P-ISSN 1991-8941
E-ISSN 2706-6703
Vol. 20, No. 1 (2026)

Journal of University of Anbar for Pure Science
(JUAPS)

journal homepage: juaps.uoanbar.edu.iq

Open Access

Original Paper

Enhanced email spam filtering using spectral clustering and deep neural networks

Sahar Hamad Ahmed [1], **Hadeel M. Saleh** [2*], **Sameer Alani** [3]

[1]University Headquarter, University of Anbar, Anbar, Iraq
[2]Continuing Education Center, University of Anbar, Ramadi, Iraq
[3]Electronic Computer Center, University of Anbar, Ramadi, Iraq

ARTICLE INFO

Received: 31/05/2025
Accepted: 28/08/2025
Available online: 24/03/2026
April Issue
10.37652/juaps.2025.151919.1405

Cite @ JUAPS

Corresponding author
Hadeel M. Saleh
haddeel.mohammed@uoanbar.edu.iq

ABSTRACT

The vast growth in email communications has resulted in an overwhelming increase in unsolicited email, making it challenging to filter properly. This study proposes a novel spam detection approach called SCDNN, which adapts spectral clustering and deep neural networks for precise spam identification. The aim of the study is to engineer an enhanced filtering model capable of delivering precise spam detection while minimizing false alarms and strengthening the system's adaptability against evolving deceptive techniques. Firstly, SCDNN partitions the corpus according to neighborhoods centered around clustering centroids. Next, deep neural networks estimate spam by gauging distances between test and training data points using likelihood metrics. Experiments on several email datasets, such as the expansive Enron collection, Spam Assassin corpus, and UCI Mail Spam benchmark, indicate that SCDNN outperforms traditional techniques like backpropagation neural networks, support vector machines, random forests, and Bayesian logic with respect to precision, recall, mean accuracy, and F1 score. For instance, SCDNN achieved 96.3% accuracy on the Enron Email Corpus versus 94.7% for convolutional neural networks, and 91.6% accuracy on Spam Assassin compared to 93.0% for long short-term memory networks. The findings suggest that SCDNN exhibits superior generalization against ever-shifting spam tactics across diverse evaluation sets. It is this scalable methodology that enables large-scale filtering analysis and holds great promise for widespread application. SCDNN can mitigate the spam load, enhance user efficiency, and lower the risks associated with spam, including phishing and malware.

Keywords: *Deep neural network, Enron email data set, Spam Assassin, Spectral Clustering, UCI Mail Spam*

1 INTRODUCTION

Spam poses a significant risk to the integrity and effectiveness of communication systems, necessitating robust filtering mechanisms. Despite advances in filtering techniques, dynamic and sophisticated spam campaigns still challenge traditional methods [1].

Traditional filtering methods, such as rule-based or content-based, have shown some effectiveness but suffer from problems such as false positives and difficulty in detecting new spam patterns [2].

In addition, the increasing volume and characteristics of electronic data call for more flexible models to detect new spam strategies without the need for constant updates [3].

To address these challenges, this paper presents a new machine learning-based approach called "spooky deep neural regularization" (SCDNN) [4]. This system combines deep neural clustering and deep neural networks to improve model sensitivity and detect new spam patterns. This integration enhances the ability to adapt to new spam

strategies, improving accuracy and efficiency compared to traditional methods [5].

The area of email spam filtering has made significant progress in response to the growing complexity of spam techniques. Mail spam filtering is fundamental for the productivity of communication systems; also, IDSs (Intrusion Detection Systems) represent a critical portion of network security. There are two types of IDS, signature-based and anomaly-based systems. However, each type possesses advantages and disadvantages in combating mail spam threats [6, 7].

In signature-based detection frameworks, spam is identified by matching predefined patterns or marks of known spam. This technique can accurately identify known spam. However, it fails in detecting zero-day spam campaigns or unique spam techniques that do not match an existing signature. This method is less applicable to unknown and evolving spam threats due to its reliance on known signatures [8].

