# Through A Comparative Analysis of SMS Spam or Ham Detection Using Machine Learning and Deep Learning

[1]Ranganath K, [2]Anuradha, [3]Ashwini, [4]Darshani, [5]Drishty Shetkar

[1,2,3,4,5]Computer Science and Engineering Department, Guru Nanak Dev Engineering College Bidar, Karnataka, India

Corresponding Author E Mail id:biradaranuradha34@gmail.com

## Abstract

**With the rapid expansion of mobile communication services, Short Message Service (SMS) has become one of the most widely used communication channels. However, the increasing volume of SMS traffic has also led to a significant rise in unsolicited spam messages, which can cause security threats such as phishing attacks, financial fraud, and information leakage. Effective spam detection systems are therefore essential to ensure secure digital communication.**

**This study presents a comparative analysis of traditional machine learning techniques and transformer-based deep learning models for SMS spam classification. Support Vector Machine (SVM) and Extreme Gradient Boosting (XGBoost) are implemented as conventional classifiers, while DistilBERT and RoBERTa are employed as advanced contextual language models. The models are trained and evaluated on a labeled SMS dataset using standard performance metrics including accuracy, precision, recall, and F1-score.**

**Experimental findings indicate substantial performance variation across models. SVM achieved an accuracy of 56.6%, while XGBoost significantly improved performance to 91.4%. Transformer-based models demonstrated superior results, with DistilBERT achieving 99.8% accuracy and RoBERTa achieving 99% accuracy. The results confirm that deep contextual language representations provide enhanced semantic understanding for spam detection tasks. This study highlights the effectiveness of transformer architectures in text classification and provides insight into the trade-offs between computational complexity and predictive performance.**

***Keywords***-SMS Spam Detection, Text Classification, Support Vector Machine, XGBoost, DistilBERT, RoBERTa, Machine Learning, Deep Learning, Natural Language Processing.

## 1. Introduction

Now a days, data-driven technologies have become an integral part of modern life. From personalized recommendations on streaming platforms to automated fraud detection in banking,Autonomous systems keeps on reshape industries and daily experiences. Among the various branches of AI, ML and DL support out as the fewest transformative fields. Both approaches focus on enabling systems to extract knowledge from data and make decisions or predictions without being explicitly programmed, yet they differ in their learning mechanisms, data requirements, and computational complexities.

Machine Learning (ML) algorithms, such as SVM and XGBoost, rely on mathematical and statistical methods to detect patterns within structured datasets. These techniques often require human intervention for tasks like feature selection and data preprocessing to achieve optimal performance. Due to their simplicity, interpretability, and efficiency, traditional ML models perform well on structured and relatively small datasets.

In contrast, Deep Learning (DL) is an advanced subset of ML that uses multi-layered neural networks to automatically extract and learn features directly from raw data. This capability enables DL models to capture complex, non-linear relationships and identify

important attributes from inputs such as text, images, or audio. Transformer-based models like DistilBERT and RoBERTa have significantly enhanced Natural Language Processing (NLP) by effectively understanding both linguistic context and semantic meaning, outperforming earlier approaches.

Choosing between ML and DL depends on factors such as dataset size, task complexity, available computational resources, and the need for interpretability. ML models are generally faster, easier to understand, and require less data, while DL models offer higher accuracy and greater adaptability when handling large, unstructured datasets.

Artificial Intelligence has become a driving force behind innovation, and its subfieldsML along with DL are at the center of technological transformation. ML along with DL are widely practical in fields for example image recognition, speech processing, sentiment analysis, and predictive analytics. However, while both share the common objective of enabling computers to learn from data, they differ in methodology, structure, and application scope. This background study explores the theoretical foundations and characteristics of both approaches and offer a summary from this four selected algorithms: XGBoost, SVM, DistilBERT, and RoBERTa.

