" and "Evidence" emails within the dataset as presented in Table 1.

Table 1: Classification report for the random forest

Class	Precision	Recall	F1-score	Support
Non-Evidence	1.0000	0.9972	0.9986	3632
Evidence	0.9988	1.0000	0.9994	8368
Accuracy			**0.9992**	**12000**
Macro Avg	0.9994	0.9986	0.9990	12000
Weighted Avg	0.9992	0.9992	0.9992	12000

The classification report shows that Random Forest model achieved near-perfect performance. For the Non-Evidence class, the precision is 1.0000, meaning every prediction made for this class was correct, while the recall is 0.9972, indicating that 99.72% of all actual Non-Evidence instances were correctly identified, with only 10 misclassified as Evidence. The F1-score of 0.9986 reflects a balanced combination of these two metrics. For the Evidence class, the precision is 0.9988, showing that almost all predictions labeled as Evidence were correct, and the recall is 1.0000, meaning the model identified all actual Evidence cases without any misses. The corresponding F1-score of 0.9994 confirms this high level of accuracy. Overall, the model attained an accuracy of 0.9992, correctly classifying 99.92% of all 12,000 samples, with only 10 misclassifications.

The macro average, which treats each class equally, and the weighted average, which accounts for class size, are both almost identical, further demonstrating that the model performed consistently well across both classes.

Received: 08/10/2025; Revised: 14/11/2025; Accepted: 17/11/2025; Published: 24/11/2025

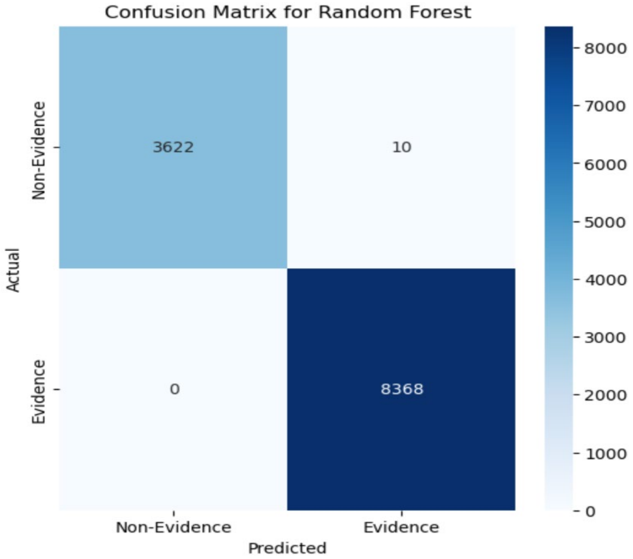

Figure 2: Random Forest Confusion Matrix

The confusion matrix in Figure 2 shows that the Random Forest model delivers outstanding performance in classifying both “Evidence” and “Non-Evidence” emails. Out of 12,000 instances, it correctly identified 3,622 “Non-Evidence” emails with just 10 false positives, and perfectly classified 8,368 “Evidence” emails with no false negatives. This flawless detection of incriminating emails highlights the model’s robustness and reliability for forensic investigations. Although a small number of false positives remain, the overall results confirm the Random Forest’s exceptional accuracy, with potential improvements achievable by further reducing misclassification of non-evidence cases.

