TRUTH SEEKER: DETECTING AND COUNTERING FAKE NEWS TO PREVENT MISINFORMATION

AI19811 PROJECT PHASE-II REPORT

Submitted by

RESHMA YASMIN M.A (2116211501080)
SHRUTHI G (2116211501099)

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

RAJALAKSHMI ENGINEERING COLLEGE (AUTONOMOUS),

CHENNAI-602 105

APRIL 2025

RAJALAKSHMI ENGINEERING COLLEGE

(an Autonomous Institution Affiliated to Anna University, Chennai)

BONAFIDE CERTIFICATE

Certified that this Phase –II Thesis titled "**TRUTH SEEKER: DETECTING AND COUNTERING FAKE NEWS TO PREVENT MISINFORMATION**" is the Bonafide work of **RESHMA YASMIN M.A (2116211501080), SHRUTHI G (2116211501099)** who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported here in does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

Dr. K. SEKAR M.E., Ph.D.,	**Dr. K. SEKAR M.E., Ph.D.,**
Professor and Head	Supervisor
Department of Artificial Intelligence and Machine Learning	Department of Artificial Intelligence and Machine Learning
Rajalakshmi Engineering College	Rajalakshmi Engineering College
Chennai- 602 105	Chennai- 602 105

Submitted for the project viva-voce examination held on ______________

Internal Examiner **External Examiner**

ACKNOWLEDGEMENT

First, we thank the almighty God for the successful completion of the project. Our sincere thanks to our chairman **Mr. S. Meganathan, B.E., F.I.E.,** for his sincere endeavor in educating us in his premier institution. We would like to express our deep gratitude to our beloved Chairperson **Dr.(Mrs.)Thangam Meganathan, M.A., M.Phil., Ph.D.,** for her enthusiastic motivation which inspired us a lot in completing this project, and Vice-Chairman **Mr. Abhay Shankar Meganathan**, **B.E., M.S.,** for providing us with the requisite infrastructure.

We also express our sincere gratitude to our college principal, **Dr. S. N. Murugesan, M.E., PhD.,** for his kind support and facilities to complete our work on time. We extend heartfelt gratitude to **Dr. K Sekar, M.E., Ph.D.,** Professor and Head of the Department of Artificial Intelligence and Machine Learning for his guidance and encouragement throughout the work. We are very glad to thank our project coordinator**, Mrs. Akshaya V, M.E,** Assistant Professor for her encouragement and support towards the successful completion of this project. We would like to sincerely thank our project guide **Dr. K Sekar, M.E., Ph.D.,** Professor and Head of the Department of Artificial Intelligence and Machine Learning for him valuable guidance throughout the project work. We extend our sincere thanks to our parents, friends, all faculty members, and supporting staff for their direct and indirect involvement in the successful completion of the project for their encouragement and support

Reshma Yasmin M.A (211501080)

Shruthi G (211501099)

DEPARTMENT VISION

To promote highly Ethical and Innovative Computer Professionals through excellence in teaching, training and research.

DEPARTMENT MISSION

- To produce globally competent professionals, motivated to learn the emerging technologies and to be innovative in solving real world problems.
- To promote research activities amongst the students and the members of faculty that could benefit the society.
- To impart moral and ethical values in their profession.

PROGRAMME EDUCATIONAL OBJECTIVES (PEOs)

PEO 1: To equip students with essential background in computer science, basic electronics and applied mathematics.

PEO 2: To prepare students with fundamental knowledge in programming languages, and tools and enable them to develop applications.

PEO 3: To encourage the research abilities and innovative project development in the field of AI, ML, DL, networking, security, web development, Data Science and also emerging technologies for the cause of social benefit.

PEO 4: To develop professionally ethical individuals enhanced with analytical skills, communication skills and organizing ability to meet industry requirements.

PROGRAM OUTCOMES (POs)

PO1: Engineering knowledge: Apply the knowledge of Mathematics, Science, Engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

PO2: Problem analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

PO3: Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

PO 4: Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

PO 5: Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

PO 6: The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

PO 7: Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

PO 8: Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

PO 9: Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

PO 10: Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

PO11: Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

PO12: Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

PROGRAM SPECIFIC OUTCOMES (PSOs)

A graduate of the Artificial Intelligence and Machine Learning Program will demonstrate

PSO 1: Foundation Skills: Ability to understand, analyze and develop computer programs in the areas related to algorithms, system software, web design, AI, machine learning, deep learning, data science, and networking for efficient design of computer-based systems of varying complexity. Familiarity and practical competence with a broad range of programming language, tools and open source platforms.

PSO 2: Problem-Solving Skills: Ability to apply mathematical methodologies to solve computational task, model real world problem using appropriate AI and ML algorithms. To understand the standard practices and strategies in project development, using open-ended programming environments to deliver a quality product.

PSO 3: Successful Progression: Ability to apply knowledge in various domains to identify research gaps and to provide solution to new ideas, inculcate passion towards higher studies, creating innovative career paths to be an entrepreneur and evolve as an ethically social responsible AI and ML professional.

COURSE OBJECTIVE

- To identify and formulate real-world problems that can be solved using Artificial Intelligence and Machine Learning techniques.
- To apply theoretical and practical knowledge of AI/ML for designing innovative, data-driven solutions.
- To integrate various tools, frameworks, and algorithms to develop, test, and validate AI/ML models.
- To demonstrate effective teamwork, project management, and communication skills through collaborative project execution.
- To instill awareness of ethical, societal, and environmental considerations in the design and deployment of intelligent systems.

COURSE OUTCOME

- **CO1:** Analyze and define a real-world problem by identifying key challenges, project requirements and constraints.
- **CO2:** Conduct a thorough literature review to evaluate existing solutions, identify research gaps and formulate research questions.

- **CO3:** Develop a detailed project plan by defining objectives, setting timelines, and identifying key deliverables to guide the implementation process.
- **CO4:** Design and implement a prototype or initial model based on the proposed solution framework using appropriate AI tools and technologies.
- **CO5:** Demonstrate teamwork, communication, and project management skills by preparing and presenting a well-structured project proposal and initial implementation results.

CO-PO-PSO Mapping

CO	**PO 1**	**PO 2**	**PO 3**	**PO 4**	**PO 5**	**PO 6**	**PO 7**	**PO 8**	**PO 9**	**PO 10**	**PO 11**	**PO 12**	**PSO 1**	**PSO 2**	**PSO 3**
CO 1	3	3	3	3	3	2	2	2	3	2	3	3	3	3	3
CO 2	3	3	3	3	3	2	-	-	2	2	2	3	3	2	2
CO 3	3	3	3	2	3	1	1	2	3	3	3	3	3	3	3
CO 4	3	3	3	3	3	2	1	2	3	2	2	3	3	3	3
CO 5	1	1	1	1	1	-	-	-	3	3	3	3	1	-	2

Note: Correlation levels 1, 2 or 3 are as defined below:

1: Slight (Low)   2: Moderate (Medium)   3: Substantial (High)

No correlation: “-”

ABSTRACT

Disinformation or fake news is a major problem in the age of technology, as it affects the public mindset and decision-making process. This project offers a deep learning-based solution for identifying fake news from both text and image data. The system combines a TF-IDF and a Passive Aggressive Classifier (PAC) for text and a Vision Transformer (ViT)-based model for images. The Web interface of the application is built using Flask, enabling users to enter news articles or upload images for authentication. For text data, the TF-IDF vectorizer retrieves useful features by transforming the textual content into numerical representations, while the PAC model classifies the news as authentic or spurious with high precision. For image data, a pre-trained ViT model, fine-tuned over a large dataset of real and fake news images, effectively identifies whether an image is genuine or manipulated. The project also includes data preprocessing techniques, such as removing stopwords, stemming, and resizing images, along with model training, testing, and visualization to maximize performance. In addition, the system offers a simple news classification module that categorizes text articles into different types, such as politics, sports, weather, or others, enhancing its versatility. The Flask-based web application is equipped with intuitive APIs that facilitate both text-based and image-based fake news detection, providing seamless user interaction. The experimental results demonstrate the effectiveness of the proposed models in accurately identifying misinformation, offering a reliable and scalable solution to combat online disinformation. Furthermore, the system can be extended by integrating real-time fact-checking APIs and expanding the dataset with diverse and multilingual content to further enhance its robustness

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

DL	Deep Learning
NLP	Natural Language Processing
CV	Computer Vision
TF-IDF	Term Frequency-Inverse Document Frequency
PAC	Passive Aggressive Classifier
ViT	Vision Transformer
SVM	Support Vector Machine
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
XAI	Explainable Artificial Intelligence
X	Feature Vector/Input
C	Class Label
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative

CHAPTER 1

INTRODUCTION

1.1 GENERAL

The exponential rise of digital media and online platforms has significantly transformed how people access and consume information. While this has made news more accessible, it has also opened the door for the rapid spread of misinformation and fake news. The consequences of such false information can be severe, impacting public opinion, influencing elections, and even inciting social tension. As digital footprints grow larger, identifying and addressing fake news has become a crucial priority to maintain the integrity of online information and safeguard public trust.

This project presents an intelligent fake news detection system that utilizes deep learning techniques to assess both textual and visual content. Traditional fact-checking methods are manual and time-consuming, making them inefficient for real-time applications. To overcome these limitations, the proposed system integrates Natural Language Processing (NLP) and Computer Vision (CV) to automatically classify news as real or fake. For textual analysis, we use Term Frequency-Inverse Document Frequency (TF-IDF) to extract features, coupled with a Passive Aggressive Classifier (PAC), known for its effectiveness in processing large volumes of news data. On the visual front, a fine-tuned Vision Transformer (ViT) model is employed to detect misleading images associated with news articles.

To enhance accessibility, the system is deployed as a Flask-based web application. This user-friendly platform allows individuals to input news articles or upload images to verify their authenticity. Additionally, the system organizes content into categories such as politics, sports, and weather, offering a more tailored and efficient user experience. By combining the power of machine learning, deep learning, and web technologies, this project delivers a robust and scalable solution to tackle the growing challenge of fake news. It provides users with quick and reliable tools to verify information in real-time, thereby promoting responsible content consumption and digital literacy.

1.1.1 PROPOSED MODEL FRAMEWORK

The proposed model framework leverages a hybrid, multimodal architecture designed to detect fake news by analyzing both textual and visual components of online content. Recognizing that misinformation often manipulates either text, images, or both, the framework integrates complementary models for each data modality to enhance detection accuracy. For the textual modality, a combination of machine learning and deep learning classifiers is employed. Naive Bayes and Support Vector Machine (SVM) serve as foundational algorithms that handle structured and concise text effectively, offering interpretability and robustness. To further strengthen the system's ability to detect complex deceptive language, Convolutional Neural Networks (CNNs) are incorporated. CNNs excel at capturing local semantic patterns, making them suitable for identifying nuanced textual features commonly found in fake news.

For the visual modality, the model integrates a Vision Transformer (ViT), which applies transformer-based self-attention mechanisms to image patches, enabling the model to capture global relationships within the image. This is particularly useful for identifying subtle visual manipulations or mismatches between image and text, such as misleading photos or edited visuals. The integration of these models forms a unified detection pipeline, where the outputs of the text and image classifiers are combined to make a final prediction. This multimodal fusion ensures that the system can flag inconsistencies across modalities, which are typical indicators of fake news. The hybrid approach thus enhances both the depth and reliability of fake content detection, making the system more robust against diverse misinformation strategies.

