Mesopotamian journal of Cybersecurity
Vol.5,No.2, **pp**. 590–605
DOI: https://doi.org/10.58496/MJCS/2025/036; ISSN: 2958-6542
https://mesopotamian.press/journals/index.php/cybersecurity

Research Article

An English-Swahili Email Spam Detection Model for Improved Accuracy Using Convolutional Neural Networks

Leshan Sankaine[1,*], John G. Ndia [2], Dennis Kaburu[3]

[1] *School of Computing and Informatics, Mount Kenya University, Thika, Kenya*

[2] *School of Computing and Information Technology, Murang'a University of Technology, Murang'a, Kenya*

[3] *School of Computing and Information Technology, Jomo Kenya University of Agriculture and Technology, Thika, Kenya.*

ARTICLE INFO

Article history

Received 02 Jan 2024
Revised 23 Mar 2025
Accepted 31 Mar 2025
Published 27 Jun 2025

Keywords

ForestPA

Key Forest-based

Classifiers

Medical

Internet of Things

User profile authentication

ABSTRACT

E-mail has become an essential tool for digital communication, facilitating global networking and information exchange. However, spam emails, particularly those in multilingual contexts, pose a significant threat to cybersecurity. In 2023, cyber-related attacks cost Africa approximately USD 10 billion, with the Kenyan economy suffering losses of USD 383 million, 45% of which resulted from phishing and spam emails. While spam detection has been extensively studied for English, low-resource languages such as Swahili lack sufficient research and datasets. Swahili is spoken by about approximately 200 million people, mainly from East Africa. The same speakers use English as a medium of communication. This, therefore, highlights the need to research English-Swahili spam detection. This study recommends a convolutional neural network (CNN)-based model to increase spam detection accuracy in English-Swahili emails. The dataset comprises 8,829 ham emails and 2,749 spam emails, totaling 11,578 messages. The model was trained and evaluated via accuracy, precision, recall, and F1-score metrics. The results indicate a 99.4% accuracy rate, 99.3% precision, 98.2% precision, and 98.7% F1 score. These findings demonstrate good performance and effectiveness.

1. INTRODUCTION

Various As digital communication continues to dominate personal and professional interactions, ensuring the safety of email systems from spam and malicious attacks has become increasingly vital. The emergence of more complex and adaptive spam strategies—particularly in bilingual environments such as English and Swahili—has led to growing interest in developing smarter spam detection mechanisms. These advanced systems aim to strengthen digital communication by precisely identifying and filtering unsolicited or harmful content, thus offering protection to users and preserving the integrity of information systems [1]. In this effort, machine learning has proven to be a powerful ally, providing sophisticated tools capable of analyzing and interpreting complex data patterns. A significant research focus lies in uncovering temporal patterns within email traffic, which helps in recognizing the evolving nature of spam activities. By applying machine learning models to historical email datasets in both English and Swahili, researchers can uncover behavioral trends over time, leading to more accurate and proactive spam detection capabilities [2].

Spam is the term for unsolicited emails that overflow inboxes, waste time and money and possibly expose recipients to scams or harmful content [3]. It may also imply the practice of sending unwanted or promotional emails to a list of recipients via email who have not requested or given permission to receive it. These could be unsolicited or promotional emails. The techniques used by spammers often involve gathering email addresses from the internet and using the username of the domain to send the messages. According to [4], a variety of techniques and technologies, including mail transfers, spoofing, botnets, open proxies, and bulk mailing programs such as mailers, are used to generate spam for commercial purposes.

In recent years, the popularity of sending spam emails has increased. According to [5], some emails contain legitimate advertising content but may also contain malicious URLs, making it impossible to tell which emails are malicious and which are not. Such mails are exploited by hackers and Phishers to extort money and damage the reputation of people. Additionally,

*Corresponding author. Email: esankaine@gmail.com

email spam wastes message speed and storage capacity and involves a variety of challenges, such as network congestion, storage capacity limitations, computing limitations that impair the effectiveness of email searches, time-consuming processes, and increased security vulnerabilities [4].

A study by [6] infers that cyber-related attacks cost the African continent an approximated cost of USD 10 billion during the year 2022, whereas the Kenyan economy lost an estimated total of USD 383 million. Various methods have been used during attacks, such as the exploitation of rogue device vulnerabilities, third-party attacks, ransomware, online fraud schemes, data espage, phishing and spam emails, social engineering, malicious software, and insider threats. Among the methods used, 45% of these attacks resulted from spamming and phishing emails. While many spam detection models with good detection results have been developed in previous research, existing approaches have focused primarily on English-language spam, often failing to effectively detect and mitigate spam in other languages [7].

According to [8], research on natural language processing (NLP) has focused mainly on twenty (20) languages out of the many others spoken, neglecting low-resource languages such as Swahili, a Bantu language that is among the languages recognized as official by the African Union (AU), Southern African Development Community (SADC) and East African Community (EAC), and a global language taught in various universities around the world [9], because they lack important training characteristics such as the quantity of native speakers or expert or supervised data and the evolution of the Swahili context through slang and other influences such as trends.

This limitation poses a significant challenge for users who communicate in multiple languages and for organizations operating in diverse linguistic environments, especially in English and Swahili.

In this study, a convolutional neural network (CNN)-based English–Swahili email spam detection model was developed. The proposed model has a high accuracy rate of 99.4%. This is an improvement over the existing similar model, which has an accuracy of 93.23%. This study uses 6,004 Swahili mails and 5,574 English mails, which is an increase in the dataset from the previous model, which employed 457 Swahili mails and 401 English mails. The increase in the dataset used implies a high level of generalization, improved accuracy, reduced overfitting, and model robustness.

2. RELATED WORKS

Substantial efforts have been dedicated to addressing concerns around the detection of spam emails in English. However, there have been increasing concerns about other languages in recent years. Furthermore, the study of multilingual spam detection is becoming increasingly popular as digital communication becomes increasingly ubiquitous, which poses a threat to cybersecurity and the user experience. Recent studies have focused heavily on the importance of language recognition as a prerequisite for multilingual spam detection. Researchers have also investigated several methods, such as character n-grams and language models, to accurately determine the language of incoming messages so that language-specific analysis can be performed later [10]. This paper summarizes past research studies in English and multilingual dimensions as follows:

The study by[3] explored the detection of multilingual spam SMSs via the naïve Bayes classifier. This research addresses a significant challenge in mobile communication by focusing on spam messages across different languages, an area that has been relatively underexplored in the literature. The authors highlight the effectiveness of the naïve Bayes classifier because of its simplicity and efficiency in text classification tasks. Their approach involves text preprocessing techniques such as tokenization, stemming, and stop-word removal, followed by feature extraction methods such as word frequency and message length analysis. To enhance the system's adaptability to different languages, they integrated a language translation module using the Google Translate API, allowing messages in multiple languages to be processed uniformly before classification. The study achieves an impressive classification accuracy of 96.13%, with high precision and recall scores, demonstrating the effectiveness of the naïve Bayes classifier in detecting spam messages. The implementation of a user-friendly interface via Python Flask further enhanced the practical applicability of the system, making it accessible for real-time spam detection. The dataset used, sourced from Kaggle, consists of 5,574 SMS messages, but the authors do not provide a breakdown of the multilingual distribution within the dataset. This raises concerns about whether the model was sufficiently trained on diverse linguistic variations or if it relied primarily on English spam messages translated into other languages. The reliance on Google Translate for language translation is another limitation, as translation errors may impact classification accuracy. Additionally, the study does not address code-switching and mixed-language messages, which are common in multilingual communication. While the study discusses related works that utilize these models, it does not provide empirical comparisons to justify the superiority of naïve Bayes for this specific task. Furthermore, the paper does not analyse false positives and false negatives, leaving gaps in understanding the model's weaknesses. Real-world testing beyond a static dataset is also lacking, and an evaluation of the computational complexity of the model would provide insights into its scalability. The study focused on SMS messages rather than email messages.

