International Journal of Emerging Trends in Computer Science and Information Technology
ISSN: 3050-9246 | https://doi.org/10.63282/3050-9246.IJETCSIT-V7I1P108
Eureka Vision Publication | Volume 7, Issue 1, 57-62, 2026

Original Article

Handling Class Imbalance in SMS Spam Datasets Using Advanced Sampling Techniques

Vempalli Mopuru Rakesh Reddy
Systems Engineer, Tata Consultancy Services.

Received On: 01/12/2025 ***Revised On: 02/01/2026*** ***Accepted On: 10/01/2026*** ***Published On: 23/01/2026***

Abstract - The prevalence of SMS spam poses significant challenges for automated messaging systems, and effective detection is often hindered by the inherent class imbalance in SMS datasets, where legitimate messages vastly outnumber spam messages. This study investigates the impact of advanced sampling techniques on improving classification performance in imbalanced SMS datasets. Techniques such as Synthetic Minority Oversampling Technique (SMOTE), Adaptive Synthetic Sampling (ADASYN), and ensemble-based resampling methods are evaluated for their effectiveness in balancing the dataset and enhancing the predictive accuracy of machine learning classifiers. Experimental results demonstrate that applying these advanced sampling strategies significantly improves spam detection rates while reducing false positives. The findings provide valuable insights for developing robust SMS spam filters and highlight the importance of addressing class imbalance in real-world text classification problems.

Keywords - Sms Spam Detection, Class Imbalance, Imbalanced Datasets, Sampling Techniques, Smote, Adasyn, Ensemble Resampling, Text Classification, Machine Learning, Spam Filtering, Predictive Accuracy, False Positive Reduction.

1. Introduction

The rapid growth of mobile communication has led to a significant increase in unsolicited SMS messages, commonly referred to as SMS spam. Efficient detection and filtering of such messages are crucial to maintaining user privacy, ensuring communication security, and enhancing overall user experience. SMS spam classification, therefore, has emerged as a critical area of research within natural language processing (NLP) and machine learning. A persistent challenge in SMS spam detection is the class imbalance in datasets, where legitimate messages (ham) significantly outnumber spam messages. This imbalance can severely affect the performance of machine learning models, often causing classifiers to be biased toward the majority class and resulting in high false negative rates for spam detection. Addressing class imbalance is essential to improve predictive accuracy and ensure reliable spam filtering. This study focuses on exploring advanced sampling techniques, including oversampling, undersampling, and hybrid methods, to mitigate the effects of imbalance and enhance the performance of SMS spam classification models.

2. Understanding Class Imbalance in SMS Spam Datasets

SMS spam datasets typically exhibit a highly skewed distribution between spam and legitimate messages (ham). In most real-world datasets, spam messages constitute a small fraction—often ranging from 10% to 20%—while the majority of messages are legitimate. For example, widely used datasets such as the UCI SMS Spam Collection contain approximately 13% spam and 87% ham messages, highlighting the extent of class imbalance. This imbalance has significant implications for machine learning model training. Models trained on such datasets tend to develop a bias toward the majority class, favoring the prediction of legitimate messages and often failing to correctly identify spam. Consequently, the recall for the minority class (spam detection) is significantly reduced, which can undermine the effectiveness of spam filtering systems. Additionally, class imbalance can lead to misleading evaluation metrics. Accuracy, for instance, may appear high even when the model fails to correctly classify spam messages, as the majority of predictions are dominated by the ham class. Therefore, addressing class imbalance is crucial to developing robust and reliable SMS spam classifiers, ensuring both high precision and recall for the minority class.

3. Overview of Sampling Techniques

Class imbalance in SMS spam datasets can be addressed using sampling techniques, which modify the distribution of the dataset to improve model learning for the minority class. These techniques can be broadly categorized into basic sampling methods and advanced sampling techniques**.**

3.1. Basic Sampling Methods

- Random Oversampling: This approach increases the number of minority class (spam) instances by randomly duplicating existing examples. While simple and effective for balancing classes, it can lead to overfitting**,** as the model may memorize repeated examples rather than generalize patterns.
- Random Undersampling: This method reduces the number of majority class (ham) instances to match the minority class. While it mitigates class imbalance, it can discard valuable information, potentially degrading model performance.
- Limitations: Both basic methods are straightforward but may fail to capture the underlying complexity of

the data. Oversampling risks redundancy and overfitting, while undersampling may lead to loss of critical information.