1.1.2 TEXT CLASSIFICATION APPROACH

To identify fake news through textual content, the system utilizes a hybrid approach that combines both traditional machine learning and deep learning methods. The pipeline begins with text preprocessing, followed by feature extraction using Term Frequency-Inverse Document Frequency (TF-IDF). The extracted features are then classified using three models: Naive Bayes, Support Vector Machine (SVM), and Convolutional Neural Network (CNN).

Naive Bayes and SVM are effective for handling short-form, structured text and provide quick and interpretable results. CNNs, on the other hand, are capable of capturing local and sequential linguistic features within longer articles, making them ideal for detecting hidden patterns that are often present in fake news. This combination ensures high accuracy and robustness in textual fake news classification.

1.1.3 IMAGE CLASSIFICATION APPROACH

Visual misinformation, such as manipulated images or deepfakes, is a growing challenge in fake news detection. To address this, the proposed system incorporates the Vision Transformer (ViT) model, a cutting-edge deep learning architecture designed for image classification tasks. Unlike traditional CNNs, ViTs use self-attention mechanisms to analyze image patches, allowing the model to capture both local and global visual features. This makes it highly effective in identifying subtle inconsistencies or edits in news-related images. The model is trained on a labeled dataset of real and fake images, enabling it to distinguish authentic visuals from those intended to mislead the public. By incorporating ViT into the system, the model gains the capability to verify not only the text but also the supporting images, improving overall detection reliability.

1.1.4 EVALUATION STANDARD

To evaluate the performance of the fake news detection system, four key metrics are utilized: Precision, Recall, F1-Score, and Accuracy.

- Precision measures the proportion of correctly identified positive cases out of all cases predicted as positive. It reflects how reliable the system is when it classifies content as fake.
- Recall determines the proportion of actual positive cases that the model successfully identifies. It highlights the model's ability to detect all fake instances.
- F1-Score is the harmonic mean of precision and recall, offering a balanced evaluation metric, especially useful when dealing with imbalanced datasets.
- Accuracy measures the overall correctness of predictions by evaluating the ratio of correctly classified instances (both real and fake) to the total number of inputs

CHAPTER 2
LITERATURE SURVEY

2.1 INTRODUCTION

The growing prevalence of fake news across digital platforms has prompted extensive research into effective detection methods. This literature survey explores a diverse range of approaches developed over the years, including machine learning, deep learning, natural language processing, and multimodal analysis. By reviewing both classical techniques and recent innovations such as transformer models, adversarial training, and user-centered frameworks, the survey highlights key contributions, trends, and challenges in the field. The studies examined offer valuable insights into the evolution of fake news detection, the integration of linguistic and visual cues, and the development of robust, adaptable systems for real-world applications.

2.2 RELATED WORK

Ahmed et al. [1] conducted a comprehensive systematic literature review focusing on machine learning techniques for fake news detection. Their study highlighted the evolution of the field, emphasizing a growing shift toward deep learning models due to their superior performance in handling complex data. The review also explored the rising significance of multimodal approaches that analyze both text and visual data, which are increasingly relevant in the age of social media. Ahmed et al. discussed key challenges such as dataset imbalance, the difficulty of obtaining high-quality labeled data, and the need for explainable models that can enhance trust and transparency. Their work provides a valuable roadmap for future research by identifying gaps in existing studies and proposing directions for model improvement and practical deployment.

Atrey et al. [2] presented a survey on multimodal fusion techniques, primarily for multimedia analysis, which has strong implications for fake news detection. They examined various methods for combining information from different modalities, including textual, visual, and audio features. Their research emphasized that integrating multiple data types can significantly enhance model performance by providing complementary evidence, which helps in identifying inconsistencies typical of fake content. The survey underscored the importance of effective fusion strategies, such as early fusion, late fusion, and hybrid methods, to optimize information extraction and classification. These insights laid foundational concepts later adopted in fake news detection systems involving multimedia content.

Castillo et al. [3] explored the credibility of information shared on Twitter by analyzing social media-specific metadata such as retweet patterns, user profiles, and temporal dynamics. Their findings showed that fake news often exhibits unique dissemination behaviors, such as rapid spreading through low-credibility accounts or bots. By using machine learning classifiers trained on metadata features, they demonstrated that non-content-based signals could be highly effective in detecting misinformation. This work was among the first to highlight the potential of social context and propagation patterns as key indicators of content reliability, paving the way for network-based fake news detection strategies.

Chen et al. [4] proposed a deep learning model that integrates Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to detect fake news from text. The CNN component captures local word-level features and syntactic structures, while the RNN captures long-range dependencies and the sequence of ideas. By combining spatial and temporal analysis, their model outperformed traditional architectures that rely solely on one method. The hybrid framework effectively handles the linguistic complexity of deceptive content and demonstrates the benefits of architectural synergy in fake news classification tasks.

Kumar and Singh [5] reviewed the role of multimodal deep learning models in fake news detection. Their study focused on how integrating textual and visual features improves the system’s ability to analyze deceptive posts, especially those involving memes or image-caption pairs. They explained the challenges of multimodal learning, such as feature alignment and modality imbalance, and offered solutions like attention mechanisms and joint embedding techniques. Their review reinforced the idea that multimodal approaches are essential for capturing the diverse and often contradictory cues found in misleading online content.

Li et al. [6] introduced a novel approach combining contrastive learning with similarity fusion for detecting fake news using multimodal data. Their method involves learning representations that emphasize the relationships between text and accompanying images. Contrastive learning helps the model distinguish between matching and non-matching text-image pairs, while similarity fusion integrates these learned representations for classification. Their approach showed notable improvements in accuracy over unimodal methods and proved effective in identifying subtle discrepancies often found in fake news, such as mismatched visuals and misleading headlines.

Malanowska et al. [7] conducted a meta-survey on digital watermarking techniques and their applicability to fake news detection. They explored how watermarking can verify the authenticity of images and videos used in online news, which is particularly relevant for combatting manipulated multimedia. The authors argued for a multidisciplinary approach that combines digital forensics with AI-based detection to ensure content integrity. Their study emphasized the preventive potential of watermarking in the fake news ecosystem, acting as a technical barrier against media tampering and helping users assess credibility at the point of consumption.

Oshikawa et al. [8] surveyed the application of Natural Language Processing (NLP) in fake news detection. They reviewed techniques like sentiment analysis, topic modeling, fact verification, and stance detection. Their work demonstrated how linguistic cues—such as sensationalism, emotional tone, or contradiction—can signal fake content. They also emphasized the importance of labeled datasets and preprocessing techniques for training effective NLP models. The survey contributed to establishing NLP as a cornerstone of fake news detection research and outlined future directions for integrating semantic understanding and contextual reasoning into detection systems.

Pan and Yang [9] presented a broad survey on transfer learning, a technique widely used in fake news detection. They explained how pre-trained models can be fine-tuned on smaller, domain-specific datasets to achieve strong performance. Their review showed that transfer learning helps overcome data scarcity and supports the development of adaptable detection systems suitable for evolving misinformation trends.

By reusing knowledge learned from large general datasets, models can quickly adapt to new topics, events, or domains where labeled data is limited.

Park and Chai [10] explored user-centered models for fake news detection that prioritize the end-user experience. Their research combined machine learning classification techniques with insights from user engagement behavior, such as click patterns and reading habits. They stressed that tailoring detection systems to user preferences enhances trust and usability. By focusing on the consumer side of the problem, their study introduced a more holistic approach that considers not just the accuracy of detection, but also how users interact with and respond to flagged content in practical scenarios.

Pate and Ibrahim [11] explored the role of fake news and hate speech in Nigeria's democratic consolidation. Their conceptual review analyzed how misinformation undermines political discourse, inflames ethnic tensions, and distorts electoral processes. They advocated for a multi-pronged response involving media literacy, legal reform, and technology-based solutions like automated fake news detection systems. Their work emphasized the broader societal impacts of misinformation and the importance of integrating sociopolitical considerations into technological interventions.

Shu et al. [12] provided an extensive review on fake news detection using data mining techniques. They categorized research efforts into content-based, social context-based, and propagation-based methods. They identified challenges like limited labeled data, the need for real-time detection, and feature engineering complexities. Their review helped shape the early research landscape by connecting fake news detection to existing data mining paradigms and promoting interdisciplinary collaboration among computer science, linguistics, and social science researchers.

Shu et al. [13] also explored the dynamic nature of misinformation on social media, emphasizing the importance of temporal analysis and adaptive models. They suggested that models must continuously learn from new data streams to remain effective. Their insights reinforced the value of incorporating user interactions, content evolution, and dissemination timelines into detection algorithms. This adaptive approach aligns with the fast-changing nature of misinformation, particularly during crisis events and breaking news situations.

Singhania et al. [14] introduced a Hierarchical Attention Network (3HAN) for fake news classification. The model uses multiple layers of attention to focus on the most relevant words, sentences, and overall document structure. This hierarchical modeling reflects the way humans process text and improves interpretability. By selectively emphasizing important textual elements, their model achieved high accuracy and offered explainable outputs, which are crucial for building user trust in AI-powered detection tools.

Subrahmanian et al. [15] conducted the DARPA Twitter Bot Challenge, which examined the role of automated bots in spreading fake news. They designed experiments to detect and neutralize bots using behavioral and content-based features. Their findings showed that bots significantly amplify misinformation and can be targeted using AI-driven countermeasures. This work highlighted the intersection of bot detection and fake news prevention and stressed the need for comprehensive tools that address both content and distribution mechanisms.

Tariq et al. [16] investigated adversarial training techniques for fake news detection. They crafted adversarial examples designed to mislead models and trained their systems to resist such attacks. This method improves the robustness and security of detection models, particularly against sophisticated misinformation strategies. Their work underscored the need for models that not only perform well in standard scenarios but also maintain reliability under adversarial conditions, making them suitable for deployment in high-stakes environments.

Volkova et al. [17] combined stylometric and contextual features in a classification model to detect fake news. Stylometric features capture writing style, such as sentence length and punctuation patterns, while contextual features reflect external elements like topic or user profile. Their hybrid approach successfully identified deceptive content by detecting inconsistencies between expected and actual writing or context. This dual-feature method broadened the scope of fake news detection by integrating deeper linguistic and behavioral cues.

Wang [18] introduced the "LIAR" dataset, which has become a benchmark for fake news detection tasks. Comprising over 12,000 labeled political statements, the dataset categorizes truthfulness on a fine-grained scale from true to pants-on-fire. The availability of such a resource enabled rigorous model training and comparison across studies. Wang’s contribution facilitated standardization in the field, allowing researchers to evaluate and improve models in a consistent manner.

Wang et al. [19] developed a soft-label framework for multi-domain fake news detection. Unlike binary classifiers, their model assigns probabilistic labels to reflect the uncertainty often present in fake news. This approach enhances generalization across domains, such as politics, health, and entertainment, which have varying stylistic and semantic characteristics. Their framework demonstrated improved performance in domain adaptation tasks, addressing a common limitation in earlier models that struggled with unseen topics.

Zhou and Chen [20] reviewed transformer-based models, such as BERT and its variants, for fake news classification. They explored how attention mechanisms and deep contextual embeddings help capture complex linguistic relationships and detect subtle deception cues.

