International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Deep Learning and Forensic Analysis for Detecting Fake Content

Dr. K. Siva Rama Prasad[1], T. Renuka[2], Ch. Chandra Sekhar Reddy[3], D. Naveen[4], J. Sudheer[5]

[1]M. Tech, Ph.D., Professor, Department of Information Technology, Kallam Haranadhareddy Institute of Technology, Chowdavaram, Guntur, Andhra Pradesh, India
[2,3,45]Student, Department of Information Technology, Kallam Haranadhareddy Institute of Technology, Chowdavaram, Guntur, Andhra Pradesh, India

Abstract
Deep Learning and Forensic Analysis for Detecting Fake Content is an intelligent deep-learning and forensic analysis platform designed to verify whether digital content is AI-generated or human-created. In today's world, people frequently encounter AI-produced text, images, audio, and videos without realizing it. This leads to challenges such as misinformation, academic plagiarism, fake identities, manipulated media, and loss of digital trust. To address these issues, the system automatically analyzes uploaded content, detects hidden patterns, and determines its authenticity with high accuracy. Users can upload text and images through a simple web interface. The deep-learning models trained on real and AI-generated datasets process the content and identify subtle indicators such as linguistic irregularities, pixel-level artifacts, frame inconsistencies, and metadata anomalies that are typically produced by generative AI models like GPT, Stable Diffusion, GANs, and deepfake architectures. Based on these forensic clues, the system instantly classifies the input as AI-generated or Real, and provides an authenticity score along with explainable insights. By automating digital content verification, this system helps users identify manipulated or AI-generated media more accurately, reduces the spread of misinformation, and strengthens digital safety. The platform provides fast, data-driven authenticity detection, enabling individuals, educators, journalists, and organizations to trust the content they consume and share.

Keywords: Deep Learning, Digital Forensics, Fake Content Identification, CNN (Convolutional Neural Networks), Content Authenticity Detection.

1. Introduction

The rapid growth of digital communication and social media platforms has significantly increased the spread of information across the world. While this has improved accessibility, it has also led to the rise of fake content such as misleading news, manipulated images, and false information. These issues can negatively impact society by influencing opinions and spreading misinformation. To address this problem, the project Deep Learning and Forensic Analysis for Detecting Fake Content focuses on developing an intelligent system that can automatically classify content as real or fake. The proposed system is designed to analyze both textual and visual data. It uses Long Short-Term Memory (LSTM) n-

etworks for text analysis, which help in understanding context, writing patterns, and semantic meaning. For image analysis, Convolutional Neural Networks (CNN) are used to extract visual features and detect manipulations. By combining these deep learning techniques, the system provides an effective solution for identifying fake content across multiple formats.

The system architecture includes modules such as data collection, preprocessing, model training, prediction, and user interface. Data is collected and cleaned before being processed by the models, which are trained to detect fake and real content. The system also includes roles like Admin and Model Builder for managing datasets and improving model performance. Users can input content and receive predictions through a user-friendly interface, making the system efficient, scalable, and useful in reducing misinformation.

2. Literature Survey

This section describes the existing techniques used for detecting fake content using deep learning and forensic analysis methods along with their advantages and limitations.

Bayar and Stamm [1] proposed a CNN-based model that learns forensic features such as noise inconsistencies, resampling artifacts, and statistical irregularities for image forgery detection. The major advantage of this approach is that it eliminates the need for manual feature extraction and provides higher accuracy compared to traditional methods. However, the model requires large datasets and high computational resources for training.

Afchar et al. [2] introduced a forensically inspired CNN architecture called MesoNet for detecting facial deepfakes. This model focuses on identifying subtle artifacts in facial regions. It performs well even on compressed videos. However, it mainly focuses on facial images and may not generalize to other types of fake content.

Gupta et al. [3] proposed a hybrid model combining CNN and LSTM for text-based fake news detection. It uses forensic linguistic features such as writing style inconsistencies and semantic analysis. While it improves accuracy, it requires large datasets and has high computational complexity.

Zhou et al. [4] developed a multimodal deep learning model combining image and text analysis. This approach improves detection accuracy by identifying inconsistencies between text and images. However, it increases system complexity and computational cost.

3. Proposed System

The proposed system, Deep Learning and Forensic Analysis for Detecting Fake Content, is designed to automatically classify content as real or fake. It analyzes both textual and visual data, making it effective for detecting fake news and manipulated images. The system provides a reliable and scalable solution to reduce misinformation in digital platforms.

The system uses Long Short-Term Memory (LSTM) networks for text analysis and Convolutional Neural Networks (CNN) for image analysis. LSTM helps in understanding contextual relationships in text, while CNN extracts visual features from images. By combining these models, the system achieves higher accuracy compared to traditional methods.

