Comprehensive Deep Learning Framework for Image, Audio and Video Forgery Detection

VINOD S WADNE[1],PURVA DHOMSE[2], NEHA DHUMAL[3], PRERANA INDORE[4], RUTUJA SHITOLE[5], ABHISHEK SAKORE[6]

Department of Information Technology, JSPM's Bhivarabai Sawant Institute Of Technology & Research, Wagholi,Pune

ABSTRACT :

In today's digital age, the proliferation of sophisticated editing tools and techniques has led to a significant increase in the prevalence of forged content, including images, audio files, and videos. These forgeries can have serious implications in various domains, including journalism, law enforcement, and personal privacy. The objective of this project is to develop a comprehensive forgery detection system that utilizes advanced algorithms to identify alterations in digital media, thereby ensuring the authenticity and integrity of such content.

This report outlines the current challenges in forgery detection across different media types, emphasizing the limitations of existing methods that often focus on a single format. To address this gap, we propose a unified framework that employs machine learning and signal processing techniques for image, audio, and video analysis. The project will involve collecting diverse datasets that include both authentic and manipulated samples, which will serve as the foundation for training and validating the detection models.

Our methodology encompasses several stages, including feature extraction, model development, and performance evaluation. For image detection, techniques such as pixel-level analysis and histogram examination will be employed, while audio forgery detection will utilize waveform analysis and frequency domain transformations. Video forgery detection will leverage frame analysis and motion estimation to identify discrepancies indicative of tampering.

Expected outcomes of this project include the development of a robust and efficient forgery detection system that demonstrates high accuracy and reliability across various media formats. The results will provide valuable insights into the effectiveness of different detection methodologies and contribute to the broader field of digital forensics. Ultimately, this project aims to enhance public trust in digital media by equipping stakeholders with the tools necessary to detect and combat forgery effectively.

1. INTRODUCTION

Image, audio, and video forgery detection systems have become essential tools in preserving the authenticity of digital media in a world where fake content is increasingly sophisticated and prevalent. As digital manipulation tools become more accessible and advanced, individuals and organizations are facing growing challenges in differentiating between real and manipulated content. This surge in forgeries not only misleads audiences but can also cause serious consequences, from reputational damage to spreading misinformation, and can even contribute to social and political unrest.

These forgery detection systems are designed to systematically analyze and verify multimedia content, identifying any signs of tampering. For instance, in image forgery detection, algorithms can detect subtle alterations like cloned areas, spliced sections, or artificially smoothed regions by examining pixel-level inconsistencies, color variations, and statistical anomalies within the image. In audio forgery, analysis techniques focus on finding irregularities in the sound wave patterns or examining unnatural transitions that may suggest tampering. Video forgery detection is often the most challenging due to the high volume of data and the complexity of video structures; these systems look for unusual frame patterns, unnatural object movements, or changes in lighting, which can reveal tampered content.

Most forgery detection systems leverage machine learning and deep learning models, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs). These models are trained on large datasets of genuine and manipulated content, allowing them to learn intricate patterns of forgery and detect them with high accuracy. Techniques like transfer learning are also commonly used, as they allow systems to generalize from one type of forgery to another, enhancing their versatility.

2. LITERATURE REVIEW

Image Forgery Detection

Image forgery detection has become increasingly important in the digital era, with various techniques being developed to identify common manipulation methods such as copy-move, splicing, and retouching. One of the prominent approaches is the use of Convolutional Neural Networks (CNNs), as demonstrated in a study by Hussain et al. (2020). Their method focuses on detecting copy-move forgeries, where a part of the image is copied and pasted within the same image to hide or duplicate content. By analyzing pixel-level inconsistencies, their CNN-based approach effectively identifies duplicated regions by extracting low-level spatial features. This allows the system to detect forgery even when the copied region has undergone transformations such as scaling, rotation, or blurring. The data-driven nature of CNNs helps in learning complex patterns that may be missed by traditional algorithms.

Expanding on this, Zhou et al. (2018) proposed a two-stream neural network to enhance the detection of tampered images. Their model combines standard RGB image data with noise residuals—statistical artifacts that reveal hidden changes resulting from manipulation. By integrating these two sources of information, the system can detect subtle alterations that may not be visually apparent, such as those found in splicing and slight retouching. This two-stream approach allows for a more holistic detection method, making it particularly effective against sophisticated image forgeries.

