© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1719602

A Review on Detection of Advance Fee Fraud Using Natural Language Processing

ABIMAJE FRIDAY[1], DR. VICTOR KULUGH[2], DR. THOMAS A. GAGA[3], DR. IBRAHIM A. YAKUBU[4], EMMANUEL DAUDA[5], MAIKORI J. E[6]
[1,2,3,4,5,6]*Department of Computer Science, Faculty of Computing, Bingham University, Karu, Nigeria*

Abstract- ***Advance Fee Fraud (AFF) remains one of the most persistent and financially damaging forms of cyber-enabled crime, with global losses running into hundreds of millions of dollars annually and a demonstrated capacity to adapt its linguistic strategies to evade automated filters. This review synthesizes the current state of research on the automatic detection of AFF messages using Natural Language Processing (NLP), with particular emphasis on the Bag-of-Words (BoW) model and the supervised machine learning classifiers commonly paired with it. Drawing on theoretical perspectives from information theory, social engineering, linguistics, and machine learning, the review traces how AFF messages have evolved from repetitive, template-driven text to highly personalized, emotionally calibrated narratives, and how detection research has responded with richer preprocessing, feature engineering, and classification strategies. It also situates BoW-based approaches against more recent transformer-based, multimodal, and adversarially robust methods, arguing that BoW retains practical relevance where interpretability, computational efficiency, and small-data performance are priorities. Three persistent gaps are identified: the limited integration of psychologically and linguistically informed features into BoW pipelines, the scarcity of cross-lingual evaluation, and the absence of systematic adversarial robustness testing for AFF-specific systems. The review concludes by outlining a theoretically grounded, lightweight BoW-based framework as a practical direction for future detection systems, particularly in resource-constrained deployment settings.***

Keywords: Advance Fee Fraud, Natural Language Processing, Bag-of-Words, Machine Learning, Text Classification, Social Engineering, Fraud Detection

I. INTRODUCTION

Advance Fee Fraud (AFF) is a cyber-enabled crime in which victims are persuaded to make upfront payments in exchange for the false promise of a larger financial reward, such as an inheritance, lottery winning, contract benefit, or business opportunity. Long associated with the so-called "419" email scams originating from Nigeria, AFF has since evolved into a transnational, multi-platform criminal enterprise that spans email, social media, instant messaging, and cryptocurrency ecosystems. The scale of the problem is considerable: reported losses in the United Kingdom alone reached £59.4 million in 2020, while the United States Federal Bureau of Investigation's Internet Crime Complaint Centre recorded over $100 million in AFF-related losses in 2019. The COVID-19 pandemic further accelerated digital fraud, with global cybercrime rising sharply as social and economic life moved online.

What makes AFF particularly difficult to counter is not merely its scale but its linguistic adaptability. Where early scam templates were repetitive and easily flagged by keyword filters, contemporary AFF messages increasingly employ personalized storytelling, pseudo-formal register, and emotionally calibrated appeals designed to establish what researchers have termed "contextual plausibility." This evolution has driven a parallel evolution in detection research, from simple rule-based filters to Natural Language Processing (NLP) pipelines built on the Bag-of-Words (BoW) model, and more recently to deep learning and transformer-based architectures.

This review synthesizes recent literature on AFF detection with a specific focus on NLP and the BoW model. It examines the conceptual and empirical evolution of AFF, the theoretical frameworks that underpin detection research, the methodological landscape of text classification techniques, and the specific empirical evidence supporting BoW-based approaches. The review closes by identifying gaps in the current literature and proposing a direction for a

lightweight, interpretable, and theoretically informed BoW-based detection framework suited to real-time and resource-constrained deployment contexts.

The Evolving Nature of Advance Fee Fraud

Recent corpus-based studies document a marked shift in the linguistic sophistication of AFF messages. Olabode and Adebayo (2021) analyzed over 5,000 AFF emails collected between 2018 and 2021 and found that contemporary messages exhibit substantially more emotional manipulation markers and narrative elaboration than earlier scam templates from the 2010s, reflecting a growing understanding of victim psychology among fraudsters. Ibrahim and Yusuf (2022) examined a corpus of 3,200 messaging-platform scam messages and found that a majority adopted formal greeting structures associated with legitimate business correspondence, while more than half incorporated personalized details such as names, locations, and professions harvested from social media profiles.

Nwachukwu and Okafor's (2021) longitudinal analysis of 8,000 scam emails spanning 2015 to 2021 identifies three broad evolutionary phases in AFF writing: an early phase (2015-2017) of repetitive, poorly written templates; a middle phase (2018-2019) marked by improved grammar and standardized formatting; and a recent phase (2020-2021) defined by personalized storytelling and adaptive language. The authors argue that this trajectory correlates with the growing adoption of machine learning-based detection systems, suggesting an adversarial dynamic in which fraudsters strategically adapt their language to evade automated filters.

This evolution can be understood through an information-theoretic lens. Building on Shannon's (1948) foundational theory of communication, Chen and Zhang (2020) found that traditional AFF templates exhibited high redundancy, making them statistically detectable through entropy-based measures, whereas contemporary personalized messages show markedly higher lexical entropy and lower compression ratios, approaching the statistical properties of legitimate correspondence. Mustapha and Abdullahi (2022) further showed that successful AFF messages strategically alternate high-information claims with low-information emotional appeals, a pattern consistent with Kahneman's (2011) dual-process account of analytical versus intuitive reasoning. At the same time, Ogunleye and Adebayo (2023) found that fraudulent messages continue to exhibit elevated repetition of specific persuasive terms, such as "urgent," "confidential," and "guaranteed," a phenomenon they term persuasive lexical clustering, even as overall lexical diversity increases. This persistent statistical signature is a key reason why frequency-based representations such as BoW remain viable detection tools despite the narrative sophistication of modern scams.

AFF has also diversified across platforms and narrative types. Adedoyin and Balogun's (2021) taxonomy of 10,000 scam emails classifies AFF narratives into inheritance and pending-funds scams, lottery and prize scams, romance-based scams, investment and business-opportunity scams, and job-offer scams, with romance-based scams showing the highest lexical diversity and the lowest template repetition, making them the most challenging for BoW-based systems. Adewale and Ogunleye (2023) documented a decline in email-based scams relative to messaging-platform scams over the past several years, with messaging-platform messages typically shorter, more informal, and more likely to include multimedia elements. Okafor and Nwachukwu (2022) further show that messages targeting Western recipients tend to be more formal and elaborate, while those targeting African recipients favor direct, transactional language and local cultural references, underscoring a segmentation strategy that increases perceived relevance within specific demographic groups. Ibrahim and Bello (2023) add that messages combining higher entropy variance with moderate redundancy achieve substantially higher conversion rates than simple template-based approaches.