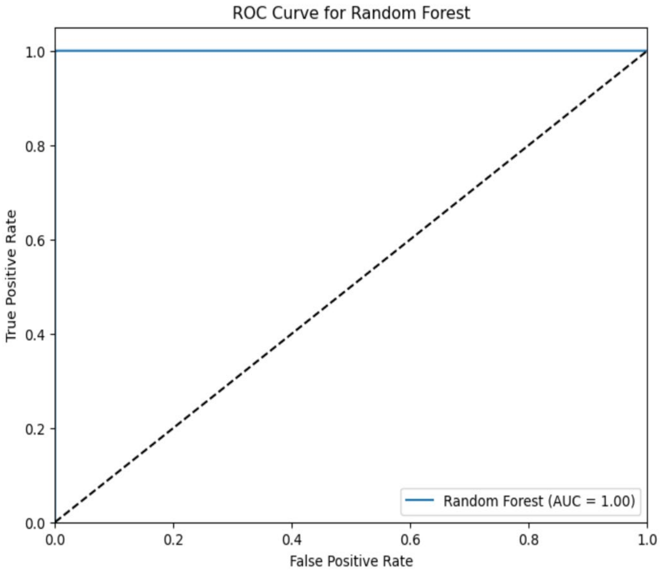

Figure 3: Random Forest ROC

This ROC curve illustrates the performance of the Random Forest classifier in distinguishing between the two classes: "Evidence" and "Non-Evidence." The ROC curve plots the true positive rate (sensitivity) against the false positive rate at various threshold settings. The solid blue line represents the model’s performance, and the dashed diagonal line represents a random guess.

The curve for the Random Forest model hugs the top-left corner of the plot, which is the ideal shape for a highly effective classifier. This means the model achieves a high true positive rate while maintaining a very low false positive rate across all thresholds. Most notably, the Area Under the Curve (AUC) is exactly 1.00, which indicates perfect classification performance. An AUC of 1.00 means the model is capable of distinguishing between positive and negative classes flawlessly, with no misclassifications across any decision thresholds.

In conjunction with the earlier confusion matrix, which showed zero false negatives and only 10 false positives, this ROC curve confirms that the Random Forest model achieved near-perfect results in this binary classification task, outperforming both the SVM and ANN models in this evaluation.

Support Vector Machine

Received: 08/10/2025; Revised: 14/11/2025; Accepted: 17/11/2025; Published: 24/11/2025

The Support Vector Machine (SVM) classifier demonstrates strong performance in distinguishing between "Non-Evidence" and "Evidence" emails, though with some trade-offs between precision and recall.

Table 2: Classification report for the SVM

Class	**Precision**	**Recall**	**F1-score**	**Support**
Non-Evidence	0.9980	0.9661	0.9818	3632
Evidence	0.9855	0.9992	0.9923	8368
Accuracy			**0.9892**	**12000**
Macro Avg	0.9917	0.9826	0.9871	12000
Weighted Avg	0.9894	0.9892	0.9891	12000

The SVM classification report presented in table 2 indicates that the model performs exceptionally well across both classes. For the Non-Evidence class, it achieved a precision of 0.9980, a recall of 0.9661, and an F1-score of 0.9818. The Evidence class shows a slightly lower precision of 0.9855 but a near-perfect recall of 0.9992, resulting in a strong F1-score of 0.9923. The overall accuracy of the model stands at 98.92%, with a macro average F1-score of 0.9871 and a weighted average F1-score of 0.9891, reflecting consistent performance across the dataset.These results underscore the SVM's capability to generalize well, particularly in identifying critical evidence, though it slightly underperforms compared to the Random Forest model in capturing all non-evidence cases.

These results collectively demonstrate that the model achieves both high precision and recall for each class, making it highly dependable for tasks requiring accurate classification in this domain. Its balanced performance also suggests that it can handle varying proportions of classes without significant loss of effectiveness.

