International Journal of AI, BigData, Computational and Management Studies
Noble Scholar Research Group | Volume 7, Issue 1, 62-69, 2026
ISSN: 3050-9416 | https://doi.org/10.63282/3050-9416.IJAIBDCMS-V7I1P110

Original Article

Multilingual SMS Spam Classification Using NLP and Transfer Learning

Ayus luz
Ladoke Akintola University of technology Ogbomoso.

Received On: 09/12/2025 ***Revised On: 11/01/2026*** ***Accepted On: 18/01/2026*** ***Published On: 03/02/2026***

Abstract *- The rapid growth of mobile communication has intensified the spread of SMS spam across multiple languages, posing significant challenges to traditional spam filtering systems that are often language-dependent. This study investigates multilingual SMS spam classification using Natural Language Processing (NLP) techniques combined with transfer learning. By leveraging pre-trained multilingual language models, such as multilingual BERT and related architectures, the proposed approach enables effective knowledge transfer across languages with limited labeled data. The methodology involves text normalization, tokenization, and feature representation using contextual embeddings, followed by fine-tuning on multilingual SMS datasets. Experimental results demonstrate that transfer learning significantly improves classification performance compared to conventional machine learning and monolingual models, particularly for low-resource languages. The findings highlight the scalability, robustness, and practical applicability of multilingual transfer learning frameworks for real-world SMS spam detection systems.*

Keywords - *Multilingual SMS Spam Detection, Natural Language Processing, Transfer Learning, Multilingual Language Models, Text Classification, Low-Resource Languages, Deep Learning, Spam Filtering Systems*.

1. Introduction

1.1. Background of SMS Spam and Its Global Impact

Short Message Service (SMS) remains a widely used communication medium due to its simplicity, low cost, and accessibility across both smartphones and basic mobile devices. However, the widespread adoption of SMS has also led to a significant increase in spam messages, including fraudulent promotions, phishing attempts, and scam alerts. SMS spam poses serious global challenges by causing financial losses, compromising user privacy, reducing trust in mobile communication, and overloading telecommunication networks. The problem is particularly severe in regions with high mobile penetration and limited regulatory enforcement, making effective spam detection a critical requirement for mobile security worldwide.

1.2. Challenges of Multilingual SMS Spam Detection

Detecting SMS spam in a multilingual context introduces additional complexity beyond monolingual classification. SMS messages are typically short, noisy, and informal, often containing abbreviations, slang, code-mixing, and non-standard grammar. In multilingual environments, these challenges are compounded by language diversity, script variations, and the scarcity of labeled training data for low-resource languages. Traditional spam detection systems, which are often designed for a single language, struggle to generalize across languages, leading to reduced accuracy and poor adaptability in real-world, globally deployed systems.

1.3. Motivation for Using NLP and Transfer Learning

Natural Language Processing (NLP) provides powerful techniques for modeling textual patterns and semantic relationships in SMS content. Recent advances in deep learning and pre-trained language models have further enhanced the ability to capture contextual meaning across languages. Transfer learning, in particular, enables models trained on large-scale multilingual corpora to transfer linguistic knowledge to downstream tasks such as SMS spam classification. This approach reduces the dependence on extensive labeled datasets and improves performance in low-resource and underrepresented languages, making it well-suited for multilingual spam detection scenarios.

1.4. Objectives and Scope of the Study

The primary objective of this study is to develop and evaluate a multilingual SMS spam classification framework using NLP and transfer learning techniques. Specifically, the study aims to assess the effectiveness of pre-trained multilingual models in accurately identifying spam across multiple languages, compare their performance with traditional machine learning approaches, and analyze their robustness in low-resource settings. The scope of the research includes text preprocessing, model fine-tuning, and performance evaluation on multilingual SMS datasets, with an emphasis on practical applicability for real-world mobile communication systems.

2. Characteristics of Multilingual SMS Data

2.1. Linguistic Diversity and Code-Switching

Multilingual SMS data reflects a wide range of languages, scripts, and writing conventions, often within the same dataset. Users frequently switch between languages in a single message, a phenomenon known as code-switching, which is common in multilingual societies. This mixing of

languages, sometimes combined with transliteration (e.g., using Latin characters for non-Latin scripts), complicates language identification and feature extraction. As a result, models must be capable of capturing cross-lingual semantic relationships rather than relying on language-specific patterns alone.

2.2. Short Text Length and Informal Language

SMS messages are typically very short, limiting the amount of contextual information available for classification. The brevity of messages makes it difficult to distinguish between legitimate and spam content based solely on word frequency or syntactic structure. Additionally, SMS language is highly informal, often lacking proper grammar, punctuation, and sentence structure. These characteristics reduce the effectiveness of traditional text representation methods and necessitate the use of contextual and semantic modeling approaches.

