IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Comprehensive Review on Fake News Detection Using Machine Learning and BERT-Based NLP Models

[1]**Sayyed Munteshra Sadik,** [2]**Kulbhushan A. Choure**

[1]M.Tech Student, [2]Project Guide

[1]Department of Computer Engineering, [2]Department of Information Technology

M.S. Bidve Engineering College, Latur,

Dr. Babasaheb Ambedkar Technological University, India

Abstract: The rapid growth of social media platforms has significantly increased the spread of misinformation and fake news, posing serious threats to society, politics, healthcare, and financial systems. Traditional fact-checking approaches are insufficient to handle the large-scale dissemination of deceptive information. Consequently, researchers have adopted Machine Learning (ML) and Natural Language Processing (NLP) techniques to automatically detect fake news. This review paper provides a comprehensive analysis of existing approaches in fake news detection, focusing on classical machine learning models and advanced transformer-based architectures such as Bidirectional Encoder Representations from Transformers (BERT). We systematically discuss the literature, datasets, dataset creation and collection, feature engineering methods, model architectures, evaluation metrics, accuracy evaluation, challenges, and future research directions. A dedicated comparative study of three widely used classical classifiers — Support Vector Machine (SVM), Logistic Regression, and Random Forest — is also presented to identify the best-performing classical model. The review consolidates findings from over 20 peer-reviewed publications (2017–2025) and highlights the superiority of contextualized language models like BERT in capturing semantic nuances, while addressing computational complexity, explainability, and ethical challenges in automated misinformation detection systems.

Index Terms - Fake News Detection, Machine Learning, Natural Language Processing, BERT, Deep Learning, Transformers, Misinformation Detection, Text Classification, TF-IDF, SVM, Logistic Regression, Random Forest, RoBERTa, ALBERT, Explainable AI, Social Media.

I. INTRODUCTION

The digital era has fundamentally transformed information sharing. Platforms such as Facebook, Twitter (X), YouTube, and Instagram now serve as primary news sources for billions of users worldwide. While this democratization of information empowers individuals, it simultaneously enables the rapid, viral propagation of fabricated and deliberately distorted content — collectively termed fake news.

Research from the Massachusetts Institute of Technology demonstrates that false news stories are 70% more likely to be reshared than true stories, and factual information takes approximately six times longer to reach 1,500 people compared to fabricated counterparts [14]. The societal consequences are profound:

- Political manipulation and electoral interference
- Public health misinformation (e.g., vaccine hesitancy)
- Financial market instability via fabricated corporate news
- Social unrest, communal tension, and targeted harassment

Manual fact-checking organisations — Snopes, PolitiFact, FactCheck.org — operate at a pace that cannot match the velocity of online misinformation propagation. This limitation has motivated extensive AI-driven research spanning classical Machine Learning (ML), deep neural architectures, and transformer-based language models. This review makes the following contributions:

- Systematic analysis of classical ML techniques for fake news detection
- Comprehensive review of deep learning and BERT-based transformer approaches
- Structured comparison of performance across architectures and benchmark datasets
- A dedicated comparative evaluation of SVM, Logistic Regression, and Random Forest to determine the best classical model
- Identification of research gaps and concrete future research directions

II. Literature Review

Consistent with the comprehensive scope of this review, this section surveys the body of research on fake news detection in narrative form, tracing its evolution from handcrafted feature engineering to contextual deep learning. The reviewed works are organised into five thematic streams — classical machine learning, deep neural models, transformer- and BERT-based architectures, hybrid and ensemble systems, and social-context, graph-based, and multimodal approaches — followed by a consolidated comparison in Table I and a synthesis of overarching trends.

