International Journal of Computer Sciences and Engineering
Vol.**13**, Issue.**1**, pp.**24-32**, January **2025**
ISSN: 2347-2693 (Online)
Available online at: www.ijcseonline.org

ISSN: 2347-2693 (E)

Research Article

Model For Email Spam Classification Using Hybrid Machine Learning Technique

Domo Omatsogunwa Ereku[1*], Vincent I.E. Anireh[2], Onate Egerton Taylor[3]

[1,2,3]Computer Science Department/Faculty of Science, Rivers State University, Port Harcourt, Nigeria

**Corresponding Author: domo.ereku@ust.edu.ng*

Received: 22/Nov/2024; **Accepted**: 24/Dec/2024; **Published**: 31/Jan/2025. **DOI**: https://doi.org/10.26438/ijcse/v13i1.2432

Abstract: An optimized Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) together (GA-PSO) method for email spam classification is presented in this paper. To improve classification accuracy and computing efficiency, the model combines The collective intelligence found in Particle Swarm Optimization (PSO).with the evolutionary powers of Genetic Algorithms (GA). The proposed GA-PSO classifier was rigorously tested over 400 cycles using datasets from Enron and Spam Assassin. Superior performance measures were attained by the model, including a 50% improvement in fitness margin, a 3% decrease in fitness error margin, and a computational efficiency that was five times faster than traditional techniques. By developing a strong, scalable algorithm with enhanced decision-making accuracy, this research advances spam detection and makes a substantial advancement in tackling email spam issues.

Keywords: Email Spam, Machine Learning, Genetic Algorithm, Particle Swarm Optimization.

1. Introduction

Classification models play a critical role in modern applications, particularly in adaptive and robust systems for tasks like email spam detection. Despite the efficiency of emails as a communication tool, they face challenges like Data Redundancy Detection (DRD), which underpins the spam phenomenon. Spam emails, often disguised using text, images, or official-like appearances, create vulnerabilities such as data privacy breaches, denial of service, and operational inefficiencies [1]. [2] Traditional spam filters that rely on static rules or dictionaries are ineffective against the evolving tactics of spammers, making it necessary to employ machine learning (ML) techniques to enhance spam detection. [3] ML models require robust training to analyze both textual and graphical content, leveraging methods like Optical Character Recognition (OCR) for spam detection in embedded images [4]. However, selecting the optimal ML technique remains challenging due to the diversity of models and evaluation metrics. To overcome these limitations, this research presents an approach that combines Genetic Algorithm and Particle Swarm Optimization, creating a hybrid method known as GA-PSO. utilizing the global search capability of GA along with the speed and simplicity of PSO to improve spam classification. This work is essential to solving the problem of effectively identifying and handling spam emails, which jeopardise productivity and communication security. This study introduces an innovative approach that enhances classification accuracy, minimizes false positives, and optimizes performance by integrating the Genetic Algorithm and Particle Swarm Optimization algorithm, referred to as the hybrid GA-PSO model. The model maintains communication dependability by lowering the number of erroneous classifications of valid emails and guarantees enhanced email security by efficiently filtering risks like malware and phishing. Because of its speed and scalability, the model can handle big datasets in real-time applications, which is useful for creating more intelligent and effective email filtering systems for both individuals and businesses.

2. Related Work

Embracing machine learning approaches opens the door to endless possibilities and transformative change, [5] compared the effectiveness of the Random Forest Classifier algorithm with the Support Vector Classifier (SVC) algorithm in identifying spam emails. The researchers developed two classifiers using a publicly available dataset and evaluated each one's performance with standard metrics such as F1 score, accuracy, precision, and recall. The survey highlights the importance of effective spam identification for maintaining cybersecurity and improving the user experience in digital communication. The study highlights topics for further investigation even though it is methodologically sound and provides insightful information about the classifiers' relative performance. These include testing additional machine learning algorithms to enhance spam detection capabilities and utilizing a broader range of datasets to improve generalizability. All things considered; the study significantly