[4] Extensively reviewed advancements in spam detection via machine learning, focusing on algorithms such as naïve Bayes, SVM, random forest, and ensemble methods. However, a critical limitation of such research is its predominant reliance on English-language datasets, such as Ling Spam and SpamAssassin, which do not represent the linguistic diversity of the global digital landscape. This lack of inclusivity highlights why limited research has been conducted on low-resource

languages such as Swahili. The models and methodologies discussed in the paper depend heavily on well-structured datasets and NLP tools, both of which are lacking in Swahili. Furthermore, spam detection techniques that leverage text-based features, such as bag-of-words and tokenization, perform optimally in languages with extensive pre-existing linguistic resources, whereas Swahili's agglutinative nature and morphological richness pose additional challenges. The absence of Swahili-specific datasets and tailored machine-learning techniques means that models trained on English-centric datasets are unlikely to be generalizable well to Swahili. Consequently, while the paper provides valuable insights into improving spam detection via advanced machine learning techniques, its applicability to Swahili and other low-resource languages remains limited, underscoring the urgent need for targeted research in this area.

[5] undertook research that focused on spam detection in WhatsApp conversations via naïve Bayes and support vector machine (SVM) machine learning techniques. The machine learning techniques used in the paper, including text preprocessing methods such as tokenization, stemming, and vectorization, rely on language-specific NLP tools that are well developed for English but underdeveloped for Swahili. This means that even if the same algorithms were applied to Swahili text, their performance would be hindered by inadequate linguistic resources. Moreover, the study assumes that common spam detection techniques, such as filtering on the basis of frequently used words and patterns, work effectively across languages. While the study demonstrates high accuracy in detecting spam in English-based chats, it implicitly highlights a major limitation in research related to low-resource languages such as Swahili. The morphological complexity of swahili presents unique challenges that require specialized models.

[6] carried out a thorough review of approaches aimed at examining domain- and header-related data in email headers. Using a group of unsupervised feature selection methods, the study presented a novel feature reduction method. The main data source was a newly created dataset with 100,000 records of spam and ham emails. The study's main conclusions were that, out of the six clustering algorithms that were tested, spectral and k-means performed satisfactorily, whereas OPTICS was the most effective, outperforming spectral and k-means by approximately 3.5%, as confirmed by thorough procedures. The performance of the other algorithms—BIRCH, HDBSCAN, and K-modes—was lower. For the three best algorithms, the average balanced accuracy was approximately 94.91%. While the study contributes valuable insights into the application of clustering methods for spam classification, several limitations emerge, particularly when considering its applicability to low-resource languages such as Swahili. One key limitation is the dataset used, which comprises 100,000 records collected from various public email corpora. Although this dataset is large and diverse, it remains heavily biased toward English-language emails. The lack of representation for low-resource languages raises concerns about the generalizability of the proposed framework. Languages such as Swahili, which exhibit morphological richness and frequent code switching, present unique challenges that clustering algorithms trained on English-centric datasets may struggle to handle effectively. Additionally, the study relies on header and domain information rather than email content, which could further limit its applicability in languages with different spam patterns and structures. Furthermore, the clustering algorithms tested, including K-means, OPTICS, and spectral clustering, rely on features that may not be as meaningful for languages with different syntactic and semantic characteristics.

[7] presented a novel approach to spam detection by leveraging multilingual BERT (M-BERT) for detecting spam across different languages and modalities (text-based and image-based). While the study contributes valuable insights, several limitations emerge, particularly when considering its applicability to low-resource languages such as Swahili. One significant limitation is the dataset used, which consists primarily of English and Chinese spam data. Although this study aims to address bilingual spam detection, it does not extend its scope to low-resource languages such as Swahili, which lack large, well-annotated spam datasets. The reliance on M-BERT, a pretrained transformer model, also presents challenges, as M-BERT’s performance is extremely dependent on the availability of quality training data for a given language. For Swahili, which has limited representation in large-scale corpora, M-BERT may not generalize well, leading to lower accuracy in spam classification. Furthermore, while the study integrates optical character recognition (OCR) for extracting text from image-based spam, it assumes that spam images contain embedded text. This approach may not be effective for Swahili spam images, which may feature different stylistic elements, nontext-based spam indicators, or mixed-language patterns that are harder to capture. Additionally, the study does not address linguistic complexities such as Swahili's rich morphology and frequent code switching with English, which could pose challenges for tokenization and feature extraction.

Traditional spam detection technologies, such as honeynet or pot-based techniques and filters based on URL lists, have limitations, according to [8], the results of . This is especially true because spammers frequently alter the appearance and features of their contents. These techniques are frequently laborious and less successful in dynamic situations. The use of machine learning (ML) and deep learning (DL) approaches has successfully solved such problems. However, because they frequently concentrate on particular categories of language, content, and account information, datasets produced utilizing APIs and web crawlers from online social networks (OSNs) may introduce bias, which might restrict their generalizability.

[11] conducted a comparative study on email spam detection via large language models (LLMs), natural language processing (NLP) models, and convolutional neural networks (CNNs). This research aimed to assess the effectiveness of

GPT-4, BERT, RoBERTa, and CNN in filtering spam emails. The study used two English-language spam email datasets from Kaggle, with 5,728 and 5,572 samples, respectively. The results revealed that BERT achieved the highest accuracy (99.39%), followed closely by GPT-4 (99.3%) and Roberta (99.04%), whereas the CNN performed the worst (76.09%). The study emphasized the importance of fine-tuning pretrained models for improved accuracy but did not explore multilingual spam detection or code-switching scenarios and was thus not effective for multilingual scenarios such as English–Swahili.

[12] provides a comprehensive review of computational models in bilingual language learning, exploring how statistical learning, connectionist modelling, and neural networks have contributed to understanding bilingual cognition. The authors discuss how various computational approaches, such as self-organizing maps (SOMBIP) and bilingual recurrent neural networks (BSRN), simulate bilingual lexicon development and second-language acquisition (L2 learning). However, most models still struggle to capture code switching, language interference, and the cognitive control mechanisms involved in bilingual processing. The review emphasized that while deep learning techniques such as transformer-based models (e.g., BERT) are promising, they still cannot fully integrate neurobiological and cognitive theories into bilingual language learning models such as English–Swahili.

[13] investigated the effectiveness of deep neural networks (DNNs) in email spam classification, comparing models such as RNN, LSTM, GRU, bidirectional RNN, bidirectional LSTM, and bidirectional GRU. The study uses two datasets: the 20 Newsgroup dataset (20,000 documents for multiclass classification) and the ENRON dataset (5,000 emails for binary spam classification). The results indicate that the CNN achieves the highest accuracy (98.5%) on the ENRON dataset, followed closely by LSTM (98.2%) and GRU (97.8%). The findings confirm that deep learning techniques significantly outperform traditional machine learning methods such as naïve Bayes and SVM in handling spam classification. However, the study is limited to English-language emails and does not explore multilingual or code-switched spam detection.

[14] proposed a universal spam detection model (USDM) using transfer learning on the BERT model, aiming to create a single spam detection model that generalizes across multiple datasets. The study utilizes four publicly available datasets—Ling-Spam (2,893 emails), Spam Text Messages (5,574 messages), Enron (32,638 emails), and SpamAssassin (6,047 emails)—to fine-tune BERT for spam classification. Their approach involves training individual models on each dataset, extracting hyperparameters, and then combining them into a single universal model. The results show that the final model achieved 97% accuracy with an F1 score of 0.96, significantly outperforming traditional spam filters. This paper highlights the advantages of pretrained transformers in NLP tasks, demonstrating that BERT-based models achieve higher precision and recall than previous deep learning architectures do. However, the study has notable limitations. It focuses exclusively on English-language datasets, ignoring multilingual spam emails or code-switching scenarios, which are crucial for real-world applications, particularly in regions where mixed-language communication is common. Additionally, BERT-based models require substantial computational resources, making them less practical for low-resource environments or real-time filtering.