3.2. Advanced Sampling Techniques

- SMOTE (Synthetic Minority Over-sampling Technique): SMOTE generates synthetic samples for the minority class by interpolating between existing minority instances. This approach reduces overfitting and enhances the classifier's ability to generalize. Variants such as Borderline-SMOTE focus on minority samples near class boundaries, and KMeans-SMOTE leverages clustering to create synthetic samples within dense regions, preserving data structure.
- ADASYN (Adaptive Synthetic Sampling): ADASYN is an extension of SMOTE that generates more synthetic samples for harder-to-learn minority examples, focusing on regions where the model struggles. This adaptive approach improves minority class recognition and overall recall.
- Cluster-based Oversampling: By dividing data into clusters before generating synthetic samples, this technique ensures that the underlying data distribution is preserved, reducing the risk of generating unrealistic examples.
- Hybrid Methods: Hybrid sampling combines oversampling and undersampling to leverage the strengths of both approaches. For example, oversampling the minority class while selectively undersampling the majority class can effectively balance SMS spam datasets without sacrificing information or introducing redundancy.
- Advanced sampling techniques have shown significant improvements in spam detection performance, particularly in enhancing recall and F1-score for the minority class while maintaining model robustness.

4. Feature Engineering and Preprocessing Considerations

Effective SMS spam classification relies not only on handling class imbalance but also on careful feature engineering and text preprocessing, which directly influence the performance of sampling techniques.

4.1. Text Preprocessing Impact on Sampling Efficacy:

Preprocessing steps such as tokenization, stopword removal, and stemming/lemmatization help standardize and clean SMS text data. These steps reduce noise and ensure that the features used for training reflect meaningful patterns. When combined with sampling techniques, preprocessing can improve the generation of synthetic samples (as in SMOTE or ADASYN) by providing more consistent and representative feature vectors. Poor preprocessing, on the other hand, can exacerbate class imbalance issues by creating noisy or redundant features that mislead the classifier.

4.2. Handling High-Dimensional Sparse Feature Spaces:

SMS messages are often transformed into high-dimensional feature spaces using techniques like TF-IDF **or** word embeddings. While these representations capture semantic and contextual information, they can also lead to sparsity, which challenges both model training and synthetic sample generation. Advanced sampling techniques must account for these high-dimensional spaces to create realistic synthetic examples that improve minority class learning without introducing noise.

4.3. Balancing Preprocessing with Sampling to Avoid Overfitting:

Oversampling or generating synthetic samples on poorly preprocessed data can lead to overfitting**,** where models memorize synthetic examples rather than generalizing to unseen data. Therefore, it is critical to carefully balance preprocessing steps with sampling strategies. Proper feature engineering ensures that the synthetic samples accurately represent the minority class distribution, enhancing model robustness and predictive accuracy for SMS spam detection.

5. Performance Evaluation Metrics

Evaluating the performance of SMS spam classifiers in imbalanced datasets requires careful consideration of appropriate metrics. Traditional accuracy can be misleading in such scenarios, as a model that predicts all messages as the majority class (ham) may achieve high accuracy despite failing to detect spam effectively. Therefore, more informative metrics are necessary to assess classifier performance on both classes.

5.1. Precision, Recall, and F1-score

- Precision measures the proportion of correctly predicted spam messages out of all messages classified as spam, reflecting the classifier's ability to avoid false positives.
- Recall (or sensitivity) measures the proportion of actual spam messages correctly identified, highlighting the model's effectiveness in detecting the minority class.
- F1-score is the harmonic mean of precision and recall, providing a balanced metric that accounts for both false positives and false negatives, making it particularly useful in imbalanced datasets.

5.2. Area under the Precision-Recall Curve (AUPRC):

The AUPRC is especially suitable for imbalanced data, as it focuses on the performance of the classifier on the minority class (spam) rather than the majority class, offering a more realistic evaluation than ROC-AUC in skewed datasets.

5.3. Matthews Correlation Coefficient (MCC):

MCC provides a single value that considers true positives, true negatives, false positives, and false negatives, offering a balanced measure even when class distributions are highly uneven. It is widely regarded as one of the most reliable metrics for imbalanced classification tasks. Using these metrics together provides a comprehensive view of model performance**,** ensuring that improvements from sampling techniques and feature engineering are correctly reflected in the evaluation of SMS spam classifiers.

6. Experimental Setup

This study evaluates the effectiveness of advanced sampling techniques in handling class imbalance for SMS spam classification using rigorous experimental procedures.