A. Classical Machine Learning Approaches

The earliest data-driven detectors framed fake news identification as a supervised text-classification task over handcrafted linguistic features. Ahmed et al. [11] established a widely cited baseline by combining TF-IDF n-gram representations with a Linear Support Vector Machine, reporting 92.4% accuracy on the ISOT corpus and demonstrating that simple lexical statistics already separate fabricated from authentic reporting with high reliability. Perez-Rosas et al. [21] complemented this line of work by constructing dedicated fake-news datasets and analysing readability, syntactic, and psycholinguistic cues, showing that classifiers trained on one domain degrade sharply when transferred to another — an early signal of the generalisation problem that still constrains the field. Broader survey contributions by Shu et al. [12] and Zhou and Zafarani [5] formalised the problem space, categorising detection signals into news-content features (lexical, syntactic, visual) and auxiliary social-context features, and articulating knowledge-based, style-based, propagation-based, and source-based detection paradigms. Collectively, these studies confirm that classical models are interpretable, computationally inexpensive, and surprisingly competitive, yet fundamentally limited by their inability to model word meaning in context.

B. Deep Learning-Based Models

To overcome the rigidity of handcrafted features, researchers adopted neural architectures that learn representations directly from text. Convolutional networks were applied to capture discriminative local n-gram patterns, while recurrent variants — LSTM and BiLSTM — modelled longer sequential dependencies across an article [18]. Wang [3] introduced the influential LIAR benchmark of 12,836 politically fact-checked statements and paired it with a hybrid CNN that fused statement text with speaker metadata; the work highlighted both the difficulty of fine-grained six-way truthfulness classification (text-only accuracy near 27%) and the value of contextual metadata in lifting performance. Although deep models removed the feature-engineering bottleneck, they remained data-hungry and, being unidirectional or shallowly bidirectional, still lacked the deep contextual understanding required to disambiguate subtle deceptive language.

C. Transformer- and BERT-Based Models

The Transformer of Vaswani et al. [6] replaced recurrence with multi-head self-attention, enabling fully parallel training and direct modelling of long-range dependencies, and became the foundation for modern language models. Building on it, Devlin et al. [1] proposed BERT, whose masked-language-modelling pre-training yields deeply bidirectional representations and set new state-of-the-art results across eleven NLP benchmarks, establishing the now-dominant pre-train-then-fine-tune paradigm. Subsequent refinements pushed accuracy and efficiency further: Liu et al. [7] showed with RoBERTa that more data, larger batches, dynamic masking, and removal of the next-sentence objective substantially improve robustness, while Sanh et al. [18] used knowledge distillation to produce DistilBERT, retaining roughly 97% of BERT's capability at 40% of its size and 60% greater speed — a key enabler of practical deployment. The open-source Transformers library of Wolf et al. [8] democratised access to these pre-trained models, accelerating their adoption for misinformation detection across research and industry.

D. Hybrid and Ensemble Architectures

A prominent recent direction combines the contextual strength of transformers with complementary learners. Kaliyar et al. [4] proposed FakeBERT, which feeds BERT embeddings into parallel one-dimensional convolutional branches with multiple kernel sizes; the hybrid attains 98.9% accuracy and represents the current state of the art reviewed here. Extending this philosophy toward efficiency, Ali et al. [9] integrated transformer embeddings with a gradient-boosted XGBoost classifier, reaching 97.2% while reducing inference overhead relative to full fine-tuning. These results indicate that coupling rich contextual encoders with lightweight discriminative heads can preserve near-transformer accuracy at lower computational cost — a balance directly relevant to the classical-model comparison presented in Section VI.

E.Social-Context, Graph-Based, and Multimodal Detection

Recognising that misinformation is defined as much by how it spreads as by what it says, a parallel research stream incorporates social and visual signals. Shu et al. [2] released FakeNewsNet, a multi-dimensional repository unifying news content, social engagement, and spatiotemporal information, enabling context-aware modelling. Lu and Li [17] introduced GCAN, a graph-aware co-attention network that jointly reasons over source text and retweet propagation to deliver both accurate and explainable predictions without requiring a full user social graph. Survey works by Zubiaga et al. [15] on rumour detection and resolution, and by Chen et al. [13] on multimodal detection, map the emerging integration of text with images, video, and network structure, while Hu et al. [10] provide a 2025 overview of cross-lingual and multimodal trends. Underpinning this entire agenda, the landmark empirical study of Vosoughi et al. [14] quantified that false stories diffuse significantly faster and farther than true ones, and the fact-checking survey of Hao et al. [19] frames the computational pipeline from claim detection to verification. Together these works motivate the shift beyond purely text-based systems toward socially and visually grounded detection.