## 2. Related Work

[1] Muhammad Salman, Muhammad Ikram, Mohamed Ali Kaafar, 2024

Salman et al. [1] tackled the challenge of evolving spammer tactics by releasing a large publicly available SMS dataset of over 68,000 messages. Their study performs a longitudinal analysis of spam patterns and benchmarks both traditional ML and deep neural network models, also evaluating commercial anti-spam tools against adversarial evasion techniques.We investigate both the linguistic and structural characteristics of the messages to compare how different machine learning modelsfrom traditional algorithms to advanced deep neural networksperform in spam detection. Additionally, we assess existing anti-spam tools and commonly used commercial solutions in terms of their effectiveness at distinguishing spam from legitimate messages.

[2] Suleiman Y. Yerima;Abul Bashar

Yerima and Bashar [2] proposed a one-class SVM-based anomaly detection framework for SMS spam that trains exclusively on legitimate messages, removing the dependency on labeled spam data. Tested on a benchmark of 5,574 messages, their system achieved 98% overall accuracy and a 100% spam detection rate with only a 3% false positive rate, outperforming conventional supervised classifiers.

[3]ShaghayeghHosseinpour;HadiShakibian

Hosseinpour and Shakibian [3] addressed the class imbalance problem in SMS spam datasets by proposing an ensemble method combining Random Forest and Logistic Regression. Their model was validated on two real-world datasets, with accuracy and AUC as evaluation metrics, demonstrating that ensemble approaches can effectively handle the uneven distribution of spam versus ham messages.

[4] SamadhanM. Nagare;Pratibha P. Dapke;SyedAhteshamuddinQuadri;Sagar B. Bandal

Nagare et al. [4] applied the Naïve Bayes classifier to the publicly available UCI SMS spam dataset for spam versus ham classification. Their work highlights that despite Naïve Bayes being a simple probabilistic method, it remains effective as a baseline for SMS spam detection when the dataset is properly preprocessed.

[5]M. Hassan Shirali-Shahreza;MohammadShirali-Shahreza

Shirali-Shahreza and Shirali-Shahreza [5] introduced a CAPTCHA-based challenge-response mechanism to filter SMS spam without relying on message content. In their method, the recipient system sends the message originator a visual question — an image of an object with a multiple-choice name prompt — and classifies the message as legitimate only if the sender answers correctly, thereby distinguishing human users from automated spam bots.

## 3.Materials and Methods

This chapter outlines the methodological framework followed to conduct the comparative analysis. It

discusses both the existing systems and the proposed system developed for this study.

This methodology emphasizes understanding and comparing the various performances for ML as well as DL methods among Unsolicited message perception from SMS data. This whole process comprises dataset collection, cleaning, and processing, model training, accuracy testing, and finally, model comparison. Primary subjective assigned for determine which traditional ML or DL gives better accuracy and reliability for SMS spam detection.

Early systems for the detection of SMS spam relied predominantly upon manual rules and traditional techniques of machine learning. These methods generally tried to identify spam messages based on specific words, phrases, or patterns that repeatedly appear in unwanted messages.