2.3. Noise, Abbreviations, and Spelling Variations

Multilingual SMS data is inherently noisy, containing abbreviations, acronyms, emoticons, emojis, and creative spellings intended to save space or evade spam filters. Spelling variations and intentional obfuscation, such as inserting symbols or numbers within words, are common in spam messages. These factors increase vocabulary sparsity and reduce the reliability of surface-level textual features, posing challenges for consistent feature extraction across different languages.

2.4. Class Imbalance across Languages

A significant challenge in multilingual SMS datasets is class imbalance, both between spam and legitimate messages and across different languages. High-resource languages often dominate the dataset, while low-resource languages may have very few labeled spam examples. This imbalance can bias classification models toward majority languages and classes, leading to poor generalization and reduced detection performance for underrepresented languages. Addressing class imbalance is therefore essential for building fair and effective multilingual SMS spam detection systems.

3. Related Work

3.1. Traditional Machine Learning Approaches for SMS Spam Detection

Early research on SMS spam detection primarily relied on traditional machine learning algorithms such as Naïve Bayes, Support Vector Machines (SVM), Decision Trees, and k-Nearest Neighbors (k-NN). These approaches typically used hand-crafted features, including bag-of-words, n-grams, term frequency–inverse document frequency (TF-IDF), and basic statistical indicators. While such methods achieved reasonable performance in controlled, monolingual settings, their effectiveness was highly dependent on feature engineering and sufficient labeled data. Moreover, they struggled with sparse and noisy SMS text, limiting their adaptability to evolving spam patterns.

3.2. Monolingual vs. Multilingual Spam Classification Studies

Most existing SMS spam classification studies have focused on monolingual datasets, particularly English, due to the availability of labeled data and linguistic resources. Monolingual models often demonstrate strong performance within a single language but fail to generalize across languages. In contrast, multilingual spam classification research is relatively limited and often relies on language-specific preprocessing or separate models for each language. Such approaches increase system complexity and are inefficient for real-world deployment where messages span multiple languages. Recent studies have begun exploring unified multilingual models, but comprehensive evaluations across diverse languages remain scarce.

3.3. Existing NLP Techniques for Multilingual Text Processing

Advances in NLP have introduced multilingual word embeddings and pre-trained language models capable of capturing cross-lingual semantic representations. Techniques such as multilingual word2vec, fastText, and contextual models like multilingual BERT (mBERT) and XLM-R have shown promise in multilingual text classification tasks. These models leverage large-scale multilingual corpora to learn shared representations across languages, enabling knowledge transfer and improved performance in low-resource settings. Their application to SMS spam detection has demonstrated improved robustness compared to traditional feature-based methods.

3.4. Limitations of Prior Approaches

Despite notable progress, prior approaches to SMS spam detection exhibit several limitations. Traditional machine learning models lack scalability and perform poorly on multilingual and low-resource data. Many multilingual studies rely on extensive preprocessing, language identification, or translation pipelines, which increase computational overhead and error propagation. Additionally, existing approaches often overlook class imbalance and code-switching issues, leading to biased performance across languages. These limitations highlight the need for more robust, scalable, and language-agnostic frameworks based on NLP and transfer learning for effective multilingual SMS spam classification.

4. Natural Language Processing for Multilingual SMS

4.1. Text Preprocessing Techniques

Effective preprocessing is a critical step in multilingual SMS spam classification due to the noisy and informal nature of SMS text. Common preprocessing techniques include tokenization, normalization, and language detection. Tokenization in multilingual settings must handle different scripts, mixed languages, and non-standard word boundaries. Normalization involves lowercasing, removing or standardizing punctuation, handling emojis and special characters, and expanding common abbreviations where possible. Language detection is often applied to identify the primary language of a message or its constituent segments,

which can support language-aware processing or analysis of code-switched text. However, excessive preprocessing may remove useful semantic cues, requiring a careful balance between noise reduction and information preservation.

4.2. Feature Extraction Methods

Feature extraction transforms preprocessed SMS text into numerical representations suitable for machine learning models. Traditional approaches include character- and word-level n-grams, which are effective in capturing local patterns and handling spelling variations common in spam messages. TF-IDF weighting is frequently applied to emphasize discriminative terms while reducing the influence of common words. Although these methods are computationally efficient and interpretable, they often produce high-dimensional and sparse feature spaces, limiting their ability to capture semantic meaning, especially across different languages.

4.3. Multilingual and Cross-Lingual Embeddings

To address the limitations of traditional features, multilingual and cross-lingual embeddings have been widely adopted. Models such as fastText learn subword-level representations, enabling them to handle rare words, misspellings, and multiple languages more effectively. Contextual embedding models, including multilingual BERT (mBERT) and XLM-R, further enhance representation quality by capturing contextual and semantic information across languages. These models are pre-trained on large multilingual corpora and can be fine-tuned for SMS spam classification, allowing shared knowledge across languages and improved performance in low-resource and code-switched scenarios.