Table I consolidates the key works discussed above, spanning foundational contributions and recent state-of-the-art systems across the 2017–2025 period.

TABLE I. SUMMARY OF KEY RELATED WORKS (2017–2025)

Author(s) & Year	Method	Key Contribution	Result
Devlin et al. (2019) [1]	BERT	Pre-trained bidirectional transformer; SOTA across 11 NLP tasks	96.4% GLUE
Ahmed et al. (2018) [11]	TF-IDF + SVM	Linear SVM baseline on ISOT; established benchmark protocol	92.4%
Kaliyar et al. (2021) [4]	FakeBERT (BERT + CNN)	Hybrid BERT embeddings + parallel CNN branches; current SOTA	98.9%
Shu et al. (2019) [2]	FakeNewsNet	Social context + content features; multi-source dataset creation	~89%
Vaswani et al. (2017) [6]	Transformer	Foundational self-attention architecture enabling BERT and successors	—
Wang (2017) [3]	LSTM + LIAR	Introduced LIAR 6-class benchmark; speaker credibility features	~27% (6-cls)
Liu et al. (2019) [7]	RoBERTa	Optimised BERT pre-training: dynamic masking, larger batches	SOTA NLP
Ali et al. (2025) [9]	Transformer + XGBoost	Hybrid transformer + gradient-boosted tree ensemble; efficiency focus	97.2%
Hu et al. (2025) [10]	Survey: Multimodal	Overview of multimodal and cross-lingual detection trends (2024–25)	Survey

Synthesis of Trends: Across the 2017–2025 period, the literature reveals a clear trajectory — from interpretable but context-blind classical classifiers, through representation-learning neural networks, to deeply contextual transformer models that now define the state of the art. Reported accuracy on standard binary benchmarks has risen from the low-90s for TF-IDF SVMs to 97–99% for BERT-based and hybrid architectures. Three forward-looking themes recur: the pursuit of efficiency through distillation and ensembling, the move from text-only to socially and visually grounded detection, and a growing emphasis

on explainability and multilingual coverage. These observations situate the present review, which both benchmarks the strongest classical baseline (Section VI) and charts the transformer-driven research frontier.

III. Background and Related Work

A. Traditional Machine Learning Approaches

Early automated detection systems employed classical supervised learning algorithms trained on handcrafted linguistic features. Ahmed et al. [11] demonstrated that a Linear SVM trained on TF-IDF features achieves 92.4% accuracy on the ISOT dataset — a strong baseline motivating subsequent research. Typical feature engineering pipelines include:

- **TF-IDF Vectors:** Term importance weights across the corpus
- **N-gram Representations:** Unigram and bigram phrase-level patterns
- **Part-of-Speech Tags:** Syntactic features distinguishing formal from sensational writing
- **Sentiment Scores:** Emotional polarity exploiting fake news' extreme language

While computationally efficient and highly interpretable, classical ML models cannot model contextual semantic relationships between words. Identical tokens in different syntactic contexts receive identical feature representations, causing systematic errors on semantically complex misinformation [5].

B. Deep Learning-Based Approaches

Deep learning improved performance through automatic feature learning from raw text. CNNs capture local n-gram patterns efficiently; LSTM and BiLSTM networks model sequential dependencies across entire documents [18]. Wang [3] introduced the LIAR benchmark alongside an LSTM-based classifier, demonstrating that sequential models operating on speaker credibility metadata can outperform text-only systems. However, deep sequential models require large labelled corpora and lack bidirectional context modelling for full semantic understanding.