Beyond linguistic evolution, AFF has professionalized into a criminal industry with specialized roles, including lead generators who harvest victim information, social engineers who craft persuasive narratives, and money mules who facilitate fund transfers. Platform diversification has also enabled coordinated, cross-channel campaigns: Singh, Thompson, and Okonkovo (2023) developed a

graph-based, cross-platform detection framework showing that analyzing temporal patterns and content evolution across platforms improved detection accuracy by 18% over single-platform approaches, consistent with reports that social media increasingly serves as the initial point of contact before fraud moves to messaging apps or email.

Natural Language Processing in Fraud Detection

NLP underpins virtually all automated approaches to fraud detection, providing the tokenization, normalization, and vectorization steps necessary to convert unstructured text into features suitable for machine learning (Liu, Chen, & Zhang, 2020; Sharma & Chen, 2023). Liu, Chen, and Zhang (2020) evaluated 50,000 emails and found that character-level features outperformed word-level features (94.3% vs. 89.7% accuracy) for detecting obfuscated phishing attempts, owing to their robustness to misspellings and character substitutions, while hybrid architectures combining character- and word-level features achieved the strongest overall performance (96.1%).

Sharma and Chen (2023) developed a deception-detection framework integrating sentiment analysis, emotion detection, and discourse-structure analysis, achieving 91.8% accuracy and showing that fraudulent messages exhibit greater sentiment volatility than legitimate communications, with rapid shifts between positive and negative emotional register serving to destabilize victims' analytical reasoning while sustaining emotional engagement. Adewale and Olabode (2021) found that domain-adapted stop-word lists and lemmatization improved detection accuracy by 7.2% on average relative to generic stemming approaches, and that selectively preserving suspicious punctuation patterns, such as excessive exclamation marks or irregular capitalization, retained useful linguistic markers of deception that standard preprocessing pipelines typically discard.

Okonkwo and Nwachukwu (2022) compared BoW, TF-IDF, Word2Vec, GloVe, and FastText representations and found that TF-IDF-weighted BoW achieved the best balance of accuracy (92.3%) and computational efficiency, even though Word2Vec achieved marginally higher accuracy (93.7%) at roughly twelve times the processing cost. Chen and Liu (2022) found that transformer-based models such as BERT achieve the highest raw accuracy (97.2%), but their computational demands, roughly 80 GPU-hours for training and 200ms per inference, limit their practicality for real-time, resource-constrained deployment, motivating a knowledge-distillation approach that recovered 96.1% accuracy at a fraction of the computational cost. Adesina and Bello (2023) further show that models trained on one platform or region can suffer accuracy degradation of 18-22% when applied elsewhere, reflecting platform-specific linguistic norms and culturally specific persuasion strategies; their proposed domain-adaptation framework, incorporating platform-specific features alongside universal deception indicators, improved cross-platform transfer performance by 16 percentage points. Ogunleye and Adebayo (2024) similarly found that standard feature-extraction techniques are vulnerable to synonym substitution (12-18% accuracy loss) and character-level perturbation (8-15% loss), and showed that adversarial training and character-level feature robustness improved resilience by 24 percentage points.

The Bag-of-Words Model: Enduring Relevance

The Bag-of-Words model represents text as an unordered collection of word occurrences, discarding grammar and word order in favor of computational simplicity (Harris & Wu, 2021; Zhang & Zhou, 2022). Despite, or perhaps because of, this simplicity, BoW remains competitive with far more sophisticated representations across a range of text classification tasks. Harris and Wu (2021) evaluated over 100,000 messages across 15 public datasets and found that TF-IDF-weighted BoW achieved an average accuracy of 91.7%, within 1.2 percentage points of BERT-based representations (92.9%), while requiring roughly 1/50th of the computational resources and 1/100th of the training time.

This efficiency advantage is particularly pronounced on small- to medium-sized datasets. Zhang and Zhou (2022) found that BoW combined with ensemble classifiers achieved 94.2% accuracy on a corpus of 20,000 financial-fraud messages, significantly

outperforming deep learning approaches on datasets under 50,000 messages, which they attribute to deep learning's tendency to overfit on limited samples. Chen and Wang (2021) found that BoW performs best on medium-length messages (93.7% accuracy on 100–500-word messages versus 88.4% on very short messages and 89.2% on very long ones), with length-normalization techniques recovering several percentage points of accuracy at both extremes. Ogunleye and Adebayo (2023) further found that BoW-based explanations were rated 3.8 times more useful and 4.2 times more trustworthy than deep-learning attention visualizations by fraud analysts, a property with direct value in regulated financial contexts where explanation requirements favor auditable models.

Nonetheless, BoW is not without limitations. Adewale and Yusuf (2022) found that BoW models trained on earlier periods show consistent performance degradation on later messages (about 3.2% accuracy loss per year) due to vocabulary shift, a problem that incremental vocabulary updating reduced to roughly 1.1% per year. Okonkwo and Abdullahi (2022) found that mutual-information-based feature selection could reduce the feature space from 10,000 to 2,000 features while improving accuracy by 1.4% and cutting training time by 78%. Eze and Okafor (2024) documented a marked cross-lingual performance gap, with BoW achieving 92.3% accuracy on English but only 81.7% on Mandarin, a gap that character n-gram BoW variants partially closed (85-88% accuracy) by bypassing language-specific word-segmentation challenges. Sharma and Patel (2023) showed that combining BoW-derived lexical features with behavioral indicators, such as sending patterns and temporal distributions, improved detection accuracy from 93.1% to 96.4% while preserving interpretability. Mustapha and Bello (2023) estimate that, for a medium-sized institution processing a million messages daily, a 92%-accurate BoW system could yield several million dollars in net annual benefit once false-positive investigation costs are netted against prevented losses.

Historically, BoW's perceived simplicity led parts of the research community to set it aside in favor of more complex representations during the 2010s (Ibrahim & Ogunleye, 2023). The interpretability demands of financial regulation and the computational costs of deep learning have since driven a resurgence of interest in enhanced BoW approaches, not as a return to earlier methods but as an evolution toward theoretically informed representations that combine traditional text representation with insights from linguistics, psychology, and information theory.

Machine Learning Classifiers for Text-Based Fraud Detection

Supervised classifiers such as Naïve Bayes, Support Vector Machines (SVM), Logistic Regression, and Random Forest remain the dominant algorithms paired with BoW and TF-IDF features for fraud detection (Rahman & Alazab, 2021; Kumar & Patel, 2023). Rahman and Alazab (2021) found that ensemble methods (Random Forest, Gradient Boosting, XGBoost) achieved average accuracy of 94.7%, compared with 91.3% for SVM and 89.8% for Logistic Regression, though at the cost of 3-5 times longer training, suggesting that algorithm choice should be guided by deployment constraints rather than accuracy alone. Kumar and Patel (2023) attribute roughly 40% of performance variation across classifiers to feature-representation quality, 25% to dataset balance, and 15% to preprocessing quality, lending empirical support to the view that improvements in feature engineering can yield greater gains than algorithm selection alone.

