Literature Review of Fake News Detection Papers

PAPER 1 : Fake News Detection Using Deep Learning Techniques

1. Methodology

The paper compares Logistic Regression (LR), Naïve Bayes (NB), Support Vector Machine (SVM), Random Forest (RF), and Deep Neural Networks (DNN) for fake news detection. The process includes dataset collection, text preprocessing (stemming, stopword removal, NLP feature extraction), and model training/testing to classify news as real or fake.

2. Novelty

- First-time comparative analysis of ML and deep learning models for fake news detection.
- DNN surpasses traditional ML models by capturing complex patterns.
- NLP-based preprocessing improves classification accuracy.

3. Advantages

- DNN achieves the highest accuracy (91%), outperforming other models.
- Fast processing time compared to conventional ML techniques.
- Effective feature extraction using NLP techniques enhances detection.

4. Metrics

- Accuracy: DNN (91%) ¿ NB (89%) ¿ SVM (79%) ¿ RF (77%) ¿ LR (75%).
- Time Efficiency: DNN (400ms) is the fastest, while LR (3750ms) is the slowest.
- Memory Usage: DNN consumes the most memory.

5. Limitations

- High memory requirement for deep learning models.
- Dependence on dataset quality for accuracy.
- Limited real-time adaptability in detecting evolving fake news patterns.

6. Future Scope

- Enhancing real-time detection for social media platforms.
- Hybrid models combining ML and DL for improved efficiency.
- Incorporating fact-checking sources for better accuracy.

This study highlights DNN as the most effective model despite its resource-intensive nature, paving the way for future advancements in real-time fake news detection.

PAPER 2 : Fake News Detection Using Machine Learning and Deep Learning Algorithms

1. Methodology

The study applies text classification techniques to detect fake news using four feature extraction methods: TF-IDF, Count Vector, N-Gram Level Vector, and Character Level Vector. Ten different machine learning (e.g., Random Forest, SVM, Naïve Bayes) and deep learning classifiers (e.g., CNN+LSTM, ANN) are tested. The dataset used comes from Kaggle with over 7,796 news articles, classified as real or fake.

2. Novelty

- Introduces CNN+LSTM for fake news detection, achieving 100% accuracy.
- Comprehensive comparison of ML vs. DL models on multiple feature extraction methods.
- Uses multiple vectorization techniques to enhance feature representation.

3. Advantages

- High accuracy (81-100%), with CNN+LSTM and AdaBoost reaching 100%.
- TF-IDF and Character-Level Vectorization improve classification performance.
- Combines multiple classifiers for robust evaluation.

4. Metrics

- Best Accuracy: CNN+LSTM (100%), AdaBoost (100%), Logistic Regression (93.37%).
- Feature Extraction Impact: Character Level Vectorization yields better results than N-Gram.
- Comparison with Previous Studies: Outperforms traditional classifiers in fake news detection.

5. Limitations

- Focuses only on text-based fake news, ignoring image or video-based misinformation.
- High computational cost for deep learning models.
- Dataset dependence: Performance may vary on different datasets.

6. Future Scope

- Extending fake news detection to multimedia content (images, videos).
- Hybrid ML-DL approaches to optimize accuracy and efficiency.
- Real-time fake news detection for social media applications.

This study highlights CNN+LSTM as the most effective model, proving deep learning's potential in automated misinformation detection.

PAPER 3 : Fake News Detection Using Machine Learning Approaches (Khanam et al., 2021)

1. Introduction

This paper explores the growing concern of fake news dissemination on social media and online platforms. Fake news has serious social, political, and economic consequences. Traditional manual fact-checking methods cannot keep up with the massive volume of online misinformation. This study aims to develop an automated fake news detection model using machine learning algorithms such as Naïve Bayes, SVM, Decision Trees, and Random Forests.

