International Journal of Advanced Multidisciplinary Research and Educational Development
Volume 2, Issue 2 | March – April 2026 | www.ijamred.com
ISSN: 3107-6513

Deepfake Detection Using Metadata Inconsistencies

Amalayana.A, III B.Sc Digital and Cyber Forensic Science, Department of Computer Science, Rathinam College Of Arts and Science, amalayana23@gmail.com
Dr. M. Ramaraj, Assistant Proffesor, Department of Computer Science, Rathinam College of Arts and Science, Coimbatore, Tamilnadu, India. ramarajm.cs@rathinam.in

Abstract*-Deepfake technology has made it increasingly difficult to distinguish between real and manipulated digital content, raising serious concerns about misinformation and security. This project presents a method for detecting deepfakes by analyzing metadata inconsistencies in images and videos. By examining details such as timestamps, device information, and file history, the system identifies anomalies that indicate tampering. Compared to traditional visual-based methods, this approach is more efficient and less computationally intensive. It provides a reliable and practical solution for enhancing digital content authenticity and supporting deepfake detection systems.*

Keywords *– Deepfake Detection, Metadata Analysis, Metadata Inconsistencies, Digital Forensics, Image and Video Forensics, Data Integrity, Cybersecurity, Fake Media Detection, Machine Learning, Multimedia Authentication, File Metadata Extraction, Anomaly Detection.*

1. Introduction

Deepfake technology has experienced rapid growth in recent years due to significant advancements in artificial intelligence, particularly in deep learning techniques such as generative adversarial networks (GANs) and autoencoders. These technologies enable the creation of highly realistic synthetic images, videos, and audio that closely mimic real-world content. As a result, deepfakes are increasingly being used in various domains, including entertainment, media production, and virtual reality. However, despite their beneficial applications, deepfakes also pose serious ethical, social, and security concerns. Malicious use of deepfake technology can lead to misinformation campaigns, identity theft, political manipulation, financial fraud, and reputational damage. The ability of deepfakes to appear authentic makes them particularly dangerous, as they can easily deceive both humans and automated systems.

Traditionally, deepfake detection methods have relied heavily on visual and behavioral analysis, such as identifying facial inconsistencies, unnatural blinking patterns, lighting mismatches, or irregular movements. While these approaches have shown effectiveness in detecting early-stage deepfakes, they are becoming less reliable as generation techniques continue to improve. Modern deepfake models are capable of producing highly refined outputs with minimal visible artifacts, making it increasingly difficult to distinguish between real and fake content using visual cues alone. This limitation highlights the need for alternative detection strategies that do not depend solely on surface-level features.

In this context, metadata analysis has emerged as a powerful and promising approach for deepfake detection. Metadata refers to the hidden or embedded information within digital media files that describes various attributes of the file, such as creation date and time, device specifications, software used for editing, compression details, file structure, and sometimes even geolocation data. Unlike visual features, metadata is often overlooked but plays a crucial role in understanding the origin and history of a media file. When a deepfake is generated or manipulated, the editing process often introduces inconsistencies or anomalies in the metadata. For example, discrepancies between creation and modification timestamps, missing or altered EXIF data, unusual encoding patterns, or mismatches between device information and file properties can serve as strong indicators of tampering.

By systematically analyzing these metadata inconsistencies, it becomes possible to detect deepfake content with a higher degree of reliability. This approach provides an additional layer of verification that complements existing visual-based detection methods. Furthermore, metadata analysis is computationally efficient and does not require extensive training data, making it suitable for real-time applications and large-scale deployment. It also offers robustness against advanced deepfakes that successfully bypass traditional detection techniques.

Despite its advantages, metadata-based detection also faces certain challenges. In some cases, metadata may be intentionally removed, modified, or forged to evade detection. Additionally, variations in file formats and compression techniques can introduce complexities in analysis. Therefore, it is essential to develop intelligent and adaptive systems capable of handling such scenarios effectively.

This research aims to address these challenges by proposing a robust deepfake detection system based on metadata inconsistencies. The system is designed to extract, analyze, and evaluate metadata features to identify anomalies that indicate manipulation. By integrating multiple metadata analysis techniques, the proposed approach enhances detection accuracy, improves reliability, and provides a scalable solution for modern cybersecurity and digital

ISSN: 3107-6513

forensic applications. Ultimately, this work contributes to strengthening trust in digital media and combating the growing threat posed by deepfake technology.

