International Journal of Computer Science and Engineering Innovations
ISSN: 3107-9458 | Volume 2 Issue 1, 25-34, January-March 2026
DOI: https://doi.org/10.64137/31079458/IJCSEI-V2I1P104

Received: 08/01/2026 Revised: 13/01/2026 Accepted: 24/01/2026 Published: 26/01/2026

Original Article

Ensemble Learning Methods for Improving SMS Spam Classification Accuracy

EZEKIEL NYONG
University of Ibadan, Nigeria.

ABSTRACT: *SMS spam classification remains a critical challenge due to the short length, informal structure, and high variability of text messages. The performance of single machine learning and deep learning models is promising; however, the class imbalance problem, overfitting issue, and poor generalization across datasets frequently occur. In this paper, we investigate how ensemble learning can enhance the peak performance and stability in SMS spam filtering. Several ensemble models (bagging, boosting, stacking and voting based) are examined using heterogeneous base learners such as Naïve Bayes, Support Vector Machines, Random Forests or neural network models. The experimental results on benchmark SMS spam datasets show that the ensemble models consistently achieve higher performance compared with single classifiers in different measures, including accuracy, precision, recall and F1-score. The results demonstrate the advantages of ensemble learning in learning complementary decision boundaries and mitigating model bias and variance. The present work demonstrates the promise of ensemble methods for the development of trustworthy, scalable SMS spam filters to be efficiently applied in a production environment.*

KEYWORDS: *SMS spam classification, Ensemble learning, Bagging, Boosting, Stacking, Voting classifiers, Machine learning, Text classification, Natural language processing, Spam detection.*

1. INTRODUCTION

1.1. BACKGROUND AND SIGNIFICANCE OF SMS SPAM CLASSIFICATION:

SMS continues to be a very popular method of communication in personal, business and transactional scenarios. Nevertheless, the sudden proliferation of mobile communication also brought a rapid increase in unsolicited and malicious SMS spam messages such as phishing scams, fraudulent sales pitches, or misleading advertisements. This kind of message is not only disruptive to users but also threatens users' security and costs. Thus, successful classification of SMS spam is required to safeguard the users' interests and trust in mobile communication systems, as well as to aid telecommunication service providers' compliance with regulations.

1.2. LIMITATIONS OF SINGLE-MODEL CLASSIFIERS:

Conventional SMS spam filtering methods were heavily dependent on using a single machine learning or deep learning model, such as Naïve Bayes, Support Vector Machines (SVM), a decision tree, or neural network models in some cases. Although these models could obtain an acceptable result, they often have their own disadvantages. Single classifiers are also vulnerable to data imbalance, noise and feature representation, and can overfit training spams or not generalize well on other datasets or new spam trends. Consequently, using a single model is not reliable and robust enough for practical SMS spam filtering systems.

1.3. MOTIVATION FOR USING ENSEMBLE LEARNING:

Ensemble classifiers mitigate the limitations of single classifiers by aggregating a set of classifiers for more accurate and robust predictions. Gains have been made by taking advantage of the diversity between base learners and using methods such as bagging, boosting, stacking and voting to decrease bias variance and classification rates. Ensemble learning is especially attractive for SMS spam filtering because of the noisy and sparse characteristics of SMS text data. By assessing the homogeneity of model coverage, we showed that combining these complementary models enables the system to cover a wide spectrum of linguistic and spam behavior, ultimately resulting in more successful detection rates and increased immunity against spam evolution.

1.4. OBJECTIVES AND SCOPE OF THE STUDY:

This study aims above all others to evaluate how ensemble learning algorithms improve the accuracy of SMS spam classification. It is intended to test, compare, and analyze several aggregation techniques by ensembles with different base classifiers and performance measures. Second, it studies the effect of ensemble learning on robustness and generalization against SMS spam datasets. The focus of the work is narrowed down to text-based SMS spam classification, using state-of-theart machine learning methods, providing an account of practical applicability and enhancements in performance compared to single-model-based approaches.

2. OVERVIEW OF SMS SPAM CLASSIFICATION

2.1. CHARACTERISTICS OF SMS DATA

SMS has specific characteristics that set it apart from all other text sources and pose a challenge to classification. The length of the message is often quite short, leading to a lack of context information and feature representations. SMS text can be noisy, with spelling mistakes, abbreviations, emojis, URLs, phone numbers and other special characters. Moreover, the SMS language is extremely informal and filled with slang, code-mixed utterances and non-standard grammar. These traits make the text pre-processing, feature extraction, and semantic interpretation difficult to treat by automated spam detection systems.

2.2. COMMON MACHINE LEARNING AND DEEP LEARNING APPROACHES

Various machine learning methods have been utilized for SMS Spam classification. The classical algorithms are Naïve Bayes, Support Vector Machines, k-Nearest Neighbors, decision trees and logistic regression, which are also often used in conjunction with features such as bag-of-words or frequency of word (TF–IDF). In the last few years, there has been a great deal of interest in deep learning methods due to their capability to learn feature representations automatically. These include convolutional neural networks (CNNs) and recurrent neural networks (RNNs), long short-term memory (LSTM) networks, and transformer-based ones. Although deep models tend to be more accurate, they typically need larger databases and higher computational power.

2.3. CHALLENGES IN ACHIEVING HIGH ACCURACY AND ROBUSTNESS

In spite of the extensive studies, it's still difficult to achieve high and consistent accuracy and robustness in SMS spam classification. Some major challenges are that spam and non-spam distribution in class is highly imbalanced, data is sparse because e-mails have short-length messages, and the content of spam changes over time. Models pre-trained on a certain dataset may not generalize well to new domains or languages. On the other hand, feature degradation due to noise and informal languages, as well as over-generalization of adversarial or obfuscated spam messages, can escape detection. Such challenges motivate the requirement of reliable algorithms, e.g. ensemble learning, which can handle real-world SMS spam data that have to be addressed in a more diverse and complex environment than synthetic data generated by lab environment conditions.

