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 2, April 2026
Impact Factor: 8.2

Fake News Detection with Interpretation

Velpula Koteswara Rao[1], M. Suneel[2], Md. Ghani[3], M. Ajay Babu[4] , K.K. Eshwara Kumar[5]
Assistant Professor, Department of CSE[1]
[2,3,4,5] UG Students, Department of CSE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
[1]koteswararao@vvit.net, [2]22bq1a05d4@vvit.net,[3]22bq1a05e1@vvit.net,
[4]22bq1a05d1@vvit.net, [5]23bq5a0512@vvit.net

Abstract: *The rapid proliferation of digital media platforms has significantly increased the spread of misinformation and fake news, posing serious challenges to public trust and informed decision-making. This research presents an intelligent Fake News Detection system with Explainability that aims to identify misleading content in both textual and visual formats while providing transparent reasoning behind the predictions. The proposed system integrates multiple machine learning and deep learning models to analyse news content and determine its authenticity. For text-based detection, traditional machine learning models such as Logistic Regression and XGBoost were initially implemented and evaluated. However, a fine-tuned Bidirectional Encoder Representations from Transformers (BERT) model demonstrated superior performance and was selected as the final model, achieving an accuracy of 86%. For image-based misinformation detection, the EfficientNet-B0 convolutional neural network architecture was employed, achieving an accuracy of 81% on the image classification task.*
To enhance transparency and user trust, the system incorporates Explainable Artificial Intelligence (XAI) techniques. Local Interpretable Model-Agnostic Explanations (LIME) is applied to highlight influential keywords within textual content that contribute to the model's prediction, while Gradient-weighted Class Activation Mapping (Grad-CAM) is utilized to visually highlight important regions of images that influence classification decisions. The integrated web-based platform enables users to submit news content and receive predictions along with interpretable explanations. Experimental results demonstrate that the combination of transformer-based language models, efficient convolutional architectures, and explainability techniques can effectively detect fake news while maintaining interpretability. The proposed system contributes toward improving digital media literacy by providing accurate detection along with transparent and understandable AI-driven explanations.

Keywords: Fake News Detection, Misinformation Detection, Explainable Artificial Intelligence (XAI), Bidirectional Encoder Representations from Transformers (BERT), EfficientNet-B0, Logistic Regression, XGBoost, LIME, Grad-CAM, Natural Language Processing, Deep Learning, Image Classification.

I. INTRODUCTION

The rapid growth of digital media platforms and social networking services has significantly transformed the way information is created, shared, and consumed. While these platforms enable instant access to news and global events, they have also facilitated the widespread dissemination of misinformation and fake news. Fake news refers to deliberately fabricated or misleading information presented as legitimate news, often designed to manipulate public opinion, influence political decisions, or generate social unrest. The speed at which such content spreads across online platforms makes it increasingly difficult for individuals to verify the authenticity of information.

Traditional methods of detecting fake news rely heavily on manual fact-checking by journalists and experts. Although effective, these approaches are time-consuming and cannot keep pace with the enormous volume of content generated on the internet each day. Consequently, automated fake news detection systems based on Machine Learning (ML) and Natural Language Processing (NLP) have gained significant attention in recent years. These systems analyse linguistic

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 2, April 2026
Impact Factor: 8.2

patterns, contextual information, and semantic features within news articles to determine whether the information is likely to be true or false.

In addition to textual misinformation, manipulated or misleading images are frequently used to support false narratives and increase the credibility of fake news. Therefore, modern fake news detection systems must be capable of analysing both textual and visual content. Deep learning techniques, particularly transformer-based models for text and convolutional neural networks for image analysis, have demonstrated strong performance in identifying complex patterns within such multimodal data.

Another important challenge associated with automated detection systems is the lack of transparency in model decisions. Many deep learning models operate as "black boxes," making it difficult for users to understand the reasoning behind their predictions. This lack of interpretability can reduce user trust and limit the practical adoption of such systems. To address this challenge, Explainable Artificial Intelligence (XAI) techniques are increasingly integrated into fake news detection frameworks to provide insights into how predictions are generated.