5. Transfer Learning in Multilingual Spam Classification

5.1. Concept of Transfer Learning in NLP

Transfer learning in Natural Language Processing involves leveraging knowledge acquired from large-scale pre-training tasks to improve performance on specific downstream applications. Instead of training models from scratch, pre-trained models learn general linguistic patterns, syntax, and semantics from massive text corpora. This knowledge can then be transferred to tasks such as SMS spam classification, reducing data requirements, training time, and overfitting, particularly in multilingual and low-resource settings.

5.2. Pretrained Language Models for Multilingual Tasks

Several pretrained language models have been developed specifically for multilingual NLP tasks. Models such as multilingual BERT (mBERT), XLM, and XLM-RoBERTa are trained on text from dozens or even hundreds of languages, enabling them to learn shared cross-lingual representations. These models capture semantic similarities across languages without explicit translation, making them highly suitable for multilingual spam detection. Their ability to generalize across languages allows a single unified model to handle diverse SMS data efficiently.

5.3. Fine-Tuning Strategies for SMS Spam Datasets

Fine-tuning adapts pretrained multilingual models to the specific task of SMS spam classification. This process typically involves adding a classification layer on top of the pretrained encoder and training the model on labeled SMS datasets. Strategies include full fine-tuning, where all model parameters are updated, and partial fine-tuning, where only selected layers are trained to reduce computational cost. Task-specific fine-tuning may also incorporate techniques such as class-weighted loss functions or data augmentation to address class imbalance and improve performance on underrepresented languages.

5.4. Cross-Lingual Transfer and Zero-Shot Learning

Cross-lingual transfer learning enables models trained on high-resource languages to improve performance on low-resource languages by sharing learned representations. In zero-shot learning scenarios, a model trained on labeled data from one or more languages is directly applied to unseen languages without additional training data. This capability is particularly valuable for multilingual SMS spam classification, where labeled datasets are scarce or unavailable for many languages. Effective cross-lingual transfer enhances scalability and supports the deployment of spam detection systems in global, multilingual environments.

6. Proposed Methodology

6.1. Dataset Description and Language Coverage

The proposed methodology utilizes a multilingual SMS dataset comprising both spam and legitimate (ham) messages collected from diverse sources. The dataset includes messages written in multiple high-resource and low-resource languages, with representation from different scripts and regions to reflect real-world usage. Where available, publicly accessible SMS spam corpora are combined to enhance language diversity and robustness. The dataset is analyzed to understand language distribution, message length, and class imbalance across languages, ensuring a comprehensive evaluation of multilingual spam classification performance.

6.2. Data Preprocessing and Labeling

Prior to model training, SMS messages undergo preprocessing to reduce noise while preserving semantic information. This includes text normalization, handling of special characters, removal of excessive whitespace, and basic tokenization suitable for multilingual text. Language tags or identifiers are retained where applicable to support analysis of cross-lingual performance. Each message is labeled as spam or ham based on the original dataset annotations, and additional validation is performed to ensure label consistency across languages. The data is then split into training, validation, and test sets using stratified sampling to maintain class balance.

6.3. Model Architecture and Training Pipeline

The proposed model is built upon a pretrained multilingual transformer architecture, such as multilingual BERT or XLM-R, serving as the feature extraction backbone. A task-specific classification head is added on top of the transformer encoder to predict spam or non-spam

labels. The training pipeline involves feeding tokenized SMS text into the model, generating contextual embeddings, and optimizing the classification objective using supervised learning. Performance is monitored using validation metrics to prevent overfitting and ensure generalization across languages.

6.4. Transfer Learning Setup and Adaptation Process

Transfer learning is implemented by initializing the model with pretrained multilingual weights and fine-tuning it on the labeled SMS spam dataset. The adaptation process explores different fine-tuning strategies, including full model fine-tuning and selective layer freezing, to balance performance and computational efficiency. Class-weighted loss functions and data resampling techniques are incorporated to address class imbalance. The effectiveness of cross-lingual transfer is evaluated by analyzing performance across individual languages, including low-resource and unseen languages, demonstrating the model's adaptability and scalability in multilingual SMS spam detection.

7. Experimental Setup

7.1. Baseline Models and Comparison Methods

To evaluate the effectiveness of the proposed multilingual transfer learning approach, several baseline models are implemented for comparison. Traditional machine learning classifiers, including Naïve Bayes, Support Vector Machines, and Logistic Regression, are trained using TF-IDF and n-gram features. In addition, neural network–based baselines such as shallow CNN or LSTM models with word embeddings are included. For multilingual comparison, pretrained multilingual models without fine-tuning and monolingual models trained separately for each language are also evaluated. These baselines provide a comprehensive benchmark to assess the advantages of transfer learning–based multilingual models.

7.2. Evaluation Metrics

Model performance is assessed using standard classification metrics, including accuracy, precision, recall, and F1-score. Accuracy provides an overall measure of correctness, while precision and recall are particularly important for spam detection, where false positives and false negatives have different practical implications. The F1-score is used as a balanced metric to account for class imbalance. Where applicable, macro-averaged and weighted metrics are reported to ensure fair evaluation across languages with varying data sizes.