2.3 INFERENCE FROM THE RELATED WORK

The surveyed literature reveals a progressive evolution in fake news detection, emphasizing the transition from traditional machine learning methods to more advanced deep learning and multimodal approaches. Early works by Castillo et al. and Shu et al. established the importance of social context and propagation patterns, laying the groundwork for metadata and network-based detection. Simultaneously, content-based methods using NLP and stylometry gained momentum, as highlighted by Oshikawa et al. and Volkova et al., proving that linguistic signals are critical indicators of fake content. Recent studies have increasingly shifted towards multimodal and hybrid techniques, as evident in the work of Kumar and Singh, Li et al., and Atrey et al. These approaches underscore the need to analyze both text and images to detect mismatches and inconsistencies in modern misinformation, especially on visually rich platforms. The incorporation of deep models such as CNNs, RNNs, and Transformers, as discussed by Chen et al. and Zhou & Chen, showcases improved performance in capturing complex semantic and contextual cues.

The use of transfer learning and domain adaptation (Pan & Yang; Wang et al.) addresses the challenge of data scarcity and domain shifts, enabling models to generalize better across topics. Additionally, robust and secure systems are being developed using adversarial training techniques (Tariq et al.), and the importance of explainability and user-centered design (Park & Chai; Singhania et al.) is increasingly recognized for practical deployment.

Overall, the related works converge on the need for an integrated, scalable, and adaptive fake news detection framework. This inference directly supports the methodology of your project, which employs both textual and image analysis using deep learning and Transformer models, integrates user interaction, and considers robustness, accuracy, and explainability.

CHAPTER 3
PROBLEM FORMULATION AND OBJECTIVIES

3.1 PROBLEM STATEMENT

In the digital age, the rapid spread of misinformation and fake news through online platforms, particularly social media, poses a significant threat to public opinion, social harmony, and democratic processes. Traditional detection methods that rely solely on text analysis are often insufficient, especially as fake content increasingly incorporates manipulated images and multimedia elements to deceive users. There is a critical need for a robust, accurate, and scalable system capable of analyzing both textual and visual information to effectively identify and classify fake news. The challenge lies in developing an integrated approach that leverages advanced machine learning and deep learning techniques to process multimodal data, handle evolving misinformation tactics, and provide reliable real-time predictions to end-users.

3.2 RESEARCH OBJECTIVES

- To develop a robust fake news detection system that leverages both textual and image-based data for enhanced classification accuracy.
- To implement machine learning and deep learning models such as SVM, Naive Bayes, CNN, and Vision Transformers (ViT) for effective content analysis.
- To preprocess and extract meaningful features from text using TF-IDF and from images through normalization and augmentation techniques.
- To explore the effectiveness of multimodal approaches by integrating text and image analysis for identifying fake news with higher reliability.
- To evaluate the system’s performance using standard metrics such as precision, recall, F1-score, and accuracy.
- To deploy the model as a web-based platform enabling real-time user interaction for fake news verification.
- To ensure the system's scalability and usability, with potential for integration into social media platforms and news aggregator tools.

CHAPTER 4
SYSTEM DESIGN

4.1 INTRODUCTION

The system design outlines the architecture for fake news detection using both text and image inputs. It consists of two parallel pipelines—one for processing textual data using CNN algorithms and another for analyzing images with transformer models. Each pipeline includes stages like preprocessing, feature extraction, and model building to accurately classify content as real or fake.

4.2 SYSTEM ARCHITECTURE

An The system comprises a modular dual-pipeline architecture designed for fake news detection, handling both text and image data independently. The textual pipeline begins with preprocessing the text dataset, followed by feature extraction and classification using a CNN-based algorithm. The user inputs a news article or snippet, which is analyzed and classified as fake or real based on the trained model. Simultaneously, the image pipeline starts with preprocessing and feature extraction from an image dataset. A transformer-based algorithm then processes the extracted features to build a separate model that also outputs a fake or real classification. This separation ensures efficient parallel processing, enhances scalability, and allows each pipeline to be optimized independently for its data type. By maintaining independent yet integrated flows for text and image, the system increases detection accuracy and adaptability, making it suitable for analyzing multimodal misinformation commonly found in social media and digital news platforms.

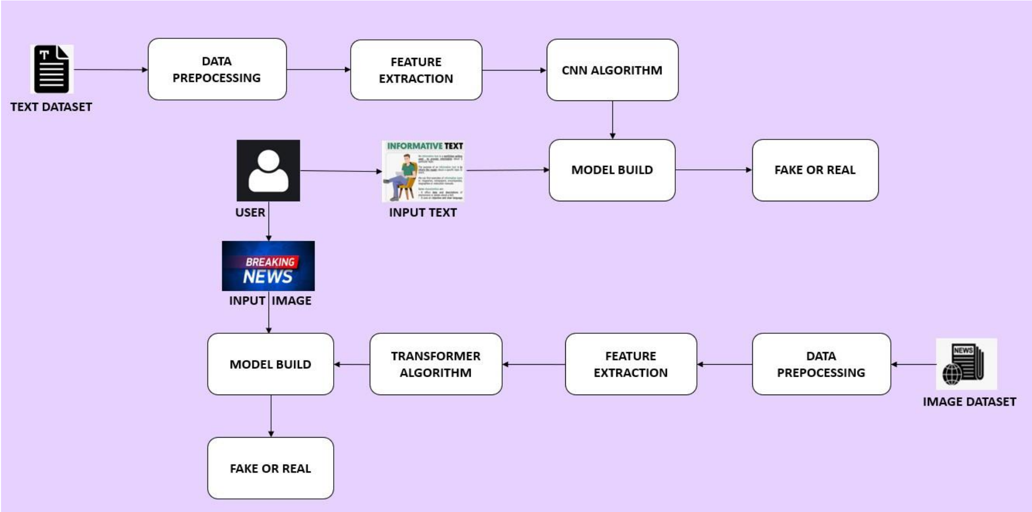

Figure 4.1 Architecture Design of the Proposed System

Figure 4.1 shows the system architecture designed which features Key Components of the Architecture Diagram. The proposed system architecture for fake news detection integrates two primary pipelines: one for text data and another for image data, enabling a multimodal detection approach.

1. **TEXT DATASET & PREPROCESSING**

The pipeline begins with a text dataset, which undergoes data preprocessing to remove noise, stopwords, and irrelevant tokens. This prepares the data for more accurate feature extraction.

2. **FEATURE EXTRACTION (TEXT)**

Key linguistic features are extracted from the cleaned text using methods like TF-IDF or word embeddings. These features represent the core semantics of the content.

3. **CNN ALGORITHM**

A Convolutional Neural Network (CNN) processes the extracted features, identifying patterns in text that are commonly found in fake news articles.

4. **IMAGE DATASET & PREPROCESSING**

Parallel to the text process, image data is also handled. Images are preprocessed to enhance clarity and remove noise, preparing them for feature extraction.

5. **FEATURE EXTRACTION (IMAGE)**

Critical visual features are extracted from the images to identify manipulations or inconsistencies.

6. **TRANSFORMER ALGORITHM**

A transformer model processes the extracted image features, learning contextual relationships for better classification.

7. **MODEL BUILD**

Both processed text and image data are used to build a unified model.

8. **USER INPUT & OUTPUT**

Users input text or images, and the model evaluates them as Fake or Real, based on learned patterns.

4.3 SYSTEM REQUIREMENTS

4.3.1 HARDWARE SPECIFICATION

The hardware system used for this project includes an Intel processor operating at a speed of 1.1 GHz, along with 8 GB of RAM and a 500 GB hard disk. This setup provides sufficient computational power and storage capacity to support data preprocessing, feature extraction, and model training for both textual and visual fake news detection tasks.

Table 4.1 Hardware Specifications

Component	**Specification**
Processor	Intel
Speed	1.1 GHz
RAM	8 GB
Hard Disk	500 GB

4.3.2 SOFTWARE SPECIFICATIONS

The software environment is based on Windows operating systems (8,10, or 11), with HTML and CSS used for the front-end interface. Python is the primary scripting language, facilitating model development and data handling. Python IDLE serves as the main development tool, offering a simple and effective platform for coding, testing, and executing scripts.

Table 4.2 Software Specifications

Component	Specification
Operating System	Windows 8/10/11
Front End	HTML, CSS
Scripts	Python Language
Tool	Python IDLE

CHAPTER 5
SYSTEM IMPLEMENTATION

5.1 SYSTEM METHODOLOGY

The fake news detection system follows a dual-modality approach that separately analyzes textual and visual content before delivering a final prediction. This methodology is structured into multiple stages, each responsible for a specific function in the fake news detection pipeline.

1. DATA COLLECTION AND PREPROCESSING

The system begins with collecting a dataset comprising news articles with associated text and images. Textual data is cleaned by removing special characters, stop words, and converting all text to lowercase. Tokenization and lemmatization are applied to standardize the language. Meanwhile, image data undergoes resizing, normalization, and augmentation (e.g., flipping, rotation) to improve model robustness and generalization.

2. TEXTUAL FEATURE EXTRACTION AND CLASSIFICATION

To convert textual information into machine-readable format, TF-IDF vectorization is used. These numerical features are then fed into multiple classifiers—Support Vector Machine (SVM), Naive Bayes, and Convolutional Neural Networks (CNN)—to predict whether the content is real or fake. Each model captures different linguistic patterns, improving the reliability of the classification.

3. IMAGE-BASED ANALYSIS WITH VISION TRANSFORMER

Images are processed using a Vision Transformer (ViT), which divides each image into patches and applies self-attention mechanisms to capture both local and global visual patterns. The ViT model is trained to detect signs of manipulation, deepfakes, or inconsistencies between images and associated text.

4. INTEGRATION AND WEB DEPLOYMENT

Both textual and visual outputs are combined to offer users a comprehensive prediction. The system is deployed on a web interface using Flask for backend processing and HTML/CSS/JavaScript for frontend design. Users can submit text and images in real time and receive a confidence-based result regarding the authenticity of the news.

5. EVALUATION AND OPTIMIZATION

The system's performance is evaluated using metrics such as accuracy, precision, recall, and F1-score. These metrics guide the model tuning process and help ensure that the system is reliable, scalable, and ready for real-world deployment.

5.2 MODULE DESCRIPTION

The proposed Fake News Detection System is divided into several key modules, each responsible for handling a distinct part of the process. These modules work collaboratively to ensure efficient and accurate classification of news content using both text and image data.

1. DATA COLLECTION MODULE

This module is responsible for gathering a multimodal dataset containing news articles with textual content and accompanying images. Sources include online repositories, verified news sites, and social media platforms. The dataset is manually labeled as real or fake to support supervised learning.

2. DATA PREPROCESSING MODULE

This module processes the raw data to make it suitable for machine learning models.

- Text preprocessing includes removing stop words, punctuation, numbers, and performing tokenization and lemmatization.
- Image preprocessing includes resizing, normalization, and augmentation to increase diversity and improve model generalization.

3. TEXT CLASSIFICATION MODULE

This module extracts features from text using TF-IDF vectorization. The transformed features are passed through multiple classifiers—Naive Bayes, SVM, and CNN—to detect fake news based on linguistic cues. Ensemble or weighted voting may be used to improve prediction reliability.

4. IMAGE ANALYSIS MODULE

Using a Vision Transformer (ViT), this module analyzes images associated with news articles. The ViT model detects inconsistencies, signs of manipulation, or deepfakes in the visual data, which often signal fake news.

5. FUSION AND DECISION MODULE

This module integrates predictions from both the text and image pipelines using a confidence-based mechanism. It determines the final classification—real or fake—based on the combined strength of both predictions.

6. WEB INTERFACE MODULE

A user-friendly web application built with Flask allows users to input news text and upload images. It processes inputs, communicates with the backend models, and returns predictions with a confidence score.

7. EVALUATION MODULE

This module calculates key performance metrics such as accuracy, precision, recall, and F1-score to assess the effectiveness of the system and guide future improvements.