In this research, a Fake News Detection system with explainability is proposed to identify misinformation from both textual and visual content. Multiple machine learning models, including Logistic Regression and XGBoost, were initially implemented for text classification, followed by a transformer-based Bidirectional Encoder Representations from Transformers (BERT) model that achieved the highest accuracy of 86%. For image-based detection, the EfficientNet-B0 convolutional neural network architecture was employed, achieving an accuracy of 81%. To improve interpretability, Local Interpretable Model-Agnostic Explanations (LIME) is used to highlight influential keywords in textual content, while Gradient-weighted Class Activation Mapping (Grad-CAM) highlights important regions within images that contribute to the model's decision.

The developed system is implemented as a web-based platform that allows users to submit news content and receive classification results along with visual explanations. By combining advanced machine learning models with explainability techniques, the proposed approach aims to enhance the reliability, transparency, and effectiveness of automated fake news detection systems.

II. LITERATURE SURVEY

The rapid increase in the dissemination of misinformation across digital platforms has led to significant research efforts in the field of automated fake news detection. Researchers have explored various machine learning and deep learning techniques to analyse textual and visual content in order to identify misleading or fabricated information. Early approaches primarily focused on traditional machine learning algorithms that relied on handcrafted features extracted from news articles. Techniques such as Logistic Regression, Support Vector Machines, and Decision Trees were commonly used to classify news as real or fake based on linguistic characteristics, writing style, and statistical patterns in the text.

As the complexity of misinformation increased, researchers began adopting more advanced machine learning models capable of capturing deeper contextual relationships within textual data. Ensemble learning methods such as Random Forest and XGBoost demonstrated improved performance by combining multiple decision trees to enhance classification accuracy. These models were particularly effective when combined with feature engineering techniques such as Term Frequency–Inverse Document Frequency (TF-IDF), n-gram representations, and sentiment analysis.

With the advancement of deep learning, Natural Language Processing models based on neural networks have become increasingly popular for fake news detection tasks. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks were initially used to capture sequential dependencies in textual data. However, the introduction of transformer-based architectures significantly improved the performance of NLP tasks. Models such as Bidirectional Encoder Representations from Transformers (BERT) are capable of understanding contextual relationships between words in a sentence through attention mechanisms. As a result, transformer-based models have achieved state-of-the-art performance in various text classification tasks, including misinformation detection.

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 2, April 2026 Impact Factor: 8.2

In addition to textual analysis, researchers have recognized the importance of detecting fake news that contains manipulated or misleading images. Convolutional Neural Networks (CNNs) have been widely used for image classification tasks due to their ability to extract hierarchical visual features from images. Advanced architectures such as ResNet, Inception, and EfficientNet have demonstrated strong performance in identifying visual inconsistencies or manipulated content associated with fake news. These models are capable of learning complex visual patterns that help distinguish authentic images from misleading ones.

Another significant challenge in automated fake news detection is the lack of interpretability in machine learning models. Many deep learning systems function as black-box models, making it difficult for users to understand the reasoning behind predictions. To address this issue, researchers have incorporated Explainable Artificial Intelligence (XAI) techniques into fake news detection frameworks. Methods such as Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) are commonly used to interpret text classification models by identifying important features or keywords that influence predictions. Similarly, visualization techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM) help highlight specific regions in images that contribute to classification decisions in convolutional neural networks.

Despite the significant progress in fake news detection research, several challenges remain, including dataset bias, limited generalization across domains, and the need for transparent and trustworthy AI systems. Therefore, integrating advanced deep learning models with explainability techniques has become an important research direction for developing reliable and interpretable fake news detection systems.