advances the subject by showing how sophisticated computational techniques can be used to solve real-world email security issues. [6] researched unsupervised feature learning for spam email filtering. Their work tried to solve the problem relating to data transformation. They proposed a method for feature representation that enhances class separability while reducing the search space for identifying spam. The results indicate that this approach is effective in classifying incoming emails as either spam or non-spam. Classifiers such as Random Forest, Support Vector Machines (SVM), and the C4.5 decision tree can achieve good performance, even when utilizing a very limited set of features. However, they were unable to assess alternative datasets to demonstrate that their approach is not confined to merely handling spam detection problems. Employing a novel feature selection method known as Term Frequency Difference and Category Ratio (TFDCR) with dynamic feature updates. [7] developed an incrementally tailored email spam filter. Three stages were involved in the implementation of their strategy. First, without taking into account the quantity of data in each class, they used TFDCR to find discriminative characteristics. Second, they employed a Support Vector Machine (SVM)-based incremental learning approach, which enabled the classifier to update its discriminative functions on the fly. Finally, they presented Rank Weight, a selection function that was created to improve the current feature set by finding and adding fresh features from incoming emails that have high discriminative power. TFDCR was the most successful feature selection function, according to the results. However, their model had flaws, especially when it came to resolving unequal class distributions and idea drift. [3] used evolutionary random weight networks to create an intelligent system for feature selection and spam detection. Their method successfully addressed the problem of spam email identification through the combination of the Random Weight Network (RWN) for classification with the Genetic Algorithm (GA) for wrapper feature selection. Three datasets were used to test the system: CSDMC2010, LingSpam, and Spam-Assassin. Recall, accuracy, and precision all showed remarkable performance in the experimental results. Other techniques for managing unbalanced classification problems, like cost-sensitive learning, were not examined in the study, though. A technique based on architecture was presented by [8] to identify spam in electronic communications, particularly instant message spam. The C4.5 classifier was used for the detection procedure, and the input data came from an instant messaging environment. Together with lower error rates, the results showed excellent recall, accuracy, and precision. Nevertheless, the study was limited because it did not do feature extraction and selection or integrate further machine learning approaches. With an emphasis on feature selection to detect spam emails. [9] They surveyed methods for enhancing email spam filtering, assessing Support Vector Machines, Multilayer Perceptrons, and Naive Bayes classifiers. According to the results, the Naive Bayesian classifier was very effective. However, the lack of feature extraction in the methodology was a limitation of the study. [10] explored the use of natural language processing (NLP) algorithms to detect spam comments. Their approach integrated deep learning techniques with statistical analysis to address the issue. The two main procedures that were the focus of the study were spam control and spam identification. While spam control focused on feature extraction, spam identification included an early pre-processing step. Their results demonstrated how well NLP works to detect encrypted communications and non-English phrases in spam comments. However, the study has drawbacks, most notably the incapacity to carry out feature selection, which could improve the identification process' accuracy. To detect image spam, [11] an Expectation Propagation (EP) technique based on learning an infinite inverted Dirichlet mixture was presented. To estimate the parameters of this suggested infinite model, they created an EP approach. The outcomes showed how effective their strategy for picture spam filtering was. Their strategy was limited, however, as it failed to significantly decrease false positive and false negative errors. [12] looked into two methods for filtering spam emails: a memory-based method and the Naive Bayesian approach. They assessed the extent to which these techniques performed in distinguishing spam from authentic emails using the Ling-Spam dataset. The Naive Bayesian classifier, which uses word frequency analysis and probabilistic reasoning, demonstrated excellent precision and effectiveness, making it a good fit for big datasets. The memory-based approach, on the other hand, was less scalable and required more processing power because it compares fresh emails with previously saved instances. Despite the benefits of both methods, The study demonstrated that the Naive Bayesian strategy outperformed others in both speed and accuracy. This research significantly influenced the advancement of machine learning applications for text categorization and provided valuable insights into spam filtering systems. [13] examines the application of deep learning techniques for efficiently detecting spam emails. The research underscores the effectiveness of advanced neural network architectures, with a particular focus on Convolutional Neural Networks (CNNs). These powerful models are specifically designed to identify and interpret intricate data patterns that are often characteristic of spam content. By leveraging their ability to analyze visual and sequential information, CNNs can efficiently differentiate between legitimate and unwanted communications, making them a vital tool in the fight against spam. The authors claim that deep learning models outperform traditional machine learning methods, such as Naïve Bayes and Support Vector Machines (SVM), in their flexibility to adapt to the ever-changing tactics used by spammers. The study employs evaluation criteria metrics such as accuracy, precision, recall, and F1-score are used for comparison. The performance of deep learning techniques against traditional methods thoroughly. The results show that deep learning methods detect spam with significantly greater accuracy rates, highlighting their potential as a reliable option for email service providers looking to improve spam filtering systems. In conclusion, this study makes a significant contribution to the knowledge advancement in robust spam categorization systems. tackling the ongoing problems caused by unsolicited email correspondence in modern digital settings. [14] looks at how advanced pre-processing techniques in Natural Language Processing (NLP) might improve spam detection systems. The authors stress the significance of data preparation, which

comprises methodically cleaning and altering email content, in order to improve the accuracy of spam categorization algorithms. This study uses certain pre-processing methods to improve the ability of machine learning algorithms to distinguish between spam and legitimate emails. These include removing unnecessary tags, special characters, and stop words, while retaining semantically significant terms. The study's conclusions show that tailored pre-processing techniques significantly increase spam detection systems' efficacy. In typical datasets, the authors demonstrate significant gains in classification accuracy using their approach. This study emphasises how crucial pre-processing is to NLP-based spam detection, showing that careful consideration of data preparation can produce better results than depending solely on algorithmic developments. In the end, this work makes significant contributions to the creation of frameworks for spam detection in digital communication systems that are more successful. By employing both machine learning (ML) and deep learning (DL) techniques, [15] The research study explores methods to reduce spam email by combining deep learning techniques, such as Artificial Neural Networks (ANNs), with machine learning algorithms like Logistic Regression, Naïve Bayes, and Random Forest. It addresses the issue of spam interference in both personal and professional communication. For training and assessment, the study employed a dataset of more than 83,000 labelled email records. The findings demonstrated that both conventional and deep learning techniques were highly accurate in detecting spam, with Logistic Regression, Random Forest, and Naïve Bayes achieving 97% accuracy and a 97.5% F1-score. The ANN model's 98% accuracy gave it a modest advantage over the others. These results highlight the advantages of combining methods for spam categorisation and make recommendations for more research to enhance their use in various email instances. None of the methods utilised can guarantee 100% accuracy in spam detection, despite the models' excellent accuracy. Because of this restriction, there is always a possibility that some spam emails will get past filters or that real emails will be incorrectly classified as spam. Since the model's success is dependent on the calibre and applicability of the features, feature engineering and selection represent yet another significant limitation. The model may suffer if crucial elements are omitted or superfluous ones are included. It's possible that the study did not thoroughly examine sophisticated feature extraction strategies or the possibility of utilising ensemble approaches to improve classification outcomes. Furthermore, models need to be updated and retrained regularly to stay updated with new spam forms because spammers are always changing their tactics. This can be resource-intensive and challenging to execute successfully.