The system consists of modules such as data collection, preprocessing, model training, prediction, and user interface. It also includes roles like Admin, Model Builder, and User for efficient management and operation. The final results are displayed through a user-friendly interface, ensuring ease of use and better understanding.

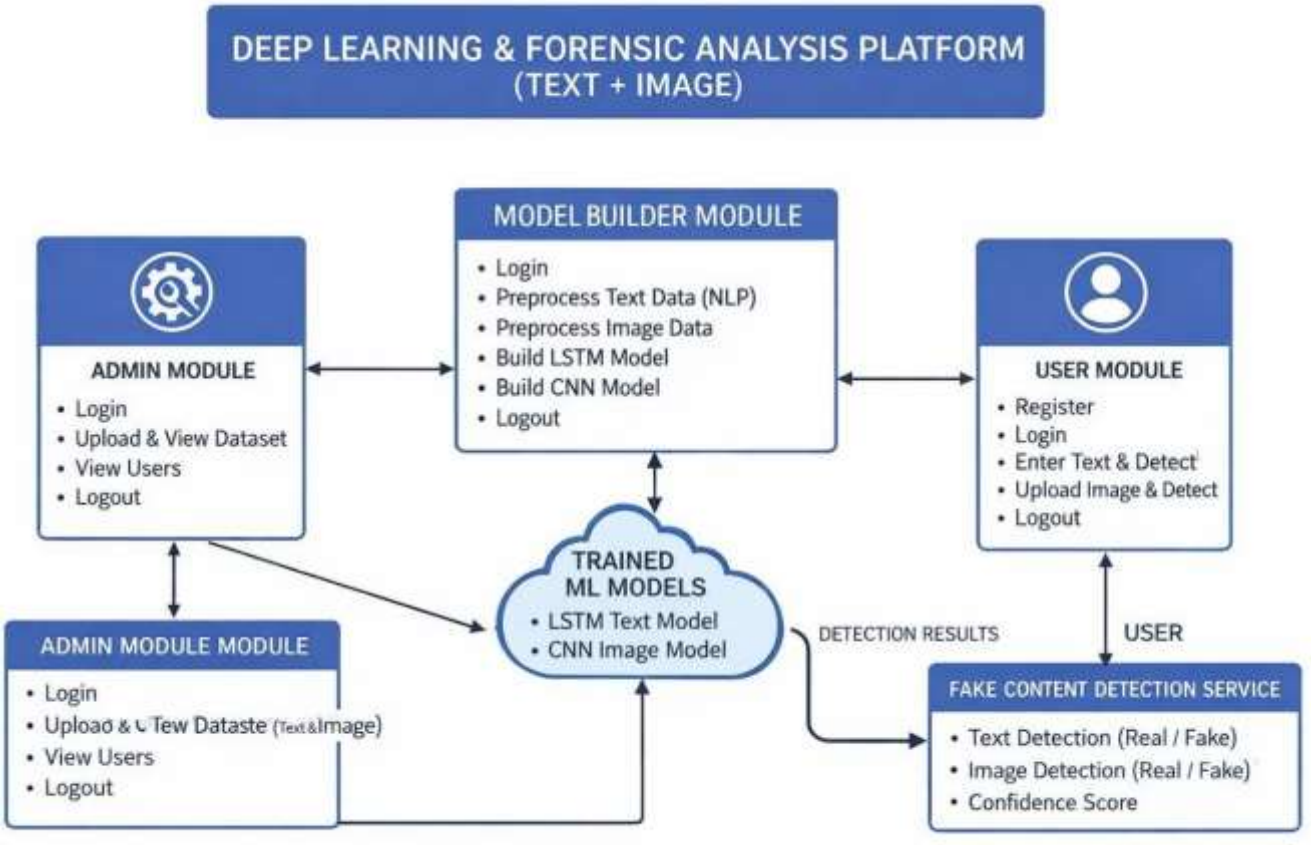

Figure 1: System Architecture

3.1 Convolutional Neural Network (CNN)

The proposed system utilizes the Convolutional Neural Network (CNN) model due to its strong capability in extracting spatial and hierarchical features from images, making it highly effective for detecting manipulated or AI-generated images. CNN automatically identifies patterns such as pixel inconsistencies, texture distortions, and deepfake artifacts that are difficult to detect manually. This model is well-suited for image classification tasks and provides high accuracy in distinguishing real and fake images.

3.2 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) model is used for detecting fake textual content because of its ability to understand sequential dependencies and contextual meaning in text data. Fake news or AI-generated text often contains contextual irregularities, exaggerated claims, or inconsistent sentence flow, which LSTM can effectively capture. This model is particularly suitable for natural language processing tasks and improves prediction accuracy by retaining important information over long sequences. LSTM also reduces the problem of vanishing gradients, making it reliable for text classification tasks.