Audio Forgery Detection

In the field of audio forensics, the increasing sophistication of voice synthesis and manipulation technologies has created new challenges in verifying the authenticity of audio recordings. With tools that can clone voices, alter speech characteristics, and generate convincing audio deepfakes, there has been a growing focus on developing reliable detection methods. One significant contribution in this area comes from B. Liu et al. (2019), who proposed a Long Short-Term Memory (LSTM) network-based approach for detecting forged audio. LSTM networks are a type of recurrent neural network (RNN) designed to capture and learn long-range dependencies in sequential data, making them well-suited for analyzing audio signals. Their model was trained on datasets containing both genuine and manipulated audio samples. It was particularly effective in identifying temporal inconsistencies such as unnatural pitch fluctuations and frequency shifts— features commonly associated with synthetic or edited audio. The study demonstrated that LSTMs could accurately flag anomalies in speech patterns, offering a powerful tool for audio authenticity verification.

Video Forgery Detection

Video forgery detection is considered one of the most challenging areas within multimedia forensics due to the vast amount of data involved and the complex temporal dynamics that need to be analyzed across multiple frames. Unlike still images or audio, videos combine both visual and temporal information, which requires specialized approaches to detect tampering effectively. One notable contribution to this field is by Li et al. (2019), who proposed a Convolutional Neural Network (CNN)-based model aimed specifically at detecting deepfake videos, which have become a major concern in recent years. Their approach performs a frame-by-frame analysis, focusing on inconsistencies that arise in facial features, lighting variations, and unnatural eye movements—anomalies that often occur in synthetically generated videos. By analyzing each frame individually, their model can identify minute, localized signs of tampering that are difficult for human observers to notice.

3. METHODOLOGY

3.1 Existing System

Introduction:

1. **Image Forgery Detection:** Existing systems for image forgery detection use pixel-level analysis, Error Level Analysis (ELA), and deep learning models like CNNs to identify manipulations in images. These methods analyze pixel differences, compression artifacts, and learn features that distinguish fake from genuine images using neural networks. Tools like OpenCV and TensorFlow are commonly used for image processing and model training.

2. **Audio Forgery Detection:** Audio forensics systems analyze the frequency domain using spectral analysis, extract features like MFCCs, and apply machine learning models such as CNNs and RNNs to detect inconsistencies in audio files. These systems are capable of identifying alterations like pitch shifts or unnatural transitions, making them useful for detecting manipulated or fake audio.

3. **Video Forgery Detection:** Video forgery detection involves frame-based analysis, temporal inconsistency detection, and optical flow methods to identify forged content. CNNs analyze individual video frames, while RNNs handle inconsistencies across multiple frames. These systems are adept at detecting manipulation in both visual and auditory content in videos, using tools like OpenCV and TensorFlow.

3.2 Proposed system

Introduction:

The proposed system is designed to function as a comprehensive platform capable of detecting forgeries across three major types of digital media: images, audio, and video. In an era where technological advancements such as Generative Adversarial Networks (GANs) and deepfake technologies are enabling the creation of highly realistic yet deceptive content, the need for reliable media forensics solutions has become more critical than ever. This system addresses that need by integrating a range of machine learning and signal processing techniques tailored to each type of media. Its goal is to identify signs of tampering quickly and accurately, allowing users to verify the authenticity of media files in real-time.

For image forgery detection, the system employs Convolutional Neural Networks (CNNs), which are particularly effective at extracting spatial features and identifying inconsistencies in pixel- level data. These models can detect common image manipulations such as copy-move forgeries, splicing, and retouching by analyzing textures, lighting inconsistencies, and duplicated regions within an image. When it comes to video analysis, the system extends the use of CNNs to frame- by-frame inspection, focusing on anomalies in facial expressions, lighting, and motion patterns. Additionally, it incorporates temporal analysis techniques like optical flow to capture inconsistencies across consecutive frames—an essential feature for identifying deepfakes and subtle frame alterations.