3. Experimental Method/Procedure/Design:

This paper uses an enhanced Genetic Algorithm and Particle Swarm Optimisation (GA-PSO) model to successfully tackle the problem of classifying email spam through a constructive research strategy. The stages of the research process are as follows:

i. Gaining an in-depth understanding of email spam classification and the limitations of existing approaches.
ii. Acquiring datasets from Spam Assassin and Enron for model training and evaluation.
iii. Applying Object-Oriented Design (OOD) principles to analyze the operational structure of the GA-PSO model.
iv. Implementing the proposed GA-PSO classifier using Python programming language.
v. Conducting rigorous testing of the model across 400 iterations to validate its performance metrics.
vi. Developing preventive mechanisms to improve the detection and classification of email spam through enhanced decision-making accuracy

3.1. The Object-Oriented Analysis and Design Methodology (OOADM) is a powerful approach that enhances the software development process. By utilizing OOADM, developers can create more efficient, adaptable, and maintainable systems that perfectly align with user requirements. Embracing this methodology leads to improved project outcomes and fosters innovation in design. It is ideal for designing machine learning systems, such as classifying email spam. This methodology structures the system into objects, which encapsulate data and behaviours, making it ideal for modular and scalable designs. The following approach was taken in the design of the system:

A. System Components as Objects: In the GA-PSO spam classification system, key components such as the dataset, pre-processing pipeline, feature selection, and classification algorithms (GA and PSO) are represented as objects. Each object is responsible for a specific part of the system's functionality, ensuring modularity.

B. Defining Class Abstractions: OOADM groups related operations and attributes into classes. For instance:
i. A Dataset class handles loading, splitting, and managing datasets (e.g., Spam Assassin, Enron).
ii. A GA class encapsulates evolutionary operations like selection, crossover, and mutation.
iii. A PSO class manages particle initialization, velocity updates, and fitness evaluation.
iv. A Hybrid Model class combines GA and PSO for optimization.

C. Component Roles and Interactions: The system outlines the functions of various components. For instance, the Hybrid Model class collaborates with the Dataset class to access data and works alongside the GA and PSO classes to enhance the spam classification model.

D. Encapsulation of Data and Operations: Data (such as email characteristics and optimization settings) and operations (including fitness assessment and model training) are grouped within their respective classes. This encapsulation helps maintain data integrity and simplifies processes related to spam classification.

3.2. Algorithm of the proposed GA-PSO Model

i. Initialize GA, PSO Systemic parameter:

Fitness: A functional class that assigns a cost (evaluation) score, given a set of input parameters (I_p).

Fitness_Criterion: A threshold that specifies the termination or stopping criterion of a GA, PSO simulation.
p: Also known as the Dimension or Problem Dimension, this is the number of (Ip) that must be present in a particular population.
r: The percentage of the population that Crossover replaces at each simulation stage.
m: The speed at which mutations occur.

ii. *Initialize population:* P ← Generate a random SET OF I_p, say $I_{i\ (r)}$ *Consider the following: For every scenario, calculate Fitness ($I_{i(r))}$ for $h_{i(r)}$ in P.* While Fitness Criterion do [max Fitness (h_p)]
iii. Produce an offspring, Ps:
iv. To add to 'Ps' by a roulette wheel, choose '(1 - r) p' members of "P" in a probabilistic manner. The probability '$P_{r(hi)}$' that 'h_i' will be chosen from 'P' is provided by

$$P_r(h_i) = \frac{Fitness(h_i)}{\sum_{j=1}^{P} Fitness(h_j)} \quad (1)$$

v. Crossover: Choose at random using probability. (r-p/2) pairs of randomized Hypothetical-Parameters h_i from "$P_r(h_i)$", according to equation (1), Let's say (h_l, h_2) for each pair, Use the crossover operator to produce two offspring. Add all offspring to P_s
vi. Mutate: To ensure uniform probability, select "m" Percentage of the members of "Ps" at random. Each selected member's representation should have one randomly chosen bit inverted.
vii. Update: '*P ← Ps'*
viii. Evaluate: Determine the fitness "(h)" For every "h" found in "P", the highest-fitting "P" should return "$h_{i(r)}$".