3. ENSEMBLE LEARNING TECHNIQUES FOR SMS SPAM DETECTION

3.1. BAGGING-BASED METHODS

Bagging (Bootstrap Aggregating) is an ensemble algorithm which enhances classification accuracy by fitting a set of base learners on the training data using various bootstrap samples from it. All models are trained on the source of slightly variant data distribution, and the outputs are combined by majority vote or probability averaging. In SMS spam detection, bagging mitigates the noise, sparsity and instability that are typical of short text messages and has produced more reliable and robust predictions.

3.2. RANDOM FOREST

Random Forest is a popular bagging-based ensemble technique that builds up an ensemble of decision trees by bootstrapping samples of the instances and randomly selecting features at each split. This randomness enhances diversity among the trees and minimizes the correlation between individual classifiers. Random Forests provide a natural model for non-linear interactions between features, which is commonly seen in tasks like SMS spam classification, and yet have some resistance to overfitting complex representations of data such as n-grams and TF–IDF.

3.3. VARIANCE REDUCTION IN SMS SPAM CLASSIFICATION

High variance is a typical problem with SMS spam classifiers when small or noisy training data are used. Since bagging ensembles decrease the variance by aggregating predictions of several models, which averages away the mistakes of individual learners. This results in better generalization performance on a variety of datasets and improved robustness to changes in spam terms, vocabulary, or message composition.

3.4. BOOSTING-BASED METHODS

The boosting-based ensemble methods concentrate on the improvement of classification accuracy by training models iteratively, and each next model focuses more on instances that are misclassified by previous ones. Compared to bagging, boosting attempts to decrease bias and variance. Application to SMS spam filtering. Specifically for the task of SMS spam detection, boosting can be especially beneficial in pinpointing subtle or changing patterns of spam that are missed by simple classifiers.

3.5. ADABOOST

AdaBoost (Adaptive Boosting) gives relatively large weights to misclassified SMS messages during training, thereby compelling the next classifiers to concentrate on hard instances. Weak learners, frequently some variant of the decision stump, are aggregated to form a strong classifier based on weighted voting. AdaBoost has proven to be effective in the classification of SMS spam, as it enhances detection on boundary-line and ambiguous spam items, but may not perform robustly against a noisy environment.

3.6. GRADIENT BOOSTING AND XGBOOST

The Gradient Boosting constructs an ensemble stage-wise with the intent to minimize a loss via gradient descent. XGBoost, developed as optimized implementation of gradient boosting, implements regularization, parallel processing and support for sparse input data. These are particularly suitable for SMS spam detection as they can model complex decision boundaries and work naturally in high-dimensional feature spaces. They tend toward higher performance than classical classifiers, particularly when feature engineering is done well.

3.7. HANDLING HARD-TO-CLASSIFY SPAM MESSAGES

Boosting is particularly good at dealing with hard-to-classify spam, which may look something like real text or use various obfuscatory techniques. Through multiple repetitions of focusing on the misclassified, boosting ensembles creates finer partitions along the line between spam and non-spam messages, leading to both higher recall and overall better performance for difficult cases.

3.8. STACKING AND BLENDING

Stacking and blending are sophisticated ensemble techniques that aggregate the predictions made by multiple base classifiers via a meta-learner. Rather than taking a flat vote, such methods use part of the data to train an additional model that will directly learn how to best combine with each output result from the base learners.

3.9. META-LEARNING FRAMEWORK

Stacking trained base classifiers on the original to-date examples are then used as input for a new classifier. The meta-learner models the interactions among base model outputs and discovers adaptive fusion rules. For SMS spam detection, the meta-learning framework can leverage the complementary advantages of various models to enhance classification performance and robustness.

3.10. COMBINING HETEROGENEOUS BASE CLASSIFIERS

Methods like stacking and blending show high performance on diverse classifiers, including Naïve Bayes, Support Vector Machines, Random Forests, and neural networks. Every model is capturing the specific characteristics of SMS text patterns, and the aggregation of these models improves generalization with less reliance on any one particular learning algorithm.

3.11. VOTING-BASED ENSEMBLES

The ensembles based on voting do not need more training stages after prediction from multiple classifiers are combined. These approaches are straightforward and low in computation, easily deployed for real-time SMS spam filtering.

3.12. HARD VOTING VS. SOFT VOTING

In a hard voting scheme, the classes' votes are directly compared, and the majority class is chosen as the output label. Soft voting involves predicting class probabilities and averaging them to get the final decision. On SMS spam, soft voting is typically better than hard voting, since it takes model confidence into account and ensures that the decision boundary is smoother.

3.13. PERFORMANCE COMPARISON

In the literature, various empirical studies confirm that ensemble methods are efficient in beating single-model-based classifiers for SMS spam detection. The lower blocks with bagging-based and voting ensembles can achieve comparable performance in a time-efficient manner, or as measured by the speed of model fusion time. The decision of the ensemble method varies according to dataset properties, resource limitations and deployment scenarios; however, ensemble learning is an efficient solution that yields a good classification accuracy for SMS spam.

4. FEATURE REPRESENTATION AND DIVERSITY IN ENSEMBLES

4.1. ROLE OF FEATURE DIVERSITY IN ENSEMBLE PERFORMANCE:

Efficient dual learning and feature representation are two important factors in ensemble-based methods. The ensemble accuracy relies not only on the diversity of base classifiers, but also on the diversity of features used for training. While models trained on different feature representations recover complementary information about the data, errors become less correlated. In feature diversity, particularly in the context of SMS spam classification due to their short and noisy properties, feature diversity allows ensembles to capture diverse linguistic patterns, signals of SPAM and contextual information for improved robustness and generalization.

4.2. TRADITIONAL FEATURES (TF-IDF, N-GRAMS)

Classical feature extraction methods, like bag-of-words (BoW), feature-based TF-IDF computation and n-grams on words or characters are commonly employed in SMS spam identification. These features perform well for capturing shallow patterns, such as frequently used spam keywords, sequences of characters and structural clues such as repeated symbols or numbers. N-grams capture the underlying pattern of misspellings and gibberish, which are typical features of spam reports. However, these methods often produce high-dimensional and sparse feature spaces that may not represent deeper semantical meanings.