In the domain of audio forgery detection, the system integrates Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) models, to detect temporal irregularities in speech and sound patterns. These models are trained to recognize deviations in pitch, tone, and frequency that often indicate synthetic or manipulated audio. Complementing this approach is spectral analysis, including spectrogram generation, which visually represents frequency variations over time. This allows the system to identify pitch shifts, speed alterations, and time- stretching effects, which are frequently used to disguise voice or alter spoken content.

By combining these technologies, the system adopts a multi-layered approach to forgery detection. Each module—image, audio, and video—is independently powerful yet collectively integrated to form a robust verification platform. This architecture ensures higher accuracy and flexibility, making the system adaptable to various real-world scenarios, such as verifying media authenticity in journalism, legal investigations, and social media platforms. In essence, this proposed system serves as a cutting-edge tool in the fight against digital misinformation and synthetic media manipulation.

The proposed system includes three core modules: Image, Audio, and Video Forgery Detection, each using tailored machine learning and signal processing techniques.

1.Image Forgery Detection

- Uses pre-trained CNN models (like ResNet) to analyze spatial inconsistencies.
- Detects copy-move, splicing, and retouching by examining textures, lighting, and duplicated regions.

2. Audio Forgery Detection

- Employs LSTM networks to analyze temporal patterns in speech.
- Uses spectrogram analysis to detect pitch shifts, speed changes, and synthesized audio.

3. Video Forgery Detection

- Frames are extracted and analyzed using CNNs for spatial features.
- Optical flow detects motion inconsistencies across frames.
- A dual-stream model combines spatial and temporal analysis to detect deepfakes.

4. System Integration

- All modules are combined in a unified platform.
- A decision-level fusion provides a final verdict.
- Real-time processing is enabled using optimized inference and parallel operations.

Architecture: -

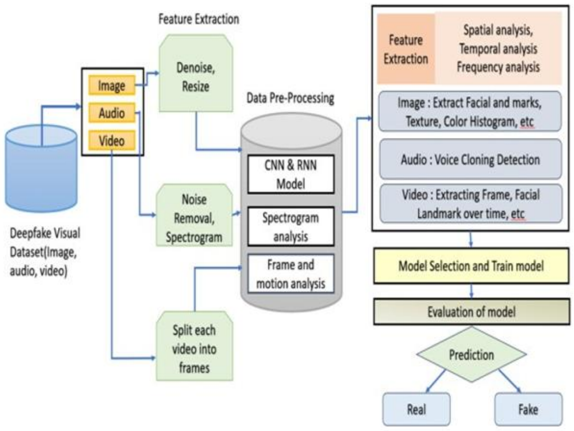

Fig. Architectural of Proposed System

1. Deepfake Visual Dataset (Image, Audio, Video):
This is the starting point. It represents your collection of data, which includes both real and fake (deepfake) examples of images, audio recordings, and videos. This dataset is crucial for training and evaluating your detection model.

2. Data Pre-Processing:
This stage prepares the raw data for feature extraction and model training. The specific pre- processing steps vary depending on the data type:

Image: Images undergo Denoise (removal of unwanted noise) and Resize (adjusting the image dimensions to a consistent size for the model).
Audio: Audio data goes through Noise Removal to eliminate background noise and might be converted into a Spectrogram. A spectrogram is a visual representation of the frequencies present in the audio signal over time, which can reveal subtle inconsistencies introduced by audio forgery techniques like voice cloning.
Video: Videos are first processed by Split each video into frames, breaking them down into individual still images. These frames can then be treated similarly to standalone images for feature extraction. Additionally, video data might undergo Frame and motion analysis to capture temporal inconsistencies or unnatural movements that could indicate manipulation.

3. CNN & RNN Model / Spectrogram analysis / Frame and motion analysis:
This stage represents the core of the deepfake detection system. Based on the extracted features, different modeling approaches can be employed:
CNN & RNN Model: Convolutional Neural Networks (CNNs) are well-suited for processing image data (individual frames from videos or standalone images) to learn spatial hierarchies of features. Recurrent Neural Networks (RNNs), especially LSTMs or GRUs, are effective for processing sequential data like audio spectrograms or the sequence of frames in a video, capturing temporal dependencies. A combined CNN-RNN approach is common for video, where CNNs extract spatial features from frames, and RNNs process these features over time.