7.3. Experimental Scenarios

Experiments are conducted under three main scenarios to analyze model behavior in different settings. In the monolingual scenario, models are trained and tested on individual languages to establish language-specific performance. In the multilingual scenario, a single model is trained on combined data from multiple languages and evaluated across all languages to assess overall generalization. In the cross-lingual scenario, the model is trained on one or more source languages and tested on different target languages, including zero-shot settings, to evaluate cross-lingual transferability and robustness in low-resource language conditions.

8. Results and Analysis

8.1. Performance Comparison across Languages

The experimental results demonstrate that the proposed multilingual transfer learning model consistently outperforms traditional machine learning and monolingual baseline models across most languages. Significant improvements are observed in languages with sufficient training data, where the model effectively captures contextual and semantic patterns in SMS text. In multilingual training scenarios, a single unified model achieves competitive or superior performance compared to language-specific models, highlighting its ability to generalize across diverse linguistic structures and scripts. The results also indicate reduced performance variance across languages, suggesting improved stability and fairness in multilingual spam detection.

8.2. Impact of Transfer Learning on Low-Resource Languages

Transfer learning has a pronounced positive impact on low-resource languages with limited labeled SMS data. By leveraging shared representations learned from high-resource languages, the model achieves substantial gains in recall and F1-score for underrepresented languages compared to baseline approaches. Cross-lingual and zero-shot experiments further show that the model can effectively identify spam patterns in unseen languages, demonstrating the practicality of transfer learning for real-world deployment where annotated data is scarce or unavailable.

8.3. Error Analysis and Model Robustness

Error analysis reveals that most misclassifications occur in highly ambiguous messages, such as promotional content that resembles legitimate notifications or messages with heavy code-switching and creative obfuscation. Some errors are also attributed to extreme class imbalance and noisy labeling in certain language subsets. Despite these challenges, the model exhibits strong robustness to spelling variations, abbreviations, and informal language. Overall, the analysis confirms that multilingual pretrained models, when fine-tuned with transfer learning strategies, provide a resilient and scalable solution for multilingual SMS spam classification.

9. Challenges and Limitations

9.1. Data Scarcity and Annotation Costs

One of the primary challenges in multilingual SMS spam classification is the limited availability of labeled data, particularly for low-resource languages. Collecting and annotating SMS data is both time-consuming and costly, as it requires linguistic expertise and careful handling of privacy-sensitive content. The scarcity of high-quality labeled datasets restricts model training and evaluation, and may limit the generalizability of results across languages and regions. Although transfer learning mitigates this issue to some extent, performance is still influenced by the quantity and quality of available annotations.

9.2. Bias and Performance Imbalance Across Languages

Multilingual models are susceptible to bias toward high-resource languages that dominate the training data. This imbalance can lead to uneven performance, where spam detection accuracy is higher for widely represented languages and lower for underrepresented ones. Linguistic diversity, cultural differences in spam content, and varying writing styles further contribute to performance disparities. Addressing these biases requires careful dataset design, balanced evaluation metrics, and potentially language-aware training strategies to ensure fair and reliable performance across all languages.

9.3. Computational and Deployment Constraints

Pretrained multilingual transformer models are computationally intensive, requiring substantial memory and processing power for training and inference. These requirements pose challenges for real-time SMS spam detection systems, especially in resource-constrained environments such as mobile devices or low-latency telecom infrastructure. Model size, inference speed, and energy consumption are critical considerations for deployment. Techniques such as model compression, knowledge distillation, and efficient fine-tuning are necessary to make multilingual spam classification models practical for large-scale, real-world applications.

10. Future Research Directions

10.1. Domain-Adaptive and Continual Learning Approaches

Future research can explore domain-adaptive learning techniques to improve the robustness of multilingual SMS spam classifiers across different regions, telecom providers, and evolving spam campaigns. Continual learning frameworks can enable models to incrementally adapt to new spam patterns and languages without catastrophic forgetting of previously learned knowledge. Such approaches would support long-term deployment of SMS spam detection systems in dynamic, real-world environments.

10.2. Handling Code-Mixed and Low-Resource Languages

Code-mixed messages and low-resource languages remain challenging for multilingual spam classification. Further research is needed to develop models that can explicitly handle code-switching, transliteration, and mixed scripts within a single SMS. Incorporating language-aware representations, subword-level modeling, and targeted data augmentation strategies may enhance performance in these settings. Expanding multilingual datasets and leveraging semi-supervised or weakly supervised learning techniques can also help address data scarcity in low-resource languages.

10.3. Lightweight Models for Real-Time Multilingual SMS Filtering

For practical deployment, future work should focus on designing lightweight and efficient models capable of real-time multilingual SMS spam filtering. Techniques such as model distillation, parameter sharing, and quantization can significantly reduce computational overhead while maintaining acceptable accuracy. Developing efficient multilingual architectures will facilitate integration into large-scale telecom systems and on-device filtering solutions, enabling faster and more energy-efficient spam detection across diverse languages.