5.2.1 TEACHER MODULE

The Teacher Module is designed to facilitate the supervision and management of fake news detection activities within the system. It begins with a Login/Authentication feature, where teachers use their credentials to securely access the platform. Once logged in, they are directed to a personalized Dashboard, which provides an overview of all user-submitted news content—both textual and visual. The dashboard also displays system-generated predictions, including whether the content is real or fake, along with user activity logs for easy monitoring.

The Content Review functionality allows teachers to manually inspect the results predicted by the models. They can verify the correctness of the classifications and provide expert feedback, which is useful for improving the model's accuracy through supervised learning. Teachers can also approve or reject content based on its authenticity, ensuring a layer of human validation.

5.2.2 KNOWDLEDGE DISTILLATION MODULE

The Knowledge Distillation Module plays a critical role in improving the efficiency and scalability of the fake news detection system. It focuses on transferring the learned knowledge from a large, complex model (often referred to as the teacher model) to a smaller, lightweight model (known as the student model). The goal of this module is to retain high accuracy and performance while reducing computational requirements, which is especially beneficial for deployment in resource-constrained environments like mobile or web applications.

In this process, the teacher model—typically a high-capacity neural network trained on large datasets—generates soft labels or probability distributions over classes. These outputs contain rich information about the decision boundaries and inter-class relationships. The student model is then trained to mimic these soft outputs instead of just the hard labels (e.g., real or fake). This enables the student model to generalize better, even with fewer parameters.

The Knowledge Distillation Module enhances model interpretability, speeds up inference time, and allows the system to maintain real-time performance without sacrificing detection accuracy. It ensures that the smaller models can still effectively identify fake news across both textual and image-based content, making the system more practical for real-world use cases.

5.2.3 CONFIDENCE BASED QUERY SELECTION MODULE

The Confidence-Based Query Selection Module is designed to enhance the efficiency and reliability of the fake news detection system by selectively identifying uncertain predictions for further review or retraining. This module operates on the principle that not all inputs are equally informative for model improvement. By focusing on low-confidence predictions, the system can dynamically refine its accuracy and robustness.

When the model analyzes a given text or image input, it generates a probability distribution across possible output classes (e.g., real or fake). The confidence score reflects how certain the model is about its prediction. If this score falls below a predefined threshold, the input is flagged by the module as a query candidate. These flagged queries are either sent for manual review, labeled through human-in-the-loop systems, or added to the training set for future fine-tuning.

This module supports active learning, allowing the system to evolve continuously by learning from the most ambiguous and challenging examples. It also helps in reducing annotation costs by prioritizing only the most impactful data points. Overall, the Confidence-Based Query Selection Module increases model reliability, improves training efficiency, and ensures adaptability to new types of fake news patterns.

CHAPTER 6

RESULTS

6.1 CLASSIFICATION PERFORMANCE METRICS

Users can submit doubtful images to the image recognition platform that is presented and receive instant suggestions regarding the authenticity of the submitted photos. The two sub-systems are then combined to be utilized in the system's user-friendly web portal. To ensure their validity, individuals can provide images through the image detection interface, while they can input or paste news content through the text detection interface. Our two-module system provides users with the freedom to identify misinformation based on visual and text data autonomously.

Table 6.1 Model Classification Performance

CLASS	PRECISION	RECALL	F1-SCORE	SUPPORT
0(Real)	1.00	0.93	0.97	120
1(Fake)	0.94	1.00	0.97	120
Accuracy		0.97		240

Table 6.1 presents the classification performance of the model across two classes: Real (0) and Fake (1). The model demonstrates strong predictive capabilities, achieving a high overall accuracy of 97% on a balanced test set of 240 samples. For the Real class, the model achieves perfect precision (1.00), indicating no false positives, with a recall of 0.93, reflecting a small number of false negatives. Conversely, the Fake class exhibits slightly lower precision (0.94) but perfect recall (1.00), implying all fake news instances were correctly identified. The F1-scores for both classes are equally high (0.97), confirming the model's balanced performance in terms of precision and recall. These results indicate that the model is highly effective in differentiating between real and fake news content.

6.2 MODEL TRAINING PROGRESS

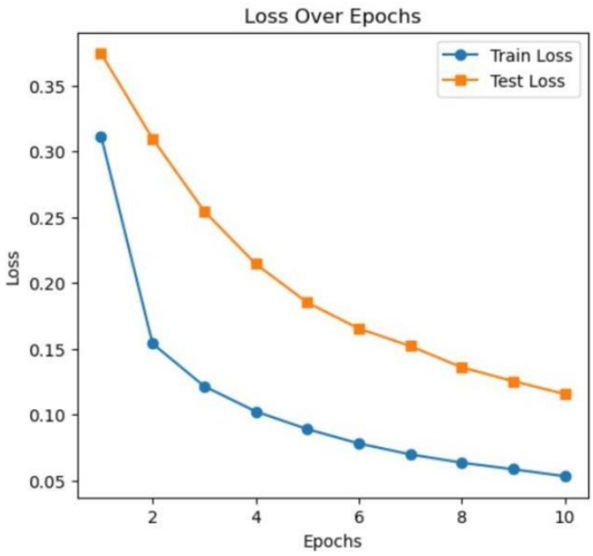

Figure 6.1 Training and Testing Loss Over Epochs

Figure 6.1 shows the training and testing loss trends over 10 epochs during model training. As depicted, both the train loss (blue line) and test loss (orange line) consistently decrease with each epoch, indicating that the model is learning effectively and generalizing well. The training loss drops more rapidly in the initial epochs, suggesting efficient convergence, while the test loss also shows a steady decline, though at a slightly higher rate than the train loss throughout. The absence of divergence between the two curves implies that the model is not overfitting and maintains a balanced performance on both seen and unseen data.

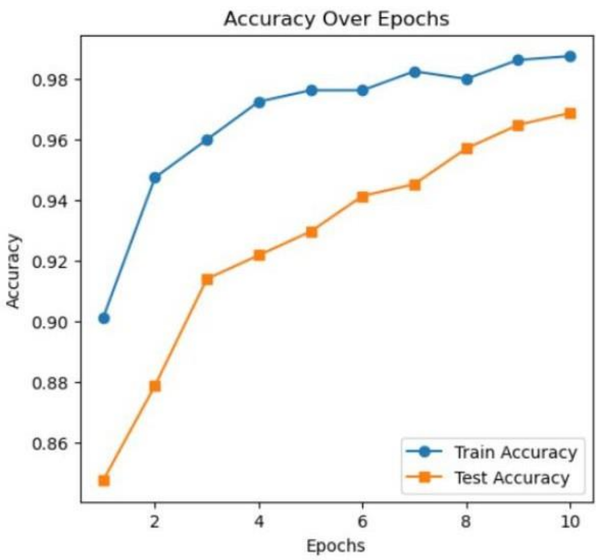

Figure 6.2 Training and Validation Accuracy plots

Figure 6.2 shows the training and testing loss trends over 10 epochs during model training. As depicted, both the train loss (blue line) and test loss (orange line) consistently decrease with each epoch, indicating that the model is learning effectively and generalizing well. The training loss drops more rapidly in the initial epochs, suggesting efficient convergence, while the test loss also shows a steady decline, though at a slightly higher rate than the train loss throughout. The absence of divergence between the two curves implies that the model is not overfitting and maintains a balanced performance on both seen and unseen data.

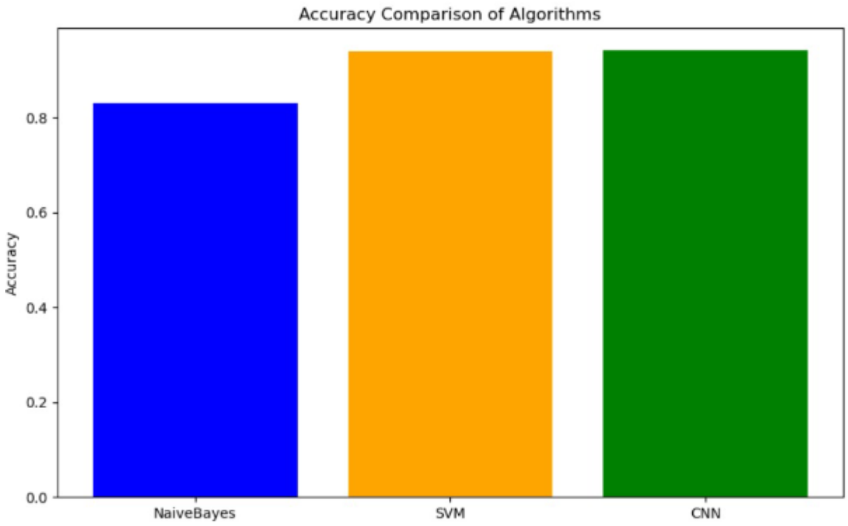

Figure 6.3 Accuracy Comparison of Different Algorithms for Fake News Detection.

Figure 6.3 shows the training and testing loss trends over 10 epochs during model training. As depicted, both the train loss (blue line) and test loss (orange line) consistently decrease with each epoch, indicating that the model is learning effectively and generalizing well. The training loss drops more rapidly in the initial epochs, suggesting efficient convergence, while the test loss also shows a steady decline, though at a slightly higher rate than the train loss throughout. The absence of divergence between the two curves implies that the model is not overfitting and maintains a balanced performance on both seen and unseen data.

6.3 CONFUSION MATRIX ANALYSIS

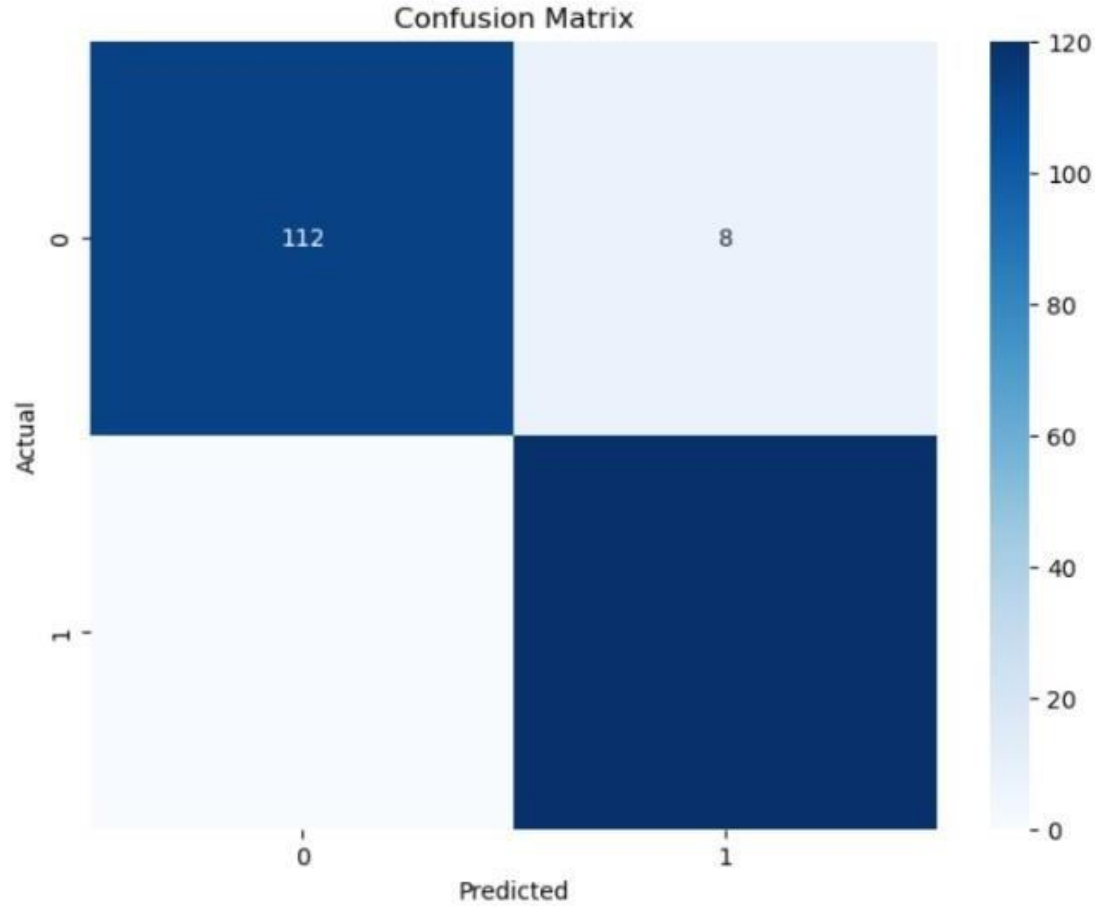

Figure 6.4 Confusion Matrix

Figure 6.4 presents the confusion matrix evaluating the performance of the fake news detection model. The matrix shows that out of the total predictions, 112 true negatives and 112 true positives were correctly classified, with only 8 false positives and 0 false negatives. This indicates that the model has a high degree of accuracy and recall, especially in detecting fake news. The absence of false negatives demonstrates the model's effectiveness in ensuring that fake news items are not incorrectly classified as real, which is crucial for minimizing misinformation spread. These results validate the model's reliability in both text and image-based fake news detection scenarios.