- Dataset Description**:** Experiments are conducted on widely used SMS spam datasets, such as the UCI SMS Spam Collection**,** which contains approximately 5,574 messages, of which 13% are spam and 87% are legitimate (ham). The dataset represents a typical imbalanced scenario in real-world SMS filtering applications.
- **Baseline Models:** To establish a reference point, baseline classifiers—including Logistic Regression, Decision Trees, Random Forests, and Support Vector Machines (SVM)**—**are trained on the original imbalanced dataset without any sampling. These models provide a benchmark for evaluating the impact of sampling strategies.

6.1. Comparison with Sampling Strategies

The study compares multiple sampling approaches:

- Basic methods: Random oversampling and undersampling.
- Advanced methods: SMOTE, ADASYN, Borderline-SMOTE, KMeans-SMOTE, and hybrid oversampling–undersampling techniques.

Each sampling method is applied prior to model training, and its impact on minority class detection (spam) is analyzed.

6.2. Cross-Validation Techniques

To ensure robust evaluation and minimize bias due to dataset splits, k-fold cross-validation is employed. In this approach, the dataset is divided into k subsets, with each subset used once as the validation set while the remaining k–1 subsets are used for training. Performance metrics—including precision, recall, F1-score, AUPRC, and MCC—are averaged across folds to provide reliable estimates of model effectiveness.

7. Analysis and Discussion

The experimental results highlight the impact of various sampling techniques on the performance of SMS spam classifiers in imbalanced datasets.

7.1. Comparative Performance of Sampling Techniques

Advanced oversampling methods such as SMOTE, ADASYN, and KMeans-SMOTE consistently improve recall and F1-score for the minority class (spam) compared to baseline models and basic sampling techniques. While random oversampling also enhances minority class detection, it often leads to overfitting due to duplicate instances. Random undersampling reduces training data, which can negatively impact overall model performance but may improve minority class recognition in some cases.

7.2. Trade-offs Between Oversampling and Undersampling

Oversampling techniques increase the representation of the minority class without discarding majority class data, preserving information but sometimes introducing synthetic noise. Undersampling, in contrast, reduces the majority class to achieve balance but risks loss of valuable information, potentially limiting the classifier's generalization ability. Hybrid approaches, which combine oversampling of the minority class with selective undersampling of the majority class, often provide the best balance**,** improving minority class detection while maintaining overall model accuracy.

7.3. Challenges in Generating Realistic Synthetic Samples

Creating synthetic SMS messages is inherently challenging due to the short, informal, and context-specific nature of SMS text. Poorly generated synthetic samples can introduce noise or unrealistic patterns, leading to model overfitting or misclassification. Clustering-based and adaptive methods like KMeans-SMOTE and ADASYN help mitigate these issues by generating samples in meaningful regions of the feature space, preserving the underlying distribution.

7.4. Recommendations for Practitioners

The choice of sampling technique should consider dataset size, imbalance ratio, and feature representation**.** For highly imbalanced datasets with sufficient minority examples, advanced oversampling (SMOTE, ADASYN) is recommended. For smaller datasets, hybrid methods can prevent overfitting while maintaining minority class performance. Additionally, careful integration with preprocessing and feature engineering is essential to ensure synthetic samples accurately reflect real-world SMS patterns.

Overall, the analysis demonstrates that handling class imbalance with advanced sampling techniques is critical for robust SMS spam detection, particularly when combined with appropriate preprocessing and evaluation metrics.

8. Future Directions

While advanced sampling techniques have significantly improved SMS spam classification, several opportunities remain for enhancing performance and adaptability in real-world applications:

8.1. Integration of Deep Learning with Sampling Techniques

Deep learning models, such as LSTM, CNN, and Transformer-based architectures, can capture complex patterns in SMS text. Combining these models with advanced sampling strategies could improve minority class detection, particularly for nuanced or context-dependent spam messages.

8.2. Use of Ensemble Learning with Sampling

Ensemble methods, such as Random Forests**,** Gradient Boosting, or stacking multiple classifiers, can further enhance robustness. When integrated with sampling techniques, ensembles can mitigate the weaknesses of individual models and reduce variance, improving overall spam detection performance.

8.3. Domain-Adaptive Sampling for Multilingual SMS Datasets

Many existing studies focus on English-language datasets, yet SMS traffic is increasingly multilingual. Developing domain-adaptive sampling approaches that account for linguistic and cultural variations can improve classifier generalization across languages and regions.