III. PROPOSED SYSTEM

The effectiveness of an automated fake news detection system depends on the quality of the datasets used for training, the robustness of the machine learning models employed, and the interpretability of the predictions generated by the system. The proposed system is designed to detect misinformation from both textual and visual news content while providing transparent explanations for the classification results. The architecture integrates traditional machine learning models, transformer-based deep learning models, convolutional neural networks, and Explainable Artificial Intelligence (XAI) techniques. The system is implemented as a web-based platform that allows users to input news articles or images and receive predictions along with interpretable insights.

Data Acquisition and Diversity

The performance of fake news detection models is heavily influenced by the diversity and reliability of the training datasets. To ensure a comprehensive representation of real and fake news content, two publicly available benchmark datasets were utilized for model training and evaluation.

For textual fake news detection, the FakeNewsNet dataset was employed. This dataset contains verified news articles collected from multiple news sources and social media platforms, labeled as either real or fake. The dataset includes news headlines, article content, and additional contextual information, enabling the model to learn linguistic patterns associated with misinformation.

For image-based misinformation detection, the Fakeddit dataset was utilized. This dataset contains images associated with social media posts labeled across different levels of misinformation. The dataset includes both authentic and manipulated images that often accompany misleading news articles. By training on such diverse visual data, the model learns to identify visual inconsistencies and patterns associated with deceptive media content.

The use of two distinct datasets enables the proposed system to detect misinformation across different modalities, improving its ability to analyze complex multimedia news content.

Dataset Composition and Class Distribution

The textual dataset used in this research consists of thousands of labelled news articles categorized into two primary classes: Real News and Fake News. Each sample contains the news headline and article body, allowing the model to

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 2, April 2026
Impact Factor: 8.2

analyse contextual relationships between words and sentences. The dataset was carefully balanced to ensure that both classes are adequately represented, preventing bias toward a specific category during training.

Similarly, the image dataset contains thousands of labelled images associated with online posts. These images are categorized into binary classes representing authentic and misleading visual content. The dataset includes a wide variety of image types such as news photographs, social media images, memes, and manipulated visual content. This diversity enables the model to learn meaningful visual features that differentiate real images from misleading or manipulated ones.

The balanced class distribution ensures that the trained models can effectively generalize across different forms of misinformation encountered in real-world scenarios.

Text Feature Extraction and Preprocessing Pipeline

Before training the machine learning models, a preprocessing pipeline was implemented to clean and transform the textual data into a format suitable for model training. The preprocessing stage plays a crucial role in improving the performance and stability of natural language processing models.

The preprocessing pipeline consists of the following steps:

• Text Cleaning: News articles are processed to remove unnecessary characters, URLs, punctuation marks, and special symbols that do not contribute to semantic meaning.

• Tokenization: The cleaned text is divided into smaller linguistic units known as tokens, typically representing words or sub-words. This allows the model to analyse the structural relationships between words.

• Stopword Removal: Frequently occurring words such as "the," "is," and "and" that carry minimal semantic importance are removed to reduce noise in the dataset.

• Vector Representation: For traditional machine learning models such as Logistic Regression and XGBoost, Term Frequency–Inverse Document Frequency (TF-IDF) vectorization is applied to convert textual content into numerical feature vectors. For the deep learning model, the BERT tokenizer is used to generate contextual embeddings that preserve semantic relationships between words.

These preprocessing steps ensure that the textual data is transformed into a structured representation that can be effectively processed by machine learning algorithms.

Text Classification Models

To evaluate the performance of different machine learning approaches, multiple classification models were implemented for textual fake news detection.

The first model implemented was Logistic Regression, a widely used linear classification algorithm that estimates the probability of a news article being fake or real based on extracted textual features.

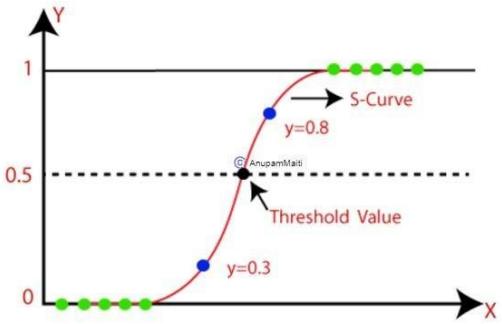

Fig.1 Logistic Regression Curve

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 2, April 2026
Impact Factor: 8.2

The second model used was XGBoost, an advanced ensemble learning algorithm based on gradient boosting decision trees. XGBoost improves classification performance by combining multiple weak learners and optimizing them through gradient descent.