6.4 FAKE NEWS DETECTION USER INTERFACE

6.4.1 UPLOAD INTERFACE

Figure 6.5 Landing Page of the Fake News Detection Web Application.

Figure 6.5 illustrates the user interface of the fake news detection system, which offers two primary functionalities: detecting fake news through image analysis and textual content analysis. The interface is visually engaging and user-friendly, featuring clearly labeled buttons for “Fake News Image” and “Fake News Text.” A bold, thematic background reinforces the seriousness of the problem, while a quote at the top emphasizes the critical nature of trust in the fight against misinformation. This dual-module design allows users to interact with the system intuitively, supporting the project's objective of making fake news detection accessible and efficient.

6.4.2 PREDICTION RESULTS PAGE

Figure 6.6 Fake news detection system classifying input text as real news.

Figure 6.6 displays the output interface of the fake news text detection module. The system processes the input text and provides a prediction—here, identifying the statement as "Looking Real News." The visual design incorporates a modern, global-themed background, reinforcing the widespread impact of misinformation. The text analysis window highlights the user-entered news content, and a prominent "Predict" button initiates the model’s classification. This output demonstrates the model's capability to effectively analyze textual claims and deliver intuitive results, aligning with the project goal of supporting trust and accuracy in digital news consumption.

Figure 6.7 Fake news detection system classifying input text as fake news.

Figure 6.7 presents an example of the system detecting and flagging potentially harmful or misleading content. The input text contains provocative language and misinformation involving public figures, which the model successfully identifies as "Looking Spam News." This classification highlights the system’s sensitivity to sensationalism and aggressive rhetoric, reinforcing its ability to distinguish spam or clickbait news from credible sources. The consistent interface design, with a predictive feedback mechanism, ensures clarity and reinforces user confidence in the model’s interpretive capability, particularly in detecting socially and politically sensitive fake news narratives.

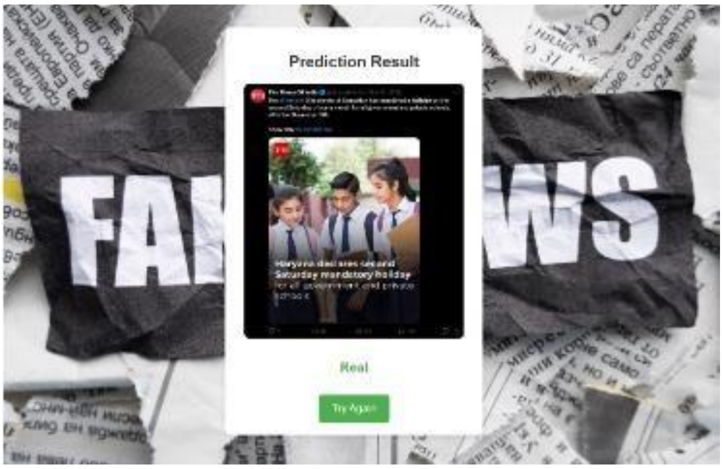

Figure 6.8 Fake news detection system classifying input image as real news.

Figure 6.8 showcases the output of the image-based fake news detection module. The system analyzes a news image featuring a school-related announcement and classifies it as “Real,” as indicated by the green label at the bottom. This prediction demonstrates the model’s effectiveness in interpreting visual content and associated text to determine the authenticity of news imagery. The background featuring the word “FAKE NEWS” torn apart adds visual emphasis to the core objective of the system—to filter truth from misinformation. This result supports the system’s utility in validating school-related news images, a common target of fake news circulation on social media.

Figure 6.9 Fake news detection system classifying input image as fake news

Figure 6.9 illustrates the system’s ability to detect fake news in image format. The image presents a satirical or exaggerated news headline claiming the removal of math from school curriculums due to difficulty. The model successfully identifies this as “Fake,” as displayed by the green label at the bottom. This outcome highlights the robustness of the fake news detection system in flagging manipulated or humorous visual content that may spread misinformation. By accurately classifying such exaggerated posts, the system proves effective in safeguarding educational integrity against misleading visuals shared online.

CHAPTER 7
CONCLUSION AND FUTUREWORK

7.1 CONCLUSION

An intensive fake news detection system based on text and image information has been developed and explained in this study. To determine whether news content is genuine or fake, the model integrates machine learning and deep learning techniques.

To capture essential text patterns, methods like CNN, SVM, and Naive Bayes are applied for text classification after feature extraction using the TF-IDF vectorizer. Image data is processed using a Transformer-based model capable of extracting intricate visual features, enhancing the system's accuracy. The proposed dual-pipeline approach ensures that text and image data are analyzed independently, providing users with reliable and accurate predictions. Metrics such as F1-score, recall, precision, and accuracy validate the model's effectiveness in identifying fake news.

The system offers significant value to social media platforms, news agencies, and readers by enabling the verification of news story accuracy. Its capability to detect misinformation can help mitigate the spread of fake news. For future enhancements, the system could incorporate video information and employ multimodal fusion techniques to further improve its accuracy and reliability. Expanding the dataset with diverse samples and integrating real-time fact-checking APIs would also strengthen the system's performance, making it more effective for real-world applications.

7.2 FUTURE WORK

The Fake News Detection System can be improved by integrating real-time fact-checking, enabling dynamic verification with verified sources. Expanding multi-lingual support and incorporating deepfake detection will enhance its ability to identify fake news across various formats. Optimizing model architectures and cloud-based deployment will improve scalability and efficiency. A mobile application can enhance accessibility, while explainable AI (XAI) will provide transparency by ex-plaining fake news classifications. These enhancements will strengthen the system’s impact in combating digital misinformation enhancing the model's resistance to adversarial attacks should be prioritized.

APPENDICES

A. SCREENSHOTS

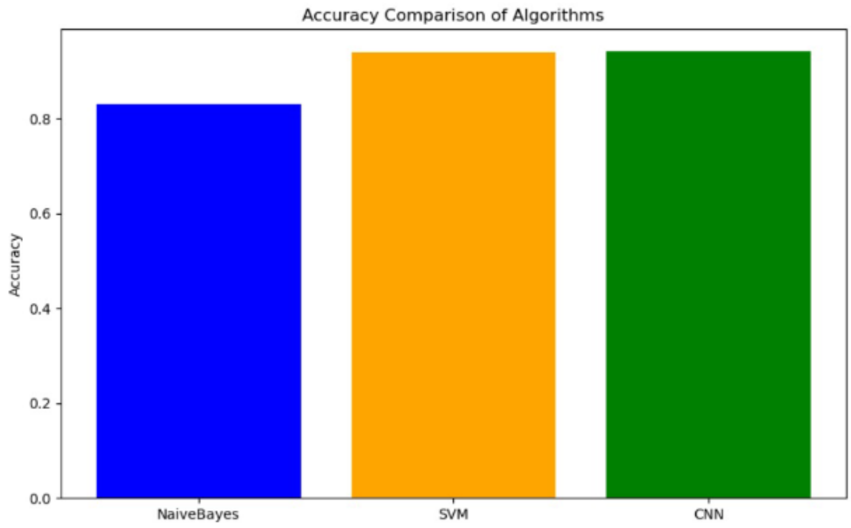

Figure A.1 Accuracy Comparison of Different Algorithms for Fake News Detection.

Figure A.2 Fake news detection system classifying input text as real news.

Figure A.3 Fake news detection system classifying input text as fake news.

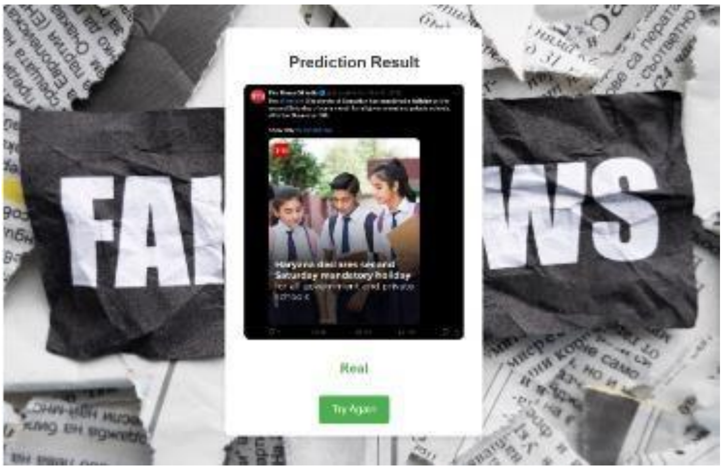

Figure A.4 Fake news detection system classifying input image as real news.

Figure A.5 Fake news detection system classifying input image as fake news

B. PAPER PUBLICATION

Team ID: B014

Team Member: 211501080 Reshma Yasmin MA

211501099 Shruthi G

Team Supervisor: Dr. K. Sekar M.E., Ph.D.,

Publication Status: Communicated

Particulars of Conference: The International Conference on Computing Technologies (ICOCT 2025)

Host Institution: Jyothi Institute of Technology

Conference Date: 13-14 June 2025

4/13/25, 11:25 AM Gmail - International Conference on Computing Technologies : Submission (2034) has been created.

Gmail RESHMA YASMIN <reshmayasmin0912@gmail.com>

International Conference on Computing Technologies : Submission (2034) has been created.

Microsoft CMT <noreply@msr-cmt.org> Fri, Apr 11, 2025 at 12:37 PM
To: reshmayasmin0912@gmail.com

Hello,

The following submission has been created.