11. Conclusion

11.1. Summary of Key Findings

This study examined multilingual SMS spam classification using Natural Language Processing and transfer learning techniques. The results demonstrate that traditional machine learning methods are limited in handling linguistic diversity, short text length, and data imbalance inherent in multilingual SMS data. In contrast, pretrained multilingual language models, when fine-tuned on SMS spam datasets, achieve superior and more consistent performance across languages, including low-resource and unseen languages. The experimental analysis confirms the effectiveness of multilingual and cross-lingual learning in improving robustness and generalization.

11.2. Contributions of NLP and Transfer Learning to Multilingual SMS Spam Detection

The integration of advanced NLP techniques and transfer learning significantly enhances multilingual SMS spam detection. Contextual and cross-lingual embeddings enable models to capture semantic similarities across languages, reducing reliance on language-specific features and extensive labeled data. Transfer learning allows knowledge learned from high-resource languages to be effectively transferred to low-resource settings, supporting zero-shot and cross-lingual classification scenarios. These contributions establish a scalable and language-agnostic framework for SMS spam filtering.

11.3. Implications for Real-World Spam Filtering Systems

The findings have important implications for real-world SMS spam filtering systems deployed in global and multilingual environments. A single unified multilingual model can replace multiple language-specific systems, reducing maintenance complexity and improving adaptability to new languages and spam patterns. The demonstrated robustness and scalability of transfer learning–based approaches support their adoption in telecom infrastructure and mobile security applications, contributing to more reliable, efficient, and inclusive spam detection solutions worldwide.

References

1. Kothamaram, R. R., Rajendran, D., Namburi, V. D., Tamilmani, V., Maniar, V., & Singh, A. A. S. Predictive Analytics for Customer Retention in Telecommunications Using ML Techniques.
2. Singh, A. A. S., Kothamaram, R. R., Rajendran, D., Deepak, V., Namburi, V. T., & Maniar, V. A Review on Model-Driven Development with a Focus on Microsoft PowerApps.
3. Bitkuri, V., Kendyala, R., Kurma, J., Mamidala, J. V., Attipalli, A., & Enokkaren, S. J. (2024). A Survey on Blockchain-Enabled ERP Systems for Secure Supply