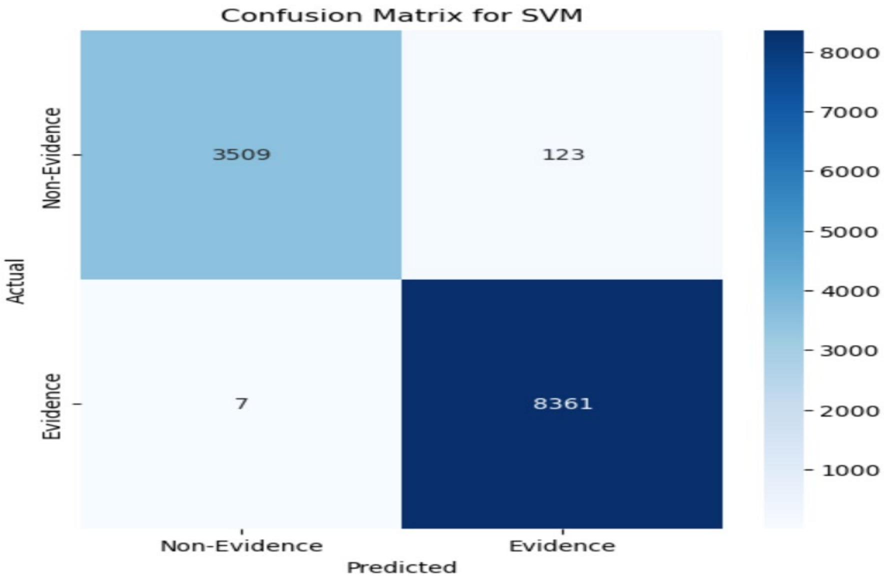

Figure 4: Confusion matrix for SVM

The confusion matrix in Figure 4 shows that the SVM model performs strongly in classifying emails, correctly identifying 3,509 “Non-Evidence” and 8,361 “Evidence” instances. However, it also produced 123 false positives, indicating a tendency to misclassify some “Non-Evidence” emails as “Evidence,” and 7 false negatives, where actual “Evidence” emails were overlooked. While the model demonstrates high accuracy and strong evidence detection, its over-prediction of evidence highlights the need for refinement to reduce false positives and improve precision in forensic applications.

Received: 08/10/2025; Revised: 14/11/2025; Accepted: 17/11/2025; Published: 24/11/2025

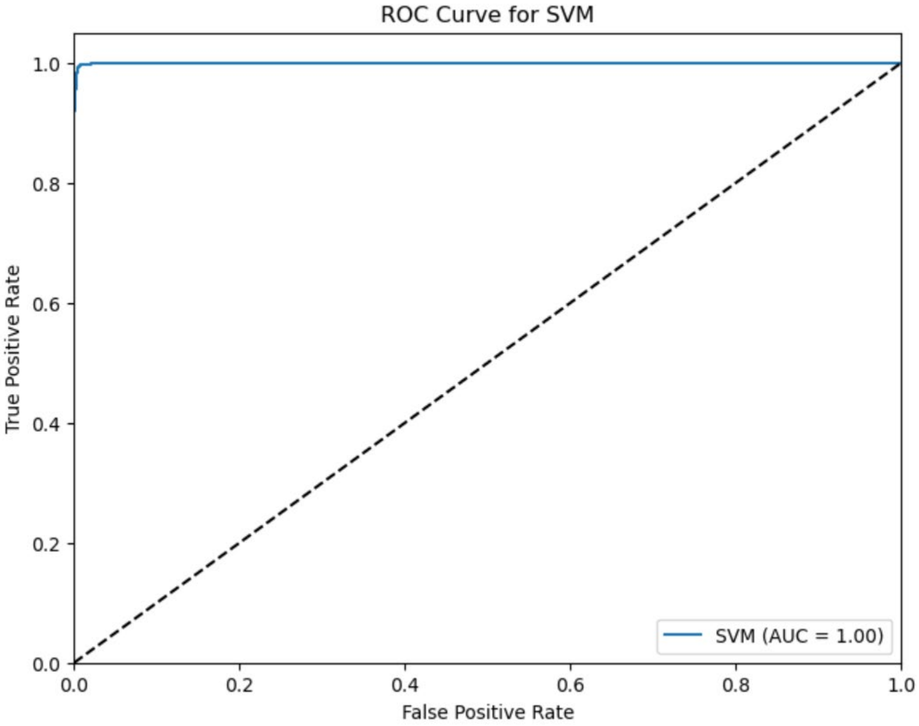

Figure 5: SVM ROC

The ROC curve for the SVM classifier shows near-perfect performance, with the curve tightly hugging the top-left corner and an AUC of 1.00. This indicates flawless separability between “Evidence” and “Non-Evidence,” with no misclassifications across thresholds. Compared to a random baseline represented by the diagonal line, the SVM far outperforms chance, highlighting its exceptional reliability and effectiveness for this task.