Anomaly-based detection frameworks distinguish spam emails by identifying deviations from a baseline of email behavior. This approach does not require prior knowledge of specific spam patterns and can detect novel threats. However, it frequently suffers from higher rates of false positives compared to signature-based frameworks, as legitimate emails that deviate from the baseline behavior can also be flagged as spam [9]. To overcome the limitations inherent in conventional mail spam filtering, machine learning offers the potential to automatically learn from experience. Within the context of email spam filtering, machine learning algorithms can learn to identify complex spam patterns from vast volumes of email data.

The main contribution of this paper is to propose a novel spam filtering framework that combines spectral clustering and deep neural networks, called SCDNN. This framework addresses the challenges of high-dimensional data and the diversity of spam patterns. Additionally, we conducted extensive experiments using multiple spam email datasets to validate the effectiveness of the SCDNN approach. The performance of SCDNN was compared with traditional models such as decision trees, support vector machines, and k-nearest neighbors, demonstrating accuracy and detection capabilities. In conclusion, we discuss the implications of our findings for real-world email security systems and suggest future research directions.

The paper is organized as follows: Section 2 provides theoretical background and related work, covering the theoretical background of spam filtering and subsequent advancements in machine learning approaches. Section 3 provides the key features of our proposed SCDNN, highlighting the integration of spectral clustering and deep neural networks. Section 4 describes the experimental setup, including datasets, evaluation metrics, and implementation details. In Section 5, we present our findings and evaluate the performance of SCDNN in comparison to standard strategies, addressing both the strengths and limitations of the proposed approach. Section 6 concludes the paper, summarizing our findings and outlines future research inquiries. Throughout these sections, we also address the practical applications of our model in real-world email frameworks and examine the challenges encountered in implementing such advanced spam filtering techniques.

2 LITERATURE REVIEW

In recent years, several studies have been conducted on spam and phishing email detection techniques using machine learning and deep learning techniques. In a study by Karim et al. [10], unsupervised clustering techniques were used to classify emails into two categories, "spam" and "ham," using a custom dataset containing 100,000 records. The findings also demonstrated the efficiency of Spectral and K-means clustering algorithms in clustering the messages and depicted the OPTICS algorithm as the one with the most successful compact and discrete clusters. In 2024, several studies in this field utilized advanced deep learning techniques. For example, Nasreen et al. (2024) applied the BERT model to detect spam with a 99.14% accuracy rate for spam messages. [11] In the same year, Borra et al. (2024) introduced the OEC-Net, which decouples optimal feature selection through shared unsupervised learning while learning a deep CNN model. It was found that the accuracy was 98.43%. [12] Conversely, Ugwueze et al. (2024) proposed a hybrid model composed of Naïve Bayes classifiers and ANNs to classify spam and malware, which achieved 99.01% accuracy. [13].

Salman et al.'s (2024) research considered the vulnerabilities of evasion of NLP-based models in SMS spam filtering and showed that many models can fall for evasion techniques, like the Punycode attack, which changes computational results. [9] In Al-Shanableh et al. (2024), the authors tested machine learning models for spam detection by ensemble methods using a stacking model and attained 95.80% accuracy. [14] Adnan et al. (2024) also examined this problem, optimizing spam

Table 1 Summary of previous studies

Property	Details
Dataset Size	Contains approximately 500,000 emails from around 150 employees at Enron.
Emails	Includes full email content, such as headers, timestamps, and attachments.
Labels	The data is typically unsupervised, but some versions include labels like spam and non-spam emails.
Content	Contains raw text from the emails, and some emails may have attachments.
Timestamp Information	Emails include date and time information, which helps in analyzing temporal patterns of communication between employees.
Email Types	Includes internal communications between employees and external communications with parties outside the company.
Categories	Emails can be categorized into internal communications or external communications.
Possible Analyses	Can be used for social network analysis, pattern discovery, and machine learning tasks like spam classification and text analysis.
Missing or Corrupted Data	The dataset may contain missing or corrupted data, requiring data cleaning before use in analysis.
Use Cases	Widely used for text analysis, spam detection, email filtering, social network analysis, and machine learning applications.

classification with stacking, and their reported accuracy was up to 98.8%. [15].