4.3. WORD EMBEDDINGS AND CONTEXTUAL REPRESENTATIONS

Word embedding methods alleviate the sparsity issue associated with the traditional features by learning to vectorise words in a dense and low-dimensional space where semantic relatedness is preserved. Word embeddings, for example, Word2Vec, GloVe, FastText, preserve semantic and syntactic similarity between words, which helps the model to generalize well in different vocabularies. In recent times, context-aware representations obtained from transformer-based models, such as BERT and its derivatives, have achieved great success in SMS spam classification by considering the sense of a word based on surrounding content. These representations are well-suited for discriminating between very similar spams that look like real SMS communications.

4.4. COMBINING MULTIPLE FEATURE EXTRACTION METHODS

It is of great importance to utilize multiple feature extraction approaches to boost the ensemble diversity and performance. For instance, conventional surface linguistic patterns as well as additional deep semantic content can be represented at the same time by exploiting TF-IDF-type features as well as word embeddings or contextual word representations. In an ensemble, different base learners can be trained on distinct sets of features, or feature representations can be concatenated to form richer inputs. In SMS spam filtering, such hybrid design feature strategies empower the ensemble of classifiers to take advantage of complementary strengths, with overall better performance in terms of performance accuracy, robustness and generalizability to new spamming tactics.

5. HANDLING CLASS IMBALANCE WITH ENSEMBLES

5.1. IMPACT OF IMBALANCED SMS DATASETS

Class imbalances are common in SMS spam classification, specifically for the case of SMS and ham versus spam (stochastic gradient descent). This imbalance can lead the learning algorithm toward over-fitting on the majority class and, while achieving high overall accuracy, generally leads to poor spam detection performance (low recall) for the spam member of the two classes. In practice, considering spam as a false negative can cause security challenges and end-user dissatisfaction. Thus, the problem of dealing with class imbalance is essential for proposing accurate and efficient SMS spam filters.

5.2. ENSEMBLE METHODS WITH SAMPLING TECHNIQUES

Ensemble learning algorithms are an efficient method to address the class imbalance problem by integrating with data-level sampling techniques. Several others, downsampling won't have that big of an impact. Unfortunately, there is no magical technique for solving the problem; if we had a method to remove these completely neutral lines, we could have done so several months ago. When combined with ensembles such as bagging or boosting, each base learner may be trained on a differently balanced data subset that better strengthens the diversity and increases minority-class recognition 13. In the context of SMS spam classification, ensemble-based sampling methods allow models to learn more discriminative spam patterns without overfitting or losing crucial information from normal messages.

5.3. COST-SENSITIVE ENSEMBLE LEARNING

Cost-sensitive ensembles learning deal with class imbalance at the algorithm level by giving higher misclassification costs to the minority class, spam. This learning objective forces classifiers to emphasize their true positive rate for training. Cost-sensitive versions of boosting and decision tree–based ensemble methods are very effective in dealing with the class imbalance problem, because they embed either class weights or cost matrices into the induction process. For targeted SMS spamming, the cost-sensitive ensembles decrease recall and F1-score on legitimate messages, but enhance performance (recall, F1-score) on spam; it can be suitable for realistic deployment where false negatives have a larger adverse effect.

6. EXPERIMENTAL DESIGN AND EVALUATION METRICS

6.1. BENCHMARK SMS SPAM DATASETS

When testing the efficacy of ensemble learning techniques on SMS-based spam tagging, it is common to apply publicly available benchmarking corpora that allow for reproducible/fair comparison. Popular datasets include SMS Spam Collection and other annotated corpora of spam and non-spam messages. These datasets are of different sizes, languages, and spam proportions, which would allow an experiment for testing the model's robustness/generalizability in a realistic situation. Multiple datasets enable analysis of the performance of ensemble models under various data and spam patterns.

6.2. EVALUATION METRICS

We use various evaluation metrics in order to have a holistic view of the classification performance. The accuracy is defined as the ratio of the total number of correctly classified messages to all tested samples, and is not an effective measure when dealing with the class imbalance problem. The precision is the ratio between the number of spam messages correctly predicted by the model (true positives) and all the spam predictions, since it reflects how many false-positive errors the classification process makes. Recall, the ratio of correctly identified spam messages to all spam messages is crucial for deployment in real-world environments. The F1-score is a weighted average of the precision and recall, and thus offers a balanced view of spam detection performance. Moreover, the Area Under the Receiver Operating Characteristic Curve (AUC) characterizes how well spam and legitimate messages are distinguished across decision thresholds, giving an insight into the general performance of ranking.

6.3. CROSS-VALIDATION AND STATISTICAL SIGNIFICANCE TESTING

Cross-validation is popular for the purpose of obtaining stable performance estimates and mitigating the bias in evaluation. Strategies like k-fold cross-validation help the model to be trained (and tested) for different data splits, making results more robust. To support the performance increase, a statistical significance comparison of ensemble methods to the baseline of classifiers is performed. Statistical tests, such as paired t-tests, Wilcoxon signed-rank tests , or McNemar's test, can determine whether they are significant (and thus due to random variation). These evaluation models, in concert, form a rigorous basis for comparing the effectiveness of ensemble learning methods to SMS spam classification.

7. PERFORMANCE ANALYSIS AND DISCUSSION

7.1. COMPARISON OF ENSEMBLE MODELS VS. SINGLE CLASSIFIERS

It is consistently shown in the experiments that it is beneficial to generate ensemble models when classifying SMS spam into several groups, rather than using a single classifier. Although the accuracy of simple methods such as Naïve Bayes or Support Vector Machines may be acceptable, their performance depends on bias and variance aspects or becomes sensitive to feature encoding. Besides, an ensemble of the classifiers, boosting, stacking, and Random Forests has higher accuracy, better recall for the spam class and F1-scores. These profits are derived from the working of ensembles to combine disparate, complementary predictive pattern information from diverse base learners in a way that decreases dependency on any one model's assumptions and limitations.