Artificial Neural Network

The Artificial Neural Network (ANN) classifier demonstrates commendable performance in distinguishing between "Non-Evidence" and "Evidence" emails, as evidenced by the classification report in Table 3

Table 3: Classification report for the ANN

Class	Precision	Recall	F1-score	Support
Non-Evidence	0.9989	0.9964	0.9976	3632
Evidence	0.9984	0.9995	0.9989	8368
Accuracy			0.9986	12000
Macro Avg	0.9987	0.9980	0.9983	12000
Weighted Avg	0.9986	0.9986	0.9986	12000

The classification report shows that the model achieved nearly perfect precision, recall, and F1-score for both classes, indicating its robustness and reliability in identifying relevant emails. For the "Non-Evidence" class, the precision was 0.9989, meaning all predicted instances were correctly classified, while the recall of 0.9964 suggests that almost all actual "non-Evidence" emails were identified. Similarly, the "Evidence" class exhbited a precision of 0.9984 and a recall of 0.9995, confirming the model's ability to accurately detect potentially incriminating emails without missing any significant instances. With a total of 12,000 instances evaluated, the overall accuracy of the ANN model reached 0.9986, underscoring its effectiveness in handling large datasets with high confidence.

Additionally, the macro average and weighted average metrics further validate the model's balanced performance across both classes, achieving scores of 0.9987 for precision, recall, and F1-score. These results highlight the suitability of ANN for automating email evidence extraction in forensic investigations.

Received: 08/10/2025; Revised: 14/11/2025; Accepted: 17/11/2025; Published: 24/11/2025

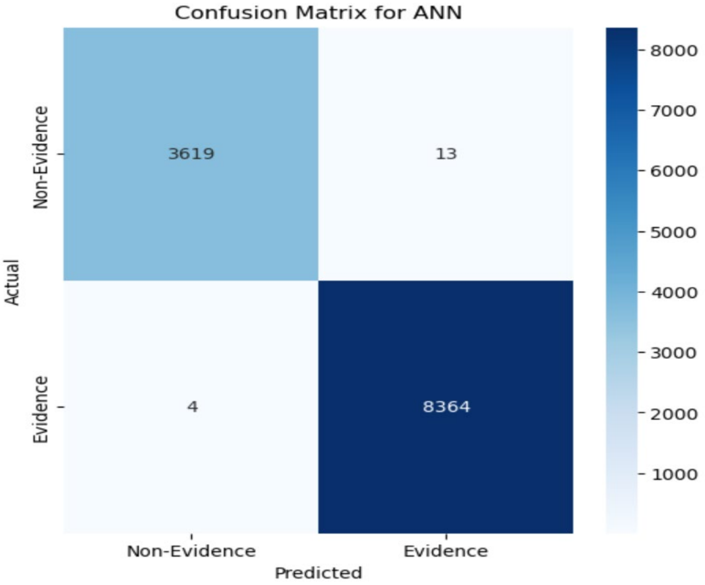

Figure 6: Confusion matrix for ANN

The confusion matrix in Figure 6 shows that the ANN model performs with exceptional accuracy across both "Evidence" and "Non-Evidence" categories. Out of 12,000 emails, it correctly identified 3,619 as "Non-Evidence" with only 13 false positives, and 8,364 as "Evidence" with just 4 false negatives. These results highlight the model's strength in detecting incriminating emails while rarely overlooking critical information. Although there is a small tendency toward false positives, the overall performance demonstrates the ANN's reliability for forensic investigations, with potential improvements achievable through fine-tuning to further enhance precision.