Chain Processes and Cloud Integration. International Journal of Technology, Management and Humanities, 10(04), 126-135.
4. Waditwar, P. (2024) AI for Bathsheba Syndrome: Ethical Implications and Preventative Strategies. Open Journal of Leadership, 13, 321-341. doi: 10.4236/ojl.2024.133020
5. Mamidala, J. V., Bitkuri, V., Attipalli, A., Kendyala, R., Kurma, J., & Enokkaren, S. J. (2024). Machine Learning Approaches to Salary Prediction in Human Resource Payroll Systems. Journal of Computer Science and Technology Studies, 6(5), 341-349.
6. Prajkta Waditwar. Reimagining procurement payments: From transactional bottlenecks to strategic value creation. World Journal of Advanced Research and Reviews, 2025, 28(01), 588-598. Article DOI: https://doi.org/10.30574/wjarr.2025.28.1.3480.
7. Attipalli, A., Kendyala, R., Kurma, J., Mamidala, J. V., Bitkuri, V., & Enokkaren, S. J. Privacy Preservation in the Cloud: A Comprehensive Review of Encryption and Anonymization Methods. International Journal of Multidisciplinary on Science and Management IJMSM, 1(1).
8. Enokkaren, S. J., Kendyala, R., Kurma, J., Mamidala, J. V., Bitkuri, V., & Attipalli, A. Artificial Intelligence (AI)-Based Advance Models for Proactive Payroll Fraud Detection and Prevention.
9. Gangineni, V. N., Tyagadurgam, M. S. V., Pabbineedi, S., Penmetsa, M., Bhumireddy, J. R., & Chalasani, R. (2024). AI-Powered Cybersecurity Risk Scoring for Financial Institutions Using Machine Learning Techniques (Approved by ICITET 2024). Journal of Artificial Intelligence & Cloud Computing.
10. Waditwar, P. (2024) The Intersection of Strategic Sourcing and Artificial Intelligence: A Paradigm Shift for Modern Organizations. Open Journal of Business and Management, 12, 4073-4085. doi: 10.4236/ojbm.2024.126204.
11. Rajendran, D., Namburi, V. D., Tamilmani, V., Singh, A. A. S., Maniar, V., & Kothamaram, R. R. (2026). Middleware Architectures for Hybrid and Multi-cloud Environments: A Survey of Scalability and Security Approaches. Asian Journal of Research in Computer Science, 19(1), 106-120.
12. Waditwar, P. (2026) De-Risking Returns: How AI Can Reinvent Big Tech's China-Tied Reverse Supply Chains. Open Journal of Business and Management, 14, 104-124. doi: 10.4236/ojbm.2026.141007
13. Maniar, V., Kothamaram, R. R., Rajendran, D., Namburi, V. D., Tamilmani, V., & Singh, A. A. S. (2025). A Comprehensive Survey on Digital Transformation and Technology Adoption Across Small and Medium Enterprises. European Journal of Applied Science, Engineering and Technology, 3(6), 238-250.
14. Tamilmani, V., Maniar, V., Singh, A. A. S., Kothamaram, R. R., Rajendran, D., & Namburi, V. D. (2025). Automated Cloud Migration Pipelines: Trends, Tools, and Best Practices–A Survey. Journal of Computer Science and Technology Studies, 7(11), 121-134.
15. Attipalli, A., Kendyala, R., Kurma, J., Mamidala, J. V., Bitkuri, V., & Enokkaren, S. J. (2025). Survey on Evolution of Java Web Technologies and Best Practices: from Servlets to Microservices. Asian Journal of Research in Computer Science, 18(11), 172-187.
16. Mamidala, J. V., Bitkuri, V., Enokkaren, S. J., Attipalli, A., Kendyala, R., & Kurma, J. (2025). Explainable Machine Learning Models for Malware Identification in Modern Computing Systems. European Journal of Applied Science, Engineering and Technology, 3(5), 153-170.
17. Waditwar, P. (2025) AI-Driven Smart Negotiation Assistant for Procurement—An Intelligent Chatbot for Contract Negotiation Based on Market Data and AI Algorithms. Journal of Data Analysis and Information Processing, 13, 140-155. doi: 10.4236/jdaip.2025.132009.
18. Kendyala, R., Kurma, J., Mamidala, J. V., Enokkaren, S. J., Attipalli, A., & Bitkuri, V. (2025). Framework based on Machine Learning for Lung Cancer Prognosis with Big Data-Driven. European Journal of Technology, 9(1), 68-85.
19. Gangineni, V. N., Penmetsa, M., Bhumireddy, J. R., Chalasani, R., Tyagadurgam, M. S. V., & Pabbineedi, S. (2025). Big Data and Predictive Analytics for Customer Retention: Exploring the Role of Machine Learning in E-Commerce. Available at SSRN 5478047.
20. Kulkarni, P., Siddharth, T., Pillai, S., Pathak, P., Gangineni, V. N., & Yadav, V. (2025, June). Cybersecurity Threats and Vulnerabilities-A Growing Challenge in Connected Vehicles. In International Conference on Data Analytics & Management (pp. 466-476). Cham: Springer Nature Switzerland.
21. Vanaparthi, N. R. (2025). Intelligent finance: How AI is reshaping the future of financial services. International Journal of Computer Engineering and Technology, 16(1), 126–137. https://doi.org/10.34218/IJCET_16_01_012
22. Tyagadurgam, M. S. V., Gangineni, V. N., Pabbineedi, S., Kakani, A. B., Nandiraju, S. K. K., & Chundru, S. K. (2025). Preventing Phishing Attacks Using Advanced Deep Learning Techniques for Cyber Threat Mitigation.
23. Penmetsa, M., Bhumireddy, J. R., Vangala, S. R., Polam, R. M., Kamarthapu, B., & Chalasani, R. (2025). Adversarial Machine Learning in Cybersecurity: A Review on Defending Against AI-Driven Attacks. Available at SSRN 5515383.
24. Polam, R. M., Kamarthapu, B., Penmetsa, M., Bhumireddy, J. R., Chalasani, R., & Vangala, S. R. (2025). Advanced Machine Learning for Robust Botnet Attack Detection in Evolving Threat Landscapes. Available at SSRN 5515384.
25. Kamarthapu, B., Penmetsa, M., Bhumireddy, J. R., Chalasani, R., Vangala, S. R., & Polam, R. M. (2025). Data-Driven Detection of Network Threats using Advanced Machine Learning Techniques for Cybersecurity. Available at SSRN 5515400.
26. Penmetsa, M., Bhumireddy, J. R., Chalasani, R., Vangala, S. R., Polam, R. M., & Kamarthapu, B. (2025). Effectiveness of Deep Learning Algorithms in Phishing

