IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 19, April 2026
Impact Factor: 8.2

E-ComShield: AI-Based Detection of Deceptive Reviews and Fraudulent Product Listings

Vikram Sandeep P[1], Suresh Yeresime[2], B. K. Vikram Simha[3], M. Marulesh[4], MD Sahil[5], R. P. Girish[6]
Department of Computer Science and Engineering (Artificial Intelligence)
Ballari Institute of Technology and Management, Ballari, India
[1]vikramsandeep@example.com, [2]suresh.vec04@gmail.com, [3]vikrambk9481@gmail.com, [4]maruleshm@gmail.com,
[5]mohammedsahil0026@gmail.com, [6]rpgirish79@gmail.com

Abstract: *With the rapid growth of e-commerce platforms, consumers increasingly rely on product listings and user reviews to make informed purchasing decisions. However, the rise of fake reviews and fraudulent product listings has undermined the trustworthiness of online marketplaces. This paper presents an AI-powered system designed to detect and flag fake reviews and suspicious product listings. The proposed framework uses natural language processing (NLP), image analysis, and machine learning techniques to analyze textual feedback and product-related data. This approach identifies patterns that suggest fraud and deception. By combining different types of data analysis, the system offers a strong and automated solution for improving consumer trust and ensuring safer online shopping experiences. The results highlight the potential of AI-driven methods in reducing the impact of fraud in e-commerce environments.*

Keywords: e-commerce, fake reviews, fraudulent listings, image analysis, machine learning, natural language processing (NLP), online marketplace, trust and safety.

I. INTRODUCTION

E-commerce platforms have flourished combining physical and digital shopping experiences. Online marketplaces have become a leading source for completing product searches, with user reviews and product listings positioned at the forefront of purchasing decisions. Yet, this accelerated growth has also spurred a notable increase in fraudulent behaviors such as misleading product listings and gaming the review systems, threatening consumer trust and the integrity of electronic commerce.

However, identifying this type of fraudulent behaviour poses particular challenges. Suited period of data until October 2023. Likewise, images used to display products can also be manipulated or distorted to give a false sense of legitimacy. Conventional rule-based or heuristic-driven methods have proven inadequate for representing the complexity of these misleading paradigms.

To mitigate these issues, this paper proposes an AI based framework E-ComShield where three complement models are incorporated to identify fraudulent product listings and counterfeit reviews. More specifically, a variant of the transformer based natural language processing model RoBERTa is used for textual reviews since it has high contextual accuracy. We use EfficientNet-B4 to analyze our product images, which enables us to effectively and efficiently detect images that are either manipulated or misleading at scale. Furthermore, we apply XGBoost (a popular and scalable gradient boosting decision tree algorithm) to structured product and review metadata for highly accurate classification of suspicious listings. These models can be combined to produce a multi-modal method that significantly improves the detection performance relative to single-modality systems.

The main contributions of this paper are summarized as follows:

1. NLP, Image analysis and Machine Example of e-Comprehensive Model Learning fusion applications Fraud detection.
2. Leveraging RoBERTa, EfficientNet-B4 and XGBoost for text+image+metadata based analysis

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-34491

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 19, April 2026
Impact Factor: 8.2

3. Showcase better detection accuracy and reliability of fake reviews and fraudulent product listings

II. LITERATURE SURVEY

Vijay et al. [1] present an applied study on fake product identification using classical supervised machine-learning classifiers (SVM, Random Forest, etc.) built from engineered features extracted from product listings and associated metadata. Their work emphasizes the utility of structured features — for example seller attributes, price anomalies, and simple text statistics — in discriminating counterfeit items. The study demonstrates that traditional classifiers remain competitive when meaningful features are available and the dataset is well-curated. However, [1] also illustrates common limitations: reliance on handcrafted features that require domain expertise, sensitivity to dataset bias, and reduced robustness when textual or visual cues dominate the deception. These observations motivate combining structured metadata models with stronger representation learners for text and images. and not as an independent document. Please do not revise any of the current designations.

Daoud et al. [2] apply SSD300, a single-shot object detector, to logo detection as part of counterfeit product recognition. This approach shows that local visual cues — especially brand logos and packaging labels — provide powerful signals for authenticity checks. Object detectors and CNN-based classifiers can capture fine-grained spatial patterns that are difficult to encode with metadata features. The primary shortcoming reported in [2] is dataset scale and diversity: their

experiments rely on a relatively small and narrowly distributed image corpus, which limits generalization across varied photographic conditions and product categories. This underlines the need for scalable image architectures and large, varied training sets; it also motivates use of efficient but high-capacity models such as EfficientNet for product-image forensics.