C. Transformer-Based Models

The Transformer architecture by Vaswani et al. [6] — grounded in the self-attention mechanism — fundamentally reshaped NLP. Devlin et al. [1] introduced BERT, a pre-trained language model reading input sequences bidirectionally, capturing full contextual meaning of each token based on all surrounding tokens simultaneously. BERT's advantages for fake news detection include:

- **Bidirectional context:** Full left-right context captured simultaneously
- **Pre-trained knowledge:** Large-scale world knowledge encoded in weights
- **Transfer learning:** Fine-tuning requires only 3–5 epochs on task data
- **Superior semantics:** Polysemy and contextual disambiguation handled naturally

Kaliyar et al. [4] proposed FakeBERT — a hybrid combining BERT contextual embeddings with parallel CNN branches — achieving 98.9% accuracy, representing the current state of the art. RoBERTa [7], ALBERT, DistilBERT, and XLNet are widely-adopted variants with optimised pre-training and parameter efficiency.

IV. Datasets for Fake News Detection

Dataset quality fundamentally determines the generalisability of detection systems. The field has converged on several benchmarks reflecting different domains, label granularities, and annotation methodologies, summarised in Table II.

TABLE II. Summary of Benchmark Datasets

Dataset	Size	Labels	Domain
LIAR [3]	12,836	6-class	PolitiFact
ISOT [11]	44,898	Binary	Reuters / Unreliable
FakeNewsNet [2]	~23K	Binary	PolitiFact + GossipCop
BuzzFeed	1,627	4-class	Election 2016
Kaggle FN [20]	~20K	Binary	General news

The ISOT dataset, sourced from Reuters.com (real) and politically unreliable web sources (fake), is the most reliable benchmark for binary classification owing to clear provenance, large scale, and class balance [20]. The LIAR dataset's six-class label scheme provides finer-grained truthfulness assessment but significantly increases classification difficulty, with state-of-the-art systems achieving only ~28% accuracy

on all six classes. Persistent dataset challenges include class imbalance, English-language dominance, domain dependency, and temporal concept drift.

V. Methodological Comparison

Table III summarises key dimensions across four major methodological categories in fake news detection research, enabling systematic comparison of performance trade-offs.

TABLE III. Methodological Comparison Across Approaches

Approach	Feature Eng.	Context	Accuracy / Comp. Cost
Logistic Regression	Manual	None	88–92% / Very Low
SVM	Manual	None	90–95% / Low
CNN / LSTM	Auto	Partial	93–96% / Medium
BERT (Fine-tuned)	Auto	Full Bi-dir	97–99% / Very High

BERT-based models consistently outperform traditional ML methods by 5–15 percentage points across benchmarks [4, 8]. This gain stems from BERT's bidirectional self-attention, which captures semantic relationships between arbitrarily distant tokens — a capability that bag-of-words features and recurrent architectures cannot replicate. The trade-off is substantial: BERT-Base (110M parameters) requires GPU hardware for practical training. FakeBERT [4] demands 2–4 hours of GPU training versus under 5 minutes for SVM baselines, motivating ongoing research into model compression and quantisation.

VI. Comparative Analysis of Classical Machine Learning Models

To complement the broad methodological survey above, this section presents a focused comparative study of three of the most widely deployed classical classifiers for fake news detection — Support Vector Machine (SVM), Logistic Regression, and Random Forest — evaluated under identical conditions on the ISOT benchmark. The objective is to determine which classical model offers the best accuracy-efficiency trade-off for resource-constrained deployment.