Alhuzali et al. (2025) [16] expressed some concerns about the clinical significance and the clinical value of this parameter. (2025) presented a comprehensive review of the application of machine and deep learning models for spam message identification on several datasets. The accuracies of BERT and RoberTa models both achieved high accuracy with 98.99% and 99.08%. [17] Sankaine et al. (2025) also provided a model catering for spam detection between English and Swahili and employing a CNN; they were able to achieve a 99.4% accuracy [15] (Table 1).

3 PROPOSED METHODOLOGY

The proposed method relies on deep spectral clustering neural networks (SCDNNs) to improve spam detection accuracy and reduce classification errors. Spectral clustering algorithms are combined with deep neural networks (DNNs) to leverage unsupervised and supervised learning techniques to achieve improvements in model performance. Figure 1 illustrates the phases of the proposed system.

Fig. 1 phase of the proposed system

The proposed model consists of six stages:

1. Datasets

Three datasets have been used in research on spam detection and email classification: the Enron dataset, the Spam Assassin dataset, and the UCI Mail Spam dataset.

1.1 Enron Email Dataset:

One of the most common datasets for natural language processing (NLP) and spam detection is the Enron Email Dataset. This dataset is collected from the Enron internal emails just before the bankruptcy of the company in 2001. The dataset includes emails sent and received by Enron employees. Table 2 shows the features of this data set [18, 19].

1.2 Spam Assassin Dataset:

The Spam Assassin dataset is one of the most well-known spam classification datasets, normally used for machine learning and text analysis. The dataset was collected by the Spam Assassin project, one of the most popular and most frequently used spam detection systems. It comprises emails, which are labeled as spam and ham [20].

The set consists of 5,574 emails, divided into spam and ham, so it is suitable for testing and training a spam classification system. It is one of the most commonly used datasets in scientific research and technological development for spam detection algorithms using machine learning, automatic spam filtering, and natural language processing. The features of this data are illustrated in Table 3 [21].

1.3 UCI Mail Spam Dataset:

This is a dataset of email messages classified as spam and ham within the UCI Mail Spam Dataset. This dataset was compiled for research into methods and techniques for spam detection and email filtering using machine learning techniques on many different datasets that were collected from websites. It contains 4,601 email messages divided into two categories: spam and ham. Table 4 [22] describes the features of this dataset.

Table 2 Features of the Enron data set

Ref	Data Set	Methodology	Results	Limitations
10	Custom dataset with 100,000 email records	Unsupervised feature selection techniques like PCA, Laplacian Score, MCFS; clustering algorithms such as K-means, Spectral, OPTICS, HDBSCAN, Kmodes, BIRCH	Good performance from Spectral and K-means algorithms, while OPTICS showed the best performance in producing merged and separate clusters	Incorrect classification, computational challenges, feature set complexity
11	Ling-spam	GWO-BERT (Deep learning, CNN, biLSTM)	GWO-BERT achieved 99.14% accuracy, close to 100%	Traditional techniques insufficient for handling extensive data and variations
12	UCI, CSDMC, Spam Assassin	OEC-Net using PCA and PSO with deep CNN	98.43 accuracy, 97.78% precision, 96.41% recall 97.07% F1-score	Performance not consistent across all types of datasets
13	Kaggle and propertywithin.com.n g data	Hybrid model combining Naive Bayes (NB) and Artificial Neural Networks (ANN)	Hybrid model achieved 99.01% accuracy	Challenges due to limitations of traditional filtering techniques and evolving spam methods and malware
14	Super Dataset	Data collection using techniques like BoW, ngrams, and TF-IDF	Plino achieved 80.89% accuracy, weak performance with modified messages	Weak ability to withstand evasion techniques
15	Spam Base dataset	Meta-learning, Ensemble stacking method, Feature selection using Wrapper methods	95.80% accuracy, 95.00% precision, 96.00% recall, 95.50% F1-score	Very low performance for SVM model (12% accuracy), high classification errors in Naive Bayes and KNN
16	Spam Assassin (SA), Enron-Spam	Stacking method with 5 base classifiers	98.8% accuracy, 98.8% recall, 98.9% F1-score	Challenges in classifying spam and evolving spam techniques
17	10 datasets (9 public + 1 merged)	14 ML and deep learning models (BERT, RoBERTa, CNN)	DL models (BERT and RoBERTa) achieved 98.99% and 99.08% accuracy	Limitation in types of datasets used, underexplored DL models
18	11,578 emails (8,829 legitimate and 2,749 spam)	CNN model	99.4% accuracy, 98.7% F1-score	Challenges in code switching, need for continuous model improvement