[15] introduced a novel approach that combines a weighted support vector machine (WSVM) with Harris hawks optimization (HHO) for spam review detection. HHO was employed for optimizing hyperparameters and feature weighting, addressing the challenge of multilingual spam reviews. The study utilized datasets in English, Spanish, and Arabic, incorporating pretrained word embeddings (BERT) alongside three-word representation methods: N-Gram-3, TF-IDF, and one-hot encoding. Four experiments were conducted, each addressing specific aspects of the problem. The proposed WSVM-HHO approach demonstrated superior performance compared with state-of-the-art algorithms, achieving accuracies of 0.88%, 0.72%, 0.90%, and 0.84% for English, Spanish, Arabic, and multilingual datasets, respectively. One of the main limitations of this study is the dataset composition. The study focuses on three relatively high-resource languages with well-developed linguistic tools and datasets, whereas languages such as Swahili remain underrepresented. This highlights the broader challenge in spam detection research—models trained on well-resourced languages may not generalize well to low-resource languages owing to differences in linguistic structure, morphology, and code-switching tendencies. Moreover, while the paper leverages pretrained embeddings such as BERT, it does not address the fact that Swahili and other low-resource languages often lack robust pretrained language models, leading to poorer spam classification accuracy. The study also assumes that the optimized WSVM-HHO model is adaptable across languages, but the effectiveness of such optimizations in a morphologically rich and underrepresented language such as Swahili remains uncertain.

[16] used text from the IWSPA-AP 2018 antiphishing joint work to create an English-Arabic parallel phishing email corpus. A balanced dataset of 1,258 emails in Arabic and English, with equal percentages of authentic and phishing emails, was utilized to assess the suggested EAPD model. According to the experimental results, the EAPD model uses a multilayer perceptron (MLP) classifier with TF-IDF to obtain 95.3% accuracy on Arabic datasets. Using a support vector machine (SVM) classifier with TF-IDF yielded an accuracy of 95.7% for English text, demonstrating the model's strong performance in multilingual phishing email detection. However, the study has notable limitations, particularly concerning its generalizability to other low-resource languages such as Swahili. The dataset used, consisting of 1,258 balanced English

and Arabic phishing and legitimate emails, is relatively small compared with the large-scale phishing corpora available for English. This limited dataset size restricts the model's ability to generalize effectively, especially in multilingual contexts. For languages such as Swahili, where labelled phishing datasets are even scarcer, this limitation becomes more pronounced. [17] conducted a study on the detection of SMS Spam in Swahili Text via deep learning approaches on a Swahili Swahili dataset and achieved an accuracy of 99% via a CNN-LSTM-LSTM hybrid model. This research was limited to SMSs and not email platforms. The dataset used is also highly imbalanced, with only 297 unique smishing messages out of over 10 million spam messages.

On the other hand, [18] conducted research on multilingual spam detection via the random forest algorithm, which targets the Tamil, English, Hindi, and Malayalam multilingual datasets and consists of 7,137 manually created and sourced from Kaggle, which consists of 1,433 spam and 5,704 ham messages, making it highly imbalanced. Since spam messages constitute only approximately 20% of the dataset, the classifier may develop a bias toward predicting spam messages. Additionally, the dataset is limited to only four languages, which, while diverse, exclude many other widely spoken and low-resource languages, such as Swahili, which also suffer from a lack of spam detection models.

[19] presents a long short-term memory (LSTM)-based model for spam detection in English and Indonesian, specifically targeting spam messages submitted through web forms on government ministry websites. The study highlights the growing challenge of web form spam, emphasizing the security risks it poses, such as phishing, malware distribution, and database overload. This research contributes valuable insights by developing a spam detection system tailored to a multilingual context and implementing data augmentation techniques to mitigate class imbalance issues. One notable limitation is the dataset composition and class imbalance. The primary dataset, the RIDA Web Form Spam Dataset, contains 4,915 messages, with a disproportionate number of spam messages (3,687) compared with no spam messages (1,228). Although the study employs data augmentation to balance the classes, artificially generated data may not always capture the nuances of real-world spam messages, particularly in diverse linguistic contexts. This issue is more pronounced for low-resource languages such as Swahili, where the scarcity of authentic labelled datasets limits the effectiveness of augmentation techniques. Additionally, the study is focused on English and Indonesian languages, restricting its applicability to English--Swahili contexts.

[20] Undertook multilingual rules for spam detection for the Chinese, English, and Vietnamese languages via SpamAssassin statistical rules. On the basis of their study of multilingual rules for spam detection in the Chinese, Vietnamese, and English languages, statistical rules may achieve 61% spam/harm classification, while the failure rate increased alarmingly to 4.9%. Additionally, [10] investigated the use of the Bayesian classifier for email spam filtering, leveraging its statistical text classification capabilities. The naïve Bayes method analyses tokens or words in spam and ham emails to calculate probabilities and classify emails as spam or legitimate. The study introduced an integrated approach that enhanced classification accuracy compared with traditional methods, achieving an improvement from 96.46% to 97.3% on a real-world dataset. This advancement highlights the effectiveness of the integrated approach in helping internet users manage and reduce spam emails. One of the main limitations of this study is its reliance on rule-based methods, which, while effective for structured spam patterns, may struggle to adapt to rapidly evolving spam techniques. The rules are manually generated and refined on the basis of predefined patterns, making them less effective against dynamic spam messages that use adversarial strategies such as misspellings, special characters, or randomized word insertions. The experiment relies on a relatively small dataset, with 705 spam messages and 653 ham messages for multilingual evaluation. This dataset may not fully capture the diversity of spam messages encountered in real-world applications, especially considering that spam varies significantly across different platforms and languages. Additionally, the study does not include any low-resource languages such as Swahili, where rule-based methods might require substantial manual effort owing to the lack of established linguistic resources.

[21] developed a multilingual SMS spam dataset and suggested a hybrid deep learning method to address the dearth of research on spam detection in non-English languages. English and four Indian languages—Tamil, Malayalam, Kannada, and Telugu—were represented in the dataset along with spam and ham communications. To classify spam, the study used a CNN-LSTM hybrid model rather than conventional feature engineering. The architecture included a word embedding layer that was fed into a Convolution1D layer, a MaxPooling layer for dimensionality reduction, and a dense layer with sigmoid activation for final classification. The efficiency of the proposed model in classifying spam messages from the multilingual SMS dataset was demonstrated via a comparison with baseline deep learning algorithms. One of the primary limitations is the dataset size and distribution. The dataset comprises 2,757 ham messages and only 525 spam messages, making it highly imbalanced (84% ham vs. 16% spam). This imbalance can lead to biased classification, where the model favours nonspam messages. Additionally, the dataset is relatively small for training deep neural networks, particularly for languages with fewer data points, such as Malayalam (190 messages) and Telugu (188 messages). The small dataset size may limit the model's generalizability, especially when applied to unseen real-world spam messages. Furthermore, the study does not extend to global low-resource languages such as Swahili. Since the structure and morphology of languages vary significantly, a model trained on one set of languages may not generalize well to others. This raises concerns about

whether the hybrid CNN-LSTM approach would work effectively for Swahili and other underrepresented languages. Additionally, the paper does not analyse the challenges of code-switching, which is common in multilingual spam messages, particularly in contexts where users frequently mix English with their native language.

Finally, one study by [22] assessed the success of machine learning algorithms in a Swahili-English email filtering system in contrast to the Gmail classifier using naïve Bayes, sequential minimal optimization (SMO), and J48 via a manually created English–Swahili dataset taken from the scholar's inbox. The findings from the study show that SMO outperforms the other algorithms, with an accuracy of 93.23%, followed by J48 (87.22%) and naïve Bayes 88.47%. This research was limited by the availability of the Swahili dataset. The dataset used in the research is composed of 457 Swahili mails and 401 English mails. This limited dataset may not fully capture the variability of real-world spam messages, particularly given the linguistic diversity and informal nature of Swahili-English code switching. A larger, more diverse dataset would provide a more robust evaluation of the classifiers. Furthermore, the study's comparison with Gmail's spam filter is limited, as it relies on manual classification rather than direct integration with Gmail's proprietary filtering system. This makes it difficult to draw definitive conclusions about the relative performance of the machine learning models compared with Gmail's filter, which likely incorporates more sophisticated deep learning techniques and continuously learns from vast amounts of real-world email data.

The various spam detection models explored in past studies are summarized in Table 1 below.