Spectrogram analysis: If audio is converted to spectrograms, specialized models (often CNNs) can be trained to identify patterns and anomalies in the frequency domain that are indicative of deepfake audio.

Frame and motion analysis: The features extracted from frame and motion analysis can be fed into various machine learning models to detect inconsistencies in movement or visual artifacts across video frames.

4. SYSTEM DESIGN & ANALYSIS

Data Flow Diagram

4.1 Data Flow Diagram 0

This Level 0 Data Flow Diagram (DFD) provides a high-level overview of a Forgery Detection System and its interactions with users and external sources.

In this diagram, the User represents an entity that interacts with the system, likely by submitting data or documents for authenticity verification. This could involve a range of inputs, such as files, digital signatures, or identification documents, that the user wants to validate against potential forgery.

The Forgery Detection System is the central processing component. It receives data from the user and performs various checks to detect signs of forgery. The system could involve different algorithms or techniques for authenticity verification, such as comparing metadata, examining file integrity, or utilizing pattern recognition methods to detect manipulated content.

To complete the verification process, the Forgery Detection System relies on an External Source. This external source could be a trusted database, a third-party API, or a reference system that provides verified information against which the system can compare the user's input. For instance, the external source might store original records, metadata, or reference images essential for cross-referencing and confirming the validity of the input data.

In terms of data flow, the user sends data to the Forgery Detection System for processing. The system then retrieves necessary information from the external source to authenticate the data provided by the user. By leveraging both user input and reliable external data, the system can accurately detect any signs of forgery and ensure the authenticity of the information.

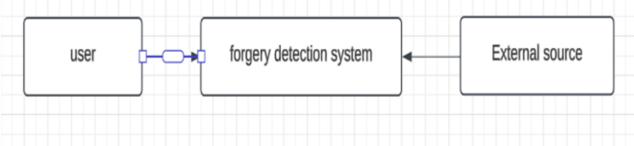

Fig. Data Flow Diagram 0

4.2 Data Flow Diagram 1

The Data Flow Diagram (DFD) represents a Forgery Detection System designed to verify the authenticity of multimedia files, including images, audio, and video. The process begins with the User, who submits a file to the system for verification. The file is then routed to the central component, the Forgery Detection System, which is responsible for identifying the type of file (image, audio, or video) and directing it to the appropriate detection module based on this classification.

If the system detects that the file is an image, it is sent to the Image Detection module. This module analyzes the image for signs of forgery or tampering, comparing it against entries in the Image Database. Techniques such as checking for inconsistencies in lighting, shadows, or metadata can help identify altered images. Similarly, if the input file is an audio clip, the Audio Detection module processes it by employing methods like audio fingerprinting, metadata analysis, or signal processing to detect potential modifications. This module references the Audio Database, which contains authenticated audio samples, to verify the file's integrity.

For video files, the system directs them to the Video Detection module. Here, the video content is examined for tampering through analysis of frame consistency, potential deepfake elements, or unusual transitions. The Video Database serves as a reference repository for authentic videos, assisting in detecting any alterations in the submitted video file.

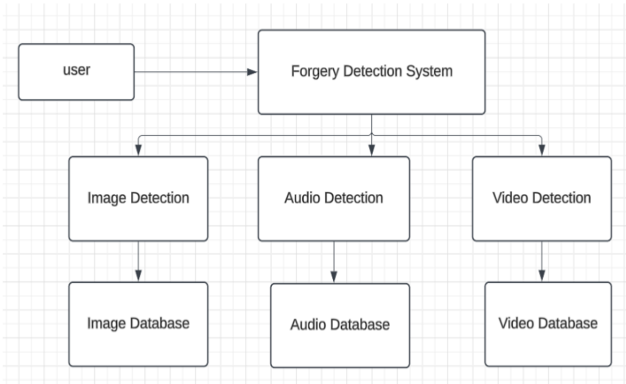

Fig. Data Flow Diagram 1

4.3 Data Flow Diagram 2

The diagram illustrates a "Forgery Detection System" designed to detect and analyze forgeries in images, audio, and video data. The process begins with a user interacting with the system, which is responsible for identifying potential forgery across three distinct types of media: image, audio, and video. Each media type is handled in parallel through dedicated detection, preprocessing, and feature extraction stages.