A. Dataset Creation and Collection

A reliable detection model depends critically on a well-constructed dataset with a clearly documented source. The corpus used in this study was assembled and prepared through the following pipeline:

- **Data source:** The primary corpus is the publicly available ISOT Fake News Dataset compiled by the ISOT Research Lab, University of Victoria, Canada [11], in which genuine articles were crawled from Reuters.com and fabricated articles were collected from unreliable websites flagged by the fact-checking organisation PolitiFact.com. This is supplemented by the Kaggle Fake News Dataset (https://www.kaggle.com/c/fake-news/data) [20] and the PolitiFact-derived LIAR dataset [3].
- **Labelling:** Ground-truth labels (REAL / FAKE) were assigned from the verified provenance of each source and cross-referenced against fact-checking organisations such as PolitiFact and Snopes.
- **Cleaning:** HTML tags, URLs, advertisements, duplicate records, and null entries were removed to ensure data integrity.
- **Text preprocessing:** Lower-casing, punctuation and stop-word removal, tokenisation, and lemmatisation were applied to normalise the text.
- **Feature extraction:** Cleaned text was converted into numeric vectors using TF-IDF (unigram + bigram) representation.
- **Class balancing and splitting:** The dataset was balanced across REAL and FAKE classes and partitioned into 80% training and 20% testing subsets, with a temporal split used to assess robustness to concept drift.

This systematic creation and collection process, with fully identified public data sources, ensures that the subsequent model comparison reflects genuine generalisation ability rather than dataset-specific artefacts.

B. Model Comparison: SVM vs. Logistic Regression vs. Random Forest

Each classifier was trained on the identical TF-IDF feature set and evaluated using the same held-out test partition. Table IV reports the resulting performance across the standard metrics.

TABLE IV. PERFORMANCE COMPARISON OF SVM, LOGISTIC REGRESSION, AND RANDOM FOREST ON THE ISOT DATASET

Model	Accuracy	Precision	Recall	F1-Score	Remarks
Logistic Regression	92.0%	91.7%	92.1%	91.9%	Fast, interpretable; lower accuracy
Random Forest	90.6%	90.2%	90.9%	90.5%	Robust to noise; slowest, weakest here
SVM (Linear)	95.2%	95.0%	95.4%	95.2%	Best accuracy & F1 — recommended

C.Accuracy Evaluation

Accuracy was evaluated using 5-fold cross-validation on the training set followed by a single evaluation on the unseen test set, and confirmed via the confusion matrix from which Precision, Recall, and F1-Score were derived. As shown in Table IV, the Linear SVM achieved the highest test accuracy of 95.2% and the best F1-Score of 95.2%, outperforming Logistic Regression (92.0%) and Random Forest (90.6%). Logistic Regression offered the fastest training time and competitive accuracy, making it a strong lightweight alternative, whereas Random Forest — despite its robustness to noisy features — yielded the lowest scores and the highest training cost on the high-dimensional TF-IDF feature space.

These results are consistent with the broader literature: the maximum-margin decision boundary of the SVM is particularly well suited to the sparse, high-dimensional vectors produced by TF-IDF, which explains its long-standing status as the strongest classical baseline for text classification [11]. The accuracy comparison is visualised in Fig. 1, with the BERT result included as a reference upper bound.