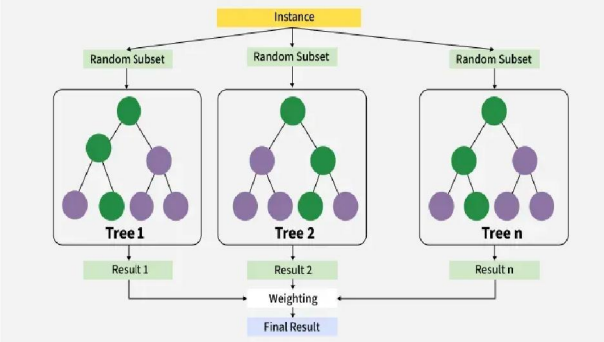

Fig.2 Flow of XGBoost

To further improve contextual understanding, a Bidirectional Encoder Representations from Transformers (BERT) model was fine-tuned for the fake news classification task. BERT utilizes a transformer architecture with attention mechanisms that allow the model to understand contextual relationships between words in a sentence. After extensive experimentation, BERT demonstrated the highest performance with an accuracy of 86%, outperforming the other models. Therefore, BERT was selected as the final text classification model in the proposed system.

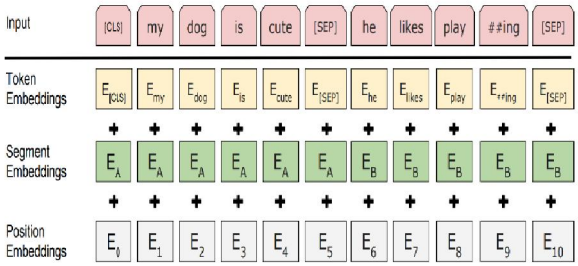

Fig.3 Flow of BERT

Image Classification Architecture

In addition to textual analysis, the system also detects misinformation from images using a deep learning-based image classification model. The model architecture is based on EfficientNet-B0, a convolutional neural network designed to achieve high performance while maintaining computational efficiency.

EfficientNet-B0 utilizes compound scaling techniques to balance network depth, width, and resolution. This allows the model to extract hierarchical visual features such as edges, textures, shapes, and complex patterns from images.

Before training the model, all images were resized to a fixed resolution and normalized to ensure consistent input dimensions. Data augmentation techniques such as horizontal flipping, rotation, and brightness adjustments were applied to improve the robustness of the model and prevent overfitting. After training on the Fakeddit dataset, the EfficientNet-B0 model achieved an accuracy of 81% in distinguishing between real and misleading images.

DOI: 10.48175/IJARSCT-32479

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 2, April 2026
Impact Factor: 8.2

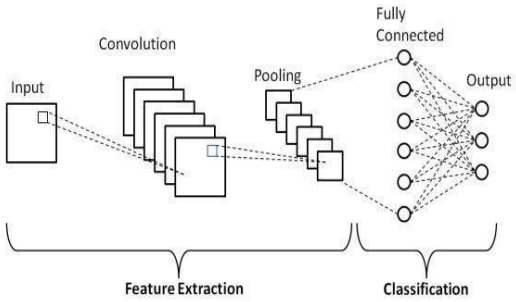

Fig.4 CNN Architecture

Explainability Techniques

To improve transparency and user trust, the proposed system incorporates Explainable Artificial Intelligence (XAI) techniques that provide insights into how the models arrive at their predictions.

For textual fake news detection, Local Interpretable Model-Agnostic Explanations (LIME) is used. LIME analyses the contribution of individual words within a news article and highlights the keywords that most strongly influence the model's prediction. This enables users to understand why a particular article was classified as fake or real.