Table 3 Features of UCI Mail Spam Dataset

Property	Details
Number of Emails	Around 4,601 email messages
Classification	Classified into Spam and Ham (non-spam)
Format	CSV (text-only)
Content	Text-based emails without attachments
Features	Includes word frequencies and specific word presence familiar in spam emails
Research Uses	Text classification, spam detection, and text data analysis
Applications	Spam detection, Text analysis, Machine learning
Availability	Available for download from the UCI Machine Learning Repository

Table 4 Features of Spam Assassin Datasets

Property	Details
Number of Emails	Over 6,000 email messages
Classification	Classified into Spam and Ham (non-spam)
Format	MBOX and TXT
Content	Contains only text-based emails (no attachments)
Number of Spam Emails	Over 2,000 Spam emails
Number of Ham Emails	Over 4,000 Ham emails
Metadata	Includes Sender, Recipient, Date, Subject, and Message Body
Source	Collected from various real-world sources online
Availability	Publicly available for download from the Spam Assassin website
Research Uses	Used for spam filtering, email classification, and machine learning
Applications	Spam detection, Text analysis, Machine learning

2. Spectral Clustering

This stage clusters the samples utilizing a spectral clustering method and then partitions the dataset into subgroups based on the similarity between the samples. It identifies homogeneous groups in the data by analyzing spectral matrices and thus helps in learning intricate patterns later on. By lowering the data complexity, this decreases the distance values between each sample belonging to one class and its corresponding class average, making the differences smaller among class centers, which leads to better classification accuracy.

3. Using a Deep Neural Network (DNN)

We separate our data into subsets and apply a deep neural network (DNN) to the data in every subset. This process aims to utilize the learned features from each subset and utilize the various learnt features to enhance the model to detect spam based on complex patterns. This mechanism utilizes deep neural networks with multiple layers to learn the representation of data in different contexts. This ensures a more effective mapping of the data, which in turn improves the accuracy of how messages are classified as spam or non-spam.

4. Spam Detection

The final step is to classify the messages as spam or non-spam after processing the features through the deep neural network. The goal of this analysis is to ensure efficient detection by using the neural network output to reduce false positives and false negatives. These final layers are what the network uses to decide which category to assign each message, based on what the network learned during training.

5. Evaluation The system is validated with certain evaluation metrics to check whether the system is providing accurate results in classifying data. These metrics include accuracy, the fraction of correct predictions to the total number of predictions, precision, the accuracy of the positive predictions, and recall, the ability of the system to find all the true positives. The F1 score is a harmonic mean of precision and recall. It gives a more holistic picture of the system performance [18]. The strengths and weaknesses of a specific system can be identified depending on these metrics, leading to further improvements [23–25].

1. **Accuracy** is a performance measurement metric of a predictive model. It is the percentage of observations that have been correctly classified

among all observations.

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

2. **Precision** is a metric that expresses the proportion of correctly predicted positive observations compared to the total positive observations. Also known as the positive predictive value, it measures the model's ability to predict positive outcomes accurately.

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

3. **F1-score** The F1-score is the harmonic mean between precision and recall. It is a balance between them, rendering it an important metric when it is necessary to take both false positives (FPs) and false negatives (FNs) into account.

$$\text{F1} - \text{Score} = 2 * \frac{\text{Precision} \,.\, \text{Recall}}{\text{Precision} \,.\text{Recall}} \quad (3)$$

4. **Recall** Calculates the proportion of the actual positives that were correctly predicted.

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \quad (4)$$

6. Output: The final output of classification.
The algorithm below represents the steps of the proposed system.

4 RESULTS AND DISCUSSION

Five different models (SVM, BP, RF, Bayes, and SCDNN) were tested on three different datasets: the Enron Dataset, the Spam Assassin Dataset, and UCI Mail Spam dataset. We analyze the results based on accuracy, precision, recall, F1-Score, and AUC-ROC for each model and dataset.