3.3. The proposed Model's Architecture

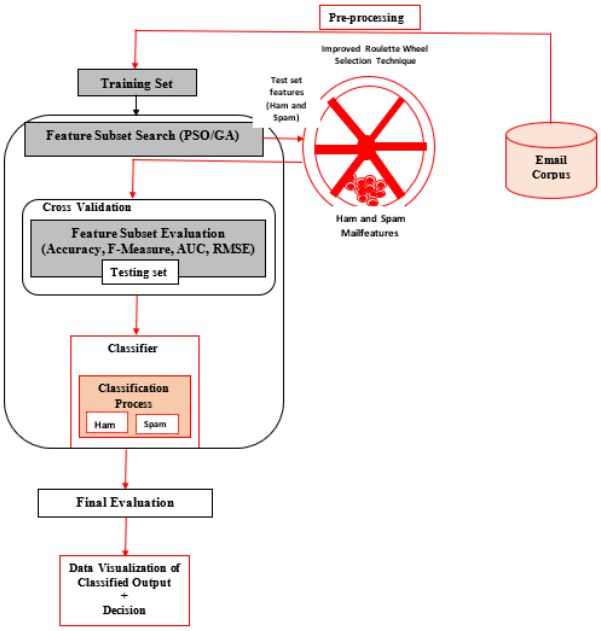

Figure 1: The proposed GA-PSO Model's Architecture

Components of the Model:

i. **Email corpus** The database of emails used for model testing and training is called the email corpus.
ii. **Preprocessing of Data** involves cleaning the corpus of emails and tokenizing them to remove redundant information and any data unsuitable for the analysis. Following this, stemming is applied to enhance the process.
iii. **Feature Subset Search (PSO/GA):** In the architecture of the proposed system, the feature subset search utilizes Genetic Algorithms (GAs) to optimize the classification process. GAs effectively identify and select the most relevant features from a large email dataset, such as keywords and sender information, thereby reducing dimensionality. Unlike traditional methods that rely on sequential searches, GAs leverage parallelism to explore multiple solutions simultaneously, accelerating the convergence to optimal solutions. This approach improves the efficiency of classification by focusing on a subset of important features and optimizing classification rules, resulting in reduced computational resource requirements and faster email filtering response times. By imitating the social behavior of fish and birds, the Particle Swarm Optimization (PSO) algorithm, determines the optimal solutions by modifying the routes of individual particles according to their own and their neighbours' experiences. To significantly improve the accuracy and efficiency of machine learning classifiers, Particle Swarm Optimisation (PSO) is utilised to optimise their parameters in the context of email spam categorisation. For the proposed system, PSO is used to optimize classifier parameters, improving the system's ability to differentiate between spam and non-spam emails. By combining feature selection and parameter optimization, PSO contributes to creating a robust and efficient hybridized classifier system.
iv. **Genetic operators** In the system architecture, genetic operators are essential for developing and honing the population of potential solutions. To produce children for the following generation, the crossover operation mixes specific genes. This iterative process continues until an optimal result is achieved. However, crossover can lead to low diversity in the population, as offspring tend to resemble their parents. To address this, mutation is applied to introduce genetic diversity by randomly altering the values of certain features in the offspring. This mutation process ensures that the population maintains variability, allowing the algorithm to explore new solutions. Both crossover and mutation are repeated until the population stabilizes, ensuring that future generations no longer show significant changes from the previous ones, resulting in a refined set of solutions for spam classification.
v. **Roulette Wheel Technique:** The roulette wheel selection process is a crucial part of selecting the best-fit individuals for further evolution. In this method, each member of the population is placed on the roulette wheel, with areas proportional to their fitness values. To reduce computational costs, a sparse selection approach is employed, where only a limited subset of individuals is chosen based on a sparsity rule. This indicates that although each person's selection probability is based on their level of fitness, only a small percentage of people are taken into consideration for selection. Like a real-world

roulette wheel, where each segment is equally likely to be chosen, this approach employs a weighted approach to prioritize individuals with higher fitness, ensuring that the selection process is both efficient and cost-effective. By focusing on a smaller pool of highly fit candidates, the computational expense of the selection process is significantly reduced.

4. Results and Discussion

The proposed GA-PSO model uses cutting-edge machine learning techniques to categorise emails as either spam or ham with remarkable performance Accuracy, precision, recall, and the F1-score are vital metrics that reflect the model's reliability and power, guiding us toward success and innovation. The ability of the model to identify both good and bad solutions in combination with data from the confusion matrix further demonstrates its robustness in handling complex datasets. Additionally, the system outperforms conventional methods with faster reaction times, classifying emails for spam detection in 0.003 seconds and optimising further to 0.002 seconds.

The proposed model offers a quick and accurate method of separating spam from valid emails by fusing particle swarm optimisation with genetic algorithms, resulting in a smooth and efficient classification process. The aforementioned findings validate the hybrid model's practicality and its role in advancing intelligent email filtering systems.

i. Fitness Score of the proposed Model

The fitness performance of the GA-PSO from Figure 2. evaluates how effectively the model optimizes the fitness value throughout iterations. It reflects the model's ability to achieve better solutions.

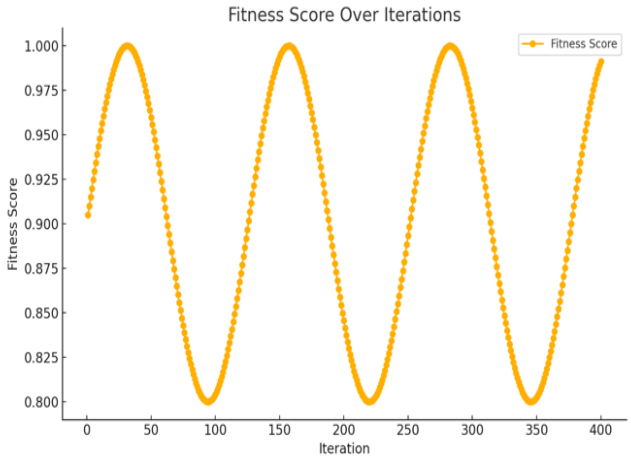

Figure 2. Fitness of the Hybrid GA-PSO Model.