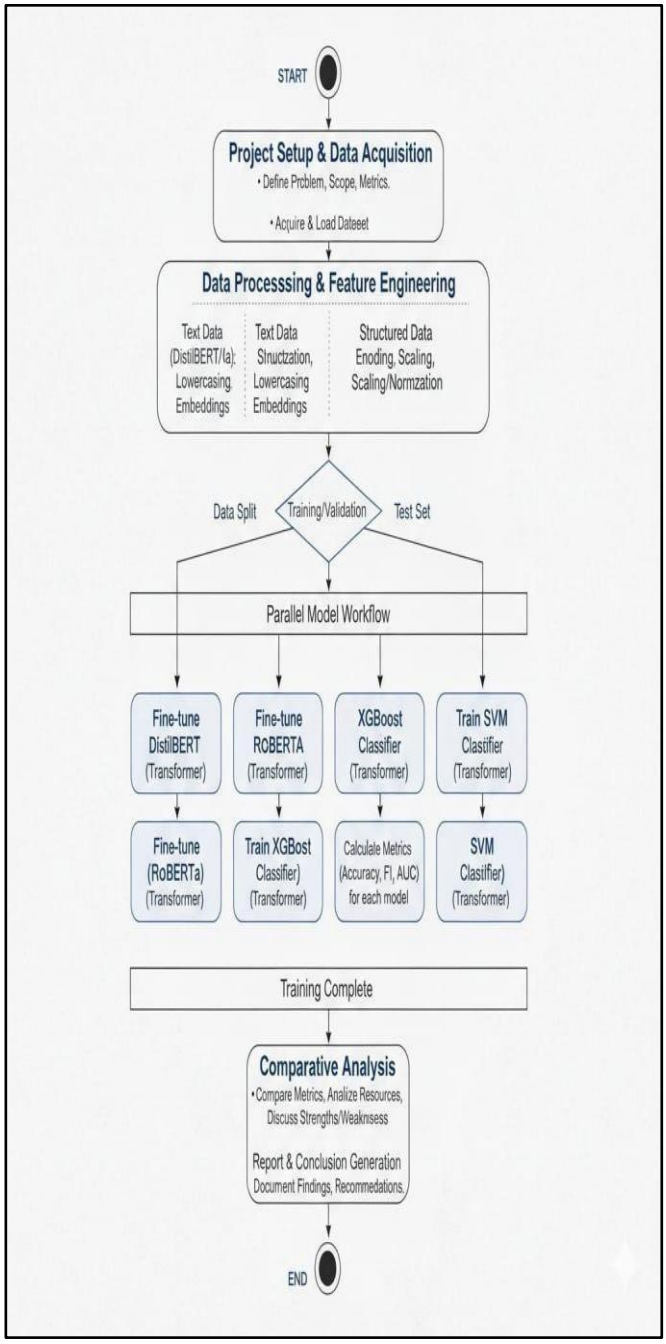



**Figure 1 : System Architecture for Spam Classification**

Detection was usually keyword- based filtering and statistical methods, where certain trigger words like "lottery," "win," or "free" defined a message as spam. This approach worked for the older forms of spam but gradually lost efficiency as spammers used more creative and masked forms of language to circumvent the filters.

ML performing be required to existed introduced to enhance accuracy and reduce reliance on

static rules. A portion of commonly used algorithms in existing systems include Bayesian classifier , SVM, and Logistic Regression. The listed techniques are developed using humanities information using a mathematical relationship to predict whether a new incoming SMS is spam or not. Features for training generally include frequency of words, message length, and special characters. The machine learning models, though performing better, are still dependent on feature engineering, where the important features or words to classify have to be decided manually.

One limitation of most existing systems is that they fail to understand message context. For instance, “free” may appear in spam messages, such as Win a free iPhone!, and in legitimate messages, such as “Free Wi-Fi available here”. Both rule-based and classic machine learning-based systems are unable to differentiate these contextual meanings, because they treat words in isolation without capturing sentence-level semantics. A further concern regarding about frameworks represents which they are not adaptive-if new types of spam messages or texting styles emerge, retraining the system with fresh data becomes necessary for it to remain effective.

Besides, predominantly old instrumentation acted as never customized in order to large-scale real-time detection. They performed well on small datasets but faced difficulties when scaling up with larger and more complex textual data. Notably, without deep representation methods for text, they failed to track hidden relationships among words and long-term dependences in messages. Therefore, while already a big step from manually detecting spam, actual frameworks static individual more shortcomings concerning accuracy, generalization, and adaptability.

Though,it exists that plenty of area in order to transformation within the period of existing systems regarding understanding context, scalability, and model adaptability. This paper proposes a system that should fill such gaps using state-of-the-art deep learning architectures to learn semantic meanings and boost accuracy of spam classification.

System design is fundamentally the utilization of system theory. This represents a structured process used to determine the interfaces, modules, and data of a system, thereby ensuring it satisfies specific demands.