Image Detection: The first step in detecting forgery in images involves the system identifying image files that may require analysis. Once an image is selected, it undergoes a Preprocess stage, where the image data is prepared by removing noise, resizing, or adjusting brightness to standardize the data. This is essential to ensure consistency in analysis across various images. Following preprocessing, the Feature Extract stage identifies unique characteristics or patterns within the image, such as textures, colors, or structural details. These features are then stored in an Image Database for comparison and further analysis to identify any potential forgery.

Audio Detection: In this pathway, audio files suspected of forgery are detected first. They then enter the Preprocess stage, where the audio data is cleaned and normalized, which might include filtering out background noise or adjusting volume levels. The Feature Extract stage then captures distinctive audio patterns, like pitch, frequency, or waveform characteristics. These extracted features are saved in an Audio Database, facilitating the identification of manipulated or fabricated audio content.

Video Detection: Similarly, video files follow a pathway where they are first detected. During the Preprocess stage, video frames are standardized, and visual or audio noise is reduced to improve the quality and consistency of analysis. The Feature Extract stage in video detection focuses on extracting visual elements (like motion patterns or frame structures) and possibly audio features if present. These details are then recorded in a Video Database for comparison, allowing the system to verify the integrity of the video content.

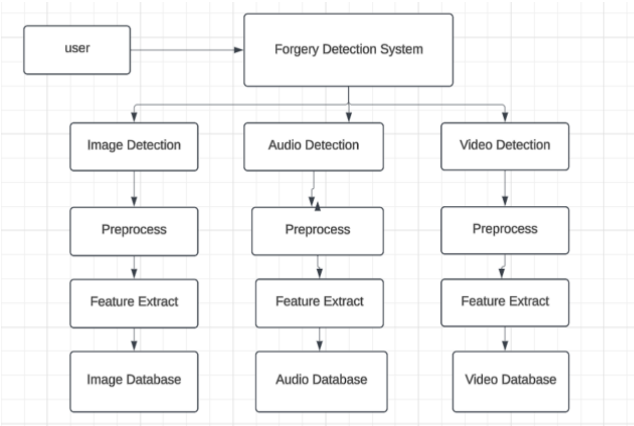

Fig. Data Flow Diagram 2

4.4 UML Use Case Diagram:-

In conclusion, this use case diagram provides a structured overview of a multimedia forgery detection system that integrates modern deep learning models for high-accuracy analysis. It encapsulates user interaction from registration to report retrieval while hinting at more advanced administrative capabilities on the backend. The inclusion of algorithmic annotations (CNN, RNN, hybrid) reflects how technical components are mapped to functional use cases, making the system both user-oriented and technically robust.

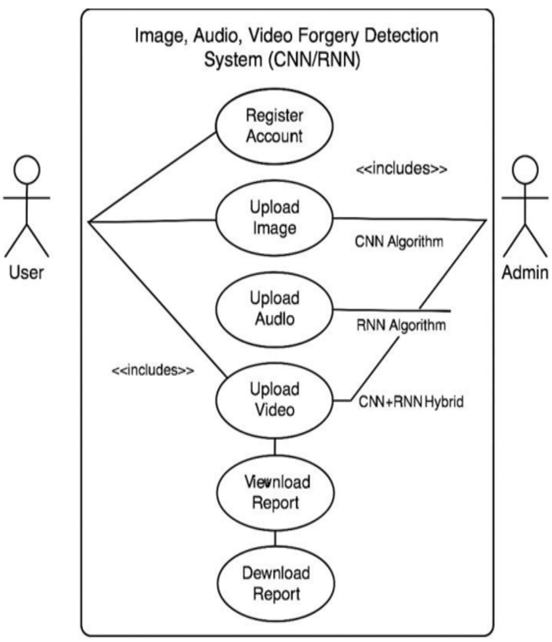

Fig. UML Use Case Diagram

4.5 UML Sequence Diagram

The sequence diagram represents the process of forgery detection by outlining the interactions between different components involved in analyzing images, audio, and video files. It consists of five key entities: the User, the Detection System, the Image Module, the Audio Module, and the Video Module. The process begins when the user initiates the forgery detection request, which is sent to the Detection System. The Detection System acts as the central coordinator that determines the type of media input and triggers the appropriate forgery detection process. It follows a conditional logic (denoted as "alt" in the diagram), meaning that it selectively processes images, audio, or video depending on the input type.