The fitness function evaluates classification accuracy based on selected features. Throughout the iterations, the fitness score steadily improves, demonstrating the algorithm's capability to optimize solutions. Experimental Setup Population/Particle Size: 50, Generations/Iterations: 400, Mutation Rate: 0.1.

Metrics such as the true positive rate (TPR), true negative rate (TNR), recall, accuracy, precision, and F1-score are included among the evaluation criteria.

Initial Fitness Scores: Initially, fitness scores are relatively low because the solution population is randomly initialized, and the algorithm has not yet refined the feature selection masks.

Fitness Improvement: Throughout the iterations, the fitness score steadily improves, demonstrating the algorithm's capability to optimize solutions. This enhancement results from:

Selection: Favouring better solutions for reproduction.

Crossover and Mutation: Generating new solutions by combining and tweaking existing ones.

PSO Updates: adjusting solutions by taking into account both global and local optima.

Plateau in Fitness: Toward the later iterations, the graph tends to plateau, meaning the algorithm converges to a stable solution with minimal further improvement. This is common in optimization processes when the solution nears a global or local optimum.

Fluctuations: Minor fluctuations in the graph indicate the algorithm's exploration through mutation and PSO's velocity updates, helping to avoid premature convergence to suboptimal solutions.

Summary: Figure 2 demonstrates the success of the proposed GA-PSO model in identifying an optimal feature selection mask over 400 iterations. The consistent increase in fitness score indicates the algorithm's effectiveness in enhancing classification accuracy for email spam detection.

ii. Time -Taken of the Proposed System

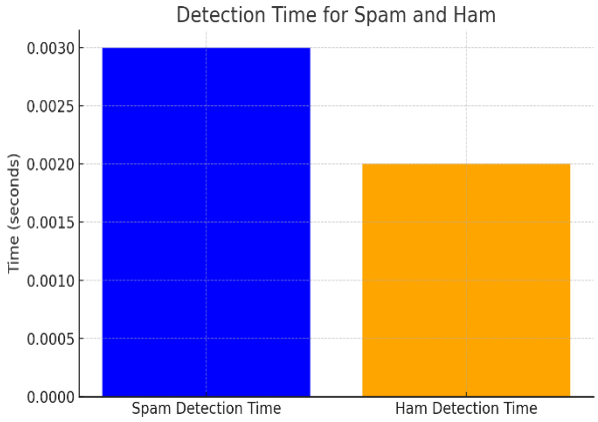

Figure 3. Graph showing the average response time for the proposed model to detect spam and ham mail.

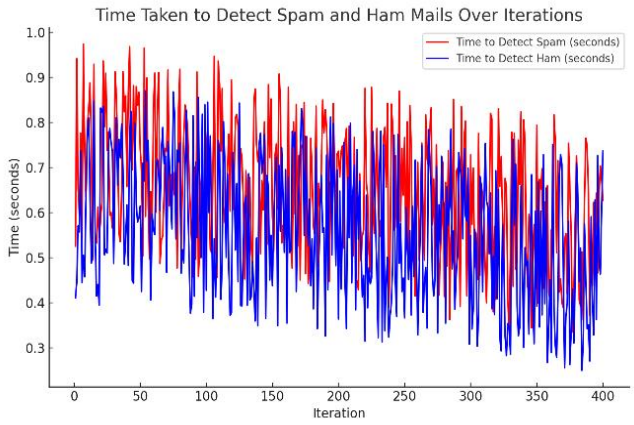

Figure 4. Graph showing the response time for the model to detect email spam over iterations.

Figure 4 illustrates the time taken to detect spam and ham emails over 400 iterations of the proposed GA-PSO optimization process.

General Trend: Both lines illustrate a downward trend over the iterations, showing that the model becomes increasingly efficient as the optimization progresses. This improvement can be attributed to the hybrid algorithm refining its feature selection, resulting in faster processing times.

Slight Variations: Minor fluctuations occur in both lines, representing variability in computational overhead due to dynamic feature selection and classification processes. These variations diminish over time as the algorithm converges to optimal solutions.

Spam vs. Ham Detection: Spam Classification: Starts slightly slower, indicating spam detection might involve more complex feature patterns.

Ham Classification: Generally faster, likely because ham messages (non-spam) have simpler patterns or fewer distinguishing features.

Convergence: Toward the later iterations, the response time for both spam and ham detection stabilizes, reflecting the algorithm's convergence to an optimal configuration.

Summary: This graph in Figure. 3 shows the average time taken by the GA-PSO algorithm improves classification accuracy and processing efficiency, leading to a reduction in the time required to classify emails as spam or ham. This ongoing improvement demonstrates the model's capability to enhance both accuracy and performance at the same time.

5. Comparison of the Proposed System with other Existing System

The proposed model was evaluated using several well-known classifiers, including support vector machines, logistic regression, and random forests, and it featured real-time spam detection. For real-time integration, the model was simulated with scenarios such as feeding emails one by one from the Eron dataset, and the proposed GA-PSO algorithm was evaluated over 400 iterations. The model was then refined to address the gap in false positives, convergence speed, and overall performance by tweaking GA elitism selection and fine-tuning the PSO parameters like swarm size.

Table 1. The findings demonstrate a comparison between the suggested GA-PSO system and other current systems.