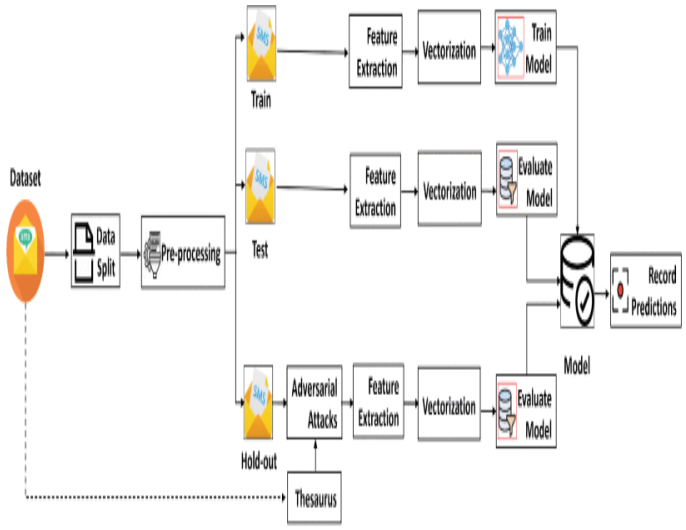



**Figure2 : System Design**

## 4.Results and Discussion

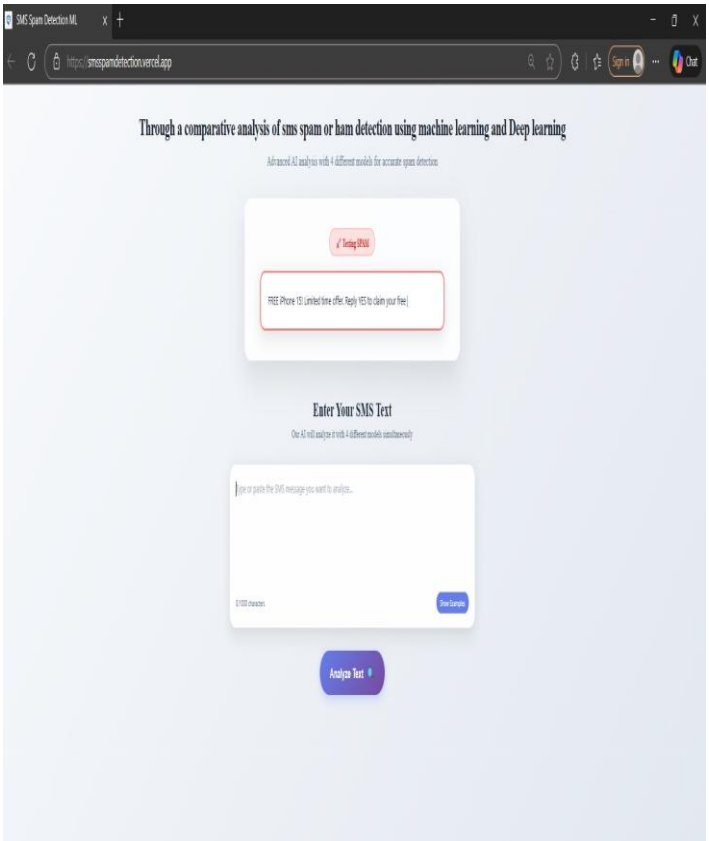


**Figure 3 : Main Page of the SMS Spam Detection System**

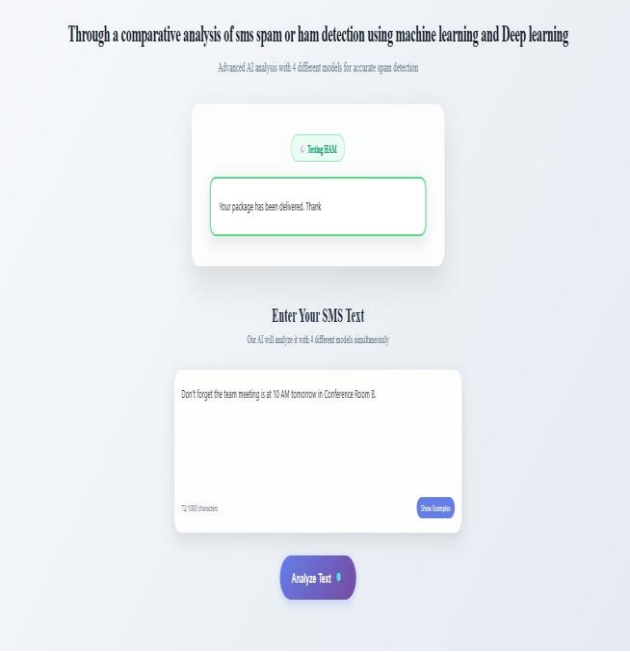