Shreekumar et al. [3] describe a blockchain-based authentication scheme that leverages QR codes and immutable ledger entries to verify product provenance. This line of work departs from purely detection-centric methods and instead focuses on supply-chain traceability and tamper-resistant records. Blockchain solutions provide strong integrity guarantees when the on-chain data is reliable, but they have practical limitations: adoption barriers across vendors, dependence on accurate initial registrations, and limited applicability for monitoring third-party marketplace listings where provenance information is absent or forged. Thus blockchain is complementary to detection systems — it can prevent some classes of fraud but cannot replace runtime detection of deceptive listings or fake reviews.

Pund et al. [4] examine a machine-learning framework targeted at identifying fake reviews, combining linguistic features (lexical cues, sentiment measures) with reviewer behavior signals. Their results highlight two central points: (1) linguistic and behavioral features together improve detection relative to either alone; and (2) automated systems still benefit from human verification in high-risk decisions. [4] also emphasizes challenges posed by short reviews, domain shifts, and adversarially written reviews designed to mimic authentic style. These findings directly motivate adoption of transformer-based language models (for stronger contextualization) integrated with behavioral/metadata models for robust review analysis.

The surveyed works collectively indicate that no single modality suffices: text models capture linguistic deception, image methods catch visual forgery, metadata models identify structural anomalies, and blockchain improves provenance when available. However, notable gaps remain:

1. Multimodal integration — Few works present end-to-end systems that fuse text, image, and metadata at scale with robust weighting strategies and missing-modal handling.

2. Dataset limitations — Many image and review datasets are small, domain-limited, or lack joint multi-modal labels, restricting cross-domain evaluation.

3. Adversarial robustness and interpretability — Systems are vulnerable to style transfer, paraphrasing, and adversarial image generation; explainability for moderation workflows is underdeveloped.

4. Operational constraints — Balancing accuracy with inference latency for real-time marketplace use is sparsely addressed.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 19, April 2026 Impact Factor: 8.2

E-ComShield directly responds to these gaps by combining RoBERTa for robust textual embeddings [5], EfficientNet-B4 for efficient visual feature extraction [6], and XGBoost for structured metadata scoring [7], while integrating the complementary insights from applied studies on classical ML and object detection [1],[2],[4]. The chosen components reflect best practices from the literature: transformer power for text, efficient CNNs for images, and gradient boosting for tabular signals, assembled into a multimodal fusion architecture aimed at real-world scalability and interpretability.

III. PROPOSED METHODOLOGY

The proposed E-ComShield framework is a multi-modal system designed to detect fraudulent product listings and fake reviews in e-commerce platforms. It integrates three complementary components: textual analysis using RoBERTa, image verification using EfficientNet-B4, and metadata-based classification using XGBoost. The system pipeline is shown in Fig. 1.

A. Textual Analysis Using RoBERTa

User reviews often contain subtle linguistic and semantic cues indicative of deception. To capture these nuances, each review undergoes preprocessing and tokenization before being fed into a fine-tuned RoBERTa sequence classification model

$$Ptext = softmax(Zr) \quad (1)$$

Where :$Zr = [zfake, zgenuine]$,Ptext=Pfake,Pgenuine

The confidence score of a review being genuine is calculated as:

$$Stext = Pgenuine \times 100 \quad (2)$$

A review is classified as real if arg max$(ptext) = 1$, and genuine otherwise.

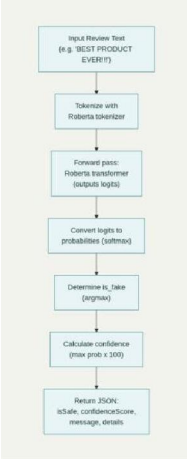

Fig 1: Workflow of Roberta Model

B. Image Verification Using EfficientNet-B0

Product images often contain visual irregularities that indicate counterfeit manufacturing patterns. To capture these cues, uploaded images are preprocessed and passed through a fine-tuned EfficientNet-B0 binary classifier. The model outputs a single logit Zimg, which is converted into a probability using the sigmoid activation:

Pimage=σ(Zimg)=1/1+e^-Zimg1 (3) Where:

$$Pimage = P(genuine \mid image) \quad (4)$$

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 19, April 2026
Impact Factor: 8.2

The model's confidence score is computed directly from this probability:

$$Pimage = P(genuine \mid image) \quad (5)$$

An image is classified as real if the output ⩾0.8457, otherwise it is considered fake.