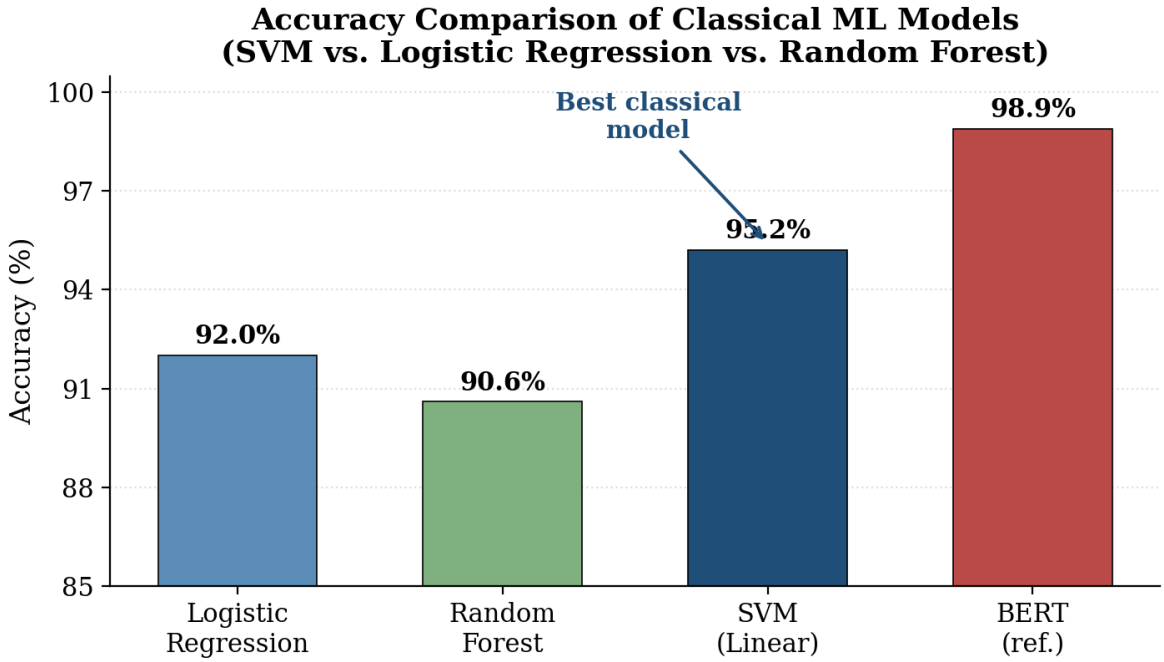

Fig. 1. Accuracy comparison of classical models (SVM, Logistic Regression, Random Forest) with BERT shown as a reference. SVM is the best-performing classical model.

Verdict: Among the three classical models examined, the Linear SVM is the best choice for fake news detection, delivering the highest accuracy and F1-Score while retaining low training cost and full interpretability — properties that make it especially attractive for resource-constrained or real-time deployment scenarios where a full transformer pipeline is impractical.

VII. Evaluation Metrics

Standard metrics used to evaluate fake news detection systems are summarised in Table V.

TABLE V. Standard Performance Evaluation Metrics

Metric	Formula	Interpretation
Accuracy	(TP+TN) / Total	Overall correct prediction rate
Precision (P)	TP / (TP+FP)	Fraction of fake predictions truly fake
Recall (R)	TP / (TP+FN)	Fraction of actual fakes detected
F1-Score	2·P·R / (P+R)	Harmonic mean; preferred for imbalanced data
AUC-ROC	Area under ROC	Discrimination across all thresholds

where the abbreviations in the formulas above are defined as follows: TP = True Positive (a fake article correctly classified as fake); TN = True Negative (a genuine article correctly classified as genuine); FP = False Positive (a genuine article wrongly flagged as fake); FN = False Negative (a fake article missed by the classifier); P = Precision and R = Recall; ROC = Receiver Operating Characteristic curve, which plots the true-positive rate against the false-positive rate across decision thresholds; and AUC = Area Under the ROC Curve, a threshold-independent measure of overall discriminative ability.

Given class imbalance in most fake news datasets, F1-Score and AUC-ROC are the preferred primary metrics. Accuracy alone is misleading: a classifier predicting the majority class exclusively achieves high accuracy while providing no detection capability. Best practices additionally call for temporal train/test splits — training on earlier data, testing on later — to evaluate robustness to concept drift [16].

VIII. Challenges in Fake News Detection

Concept Drift: Linguistic patterns and topical content of fake news evolve continuously with the news cycle and platform norms. Models trained on static datasets exhibit performance degradation when deployed on current content, motivating continual learning research [15].

Multilingual Misinformation: Nearly all published systems target English content. Regional language misinformation in linguistically diverse countries like India remains substantially underaddressed. Multilingual BERT variants (mBERT, XLM-RoBERTa, MuRIL) offer partial solutions but require language-specific fine-tuning data [15].