**Figure 4 : Sample Ham Message Input**

**Figure 5 : Ham Message Detection Result**

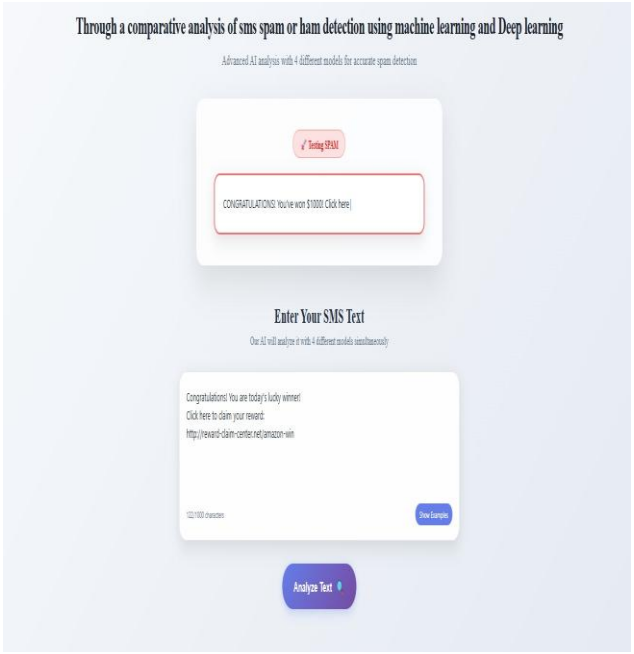
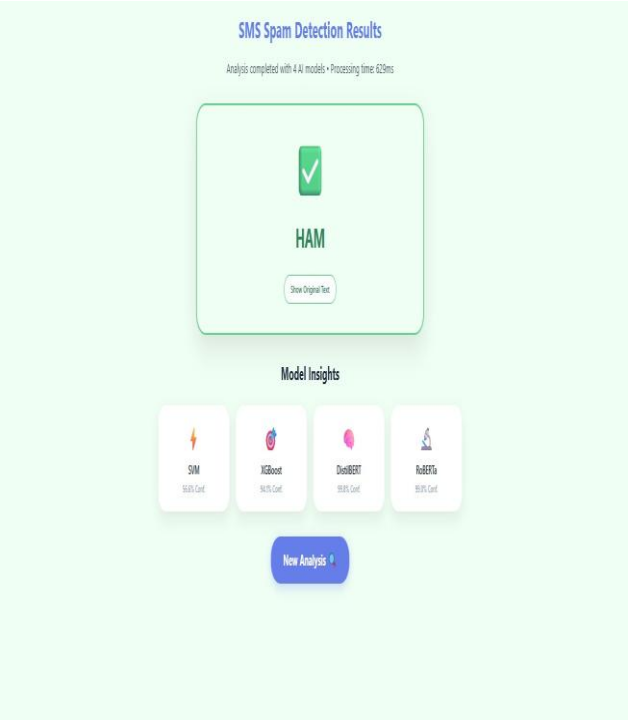
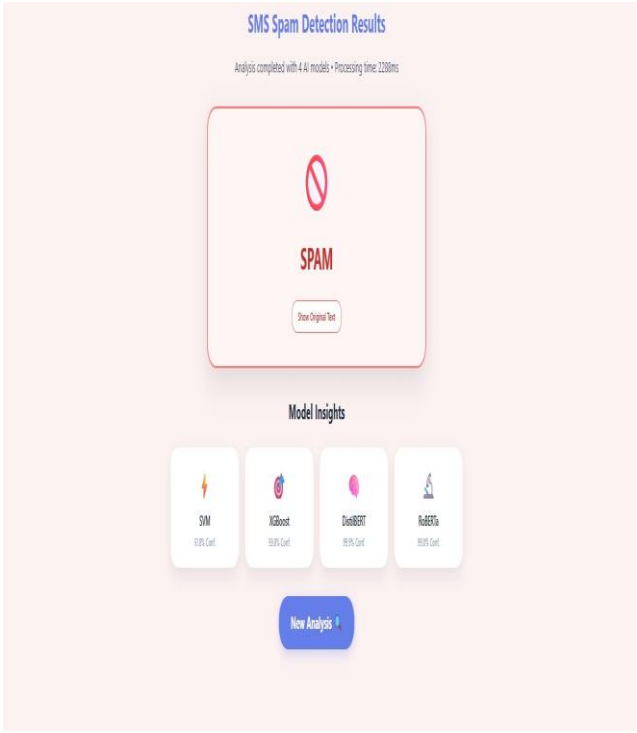