2. Related Work

2.1 Fake News on Social Media

- Social media platforms, including Facebook, Twitter, and Reddit, contribute significantly to fake news spread.
- Studies show that fake news spreads faster than real news due to sensationalism and user engagement.
- Tacchini et al. (2017) classified 15,500 Facebook posts using Logistic Regression and achieved 99% accuracy.

2.2 Role of Natural Language Processing (NLP)

- Sentiment analysis is widely used to determine the emotional tone of news articles.
- Named Entity Recognition (NER) helps detect entities like names and places in news articles.
- Part-of-Speech (POS) tagging improves feature extraction for machine learning models.

2.3 Data Mining for Fake News Detection

- Supervised Learning: Uses labeled datasets to train classifiers.
- Unsupervised Learning: Detects patterns in unlabeled data, such as clustering similar news articles.

2.4 Machine Learning Classifiers Used in Fake News Detection

Algorithm	Strengths	Weaknesses
Decision Tree	Simple, interpretable	Prone to overfitting
Random Forest	Reduces overfitting, improves accuracy	Computationally expensive
Support Vector Machine (SVM)	Works well for high-dimensional data	Requires feature selection
Naïve Bayes	Fast and efficient for text classification	Assumes feature independence
K-Nearest Neighbors (KNN)	Simple to implement	Slow for large datasets

3. Challenges in Fake News Detection

- Context Understanding: Sarcasm and satire can confuse models.
- Data Imbalance: More real news articles exist than fake ones, affecting classifier performance.
- Real-Time Detection: Models must be fast enough for real-time news verification.

4. How This Study is Different from Existing Research

Feature	**Previous Research**	**This Study**
Model Complexity	Focused on deep learning	Evaluates multiple ML classifiers
Feature Extraction	Used only TF-IDF	Uses POS tagging, sentiment analysis, and lexical features
Computational Efficiency	High resource requirements	Optimized for real-time use

5. Conclusion

The study evaluates multiple machine learning models for fake news detection and identifies Random Forest and SVM as the most effective classifiers. It highlights the need for hybrid models combining statistical techniques and NLP for better performance.

PAPER 4 : Fake News Detection Using Machine Learning

Introduction

The paper explores the issue of fake news detection using computational techniques, particularly linear classification models. It reviews existing research on natural language processing (NLP), machine learning (ML), and statistical approaches in fake news identification. The literature review highlights key advancements, challenges, and gaps in the field.

Existing Approaches to Fake News Detection

Several researchers have proposed different methods for detecting fake news. The primary techniques discussed in the paper include:

1. Natural Language Processing (NLP)-Based Approaches

- Studies by [Author et al.] focused on text analysis using word frequency, sentiment analysis, and linguistic patterns to classify news as real or fake.
- These approaches are useful but often struggle with deep contextual understanding.

2. Machine Learning Models

- Previous works by [Author et al.] applied traditional ML classifiers such as Support Vector Machines (SVM), Naïve Bayes, and Decision Trees for fake news classification.
- These models perform well when trained with large datasets but may fail to generalize to new data.

3. Deep Learning-Based Approaches

- Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have been used in fake news detection to capture complex relationships between words.
- Research by [Author et al.] demonstrated how transformers like BERT improve classification accuracy.
- However, deep learning models require extensive labeled data and high computational power.

Challenges in Fake News Detection

- Misinformation Evolution: Fake news strategies change over time, making it difficult for models to remain effective.
- Data Imbalance: Real news articles significantly outnumber fake ones, leading to biased models.

- Context Understanding: Many models struggle with sarcasm, satire, and nuanced misinformation.

Research Gap and Need for Improvement

While prior research has provided strong foundations, the paper highlights the following gaps:

- Limited work has been done on integrating matrix-based methods like Singular Value Decomposition (SVD) for feature extraction.
- Most existing studies focus on English-language fake news detection, with less work in multilingual environments.
- There is a need for a hybrid approach combining statistical techniques and AI models for better accuracy.