Track Name: ICOCT2025

Paper ID: 2034

Paper Title: Hybrid-CNN Classifier for Fake News Detection: A Multimodal Defense Against Digital Misinformation

Abstract:
The rapid dissemination of fake news across online platforms poses a serious threat to public trust, social cohesion, and informed decision-making. To address this growing concern, the present work introduces a dual-modal fake news detection system that synergizes textual and visual analysis for enhanced accuracy and reliability. At the core of the system is a Hybrid-CNN Classifier, designed to process and classify text using convolutional neural networks for feature extraction, followed by classification via Support Vector Machine (SVM) and Naive Bayes to optimize decision boundaries and improve classification robustness. The system integrates an image classification model based on Transformer architecture, enabling the identification of manipulated or misleading visuals often used to support fabricated narratives. Both modules function independently to assess different types of input—text or image—offering flexibility and precision across content formats. As a web-based platform, users can submit text or images to instantly verify the authenticity of news content. This multimodal approach significantly improves detection accuracy by examining both the language and visual cues of online content. Experimental results validate the model's effectiveness, offering a scalable, real-time solution to combat the proliferation of misinformation in the digital age.

Created on: Fri, 11 Apr 2025 07:07:34 GMT

Last Modified: Fri, 11 Apr 2025 07:07:34 GMT

Authors:
- pksekarksr@gmail.com (Primary)
- reshmayasmin0912@gmail.com
- gshruthi2002@gmail.com
- tasdheepa@gmail.com

Secondary Subject Areas: Not Entered

Submission Files:
Fake_news.pdf (4 Mb, Fri, 11 Apr 2025 07:07:26 GMT)

Submission Questions Response: Not Entered

Thanks,
CMT team.

To stop receiving conference emails, you can check the 'Do not send me conference email' box from your User Profile.

Microsoft respects your privacy. To learn more, please read our Privacy Statement.

https://mail.google.com/mail/u/0/?ik=5185f5cb34&view=pt&search=all&permmsgid=msg-f:1829089058255138570&simpl=msg-f:1829089058255138570 1/2 turnitin

Similarity Report ID: oid:29427:448382887

PAPER NAME

PAPER_CONFERENCE.pdf

WORD COUNT	CHARACTER COUNT
4409 Words	**25591 Characters**
PAGE COUNT	FILE SIZE
7 Pages	**5.0MB**
SUBMISSION DATE	REPORT DATE
Apr 12, 2025 1:36 PM UTC	**Apr 12, 2025 1:36 PM UTC**

12% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 6% Internet database
- 5% Publications database
- Crossref database
- Crossref Posted Content database
- 7% Submitted Works database

Excluded from Similarity Report

- Bibliographic material
- Quoted material

Hybrid-CNN Classifier for Fake News Detection: A Multimodal Defense Against Digital Misinformation

Sekar K
Department of AIML
Rajalakshmi Engineering College
Chennai,India
pksekarksr@gmail.com

Reshma Yasmin M.A
Department of AIML
Rajalakshmi Engineering College
Chennai, India
reshmayasmin0912@gmail.com

Shruthi G
Department of AIML
Rajalakshmi Engineering College
Chennai, India
gshruthi2002@gmail.com

Dheepa T
Department of Computer Science and Engineering
Kongunadu College of Engineering and Technology
Tiruchirappalli, India
tasdheepa@gmail.com

***Abstract*—The widespread circulation of false information through digital media presents a significant challenge to public confidence, societal harmony, and informed decision-making. To tackle this issue, this work proposes a dual-input fake news detection framework that leverages both textual and image- based analysis for improved performance and trustworthiness. Central to the system is a Hybrid-CNN model that extracts features from text using convolutional layers and subsequently classifies them using Support Vector Machines (SVM) and Naive Bayes algorithms, enhancing classification precision. Alongside, a Transformer-based image analysis module is incorporated to detect manipulated or deceptive visuals frequently used in misleading content. These components operate independently, enabling accurate assessment of both text and images, thereby ensuring adaptability across content types. Offered as an online application, the platform allows users to verify the credibility of news through text or image submissions in real time. By analyzing both linguistic patterns and visual elements, the system achieves higher accuracy in identifying misinformation. The proposed solution demonstrates strong performance in experimental evaluations, presenting a robust and scalable method to counteract the spread of fake news in today's digital landscape.**

***Index Terms*—Misinformation detection, Convolutional Neural Network (CNN), Support Vector Machine (SVM), Naive Bayes classifier, Transformer-based models.**

I. Introduction

The rapid evolution of social media and digital platforms has fundamentally transformed how information is created, consumed, and shared. However, this transformation has also facilitated the widespread distribution of false information, posing risks to democratic systems, societal harmony, and the credibility of public discourse [1]. Fake news refers to content that is deliberately fabricated or deceptive but presented in the format of legitimate news. It is often used to manipulate public perception, push political narratives, or generate online engagement for profit [2]. Its global impact has been evident in critical areas such as elections, financial markets, and public health, particularly during events like the COVID-19 pandemic [3]. Traditional methods for identifying misinformation rely heavily on manual fact-checking, a time-consuming and resource-intensive process that struggles to scale with the high volume and speed of digital content generation. This has driven a strong need for automated detection systems that leverage the advancements in machine learning (ML) and deep learning (DL) technologies [4], [5]. Text-based fake news detection has largely been enabled by natural language processing (NLP) techniques, which analyze sentiment, linguistic structure, and semantic meaning to determine authenticity [6]. Popular models such as Convolutional Neural Networks (CNNs), Support Vector Machines (SVMs), and Naive Bayes classifiers have shown promising results in extracting meaningful patterns from news articles for classification purposes [7], [8]. Nonetheless, these models often fall short in scenarios where misleading visual elements—such as doctored images or deepfakes—are embedded to reinforce credibility [9]. To overcome this limitation, researchers have begun exploring multimodal approaches that combine textual and visual data for a more holistic understanding of content [10], [11]. By examining both the textual narrative and the accompanying imagery, these systems significantly enhance detection capabilities. Transformer-based models, known for their powerful self-attention mechanisms and parallel processing capabilities, have emerged as highly effective tools in tackling complex image classification challenges [12]. Motivated by this need, the current work presents a dual-modal fake news detection framework that unifies both text and image analysis. Unlike conventional systems that treat these modalities separately, the proposed method integrates them to better capture the complexity of misinformation and improve overall prediction accuracy [13]. For the textual component, this study evaluates the performance of SVM, Naive Bayes, and CNN models, while for the visual component, a Transformer-based classifier is employed. Users can interact with the system through a user-friendly web interface, where they can submit news text or related images for immediate verification. This not only automates the evaluation process but also equips users with tools to critically assess the authenticity of online content [14]. Overall, the proposed framework contributes to the fight against misinformation by providing a scalable, effective, and intelligent solution based on modern ML and DL methods. Through the combined use of linguistic and visual features, it enhances the reliability and robustness of fake news detection, offering a foundation for future innovation in digital verification technologies [15]. As disinformation techniques continue to evolve—especially with the rise of AI-generated content such as deepfakes and synthetic media—there is an urgent need for flexible systems capable of adapting to new threats. This research aims to address that challenge by delivering a scalable and adaptive architecture capable of handling diverse forms of deceptive content across textual and visual domains.

II. RELATED WORKS

With the surge in misinformation across social networking platforms, considerable research has been directed toward developing reliable fake news detection systems. Early strategies predominantly concentrated on analyzing textual content, utilizing linguistic features such as word usage, sentiment orientation, and syntactic structures. These approaches commonly employed machine learning algorithms including Naive Bayes, Support Vector Machines (SVM), and decision trees to classify content as authentic or deceptive [1], [2]. While effective to an extent, these models often lacked the ability to interpret nuanced meanings and deeper contextual relationships within the text. The emergence of deep learning introduced more advanced models like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), which significantly improved performance by capturing semantic and contextual nuances in textual data [3]. In addition to linguistic cues, some researchers began leveraging auxiliary information such as user engagement patterns, social context, and content dissemination pathways to improve model accuracy. Wang et al. [4], for instance, employed adversarial training techniques to increase robustness against manipulated inputs, while other studies incorporated behavioral features to enhance model effectiveness [5]. Parallel to these text-centric approaches, image-based fake news detection has gained traction in re- cent years. Recognizing that fabricated news often includes manipulated visuals to increase credibility, researchers have utilized CNNs and other deep learning architectures to detect such tampered imagery [6]. However, one major limitation of these models is their isolated operation from text-based systems, which reduces their overall effectiveness in handling multimodal misinformation. To overcome these limitations, multimodal approaches have emerged that combine textual and visual data. Li et al. [7] proposed a model integrating similarity fusion and contrastive learning, which improved detection accuracy by jointly analyzing both modalities. Furthermore, the use of Transformer-based architectures and attention mechanisms has enabled more effective alignment between textual and visual representations, leading to enhanced performance in recent studies [8]. Despite notable progress, challenges persist in effectively fusing text and image modalities and ensuring generalization across varied datasets. Building upon this foundation, the current study presents a unified framework that integrates a Transformer-based image classifier with traditional machine learning techniques for text analysis. This dual-modal strategy is designed to provide a more comprehensive and robust solution for detecting fake news.

III. DEVELOPED SYSTEM

The proposed system aims to identify misleading content in both textual and visual formats by leveraging a blend of machine learning algorithms and deep learning models. Built as a modular, web-based application, the framework is designed for separate processing of text and images, while also being adaptable for future integration of multimodal analysis. The system is organized into five main components: evaluation metrics to measure the effectiveness of the classifiers, integration modules to manage the communication between system parts, text analysis and classification using linguistic features to determine authenticity, image processing and classification using Transformer-based models to detect manipulated visuals, and an architectural overview that outlines the flow and structure of the entire platform. This flexible and scalable architecture ensures the system is capable of evolving alongside the growing complexity of online misinformation.

A. *System Architecture Overview*

The system's architecture is designed to handle both textual and visual inputs simultaneously for detecting fake news, with two parallel processing pipelines—one dedicated to text and the other to images—as illustrated in Fig. 1. Users can submit either type of input or both, and each is analyzed independently by specialized models. When a user enters news text, it is first processed through a text pipeline that begins with data cleaning, where unnecessary elements such as punctuation, stop words, and special characters are removed to prepare the data for analysis. To extract relevant information from the cleaned text, the Term Frequency-Inverse Docu- ment Frequency (TF-IDF) technique is applied, converting the text into numerical vectors that highlight the significance of individual words within the content. These vectors are then classified using models such as CNN, SVM, or Naive Bayes to determine the authenticity of the news, producing a result that indicates whether it is genuine or fabricated. In the image pipeline, users upload a news-related visual, which is subjected to preprocessing steps like normalization of pixel values and resizing for consistent model input. Key visual patterns are then identified using feature extraction, and these are passed through a Transformer-based model optimized for high-performance image classification. The model evaluates the visual data and outputs a prediction indicating if the image is associated with real or fake news. The overall system structure enables both modalities to function independently in real time, supporting a robust and extendable framework for misinformation detection.

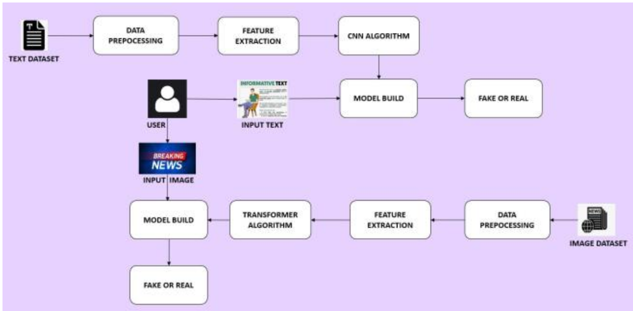

Fig. 1. Overview of the Developed System Architecture for Detecting Fake News from Textual and Visual Inputs.