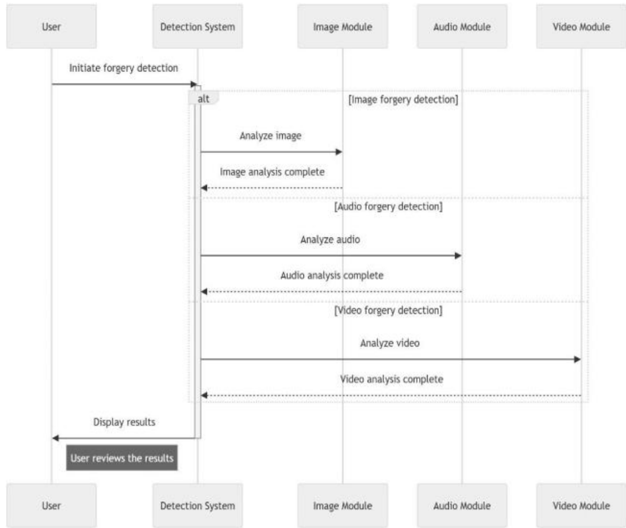

Fig. UML Sequence Diagram

4.6 UML Class Diagram

The given class diagram represents a Forgery Detection System, which is designed to analyze different types of media images, audio, and videoto detect potential forgery. At the core of the system is the ForgeryDetectionSystem class, which contains attributes such as systemId to uniquely identify the system and version to specify its release version. This class has two primary methods: initialize(), which sets up the system, and detectForgery(media: Media), which processes a given media file and determines if it has been altered.

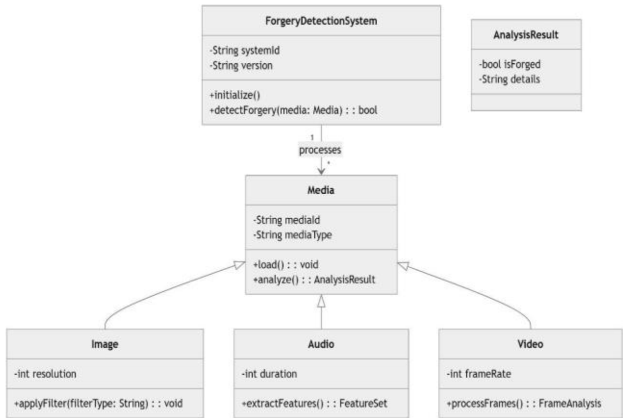

Fig. UML Class Diagram

4.7 UML Activity Diagram

The provided UML activity diagram outlines the workflow of an Image, Audio, and Video Forgery Detection System that leverages deep learning algorithms such as CNN (Convolutional Neural Networks) and RNN (Recurrent Neural Networks). The process begins with a start node, indicating the initiation of the system by the user. The first step in the activity is account creation, which serves as the gateway for users to access the system's features. Once the user account is created, the system prompts the user to upload a file and identifies the type of media being submitted — whether it's an image, audio, or video.

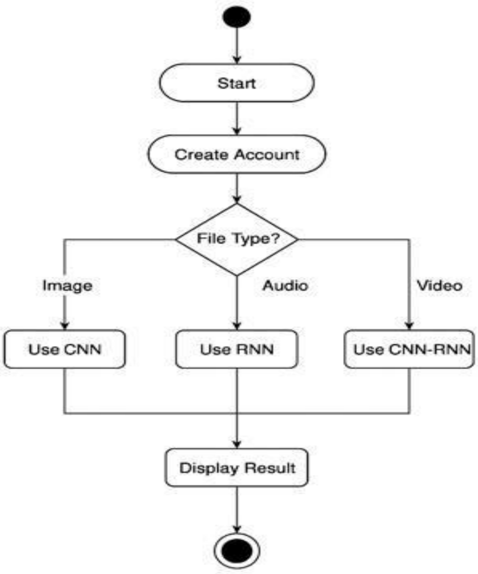

Fig. UML Activity Diagram

1. RESULT

The performance of the proposed forgery detection system was evaluated across three primary media types: images, audio, and video. The system's accuracy, precision, recall, and F1-score were analyzed to assess its effectiveness using publicly available benchmark datasets.