8.4. Exploration of Cost-Sensitive Learning

As an alternative or complement to sampling, cost-sensitive learning assigns higher penalties to misclassifying the minority class (spam), allowing models to learn effectively from imbalanced datasets without generating synthetic samples. Integrating cost-sensitive approaches with existing sampling techniques could further improve detection accuracy and recall. These future directions highlight the potential for more sophisticated, adaptive, and scalable solutions for SMS spam detection, ensuring that classifiers remain effective across diverse datasets, languages, and emerging spam strategies.

9. Conclusion

This study highlights the critical role of handling class imbalance in SMS spam datasets for building effective and reliable spam detection systems. Experimental analysis demonstrates that advanced sampling techniques—including SMOTE, ADASYN, cluster-based oversampling, and hybrid methods—significantly improve the detection of minority class instances, enhancing recall and F1-score while mitigating the limitations of baseline models and basic sampling approaches. Careful integration of feature engineering, preprocessing, and sampling ensures that synthetic samples accurately reflect real-world SMS patterns, reducing the risk of overfitting and improving overall model robustness. Evaluation using metrics beyond accuracy, such as precision, recall, F1-score, AUPRC, and MCC, provides a comprehensive understanding of model performance in imbalanced scenarios. The findings underscore the importance of addressing class imbalance for deploying reliable SMS spam filters in real-world applications. By adopting appropriate sampling strategies and robust evaluation practices, practitioners can improve spam detection rates, protect users from unwanted messages, and enhance the overall effectiveness of automated messaging systems.

References

[1] 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.

[2] 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.

[3] 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.

[4] 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

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] Nandiraju, S. K. K., & Chundru, S. K. Enhancing Cybersecurity: Zero-Day.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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

[34] 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

[35] 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

[36] 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.

[37] 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.

[38] 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.

[39] 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.

[40] 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.

[41] Oliveira, T., & Martins, M. F. (2011). Literature review of information technology adoption models at firm level. The Electronic Journal of Information Systems Evaluation, 14(1), 110–121.

[42] Rogers, E. M. (2003). Diffusion of innovations (5th ed.). Free Press.

[43] Schumacher, A., Erol, S., & Sihn, W. (2016). A maturity model for assessing Industry 4.0 readiness and maturity of manufacturing enterprises. Procedia CIRP, 52, 161–166.

[44] Tornatzky, L. G., & Fleischer, M. (1990). The processes of technological innovation. Lexington Books.

[45] Vial, G. (2019). Understanding digital transformation: A review and a research agenda. The Journal of Strategic Information Systems, 28(2), 118–144.

[46] Pol, N. U. R., Ghezzi, A., Balocco, R., & Rangone, A. (2023). Understanding SMEs digitalization: A literature review of maturity models. Proceedings of the European Conference on Innovation and Entrepreneurship. DOI:10.34190/ecie.18.2.1823

[47] Silva, M., Mamede, R., & Santos, P. (2024). A new proposed model to assess the digital organizational readiness to maximize the results of the digital transformation in SMEs. Journal of Innovation & Knowledge.

[48] (Silva, Mamede & Santos). (2024). EconStor. A new proposed model to assess the digital organizational readiness to maximize the results of the digital transformation in SMEs.

[49] Soomro, M. A., Hanafiah, M. H. B., & Abdullah, N. L. (2020). Digital readiness models: A systematic literature review. Journal/Conference Publication.

[50] Williams, C. A., Schallmo, D., Lang, K., & Boardman, L. (2019). Digital maturity models for small and medium-sized enterprises: A systematic literature review. ISPIM Innovation Conference Proceedings.

[51] (Author Unknown). (2024). Assessment of organizational readiness for digital transformation in SMEs. Procedia Computer Science, 204, 362–369.

[52] Various Authors. (2024). Toward SMEs digital transformation success: A systematic literature review. Information Systems and e-Business Management, 22, 667–719.

[53] Haryanti, et al. (2025). Sustainable digital transformation roadmaps for SMEs: A systematic literature review. Sustainability, 16(19), 8551.

[54] (If accessible) Egodawele, M., Sedera, D., & Bui, V. (2022). A systematic review of digital transformation literature (2013–2021) and the development of an overarching model to guide future research. ArXiv Preprint.

[55] Gonzalez-Varona, J. M., Lopez-Paredes, A., Poza, D., & Acebes, F. (2024).