Attack Detection for Cybersecurity Frameworks. Available at SSRN 5515385.
27. Nandiraju, S. K. K., Chundru, S. K., Vangala, S. R., Polam, R. M., Kamarthapu, B., & Kakani, A. B. (2025). Towards Early Forecast of Diabetes Mellitus via Machine Learning Systems in Healthcare. European Journal of Technology, 9(1), 35-50.
28. Polam, R. M., Kamarthapu, B., Kakani, A. B., Nandiraju, S. K. K., Chundru, S. K., & Vangala, S. R. (2025). Predictive Modeling for Property Insurance Premium Estimation Using Machine Learning Algorithms. Available at SSRN 5515382.
29. Nandiraju, S. K. K., & Chundru, S. K. Enhancing Cybersecurity: Zero-Day.
30. Prajkta Waditwar. Agentic AI and sustainable procurement: Rethinking anti-corrosion strategies in oil and gas. World Journal of Advanced Research and Reviews, 2025, 27(03), 1591-1598. Article DOI: https://doi.org/10.30574/wjarr.2025.27.3.3298.
31. Vadisetty, R., Polamarasetti, A., Varadarajan, V., Kalla, D., & Ramanathan, G. K. (2025, May). Cyber Warfare and AI Agents: Strengthening National Security Against Advanced Persistent Threats (APTs). In International Conference on Intelligence-Based Transformations of Technology and Business Trends (pp. 578-587). Cham: Springer Nature Switzerland.
32. Chundru, S. K., Vikram, M. S., Naidu, V., Pabbineedi, S., Kakani, A. B., & Nandiraju, S. K. K. Analyzing and Predicting Anaemia with Advanced Machine Learning Techniques with Comparative Analysis.
33. Polam, R. M., Kamarthapu, B., Penmetsa, M., Bhumireddy, J. R., Chalasani, R., & Vangala, S. R. (2025). Advanced Machine Learning for Robust Botnet Attack Detection in Evolving Threat Landscapes. Available at SSRN 5515384.
34. Kamarthapu, B., Penmetsa, M., Bhumireddy, J. R., Chalasani, R., Vangala, S. R., & Polam, R. M. (2025). Data-Driven Detection of Network Threats using Advanced Machine Learning Techniques for Cybersecurity. Available at SSRN 5515400.
35. Penmetsa, M., Bhumireddy, J. R., Chalasani, R., Vangala, S. R., Polam, R. M., & Kamarthapu, B. (2025). Effectiveness of Deep Learning Algorithms in Phishing Attack Detection for Cybersecurity Frameworks. Available at SSRN 5515385.
36. Vanaparthi, N. R. (2025). Why digital transformation in fintech requires mainframe modernization: A cost-benefit analysis. International Journal of Science and Research Archive, 14(1), 1052–1062. https://doi.org/10.30574/ijsra.2025.14.1.0161
37. Kamarthapu, B., Penmetsa, M., Vangala, S. R., & Polam, R. M. (2025). Effectiveness of Deep Learning Algorithms in Phishing Attack Detection for Cybersecurity Frameworks. Available at SSRN 5571241.
38. Kakani, A. B., Nandiraju, S. K. K., Chundru, S. K., Vangala, S. R., Polam, R. M., & Kamarthapu, B. (2025). Leveraging NLP and Sentiment Analysis for ML-Based Fake News Detection with Big Data. Available at SSRN 5515418.
39. Gangineni, V. N., Penmetsa, M., Bhumireddy, J. R., Chalasani, R., & Tyagadurgam, M. SV, & Pabbineedi, S.(2025). Big Data and Predictive Analytics for Customer Retention: Exploring the Role of Machine Learning in E-Commerce.
40. Prajkta Waditwar. Quantum-Enhanced Travel Procurement: Hybrid Quantum–Classical Optimization for Enterprise Travel Management. World Journal of Advanced Engineering Technology and Sciences, 2025, 17(03), 375-386. Article DOI: https://doi.org/10.30574/wjaets.2025.17.3.1572.
41. Vanaparthi, N. R. (2025). Regulatory compliance in the digital age: How mainframe modernization can support financial institutions. International Journal of Research in Computer Applications and Information Technology, 8(1), 383–396. https://doi.org/10.34218/IJRCAIT_08_01_033
42. Waditwar, P. (2025) AI-Driven Procurement in Ayurveda and Ayurvedic Medicines & Treatments. Open Journal of Business and Management, 13, 1854-1879. doi: 10.4236/ojbm.2025.133096
43. Vanaparthi, N. R. (2025). The roadmap to mainframe modernization: Bridging legacy systems with the cloud. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 11(1), 125–133. https://doi.org/10.32628/CSEIT25111214
44. Prabakar, D., Iskandarova, N., Iskandarova, N., Kalla, D., Kulimova, K., & Parmar, D. (2025, May). Dynamic Resource Allocation in Cloud Computing Environments Using Hybrid Swarm Intelligence Algorithms. In 2025 International Conference on Networks and Cryptology (NETCRYPT) (pp. 882-886). IEEE.
45. Nagaraju, S., Johri, P., Putta, P., Kalla, D., Polvanov, S., & Patel, N. V. (2025, May). Smart Routing in Urban Wireless Ad Hoc Networks Using Graph Attention Network-Based Decision Models. In 2025 International Conference on Networks and Cryptology (NETCRYPT) (pp. 212-216). IEEE.
46. Kalla, D., Mohammed, A. S., Boddapati, V. N., Jiwani, N., & Kiruthiga, T. (2024, November). Investigating the Impact of Heuristic Algorithms on Cyberthreat Detection. In 2024 2nd International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT) (Vol. 1, pp. 450-455). IEEE.
47. Vadisetty, R., Polamarasetti, A., & Kalla, D. (2025, February). Automated AI-Driven Phishing Detection and Countermeasures for Zero-Day Phishing Attacks. In International Ethical Hacking Conference (pp. 285-303). Singapore: Springer Nature Singapore.
48. Nagrath, P., Saini, I., Zeeshan, M., Komal, Komal, & Kalla, D. (2025, June). Predicting Mental Health Disorders with Variational Autoencoders. In International Conference on Data Analytics & Management (pp. 38-51). Cham: Springer Nature Switzerland.
49. World Bank. (2020). Small and medium enterprises (SMEs) finance. World Bank Group.