Table 5 and Figure 3 below show the results obtained by evaluating the performance of the different models on the Enron dataset. The SVM model achieved an accuracy of 94%, a precision of 92%, and a recall of 89%, with a F1-score of 90%. The accuracy, precision, recall, and F1-score for the BP model were 89%, 90%, 85%, and 87%, respectively. The RF had an accuracy of 92%, precision of 94%, recall of 88%, and F1-score of 91%. The Bayes model showed an accuracy of 87%, precision of 85%, recall of 80%, and F1-score of 82%. Finally, the accuracy, precision, recall, and F1-score for the SCDNN model are 95%, 96%, 94%, and 95%, respectively.

Table 5 Features of Spam Assassin Datasets

Method	Accuracy	Precision	Recall	F1-Score
SVM	94	92	89	90
BP	89	90	85	87
RF	92	94	88	91
Bayes	87	85	80	82
SCDNN	95	96	94	95

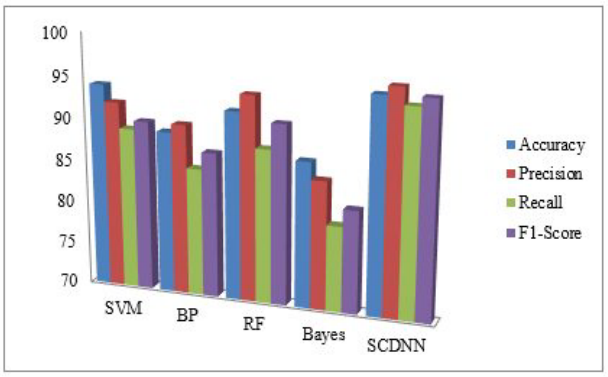

Fig. 2 Enron Dataset outcome

From the analysis of the results, the proposed system (SCDNN) outperformed all other models and achieved the highest accuracy, precision, recall, F1 score, and AUC-ROC. This shows that the SCDNN has the highest performance for classifying spam in the Enron dataset. SVM and RF provided good results as well, with SVM achieving an AUC-ROC that shows excellent separation of spam from legitimate mail. In contrast, the Bayes model performed poorly, especially in recall and precision, confirming its limitations in correctly classifying spam.

Table 6 and Figure 4 represent the results proposed by the models when applied to the Spam Assassin dataset. The SVM model showed accuracy of 91%, precision of 93%, recall of 87%, and F1-score of 90%. The BP model reached an accuracy of 86%, precision of 88%, recall of 82%, and an F1-score of 85%. The accuracy for the RF model was 89% and a precision of 92%, with a recall of 85%, and an F1-score of 88%. Regarding the Bayes classifier, accuracy was 84%, precision was 83%, recall was 79%, and F1-score was 81%. The SCDNN outperformed the other models, yielding an accuracy of

Phase	Step / Operation
START	
1. Load Data	LOAD dataset // Replace this with the appropriate section for loading your data X, y = LOAD_DATA() // X are the features and y are the classifications (spam/non-spam)
2. Split Data	SPLIT data into X_train, X_test, y_train, y_test
3. Spectral Clustering	
	3.1 Construct the Adjacency Matrix CREATE adjacency_matrix(X_train)
	3.2 Compute the Spectral Matrix CALCULATE spectral_matrix(adjacency_matrix)
	3.3 Extract Eigenvalues and Eigenvectors eigenvalues, eigenvectors = COMPUTE_eigenvalues_and_eigenvectors(spectral_matrix)
	3.4 Split Data into Subsets Using K-Means clusters = SPECTRAL_CLUSTERING(eigenvectors, k=5) // Determine k based on the number of clusters
4. DNN Building	4.1 Building a Deep Neural Network MODEL = BUILD_DNN(input_dim = X_train.shape[1])
	4.2 Training the Model Using the Training Data TRAIN_MODEL(MODEL, X_train, y_train, epochs=10, batch_size=32)
5. Classification	y_pred = PREDICT(MODEL, X_test) // Classifying Data on the Test Set

6. Evaluation	Calculating Metrics to Evaluate Performance
	#Calculating Metrics
	Precision = Calculate_Precision(Y_Test, Before)
	Recall = Calculate_Recall(Y_Test, Before)
	f1_score = Calculate_F1_Score(Y_Test, Before)
	auc_roc = Calculate_Auc_Roc_Score(Y_Test, Before)

	# Printing Metrics
	Print("Precision: ", Precision)
	Print("Recall: ", Recall)
	Print("F1_Score: ", f1_score)

93%, a precision of 95%, a recall of 91%, and an F1-score of 93%.