7.2. ROBUSTNESS TO NOISE AND ADVERSARIAL VARIATIONS

SMS spam corpus is noisy and can be adversarily manipulated (e.g., by using obfuscation, misspelling or by the appearance of deceptive lexical patterns). Combining classifiers is more resilient to noise and adversarial perturbations than individual ones. The bagging-based approach decreases sensitivity to noisy samples by variance reduction, and the boosting-based approach increases accuracy for the difficult and marginal spam examples. Stacking and feature-diverse ensembles also contribute to robustness by integrating models trained over diverse representations, thus making it harder for adversarial spam to bypass detection from all the parts at once.

7.3. COMPUTATIONAL COMPLEXITY AND SCALABILITY CONSIDERATIONS

Although these ensemble methods have superior performance, they bring in extra computational cost since we need to train and predict from multiple models. Methods like Random Forests or boosting demand more memory and runtime compared to single classifiers, especially for high-dimensional SMS feature spaces. Nevertheless, the majority of the ensemble approaches can be naturally parallelized and thus can be trained efficiently on current computational platforms. Ensemble methods based on voting strike a good balance between performance and computational overhead, while optimized realizations, like XGBoost, enhance scalability. In real-world SMS spam filtering systems, the selection of ensemble methods will consider a compromise between classification performance and the deployment constraints, such as RT processing and resource limitations.

8. CHALLENGES AND LIMITATIONS

8.1. INCREASED TRAINING AND INFERENCE COSTS

An important disadvantage of ensemble methods is the added computational expense for training and testing multiple models. Ensembles are much more expensive in computation and memory compared to single classifiers, especially for complicated paradigms, namely boosting and stacking. The scaling costs of this expensive process increase even more when we use the spam classification task using high-dimensional feature spaces. When performing inference, combining predictions from several base learners can lead to latency issues, which is not desirable, especially for large-scale or real-time filtering systems.

8.2. MODEL INTERPRETABILITY ISSUES

Ensemble model performance is often increased at the expense of interpretability. Simple classifiers like Naïve Bayes or decision trees have transparent decision-making, while ensembles (those with many base learners/ deep models) are often less interpretable. It can be hard to see why a certain SMS is being detected as spam, which may undermine confidence and make it difficult to debug or comply with regulations. This lack of interpretability is problematic from the point of view of practical deployments as interpretability rises in importance.

8.3. DEPLOYMENT CONSTRAINTS FOR REAL-TIME SMS FILTERING

Real-time (SMS) spam filtering systems should perform under tight latency, memory and energy constraints especially in mobile/edge-computing environments. Utilizing ensemble models in this environment may be problematic due to higher computational and storage needs. Worse, repeated updates to the model may be required in order to adapt to changes in spam patterns, thus making deployment and maintenance cumbersome. These limitations underscore the importance of appropriate model selection, optimization considerations and perhaps consideration of light-weight or hybrid ensemble models that balance merit performance while meeting operational real-time constraints.

9. FUTURE RESEARCH DIRECTIONS

9.1. HYBRID ENSEMBLES COMBINING DEEP AND TRADITIONAL MODELS

In the future, hybrid ensemble architectures that combine deep learning and classical machine learning can be studied. While deep models (transformer-based) can capture rich semantic and contextual information, traditional models could have an advantage in efficiency and robustness through handcrafted features. Integrating these methods as an ensemble can take advantage of their complementary properties, leading to enhanced performance for the SMS spam detection task in both increased accuracy and better generalization over various message types.

9.2. LIGHTWEIGHT ENSEMBLES FOR MOBILE AND REAL-TIME SYSTEMS

Since the SMS spam filtering will be generally applied to resource-limited devices, the demand for lightweight ensembles that trade-off performance and computational cost has been consistently increasing. Regarding future work, model compression or pruning, knowledge distillation and selective ensemble activation could be used to mitigate training and inference costs. Efficient ensembles for mobile and real-time systems will be designed to again allow for scalable low-latency spam-detection with only negligible performance degradation.

9.3. ADAPTIVE AND ONLINE ENSEMBLE LEARNING

Spam trends change quickly, and it is important to have models that can handle new and unseen kinds of spam. Adaptive and online ensemble learning approaches, which update base learners gradually when new data arrives at the system, might be an interesting line of research. These approaches are able to keep the high detection performance for a long time with low retraining cost. Adaptive ensembles for SMS spam classification can effectively deal with concept drifts and emerging spam patterns.

9.4. MULTILINGUAL AND CROSS-DOMAIN ENSEMBLE APPROACHES

Since the use of SMS communication is worldwide, multilingual and cross-domain spam classification problems must be solved in future work. Ensemble methods which combine models trained on diverse languages, domains or feature representations can enhance robustness and transferability. Transfer learning and domain adaptation can also be applied to the ensemble systems to extend their performance on low-resource languages and several communication environments, extending the applicability for SMS spam detection.

10. CONCLUSION

10.1. SUMMARY OF KEY FINDINGS

The work in this paper focused on the contribution of ensemble learning techniques to the improvement of performance for SMS spam classification. The scope of analysis included several ensemble methods: bagging, boosting, stacking and voting methods and numerous feature representations, as well as ways to appropriately deal with class imbalance. In a variety of scenarios, experiments showed that ensembles consistently improve the performance measured across common metrics, where recall and F1-score relating to the spam class are significantly better when compared with using single classifiers. It also demonstrated that multiple feature diversity and classifier heterogeneity contribute to obtaining robust and generalizable performance.

10.2. EFFECTIVENESS OF ENSEMBLE LEARNING IN IMPROVING SMS SPAM CLASSIFICATION ACCURACY

Ensemble learning was found to be very successful in enhancing SMS spam classification accuracy by decreasing bias and variance, and simultaneously integrating diverse decision patterns from multiple models. Bagging increased stability and noise robustness, boosting enhanced detection of difficult-to-classify spam messages, and stacking achieved the best performance by an optimal combination of classifiers. In the end, ensemble learning presented an effective solution to the issues inherent in SMS data that is short in length, subjected to noise and evolving spamming habits.