Conclusion

The literature review concludes that while significant progress has been made in fake news detection, challenges remain. The proposed research aims to build on existing methods by incorporating a matrix-based classification approach for improved efficiency and accuracy.

PAPER 5 : Performance Comparison of Machine Learning Classifiers for Fake News Detection

Methodology

- Evaluates seven machine learning models:
  - Support Vector Machine (SVM), Logistic Regression, Decision Trees, Random Forest, Gradient Boosting, XGBoost, and Neural Networks.
- Uses TF-IDF (Term Frequency-Inverse Document Frequency) and Word Embeddings for feature extraction from news text.
- Splits the dataset into training (80%) and testing (20%) sets, applying supervised learning to train models.
- Compares models based on accuracy, computational efficiency, and generalization capability.

Novelty

- Unlike prior research that focuses on deep learning models, this study provides a comparative analysis of traditional ML classifiers to identify the best balance between accuracy and efficiency.
- Identifies SVM with TF-IDF as the best-performing model (94% accuracy), showing that deep learning is not always necessary for effective fake news detection.

- Explores lightweight models suitable for low-resource environments, making it applicable for real-time fake news monitoring.

Advantages

- SVM with TF-IDF outperforms complex models while requiring significantly less computational power.
- Fast training and inference time, making it ideal for real-world applications.
- The study provides a clear comparison of models, helping future researchers select the most effective approach.

Metrics Used

- Accuracy: Measures overall correctness of predictions.
- Precision & Recall: Evaluates how well the model differentiates real and fake news.
- F1-score: A balanced metric that considers both precision and recall.

Limitations

- The dataset consists only of English-language news, making it less generalizable to multilingual or regional misinformation.
- No image-based fake news detection, meaning it cannot handle misleading images used alongside text.
- Does not analyze contextual relationships in news articles, which deep learning models handle better.

Future Scope

- Expansion to multilingual datasets to improve performance on non-English fake news.
- Integration of image analysis to detect manipulated images used to spread fake news.
- Testing with real-time news feeds to evaluate how well the model performs in continuously evolving misinformation landscapes.

PAPER 6 : Fake News Detection Enhancement with Data Imputation

Methodology

- Introduces data imputation techniques to handle missing values in fake news datasets.

- Uses categorical feature imputation (most frequent value) and numerical imputation (mean value) to replace missing data.
- Applies feature extraction using TF-IDF vectorization to convert text into numerical format.
- Compares the performance of SVM, Decision Trees, Multi-Layer Perceptron (MLP), and Gradient Boosting on both raw and imputed datasets.
- Splits data into 80% training and 20% testing for model evaluation.

Novelty

- Most fake news studies ignore missing values, leading to biased models. This study improves detection accuracy by implementing data imputation techniques before classification.
- Demonstrates that MLP (Multi-Layer Perceptron) performs best (45.7% accuracy) after imputation, highlighting the impact of improved data quality.
- The study shows that machine learning models perform better when missing values are properly handled, making it valuable for real-world fake news datasets.

Advantages

- Improved model accuracy (16% increase) compared to classifiers without imputation.
- Enhances data reliability, making classification more robust.
- Effective even with small datasets, where missing values significantly impact performance.
- Reduces bias in machine learning models, improving the overall prediction quality.

Metrics Used

- Accuracy: Measures classification performance.
- Precision & Recall: Evaluates how well the model identifies fake vs. real news.
- F1-score: Provides a combined measure of precision and recall to assess overall model effectiveness.

Limitations

- Computationally expensive: Imputation adds extra processing time, making it slower than standard ML models.
- Limited dataset size: The study lacks validation on larger, real-world datasets, which could impact scalability.
- Does not address contextual understanding: The model works purely on feature extraction and lacks deep linguistic analysis.

Future Scope

- Combining imputation techniques with deep learning to improve accuracy while handling missing data more effectively.
- Testing on larger datasets to ensure the approach is scalable for real-world applications.
- Incorporating social media metadata (such as user credibility scores) to improve fake news classification.