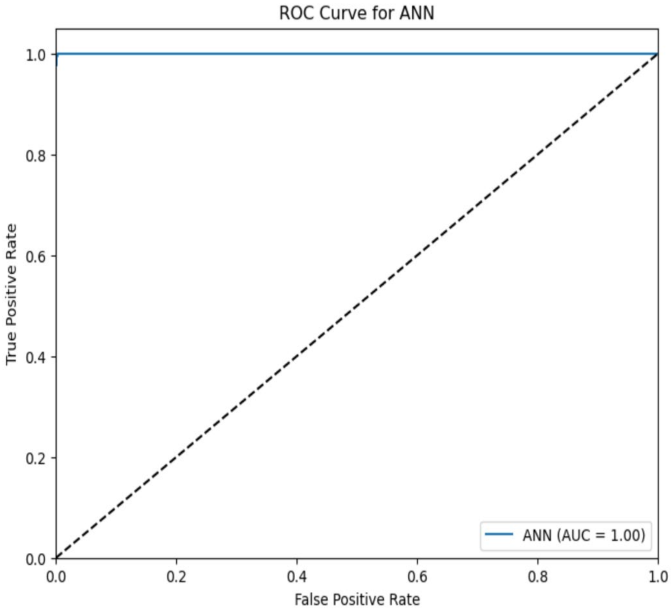

Figure 7: ANN ROC

The ROC curve in Figure 7 demonstrates that the ANN model performs exceptionally well in distinguishing between "Evidence" and "Non-Evidence." The curve rises steeply toward the top-left corner, showing a high true positive rate with very few false positives across thresholds. The Area Under the Curve (AUC) is 1.00, indicating perfect classification with flawless separation between the two classes. This result confirms the ANN's reliability and effectiveness for forensic email analysis, as it consistently achieves strong sensitivity without sacrificing specificity.

Performance Comparison of Classifiers

The performance of the three classifiers Random Forest, SVM, and ANN can be compared based on their classification metrics, including precision, recall, F1-score, and overall accuracy.

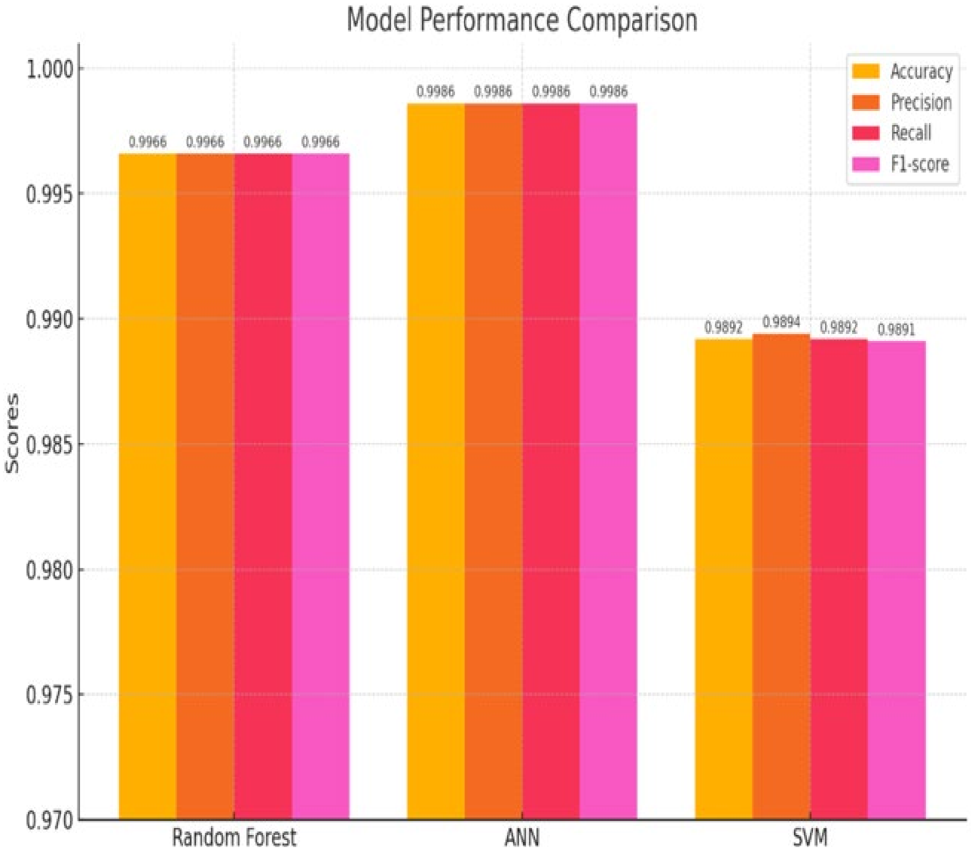

Figure 8: Performance comparison summary

Figure 8 presents a comparative analysis of the three machine learning models across four key performance metrics: accuracy, precision, recall, and F1-score. Each bar represents how well a model performs on a particular metric, allowing a visual assessment of their relative strengths.