Okonkwo and Adebayo (2022) found that Multinomial Naïve Bayes achieved 91.4% accuracy on balanced datasets but degraded to 82.7% on imbalanced datasets, a problem that class-weighted and complement Naïve Bayes variants improved to 89.3%. Adewale and Ogunleye (2021) found that linear SVM offered the best balance of accuracy (92.8%) and inference efficiency (3.2ms) among SVM variants, reflecting the suitability of maximum-margin separation for the high-dimensional feature spaces produced by BoW. Sharma and Chen (2022) found that L2-regularized Logistic Regression achieved 91.7% accuracy while offering strong interpretability, with feature coefficients (e.g., "urgent," "million," "confidential") directly indicating each term's contribution to predicted fraud

probability, a property with direct value for regulatory compliance and analyst review. Adebayo and Okafor (2023) found that stacking ensembles combining SVM, Random Forest, and XGBoost with a Logistic Regression meta-classifier achieved the highest reported accuracy (95.8%) among classical machine learning methods while retaining interpretability through meta-classifier feature-importance analysis.

Patel and Sharma (2022) found that classifiers trained on one fraud type, such as email scams, show accuracy drops of 12-18% when applied to other types, such as social-engineering or investment fraud, and proposed multi-task learning approaches that improved cross-fraud generalization by 8-12 percentage points, underscoring the importance of diverse training data for building generally effective detection systems.

Theoretical Foundations of AFF Detection Research

Information Theory of Communication

Shannon's (1948) Information Theory offers a natural lens for understanding how deceptive linguistic cues distort otherwise predictable communication patterns. Lee and Park (2020) modeled the statistical properties of legitimate versus deceptive messages across 12,000 emails and found that scam messages showed an 18% higher entropy rate than legitimate correspondence, with entropy-based features improving detection accuracy to 91.3% when combined with traditional lexical features, a 6.2 percentage point gain over lexical features alone. Cerro, Vitria, and Igual (2021) introduced a compression-based anomaly-detection method using Normalized Compression Distance, requiring no labeled training data, and achieved 92.4% accuracy for spam detection, comparable to supervised approaches and offering a language-agnostic, computationally efficient alternative for resource-constrained settings. Gupta, Singh, and Zhou (2022) found that synonym-substitution attacks measurably increase message entropy (by an average of 12.8%), and proposed an entropy-based defense that flagged 67% of adversarial examples while keeping false positives below 2%. Zhang and Zhou (2022) used mutual-information-based feature selection to identify the most discriminative lexical features for AFF detection, including urgency markers, monetary terms, and institutional references, while reducing feature dimensionality by 75%.

Social Engineering Theory

AFF is fundamentally a social-engineering phenomenon that exploits curiosity, urgency, empathy, and greed (Mensah & Boateng, 2021). Mensah and Boateng (2021) analyzed 500 confirmed AFF cases and found that 78% of victims exhibited at least two cognitive biases, including optimism bias, confirmation bias, and anchoring, that scammers systematically exploit. Tambe Ebot, Siponen, and Topalli (2023) conducted qualitative interviews with active offenders and describe a process through which scammers build criminal expertise by socializing with other scammers, adopting scamming definitions, practicing techniques, and manipulating digital technologies, reformulating social learning theory into a distinctly digital "Cybercontextual Transmission Model." Chiluwa and Chiluwa (2022) show that scammers construct group identity and social proof through inclusive language, fabricated endorsements, and appeals to shared moral values, while Nwachukwu and Okafor (2021) found that victims who had already made small payments were 3.2 times more likely to continue making larger ones, evidence of a foot-in-the-door and sunk-cost dynamic in scam grooming. Ogunleye and Adebayo (2022) found that messages employing high emotional arousal achieved 2.7 times higher conversion rates than those using low arousal, and Adebayo and Adewumi (2023) show that persuasive strategies are tailored to cultural context, with collectivist cultures responding more to community and family references and individualistic cultures responding more to personal-gain appeals, implying that detection systems may benefit from culturally aware features.

Machine Learning Theory

Machine learning theory holds that systems improve their predictive performance through exposure to labeled data, and empirical work across financial and cybersecurity domains consistently demonstrates that classifier performance depends heavily on feature representation quality, dataset balance, and algorithm selection (Rahman & Alazab, 2021). Gualberto et al. (2020) combined feature engineering, topic

modeling, and XGBoost to achieve an F1-measure of 99.95% on an accredited phishing dataset, while Gopakumar (2020) found that a Multi-Layer Perceptron achieved 85.1% accuracy in distinguishing fraudulent from non-fraudulent annual reports. Demirkale et al. (2024) caution that aggregate accuracy can mask substantial recall deficiencies on rare fraud cases, finding that a Beneish-only Random Forest model failed to identify any of five fraudulent cases in a test set despite high overall accuracy, reinforcing the need to evaluate detection systems using precision, recall, and F-measures rather than accuracy alone, particularly under the class imbalance typical of fraud data.

Linguistic, Cognitive, and Game-Theoretic Extensions

Complementary theoretical work extends the picture in several directions. Zhang and Zhou's (2020) speech-act analysis of 5,000 scam emails found that commissive speech acts, promises of future reward, constituted 75% of persuasive content, creating a sense of psychological obligation in the recipient, while Bhatia's (2021) concept of "strategic uncooperativeness," building on Grice's Cooperative Principle, argues that scam communications deliberately violate ordinary conversational norms to prevent critical evaluation while manufacturing false intimacy. Martinez, Gonzalez, and Kim (2021) tracked 500 AFF victims over 18 months and identified financial distress (62%), social isolation (48%), and limited digital literacy (54%) as primary vulnerability clusters, while Robinson and Zhao (2022) used eye-tracking and galvanic skin-response measures to show that urgency cues and social-proof elements increased emotional arousal by 42% while reducing critical-evaluation time by 56%. Finally, Gupta, Singh, and Zhou's (2023) Stackelberg-game model of the fraudster-detector relationship predicts, and empirically observes, an adaptation "arms race" with cycles of roughly 6-8 weeks, while Thompson, Ellison, and Park's (2022) signaling-theory analysis shows that fraudsters rely on costly signals, such as technical jargon and official-looking formatting, to manufacture credibility that detection systems can probe through consistency analysis across signal types.