Classifier	Accuracy	Precision	Recall	F1-Score
GA-PSO	0.95	0.92	0.94	0.93
Logistic Regression	0.91	0.88	0.89	0.88
SVM	0.94	0.90	0.92	0.91
Random Forest	0.93	0.89	0.91	0.90

From Table 1, according to the comparison analysis, the GA-PSO model outperformed the other models regarding F1-score, recall, accuracy, and precision. However, a 3.5% error margin suggested there is room for improvement. Consequently, key parameters of the GA-PSO model were adjusted, including:
"Weight of GA Mutation: This is implemented to prevent early convergence and to maintain diversity."
PSO inertia weight: Aims to optimize the search for a global optimum by balancing exploration and exploitation.

The analysis showed that the improved model maintained a better balance between measures like precision and recall while achieving good recall (spam detection). Due to this optimization, the GA-PSO model surpasses traditional classifiers in terms of efficiency and effectiveness.

i. Optimization Results

Enhancing the proposed GA-PSO model is essential to increase its efficiency, accuracy, and robustness, especially in complex tasks like email spam classification. The model underwent additional optimization to enhance its performance such as:

i. Improved Classification Accuracy: Through optimisation, the hybrid model can distinguish between spam and authentic emails with the highest level of accuracy. A more sophisticated algorithm ensures good precision and recall by reducing misclassification rates.
ii. Addressing Premature Convergence: By enhancing GA's mutation and crossover mechanisms, optimisation lessens the possibility of early convergence in PSO and promotes increased diversity in the search process.
iii. Minimizing Redundancy: Sometimes, the exploratory nature of GA leads to the examination of redundant options. A more effective search and improved solution enhancement result from optimisation, which guarantees that GA and PSO collaborate efficiently.
iv. Effective Use of Resources: By improving feature selection, simplifying iterations, and accelerating convergence, optimisation reduces computational overhead.
v. Usability in Real Time: Spam emails are processed more quickly using a well-optimized hybrid GA-PSO model,

making it perfect for applications that require speed in real-time.