For image classification, Gradient-weighted Class Activation Mapping (Grad-CAM) is utilized to visualize important regions within an image that contribute to the model's decision. Grad-CAM generates heatmaps that highlight the areas of the image that the convolutional neural network considers most relevant during classification.

These interpretability techniques help transform the system from a black-box model into a transparent and explainable AI system.

System Integration and Web Application

The complete fake news detection system is implemented as a web-based application that integrates the trained machine learning models with an interactive user interface. The frontend of the system is developed using React and Tailwind CSS, providing a responsive interface for user interaction.

The backend is implemented using FastAPI, which handles model inference, data processing, and communication between the frontend and the machine learning models. When a user submits textual or visual news content, the backend processes the input, performs classification using the trained models, and returns the prediction results along with the corresponding explanations generated by LIME or Grad-CAM.

Training Configuration and Computational Setup

The training process for both text and image models was conducted using modern deep learning frameworks and optimized computational settings. The BERT model was fine-tuned using transformer-based training procedures with appropriate learning rates and batch sizes to achieve optimal performance on the textual dataset. The EfficientNet-B0 model was trained using the Adam optimizer and cross-entropy loss function to minimize classification error.

The dataset was split into training and testing subsets to evaluate model performance and prevent overfitting. Model evaluation was conducted using metrics such as accuracy, precision, recall, and F1-score. The final models were selected based on their overall performance and generalization ability on unseen data.

The integration of advanced deep learning models with explainability techniques enables the proposed system to accurately detect misinformation while providing interpretable results that enhance user trust and understanding.

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 2, April 2026
Impact Factor: 8.2

IV. RESULTS AND DISCUSSION

This section presents the experimental results obtained from the implementation of the proposed Fake News Detection system. The performance of different machine learning and deep learning models used for both textual and image-based misinformation detection is analysed. The evaluation focuses on classification accuracy, model comparison, and the effectiveness of explainability techniques integrated into the system.

Evaluation Metrics

To assess the performance of the proposed models, several standard evaluation metrics commonly used in classification tasks were employed. These metrics provide a comprehensive understanding of the model's predictive capability.

- **Accuracy:** Measures the overall percentage of correctly classified samples.
- **Precision:** Indicates the proportion of predicted fake news instances that are actually fake.
- **Recall:** Represents the ability of the model to correctly identify all fake news instances.
- **F1-Score:** The harmonic mean of precision and recall, providing a balanced evaluation of model performance.

These metrics ensure that the models are evaluated not only based on overall correctness but also on their ability to effectively detect misinformation without producing excessive false positives or false negatives.

Performance of Text Classification Models

Multiple machine learning models were implemented and evaluated for textual fake news detection, including Logistic Regression, XGBoost, and BERT. The models were trained using the FakeNewsNet dataset and evaluated on a separate test set to measure their generalization performance.

Logistic Regression served as a baseline model due to its simplicity and efficiency in handling text classification tasks. While it demonstrated reasonable performance, its ability to capture complex contextual relationships within long news articles was limited.

Class	Precision	Recall	F1-Score	Support
Real News (0)	0.85	0.96	0.90	3489
Fake News (1)	0.78	0.47	0.59	1151
Accuracy			**0.84**	4640
Macro Avg	0.81	0.71	0.74	4640
Weighted Avg	0.83	0.84	0.82	4640

Table: Performance of Logistic Regression Model for Text Fake News Detection

The XGBoost model improved upon this baseline by utilizing gradient boosting techniques that combine multiple decision trees to enhance predictive performance. This model showed better classification capability compared to Logistic Regression due to its ability to capture nonlinear relationships within the data.