TABLE I. RELATED RESEARCH ON AUTOMATED SPAM DETECTION MODELS

Reference	Research Area	Language	Dataset/Domain	Algorithm Used	Outcome
[3]	A comparative study on email spam detection using Large Language Models (LLMs), Natural Language Processing (NLP) models, and Convolutional Neural Networks (CNNs).	English	datasets from Kaggle, with 5,728 and 5,572 samples, respectively	GPT-4, BERT, RoBERTa, and CNN	The results disclosed that BERT attained the highest accuracy (99.39%), followed closely by GPT-4 (99.3%) and Roberta (99.04%), while CNN performed the worst (76.09%)
[4]	A review of how advanced machine learning can improve spam detection.	English	Ling Spam dataset of SpamAssassin	Naïve Bayes; Boosting and AdaBoost; Random Forest Classifier; K-Nearest Neighbors (KNN); Decision Tree;	The Random Forest Algorithm proved to be more effective among the other four algorithms.
[5]	Chat Analysis and Spam Detection of WhatsApp messages using Machine Learning.	English	Whatsapp Chats.	Naïve Bayes, SVM, and Maximum Entropy.	The algorithms obtained accuracies of 0.95%, 0.97%, and 0.91%; SVM proved more effective.
[6]	Effective Email Grouping into Spam and Ham: The Basis Research of an All-Inclusive Unsupervised System.	English.	A manual dataset was produced using freely accessible raw email corpora.	Multialgorithm clustering approach.	Test accuracy with a 60-40 split was 97. 44% and 94. 57%, respectively.
[7]	Efficacy of noncontextualized word embeddings (Word2Vec, GloVe) against contextualized word embeddings (BERT, ELMo) for social media spam detection	English, Chinese	Twitter—which came from Kaggle—and YouTube—which included remarks from the UCI DL Repository	RNN	When compared to Word2Vec and GloVe, BERT and ELMo consistently produced much higher accuracy, precision, and recall.
[10]	Integrated Spam Detection for Multilingual Emails.	Multilingual	Mail datasets were collected from Gmail and Yahoo.	Bayesian Classifier.	Between 95%-97% accuracy was recorded on different datasets.
[12]	A comprehensive review of computational models in bilingual language learning	Bilingual lexicon development and second-language acquisition (L2 learning)	Bilingual	BERT	The review emphasizes that while deep learning techniques are promising but cannot fully integrate neurobiological and

Reference	Research Area	Language	Dataset/Domain	Algorithm Used	Outcome
					cognitive theories into bilingual language learning models.
[13]	Effectiveness of deep neural networks (DNNs) in email spam classification	English	Two datasets: the 20 Newsgroup dataset (20,000 documents for multiclass classification) and the ENRON dataset (5,000 emails for binary spam classification	RNN, LSTM, GRU, Bidirectional RNN, Bidirectional LSTM, and Bidirectional GRU	The results indicate that CNN achieves the highest accuracy (98.5%) on the ENRON dataset, followed closely by LSTM (98.2%) and GRU (97.8%).
[14]	Universal Spam Detection Model (USDM) using transfer learning on the BERT model The study utilizes)—to fine-tune BERT for spam classification.	English	four publicly available datasets—Ling-Spam (2,893 emails), Spam Text Messages (5,574 messages), Enron (32,638 emails), and SpamAssassin (6,047 emails)	BERT	The results show that the final model achieved 97% accuracy with an F1-score of 0.96, significantly outperforming traditional spam filters.
[15]	A multilingual spam review detection based on pretrained word embedding and weighted swarm support vector machines	English, Spanish, Arabic	English, Spanish, Arabic, and Multilingual datasets	Weighted Support Vector Machine (WSVM) with Harris Hawks Optimization (HHO)	The proposed WSVM-HHO approach demonstrated superior performance compared to state-of-the-art algorithms, achieving accuracies of 0.88%, 0.72%, 0.90%, and 0.84% for English, Spanish, Arabic, and multilingual datasets, respectively
[16]	Enhancing cybersecurity using machine learning and natural language Processing for Arabic phishing email detection	English and Arabic	1258 emails from the IWSPA-AP 2018 dataset	• ML and DL Classifiers.	MLP classifier combined with TF-IDF, the EAPD achieved an accuracy of 95.3 percent on Arabic datasets. The English text, on the other hand, achieved a 95.7 percent accuracy when paired with the SVM classifier and TF-IDF.
[17]	Detection of SMS Spam in Swahili Text by Use of Deep Learning Approaches	Swahili	Swahili dataset.	CNN-LSTM-LSTM hybrid model	On the Swahili dataset, the CNN-LSTM-LSTM hybrid model achieved the utmost accuracy of approximately 99%, while CNN-BiLSTM outperformed with an accuracy of 98.38 on the UCI dataset.
[18]	Random Forest-Based Multilingual Spam Detection.	Tamil, English, Hindi and Malayalam	The multilingual dataset is made up of 7137 messages from Kaggle and 7137 messages that were created by the user.	Random forest.	The model achieved an accuracy of approximately 90%.
[19]	Using a Long Short-Term Memory Algorithm on Ministry Websites to Identify Spam	English and Indonesian	RIDA Web Form Spam Dataset, SpamAssassin Email Dataset, Bahasa	Long Short-Term Memory (LSTM)	The study achieved accuracy rates of 82.4% for the RIDA Web Form Spam

Reference	Research Area	Language	Dataset/Domain	Algorithm Used	Outcome
			Indonesia dataset, and UCI SMS Spam Collection Dataset.		Dataset, 85.3% on the SpamAssassin Email Dataset, and 96.1% on the Bahasa Indonesia dataset.
[20]	Multilingual Rules for Spam Detection	Chinese, Vietnamese, and English.	SpamAssassin	Statistical rules.	Not more than 61% of detections were made, and up to 4.9% of alarms were not successful.
[21]	hybrid deep learning method for multilingual spam SMS detection.	English, Tamil, Malayalam, Kannada, and Telugu.	There are 2,274 messages in English, 346 in Tamil, 284 in Kannada, 190 in Malayalam, and 188 in Telugu.	The CNN-LSTM model combines two deep learning models: the Convolutional Neural Network (CNN) and the Long Short-Term Memory (LSTM).	CNN-LSTM model had better performance than other models.
[22]	Assessment of a Machine Learning Algorithm's Performance in a Swahili-English Email Filtering System in Comparison to Gmail Classifier.	English, Swahili.	Manually created English-Swahili dataset.	Naïve Bayes, Sequential Minimal Optimization (SMO) and J48.	The results depicted that SMO gives good results equated to other algorithms with an accuracy of 0.93% followed by Naïve Bayes 0.88% and J48 0.87%, respectively.

On the basis of the literature summarized in Table 1 above, several studies have been conducted on spam emails, with results from the findings based on the characteristics of fraudulent mail. Research on other languages has yielded several techniques and algorithms. However, from the foregoing research, it is realized that research on English–Swahili contexts is limited. The applicability of the existing algorithms cannot be measured owing to the limited Swahili datasets and the evolution of the Swahili context through slang and other influences, such as trends. Furthermore, the efficiency of existing models is characterized by limited accuracy, unreliability, and underperformance. According to [23], the detection of email spam by existing spam detection methods, such as naïve Bayes and SVM, has several limitations, such as the following:

i. Limited Multilingual Capabilities—Most spam detection models are trained on monolingual datasets, making them ineffective in detecting English–Swahili spam, code switching, and transliterations.
ii. Contextual Understanding Deficiencies – traditional contemporary techniques primarily capture local word features but struggle with long-range dependencies and semantic meaning, leading to false positives in emails with misleading spam terms.
iii. High computational resources limit their feasibility in real-time applications.
iv. Dataset Limitations—Many existing models lack diverse bilingual datasets, making them less effective in practical multilingual communication environments.

Our study aims to address the identified gaps in an English–Swahili multilingual spam detection setup by developing an optimized CNN-based model for English–Swahili spam detection, incorporating advanced preprocessing techniques to handle code-switching and bilingual text structures.

3. THEORETICAL FRAMEWORK

The paper borrows from general deterrence theory (GDT), which is used primarily in the context of criminology and legal studies. This theory argues that the fear of punishment prevents people from committing crimes. In other words, penalties deter (discourage) people from showing lawless behavior [24]. During the study, the GDT was applied to design a model founded on a supervised machine learning classifier to learn and identify characteristics that differentiate email spam from legitimate emails on the basis of historical data of malicious activities. The theory uses four dimensions, namely, deterrence, prevention, detection, and remedy, to realize information security (IS), as shown in Figure 1 below.

Deterrence emphasizes the importance of physical security for computer devices and strict adherence to information security policies, which play crucial roles in mitigating the risks associated with spam email attacks.