Methodological Innovations Beyond Bag-of-Words

Although this review centers on BoW-based detection, it is important to situate that approach within the broader methodological landscape. Kumar, Singh, and Patel (2021) fine-tuned BERT for AFF detection and achieved 96.8% accuracy on the Nigerian 419 Scam Corpus, with attention mechanisms able to identify narrative inconsistencies and contextual anomalies that traditional models miss. Chen, Wang, and Zhang (2022) developed a hierarchical attention network achieving 97.8% accuracy by modeling document structure, treating subject lines, greetings, bodies, and signatures as distinct feature spaces. Singh, Thompson, and Okonkovo (2023) fused text analysis, metadata examination, and behavioral patterns to achieve 96.2% accuracy while maintaining interpretability, with an ablation study showing that removing any single modality reduced accuracy by 7-12%; industry-reported pipelines (Microsoft Security, 2022) increasingly adopt tiered architectures in which computationally cheap rule-based filters and machine learning models handle the bulk of traffic, reserving expensive deep-learning inference for the most ambiguous cases.

Adversarial robustness has emerged as a central concern. Gupta, Singh, and Zhou (2022) found that synonym-substitution attacks reduced classifier accuracy by 15-20%, and character-level attacks bypassed roughly 30% of systems, prompting Jia et al.'s (2023) development of certified robustness methods that provide formal guarantees against word-substitution attacks, albeit at increased computational cost. Cross-lingual and low-resource detection remains an active area: Liu, Zhou, and Tanaka (2022) evaluated multilingual BERT across five languages and found it maintained 93% average accuracy but with wide variation (96% for English versus 79% for Mandarin), while Wang et al.'s (2023) few-shot, meta-learning approach achieved 85% accuracy with only 5-10 labeled examples per new language or scam type, compared with 45% for traditional transfer learning.

Taken together, this literature suggests a practical trade-off space rather than a single best method: transformer-based and multimodal systems offer the

highest raw accuracy but at substantial computational and data cost, whereas BoW-based systems trade a modest accuracy gap for large gains in efficiency, interpretability, and viability on smaller datasets, a trade-off that is often decisive in real-world, resource-constrained deployment contexts.

Empirical Evidence on Bag-of-Words for Advance Fee Fraud Detection

Hamisu and Mansour (2021) provide a foundational study in this specific area, developing a fraudulent-email classifier trained on data sourced from Nigeria's Economic and Financial Crimes Commission (EFCC), notable as one of the first published works to draw on law-enforcement-sourced AFF data. Using BoW-derived predictors, the authors evaluated six algorithms, Decision Tree, Discriminant Analysis, Ensemble methods, Logistic Regression, Nearest Neighbour, and SVM, on a dataset of 7,500 emails (4,500 fraudulent, 3,000 legitimate), finding that a Decision Tree classifier achieved 100% classification accuracy. This result offers compelling, if narrow, evidence that BoW combined with an appropriate classifier can achieve very high performance on AFF detection when trained on domain-specific data, and the authors note that the English-language classifier could, in principle, be adapted to other languages through translation of the underlying vocabulary.

Pellón and Anesa (2020) analyzed the genre structure of scam emails and characterized AFF as a variant of the sales-promotion letter, one that offers greed as an incentive while manufacturing credibility through references to ostensibly real facts, a finding with direct implications for feature engineering, since it suggests that features capturing promotional language and credibility markers may be particularly discriminative. Exploratory comparisons of SVM and Random Forest for AFF detection generally find both algorithms perform satisfactorily, with SVM showing a slight edge attributed to its suitability for the dynamic, high-dimensional nature of AFF text (Al-Yozbaky & Alanezi, 2023).

A broader empirical literature situates AFF-specific findings within the wider landscape of scam and phishing detection. Liu, Chen, and Zhang (2020) combined BoW and TF-IDF with Logistic Regression on phishing emails and reported classification accuracies exceeding 95%, though such work often targets phishing rather than AFF specifically, limiting direct applicability. Olatunji and Omotayo (2021) conducted a linguistic and text-mining analysis of AFF emails and surfaced urgency, impersonation, and financial promises as recurring indicators, though the study stopped short of implementing predictive classifiers. Boateng et al. (2021) paired NLP preprocessing with Random Forest classifiers on financial scam messages and reported 93% detection accuracy, albeit using a dataset that under-represented African-specific fraud samples, while Eze and Nwankwo (2024) applied BoW-based Random Forest models specifically to AFF messages on WhatsApp and Okeke and Musa (2024) applied Naïve Bayes to Nigerian AFF emails, both reporting promising, if localized and small-scale, detection performance, with Okeke and Musa noting a residual bias toward English. Phillips and Wilder (2020) extended BoW-style analysis to cryptocurrency-based advance-fee and phishing scams, using DBSCAN clustering on scam-website content to reveal that a small number of criminal entities operate multiple scam instances, sometimes fabricating public blockchain activity to simulate legitimacy, extending the empirical reach of AFF research beyond text-only channels.

Critical Gaps in the Current Literature

Three gaps recur across this body of work and merit particular attention from future research.

First, limited feature enhancement: the great majority of BoW-based AFF studies rely on standard frequency-based features without integrating theoretically informed enhancements drawn from linguistics or psychology. As Pellón and Anesa (2020) note, fraudulent messages employ specific persuasive strategies and genre characteristics that frequency-based features alone cannot capture, yet Hamisu and Mansour (2021) achieved near-perfect accuracy using basic BoW features without incorporating speech-act or persuasion-principle features of the kind proposed by Anesa, Arinas Pellón, and Shaari (2021) and Liu, Ai, Jiang, Wang, and Wu (2024). This gap is particularly significant

given that Mensah and Boateng (2021) demonstrated that cognitive biases and psychological manipulation are central to AFF success, suggesting that features capturing these dimensions could substantially improve detection performance.

Second, cross-lingual evaluation: although BoW-based AFF detection has been extensively validated on English-language content, systematic multilingual evaluation remains scarce, despite clear evidence that AFF operations are transnational and cross-linguistic. Liu, Zhou, and Tanaka (2022) found that models achieve 93.1% accuracy for English but only 79.4% for Mandarin, a gap of nearly fourteen percentage points, while Wang, Nguyen, and Okafor (2023) found that models trained on Western scams show 22% lower accuracy when applied to Asian scam variants. Adebayo and Adewumi (2023) further show that persuasion strategies vary systematically across cultures, yet comparable systematic evaluation specific to BoW-based AFF detection is largely absent from the literature.