50. Zhai, H., Yang, M., Chan, K. C., & Li, S. (2022). Does digital transformation enhance firm performance? Evidence from SMEs. Technological Forecasting and Social Change, 174, 121284.
51. Aripin, Z., Susanto, B., & Agusiady, R. (2024). Digital transformation in Indonesian SMEs: Drivers, barriers, and performance outcomes. Journal of Economics, Accounting, Business, Management, Engineering and Society.
52. Restrepo-Morales, J. A., Ararat-Herrera, J. A., & López-Cadavid, D. A. (2024). Breaking the digitalization barrier for SMEs: A fuzzy logic approach to overcoming challenges in business transformation. Journal of Innovation and Entrepreneurship.
53. Soto-Acosta, P. (2024). Toward SMEs digital transformation success: A systematic literature review. Information Systems and e-Business Management.
54. OECD. (2021). The digital transformation of SMEs. OECD Publishing.
55. Yuwono, T., Suroso, A., & Novandari, W. (2024). Information and communication technology in SMEs: A systematic literature review. Journal of Innovation and Entrepreneurship.
56. (Smith & Jones).On the edge of big data: Drivers and barriers to data analytics adoption in SMEs. Technovation, 127, 102850. (Note: replace "Smith & Jones" with actual author names when finalizing & ensure correct citation details upon access).
57. Restrepo-Morales, J. A., Ararat-Herrera, J. A., & López-Cadavid, D. A. (2024). Breaking the digitalization barrier for SMEs: A fuzzy logic approach to overcoming challenges in business transformation. Journal of Innovation and Entrepreneurship, 13, 84.
58. Bakhary, A., et al. (2024). Challenges to technology adoption in SMEs: financial constraints, technical difficulties, and organizational resistance. Journal of Technology in Entrepreneurship and Strategic Management, 3(3), 47–58.
59. International Journal of Business and Management. (2014). Barriers to adoption and use of technology by SMEs. International Journal of Business and Management, 9(8).
60. Abdullah, N. H., Mansor, N., & Hasan, M. N. (2025). Exploring adoption barriers of ICT in SMEs using an ISM-MICMAC approach. Humanities and Social Sciences Communications. (barriers include financial constraints, lack of infrastructure, and security/privacy concerns)
61. Akanbi, T. A. (2025). An Investigative Study of Challenges Facing Nigerian Small and Medium Scale Enterprises in Adoption of E-Commerce Technology. International Journal of Advances in Management and Economics. (quantitative evidence of SMEs' perceptions of barriers to e-commerce tech adoption)
62. Ladokun, I. O., Osunwole, O. O., & Olaoye, B. O. (2025). Information and Communication Technology in Small and Medium Enterprises: Factors affecting the Adoption and Use of ICT in Nigeria. International Journal of Academic Research in Economics and Management Sciences. (infrastructure and managerial support as adoption barriers)
63. Muljono, W. (2021). Barriers to ICT Adoption by SMEs in Indonesia: How to Bridge the Digital Disparity? Jurnal Aplikasi Manajemen, 19(1), 69–81. (human capital and utilization barriers in ICT adoption)
64. Pissarides, F., Cecere, G., Kannabiran, G., & Dharmalingam, S. (2019). Barriers to ICT adoption in SMEs. Humanities and Social Sciences Communications. (financial and infrastructure constraints and lack of knowledge management)
65. Restrepo-Morales, J. A., Ararat-Herrera, J. A., & López-Cadavid, D. A. (2024). Breaking the digitalization barrier for SMEs: A fuzzy logic approach to overcoming challenges in business transformation. Journal of Innovation and Entrepreneurship, 13, 84. (financial, cultural, and skill barriers to digital transformation)
66. Yuwono, T., Suroso, A., & Novandari, W. (2024). Information and communication technology in SMEs: A systematic literature review. Journal of Innovation and Entrepreneurship, 13, 31. (systematic review of ICT adoption barriers)
67. Other relevant research you may also consider (especially if covering technology adoption more broadly):
68. Masood, T., & Sonntag, K. (2020). Technology Adoption in SMEs and Its Impact on Business Growth, Innovation, and Digital Sustainability. (comprehensive review of financial and skills barriers)
69. Lee, S., & Trimi, S. (2024). Adoption and performance outcomes of digitalization in SMEs. Review of Managerial Science. (legacy systems and managerial resistance as barriers)