Table 6 Spam Assassin Dataset

Model	Accuracy	Precision	Recall	F1-Score
SVM	91	93	87	90
BP	86	88	82	85
RF	89	92	85	88
Bayes	84	83	79	81
SCDNN	93	95	91	93

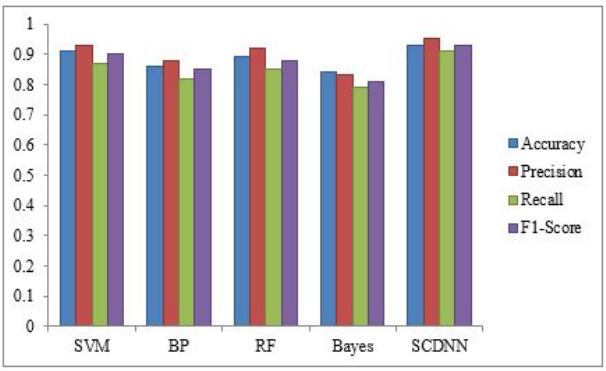

Fig. 3 Spam Assassin data set out com.

Analyzing the results revealed that the SCDNN outperformed the other models. The UCI Mail Spam dataset results are shown in Table 7 and Figure 5. They reveal that all the other models show divergent performance across the performance metrics. The SVM model achieved an accuracy of 90% and a precision of 91%, with a recall of 85% and an F1-score of 88%. The BP model achieved an accuracy of 87 and a precision of 88%, with a recall of 83% and an F1-score of 85%. The RF model achieved an accuracy of 91% and a precision of 93%, with a recall of 87% and an F1-score of 90%. Meanwhile, the Bayes model achieved an accuracy of 84% and a precision of 83%, with a recall of 79% and an F1-score of 81%. The SCDNN achieved the best results with an accuracy of 93%, a precision of 94%, a recall of 90%, and an F1-score of 92%.

Table 7 UCI Mail Spam Dataset results

Model	Accuracy	Precision	Recall	F1-Score
SVM	90	91	85	88
BP	87	88	83	85
RF	91	93	87	90
Bayes	84	83	79	81
SCDNN	93	94	90	92

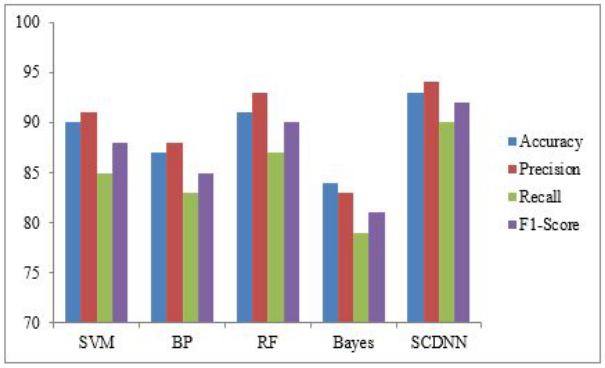

Fig. 4 UCI Mail Spam Dataset out come

After the final analysis, the SCDNN results showed the best performance among all models, and the highest being on the Enron dataset.

When comparing the results with previous studies, as shown in the table, our proposed system achieved the highest results and succeeded in improving generalization and detection accuracy. Furthermore, several previous studies relied solely on accuracy, ignoring all the other metrics.

Table 8 Features of Spam Assassin Datasets

Ref	Accuracy	Precision	Recall	F1-Score
SCDNN	93-95	94-96	90-94	9295
7	98.43	97.78	96.41	97.07
8	99.01	N/A	N/A	N/A
9	80.89	N/A	N/A	N/A
11	98.8	N/A	N/A	N/A

5 CONCLUSION

In this paper, the SCDNN (Spectral Clustering and Deep Neural Networks) method is proposed as a novel and effective approach for spam filtering. SCDNN combines the benefits of the unsupervised STC and DNN architectures and thus can deal with the complex and sophisticated data found in spam. Experimental results on benchmark datasets, such as the ENRON Email Corpus,

the Spam Assassin, and UCI Mail Spam, show that SCDNN achieves better performance than competing state-of-the-art models, including CNN, LSTM, and Random Forest, in terms of accuracy, recall, and precision. This high performance demonstrates the potential of SCDNN's for a prompt and robust process of multi-dimensional data and complex email structures.