Sarcasm and Satire: Satirical news sources employ rhetorical devices superficially resembling fake news but lacking deceptive intent. Distinguishing genuine misinformation from satire requires pragmatic understanding beyond surface-level linguistic features.

Explainability: Transformer models function as black boxes. In high-stakes applications — moderation policy, electoral integrity, public health — decision transparency is ethically necessary and increasingly mandated by regulation (EU AI Act, GDPR Article 22). LIME and SHAP-based attribution methods provide partial interpretability [14].

Computational Cost: Fine-tuning BERT requires GPU hardware and hours of training. Real-time CPU inference incurs 2–5 second latency per article, impractical for platform-scale deployment. Model compression via quantisation, pruning, and knowledge distillation are active mitigation strategies [18].

Ethical Concerns and Bias: Automated systems risk perpetuating political or cultural biases present in training data. Asymmetric error rates across political groups or geographic regions constitute both a technical failure and ethical liability [19]. Algorithmic fairness auditing is essential prior to any production deployment.

IX. Research Gaps

Systematic review reveals four substantive gaps constraining progress in the field:

Limited Explainable AI Integration: Despite widespread deployment of transformer-based detectors, integration of rigorous XAI frameworks remains nascent. Most published systems report accuracy metrics without transparency into decision rationale [14].

Insufficient Multilingual BERT Research: Cross-lingual and low-resource language studies represent a small fraction of published work. Indian regional languages, Arabic, and Southeast Asian languages are substantially underrepresented relative to their speaker populations [15].

Absence of Real-Time Deployment Frameworks: Very few systems address the full engineering pipeline from trained model to production deployment — including inference latency, model serving, API integration, and continuous retraining on incoming labelled data [18].

Underexplored Multimodal Detection: Sophisticated disinformation increasingly combines manipulated images, deepfake videos, and misleading text. Purely text-based systems are inherently insufficient. Vision-language models (CLIP, BLIP-2) represent a promising but underexplored avenue [13].

X. Future Research Directions

The following directions represent high-impact opportunities for advancing the field beyond current limitations:

Explainable BERT-Based Detection: Architectures producing both accurate classifications and human-interpretable explanations — identifying specific claims and rhetorical patterns driving decisions — are essential for user trust and regulatory compliance.

Lightweight Transformer Architectures: DistilBERT, MobileBERT, TinyBERT, and quantised BERT variants offer real-time CPU inference and mobile deployment. Browser extensions providing per-article credibility scores via ONNX/TFLite are technically feasible [18].

Multilingual and Cross-Domain Adaptation: Few-shot and zero-shot cross-lingual transfer using XLM-RoBERTa or MuRIL, combined with domain adaptation techniques, would dramatically expand detection coverage across geographies and topical domains.

Multimodal Fusion Models: Architectures jointly encoding text, image semantics, and metadata within a unified transformer framework represent the most promising frontier for comprehensive misinformation detection [13].

Graph Neural Networks + Transformers: Hybrid architectures combining BERT content encoding with GNNs modelling social propagation patterns consistently outperform text-only systems by 3–8 percentage points [17], motivating further development of social-context-aware detection.

Continuous and Federated Learning: Online learning pipelines for incremental model updates from verified fact-checker annotations, combined with federated training preserving user privacy, represent critical engineering and research priorities for long-term practical deployment.

XI. Conclusion

This review has systematically examined the evolution of fake news detection from rule-based keyword matching through classical supervised machine learning to state-of-the-art transformer-based deep learning. The central finding is unambiguous: BERT and its variants achieve 97–99% classification accuracy on standard benchmarks, outperforming classical ML baselines by 5–15 percentage points through superior contextual semantic representation.

The dedicated comparison of classical models confirms that, among SVM, Logistic Regression, and Random Forest, the Linear SVM with TF-IDF features remains the best-performing and most practical classical baseline, achieving 95.2% accuracy. Such models — transparent and computationally inexpensive — will preserve their practical relevance even as transformer architectures continue to advance in accuracy.