Prevention involves the use of convolutional neural networks (CNNs) to filter out spam messages before they reach users' inboxes. By training CNNs on datasets containing both English and Swahili spam messages, the model can learn to recognize patterns and features indicative of spam, improving filtering accuracy.

Detection relies on machine learning (ML) techniques to classify emails as either spam (nonham) or legitimate (ham), enabling an automated and efficient identification process.
Remedy focuses on the actions taken once spam has been detected, including blocking the sender, reporting the spam, or deleting the message. The proposed model not only detects spam but also prevents future occurrences by blocking originating addresses and reporting offenders to law enforcement agencies, thereby enhancing overall information security.

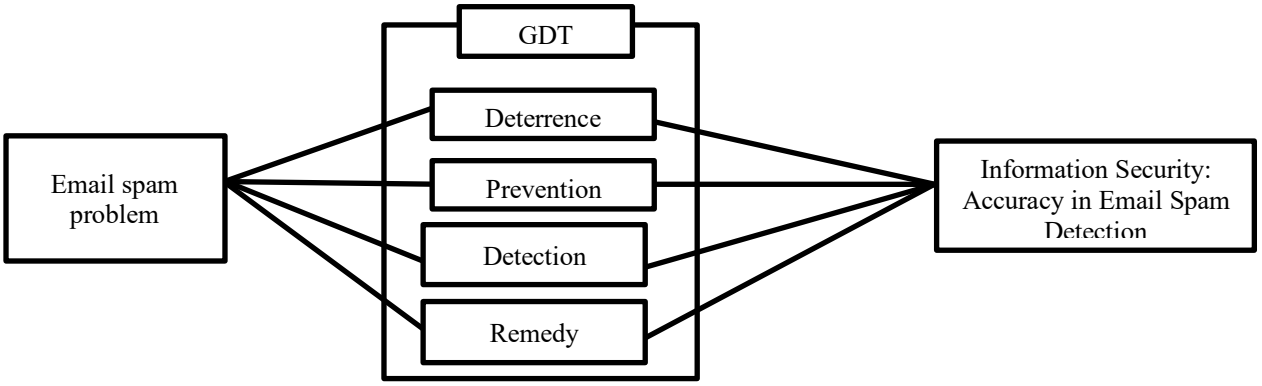

Fig. 1. Theoretical framework for the proposed Model

Our experimental setup involved a systematic process, as depicted in Figure 2 below. This begins with data collection, where relevant emails are gathered for training and evaluation. Next, the preprocessing stage involves cleaning and normalizing the text by removing unnecessary symbols, tokenizing words, and extracting key features. Once the data are prepared, the model training phase takes place, where a convolutional neural network (CNN) is trained via labelled email datasets. After training, the model proceeds to text classification, where it categorizes emails as spam or legitimate (ham). The next step involves performance evaluation, where metrics such as accuracy, precision, recall, and F1 score are used to assess the model's effectiveness. On the basis of this evaluation, the model is then used to make predictions on new email data. If the predictions are incorrect, the model is refined and retrained.

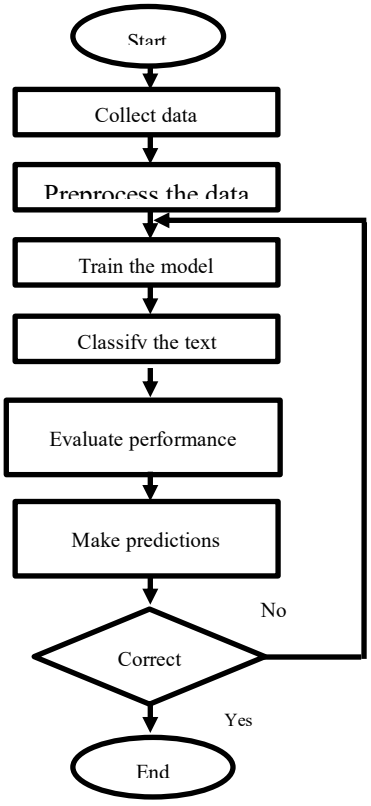

Fig. 2. Model development life cycle

3.1. Dataset creation

We used a bilingual dataset comprising 11,578 email samples. The dataset consists of 8,829 ham emails and 2,749 spam emails, ensuring a balanced representation of legitimate and spam content in both languages, as depicted in Table 2 below:

TABLE II. DISTRIBUTION OF HAM AND SPAM MESSAGES

Row Labels	Number of Ham Messages	Number of Spam Messages	Grand Total
English	4,826	748	5,574
Swahili	4,003	2,001	6,004
Grand Total	**8,829**	**2749**	**11,578**

3.2. Data Preprocessing

Before training the English–Swahili CNN model, we applied a series of preprocessing steps to the dataset:

i. Text Tokenization: In this stage, the email text messages were converted into numerical sequences via Keras' Tokenizer, with a vocabulary size limited to 5000 words. This approach enhances computational efficiency, reduces noise by filtering out rare words that contribute minimally to classification, and ensures that the model focuses on frequently occurring, meaningful terms. Consequently, this helps prevent overfitting to uncommon words and enhances generalization.

ii. Padding sequences: In this stage, all the input text data were standardized to a fixed length of 100 words by padding techniques.

iii. Data Splitting: In this stage, we employed 5-fold cross-validation to systematically evaluate the model's performance across different training and testing subsets.

iv. Tokenizer Preservation: Last, the tokenizer was saved as a .pkl file to facilitate future predictions.

3.3. CNN Architecture

The proposed CNN model was developed on the basis of the following layers:

i The embedding layer (128-dimensional vectors) converts words into dense vector representations for better contextual understanding.

ii The 1D convolutional layer (64 filters, kernel size = 5, ReLU activation), which captures local patterns and important features in the text sequences.

iii The global max pooling layer reduces dimensionality while preserving the most relevant text features.

iv Fully connected layers have three sublayers, namely, the dense layer (10 neurons, ReLU activation), which learns deeper representations of the text; the dropout layer (50%), which prevents overfitting by randomly deactivating neurons during training; and the output layer (1 neuron, sigmoid activation), which outputs a probability score that determines whether an email is spam or not spam.

3.4. Model compilation

Once the CNN architecture was defined, the model was compiled to establish optimization, loss measurement, and evaluation metrics via adaptive moment estimation (Adam) and a loss function (binary cross-entropy). The Adam optimizer was selected because of its ability to dynamically adjust learning rates for individual parameters, making it well suited for handling text data with varying patterns, whereas the loss function–binary cross-entropy was used since our task involves binary classification (spam vs. non spam). The binary cross-entropy loss function (log loss) was used to penalize incorrect predictions and encourage the model to produce probability scores close to 0 or 1 [23]. This function is particularly effective when combined with the sigmoid activation function in the output layer, as indicated by the formula below:

$$\text{Loss} = -\text{N1i} = 1\sum\text{N}[\text{yilog}(\text{y\^i}) + (1 - \text{yi})\text{log}(1 - \text{y\^i})] \quad (1)$$

where

where yi is the actual label (1 for spam, 0 for non-spam).

where y^i is the predicted probability of being spam.

N is the total number of samples in the dataset.

The proposed model was evaluated via four key performance metrics, namely, accuracy, precision, recall, and F1 score, each of which provides a distinct perspective on the model's effectiveness. By dividing the total number of emails processed by the proportion of correctly classified emails (spam and non-spam), accuracy quantifies how accurate the model's predictions are overall. It is calculated with the following formula:

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \text{ [25]} \quad (2)$$

Precision measures the proportion of emails that were correctly classified as spam, providing information about the accuracy of the model's positive classifications. It is determined by the following formula:

$$Precision = TP/(TP + FP)[26] \quad (3)$$

Recall evaluates the model's ability to correctly identify all spam emails, measuring the proportion of actual spam messages that were correctly classified. It is calculated as:

$$Recall = TP/(TP + FN)[27] \quad (4)$$

Finally, the F1 score provides a balanced measure of the model's performance by considering both precision and recall. It is the harmonic mean of these two metrics and is given by the following formula:

$$F1\ score = 2 \times ((Precision \times Recall)/(Precision + Recall))\ [28] \quad (5)$$

For all these metrics:

False positive (FP) = Number of Hams incorrectly classified by the model;
False negative (FN) = The quantity of spam that the model misclassified;
True positive (TP) = The quantity of spam that the model correctly classified;
True negative (TN) = The number of Hams that the model correctly classified.