10.3. IMPLICATIONS FOR REAL-WORLD SPAM DETECTION SYSTEMS

The results have significant implications for real life SMS spam detection systems' design and deployment. The ensemble-based models are more robust and adaptive to adversarial spamming, and thus suitable for large-sized secure systems.

Although computational and interpretability issues persist, thoughtful choices and tuning of ensembles can accommodate the tradeoff between performance and practical deployment. Therefore, ensemble learning is a practical and efficient way to implement robust, accurate and scalable SMS spam filtering systems in the real world.

REFERENCES

[1] B. Narra, V. Kumar, Hari, Navya Vattikonda, A. K. Gupta, and Achuthananda Reddy Polu, "The Integration of Artificial Intelligence in Software Development: Trends, Tools, and Future Prospects," *Zenodo (CERN European Organization for Nuclear Research)*, Apr. 2024, doi: https://doi.org/10.5281/zenodo.17099155

[2] A. K. Gupta., "Leveraging deep learning models for intrusion detection systems for secure networks," Journal of Computer Science and Technology Studies, vol. 6, no. 2, pp. 199-208, 2024. Doi: https://doi.org/10.32996/jcsts.2024.6.2.22

[3] R. P. Achuthananda et al., "Evaluating machine learning approaches for personalized movie recommendations: A comprehensive analysis," Journal of Contemporary Education Theory & Artificial Intelligence, pp. 1-8, 2024.

[4] A. Reddy Polu et al., "Analyzing The Role of Analytics in Insurance Risk Management: A Systematic Review of Process Improvement and Business Agility," IJASTR-International Journal of Applied Science and Technical Research, vol. 1, no. 1, 2024, doi: https://doi.org/10.5281/zenodo.15209484.

[5] V. Tamilmani et al., "A Review of Cyber Threat Detection in Software-Defined and Virtualized Networking Infrastructures," International Journal of Technology, Management and Humanities, vol. 10, no. 4, pp. 136-146, 2024. Doi: https://doi.org/10.21590/ijtmh.10.04.15

[6] R. R. Kothamaram, "Predictive Analytics for Customer Retention in Telecommunications Using ML Techniques," vol. 1, no.1, pp. 45-58, 2024. Doi: https://doi.org/10.71141/30485037/V1I1P107

[7] A. A. S. Singh et al., "A Review on Model-Driven Development with a Focus on Microsoft PowerApps," International Journal of Humanities Science Innovations and Management Studies, vol. 1, no. 1, pp. 43-56, 2024. Doi: https://doi.org/10.64137/30508509/IJHSIMS-V1I1P105

[8] None Varun Bitkuri, None Raghuvaran Kendyala, None Jagan Kurma, N. Jaya, None Avinash Attipalli, and J. Enokkaren, "A Survey on Blockchain-Enabled ERP Systems for Secure Supply Chain Processes and Cloud Integration," International Journal of Technology Management and Humanities, vol. 10, no. 02, pp. 52–65, Jun. 2024, doi: https://doi.org/10.21590/ijtmh.2024100209.

[9] P. Waditwar, "AI for Bathsheba Syndrome: Ethical Implications and Preventative Strategies," Open Journal of Leadership, vol. 13, no. 03, pp. 321–341, 2024, doi: https://doi.org/10.4236/ojl.2024.133020.

[10] Jaya Vardhani Mamidala et al., "Machine Learning Approaches to Salary Prediction in Human Resource Payroll Systems," Journal of Computer Science and Technology Studies, vol. 7, no. 10, pp. 528–536, Oct. 2025, doi: https://doi.org/10.32996/jcsts.2025.7.10.52.

[11] Prajkta Waditwar, "Reimagining procurement payments: From transactional bottlenecks to strategic value creation," World Journal of Advanced Research and Reviews, vol. 28, no. 1, 588-598, 2025, doi: https://doi.org/10.30574/wjarr.2025.28.1.3480.

[12] J. Enokkaren, "Privacy Preservation in the Cloud: A Comprehensive Review of Encryption and Anonymization Methods," International Journal of Multidisciplinary on Science and Management, vol. 1, no. 1, pp. 35-44, 2024, doi: https://doi.org/ 10.71141/30485037/V1I1P106

[13] S. J. Enokkaren et al., "Artificial Intelligence (AI)-Based Advance Models for Proactive Payroll Fraud Detection and Prevention," International Journal of Machine Learning, AI & Data Science Evolution. Vol. 1, no. 1, pp. 1-11, 2024.

[14] M. S. V. Tyagadurgam, "AI-Powered Cybersecurity Risk Scoring for Financial Institutions Using Machine Learning Techniques," Journal of Artificial Intelligence & Cloud Computing, pp. 1–9, Dec. 2024, doi: https://doi.org/10.47363/jaicc/2024(3)452.

[15] P. Waditwar, "The Intersection of Strategic Sourcing and Artificial Intelligence: A Paradigm Shift for Modern Organizations," *Open Journal of Business and Management*, vol. 12, no. 06, pp. 4073–4085, 2024, doi: https://doi.org/10.4236/ojbm.2024.126204.

[16] D. Rajendran, Venkata Deepak Namburi, Vetrivelan Tamilmani, A. Arjun, V. Maniar, and Rami Reddy Kothamaram, "Middleware Architectures for Hybrid and Multi-cloud Environments: A Survey of Scalability and Security Approaches," *Asian Journal of Research in Computer Science*, vol. 19, no. 1, pp. 106–120, Jan. 2026, doi: https://doi.org/10.9734/ajrcos/2026/v19i1808.

[17] "De-Risking Returns: How AI Can Reinvent Big Tech's China-Tied Reverse Supply Chains," *Scirp.org*, 2026. https://www.scirp.org/journal/papercitationdetails?paperid=148354&JournalID=2447 (accessed Feb. 07, 2026).