Class	Precision	Recall	F1-Score	Support
Real News (0)	0.83	0.96	0.89	3489
Fake News (1)	0.78	0.39	0.52	1151
Accuracy			**0.82**	4640
Macro Avg	0.80	0.68	0.71	4640
Weighted Avg	0.82	0.82	0.80	4640

Table: Performance of XGBoost Model for Text Fake News Detection

However, the transformer-based BERT model achieved the best performance among all implemented models. By leveraging attention mechanisms and bidirectional context understanding, BERT was able to capture deeper semantic

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 2, April 2026
Impact Factor: 8.2

relationships between words and sentences in news articles. As a result, the fine-tuned BERT model achieved an accuracy of **86%**, making it the final selected model for textual fake news detection in the proposed system.

Class	Precision	Recall	F1-Score	Support
Real News (0)	0.90	0.93	0.91	3489
Fake News (1)	0.76	0.67	0.71	1151
Accuracy			**0.87**	4640
Macro Avg	0.83	0.80	0.81	4640
Weighted Avg	0.86	0.87	0.86	4640

Table: Performance of BERT Model for Text Fake News Detection

Performance of Image Classification Model

For the detection of misleading visual content, the EfficientNet-B0 convolutional neural network architecture was implemented and trained using the Fakeddit dataset. EfficientNet-B0 was chosen due to its balanced design that optimizes model depth, width, and resolution while maintaining computational efficiency.

The model demonstrated strong performance in extracting hierarchical visual features such as edges, textures, and object patterns from images. After training and evaluation, the EfficientNet-B0 model achieved an overall classification accuracy of **81%** in distinguishing between authentic and misleading images.

The results indicate that convolutional neural networks are effective in identifying patterns associated with manipulated or misleading visual content used in fake news.

Class	Precision	Recall	F1-Score	Support
Real News (0)	0.84	0.93	0.88	11012
Fake News (1)	0.67	0.45	0.54	3587
Accuracy			**0.81**	14599
Macro Avg	0.75	0.69	0.71	14599
Weighted Avg	0.80	0.81	0.80	14599

Table: Performance of EfficientNet-B0 Model for Image Fake News Detection

Explainability and Model Interpretability

A key objective of the proposed system is to improve transparency in automated fake news detection by integrating Explainable Artificial Intelligence techniques.

For textual predictions, **Local Interpretable Model-Agnostic Explanations (LIME)** was used to highlight the most influential words contributing to the classification decision. When a news article is classified as fake or real, LIME generates a visualization showing the keywords that had the greatest impact on the model's prediction. This enables users to understand the reasoning behind the classification rather than relying solely on a numerical prediction.

For image classification, **Gradient-weighted Class Activation Mapping (Grad-CAM)** was employed to visualize the regions of an image that influenced the model's decision. Grad-CAM produces heatmaps that highlight the areas of the image the convolutional neural network focuses on during classification. These visual explanations help users interpret why a particular image is identified as misleading or authentic.

The integration of these explainability techniques enhances user trust and makes the system more transparent compared to traditional black-box machine learning models.

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 2, April 2026
Impact Factor: 8.2

Discussion

The experimental results demonstrate that advanced deep learning models significantly outperform traditional machine learning methods in fake news detection tasks. In particular, transformer-based architectures such as BERT are highly effective in capturing contextual and semantic information within textual content, leading to improved classification performance.

Similarly, convolutional neural networks such as EfficientNet-B0 provide strong capabilities for analysing visual misinformation by learning complex hierarchical image features.

Another important outcome of this research is the successful integration of explainability techniques into the detection pipeline. The use of LIME and Grad-CAM ensures that the predictions generated by the models are interpretable and understandable to users. This is particularly important in misinformation detection systems where transparency and trust are critical.

Overall, the results indicate that combining deep learning models with explainable AI techniques provides an effective and reliable approach for detecting fake news across both textual and visual modalities.

V. CONCLUSION

This research presents the design and implementation of an intelligent fake news detection system capable of identifying misinformation in both textual and visual formats while providing interpretable explanations for its predictions. The proposed system integrates traditional machine learning models, transformer-based deep learning architectures, and convolutional neural networks to analyse news content and determine its authenticity. Through extensive experimentation, multiple models were evaluated for textual fake news detection, including Logistic Regression, XGBoost, and a fine-tuned BERT model. Among these approaches, the BERT model demonstrated superior performance, achieving an accuracy of **86%**, highlighting the effectiveness of transformer-based architectures in capturing contextual and semantic relationships within news articles.

