

Arabic SMS Spam filter Using Machin Learning

Shahd Dairi

Table of contents:

- Problem statement
- Research objectives
- Literary reviews
- Critical Analysis
- Connections
- Conflicts Between the Findings
- Synthesis Matrix for Literature Review
- References

Problem Statement:

The rapid growth of mobile communication has led to an increasing reliance on Short (SMS) as a primary means of information exchange. However, this growth has been accompanied by a significant rise in unsolicited and harmful messages, commonly referred to as spam. Spam poses significant threats to user privacy and information security (Das et al.,2026). These messages often include advertisements, phishing attempts, and fraudulent content, posing risks to users and consuming device resources such as storage and processing capacity. The Significance of a spam filtering solution using machine learning algorithms incapacity. users time and effort in filtering spam from non-spam messages. Spam detection in text-based messages essentially involves the use of machine learning for the identification of unwanted or malicious messages in texts such as emails, SMS (Ahmad et al., 2025). should focus on real-life data for training experiments and models, rather than manually generated datasets (Ruchi et al.,2024). as they can be trained on categorized data to predict new message classifications and spam characteristics such as keywords, punctuation, and requests for sensitive data. The Arabic language, with its rich morphology and complex structure, presents unique challenges for text classification tasks such as spam detection (Al Saidat et al., 2024). However, many existing solutions and models are in English and ineffective when dealing with Arabic SMS messages.

The research problem focuses on SMS spam messages in Arabic. Given some limitations on SMS filtering models, the aim of this search is a clear need to develop a robust, efficient model, SMS filtering by machine learning models.

Research objectives:

The goal is to develop an effective machine learning-based filtering model capable of accurately distinguishing spam and non-spam incoming messages based on their content and characteristics enhancing user protection from spam and potentially harmful content.

- 1- To develop a machine learning-based model capable of classifying SMS messages into spam and ham categories
- 2-To train the proposed model using a suitable SMS dataset to ensure effective learning of message patterns
- 3-To apply appropriate data preprocessing techniques, including text cleaning, tokenization, and feature extraction, to improve model performance.
- 4-To evaluate the performance of the model using measurable metrics such as accuracy, precision, recall, and F1-score.
- 5-To compare the performance of different machine learning algorithms to identify the most effective approach for SMS spam detection.

6- To implement the final model within a simple and user-friendly interface for practical usage.

Literary reviews:

1. An Intelligent Framework Based on Deep Learning for SMS and e-mail Spam Detection

In this paper explores the application of multiple machine learning classifiers and a deep learning classifier to detect spam in SMS and email datasets with improved accuracy. Experiments were conducted on real datasets to evaluate the performance of the proposed system. The researcher found that the support vector machine (SVM) classifier outperformed other models, indicating its effectiveness for classification tasks. The study's findings suggest that the proposed system offers enhanced accuracy compared to existing models for detecting spam in SMS and email messages. The SVM classifier emerged as the top-performing model, showcasing its suitability for classification purposes in this context. (Maqsood, U. 2023)

2. A Novel Approach for Arabic SMS Spam Detection Using Hybrid Deep Learning Techniques

This study aimed to detect potentially harmful and spam messages in Arabic SMS communications using deep learning techniques. The researchers proposed a hybrid deep learning model that combines Convolutional Neural Networks (CNN) with Bidirectional Long Short-Term Memory (Bi-LSTM) networks to improve spam detection performance. The proposed model was applied to a dataset consisting of Arabic SMS messages that included both spam and legitimate (ham) messages. Prior to training, the dataset underwent several preprocessing steps, such as text cleaning, tokenization, and padding, in order to prepare the textual data for the deep learning model. The hybrid model was trained using the Adam optimizer and evaluated using several performance metrics, including accuracy, precision, recall, and F1-score. Additionally, an early stopping technique was implemented during training to prevent overfitting. The experimental results showed that the proposed model achieved high classification performance, with an accuracy of 0.9699, precision of 0.9739, recall of 0.9675, and an F1-score of 0.9707. These results demonstrate the effectiveness of the hybrid CNN–BiLSTM model in accurately detecting spam messages in Arabic SMS datasets. The study concluded that the proposed approach provides a robust and reliable solution for Arabic SMS spam classification. (Al Saidat et al., 2024)