B. Textual Data Processing and Classification

Text data plays a vital role in detecting fake news, as genuine and fabricated articles usually differ significantly in terms of vocabulary usage, sentence structure, and writ- ing style. The system begins by collecting labeled datasets comprising both real and misleading news samples. Before analysis, the data undergoes several cleaning steps to ensure its quality and suitability. This involves stripping out HTML

tags, unnecessary symbols, and numeric values that do not offer meaningful contributions to the classification task. To

maintain uniformity, all characters are converted to lowercase, and common stop words like ”the,” ”is,” and ”and” are removed, as they provide little value for prediction. The text is then broken down into smaller components or tokens—a step known as tokenization—which simplifies further analysis. Additionally, techniques like stemming and lemmatization are applied to reduce words to their root forms, allowing the model to interpret different variations of a word as the same base concept. After preprocessing, the next essential stage is transforming this cleaned text into a numerical format that machine learning algorithms can understand. For this, the system uses Term Frequency-Inverse Document Frequency (TF-IDF), a method that evaluates the relevance of words by weighing their occurrence within an article against their distribution across the dataset. Words that appear frequently in specific documents but not across the entire corpus are given higher importance, thus improving their ability to help the model distinguish between factual and false content. This strategic focus on the most informative terms enhances the precision and overall performance of the fake news detection system.

$$TF\text{-}IDF(t,d) = TF(t,d) \times \log \frac{N}{DF(t)} \quad (1)$$

Where:

- $TF(t, d)$ is the frequency of term t in document d
- N is the total number of documents
- $DF(t)$ is the number of documents containing term t

The system incorporates text-based fake news detection and performs a comparative analysis of three different machine learning algorithms. Support Vector Machine (SVM) is one of the supervised learning methods employed, which functions by learning from labeled training data to predict class labels for new, unseen inputs. SVM identifies the optimal decision boundary—also known as a hyperplane—that best separates the two categories: real news and fake news. It aims to maximize the distance between this boundary and the closest data points from each class, which improves the model’s ability to generalize when handling new data. The mathematical formulation below represents the optimization problem that SVM is designed to solve:

$$\min_{w,b} \frac{1}{2}\|w\|^2 \quad \text{subject to} \quad y_i(w \cdot x_i + b) \geq 1 \quad (2)$$

Using this formulation, the model is designed to identify the hyperplane that offers the greatest separation between the two classes. Given SVM’s ability to effectively handle high-dimensional feature spaces like those generated from textual data, it proves to be a strong candidate for fake news detection.

The second algorithm under consideration is the Naive Bayes classifier—a straightforward yet powerful probabilistic technique. It is grounded in Bayes’ theorem and determines the likelihood that a data instance belongs to a particular class C, given a set of features X. The relationship is typically expressed as:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \quad (3)$$

Naive Bayes operates under the assumption that all features are independent, a simplification that not only speeds up computation but also makes it highly scalable for large datasets. This characteristic proves advantageous in text classification tasks, where word frequencies are crucial indicators of whether a news article is credible or fabricated.

The third algorithm utilized in this system is the Convolutional Neural Network (CNN). Although originally developed for image recognition, CNNs have been effectively adapted for text-based tasks. In this context, the input text is transformed into one-dimensional matrices representing word sequences. Convolutional layers are used to detect local patterns or phrases that are often associated with misleading content. Pooling layers then refine these features by reducing dimensionality while preserving essential information. Finally, fully connected layers carry out the classification based on the extracted patterns, producing a prediction about the authenticity of the news article. All three models are individually trained and evaluated. Their effectiveness is measured using performance metrics such as accuracy, precision, recall, and F1-score. The algorithm that achieves the most favorable results across these parameters is selected for deployment within the system.

C. Image Data Processing and Classification

Images often serve as impactful components in news media, enhancing the perceived authenticity of the content. However, with advancements in technology like deepfakes and manipulated visuals, the misuse of images to propagate misinformation has become a significant concern. To combat this, the system incorporates an image analysis module designed to detect subtle signs of manipulation that might not be immediately visible. The process begins with image preprocessing, where all input visuals are resized to a consistent dimension to ensure uniformity across the dataset. Pixel values are normalized to a 0–1 range, which supports efficient model learning. Additionally, data augmentation techniques such as flipping, zooming, and rotating are applied to artificially expand the dataset and promote better model generalization while minimizing overfitting. Once preprocessed, a Vision Transformer (ViT) model is used to extract features, leveraging its strength in identifying complex spatial patterns and capturing long-range dependencies within the image. The ViT model processes the image by dividing it into fixed-size patches, converting them into embeddings, and then feeding them into a self-attention mechanism, which enables the model to weigh the importance of different parts of the image. These deep representations help the system accurately classify whether an image has been tampered with or is authentic.

$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (4)$$

Where:

- Q, K, and V are the Query, Key, and Value matrices
- d_k is the dimension of the key vectors

By processing all patches in parallel, the Vision Transformer is capable of rapidly learning both local and global patterns within the image. Its attention mechanism enables the model to focus on key regions that are most informative, allowing it to detect subtle alterations and manipulations. This targeted focus significantly boosts the system's ability to discern authenticity, thereby improving the overall accuracy of fake news detection through reliable classification of images as either real or fake.

D. System Integration and Web Deployment

To effectively identify false news, the proposed system integrates both text and image analysis algorithms into a unified, user-friendly web platform. This platform offers two key functionalities that allow users to verify the authenticity of news content: validating either the textual portion of a news article or any accompanying image. This dual capability ensures that the system remains versatile and relevant for everyday use.

Users can input a news article directly into the text analysis module, where it is processed using one of the trained models—Convolutional Neural Network (CNN), Support Vector Machine (SVM), or Naive Bayes classifier. The system then generates a prediction, classifying the input as either genuine or fake, with the results clearly displayed on the interface.

Alternatively, users may upload a relevant image through the image analysis module. This input is analyzed by a Transformer-based model, known for its ability to capture complex visual dependencies. The system evaluates the image's features and delivers a prediction that indicates whether the image is real or manipulated.

can handle large volumes of data efficiently, making it scalable and well-suited for high-demand environments requiring real-time fake news detection.

The system backend is developed using Python with the Flask framework handling server-side functionality. On the frontend, HTML, CSS, and JavaScript are used to create a responsive and interactive user interface. Machine learning tasks are facilitated by libraries such as TensorFlow, Scikit-learn, and Keras. With the entire model deployed on a cloud server, the platform guarantees smooth performance, high accessibil- ity, and scalability across various devices and user loads.

E. Evaluation Metrics

The fundamental purpose of designing a system to detect fake news is to have it accurately recognize and classify news content with good accuracy. The performance of such a system has to be tested with the measures specified in this subsection.

1) Precision: Precision measures the proportion of positive predictions that are actually correct. It is useful when the cost of false positives is high.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

2) Recall: Recall indicates the proportion of actual positives that were correctly identified by the model. It is important when the cost of false negatives is high.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

3) F1-score: The F1-score represents the harmonic mean of precision and recall. It provides a balanced measure when both false positives and false negatives are considered important.

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

4) Accuracy: Accuracy evaluates the overall correctness of the model by calculating the ratio of correctly predicted instances to the total instances.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

Where:

Both modules are designed to operate independently and in parallel, enabling real-time processing without delay. This parallelism and modular architecture ensure that the system

- TP: True Positive
- TN: True Negative
- FP: False Positive
- FN: False Negative

IV. RESULTS AND DISCUSSION

The effectiveness of the proposed method for detecting false news was assessed using both textual and image-based datasets. Table 1 presents the evaluation metrics for the text classification model, demonstrating its strong capability in differentiating between real and fake news. The model achieved an accuracy of 97For Class 0 (Real News), the F1-score reached 0.97, supported by a precision of 1.00 and a recall of 0.93. In the case of Class 1 (Fake News), the model exhibited high effectiveness with a precision of 0.94 and a perfect recall of 1.00. These results highlight the model's robustness in identifying fake content accurately. Performance across both classes remains well-balanced, as indicated by the consistently high macro and weighted averages of precision, recall, and F1-score, each standing at 0.97. The confusion matrix in Fig. 2 further visualizes the model's classification outcomes, clearly distinguishing correct and incorrect predictions.

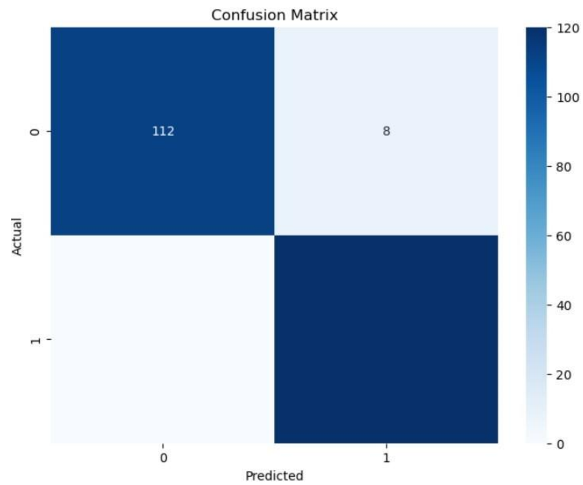

Fig. 2. Confusion Matrix

Further performance metrics are detailed in Table 1, showing precision, recall, and F1-score for both classes.

TABLE I
CLASSIFICATION REPORT OF THE PROPOSED MODEL

Class	Precision	Recall	F1-Score	Support
0 (Real)	1.00	0.93	0.97	120
1 (Fake)	0.94	1.00	0.97	120
Accuracy		0.97		240
Macro Avg	0.97	0.97	0.97	240
Weighted Avg	0.97	0.97	0.97	240

The image-based fake news detection module, built using the Vision Transformer (ViT) framework, also demonstrated dependable performance. It effectively identified manipulated images, including those with subtle inconsistencies such as irregular lighting and texture variations. Figure 3 displays the training progression of the ViT model over ten epochs, highlighting trends in accuracy and loss. The steady rise in accuracy, coupled with a gradual decline in loss values, reflects the model's learning efficiency and effectiveness in handling image-based classification tasks.

Users can submit doubtful images to the image recognition platform that is presented in Figure 4 and receive instant suggestions regarding the authenticity of the submitted photos.The output will be displayed as real and fake news as presented in Figure 4 and Figure 5.Figure 4 displays the output for real news content, whereas Figure 5 shows the output for fake news content.

The two sub-systems are then combined to be utilized in the system's user-friendly web portal. Figure 6 illustrates the accuracy of the system's algorithms, highlighting how well the

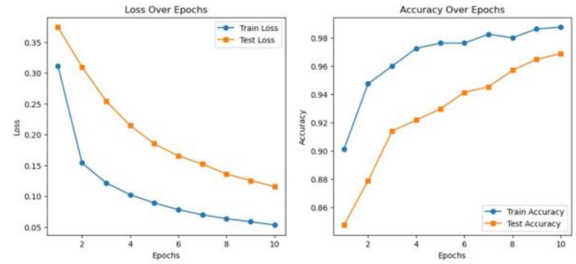

Fig. 3. Loss and Accuracy Curves of the Vision Transformer (ViT) Model Over Epochs

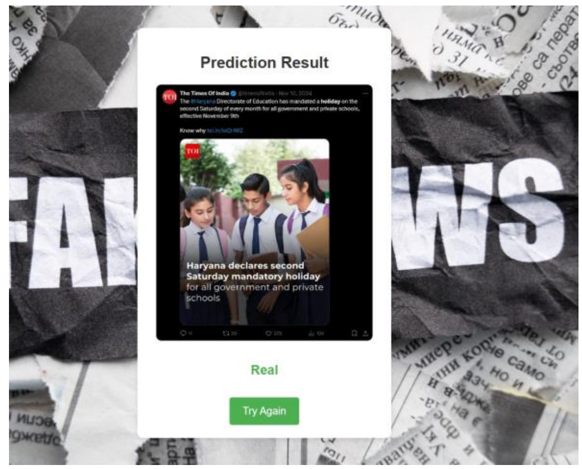

Fig. 4. Output of the system identifying the input image as authentic news content.