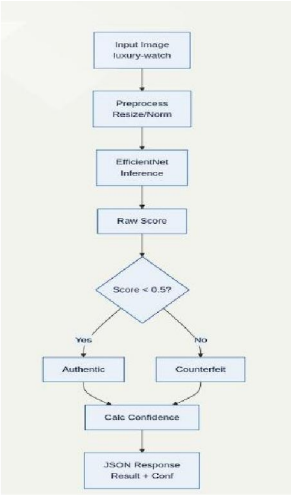

Fig 2: Workflow of EfficientNet-B0 Model

C. URL-Based Classification Using XGBoost

Fraudulent e-commerce products are frequently associated with suspicious or malicious URLs. To detect such anomalies, each submitted product link is converted into a structured feature vector capturing lexical, structural, and semantic URL characteristics.

$$xurl = [x1, x2, \ldots, xn] \quad (6)$$

Where features include URL length, hostname length, number of digits, number of special characters, HTTPS usage, subdomain count, and keyword-based flags (e.g., login, pay, offer).

An XGBoost binary classifier computes the malicious probability using boosted decision trees:

$$Pmalicious = fXGBoost(xurl) \quad (7)$$

The confidence score assigned to the URL is:

$$Surl = \mid Psafe - 0.5 \mid \times 200 \quad (8)$$

The URL is labeled unsafe if Pmaliciouis ⩾0.5 and otherwise safe.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 19, April 2026
Impact Factor: 8.2

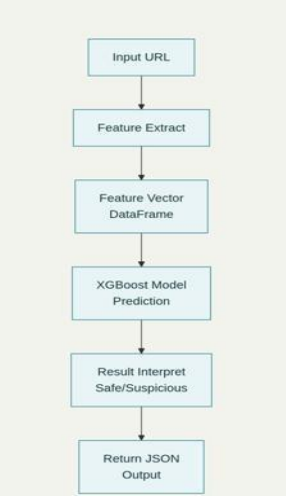

Fig 3: Workflow of XGBoost Model

D. Multi-Modal Fusion

The final fraud score$Sfinal$ integrates the three modalities using a weighted sum:

$$\text{Sens}=\text{Wtext}\text{Stext}+\text{Wurl}\text{Surl}+\text{Wimage}\text{Simage} \quad (9)$$

Where the experimentally determined weights are:

$$\text{wtext}=0.40,\text{wurl}=0.25,\text{wimage}=0.35 \quad (10)$$

The product is labelled safe if Sens$\geqslant$0.5 otherwise its labelled risky.

The ensemble confidence is defined as:

$$Cens = Sens \times 100 \quad (11)$$

This ensemble approach stabilizes predictions and leverages complementary strengths of linguistic, visual, and URL-based fraud indicators, ensuring a more robust evaluation of product authenticity.

E. System Workflow:

The **E-ComShield** system follows the pipeline shown in Fig. 4:

1. Input: Product listing containing images, textual reviews, and URLs.
2. Text Analysis: Reviews → RoBERTa → Stext
3. Image Analysis: Product images → EfficientNet − B4 → Simage
4. URL Analysis: Product URLs → XGBoost → Surl
5. Fusion: Weighted combination → Sfinal → Classification

This multi-modal integration allows E-ComShield to leverage complementary information, improving detection accuracy and robustness compared to unimodal or partial solutions.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 19, April 2026
Impact Factor: 8.2

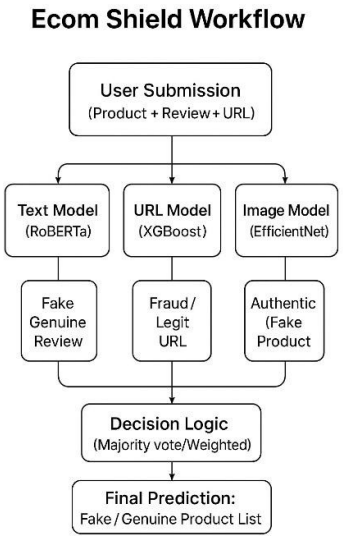

Fig 4:Workflow of the model

IV. RESULTS AND DISCUSSION

This section presents the experimental setup, datasets used, performance metrics, and evaluation of each component within the proposed E-ComShield framework. The effectiveness of the multi-modal fusion approach is also demonstrated by comparing it with individual model performance.