[18] V. Maniar, R. R. Kothamaram, D. Rajendran, V. D. Namburi, V. Tamilmani, and A. A. S. Singh, "A Comprehensive Survey on Digital Transformation and Technology Adoption Across Small and Medium Enterprises," *European Journal of Applied Science, Engineering and Technology*, vol. 3, no. 6, pp. 238–250, Dec. 2025, doi: https://doi.org/10.59324/ejaset.2025.3(6).18.

[19] Vetrivelan Tamilmani, V. Maniar, A. Arjun, Rami Reddy Kothamaram, D. Rajendran, and Venkata Deepak Namburi, "Automated Cloud Migration Pipelines: Trends, Tools, and Best Practices – A Survey," *Journal of Computer Science and Technology Studies*, vol. 7, no. 11, pp. 121–134, 2025, doi: https://doi.org/10.32996/jcsts.2025.7.11.14.

[20] A. Attipalli, R. Kendyala, J. Kurma, J. V. Mamidala, V. Bitkuri, and S. J. Enokkaren, "Survey on Evolution of Java Web Technologies and Best Practices: from Servlets to Microservices," *Asian Journal of Research in Computer Science*, vol. 18, no. 11, pp. 172–187, Nov. 2025, doi: https://doi.org/10.9734/ajrcos/2025/v18i11786.

[21] Jaya Vardhani Mamidala *et al.*, "Explainable Machine Learning Models for Malware Identification in Modern Computing Systems," *European Journal of Applied Science Engineering and Technology*, vol. 3, no. 5, pp. 153–170, Oct. 2025, doi: https://doi.org/10.59324/ejaset.2025.3(5).13.

[22] Prajkta Waditwar, "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*, vol. 13, no. 02, pp. 140–155, Jan. 2025, doi: https://doi.org/10.4236/jdaip.2025.132009.

[23] Raghuvaran Kendyala, Jagan Kurma, Jaya Vardhani Mamidala, Sunil Jacob Enokkaren, Avinash Attipalli, and Varun Bitkuri, "Framework based on Machine Learning for Lung Cancer Prognosis with Big Data-Driven," *European Journal of Technology*, vol. 9, no. 1, pp. 68–85, Oct. 2025, doi: https://doi.org/10.47672/ejt.2787.

[24] A. B. Kakani, S. K. K. Nandiraju, S. K. Chundru, S. R. Vangala, R. M. Polam, and B. Kamarthapu, "Big Data and Predictive Analytics for Customer Retention: Exploring the Role of Machine Learning in E-Commerce," *SSRN Electronic Journal*, 2025, doi: https://doi.org/10.2139/ssrn.5515281.

[25] P. Kulkarni, T. Siddharth, S. Pillai, P. Pathak, V. N. Gangineni, and V. Yadav, "Cybersecurity Threats and Vulnerabilities - A Growing Challenge in Connected Vehicles," *Lecture Notes in Networks and Systems*, pp. 466–476, Nov. 2025, doi: https://doi.org/10.1007/978-3-032-03558-5_39.

[26] N. R. Vanaparthi, "INTELLIGENT FINANCE: HOW AI IS RESHAPING THE FUTURE OF FINANCIAL SERVICES," INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING AND TECHNOLOGY, vol. 16, no. 1, pp. 126–137, Jan. 2025, doi: https://doi.org/10.34218/ijcet_16_01_012.

[27] M. Sai, Venkataswamy Naidu Gangineni, Sriram Pabbineedi, Ajay Babu Kakani, K. Kireeti, and Sandeep Kumar Chundru, "Preventing Phishing Attacks Using Advanced Deep Learning Techniques for Cyber Threat Mitigation," *Journal of Data Analysis and Information Processing*, vol. 13, no. 03, pp. 314–330, Jan. 2025, doi: https://doi.org/10.4236/jdaip.2025.133020.

[28] M. Penmetsa, J. R. Bhumireddy, R. Chalasani, S. R. Vangala, R. M. Polam, and B. Kamarthapu, "Adversarial Machine Learning in Cybersecurity: A Review on Defending Against AI-Driven Attacks," *European Journal of Applied Science, Engineering and Technology*, vol. 3, no. 4, pp. 4–14, Jun. 2025, doi: https://doi.org/10.59324/ejaset.2025.3(4).01.

[29] Ram Mohan Polam, Bhavana Kamarthapu, Mitra Penmetsa, Jayakeshav Reddy Bhumireddy, R. Chalasani, and Srikanth Reddy Vangala, "Advanced Machine Learning for Robust Botnet Attack Detection in Evolving Threat Landscapes," *Asian Journal of Research in Computer Science*, vol. 18, no. 8, pp. 1–14, Aug. 2025, doi: https://doi.org/10.9734/ajrcos/2025/v18i8735.

[30] Bhavana Kamarthapu, Mitra Penmetsa, Jayakeshav Reddy Bhumireddy, R. Chalasani, Srikanth Reddy Vangala, and Ram Mohan Polam, "Data-Driven Detection of Network Threats Using Advanced Machine Learning Techniques for Cybersecurity," *International Journal of Applied Information Systems*, vol. 13, no. 1, pp. 37–44, Aug. 2025, doi: https://doi.org/10.5120/ijais2025452028.

[31] M. Penmetsa, J. R. Bhumireddy, R. Chalasani, S. R. Vangala, R. M. Polam, and B. Kamarthapu, "Effectiveness of Deep Learning Algorithms in Phishing Attack Detection for Cybersecurity Frameworks," *SSRN Electronic Journal*, 2025, doi: https://doi.org/10.2139/ssrn.5515385.

[32] Nandiraju Nandiraju, Sandeep Kumar Chundru, Srikanth Reddy Vangala, Ram Mohan Polam, Bhavana Kamarthapu, and Kakani Kakani, "Towards Early Forecast of Diabetes Mellitus via Machine Learning Systems in Healthcare," *European Journal of Technology*, vol. 9, no. 1, pp. 35–50, 2025, doi: https://doi.org/10.47672/ejt.2729.