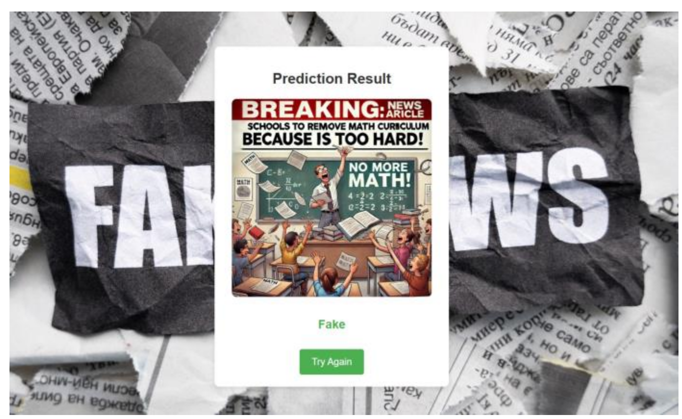

Fig. 5. Output of the system identifying the input image as falsified news content.

CNN, SVM, and Naive Bayes models fare. To ensure their validity, individuals can provide images through the image detection interface, while they can input or paste news content through the text detection interface, as illustrated in Figure 7 and Figure 8. Figure 7 displays the output for real news content, whereas Figure 8 shows the output for fake news content.

Our two-module system provides users with the freedom to identify misinformation based on visual and text data

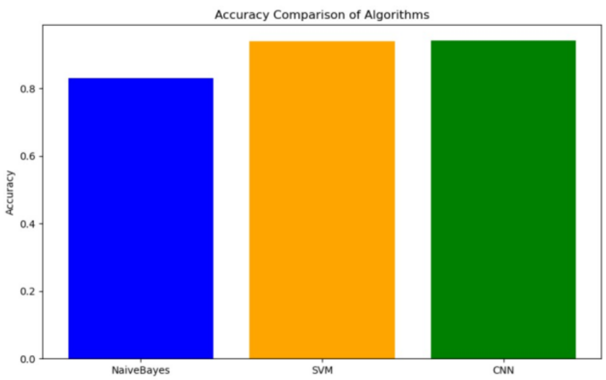

Fig. 6. Comparative analysis of algorithm performance based on accuracy metrics.

Fig. 7. System output showing classification of the input text as genuine news.

Fig. 8. System output showing classification of the input text as false news.

autonomously. Combining the Vision Transformer model for images and the Passive Aggressive Classifier for text improves the system's resilience. Expanded sets and inclusion of fact-checking APIs would be the major goals of future endeavors to improve model accuracy and real-time prediction.

V. CONCLUSION

This study presents a robust fake news detection system that leverages both textual and visual data for comprehensive analysis. The model combines traditional machine learning and advanced deep learning techniques to evaluate the authenticity of news content. For text classification, algorithms such as CNN, SVM, and Naive Bayes are employed following feature extraction using the TF-IDF vectorizer to effectively capture important linguistic patterns. To improve detection based on images, a Transformer-based architecture is utilized for its strength in extracting complex visual features. The dual-stream design enables independent processing of text and image inputs, resulting in more accurate and reliable predictions. Evaluation metrics—including F1-score, precision, recall, and accuracy—demonstrate the system's strong performance in detecting false information. This solution holds significant value for social media platforms, news organizations, and end-users by offering an effective tool to validate the authenticity of news content. Future advancements may include integration of video analysis and multimodal fusion techniques to further enhance detection capabilities.

REFERENCES

[1] U. A. Pate and A. M. Ibrahim, ”Fake News, Hate Speech and Nigeria's Struggle for Democratic Consolidation: A Conceptual Review,” Handbook of Research on Politics in the Computer Age, pp. 89-112, 2020.
[2] A. Malanowska, W. Mazurczyk, T. K. Araghi, D. Meg´ıas and M. Kuribayashi, ”Digital Watermarking—A Meta-Survey and Techniques for Fake News Detection,” IEEE Access, vol. 12, pp. 36311-36345, 2024.
[3] Y. Li, K. Jia and Q. Wang, ”Multimodal Fake News Detection Based on Contrastive Learning and Similarity Fusion,” IEEE Access, vol. 12, pp. 155351-155364, 2024.
[4] M. Park and S. Chai, ”Constructing a User-Centered Fake News Detection Model by Using Classification Algorithms in Machine Learning Techniques,” IEEE Access, vol. 11, pp. 71517-71527, 2023.
[5] D. Wang, W. Zhang, W. Wu and X. Guo, ”Soft-Label for Multi-Domain Fake News Detection,” IEEE Access, vol. 11, pp. 98596-98606, 2023.
[6] A. Tariq, A. Mehmood, M. Elhadef and M. U. G. Khan, ”Adversarial Training for Fake News Classification,” IEEE Access, vol. 10, pp. 82706-82715, 2022.
[7] B. Shu, S. Sliva, H. Wang, J. Tang and B. Liu, ”Fake News Detection on Social Media: A Data Mining Perspective,” ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22-36, 2017.
[8] P. K. Atrey, M. A. Hossain, A. El Saddik, and M. S. Kankanhalli, ”Multimodal Fusion for Multimedia Analysis: A Survey,” Multimedia Systems, vol. 16, no. 6, pp. 345-379, 2010.
[9] Z. Zhou and L. Chen, ”Fake News Detection with Transformer Models: A Review,” International Conference on Artificial Intelligence and Computer Engineering, pp. 275-280, 2023.
[10] S. J. Pan and Q. Yang, ”A Survey on Transfer Learning,” IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345-1359, 2010.
[11] R. Oshikawa, J. Qian, and W. Wang, ”A Survey on Natural Language Processing for Fake News Detection,” Proceedings of ACL, pp. 74-84, 2018.
[12] W. Y. Wang, ”Liar, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection,” Proceedings of the 55th ACL, pp. 422-426, 2017.
[13] V. S. Subrahmanian et al., ”The DARPA Twitter Bot Challenge,” Computer, vol. 49, no. 6, pp. 38-46, 2016.
[14] H. Chen, D. Kazerooni, and K. Hsu, ”Deep Learning for Fake News Detection using CNN and RNN,” IEEE International Conference on Big Data (Big Data), pp. 5378-5383, 2021.
[15] C. Castillo, M. Mendoza, and B. Poblete, ”Information Credibility on Twitter,” Proceedings of the 20th International Conference on World Wide Web (WWW), pp. 675-684, 2011.

[16] P. Singhania, B. Fernandez and S. Rao, ”3HAN: A Deep Neural Network for Fake News Detection,” IEEE International Conference on Data Mining Workshops (ICDMW), pp. 889-896, 2020.
[17] A. A. A. Ahmed, A. Aljabouh, P. K. Donepudi, and M. S. Choi, ”Detecting Fake News Using Machine Learning: A Systematic Literature Review,” arXiv preprint arXiv:2102.04458, 2021.
[18] J. Shu, D. Sliva, H. Wang, J. Tang, and B. Liu, ”Fake News Detection on Social Media: A Data Mining Perspective,” ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22-36, 2017.
[19] T. Volkova, K. Shaikh, and B. L. Gordon, ”Combining Stylometric and Contextual Features for Fake News Detection,” IEEE International Conference on Data Mining Workshops (ICDMW), pp. 821-828, 2017.
[20] A. Kumar and A. Singh, ”Fake News Detection Using Multimodal Deep Learning Models: A Review,” Journal of Information Science, vol. 49, no. 2, pp. 172-185, 2023.

REFERENCES

[1] A. A. A. Ahmed, A. Aljabouh, P. K. Donepudi, and M. S. Choi, "Detecting Fake News Using Machine Learning: A Systematic Literature Review," *arXiv preprint arXiv:2102.04458*, 2021.

[2] Atrey, P. K., Hossain, M. A., El Saddik, A., & Kankanhalli, M. S. (2010). Multimodal Fusion for Multimedia Analysis: A Survey. Multimedia Systems, 16(6), 345-379.

[3] Castillo, C., Mendoza, M., & Poblete, B. (2011). Information Credibility on Twitter. Proceedings of the 20th International Conference on World Wide Web (WWW), 675-684.

[4] Chen, H., Kazerooni, D., & Hsu, K. (2021). Deep Learning for Fake News Detection using CNN and RNN. IEEE International Conference on Big Data (Big Data), 5378-5383.

[5] Kumar, A., & Singh, A. (2023). Fake News Detection Using Multimodal Deep Learning Models: A Review. Journal of Information Science, 49(2), 172-185.

[6] Li, Y., Jia, K., & Wang, Q. (2024). Multimodal Fake News Detection Based on Contrastive Learning and Similarity Fusion. IEEE Access, 12, 155351-155364.

[7] Malanowska, A., Mazurczyk, W., Araghi, T. K., Megías, D., & Kuribayashi, M. (2024). Digital Watermarking—A Meta-Survey and Techniques for Fake News Detection. IEEE Access, 12, 36311-36345.

[8] Oshikawa, R., Qian, J., & Wang, W. (2018). A Survey on Natural Language Processing for Fake News Detection. Proceedings of ACL, 74-84.

[9] Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345-1359.

[10] Park, M., & Chai, S. (2023). Constructing a User-Centered Fake News Detection Model by Using Classification Algorithms in Machine Learning Techniques. IEEE Access, 11, 71517-71527.

[11] Pate, U. A., & Ibrahim, A. M. (2020). Fake News, Hate Speech, and Nigeria's Struggle for Democratic Consolidation: A Conceptual Review. Handbook of Research on Politics in the Computer Age, 89-112.

[12] Shu, J., Sliva, D., Wang, H., Tang, J., & Liu, B. (2017). Fake News Detection on Social Media: A Data Mining Perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22-36.

[13] Shu, B., Sliva, S., Wang, H., Tang, J., & Liu, B. (2017). Fake News Detection on Social Media: A Data Mining Perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22-36.

[14] Singhania, P., Fernandez, B., & Rao, S. (2020). 3HAN: A Deep Neural Network

for Fake News Detection. IEEE International Conference on Data Mining Workshops (ICDMW), 889-896.

[15] Subrahmanian, V. S., Azaria, A., Durst, S., Kagan, V., Galstyan, A., Lerman, K., ... & Menczer, F. (2016). The DARPA Twitter Bot Challenge. Computer, 49(6), 38-46.

[16] Tariq, A., Mehmood, A., Elhadef, M., & Khan, M. U. G. (2022). Adversarial Training for Fake News Classification. IEEE Access, 10, 82706-82715.

[17] Volkova, S., Shaikh, K., & Gordon, B. L. (2017). Combining Stylometric and Contextual Features for Fake News Detection. IEEE International Conference on Data Mining Workshops (ICDMW), 821-828.

[18] Wang, D., Zhang, W., Wu, W., & Guo, X. (2023). Soft-Label for Multi-Domain Fake News Detection. IEEE Access, 11, 98596-98606.

[19] Wang, W. Y. (2017). Liar, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection. Proceedings of the 55th ACL, 422-426.

[20] Zhou, Z., & Chen, L. (2023). Fake News Detection with Transformer Models: A Review. International Conference on Artificial Intelligence and Computer Engineering, 275-280.