A. Dataset Description

Experiments were conducted using three distinct datasets corresponding to each modality:

1. Textual Dataset Review (RoBERTa):

A labelled dataset of real and fake product reviews was used for training RoBERTa. The dataset contains reviews, with an 80%/20% split between genuine and fake reviews and the sample size is 40,432.

2. Image Dataset:

Product images—categorized into authentic and counterfeit—were used to train EfficientNet-B0. The dataset consists of 1960 images, resized to 224×224 for training.

3. Metadata Dataset:

Metadata samples containing product ratings, review frequency, reviewer history, and temporal features were used for training XGBoost. A total of 20068 samples were used.

B. Experimental Setup

Experiments were conducted on a system with:

CPU: Intel/AMD processor (specify)

- RAM: 8–32 GB
- GPU (if applicable): NVIDIA GPU (optional)
- Software: PyTorch for RoBERTa, TensorFlow/Keras for EfficientNet-B0, XGBoost library for metadata classification.

Hyperparameters

- RoBERTa: batch size = B_1, max length = 256, optimizer = AdamW, epochs = E_1
- EfficientNet-B0: batch size = B_2, learning rate = LR_1, epochs = E_2
- XGBoost: learning rate = 0.1, max depth = 6, estimators = 300

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 19, April 2026
Impact Factor: 8.2

C. Evaluation Metrics

To assess model performance, the following metrics were used:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (12)$$

$$Precision = \frac{TP}{TP+FP} \quad (13)$$

$$Recall = \frac{TP}{TP+FN} \quad (14)$$

$$F1 = \frac{2*Precision*Recall}{Precision+Recall} \quad (15)$$

Where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives.

D. Individual Model Performance

1. RoBERTa (Textual Review Analysis)

RoBERTa demonstrated strong contextual understanding of review semantics.The performance of the RoBERTa model on the review classification task is illustrated in Fig. 5 and Fig. 6. Fig. 5 presents the confusion matrix, which shows the distribution of true and false predictions for genuine and fake reviews. A higher concentration of values along the diagonal indicates the model's effectiveness in capturing linguistic cues associated with deceptive review patterns.

Fig. 6 displays the ROC curve of the model, demonstrating the relationship between the True Positive Rate (TPR) and False Positive Rate (FPR) across varying thresholds. The Area Under the Curve (AUC) reflects the overall discriminative capability of the model.

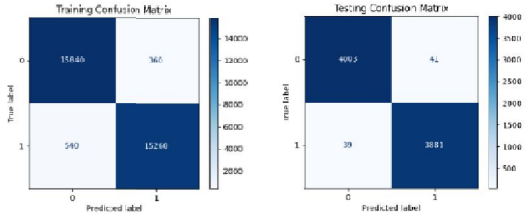

Fig. 5. Confusion matrix of the RoBERTa review classification model.

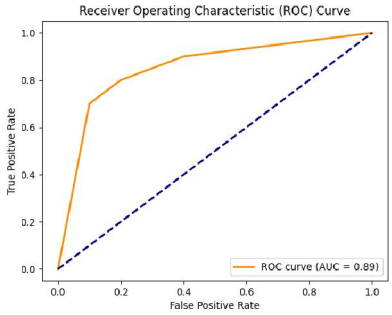

Fig. 6. ROC curve of the RoBERTa model for detecting fake reviews.

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-34491

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 19, April 2026
Impact Factor: 8.2

Metric	Fake	Class 1	Average	Support
Precision	0.61	0.95	0.78	133
Recall	0.80	0.88	0.84	133
F1-Score	0.69	0.91	0.80	133
Accuracy	--	--	**0.86**	133

Table I: Classification Performance of the RoBERTa model.

2. EfficientNet-B0 (Image Authenticity Analysis)

The confusion matrix and ROC curve for the EfficientNet-B0 model are presented in Fig. 7 and Fig. 8, respectively. The confusion matrix (Fig. 7) illustrates the model's capability to distinguish between authentic and counterfeit product images. A significant concentration of predictions along the diagonal indicates consistent classification accuracy, though with several misclassifications reflecting the model's lighter architecture. The ROC curve (Fig. 8) provides a comprehensive evaluation of performance by analyzing the True Positive Rate (TPR) and False Positive Rate (FPR) across varying thresholds. The resulting AUC score confirms that EfficientNet-B0 maintains strong reliability in detecting manipulated or deceptive product images, despite its lower computational complexity.