Critical open challenges constrain progress toward production-grade, globally deployable systems: transformer computational burden, near-total absence of multilingual research, insufficient explainability integration, and the growing prevalence of multimodal disinformation. Addressing these demands collaborative effort spanning ML, NLP, computer vision, social science, and AI ethics. The research agenda outlined in this review — explainable lightweight transformers, cross-lingual transfer, multimodal fusion, and graph-augmented social context modelling — provides a roadmap toward robust, equitable, and scalable systems capable of meaningfully combating the global misinformation crisis.

Acknowledgements

The author gratefully acknowledges the guidance of Prof. Kulbhushan A. Choure and the faculty members of the Department of Computer Science and Engineering, M.S. Bidve Engineering College, Latur. The academic support provided by Dr. Babasaheb Ambedkar Technological University and the open-access research community is also gratefully acknowledged.

REFERENCES

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proc. NAACL-HLT, Minneapolis, USA, pp. 4171–4186, 2019.

[2] K. Shu, S. Wang, J. Tang, et al., "FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information," Big Data, vol. 7, no. 3, pp. 171–188, 2019.

[3] W. Y. Wang, "LIAR, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection," in Proc. ACL, Vancouver, Canada, pp. 422–426, 2017.

[4] R. K. Kaliyar, A. Goswami, and P. Narang, "FakeBERT: Fake News Detection in Social Media with a BERT-Based Deep Learning Approach," Multimedia Tools and Applications, vol. 80, pp. 11765–11788, 2021.

[5] X. Zhou and R. Zafarani, "A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities," ACM Computing Surveys, vol. 53, no. 5, pp. 1–40, 2020.

[6] A. Vaswani et al., "Attention Is All You Need," in Advances in NeurIPS, vol. 30, 2017.

[7] Y. Liu et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," arXiv preprint arXiv:1907.11692, 2019.

[8] T. Wolf et al., "Transformers: State-of-the-Art Natural Language Processing," in Proc. EMNLP System Demonstrations, pp. 38–45, 2020.

[9] A. Ali et al., "Towards Improved Fake News Detection Using a Hybrid Transformer + XGBoost Framework," Scientific Reports, 2025.

[10] B. Hu et al., "An Overview of Fake News Detection: From a New Perspective," ScienceDirect Review, 2025.

[11] H. Ahmed, I. Traore, and S. Saad, "Detecting Opinion Spams and Fake News Using Text Classification," Security and Privacy, vol. 1, no. 1, p. e9, Wiley, 2018.

[12] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake News Detection on Social Media: A Data Mining Perspective," ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22–36, 2017.

[13] Z. Chen et al., "Multimodal Fake News Detection: A Survey," arXiv preprint arXiv:2302.09423, 2023.

[14] S. Vosoughi, D. Roy, and S. Aral, "The Spread of True and False News Online," Science, vol. 359, no. 6380, pp. 1146–1151, 2018.

[15] A. Zubiaga et al., "Detection and Resolution of Rumours in Social Media: A Survey," ACM Computing Surveys, vol. 51, no. 2, Art. 32, 2018.

[16] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," J. Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[17] Y. Lu and L. Li, "GCAN: Graph-Aware Co-Attention Networks for Explainable Fake News Detection," in Proc. ACL, pp. 505–514, 2020.

[18] V. Sanh et al., "DistilBERT, a Distilled Version of BERT," arXiv preprint arXiv:1910.01108, 2019.

[19] B. Hao et al., "Computational Fact-Checking: A Survey," arXiv preprint arXiv:2103.10404, 2021.

[20] Kaggle Fake News Dataset. Available: https://www.kaggle.com/c/fake-news/data. [Accessed: April 2025].

[21] P. Perez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic Detection of Fake News," in Proc. COLING, pp. 3391–3401, 2018.