3.5. Experimental Setup

3.5.1. Cross-Validation Approach

To ensure robust model evaluation, we implemented 5-fold cross-validation, dividing the dataset into five subsets. Each subset was used as a test set once, while the remaining four served as training data. The process was repeated five times, and the overall performance was averaged.

3.5.2. Model training

The model training process was structured to optimize learning efficiency while preventing overfitting. A batch size of 32 was selected, meaning that the dataset was divided into smaller subsets containing 32 samples each. This choice ensures a balance between computational efficiency and training stability. The model underwent training for 5 epochs, meaning that it completed five full iterations over the dataset. This number was chosen to allow the model to learn meaningful patterns from the data while avoiding excessive training that could lead to overfitting. Additionally, early stopping was implemented as a safeguard to improve generalizability. The training was automatically halted if the validation loss failed to improve for two consecutive epochs, thereby conserving computational resources and ensuring that the model did not memorize training data but rather learned patterns applicable to unseen data.

3.6. Data Analysis and Presentation

We employed Explanatory Data Analysis (EDA), a technique that evaluates data and identifies trends and patterns or confirms hypotheses via statistical summaries and graphical representations. According to [29], there are three types of EDAs:

i. Univariate EDA involves examining one variable at a time to find the simplest patterns within it.
ii. Bivariate EDA is a method for analysing data that uses graphical representations and statistical summaries to find trends or patterns or to verify assumptions.
iii. Multivariate EDA entails examining three or more variables at a time to understand the relationships between numerous variables and recognize any complex patterns or outliers that might exist.

The study employed univariate EDA to calculate and visualize the distribution of spam vs. nonspam emails via a pie chart. Univariate EDA was also used to compute and plot the average word length for spam and nonspam emails. Bivariate EDA was used to explore the relationship between email classification (spam vs. nonspm) and average word length, which involves two variables. This analysis uses bar charts to compare the average word length across categories.

3.7. Model Deployment & Testing

Upon completion of training, the final model was saved and tested via real email messages. Each input text was tokenized, padded, and passed through the model for classification. A threshold of **0.3** was applied to determine whether an email was classified as spam or nonspam.

4. FINDINGS AND DISCUSSION

The model was developed through trials and evaluated via metrics. The outcomes are presented in proportion to the study's objective. This was compared against results from previous studies based on the literature presented in section 2 above. Related research has demonstrated that traditional machine learning models achieve accuracies ranging from 92% to 98% depending on the size of the dataset used. In contrast, this study achieved an average accuracy of 99%, as indicated in Tables 3 to 7 below:

In Table 3 below, the model starts with a training accuracy of 82.74% and a training loss of 0.4015. The validation accuracy is notably high at 99.27%, with a validation loss of 0.0249, suggesting good initial generalizability.

TABLE III. TRAINING PROGRESS OVERVIEW 1/5

Epoch	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1.	82.74%	0.4015	99.27%	0.0249
2.	97.88%	0.0479	99.48%	0.018
3.	98.42%	0.031	99.40%	0.021
4.	98.73%	0.0277	99.44%	0.0224
5.	82.74%	0.4015	99.27%	0.0249

In Table 4 below, the model achieved 83.74% training accuracy in the first epoch, which increased to 98.99% by the third epoch, accompanied by a decrease in training loss from 0.4055 to 0.0336. The validation accuracy remained high, at approximately 99%, although the validation loss fluctuated slightly, ending at 0.0352 in the third epoch.

TABLE IV. TRAINING PROGRESS OVERVIEW 2/5

Epoch	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1.	83.74%	0.4055	99.09%	0.0284
2.	98.50%	0.0536	98.92%	0.0292
3.	98.99%	0.0336	99.01%	0.0352

In Table 5 below, the model begins with 81.92% training accuracy in the first epoch and improves to 99.35% by the third epoch, with training loss decreasing from 0.397 to 0.0505. The validation accuracy also remained robust, fluctuating slightly but close to 99.31%, with the validation loss increasing slightly to 0.0368.

TABLE V. TRAINING PROGRESS OVERVIEW 3/5

Epoch	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1.	81.92%	0.397	99.35%	0.0317
2.	98.65%	0.0744	99.27%	0.0343
3.	99.35%	0.0505	99.31%	0.0368

In Table 6, the training accuracy started at 84.49% in the first epoch and increased steadily to 99.16% by the fifth epoch, whereas the training loss decreased significantly from 0.3815 to 0.0249. The validation accuracy also improved, reaching 99.52%, with the validation loss remaining low but fluctuating slightly, ending at 0.0309.

TABLE VI. TRAINING PROGRESS OVERVIEW 4/5

Epoch	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1.	84.49%	0.3815	99.35%	0.0275
2.	98.28%	0.056	99.40%	0.0222
3.	99.10%	0.034	99.52%	0.0218
4.	99.13%	0.0272	99.48%	0.0269
5.	99.16%	0.0249	99.52%	0.0309

Finally, in Table 7, the model begins with 85.34% training accuracy, which increases to 99.26% by the fifth epoch, whereas the training loss decreases from 0.3936 to 0.0246. The validation accuracy peaked at 99.65% during the fourth epoch, with the validation loss decreasing steadily to a final value of 0.0212.

TABLE VII. TRAINING PROGRESS OVERVIEW 5/5

Epoch	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1.	85.34%	0.3936	99.52%	0.0221
2.	98.39%	0.0542	99.40%	0.0194
3.	98.83%	0.0354	99.52%	0.0176
4.	99.03%	0.0358	99.65%	0.0179
5.	99.26%	0.0246	99.57%	0.0212

As depicted in Tables 3 to 7 above, the training accuracy increased from 85% in the first epoch to 99.26% by the fifth epoch, indicating that the model learns effectively over time. Similarly, the training loss decreases from 0.39 in the first epoch to 0.02 by the fifth epoch, indicating that the model minimizes the error in its predictions as it trains. This means that the model is learning effectively.

The model's accuracy started at 98.36% and remained consistently high, reaching 99.91% in the third epoch and stagnating for the 4th and 5th epochs. Hence, this model can generalize well-undetected data.
The accuracy, precision, recall, and F1 score were used to assess the model's performance, as indicated in Tables 8 and 9 and Figure 3:

TABLE VIII. PERFORMANCE METRICS OF THE EMAIL SPAM DETECTION MODEL

Metric	Value	Remarks
Accuracy	0.9940	This shows that 99.40% of the predictions were correct.
Precision	0.99.30	This indicates that 99.30% of the instances predicted as positive were correct.
Recall	0.9815	This shows that 98.15% of the actual positive instances were correctly identified.
F1 Score	0.9872	This is a balance between precision and recall, with a value of 0.99 inferring a strong balance between these metrics.

TABLE IX. CONFUSION MATRIX FOR THE EMAIL SPAM DETECTION MODEL

	Predicted ham	Predicted spam
Ham	1,762 (True Negative)	4 (False Positive)
Spam	7(False Negative)	542 (True Positive)