3. SMS Scam Detection Application Based on Optical Character Recognition for Image Data Using Unsupervised and Deep Semi-Supervised Learning

This research paper explores a method for developing a model designed to detect smishing messages via SMS by understanding the relationships between contextual factors, textual content, and images, which are often overlooked in previous studies. The research utilized Optical Character Recognition (OCR) to analyze a dataset containing both legitimate and spam SMS messages. The study evaluated two main approaches. The first approach employed deep learning models, including Recurrent Neural Networks (RNN-Flatten), Long Short-Term Memory (LSTM), and Bidirectional LSTM (Bi-LSTM). The second approach involved traditional machine learning techniques such as K-Nearest Neighbors (KNN) and K-means, along with feature extraction methods including Term Frequency–Inverse Document Frequency (TF-IDF) and Principal Component Analysis (PCA). The experimental results showed that the KNN-Flatten model achieved the best performance, reaching an accuracy of 94.13%, indicating its potential effectiveness for real-time smishing detection systems. However, the study also noted that the model may face scalability challenges due to its computational complexity, which could limit its performance when applied to large-scale datasets. (Shinde, 2024)

4. SMART FILTERS FOR SMS SPAM: A MACHINE LEARNING APPROACH TO SMS CLASSIFICATION

This study focuses on developing an accurate system for detecting and classifying spam messages using machine learning techniques. The researchers evaluated several classification algorithms and incorporated multiple features to enhance the model's performance. The methodology involved applying five machine learning classifiers—Logistic Regression (LR), Naïve Bayes Classifier (NBC), Support Vector Machine (SVM), Random Forest (RF), and Gradient Boosting Machine (GBM). The models were trained and tested using the SMS Spam Collection dataset obtained from Kaggle. In addition, the study utilized eleven different features, including links, character count, date information, email addresses, emojis, and phone numbers, to improve the effectiveness of the classification process. The performance of the models was evaluated using several metrics such as accuracy, sensitivity, precision, recall, F1-score, and specificity. The results indicated that the Light Gradient Boosting Machine (LGBM) model achieved the best performance among the tested algorithms. It obtained a sensitivity, precision, recall, and F1-score of 100%, along with a specificity score of 100%, resulting in an overall accuracy of 100%. The study concluded that the proposed approach significantly outperformed many recent methods in the field of SMS spam detection. (Nawaz et al ., 2025)

5. Machine Learning-Based Optimization Algorithms for Spam SMS Classification

This study proposed a novel framework for SMS spam detection that integrates machine learning techniques with optimization algorithms. The researchers combined the Golden Search Optimizer (GSO) and the Whale Optimization Algorithm (WOA) to develop a hybrid optimization approach known as GSOWOA. This optimization

method was integrated with the Random Forest (RF) classifier in order to enhance classification performance and reduce convergence time.

The proposed model focused on improving feature selection and optimizing model parameters to achieve more accurate spam message classification. Experimental results the model achieved an overall accuracy of 98.9% and a precision of 97.85%, while also improving convergence speed by approximately 35% compared with conventional methods.

Furthermore, the GSOWOA approach showed strong resistance to overfitting and maintained high computational efficiency. These results indicate that the proposed framework is effective for real-time SMS spam detection, particularly in environments with limited computational resources such as edge devices. (Das et al.,2026)

6. SMS Spam Detection with NLP and Ensemble Learning for Enhanced

Classification

This study proposed a scalable SMS spam detection framework that integrates Natural Language Processing (NLP) techniques with an ensemble approach combining both machine learning and deep learning models. The system applied preprocessing steps such as tokenization, stemming, and lemmatization, while feature extraction was performed using TF-IDF and CountVectorizer. Several classifiers, including Random Forest, Support Vector Machine, Extra Trees, Bi-LSTM, GRU, and a transformer-based model, were combined using a soft-voting ensemble. The results showed that the proposed framework achieved a high accuracy of 98.92% and an F1-score of 0.987. In addition, Explainable AI techniques such as SHAP and LIME were integrated to enhance model interpretability. The system was deployed using a Flask API to enable near real-time SMS classification, making it suitable for practical applications in sectors such as banking and e-commerce. (Yadav et al., 2025)