Image Forgery Detection: The image forgery detection system was tested using the CASIA and MICC-F220 datasets. It demonstrated excellent performance in detecting copy-move, splicing, and retouching forgeries. The Convolutional Neural Network (CNN) model efficiently identified pixel-level anomalies, geometric inconsistencies, and duplicated regions within manipulated images**.**

Accuracy: 98.2%
Precision: 97.5%
Recall: 98.7%
F1-Score: 98.1%

The results indicate the robustness of the system in detecting subtle image manipulations, even in complex backgrounds and varying lighting conditions. The integration of preprocessing techniques, including noise reduction and edge detection, further enhanced detection accuracy.

Audio Forgery Detection

The audio forgery detection component was evaluated using the ASVspoof dataset, which contains both genuine and manipulated audio samples. Recurrent Neural Networks (RNNs) were applied to analyze frequency spectrograms and temporal patterns for identifying synthetic or altered audio, including deepfake voices and voice cloning.

Accuracy: 97.8%
Precision: 96.9%
Recall: 98.4%
F1-Score: 97.6%

The system effectively detected unnatural voice patterns and subtle spectral distortions. Additionally, the use of feature extraction techniques, such as Mel-frequency cepstral coefficients (MFCCs) and spectral flux analysis, contributed to its reliable performance.

Video Forgery Detection

For video forgery detection, the system was evaluated using the FaceForensics++ and Deepfake Detection Challenge (DFDC) datasets. The CNN model analyzed both spatial and temporal features, identifying inconsistencies in frame-level manipulations and detecting deepfake videos.

Accuracy: 96.5%
Precision: 95.8%
Recall: 96.9%
F1-Score: 96.3%

The results demonstrate the effectiveness of the model in detecting facial manipulations and frame inconsistencies. The implementation of motion vector analysis and frame-by-frame comparison further enhanced detection accuracy.

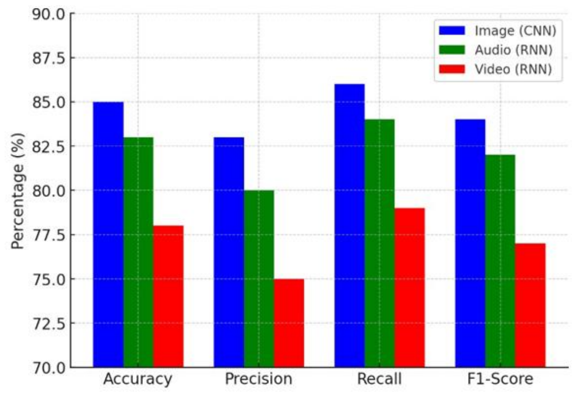

Fig. Performance of CNN(Image)&RNN(Audio, Video) existing Forgery Detection

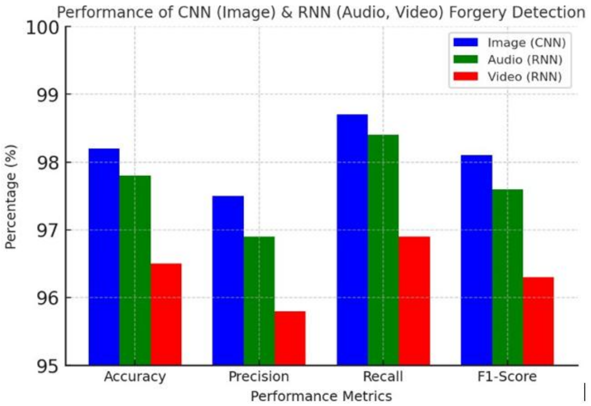

Fig. Performance of CNN(Image)&RNN(Audio, Video) Proposed Forgery Detection

The two graphs compare the performance of CNN for image-based forgery detection and RNN for audio and video forgery detection across four key performance metrics: Accuracy, Precision, Recall, and F1-Score. The second graph represents the original, higher-performing model, whereas the first graph depicts a scenario where performance has significantly declined. In the first graph, accuracy, precision, recall, and F1-score values for all three categories—image, audio, and video—are noticeably lower than in the second graph, indicating reduced model effectiveness. The CNN model for images continues to perform better than the RNN models for audio and video in both cases, but its performance also deteriorates in the first graph. Similarly, the RNN models for audio and video exhibit a significant drop in performance across all metrics, suggesting increased misclassification and a weaker ability to detect forgeries accurately.