[33] R. M. Polam, B. Kamarthapu, A. B. Kakani, S. K. K. Nandiraju, S. K. Chundru, and S. R. Vangala, "Predictive Modeling for Property Insurance Premium Estimation Using Machine Learning Algorithms," *SSRN Electronic Journal*, 2025, doi: https://doi.org/10.2139/ssrn.5515382.

[34] S. K. K. Nandiraju, S. K. Chundru, M. S. V. Tyagadurgam, V. N. Gangineni, S. Pabbineedi, and A. B. Kakani, "Enhancing Cybersecurity: Zero-Day Attack Detection in Network Traffic with Deep Learning Model," *Asian Journal of Research in Computer Science*, vol. 18, no. 7, pp. 262–273, Aug. 2025, doi: https://doi.org/10.9734/ajrcos/2025/v18i7734.

[35] P. Waditwar, "Agentic AI and sustainable procurement: Rethinking anti-corrosion strategies in oil and gas," *World Journal of Advanced Research and Reviews*, vol. 27, no. 3, pp. 1591–1598, Sep. 2025, doi: https://doi.org/10.30574/wjarr.2025.27.3.3298.

[36] Rahul Vadisetty, Anand Polamarasetti, V. Varadarajan, D. Kalla, and G. K. Ramanathan, "Cyber Warfare and AI Agents: Strengthening National Security Against Advanced Persistent Threats (APTs)," *Communications in computer and information science*, pp. 578–587, Oct. 2025, doi: https://doi.org/10.1007/978-3-032-07373-0_43.

[37] S. K. Chundru, M. S. V. Tyagadurgam, V. N. Gangineni, S. Pabbineedi, A. B. Kakani, and S. K. K. Nandiraju, "Analyzing and Predicting Anaemia with Advanced Machine Learning Techniques with Comparative Analysis," *International Journal of Applied Information Systems*, vol. 13, no. 1, pp. 28–36, Aug. 2025, doi: https://doi.org/10.5120/ijais2025452027.

[38] R. M. Polam, B. Kamarthapu, M. Penmetsa, J. R. Bhumireddy, R. Chalasani, and S. R. Vangala, "Advanced Machine Learning for Robust Botnet Attack Detection in Evolving Threat Landscapes," *SSRN Electronic Journal*, 2025, doi: https://doi.org/10.2139/ssrn.5515384.

[39] Bhavana Kamarthapu, Mitra Penmetsa, Jayakeshav Reddy Bhumireddy, R. Chalasani, Srikanth Reddy Vangala, and Ram Mohan Polam, "Data-Driven Detection of Network Threats Using Advanced Machine Learning Techniques for Cybersecurity," *International Journal of Applied Information Systems*, vol. 13, no. 1, pp. 37–44, Aug. 2025, doi: https://doi.org/10.5120/ijais2025452028.

[40] M. Penmetsa, J. R. Bhumireddy, R. Chalasani, S. R. Vangala, R. M. Polam, and B. Kamarthapu, "Effectiveness of Deep Learning Algorithms in Phishing Attack Detection for Cybersecurity Frameworks," *SSRN Electronic Journal*, 2025, doi: https://doi.org/10.2139/ssrn.5515385.

[41] Narasimha Rao Vanaparthi, “Why digital transformation in fintech requires mainframe modernization: A cost-benefit analysis,” *International Journal of Science and Research Archive*, vol. 14, no. 1, pp. 1052–1062, Jan. 2025, doi: https://doi.org/10.30574/ijsra.2025.14.1.0161.

[42] Ajay Babu Kakani, K. Kireeti, Sandeep Kumar Chundru, Srikanth Reddy Vangala, Ram Mohan Polam, and Bhavana Kamarthapu, “Leveraging NLP and Sentiment Analysis for ML-Based Fake News Detection with Big Data,” *SSRN Electronic Journal*, Jan. 2025, doi: https://doi.org/10.2139/ssrn.5515418.

[43] Prajkta Waditwar, “Quantum-Enhanced Travel Procurement: Hybrid Quantum–Classical Optimization for Enterprise Travel Management,” *World Journal of Advanced Engineering Technology and Sciences*, vol. 17, no. 3, pp. 375–386, Dec. 2025, doi: https://doi.org/10.30574/wjaets.2025.17.3.1572.

[44] Narasimha Rao Vanaparthi, “REGULATORY COMPLIANCE IN THE DIGITAL AGE: HOW MAINFRAME MODERNIZATION CAN SUPPORT FINANCIAL INSTITUTIONS,” vol. 8, no. 1, pp. 383–396, Jan. 2025, doi: https://doi.org/10.34218/IJRCAIT_08_01_033.

[45] P. Waditwar, “AI-Driven Procurement in Ayurveda and Ayurvedic Medicines & Treatments,” *Open Journal of Business and Management*, vol. 13, no. 03, pp. 1854–1879, 2025, doi: https://doi.org/10.4236/ojbm.2025.133096.

[46] N. Rao, “The Roadmap to Mainframe Modernization: Bridging Legacy Systems with the Cloud,” *International Journal of Scientific Research in Computer Science Engineering and Information Technology*, vol. 11, no. 1, pp. 125–133, Jan. 2025, doi: https://doi.org/10.32628/cseit25111214.

[47] D. Prabakar, N. Iskandarova, N. Iskandarova, D. Kalla, K. Kulimova, and D. Parmar, “Dynamic Resource Allocation in Cloud Computing Environments Using Hybrid Swarm Intelligence Algorithms,” *2025 International Conference on Networks and Cryptology (NETCRYPT)*, pp. 882–886, May 2025, doi: https://doi.org/10.1109/netcrypt65877.2025.11102314.

[48] Subuddi Nagaraju, Prashant Johri, Prakash Putta, D. Kalla, Sultonmakhmud Polvanov, and N. V. Patel, “Smart Routing in Urban Wireless Ad Hoc Networks Using Graph Attention Network-Based Decision Models,” pp. 212–216, May 2025, doi: https://doi.org/10.1109/netcrypt65877.2025.11102255.