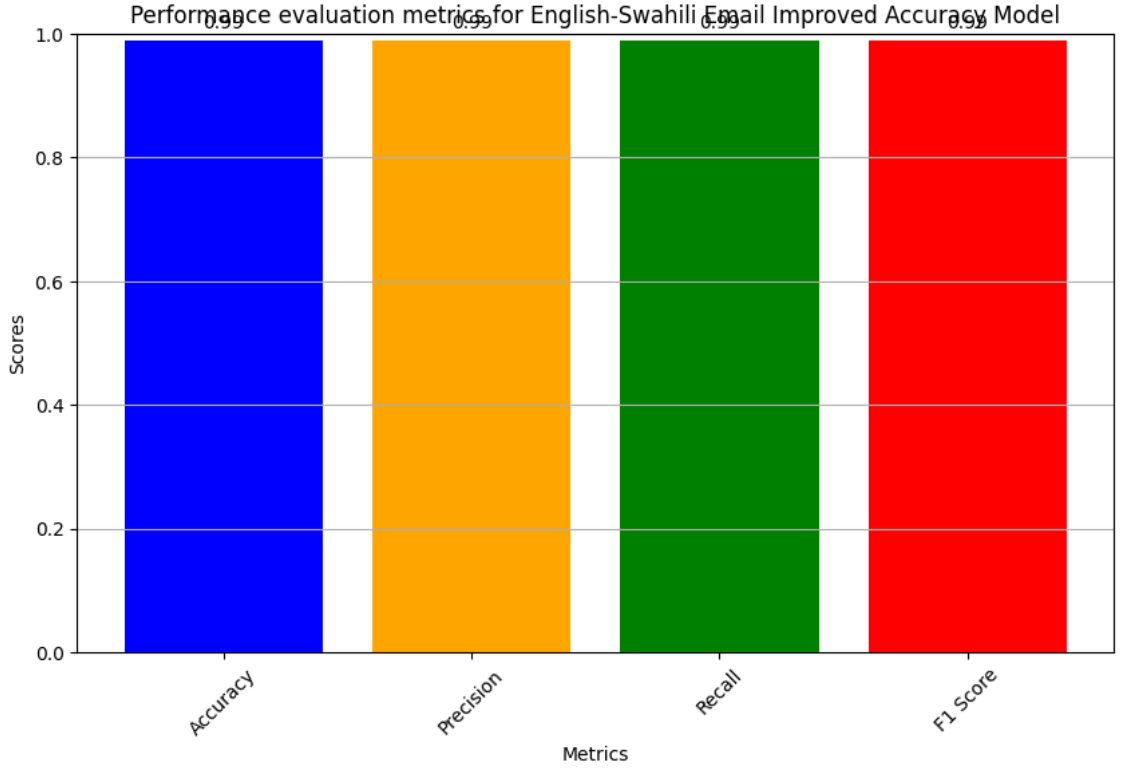

Fig. 3. Model performance evaluation metrics

The above results indicate that the CNN is highly effective in classifying bilingual spam emails**.** It achieved an accuracy of 99.4%, high precision (99.3%), recall (98.15%), and F1 score (98.72%)**.** The low false positive and false negative rates further affirm the model's reliability. Comparisons with traditional machine learning models such as naïve Bayes and SVM confirm the CNN's superior performance in detecting spam in a bilingual setting. In general, these metrics indicate that the model is highly accurate, with outstanding precision and recall and very few misclassifications. The confusion matrix indicates a need for model refinement with an emphasis on reducing false negatives.
The model initially employed a classification threshold of 0.3, which resulted in a greater number of emails being labelled spam. While this approach minimizes false negatives, it also increases false positives. To refine the classification process, a receiver operating characteristic (ROC) analysis was conducted, and an ROC curve was plotted to visually assess the model's ability to distinguish between spam and nonspam emails. The ROC curve illustrates the relationship between the true positive rate (Recall) and the false positive rate (1 - specificity) at various threshold values. The model's overall performance was assessed via the area under the curve (AUC), where a higher AUC signifies stronger spasm detection capabilities. The optimal threshold (0.99) was determined via [30], which optimally balances spam detection while

minimizing false positives. Although this threshold enhances precision, it may slightly reduce recall, depending on the distribution of spam emails in real-world scenarios.

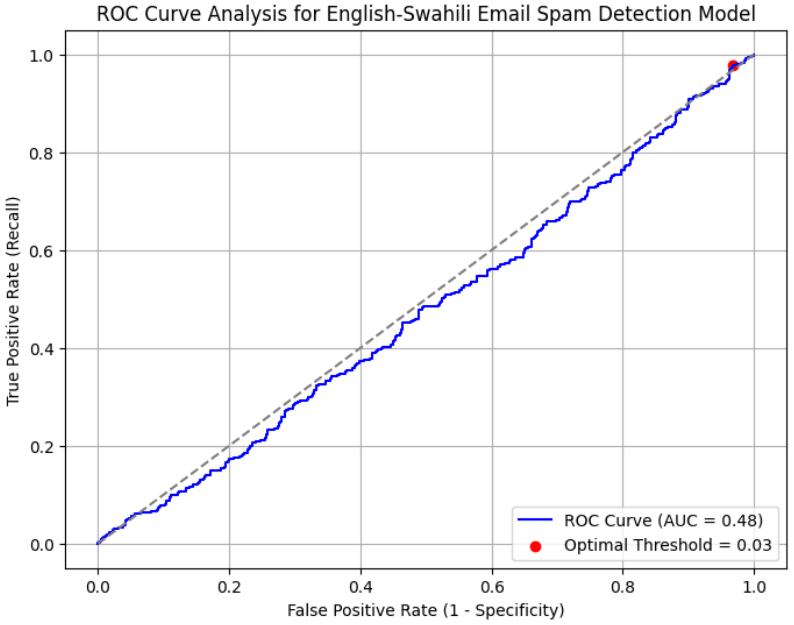

Fig. 4. ROC analysis for the proposed model

One distinguishing characteristic of spam emails is their average word length, which tends to be greater than that of nonspam emails. While legitimate emails are generally concise and well structured, spam messages often contain long, misleading phrases designed to capture the recipient's attention or evade detection. This difference in language structure and complexity plays a crucial role in spam classification, as spam emails frequently use unnecessary verbosity, keyword stuffing, and deceptive wording to bypass filtering mechanisms. As depicted in Figure 5 below, the proposed model effectively utilizes word length variations as a key feature in distinguishing between spam and nonspam emails. Longer words and inflated sentence structures are often indicators of fraudulent or promotional content, which makes word-length analysis a useful component of the classification process.

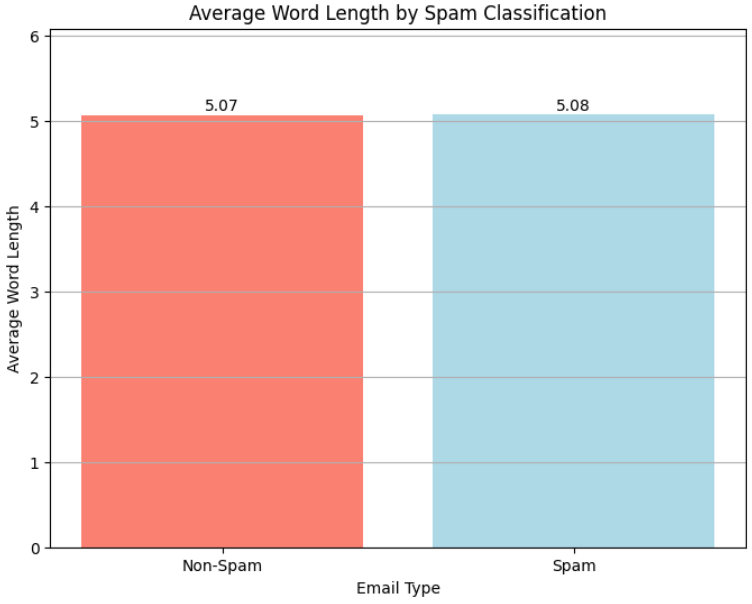

Fig. 5. Average word length by spam

The average word length for spam messages in our dataset was approximately five words. It follows that spam messages in our dataset had fewer words.

5. CONCLUSION AND FUTURE WORK

This study presents a CNN-based English–Swahili spam detection model. The study demonstrated the model's effectiveness in classifying English–Swahili bilingual emails with a 99.4% accuracy rate, 99.3% precision, 98.2% precision, and 98.7% F1 score. The findings demonstrate that the proposed good performance and effectiveness. The proposed model significantly improves email security in multilingual contexts, addressing gaps in existing monolingual spam filters. Generally, the results show that the model is a promising tool for enhancing email communication security. Thus, the research findings underscore the importance of ongoing improvements and adaptations to maintain high detection accuracy in dynamic email spam landscapes.

The study confirmed that an accurate, effective, and efficient spam detection model can be developed for email communication in diverse dialects. The use of supervised deep learning techniques was successful in identifying spam across English and Swahili. Additionally, the results highlighted the significance of continuous model modification since spammers use tactics that keep evolving. This notwithstanding, the development of an improved spam detection model for English-Swahili emails represents a significant step toward combating email spam.

To further increase spam detection accuracy, future work should explore hybrid deep learning architectures that combine CNNs with bidirectional long short-term memory (BiLSTM) or transformer-based models to enhance feature extraction. Additionally, expanding the dataset to include real-world spam emails from diverse domains enhances model robustness. These improvements ensure greater adaptability, reduced misclassification rates, and better spam filtering in English-Swahili email communication.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Funding

No funding was received for this work.