The main benefit of SCDNN is its capability to increase spam classification accuracy and reduce spam message overhead, thereby achieving higher user efficiency and better security against threats such as phishing and malware attacks. By integrating spectral clustering with deep neural systems, SCDNN demonstrates its capability to overcome the limitations of strategies heavily reliant on labeled information, making it suitable for addressing complex spam problems.

Due to the satisfactory performance, SCDNN could potentially be implemented in real-world email systems to overcome the increasing spam challenge. A promising future for SCDNN technology is anticipated, with potential applications in real-time email classification. In addition, further investigation is required to evaluate SCDNN in complex scenarios (cloud computing, real-time systems) and optimize it for big data. Considering these advancements, SCDNN technology may play a significant role in enhancing the performance and security of email.

Acknowledgement

N/A

Funding source

No funds received.

Data availability

N/A

DECLARATIONS

Conflict of interest

There are no competing interests.

Consent to publish

N/A

Ethical approval

N/A

REFERENCES

[1] Zhang C. Enhancing Spam Filtering: A Comparative Study of Modern Advanced Machine Learning Techniques. ITM Web of Conferences. 2025;70:04013. 10.1051/itmconf/20257004013

[2] Meenakshi. Spam Content Filtering in Online Social Networks. Wiley; 2025. 10.1002/9781394272464.ch11

[3] Jain V. Intelligent Email Spam Detection: A Machine Learning-Based Approach. In: 2025 5th International Conference on Trends in Material Science and Inventive Materials (ICTMIM). IEEE; 2025. p. 1574–1579. 10.1109/ictmim65579.2025.10988140

[4] Samarthrao KV, Rohokale VM. Enhancement of email spam detection using improved deep learning algorithms for cyber security. Journal of Computer Security. 2021;30(2):231–264. 10.3233/jcs-200111

[5] Gupta S, Barigidad S, Hussain S, Dubey S, Kanaujia S. Hybrid Machine Learning for Feature-Based Spam Detection. In: 2025 2nd International Conference on Computational Intelligence, Communication Technology and Networking (CICTN). IEEE; 2025. p. 801–806. 10.1109/cictn64563.2025.10932459

[6] Yin B, Hu Q, Zhu Y, Zhou K. Semi-supervised learning for shale image segmentation with fast normalized cut loss. Geoenergy Science and Engineering. 2023;229:212039. 10.1016/j.geoen.2023.212039

[7] Han X, Tong X, Fan Y. Eigen Selection in Spectral Clustering: A Theory-Guided Practice. Journal of the American Statistical Association. 2021;118(541):109–121. 10.1080/01621459.2021.1917418

[8] Kufel J, Bargieł-Łączek K, Kocot S, Koźlik M, Bartnikowska W, Janik M, et al. What Is Machine Learning, Artificial Neural Networks and Deep Learning?—Examples of Practical Applications in Medicine. Diagnostics. 2023;13(15):2582. 10.3390/diagnostics13152582

[9] Markevych M, Dawson M. A Review of Enhancing Intrusion Detection Systems for Cybersecurity Using Artificial Intelligence (AI). International conference KNOWLEDGE-BASED ORGANIZATION. 2023;29(3):30–37. 10.2478/kbo-2023-0072

[10] Karim A, Azam S, Shanmugam B, Kannoorpatti K. Efficient Clustering of Emails Into Spam and Ham: The Foundational Study

of a Comprehensive Unsupervised Framework. IEEE Access. 2020;8:154759–154788. 10.1109/access.2020.3017082

[11] Nasreen G, Murad Khan M, Younus M, Zafar B, Kashif Hanif M. Email spam detection by deep learning models using novel feature selection technique and BERT. Egyptian Informatics Journal. 2024;26:100473. 10.1016/j.eij.2024.100473