3.3 Frequency-Domain Forensic Analysis

Frequency-Domain Forensic Analysis is employed to enhance the detection of AI-generated and manipulated images by analyzing hidden patterns in the frequency spectrum. Unlike spatial-domain methods that focus only on visible pixel information, frequency-domain techniques examine underlying periodic artifacts and generative fingerprints. AI-generated images often contain subtle inconsistencies

in high-frequency components that are not noticeable to the human eye but can be effectively captured in the frequency domain. This approach improves detection robustness, increases accuracy against highly realistic deepfakes, and strengthens the overall reliability of the fake content detection system.

4. Results and Discussions

The output of the Deep Learning and Forensic Analysis for Detecting Fake Content provides accurate and meaningful results based on the analysis of user input, which can be either text or image data. After the user submits the input, the system processes the data using deep learning models such as LSTM for text and CNN for images. The final output is presented in a clear and understandable format, indicating whether the given content is Real or Fake. This classification helps users quickly identify the authenticity of the content.

In addition to the classification result, the system also generates a confidence score, which represents the probability of the prediction. This score provides insight into how confident the model is about its decision, thereby increasing user trust in the system. The output is displayed through a user-friendly interface, ensuring that even non-technical users can easily understand the results. For image inputs, the system may also highlight suspicious patterns or inconsistencies, while for text inputs, it reflects linguistic analysis performed by the model.

The system also produces outputs related to model performance and system functionality. For the Admin and Model Builder roles, outputs include trained models, evaluation metrics such as accuracy, and processed datasets. These outputs help in improving the system's performance over time. Additionally, all prediction results can be stored for future reference, enabling analysis of past data and continuous improvement of the detection models. Overall, the outputs of the system are designed to be accurate, informative, and useful for both users and administrators.

5. Conclusion

The Fake Content Detection system successfully demonstrates how deep learning and digital forensic techniques can be combined to identify whether content is real or AI-generated. By supporting both text and image analysis within a single platform, the system provides a comprehensive solution for verifying digital content. It delivers accurate predictions along with confidence scores and explainable insights, such as highlighted text features and probability distributions, which enhance transparency and user trust. The user-friendly interface further ensures that the system can be easily used by individuals without technical expertise.

In conclusion, this project plays an important role in addressing the growing challenge of misinformation and AI-generated content in the digital world. It not only helps users verify authenticity but also promotes responsible content consumption. With further improvements such as expanding datasets, incorporating video detection, and optimizing model performance, the system can be developed into a more robust and scalable solution for real-world applications in media, education, and cybersecurity.

6. Appendix

The appendix provides supporting technical details related to the development of the Deep Learning and Forensic Analysis for Detecting Fake Content. The application was developed using Python with deep learning libraries for building the detection models. The development and implementation were carried

out using Visual Studio Code as the primary coding and debugging environment.
A MySQL database was configured to store user information, datasets, and prediction results. The user interface was designed using HTML, CSS, and JavaScript to provide a simple and interactive experience for users. The backend was developed using the Django framework to handle data processing, model integration, and communication between different system components.
The system accepts inputs in the form of text content or images uploaded by users. It processes the input using preprocessing techniques and applies LSTM for text analysis and CNN for image analysis to detect whether the content is real or fake. The output includes classification results along with confidence scores, which are displayed clearly to the user.
System testing was performed to verify major functionalities such as user authentication, dataset handling, preprocessing, model prediction, and result generation. The system requires a computer with basic hardware specifications and a stable internet connection for proper operation. Overall, the appendix provides essential technical insights into the implementation and functioning of the system.

7. Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this project. The research work carried out in this project is completely independent and has not been influenced by any external organization, sponsor, or individual. The results and findings presented are based solely on the implementation and analysis performed by the authors.

8. Acknowledgement

Sincere gratitude is expressed to guide Dr. K. Siva Rama Prasad for his valuable guidance, continuous support, and encouragement throughout the development of this project. His expertise and suggestions greatly contributed to the successful completion of this work.
Thanks are also extended to the faculty members of the Department of Information Technology for providing the necessary resources and support. Appreciation is expressed to friends and family for their encouragement and support during the project development.

9. References

1. Zellers R., Holtzman A., Rashkin H., Bisk Y., Farhadi A., Roesner F., Choi Y., “Defending Against Neural Fake News”, Advances in Neural Information Processing Systems (NeurIPS), 2019.
2. Nataraj L., Mohammed T., Chandrasekaran S., Bappy J., Roy-Chowdhury A., Manjunath B., “Detecting GAN Generated Fake Images Using Co-occurrence Matrices”, 2019.
3. Rossler A., Cozzolino D., Verdoliva L., Riess C., Thies J., Nießner M., “FaceForensics++: Learning to Detect Manipulated Facial Images”, 2019.