Acknowledgement

The authors would like to thank the paper reviewers for their contribution to the publication of this paper.

References

[1] J. Sharma, Sonia, K. Kumar, P. Jain, R. H. C. Alfilh, and H. Alkattan, "Enhancing Intrusion Detection Systems with Adaptive Neuro-Fuzzy Inference Systems," *Mesopotamian J. CyberSecurity*, vol. 5, no. 1, pp. 1–10, 2025, doi: 10.58496/MJCS/2025/001.

[2] A. M. Salman, B. T. Al-nuaimi, A. A. Subhi, H. Alkattan, and R. H. C. Alfilh, "Enhancing Cybersecurity with Machine Learning : A Hybrid Approach for Anomaly Detection and Threat Prediction," vol. 5, no. 1, pp. 202–215, 2025.

[3] K. Aparna and S. Halder, "Detection of Multilingual Spam SMS Using NaïveBayes Classifier," *5th IEEE Int. Conf. Cybern. Cogn. Mach. Learn. Appl. ICCCMLA 2023*, pp. 89–94, 2023, doi: 10.1109/ICCCMLA58983.2023.10346960.

[4] I. Panda and S. Dash, "A Review on Enhancing Spam detection With Advance Machine learning," vol. 12, no. 1, pp. 17–22, 2024.

[5] D. Teja, S. K. Kumar, and D. M. Chandra, "Chat Analysis and Spam Detection of Whatsapp using Machine Learning Chat Analysis and Spam Detection of Whatsapp using Machine Learning," no. November, 2023.

[6] A. Karim, S. Azam, B. Shanmugam, and K. Kannoorpatti, "Efficient Clustering of Emails into Spam and Ham: The Foundational Study of a Comprehensive Unsupervised Framework," *IEEE Access*, vol. 8, pp. 154759–154788, 2020, doi: 10.1109/ACCESS.2020.3017082.

[7] J. Cao and C. Lai, "A bilingual multi-type spam detection model based on M-BERT," *Proc. - IEEE Glob. Commun. Conf. GLOBECOM*, vol. 2020-Janua, 2020, doi: 10.1109/GLOBECOM42002.2020.9347970.

[8] S. Rao, A. K. Verma, and T. Bhatia, "A review on social spam detection: Challenges, open issues, and future directions," *Expert Syst. Appl.*, vol. 186, no. March, p. 115742, 2021, doi: 10.1016/j.eswa.2021.115742.

[9] Z. Zhang, Z. Deng, W. Zhang, and L. Bu, "MMTD: A Multilingual and Multimodal Spam Detection Model Combining Text and Document Images," *Appl. Sci.*, vol. 13, no. 21, p. 11783, 2023, doi: 10.3390/app132111783.

[10] A. Iyengar, G. Kalpana, S. Kalyankumar, and S. Gunanandhini, "Integrated SPAM detection for multilingual emails," *2017 Int. Conf. Inf. Commun. Embed. Syst. ICICES 2017*, no. Icices, pp. 2–5, 2017, doi: 10.1109/ICICES.2017.8070784.

[11] K. I. Roumeliotis, N. D. Tselikas, and D. K. Nasiopoulos, "Next-Generation Spam Filtering: Comparative Fine-Tuning of LLMs, NLPs, and CNN Models for Email Spam Classification," *Electron.*, vol. 13, no. 11, pp. 1–24, 2024, doi: 10.3390/electronics13112034.

[12] P. Li and Q. Xu, *Computational Modeling of Bilingual Language Learning: Current Models and Future Directions*, vol. 73, no. December 2023. 2023. doi: 10.1111/lang.12529.

[13] V. R. Chirra, H. D. Maddiboyina, Y. Dasari, and R. Aluru, "Review of Computer Engineering Studies Performance Evaluation of Email Spam Text Classification Using Deep Neural Networks," vol. 7, no. 4, pp. 91–95, 2020.

[14] V. S. Tida and S. H. Hsu, "Universal Spam Detection using Transfer Learning of BERT Model," *Proc. 55th Hawaii Int. Conf. Syst. Sci.*, 2022, doi: 10.24251/hicss.2022.921.

[15] A. M. Al-Zoubi, A. M. Mora, and H. Faris, "A Multilingual Spam Reviews Detection Based on Pre-Trained Word Embedding and Weighted Swarm Support Vector Machines," *IEEE Access*, vol. 11, no. June, pp. 72250–72271, 2023, doi: 10.1109/ACCESS.2023.3293641.

[16] S. Salloum, *Enhancing Cybersecurity: Machine Learning and Natural Language Processing for Arabic Phishing Email Detection*. 2024. [Online]. Available: https://salford-repository.worktribe.com/output/2363839

[17] I. S. Mambina, J. D. Ndibwile, D. Uwimpuhwe, and K. F. Michael, "Uncovering SMS Spam in Swahili Text Using Deep Learning Approaches," *IEEE Access*, vol. 12, no. January, pp. 25164–25175, 2024, doi: 10.1109/ACCESS.2024.3365193.

[18] R. Priyangka, "Multilingual Spam Detection Using Random Forest," vol. 12, no. 04, pp. 336–338, 2023.

[19] R. F. Busyra and A. S. Girsang, "Applying Long Short-Term Memory Algorithm for Spam Detection on Ministry Websites," *J. Syst. Manag. Sci.*, vol. 14, no. 2, pp. 1–20, 2024, doi: 10.33168/JSMS.2024.0201.

[20] M. Tuan Vu, Q. A. Tran, F. Jiang, and V. Q. Tran, "Multilingual Rules for Spam Detection," *J. Mach. to Mach. Commun.*, vol. 1, no. 2, pp. 107–122, 2015, doi: 10.13052/jmmc2246-137x.122.

[21] E. Ramanujam, K. Shankar, and A. Sharma, "Multi-lingual Spam SMS detection using a hybrid deep learning technique," *Proc. - 2022 IEEE Silchar Subsect. Conf. SILCON 2022*, pp. 1–6, 2022, doi: 10.1109/SILCON55242.2022.10028936.

[22] R. A. Omar and A. Tjahyanto, "Evaluation of the performance of a machine learning algorithms in Swahili-English emails filtering system relative to Gmail classifier," *2018 Int. Conf. Inf. Commun. Technol. ICOIACT 2018*, vol. 2018-Janua, pp. 266–269, 2018, doi: 10.1109/ICOIACT.2018.8350713.

[23] E. H. Tusher, M. A. Ismail, M. A. Rahman, A. H. Alenezi, and M. Uddin, "Email Spam: A Comprehensive Review of Optimize Detection Methods, Challenges, and Open Research Problems," *IEEE Access*, vol. 12, no. October, pp. 143627–143657, 2024, doi: 10.1109/ACCESS.2024.3467996.

[24] L. Forrester and C. Ruiz, "Classical Theories of Criminology: Deterrence," *Introd. to Criminol. Crim. Justice*, 2024.

[25] L. H. Son, A. Kumar, S. R. Sangwan, A. Arora, A. Nayyar, and M. Abdel-Basset, "Sarcasm detection using soft attention-based bidirectional long short-term memory model with convolution network," *IEEE Access*, vol. 7, pp. 23319–23328, 2019, doi: 10.1109/ACCESS.2019.2899260.

[26] S. Biere and M. B. Analytics, "Hate Speech Detection Using Natural Language Processing Techniques," *Vrije Univ. Amsterdam*, p. 30, 2018.

[27] P. Teja Nallamothu and M. Shais Khan, "Machine Learning for SPAM Detection," *Asian J. Adv. Res.*, vol. 6, no. 1, pp. 167–179, 2023.

[28] R. Y. Choi, A. S. Coyner, J. Kalpathy-Cramer, M. F. Chiang, and J. Peter Campbell, "Introduction to machine learning, neural networks, and deep learning," *Transl. Vis. Sci. Technol.*, vol. 9, no. 2, pp. 1–12, 2020, doi: 10.1167/tvst.9.2.14.

[29] L. Anselin, *An Introduction to Spatial Data Science with GeoDa: Volume 1: Exploring Spatial Data*. CRC Press, 2024.

[30] W. J. Youden, "Statistical Techniques," *NBS Spec. Publ.*, no. 300–301, p. 421, 1969.