**Figure 6 : Spam Message Input Screen**



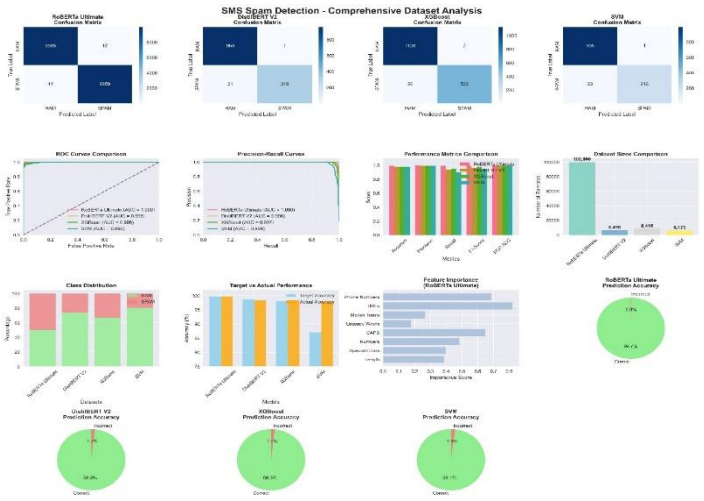



**Figure 7 :Spam Detection Output**

Validation results help identify overfitting or underfitting issues, ensuring the model generalizes well.

To evaluate model performance objectively, the following metrics are used:

- **Accuracy:** The fraction of trueassumed samples to total samples.
- **Precision:** The fraction of true assumedgood remarksresult in beneficial results.
- **Recall:** The fraction of true assumed positives Within the entirety of actual positives.
- **F1-Score:** The thebalanced average of Positive predictive value as well as True positive rate, distributing a balanced performance measure.

For visual analysis, confusion matrices and bar charts can be plotted to display model results side by side.

Once all models are evaluated, their performance metrics are compared. The comparative study highlights which model performs best under the same experimental conditions.

- RoBERTa and DistilBERT generally outperform traditional models when handling textual or contextual data due to their deep learning capabilities.
- XGBoost often excels in structured, numerical datasets because of its boosting approach.
- SVM, though simpler, provides a solid baseline for smaller datasets and linear classification tasks.

This comparison ensures a fair analysis based on model complexity, computation time, and prediction quality.