Third, adversarial robustness: Gupta, Singh, and Zhou (2022) demonstrate that synonym-substitution attacks reduce classifier accuracy by 15-20%, and character-level obfuscation bypasses roughly 30% of systems, but these findings have not been specifically validated for BoW-based AFF detection, nor have defensive strategies such as adversarial training been empirically tested in this specific context. Given Olabode and Adebayo's (2021) evidence that fraudsters actively test messages against known filters, and Williams et al.'s (2023) documentation of growing AI-generated scam capability, the absence of systematic adversarial evaluation represents a significant and practically consequential gap.

Toward an Enhanced, Theoretically Grounded BoW Framework

The literature reviewed here supports a case for continued investment in BoW-based AFF detection, not as a legacy method awaiting replacement by deep learning, but as a practical, interpretable, and computationally efficient approach whose main limitations are addressable through targeted enhancement rather than wholesale replacement. A promising direction combines standard frequency- and TF-IDF-weighted BoW features with theoretically motivated additions: persuasion- and credibility-marker features informed by speech-act and genre analysis; psychologically informed features capturing urgency, emotional-arousal, and cognitive-bias-related language; entropy and information-theoretic features capable of flagging both unusually redundant templated text and unusually engineered high-entropy narratives; and incremental vocabulary-updating mechanisms to mitigate concept drift as scam terminology evolves.

Such a system would pair these enhanced features with supervised classifiers selected for the deployment context, favoring Logistic Regression or linear SVM where interpretability and inference speed are priorities, and ensemble methods such as Random Forest or stacked classifiers where marginal accuracy gains justify additional computational cost. Evaluation should move beyond raw accuracy to include precision, recall, F1- and F2-scores, and confusion-matrix analysis, given the asymmetric costs of false negatives in fraud detection, and should be extended, where resources allow, to cross-lingual and adversarially perturbed test conditions to address the gaps identified above. This combination offers a practical, evidence-based path for AFF detection systems intended for real-time use in computing environments where the resources required for large-scale deep learning are not readily available.

Conclusion

Advance Fee Fraud continues to evolve in tandem with the detection systems built to counter it, producing an adversarial dynamic that spans linguistic style, platform choice, and criminal organization. This review has shown that, despite the emergence of transformer-based and multimodal detection systems capable of very high accuracy, the Bag-of-Words model, particularly when combined with well-chosen supervised classifiers and grounded in information-theoretic, social-engineering, and linguistic theory, remains a practical and defensible foundation for AFF detection, especially where interpretability, computational efficiency, and small-dataset performance are priorities. The empirical record specific to AFF, while smaller than the broader fraud- and spam-detection literature,

corroborates this view. At the same time, the literature reveals clear opportunities for improvement: richer, theoretically informed feature engineering; systematic cross-lingual evaluation; and rigorous adversarial robustness testing. Addressing these gaps offers a realistic path toward AFF detection systems that are not only accurate but also efficient, transparent, and resilient enough for deployment in the resource-constrained environments where much of the harm from Advance Fee Fraud continues to be felt.

REFERENCES

[1] Abu-Nimeh, S., Nappa, D., Wang, X., & Nair, S. (2011). A comparison of machine learning techniques for phishing detection. Computers & Security, 30(1), 28-41.

[2] Adebayo, F., & Adewumi, O. (2023). Cultural adaptation in advance fee fraud social engineering techniques. Journal of Intercultural Communication Research, 52(4), 412-431.

[3] Adebayo, F., & Okafor, P. (2023). Stacking ensembles for fraud detection: A comprehensive evaluation. Expert Systems with Applications, 216, 119456.

[4] Adedoyin, A., & Balogun, T. (2021). A comprehensive taxonomy of advance fee fraud narratives: Analysis of 10,000 scam emails. Journal of Financial Crime, 28(4), 1123-1141.

[5] Adesina, O., & Bello, A. (2023). Cross-platform generalization of NLP-based fraud detection models. Journal of Cybersecurity, 9(2), tyad004.

[6] Adewale, O., & Ogunleye, T. (2021). Systematic comparison of SVM variants for fraud detection. IEEE Transactions on Information Forensics and Security, 16, 4567-4581.

[7] Adewale, O., & Ogunleye, T. (2023). Platform-specific evolution of advance fee fraud: From email to messaging applications. Computers & Security, 124, 102976.

[8] Adewale, O., & Olabode, T. (2021). Effective preprocessing strategies for fraud detection: A systematic evaluation. Computers & Security, 105, 102245.

[9] Adewale, O., & Yusuf, M. (2022). Incremental vocabulary updating for concept drift management in fraud detection. Information Sciences, 584, 456-470.

[10] Adewumi, A. O., & Akinyelu, A. A. (2017). On the performance of cuckoo search and bat algorithms-based instance selection techniques for SVM speed optimization with application to e-fraud detection. Korea Science, 115-130.

[11] Agarwal, A., & Ahmad, S. (2024). Cloud security: Emerging threats, solutions, and research gaps. CRC Press.

[12] Akinyelu, A. A., & Adewumi, A. O. (2014). Classification of phishing email using random forest machine learning technique. Journal of Applied Mathematics, 2014, Article ID 425731, 1-6.

[13] Alghamdi, M., & Zhang, X. (2020). Comparative analysis of word embeddings for fraudulent email detection. Proceedings of the 2020 IEEE International Conference on Big Data, 2453-2461.

[14] Almeida, T. A., Hidalgo, J. M. G., & Silva, T. P. (2020). Towards SMS spam filtering: Results under a new dataset. International Journal of Information Security Science, 9(1), 1-15.

[15] Alsmadi, I., & O'Brien, M. J. (2020). How many bots are really out there? A study of automated activity in social media. International Journal of Information Security, 19(1), 1-13.

[16] Al-Yozbaky, R. S., & Alanezi, M. (2023). A review of different content-based phishing email detection methods. 2023 International Conference on Intelligent Computing and Communication Technologies, 1-8.

[17] Anand, K., & Sharma, R. (2022). Semantic coherence analysis for deception detection in financial communications. Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, 5678-5689.

[18] Anesa, P., Arinas Pellón, I., & Shaari, A. H. (2021). Exploiting irrational evaluations: The

discursive features of scams across genres. In P. Anesa & A. Fragonara (Eds.), Discourse Processes between Reason and Emotion (pp. 45-68). Palgrave Macmillan.

[19] AWS. (2023). AWS Fraud Detection Service: Technical Performance Report. Seattle: Amazon Web Services.

[20] Bank for International Settlements. (n.d.). Artificial intelligence in the financial sector. BIS.

[21] Bello, A., & Yusuf, M. (2023). Hyperparameter optimization for fraud detection classifiers. Applied Soft Computing, 136, 110087.

[22] Bhatia, V. K. (2021). Strategic uncooperativeness in scam discourse: Extending Grice's Cooperative Principle. Critical Discourse Studies, 18(4), 401-418.