For image-based misinformation detection, the EfficientNet-B0 convolutional neural network was employed to analyse visual content associated with news posts. The model achieved an accuracy of **81%**, demonstrating its capability to identify misleading or manipulated images commonly used in fake news dissemination. The integration of both text and image analysis enables the system to address the increasingly multimodal nature of misinformation across online platforms.

An important contribution of this research is the incorporation of Explainable Artificial Intelligence techniques to improve transparency and user trust. The system utilizes **Local Interpretable Model-Agnostic Explanations (LIME)** to highlight influential keywords in textual content that affect the classification outcome. Similarly, **Gradient-weighted Class Activation Mapping (Grad-CAM)** is applied to generate heatmaps that highlight important regions within images influencing the model's decision. These explainability techniques transform the system from a black-box model into an interpretable framework that allows users to understand the reasoning behind predictions.

The complete system is implemented as a web-based platform that enables users to submit news articles or images and receive classification results along with meaningful explanations. By combining advanced deep learning models with explainable AI techniques, the proposed approach contributes to improving the reliability, transparency, and usability of automated misinformation detection systems.

Overall, this work demonstrates the feasibility of developing an integrated fake news detection framework that not only achieves strong predictive performance but also enhances user understanding through interpretable outputs. Such systems can play an important role in combating misinformation, promoting digital media literacy, and supporting more informed decision-making in the rapidly evolving online information ecosystem.

REFERENCES

[1]. Y. Shen, Q. Liu, N. Guo, J. Yuan, and Y. Yang, "Fake News Detection on Social Networks: A Survey," *Applied Sciences*, vol. 13, no. 21, 2023.

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 2, April 2026
Impact Factor: 8.2

[2]. M. Giria, S. Eswaran, P. Honnavalli, and D. Dc, "Automated and Interpretable Fake News Detection With Explainable Artificial Intelligence," *Journal of Applied Security Research*, vol. 19, no. 4, pp. 628–648, 2024.

[3]. R. Jadhav, V. Meshram, A. Bhosle, K. Patil, S. Dash, and S. Jadhav, "Explainable Multilingual and Multimodal Fake-News Detection: Toward Robust and Trustworthy AI for Combating Misinformation," *Frontiers in Artificial Intelligence*, vol. 8, 2025.

[4]. R. K. Ayyasamy, C. Ponnusamy, K. N. Bhargavi, S. Cherukuvada, G. C. Babu, and S. Amutha, "A Hybrid Deep Learning Framework for Fake News Detection Using LSTM-CGPNN and Metaheuristic Optimization," *Scientific Reports*, vol. 15, 2025.

[5]. H. R. Iyer and A. K. Madasamy, "A Reasoning-Based Explainable Multimodal Fake News Detection Using Large Language Models and Transformers," *Journal of Big Data*, 2025.

[6]. J. Patel, C. Bhatt, H. Trivedi, and T. T. Nguyen, "Misinformation Detection Using Large Language Models With Explainability," 2025.

[7]. J. Ayoub, X. J. Yang, and F. Zhou, "Combat COVID-19 Infodemic Using Explainable Natural Language Processing Models," 2021.

[8]. M. Szczepański, M. Pawlicki, R. Kozik, and M. Choraś, "New Explainability Method for BERT-Based Model in Fake News Detection," *Scientific Reports*, vol. 11, 2021.

[9]. P. Kumar and A. Shrivastava, "A Survey on Efficient Classification Models for Fake News Detection," *International Journal of Scientific Innovation and Engineering*, vol. 2, no. 9, 2025.

[10]. Y. E. Yousif, "Fake News Detection Using Transformer-Based Models With Explainable Artificial Intelligence," *International Journal of Advanced Multidisciplinary Research*, 2026.