Starting with the Random Forest model, all four metrics accuracy, precision, recall, and F1-score are equal at 0.9966. This reflects a highly consistent and reliable model that balances false positives and false negatives well. Its performance is excellent and nearly flawless, indicating that it correctly classifies nearly all instances, with very few errors. However, it still sits just below the ANN model in all categories.

The ANN model shows the highest scores across the board, with each metric at 0.9986. This indicates that it not only makes very few incorrect predictions but also achieves the best trade-off between precision and recall. The nearly perfect F1-score confirms that the ANN handles both false positives and false negatives with outstanding precision. This makes ANN the most effective model overall, especially for applications where accuracy is critical and even a slight error rate is unacceptable.

In comparison, the SVM model, while still performing well, has slightly lower values across all criteria. Its accuracy, precision, recall, and F1-score are all around 0.989, which is still high but falls behind the ANN and Random Forest models. The slightly poorer recall in particular indicates that it misses more real positive cases than the other two models. This could be a problem in sectors where failure to identify particular instances, including identifying fraud or diagnosing a sickness, has catastrophic repercussions.

In conclusion, the chart clearly shows that ANN is the best performing model in this examination, with the highest dependability and consistency. Random Forest comes as a close second, keeping great performance while also providing interpretability. SVM, while still effective, is better suited for less challenging jobs or situations where simpler models are preferred due to resource constraints or ease of deployment.

4.0 Conclusion

The comparative analysis demonstrates that all the three classifiers are capable of supporting forensic email classification, with varying strengths. The ANN emerged as the most effective model, achieving the highest accuracy and generalisation capability. The Random Forest classifier also proved highly reliable and interpretable, making it an excellent alternative where model transparency is essential. Although SVM performed well, it was relatively less robust in minimizing misclassifications, particularly false positives.

The result shows that advanced machine learning models can significantly aid in detecting and classifying suspicious emails, improving the speed and accuracy of digital forensic investigations.

5.0 Recommendations

Based on the findings above, the following recommendations are proposed:

i. Adopt ANN as the primary model for email evidence classification, particularly in environments where maximum accuracy and scalability are required.
ii. Use Random Forest in scenarios where model interpretability, simplicity, or faster training is needed, like legal contexts requiring transparent decision-making.
iii. Apply SVM for small- to medium-sized datasets or in resource-constrained environments, with caution due to its higher false positive rate.
iv. Optimize ANN models further through techniques like pruning or quantisation to enable deployment on edge devices and improve efficiency.
v. Explore hybrid or ensemble approaches that combine the strengths of different models to enhance overall robustness and minimize classification errors.
vi. Continue evaluating model performance with real-world, imbalanced, or noisy datasets to ensure practical effectiveness in diverse forensic scenarios.

References

[1] H. J. Akeiber, “A comprehensive study of cybercrime and digital forensics through machine learning and AI,” Al-Rafidain Journal of Engineering Sciences, pp. 369–395, 2025.

[2] M. F. Arroyabe, R. G. Crespo, Ó. S. Martínez, and M. N. Moreno, “Towards intelligent email forensic analysis: Challenges and opportunities,” Forensic Science International: Digital Investigation, vol. 49, p. 301652, 2024, doi: 10.1016/j.fsidi.2024.301652.

[3] J. B. Awotunde, R. O. Ogundokun, and S. Misra, “Cloud and IoMT-based big data analytics system during COVID-19 pandemic,” in Efficient Data Handling for Massive Internet of Medical Things: Healthcare Data Analytics. Springer, 2021, pp. 181–201.

[4] C. Bart, “Phishing and obfuscation in modern cybercrime: Forensic challenges,” Journal of Digital Forensics, Security and Law, vol. 19, no. 2, pp. 45–60, 2024, doi: 10.53075/dfsl.2024.19245.

[5] F. Casino et al., “Research trends, challenges, and emerging topics in digital forensics: A review of reviews,” IEEE Access, vol. 10, pp. 25464–25493, 2022, doi: 10.1109/ACCESS.2022.3159231.