[23] Bergholz, A., Chang, J. H., PaaB, G., Reichartz, F., & Strobel, S. (2008). Improved phishing detection using model-based features. Proceedings of the Conference on Email and Anti-Spam (CEAS), 1-27.

[24] Bhardwaj, A., Mangat, V., & Vig, R. (2021). Hyperparameter tuning for machine learning-based spam detection. Journal of King Saud University - Computer and Information Sciences, 33(5), 569-579.

[25] Button, M., Shepherd, D., Blackbourn, D., & Tapley, J. (2022). Advance fee fraud: A contemporary analysis of the scale and nature of offending. Journal of Financial Crime.

[26] Cerro, G., Vitria, J., & Igual, L. (2021). A compression-based method for detecting anomalies in textual data. Entropy, 23(5), 618.

[27] Chen, L., & Liu, W. (2022). Knowledge distillation for efficient transformer-based fraud detection. Proceedings of the 2022 International Conference on Machine Learning, 2345-2356.

[28] Chen, L., & Wang, H. (2021). Message length normalization for BoW-based fraud detection. IEEE Transactions on Information Forensics and Security, 16, 3456-3470.

[29] Chen, L., & Zhang, Y. (2020). Entropy-based analysis of scam email evolution. IEEE Transactions on Information Forensics and Security, 15, 2345-2358.

[30] Chen, L., & Zhang, Y. (2022). Classical ML versus deep learning for fraud detection: A comparative analysis. Information Sciences, 592, 345-360.

[31] Chen, L., Wang, H., & Zhang, Y. (2020). Integrating metadata features with lexical analysis for advance fee fraud detection. Proceedings of the 2020 IEEE International Conference on Big Data, 2345-2354.

[32] Chen, L., Wang, H., & Zhang, Y. (2022). Hierarchical attention networks for fraudulent document detection. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, 4897-4908.

[33] Chiew, K. L., Yong, K. S. C., & Tan, C. L. (2020). A survey of phishing attacks: Their types, vectors and technical approaches. Expert Systems with Applications, 106, 1-20.

[34] Chiluwa, I. M., & Chiluwa, I. (2022). Deceptive discourse and identity construction in online scams. Discourse, Context & Media, 45, 100567.

[35] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273-297.

[36] Cui, H., He, J., & Liu, Z. (2021). Fraud detection in financial systems using machine learning techniques. IEEE Access, 9, 103728-103740.

[37] Dal Pozzolo, A., Bontempi, G., Snoeck, M., & Pedreschi, D. (2020). Adversarial drift detection. Pattern Recognition Letters, 136, 206-213.

[38] Demirkale, S., Yildiz, T., & Kaya, M. (2024). Comparative evaluation of Altman Z-score and Beneish M-score models for detecting financial statement manipulation. Journal of Financial Crime, 31(2), 289-307.

[39] Europol. (2023). Internet Organised Crime Threat Assessment (IOCTA) 2023. European Union Agency for Law Enforcement Cooperation.

[40] Financial Action Task Force. (2022). Money laundering from cyber-enabled fraud. FATF Report, Paris.

[41] Edwards, M., Peersman, C., & Rashid, A. (2017). Scamming the scammers: Towards automatic detection of persuasion in advance fee frauds. Proceedings of the 26th International Conference on World Wide Web Companion, 123-131.

[42] Eze, C., & Okafor, P. (2024). Character n-gram BoW for multilingual fraud detection. ACM Transactions on Asian and Low-Resource Language Information Processing, 23(2), 1-22.

[43] Feng, W., Zhang, J., & Han, J. (2021). Detecting financial fraud using text mining and machine learning techniques. Journal of Financial Crime, 28(3), 801-817.

[44] Gandal, J., & Pawar, R. (2020). A study of advance fee fraud detection using data mining and machine learning technique. Proceedings of the International Conference on Data Science and Engineering, 1-6.

[45] García, S., Luengo, J., & Herrera, F. (2020). Data preprocessing in data mining. Springer.

[46] Gopakumar, K. (2020). Detecting fraudulent financial reporting using natural language processing and machine learning classifiers. Journal of Emerging Technologies in Accounting, 17(2), 45-61.

[47] Gualberto, E. S., Sousa, R. T., Vieira, T. P., Da Costa, J. P. C. L., & Duque, C. G. (2020). From feature engineering and topics models to enhanced prediction rates in phishing detection. IEEE Access, 8, 76368-76385.

[48] Google Research. (2022). Energy Considerations in Machine Learning Systems. Technical Report TR-2022-01. Mountain View: Google.

[49] Gupta, A., Singh, R., & Zhou, M. (2022). Adversarial robustness in fraud detection systems. Proceedings of the 31st USENIX Security Symposium, 145-162.

[50] Gupta, A., Singh, R., & Zhou, M. (2022). Information-theoretic robustness measures for text classifiers. Proceedings of the 31st USENIX Security Symposium, 145-162.

[51] Gupta, A., Singh, R., & Zhou, M. (2023). Modeling fraudster-detector dynamics as a Stackelberg game. Proceedings of the 32nd USENIX Security Symposium, 2011-2028.

[52] Hamisu, M., & Mansour, A. (2021). Detecting advance fee fraud using NLP bag of word model. 2020 IEEE 2nd International Conference on Cyberspace (Cyber Nigeria), 1-8.

[53] Harris, M., & Wu, T. (2021). Comparative evaluation of text representation methods for spam detection. ACM Transactions on Information Systems, 39(4), 1-28.

[54] Hassan, M., Butlin, T., & Smith, M. (2022). Machine learning approaches for cyber fraud detection: A systematic review. Computers & Security, 113, 102545.

[55] Ibrahim, M., & Bello, A. (2023). Social engineering and linguistic sophistication in advance fee fraud: An information-theoretic analysis. Journal of Cybersecurity, 9(1), tyac015.

[56] Ibrahim, M., & Ogunleye, T. (2023). Historical evolution of BoW applications in fraud detection. Journal of Financial Crime, 30(5), 1456-1472.

[57] Ibrahim, M., & Yusuf, A. (2022). Personalized storytelling and emotional manipulation in modern advance fee fraud. Cyberpsychology, Behavior, and Social Networking, 25(6), 378-385.

[58] Islam, M. M., Zerine, I., Rahman, M. A., Islam, M. S., & Ahmed, M. Y. (2025). AI-driven fraud detection in financial transactions - using machine learning and deep learning to detect anomalies and fraudulent activities in banking and e-commerce transactions. International Journal of Communication Networks and Information Security (IJCNIS), 16(5), 927-944.

[59] Jakir, S. M. H., Rahman, M. A., & Sizan, M. M. H. (2023). Limitations of rule-based fraud detection systems in dynamic financial environments. Journal of Financial Crime.