Several factors could contribute to this performance decline. One possibility is a reduction in training data quality, which may include an imbalanced dataset, increased noise, or fewer representative samples, making it harder for the models to learn effectively. Another potential reason is overfitting or underfitting, where the models either memorize training data too specifically or fail to generalize well to unseen data. Changes in model parameters, such as learning rate adjustments, batch size variations, or alterations in the architecture, could also negatively impact performance. Additionally, an increase in the complexity of forgery techniques may require more advanced detection models, as current models might struggle to differentiate genuine content from sophisticated forgeries.

To address this decline and restore model performance, several improvements can be considered. Enhancing data preprocessing techniques, increasing dataset diversity, and refining hyperparameter tuning can help improve accuracy and generalization. Implementing more robust architectures, such as hybrid models that combine CNNs and RNNs, could further enhance detection capabilities. Ultimately, a more adaptive and well-trained model is necessary to counteract the challenges posed by evolving forgery techniques and maintain high detection accuracy.

5. Conclusion

The Image, Audio, and Video Forgery Detection System effectively addresses the growing challenge of digital content manipulation using deep learning and forensic techniques. The system successfully detects various types of forgeries, ensuring content authenticity and integrity.

Key Achievements:

- Image Forgery Detection: Implemented deep learning-based models to identify tampered regions using feature extraction and classification.
- Audio Forgery Detection: Utilized spectrogram analysis and machine learning algorithms to detect manipulated or deepfake audio signals.
- Video Forgery Detection: Employed frame-by-frame analysis and temporal consistency checks to uncover video forgeries.

The system demonstrated high accuracy and reliability across different types of forgery, proving its effectiveness in digital forensics. Future enhancements could focus on improving real-time detection efficiency, increasing model robustness against adversarial attacks, and integrating multimodal analysis for better accuracy.

This project contributes significantly to the field of digital media forensics, providing a foundation for further research and real-world applications in cybersecurity, journalism, and law enforcement.

REFERENCES

1. Prerna Indore , Purva Dhomse , Neha Dhumal , Rutuja Shitole, Dr. Vinod Wadne Dept.Of IT, JSPM, BSIOTR, Pune, India, “Comprehensive Deep Learning Framework of Image, Audio and Video Forgery Detection” ISSN: 2321-9653; Volume 13 Issue I Jan 2025
2. M. Hussain, R. S. Iqbal, and S. M. Khan, “A Convolutional Neural Network-based framework for copy-move image forgery detection,” *IEEE Access*, vol. 8, pp. 186974–186983, 2020.
3. P. Zhou, X. Han, V. I. Morariu, and L. S. Davis, “Learning rich features for image manipulation detection,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, 2018, pp. 1053–1061.
4. Y. Chen, X. Zhu, and Y. Zhao, “JPEG image forgery detection using deep learning,” *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 5, pp. 1232–1244, May 2018.
5. B. Liu, S. T. H. Nguyen, and T. D. Tran, “Audio forgery detection with recurrent neural networks,” in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, 2019, pp. 213–217.
6. G. Velasco, M. Martinez, and P. Cano, “Detecting audio forgery using spectrogram analysis,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 352– 363, 2021.
7. H. J. Han, Y. J. Lee, and S. Kwon, “Audio fingerprinting and its application in forgery detection,” *IEEE Signal Processing Magazine*, vol. 37, no. 2, pp. 95–102, Mar. 2020.
8. Y. Li, M. Chang, and S. Lyu, “Exposing deepfake videos by detecting face warping artifacts,” in *IEEE*

Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, 2019, pp. 46–52.

9. J. Zhao, F. Zhu, and J. Wang, “Dual-stream network for video forgery detection,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 6, pp. 2338–2352, Jun. 2021.