After comparison, the best-performing model is identified created on the highest accuracy and balanced F1-score. Graphical representations such as precision vs. model type and confusion matrices provide a clear understanding of how each model behaves. The results are also discussed in relation to dataset characteristics and computational requirements.

## 5. Conclusion

This comparative study successfully evaluated the execution of many diverse machine learning models SVM (Support Vector Machine), XGBoost, and the transformer-based DistilBERT and RoBERTaon the task of binary SMS spam classification (spam vs. ham).

The results clearly demonstrated a performance hierarchy, typically with the sophisticated transformer models, RoBERTa and DistilBERT, achieving the peak overall precision along with F-scores. RoBERTa often maintained a slight edge due to its extensive pre-training and robust architecture, proving most effective at capturing complex semantic nuances.

Conversely, XGBoost offered a compelling balance, delivering strong presentation comparable to traditional deep learning methods while using significantly less training time and computational resources. The SVM, though the simplest, provided a solid, interpretable baseline, often excelling in situations where feature engineering was highly optimized.

In conclusion, the choice of model hinges on the project's priorities. For absolute peak performance, RoBERTa is the recommended choice. However, when balancing accuracy with efficiency and resource constraints, XGBoost provides the most practical and efficient solution.

## References

[1]M. Salman, M. Ikram, and M. A. Kaafar, "Investigating evasive techniques in SMS spam filtering: a comparative analysis of machine learning models," *IEEE Access*, vol. 12, pp. 24306–24324, 2024, doi: 10.1109/ACCESS.2024.3364671.

[2] S. Y. Yerima and A. Bashar, "Semi-supervised novelty detection with one class SVM for SMS spam detection," in *Proc. 29th Int. Conf. Systems, Signals and Image Processing (IWSSIP)*, Sofia, Bulgaria, Jun. 2022, pp. 1–4, doi: 10.1109/IWSSIP55020.2022.9854496.

[3] S. Hosseinpour and H. Shakibian, "An ensemble learning approach for SMS spam detection," in *Proc. 9th Int. Conf. Web Research (ICWR)*, Tehran,Iran,May2023, pp. 125–128, doi: 10.1109/ICWR57742.2023.10139070.

[4] S. M. Nagare, P. P. Dapke, S. A. Quadri, and S. B. Bandal, "Short message service (SMS) mobile spam detection using Naïve Bayes," in *Proc. 5th Int. Conf. Mobile Computing and Sustainable Informatics (ICMCSI)*, 2024, doi: 10.1109/ICMCSI61536.2024.00016.

[5] M. H. Shirali-Shahreza and M. Shirali-Shahreza, "An anti-SMS-spam using CAPTCHA," in *Proc. 2008 ISECS Int. Colloq. Computing, Communication, Control and Management (CCCM)*, Guangzhou, China, vol. 2, Aug. 2008, pp. 318–321.

[6] T. Xia and X. Chen, "A discrete hidden Markov model for SMS spamdetection,"Applied Sciences, vol. 10, no. 14, p. 5011, 2020, doi: 10.3390/app10145011.

[7] M. Nivaashini, R. S. Soundariya, A. Kodieswari, and P. Thangaraj, "SMSspamdetection using deep neural network," International Journal of Pure andAppliedMathematics, vol. 119, no. 18, pp. 2425–2436, 2018.

[8] M. Gupta, A. Bakliwal, S. Agarwal, and P. Mehndiratta, "A comparative studyofspam SMS detection using machine learning classifiers," in Proc. 11th Int. Conf.Contemporary Computing (IC3), Aug. 2018, pp. 1–6.

[9] G. S. Sravya, G. Pradeepini, and Vaddeswaram, "Mobile SMS spamfiltertechniquesusing machine learning techniques," International Journal of Scientific &TechnologyResearch, vol. XX, no. X, pp. XX–XX.

[10] F. Wei and T. Nguyen, "A lightweight deep neural model for SMS spamdetection,"in Proc. Int. Symp. Networks, Computers and Communications (ISNCC), Oct. 2020, pp.1–6.