Critical Analysis:

One of the main strengths of these studies lies in their diverse methodologies. Furthermore, some studies presented a new and innovative approach: hybrid methods, which demonstrated clear progress over traditional classification methods.

However, some notable limitations can be identified. Most studies relied on English, with limited Arabic coverage, except for one study. This indicates a gap, possibly due to the complexity of the Arabic language.

Methodologically, the studies lacked detailed explanations of the data processing methods and steps, as well as the division of datasets, which could affect the reproducibility of results.

Regarding data sources, all studies included textual data. While textual data is easy to use and allows for comparisons, it lacked multimedia, with only one study focusing on text messages and images.

connections:

1-Detecting And Classifying: All studies focus on one problem which is detecting and classifying SMS spam messages using Artificial Intelligence.

2-Machine Learning and Text Analysis techniques were used to facilitate the task of classifying and analyzing words using natural language processing (NLP) techniques such as Tokenization, Feature extraction and TF-IDF.

3-Machine Learning Algorithms: Many studies used common machines learning algorithms including Support Vector Machine (SVM) algorithms, including ^[1]_[SEP] Random Forest (RF) and Naïve Bayes.

4-Performance Metrics: The studies evaluated their models using similar performance metrics, to determine the success of the study and to facilitate comparison between the Accuracy, Precision, Recall ^[1]_[SEP] models used such as comparison between the models and F1-Score.

5- Improve Accuracy: All studies aim to improve the accuracy of spam detection compared to traditional approaches.

Conflicts Between the Findings:

1. Types of Technologies: The studies varied considerably in the technologies used and can be divided into three categories:

Studies that relied primarily on traditional machine learning techniques. Other studies that used deep learning models. While some studies adopted hybrid approaches that combined multiple technologies using crowdsourcing or integrating optimization algorithms with machine learning models

2- Data Types: Most studies primarily analyzed only SMS text messages. However, one study analyzed both images and text messages simultaneously using Optical Character Recognition (OCR) technology.

3- Methods for Improving Model Performance: Some studies have focused on feature engineering, using 11 different features. Other studies have addressed optimization techniques to enhance model performance. Still other studies have improved performance through group learning.

4- Model Accuracy: Accuracy levels vary between studies, ranging from 100% to 94.13%. This variation stems from several factors, including the models and techniques used

Synthesis matrix for literature review:

Study	Author (s)	Year	Key Findings	Methodology	Key Findings	Weaknesses / Limitations
An Intelligent Framework Based on Deep Learning for SMS and e-mail Spam Detection	Umair Maqsood	2023	The SVM classifier achieved the best performance among the tested models, demonstrating high effectiveness in spam classification.	Applied several machine learning classifiers and a deep learning model to detect spam in SMS and email datasets.	The SVM classifier achieved the best performance among the tested models, demonstrating high effectiveness in spam classification.	Limited details about dataset diversity and generalization across different languages or datasets.
A Novel Approach for Arabic SMS Spam Detection Using Hybrid Deep Learning Techniques	Mohammed Rasol Al Saidat, Suleiman Y. Yerima, Khaled Shaalan	2024	The hybrid model achieved 96.99% accuracy, showing strong capability in detecting Arabic spam messages.	Proposed a hybrid deep learning model (CNN + Bi-LSTM) for Arabic SMS spam detection. Preprocessing included text cleaning, tokenization, and padding.	The hybrid model achieved 96.99% accuracy, showing strong capability in detecting Arabic spam messages.	translated data quality and the high computational overhead necessitated by the hybrid model's complexity.
SMS Scam Detection Application Based on Optical Character Recognition for Image Data Using Unsupervised and Deep Semi-Supervised Learning	Anjili Shinde	2024	The KNN-Flatten model achieved the best performance with 94.13% accuracy, suitable for real-time smishing detection.	Used OCR technology to analyze SMS messages containing text and images. Combined deep learning models (RNN, LSTM, Bi-LSTM) with traditional ML and feature extraction (TF-IDF, PCA).	The KNN-Flatten model achieved the best performance with 94.13% accuracy, suitable for real-time smishing detection.	The complexities of the system's structure, maintenance requirements, limited resources, and a lack of focus on minute details such as symbols and numbers in text messages.
Smart filters for SMS spam: A machine learning approach to SMS classification	Ishrat Nawaz et al.	2025	The LGBM model achieved 100% accuracy and performance metrics, outperforming other classifiers.	Evaluated five ML classifiers (LR, NBC, SVM, RF, GBM) using the SMS Spam Collection dataset and 11 engineered features.	The LGBM model achieved 100% accuracy and performance metrics, outperforming other classifiers.	Lack of available text data
Machine learning-based optimization algorithms for spam	Lipsa Das, Laxmi Ahuja, Adesh Pandey	2026	Achieved 98.9% accuracy and improved convergence speed by	Developed a hybrid optimization model combining Golden Search Optimizer (GSO)	Achieved 98.9% accuracy and improved convergence speed by 35%, demonstrating	It focused primarily on text messages only, without considering multimedia, and the performance of the

SMS classification			35%, demonstrating strong resistance to overfitting.	and Whale Optimization Algorithm (WOA) with Random Forest.	strong resistance to overfitting.	GSOWOA algorithm was not studied extensively.
SMS spam detection with NLP and ensemble learning for enhanced classification	Avinash Yadav, Rahul Yadav, Kiruthika M	2025	Achieved 98.92% accuracy and 0.987 F1-score, with interpretable predictions and near real-time deployment.	Proposed an ensemble spam detection system integrating NLP preprocessing, ML classifiers, deep learning models, and explainable AI (SHAP, LIME).	Achieved 98.92% accuracy and 0.987 F1-score, with interpretable predictions and near real-time deployment.	System complexity and integration of multiple models may increase implementation difficulty.

References

1. Ahmad, M, Rozlan, M., Yusri, M. (2025). Spam Detection Using Machine Learning Techniques. IEEE.
<https://www.techrxiv.org/doi/full/10.36227/techrxiv.173933255.51566942/v1>
2. Ruch, A, Anshita, D. Surya, K. (2024). A novel approach for spam detection using natural language processing with AMALS models. IEEE Access.
3. Maqsood, U. (2023). An intelligent framework based on deep learning for SMS and e-mail spam detection. Wireless Communications and Mobile Computing. Wiley online library. <https://doi.org/10.1155/2023/6648970>
4. Al Saidat, M. R., Yerima, S. Y., & Shaalan, K. (2024). A novel approach for Arabic SMS spam detection using hybrid deep learning techniques. Procedia Computer Science.
<https://www.sciencedirect.com/science/article/pii/S187705092403000X>
5. Shinde, A. (2024). SMS scam detection application based on optical character recognition for image data using unsupervised and deep semi-supervised learning. Sensors. <https://www.mdpi.com/1424-8220/24/18/6084>
6. Nawaz, I., Khosa, S., Fatima, R., Saeed, M., & Hashmi, M. S. A. (2025). Smart filters for SMS spam: A machine learning approach to SMS classification. Spectrum of engineering sciences.
<https://thesesjournal.com/index.php/1/article/view/333>
7. Yadav, A., Yadav, R., & Kiruthika, M. (2025). SMS spam detection with NLP and ensemble learning for enhanced classification. IEEE.
<https://ieeexplore.ieee.org/document/11159039>
8. Das, L., Ahuja, L., & Pandey, A. (2026). Machine learning-based optimization algorithms for spam SMS classification. Engineering, Technology & Applied Science Research.
<https://www.etasr.com/index.php/ETASR/article/view/14670>