[12] Borra SR, Yukthika M, Bhargavi M, Samskruthi M, Saisri PV, Akhila Y, et al. OEC Net: Optimal feature selection-based email classification network using unsupervised learning with deep CNN model. e-Prime - Advances in Electrical Engineering, Electronics and Energy. 2024;7:100415. 10.1016/j.prime.2023.100415

[13] Ugwueze WO, Anigbogu SO, Asogwa EC, Asogwa DC, Anigbogu KS. Enhancing Email Security: A Hybrid Machine Learning Approach for Spam and Malware Detection. World Journal of Advanced Engineering Technology and Sciences. 2024;12(1):187–200. 10.30574/wjaets.2024.12.1.0160

[14] Salman M, Ikram M, Kaafar MA. Investigating Evasive Techniques in SMS Spam Filtering: A Comparative Analysis of Machine Learning Models. IEEE Access. 2024;12:24306–24324. 10.1109/access.2024.3364671

[15] Al-shanableh N, Alzyoud MS, Nashnush E. ENHANCING EMAIL SPAM DETECTION THROUGH ENSEMBLE MACHINE LEARNING: A COMPREHENSIVE EVALUATION OF MODEL INTEGRATION AND PERFORMANCE. Communications of the IIMA. 2024;22(1). 10.58729/1941-6687.1451

[16] Adnan M, Imam MO, Javed MF, Murtza I. Improving spam email classification accuracy using ensemble techniques: a stacking approach. International Journal of Information Security. 2023;23(1):505–517. 10.1007/s10207-023-00756-1

[17] Alhuzali A, Alloqmani A, Aljabri M, Alharbi F. In-Depth Analysis of Phishing Email Detection: Evaluating the Performance of Machine Learning and Deep Learning Models Across Multiple Datasets. Applied Sciences. 2025;15(6):3396. 10.3390/app15063396

[18] Sankaine L, Ndia JG, Kaburu D. An English-Swahili Email Spam Detection Model for Improved Accuracy Using Convolutional Neural Networks. Mesopotamian Journal of CyberSecurity. 2025;5(2):590-605

[19] Pathan MS, Dhyani A. Spam Detection in Emails: A Comprehensive Study and Implementation Approach. Journal of Smart Sensors and Computing. 2025;1(1):1–7. 10.64189/ssc.25204

[20] Jazzar M, F Yousef R, Eleyan D. Evaluation of Machine Learning Techniques for Email Spam Classification. International Journal of Education and Management Engineering. 2021;11(4):35–42. 10.5815/ijeme.2021.04.04

[21] Al-augby S, Alyasiri H, Abdulkadhim FG, Oleiwi ZC. A Stacked Ensemble Classifier for Email Spam Detection via an Evolutionary Algorithm. Mesopotamian Journal of CyberSecurity. 2025;5(2):657-70

[22] Shrivastava A, Dubey R. Classification of Spam Mail using different machine learning algorithms. In: 2018 International Conference on Advanced Computation and Telecommunication (ICACAT). IEEE; 2018. p. 1–10. 10.1109/icacat.2018.8933787

[23] Saleh HM, Marouane H, Fakhfakh A. Improves Intrusion Detection Performance InWireless Sensor Networks Through Machine Learning, Enhanced By An Accelerated Deep Learning Model With Advanced Feature Selection. Iraqi Journal for Computer Science and Mathematics. 2024;5(3). 10.52866/ijcsm.2024.05.03.050

[24] Mohammadi K, Islam A, Belhaouari SB. Zooming Into Clarity: Image Denoising Through Innovative Autoencoder Architectures. IEEE Access. 2024;12:98816–98834. 10.1109/access.2024.3424972

[25] Jasim IS, Deniz Duru A, Shaker K, Abed BM, Saleh HM. Evaluation and measuring classifiers of diabetes diseases. In: 2017 International Conference on Engineering and Technology (ICET). IEEE; 2017. p. 1–4. 10.1109/icengtechnol.2017.8308165

How to cite this article

Ahmed SH, Saleh HM, Alani S. Enhanced Email Spam Filtering Using Spectral Clustering and Deep Neural Networks. Journal of University of Anbar for Pure Science. 2026; 20(1):292-300. doi:10.37652/juaps.2025.151919.1405