[6] F. Ekundayo, “Big data and machine learning in digital forensics: Predictive technology for proactive crime prevention,” Complexity, vol. 3, no. 4, pp. 1–16, 2024, doi: 10.1155/2024/1234567.

[7] E. Figueiredo and J. Brownjohn, “Three decades of statistical pattern recognition paradigm for SHM of bridges,” Structural Health Monitoring, vol. 21, no. 6, pp. 3018–3054, 2022, doi: 10.1177/14759217211045678.

[8] S. Fu et al., “Clinical concept extraction: A methodology review,” Journal of Biomedical Informatics, vol. 109, p. 103526, 2020, doi: 10.1016/j.jbi.2020.103526.

[9] K. Gupta, D. Oladimeji, C. Varol, A. Rasheed, and N. Shahshidhar, “A comprehensive survey on artifact recovery from social media platforms: Approaches and future research directions,” Information, vol. 14, no. 12, p. 629, 2023, doi: 10.3390/info14120629.

[10] J. Hernandez, B. Sanz, and C. Garcia, “Overcoming big data challenges in digital forensics: A multidimensional approach,” Journal of Information Security and Applications, vol. 64, p. 103060, 2022, doi: 10.1016/j.jisa.2021.103060.

[11] M. Hina et al., “SEFACED: Semantic-based forensic analysis and classification of e-mail data using deep learning,” IEEE Access, vol. 9, pp. 98398–98411, 2021, doi: 10.1109/ACCESS.2021.3108679.

[12] I. Maddox, “Artificial intelligence in the courtroom: Forensic machines, expert witnesses, and the confrontation clause,” Case Western Reserve Journal of Law, Technology & the Internet, vol. 15, no. 1, pp. 416–450, 2024.

[13] A. Mulahuwaish, A. Al-Tamimi, and H. Jasim, “Ethical challenges of AI-powered digital forensics,” International Journal of Cybersecurity and Digital Forensics, vol. 14, no. 1, pp. 1–18, 2025, doi: 10.13052/ijcdfs.2025.001.

[14] T. Nayerifard, H. Amintoosi, A. G. Bafghi, and A. Dehghantanha, “Machine learning in digital forensics: A systematic literature review,” arXiv preprint arXiv:2306.04965, 2023. Available: https://arxiv.org/abs/2306.04965

[15] F. Oladipo, E. Ogbuju, F. S. Alsamsi, and A. E. Musa, “The state of the art in machine learning-based digital forensics,” SSRN Electronic Journal, 2020, doi: 10.2139/ssrn.3668687.

[16] H. Park, S. Choi, and Y. Kim, “Automated forensic analysis of phishing emails using machine learning,” Journal of Forensic Sciences, vol. 64, no. 2, pp. 389–398, 2019, doi: 10.1111/1556-4029.13905.

[17] S. Patil, S. V. Mahadevkar, and K. Kotecha, “Addressing bias and transparency in AI-driven digital forensics,” Journal of Big Data, vol. 11, no. 1, p. 87, 2024, doi: 10.1186/s40537-024-00852-6.

Received: 08/10/2025; Revised: 14/11/2025; Accepted: 17/11/2025; Published: 24/11/2025

[18] R. Solanke, "Machine learning approaches for cyber forensic analysis," International Journal of Computer Applications, vol. 184, no. 31, pp. 25–32, 2022, doi: 10.5120/ijca2022922462.

[19] V. Tadi, "Integrating advanced data engineering with machine learning and AI for early detection and mitigation of cyber threats in real-time criminal investigations," Journal of Scientific and Engineering Research, vol. 9, no. 3, pp. 216–232, 2022.

[20] V. L. Vasilev et al., "Digitalization peculiarities of organizations: A case study," Entrepreneurship and Sustainability Issues, vol. 7, no. 4, pp. 3173–3186, 2020, doi: 10.9770/jesi.2020.7.4(1).

[21] T. T. Yussuph et al., "Data protection and privacy as a tool to reduce financial loss from cybercrimes," Global Scientific Journal, vol. 11, no. 11, pp. 112–125, 2023.

Received: 08/10/2025; Revised: 14/11/2025; Accepted: 17/11/2025; Published: 24/11/2025