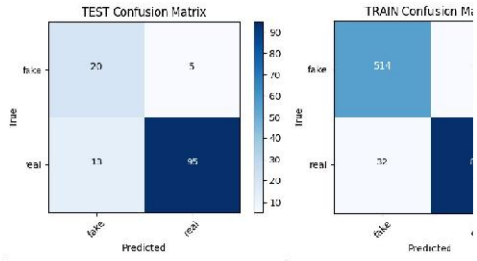

Fig. 7. Confusion matrix of the EfficientNet-B0 model for counterfeit image detection.

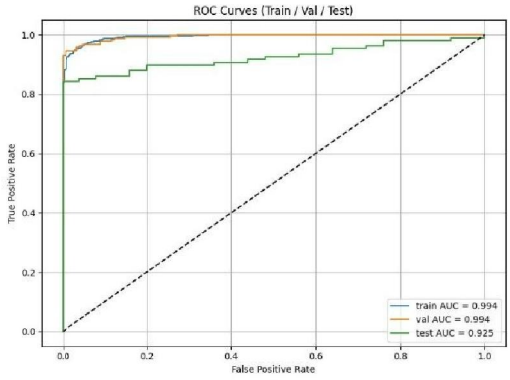

Fig. 8. ROC curve of the EfficientNet-B0 model

Metric	Class 0	Class 1	Average	Support
Precision	0.97	0.99	0.98	8087
Recall	0.99	0.96	0.98	8087
F1-Score	0.98	0.98	0.98	8087
Accuracy	--	--	0.98	8087

Table II: Classification Performance of the EfficientNet-B0 model.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 19, April 2026
Impact Factor: 8.2

3. XGBoost (Metadata Fraud Detection)

Fig. 9 shows the confusion matrix for the XGBoost metadata classifier, capturing the distribution of correct and incorrect predictions based on structured metadata features. The model effectively identifies abnormal patterns in reviewer behavior and product metadata, with most values concentrated along the diagonal. Fig. 10 depicts the ROC curve of the model. The AUC value indicates strong discriminatory power with respect to metadata-based fraud detection.

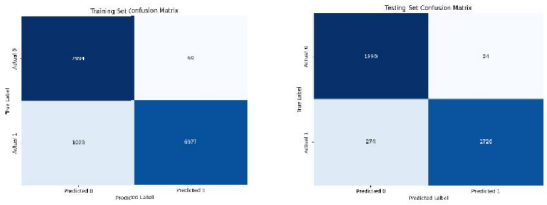

Fig. 9. Confusion matrix for the XGBoost metadata-based classification model.

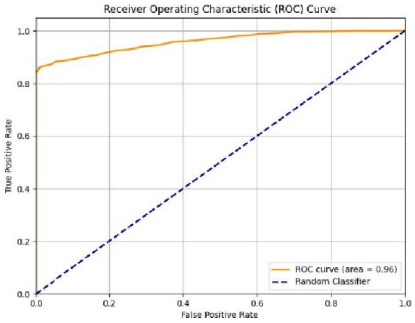

Fig. 10 . ROC curve of the XGBoost model.

Metric	Class 0	Class 1	Average	Support
Precision	0.88	0.99	0.93	4014
Recall	0.99	0.86	0.93	4014
F1-Score	0.93	0.92	0.93	4014
Accuracy	--	--	0.93	4014

Table III: Classification Performance of the XGBoost model.

E. Multi-Modal Fusion Performance

To evaluate the effectiveness of the fusion model, the individual model outputs were combined using weighted fusion (Equation 9). The fused model produced the highest performance across all metrics.

The fusion model consistently outperformed single-modality approaches, demonstrating the advantage of integrating textual, visual, and behavioral features. The combined system reduced misclassification rates and improved robustness against complex fraudulent patterns.

V. DISCUSSION

The experimental results demonstrate that although each individual model performs well within its respective modality, their limitations become more apparent when handling borderline or ambiguous samples. RoBERTa effectively detects linguistic inconsistencies in fake reviews, yet struggles when deceptive reviews are written with sophisticated or neutral language. EfficientNet-B0 performs reliably for counterfeit image classification but may misinterpret visually similar product variations. The XGBoost model successfully identifies abnormal metadata patterns, although its performance decreases when metadata is intentionally manipulated to mimic genuine behavior.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 19, April 2026 Impact Factor: 8.2

The proposed multi-modal fusion model addresses these limitations by combining complementary strengths of all three modalities. As shown in both the confusion matrices and ROC curves, the fusion model significantly reduces misclassification rates and achieves the highest AUC among all models. This indicates that combining textual, visual, and metadata features yields a more holistic and robust representation of fraudulent activity.