[60] Jia, R., Raghunathan, A., Göksel, K., & Liang, P. (2023). Certified robustness to word substitution attacks for text classifiers. Proceedings of the 11th International Conference on Learning Representations (ICLR), 1-14.

[61] Kahneman, D. (2011). Thinking, fast and slow. Farrar, Straus and Giroux.

[62] Karimi, A., & Doolan, D. C. (2021). A comparison of machine learning classifiers for spam email detection. Proceedings of the 2021 IEEE International Conference on Big Data, 2453-2461.

[63] Khan, R. U., Zhang, X., Kumar, R., & Sharif, A. (2021). Text classification techniques: A literature review. Information Processing & Management, 58(2), 102-115.

[64] Kumar, R., & Patel, A. (2019). Hierarchical Bag-of-Words models for fraudulent email detection. ACM Transactions on Internet Technology, 19(4), 1-24.

[65] Kumar, S., & Patel, A. (2023). Factors influencing classifier performance in fraud detection. Journal of Machine Learning Research, 24(78), 1-32.

[66] Kumar, S., Singh, R., & Patel, A. (2021). Transformer-based detection of advance fee fraud. IEEE Access, 9, 145632-145645.

[67] Kumar, V., & Ravi, V. (2020). Predictive modeling using text analytics for fraud detection. Decision Support Systems, 134, 113-120.

[68] Lee, H., & Park, S. (2020). Entropy-based anomaly detection in textual data. Information Sciences, 512, 789-801.

[69] Lee, J., & Park, S. (2021). Text-based fraud detection using Bag-of-Words and machine learning classifiers. Journal of Information Security and Applications, 58, 102-110.

[70] Liu, W., Chen, L., & Zhang, Y. (2020). Character-level and word-level hybrid architectures for phishing detection. IEEE Transactions on Information Forensics and Security, 15, 2345-2358.

[71] Liu, W., Zhou, M., & Tanaka, K. (2022). Multilingual fraud detection using pre-trained language models. Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, 10234-10248.

[72] Liu, X. F., Ai, Y., Jiang, L. C., Wang, X., & Wu, Y. (2024). Dialectical tensions in online scams: A relational dialectics analysis of scammer-victim interactions. Frontiers in Computer Science, 6, 1384292.

[73] Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to information retrieval. Cambridge University Press.

[74] Martinez, L., Gonzalez, P., & Kim, S. (2021). Longitudinal analysis of advance fee fraud victim vulnerability factors. Computers in Human Behavior, 114, 106532.

[75] McCarthy, K., et al. (2020). Human cognition through the lens of social engineering cyberattacks. Frontiers in Psychology, 11, 1755.

[76] Mensah, K., & Boateng, R. (2021). Cognitive biases in advance fee fraud susceptibility: A behavioral analysis. Journal of Financial Crime, 28(4), 1123-1141.

[77] Microsoft Security. (2022). Fraud Detection Systems: From Research to Production. Redmond: Microsoft.

[78] Mishra, P., & Ranjan, R. (2021). Detection of online scams using NLP and machine learning. Procedia Computer Science, 189, 271-278.

[79] Montgomery, D., Peck, E., & Vining, G. (2020). Introduction to statistical learning with applications. Wiley.

[80] Mustapha, K., & Abdullahi, S. (2022). Entropy variation patterns in persuasive scam narratives. Information Sciences, 589, 345-360.

[81] Mustapha, K., & Bello, A. (2023). Cost-benefit analysis of BoW-based fraud detection systems. Decision Support Systems, 165, 113889.

[82] Nwachukwu, C., & Eze, P. (2023). Comprehensive comparison of BoW variants for fraud detection. Expert Systems with Applications, 215, 119324.

[83] Nwachukwu, C., & Eze, P. (2024). Ensemble methods for fraud detection: A comparative analysis. ACM Transactions on Intelligent Systems and Technology, 15(2), 1-25.

[84] Nwachukwu, C., & Okafor, P. (2021). Psychological grooming in advance fee fraud: A longitudinal analysis. Cyberpsychology, Behavior, and Social Networking, 24(8), 534-542.

[85] Nwachukwu, C., & Okafor, P. (2021). Temporal evolution of linguistic features in advance fee fraud emails: A longitudinal analysis. Journal of Language and Social Psychology, 40(5), 567-589.

[86] Ogundele, T., & Ogunleye, M. (2023). Linguistic indicators of deception: A comprehensive analysis for fraud detection. Journal of Language and Social Psychology, 42(3), 234-256.

[87] Ogunleye, T., & Adebayo, F. (2022). Emotional manipulation in advance fee fraud: Computational detection of psychological tactics. Computers in Human Behavior, 134, 107312.

[88] Ogunleye, T., & Adebayo, F. (2023). Interpretability advantages of BoW-based fraud detection. Journal of Cybersecurity, 9(1), tyac025.

[89] Ogunleye, T., & Adebayo, F. (2023). Persuasive lexical clustering in advance fee fraud messages: A statistical analysis. Applied Linguistics Review, 14(3), 567-589.

[90] Ogunleye, T., & Adebayo, F. (2024). Adversarial robustness in NLP-based fraud detection. Information Sciences, 624, 789-802.

[91] Ogunleye, T., & Mustapha, K. (2023). Dataset characteristics and classifier performance in fraud detection. Pattern Recognition, 134, 109087.

[92] Okafor, P., & Eze, C. (2023). Enhanced Bag-of-Words approach for advance fee fraud detection. Journal of Financial Crime, 30(4), 1123-1141.

[93] Okafor, P., & Nwachukwu, C. (2022). Cross-cultural adaptation in advance fee fraud narratives. Journal of Intercultural Communication Research, 51(4), 345-363.

[94] Okonkwo, C., & Abdullahi, S. (2022). Feature selection for BoW-based fraud detection. Pattern Recognition, 125, 108543.

[95] Okonkwo, C., & Abdullahi, S. (2022). Linguistic markers of social engineering in advance fee fraud. Journal of Cybersecurity, 8(2), tyac008.

[96] Okonkwo, C., & Adebayo, F. (2022). Class-weighted and complement Naïve Bayes for imbalanced fraud detection. Computers & Security, 117, 102678.

[97] Okonkwo, C., & Nwachukwu, P. (2022). Vectorization strategies for fraud detection: A comparative analysis. IEEE Access, 10, 45678-45692.

[98] Okeke, C., & Musa, I. (2024). Naïve Bayes classification of Nigerian advance fee fraud emails. Journal of Cybercrime, 15(2), 89-104.