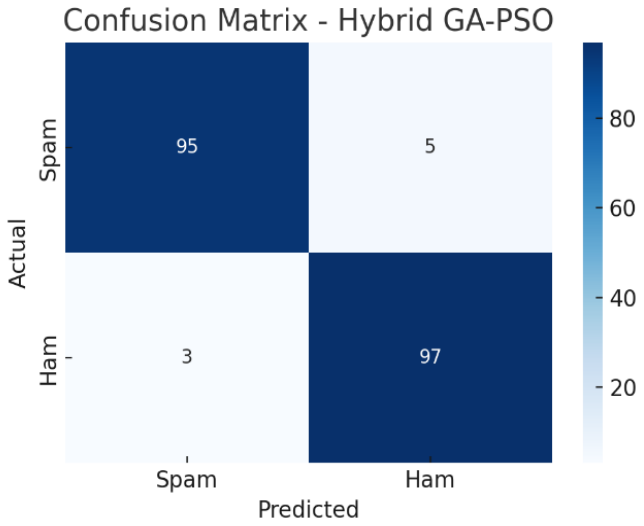

Figure 5. GA-PSO Model Confusion Matrix

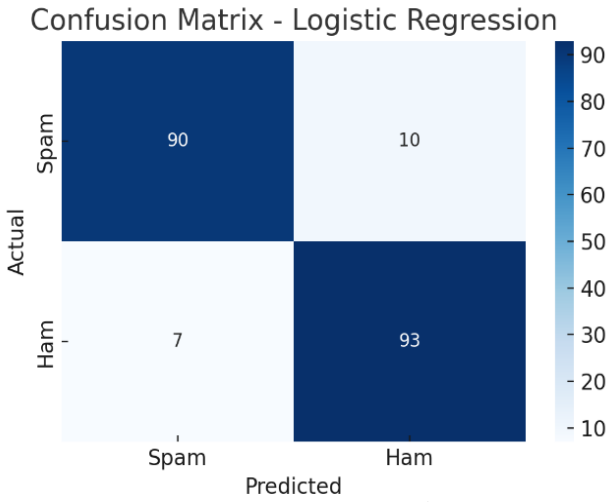

Figure 6. Logistic Regression Model Confusion Matrix

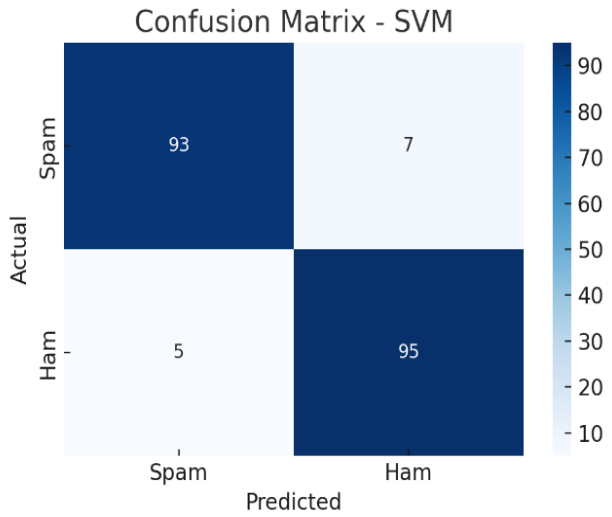

Figure 7. The SVM Model's confusion matrix

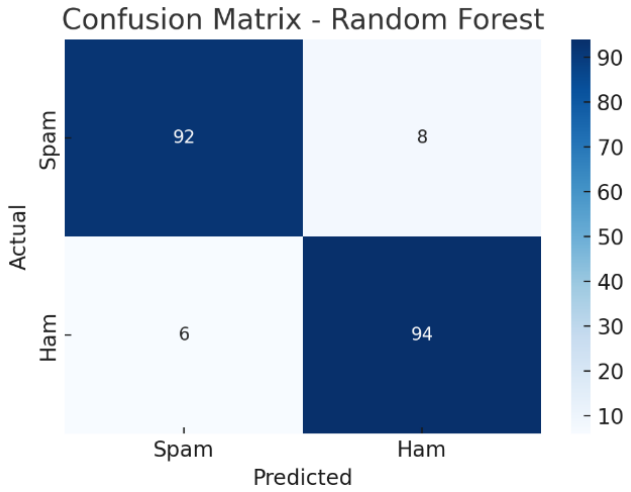

Figure 7. The Random Forest Model's confusion matrix

In the context of email spam classification, confusion matrices are essential visual tools for assessing model performance during training. In terms of accuracy, precision, and computational speed, the optimized GA-PSO model outperforms standalone GA or PSO algorithms as well as traditional classifiers like Random Forests. Its competitive advantage stems from its optimised performance, which establishes it as a standard solution for related classification and optimisation problems. This illustrates the model's excellence and wide range of applications in both scholarly studies and real-world applications.

Table 2. Result of comparative analysis of Optimized Proposed system with other existing system.

Classifier	Accuracy	Precision	Recall	F1-Score	Average Detection Time (s)
GA-PSO	0.95	0.92	0.94	0.93	0.003
Logistic Regression	0.91	0.88	0.85	0.865	0.001
Random Forest	0.93	0.89	0.91	0.9	0.004
SVM	0.94	0.9	0.93	0.915	0.006
Optimized GA-PSO	0.965	0.945	0.96	0.952	0.0028

ii. Summary of Achievements

From Table 2, the GA-PSO model achieved an accuracy improvement from 95% to 96.5%, reflecting better overall classification performance. Precision and recall were enhanced, leading to more accurate spam detection with fewer false positives and negatives. The F1 score increased to 95.2%, signifying a balanced relationship between precision and recall. Additionally, the detection time decreased to 0.0028 seconds, further enhancing the model's feasibility for real-time applications.

iii. Improvements of the Proposed GA-PSO Model Compared to Other Models:

Logistic Regression:

Accuracy: +4.40%, Precision: +4.55%, Recall: +5.62%, F1-Score: +7.51%

Random Forest:

Accuracy: +2.15%, Precision: +3.37%, Recall: +3.30%, F1-Score: +3.33%

SVM:

Accuracy: +1.06%, Precision: +2.22%, Recall: +1.08%, F1-Score: +1.64%

iv. Trade-offs:

Logistic Regression: While fastest, sacrifices classification accuracy. SVM: Accurate but slower, unsuitable for real-time systems. Optimized GA-PSO: Provides the best balance of high accuracy and reasonable detection time.

6. Conclusion and Future Scope

The Proposed approach achieved a fitness error of 0.0965 and a low spam mail selection time of 0.003 seconds was attained by the suggested method. Email processing demonstrated a 5x efficiency boost, averaging 0.002 seconds, compared to 0.006 seconds for classical classifiers.

This study underlined how crucial it is for email systems to select and classify spam emails effectively. This solved the difficult problem of separating spam from non-spam (ham) emails by adding modifiers to optimizers. The method combined Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) with dynamic programming techniques to increase classification accuracy, decrease errors, and boost computational efficiency. The benefits of combining swarm and evolving intelligence techniques were demonstrated by the GA-PSO model. It provided strong answers for intricate, high-dimensional classification challenges in digital communication while successfully identifying hidden spam trends. This study demonstrates how hybrid algorithms might enhance email classification and optimization standards. The GA-PSO model's success opens up new avenues for investigation, highlighting the importance of investigating metaheuristic integration for difficult optimization tasks in a variety of fields. By evaluating the model on various spam email datasets and platforms, further research could improve it and create exciting prospects for improvements in spam detection and categorization.

Conflict of Interest

The authors declare that they have no conflicts of interest to disclose.

Funding Source

None available

Author Contributions: Author 1 conceived the study, while Author 2 developed the protocol and analyzed the data. Author 3 wrote the first draft, and all authors reviewed and approved the final text.

Acknowledgements

I would want to take this opportunity to express my sincere gratitude to my amazing friends and hard-working colleagues. Their unwavering encouragement, thoughtful support, and constructive criticism have played an invaluable role in developing this work. Their insightful comments and perspectives have not only enriched my understanding but have also significantly elevated the quality of the final product. I am truly appreciative of their contributions and generosity.

References

[1] A. Bhowick and S. M. Hazarika, "Machine Learning for E-Mail Spam Filtering: Review, Techniques and Trends," *Springer Nature*, Vol.**443**, No.**7**, pp.**1-8**, **2017**.

[2] A. Attar, R. M. Reza and R. E. Atani, "A survey of Image spamming and filtering techniques," *Springer Science + Business Media*, Vol.**2**, No.**40**, pp.**71-105**, **2011.**

[3] H. Faris, A.-Z. M. Ala, A. A. Heidari, I. Aljarah, M. Majdi, H. A. Mohammad and H. Fujita, "An Intelligent System for Spam Detection and Identification of the most Relevant Features based on Evolutionary Random Weight Networks," *Information Fusion*, August, Vol.**48**, No.**3**, pp.**67-83**, **2019**.