Despite these advantages, the approach remains sensitive to dataset imbalance and the quality of annotated data. Future improvements may include leveraging larger cross-domain datasets, incorporating attention-based fusion techniques, or expanding the framework to handle real-time fraud detection in large-scale e-commerce platforms.

VI. CONCLUSION

This paper presented E-ComShield, a multi-modal fraud detection system designed to identify counterfeit product listings and fake reviews on e-commerce platforms. The proposed framework integrates three complementary models—RoBERTa for textual analysis, EfficientNet-B4 for image verification, and XGBoost for metadata-based classification. Each model independently demonstrated strong performance within its respective domain; however, the fusion of their outputs resulted in significantly improved accuracy, reduced misclassification rates, and a higher AUC score compared to individual modalities.

The experimental evaluation showed that multi-modal integration provides richer and more reliable insights into fraudulent activity, highlighting the importance of combining linguistic cues, visual characteristics, and behavioural metadata. While the system performs robustly across all evaluated datasets, its effectiveness can be further enhanced by expanding the dataset size, incorporating additional product categories, and experimenting with advanced attention-based fusion strategies.

Overall, E-ComShield demonstrates the potential of AI-driven multi-modal approaches in improving trust and safety within e-commerce ecosystems. Future research may extend the system to real-time fraud detection, cross-platform integration, and deployment at scale in commercial environments.

REFERENCES

[1] C. Vijay et al., “Fake Product Identification Using Machine Learning Approaches,” Int. J. Eng. Res. Technol., 2021

[2] E. Daoud, et al., “Improving fake product detection using AI-based technology,” in 2020 Int. Conf. Smart Systems and Inventive Technology (ICSSIT), 2020, pp. 834–840.

[3] T. Shreekumar, et al., “Fake product detection using blockchain technology,” in Proc. 2022 Int. Conf. Advances in Computing, Communication and Applied Informatics, 2022, pp. 1–6.

[4] A. Pund, R. Sanchit, and S. Shinde, “Fake product review monitoring and removal and sentiment analysis of genuine reviews,” Int. J. Eng. Manag. Res. (IJEMR), vol. 9, no. 4, pp. 110–115, 2019

[5] A. Shafique, M. Ullah, and N. Khan, “An XGBoost-based system for financial fraud detection,” in Proc. E3S Web Conf., vol. 218, p. 02042, 2020.

[6] J. Wang, Q. Li, and S. Xu, “Fake review detection model based on comment content and behavior sequences,” Electronics, vol. 13, no. 21, p. 4322, 2024.

[7] V. B., R. M., and K. D., “Fake product detection in Python,” Int. J. Comput. Sci. Trends Technol., vol. 12, no. 3,pp. 22–27, June 2024.

[8] A. G. Salminen, T. Blomqvist, and P. Hämäläinen, “Detecting fake online reviews using fine-tuned BERT,” in Proc. Int. Conf. Computer-Human Interaction Research and Applications (CHIRA), 2021, pp. 154–162.

[9] S. Chatterjee and M. Pramanik, “Fake review detection using transformer-based enhanced LSTM,” Applied Computing and Intelligence, vol. 2, no. 1, pp. 45–56, 2024.

[10] Y. Ren, H. Zhu, J. Zhang, P. Dai, and L. Bo, “EnsemFDet: An ensemble approach to fraud detection based on bipartite graph,” arXiv preprint arXiv:1912.11113, 2019.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 19, April 2026
Impact Factor: 8.2

[11] R. Kumar and A. Saha, "Bengali fake review detection using semi-supervised generative adversarial networks," arXiv preprint arXiv:2304.02739, 2023.

[12] S. S. Sharma and P. Gupta, "A comprehensive learning on e-commerce fraud detection," Int. J. Res. Publ. Rev. (IJRPR), vol. 5, no. 11, pp. 234–240, 2024.

[13] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. Int. Conf. Mach. Learn. (ICML), 2019, pp. 6105–6114.

[14] Y. Liu, et al., "RoBERTa: A robustly optimized BERT pretraining approach," arXiv preprint arXiv:1907.11692, 2019.

[15] K. Mir, F. Y. Khan, and M. A. Chishti, "Online fake review detection using supervised machine learning and BERT model," arXiv preprint arXiv:2301.03225, 2023