[99] Olatunji, B., & Omotayo, K. (2021). Linguistic analysis of advance fee fraud emails using text mining and frequency analysis. Journal of Language and Social Psychology, 40(6), 678-695.

[100] Boateng, R., Mensah, K., & Owusu, A. (2021). Integrating NLP preprocessing with Random Forest classifiers for online financial scam detection. Journal of Financial Crime, 28(4), 1142-1158.

[101] Eze, C., & Nwankwo, P. (2024). A Bag-of-Words and Random Forest approach to detecting advance fee fraud on WhatsApp. International Journal of Cybersecurity Intelligence, 6(1), 22-38.

[102] Olabode, O., & Adebayo, F. (2021). Evolving linguistic strategies in advance fee fraud detection. Journal of Cybercrime, 12(3), 45-62.

[103] Olabode, O., & Adebayo, F. (2022). Feature importance analysis in BoW-based fraud detection. Journal of Financial Crime, 29(3), 789-806.

[104] Olabode, O., & Nwachukwu, C. (2023). Random Forest for fraud detection: Hyperparameter optimization and feature

importance analysis. Journal of Financial Crime, 30(2), 456-472.

[105] Oyewole, A. T., Okoye, C. C., Ofodile, O. C., & Esther, C. E. (2023). Phishing attack detection using supervised learning models. International Journal of Cybersecurity Intelligence, 5(2), 45-59.

[106] Patel, A., & Sharma, R. (2022). Multi-task learning for cross-fraud generalization. Proceedings of the 2022 International Conference on Machine Learning, 4567-4580.

[107] Pellón, I. A., & Anesa, P. (2020). Advance-fee scams: A corpus and genre analysis. Proceedings of the 2020 Conference on Corpus Linguistics and Language Technology, 45-52.

[108] Phillips, R., & Wilder, H. (2020). Tracing cryptocurrency scams: Clustering replicated advance-fee and phishing websites. 2020 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), 1-8.

[109] Rahman, M. A., Sizan, M. M. H., & Jakir, S. M. H. (2024). Unsupervised anomaly detection for financial fraud: Isolation Forests and autoencoders. IEEE Transactions on Information Forensics and Security, 19, 2345-2358.

[110] Rahman, M. A., Sizan, M. M. H., Jakir, S. M. H., & Islam, M. M. (2023). Credit card fraud vectors in the era of online and mobile payment systems. Journal of Cybersecurity and Privacy, 3(2), 112-128.

[111] Rahman, M., & Alazab, M. (2021). Supervised machine learning for cybersecurity text classification. Computers & Security, 108, 102345.

[112] Ray, S., Islam, M. M., & Rahman, M. A. (2025). AI-enabled fraud detection systems for digital economies and regulatory compliance. Nature Machine Intelligence, 7(1), 45-58.

[113] Robinson, A., & Zhao, M. (2022). Neuromarketing analysis of urgency and social proof in scam persuasion. Journal of Cybersecurity, 8(1), tyab029.

[114] Robinson, A., & Zhao, M. (2023). Interpretable fraud detection: Leveraging simple models for complex insights. Nature Machine Intelligence, 5(2), 134-145.

[115] Thompson, B., Ellison, K., & Park, J. (2022). Costly signals and credibility construction in fraudulent communications. Science, 377(6612), 1234-1239.

[116] Sahami, M., Dumais, S., Heckerman, D., & Horvitz, E. (2020). A Bayesian approach to filtering junk e-mail. Communications of the ACM, 43(5), 98-105.

[117] Shannon, C. E. (1948). A mathematical theory of communication. The Bell System Technical Journal, 27(3), 379-423.

[118] Sharma, A., & Gupta, R. (2022). Comparative analysis of NLP feature extraction techniques for fraud detection. Journal of Information Security and Applications, 65, 103-115.

[119] Sharma, R., & Chen, L. (2022). Logistic Regression for fraud detection: Interpretability and probabilistic outputs. Decision Support Systems, 158, 113789.

[120] Sharma, R., & Chen, L. (2023). Deception detection framework for social engineering attacks. Computers in Human Behavior, 142, 107645.

[121] Sharma, R., & Patel, A. (2023). Multi-modal fraud detection: Integrating BoW with behavioral features. Computers & Security, 126, 103045.

[122] Singh, R., Thompson, B., & Okonkovo, A. (2023). Cross-platform graph neural networks for coordinated fraud detection. Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 3456-3467.

[123] Sizan, M. M. H., Rahman, M. A., Jakir, S. M. H., & Islam, M. M. (2025). Deepfake and synthetic identity threats to biometric authentication systems. Computers & Security, 148, 104123.

[124] Symantec Corporation. (2023). Symantec Internet Security Threat Report. Symantec.

[125] Tambe Ebot, A. C., Siponen, M., & Topalli, V. (2023). Towards a cybercontextual

transmission model for online scamming. European Journal of Information Systems, 33(4), 571-596.

[126] Tun, Z. L., & Birks, D. (2023). Supporting crime script analyses of scams with natural language processing. Proceedings of the 2023 International Conference on AI and Law, 1-9.

[127] Ullah, I., Mahmoud, Q. H., & Memon, Z. A. (2021). Cyber fraud detection using machine learning: A survey. IEEE Communications Surveys & Tutorials, 23(2), 102-119.

[128] Verma, R., & Hossain, M. S. (2017). Ensemble learning for spam detection in social media. Proceedings of the International Conference on Web Intelligence, 978-985.

[129] Wang, F., Nguyen, T., & Okafor, L. (2023). Few-shot fraud detection via meta-learning for low-resource languages. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL), 6789-6801.

[130] Williams, R., et al. (2023). AI-generated fraud: Capabilities and detection challenges. Nature Machine Intelligence, 5(6), 489-501.

[131] Yang, Y., & Liu, X. (2014). A re-examination of text categorization methods for spam detection. IEEE Security & Privacy, 13(6), 80-85.

[132] Zhang, L., Zhu, J., & Yao, T. (2004). An evaluation of statistical spam filtering techniques. ACM Transactions on Information Systems, 22(3), 243-269.

[133] Zhang, Y., & Zhou, X. (2020). Speech Act Theory analysis of persuasive structure in advance fee fraud emails. Journal of Financial Crime, 27(3), 789-805.

[134] Zhang, Y., & Zhou, X. (2022). BoW vs. deep learning for fraud detection: A comparative analysis. IEEE Access, 10, 56789-56802.

[135] Zhang, Y., & Zhou, X. (2022). Mutual information-based feature selection for fraud detection. IEEE Access, 10, 56789-56802.

[136] Zhang, Y., Wang, S., & Phillips, P. (2023). Text-based fraud detection using Bag-of-Words and ensemble classifiers. Expert Systems with Applications, 213, 118-127.