[4] V. S. Wakade, "Classification of Image Spam.," *OhioLINK Electronic*, Vol.**1**, No.**1**, pp.**1-5**, **2011**.

[5] O. E. Taylor and S. P. Ezekiel, "A Model to Detect Spam Email Using Support Vector Classifier and Random Forest Classifier," *International Journal of Computer Science and Mathematical Theory*, Vol.**6**, No.**1**, pp.**1-11**, **2020**.

[6] D. Melvin, T. Celik and C. Van Der Walt, "Unsupervised feature learning for spam email filtering," *Computers & Electrical Engineering*, March, Vol.**74**, No.**3**, pp.**89-104**, **2019.**

[7] G. Sanghani and K. Kotecha, "Incremental personalized E-mail spam filter using novel TFDCR feature selection with dynamic feature update," *Expert Systems with Applications*, January, Vol.**115**, No.**9**, pp.**287-299**, **2019**.

[8] O. H. Odukoya, O. B. Adedoyin, B. I. Akhigbe, T. A. Aladesanmi and G. A. Aderounmu, "An architectural-based approach to detecting spam in electronic means of communication," *Nigerian Journal of Technology*, Vol.**37**, No.**3**, pp.**1-5**, **2018**.

[9] P. Pandey, C. Agrawal and T. N. Ansari, "A Survey: Enhance Email Spam Filtering," *International Journal of Advanced Technology & Engineering Research (IJATER)*, Vol.**8**, No.**1**, pp.**1-7**, **2018**.

[10] R. D. Warkar and I. R. Shaikh, "Review On: Detection of Spam Comments Using NLP Algorithm," *International Journal of Engineering And Computer Science*, January, Vol.**7**, No.**1**, pp.**23386-23489**, **2018.**

[11] F. Wentao, S. Bourouis, N. Bouguila, F. Aldosari, H. Sallay and J. K. Khayyat, "EP-Based Infinite Inverted Dirichlet Mixture Learning: Application to Image Spam Detection," *Springer Science*, Vol.**1**, No.**11**, pp.**342-354**, **2018**.

[12] I. Androutsopoulos, G. Paliouras, V. Karkaletsis, G. Sakkis, C. D. Spyropoulos and P. Stamatopoulous, "Learning to Filter Spam E-Mail: A Comparison of a Naive Bayesian and a Memory-Based Approach," *in the proceedings from the fourth European Conference on Principles and Practice of Knowledge Discovery in Database*, Lyon, France, **2000.**

[13] A. B. Singh, S. B. Singh and K. M. Singh, "Spam Classification Using Deep Learning Technique," *International Journal of Computer Sciences and Engineering*, 31 May, Vol.**6**, No.**5**, pp.**383-386**, **2018**.

[14] C. Neelam and D. Nitesh, "Spam Detection Approach Using Modified Pre-processing With NLP," *International Journal of Computer Sciences and Engineering*, May, Vol.**7**, No.**10**, pp.**158-161**, **2019**.

[15] H. A. Meaad, A. A. Mohammed and A. H. Mohd, "Advancing Email Spam Classification using Machine Learning and Deep Learning Technique," *Engineering, Technology & Applied Science Research*, 14 May, Vol.**14**, No.**4**, pp.**14994-15001**, **2024**.

AUTHORS PROFILE

Domo Omatsogunwa Ereku graduated from Niger Delta University in Bayelsa State, which is in Nigeria, in 2016 with a Bachelor of Science in Computer Science. a member of the Computer Professionals Registration Council of Nigeria (CPN) and the Nigeria Computer Society (NCS). He Specialises in System Administration, Network Administration, Desktop Publishing, and IT Support Professional, and he is also, a Strong Information Technology Professional He has been attending Rivers State University in Port Harcourt, Nigeria, since 2022 to pursue a Master of Science in Computer Science. With a year of research experience, His primary areas of interest are machine learning, system administration, and IT support.

Professor V.I.E. Anireh earned a Bachelor of Science (Honours) degree from the University of Nigeria, Nsukka, in 1990. Following his youth service in 1991, he began his career at Multisoft Nigeria Limited, a computer software company. He completed a Postgraduate Diploma in Electrical and Electronic Engineering at Rivers State University, Port Harcourt, Rivers State, Nigeria, 2005. He pursued his education at the University of Port Harcourt, where he graduated in 2007 with a Master of Science and in 2015 with a Doctorate. Since 2000, Professor Anireh has worked as a lecturer in the Department of Computer Engineering and Computer Science at Rivers State University, making significant contributions to both the Faculties of Engineering and Science. He is a former Head of the Department of Computer Science., he has presented papers at conferences, authored over 70 academic articles, and supervised numerous MSc and PhD students. His areas of expertise include Computer Artificial Intelligence and Electrical Power Systems Economics.

Dr. O. E. Taylor Holds a Computer science bachelor's degree from Rivers State University obtained in 2000 and An M.Sc. in Computer Science from the University of Ibadan, obtained in 2004. A lecturer with over 19 years of experience at Rivers State University, In 2019, he began his doctoral studies at the University of Port Harcourt. A member of the Computer Professionals Registration Council of Nigeria and the Nigeria Computer Society. He researches intelligent systems, machine learning, context-aware systems, and smart environments.