[49] D. Kalla, A. S. Mohammed, V. N. Boddapati, N. Jiwani, and T. Kiruthiga, “Investigating the Impact of Heuristic Algorithms on Cyberthreat Detection,” *2024 2nd International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT)*, pp. 450–455, Nov. 2024, doi: https://doi.org/10.1109/icaiccit64383.2024.10912106.

[50] Rahul Vadisetty, Anand Polamarasetti, and D. Kalla, “Automated AI-Driven Phishing Detection and Countermeasures for Zero-Day Phishing Attacks,” *Lecture notes in networks and systems*, pp. 285–303, Jan. 2026, doi: https://doi.org/10.1007/978-981-96-8632-2_16.

[51] Preeti Nagrath, I. Saini, M. Zeeshan, Komal, Komal, and D. Kalla, “Predicting Mental Health Disorders with Variational Autoencoders,” *Lecture notes in networks and systems*, pp. 38–51, Oct. 2025, doi: https://doi.org/10.1007/978-3-032-03751-0_4.

[52] World Bank, Small and medium enterprises (SMEs) finance, World Bank Group, 2020. [Online]. Available: https://www.worldbank.org/en/topic/smefinance

[53] 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.

[54] Zaenal Aripin, B. Susanto, and R. Agusiady, “DIGITAL TRANSFORMATION IN INDONESIAN SMES: DRIVERS, BARRIERS, AND PERFORMANCE OUTCOMES,” *Journal of Economics, Accounting, Business, Management, Engineering and Society*, vol. 1, no. 11, pp. 1–8, 2024, Accessed: Feb. 07, 2026. [Online]. Available: https://kisainstitute.com/index.php/kisainstitute/article/view/43

[55] J. A. Restrepo-Morales, J. A. Ararat-Herrera, D. A. López-Cadavid, and A. Camacho-Vargas, “Breaking the digitalization barrier for SMEs: a fuzzy logic approach to overcoming challenges in business transformation,” *Journal of Innovation and Entrepreneurship*, vol. 13, no. 1, Nov. 2024, doi: https://doi.org/10.1186/s13731-024-00429-w.

[56] G. H. Sagala and D. Őri, “Toward SMEs digital transformation success: a systematic literature review,” *Information Systems and e-Business Management*, vol. 22, Jul. 2024, doi: https://doi.org/10.1007/s10257-024-00682-2.

[57] OECD, “The Digital Transformation of SMEs,” *OECD*, Feb. 03, 2021. https://www.oecd.org/en/publications/the-digital-transformation-of-smes_bdb9256a-en.html

[58] T. Yuwono, A. Suroso, and W. Novandari, “Information and communication technology in SMEs: a systematic literature review,” *Journal of innovation and entrepreneurship*, vol. 13, no. 1, May 2024, doi: https://doi.org/10.1186/s13731-024-00392-6.

[59] T. Justy, E. Pellegrin-Boucher, D. Lescop, J. Granata, and S. Gupta, “On the edge of Big Data: Drivers and barriers to data analytics adoption in SMEs,” *Technovation*, vol. 127, p. 102850, Sep. 2023, doi: https://doi.org/10.1016/j.technovation.2023.102850.

[60] Saleh Alarifi, “SMEs’ Resilience Toward Cyberattacks in Saudi Arabia: A Review Paper,” *International Journal of Scientific Research and Management (IJSRM)*, vol. 13, no. 12, pp. 10162–10175, Jan. 2025, doi: https://doi.org/10.18535/ijsrm/v13i12.em13.

[61] G. B. Thapa and Dr. S. Thapaliya, “Cybersecurity Challenges in Small and Medium Enterprises (SMES) in Nepal,” *International Journal of Multidisciplinary and Innovative Research*, vol. 02, no. 06, Jun. 2025, doi: https://doi.org/10.58806/ijmir.2025.v2i6n05.

[62] S. Chaudhary, Vasileios Gkioulos, and D. W. Goodman, “Cybersecurity Awareness for Small and Medium-Sized Enterprises (SMEs): Availability and Scope of Free and Inexpensive Awareness Resources,” *Lecture Notes in Computer Science*, pp. 97–115, Jan. 2023, doi: https://doi.org/10.1007/978-3-031-25460-4_6.

[63] M. Wallang, M. D. K. Shariffuddin, and M. Mokhtar, "CYBER SECURITY IN SMALL AND MEDIUM ENTERPRISES (SMEs)," *Journal of Governance and Development (JGD)*, vol. 18, no. 1, pp. 75–87, Dec. 2022, doi: https://doi.org/10.32890/jgd2022.18.1.5.

[64] Carlos Rombaldo Junior, I. Becker, and S. D. Johnson, "Unaware, Unfunded and Uneducated: A Systematic Review of SME Cybersecurity," *arXiv (Cornell University)*, Sep. 2023, doi: https://doi.org/10.48550/arxiv.2309.17186.

[65] A. Chidukwani, S. Zander, and P. Koutsakis, "Cybersecurity Preparedness of Small-to-Medium Businesses: A Western Australia Study with Broader Implications," Computers & Security, vol. 145, pp. 104026–104026, Jul. 2024, doi: https://doi.org/10.1016/j.cose.2024.104026.

[66] C. K. Tan, S. K. Khan, and U. F. Khattak, "Exploration of The Impact of Cyber Situational Awareness On Small and Medium Enterprises (SMEs) in Malaysia," *Journal of Informatics and Web Engineering*, vol. 4, no. 1, pp. 292–306, Feb. 2025, doi: https://doi.org/10.33093/jiwe.2025.4.1.21

[67] M. Awan, A. Alam, and M. Kamran, "Cybersecurity Challenges in Small and Medium Enterprises: A Scoping Review," *Journal of Cyber Security and Risk Auditing*, vol. 2025, no. 3, pp. 89–102, Jul. 2025, doi: https://doi.org/10.63180/jcsra.thestap.2025.3.7