2.Related Works

Over the past few years, deepfake detection has become an important area of research, mainly because of how quickly AI-generated media is improving. In the early stages, most detection methods focused on identifying visible flaws in fake content. Researchers looked for things like unnatural facial expressions, strange eye blinking patterns, mismatched lip movements, or lighting inconsistencies. These techniques worked reasonably well at first, but as deepfake technology advanced, such obvious visual errors became less common. Modern deepfakes are now so refined that it is often difficult to detect them just by looking at the content.

To improve detection, researchers started using machine learning techniques. Algorithms like Support Vector Machines (SVM), Random Forest, and other classifiers were trained using features extracted from images and videos. These features included texture details, pixel variations, and frequency patterns. While this approach improved detection accuracy compared to basic visual inspection, it still depended heavily on how well the features were chosen. If important features were missed, the model's performance would drop.

With further advancements, deep learning methods became more popular. Models such as Convolutional Neural Networks (CNNs) are capable of automatically learning patterns directly from raw data, without the need for manual feature selection. These models have shown high accuracy in detecting deepfakes, especially when trained on large datasets. However, they come with their own challenges. They require a lot of computational power, large amounts of labeled data, and longer processing time, which makes them less suitable for real-time applications.

Recently, researchers have started exploring metadata-based detection as an alternative approach. Instead of focusing only on what we see in an image or video, this method looks at the hidden information stored within the file. Metadata includes details like when the file was created, what device was used, how it was processed, and how it was encoded. When a media file is manipulated or generated artificially, these details often become inconsistent. For example, the timestamp might not match, camera information may be missing, or the encoding pattern may look unusual. These small inconsistencies can provide strong clues that the content has been altered.

To achieve better results, some studies have combined both visual and metadata-based techniques. These hybrid approaches use the strengths of both methods—visual analysis helps detect visible anomalies, while metadata analysis verifies the file's authenticity at a deeper level. This combination often leads to more accurate and reliable detection.

However, there are still challenges that need to be addressed. In some cases, metadata can be removed or intentionally modified to hide traces of manipulation. Also, different file formats and editing tools can make metadata analysis more complex. As deepfake technology continues to evolve, detection methods must also keep improving to stay effective.

Overall, while significant progress has been made in detecting deepfake media, there is still no perfect solution. Combining different techniques and developing more adaptive systems will be essential to effectively tackle the growing threat of deepfakes in today's digital world.

3.System Design

The system is designed to detect deepfake media by focusing on metadata inconsistencies instead of relying only on visual details. It works in a simple step-by-step manner, starting from taking an image or video as input and then analyzing it to check for any hidden irregularities. The system extracts important metadata such as timestamps, device information, and file properties, and then examines this data for unusual patterns or mismatches. Based on these findings, it determines whether the media is real or fake. This approach makes the system both efficient and reliable, especially when dealing with highly realistic deepfakes that are difficult to detect visually.

Key Steps in the System

•Input Collection – The user uploads an image or video file into the system.
•Preprocessing – The file is checked and prepared for analysis.
•Metadata Extraction – Important hidden data (EXIF, timestamps, device info) is collected.
•Metadata Analysis – The system checks for inconsistencies or unusual patterns.
•Detection Techniques – Various methods are applied to identify manipulation.
•Classification – The media is classified as Real or Fake.
•Result Display – The final output is shown to the user.

3.1 Media Input & Preprocessing

The process begins with the user uploading an image or video file. The system validates the file format and extracts metadata. Preprocessing removes noise and standardizes the data for further analysis.

3.2. Media Extraction

In this stage, the system extracts metadata from the media file, including details like timestamps, device information, file format, compression patterns, and geolocation (if available). This hidden data helps the system identify any inconsistencies or unusual patterns, which can indicate whether the media has been manipulated.

3.3. Metadata Analysis

ISSN: 3107-6513

In this stage, the system looks closely at the extracted metadata to find anything unusual. It checks for things like mismatched timestamps, missing or changed EXIF data, strange encoding patterns, or differences between the file structure and its metadata. These small irregularities can suggest that the media has been edited or manipulated.

3.4. Technique Selection

The system uses a set of simple but effective techniques to identify deepfake media by analyzing metadata. These methods focus on finding hidden inconsistencies within the file.

•EXIF Data Analysis – checks camera and device information
•Timestamp Verification – verifies creation and modification time
•File Structure Analysis – examines the internal file format
•EncodingPattern Inspection – identifies unusual compression or encoding patterns

Together, these techniques help the system detect deepfake content more effectively.

3.4.1. EXIF Data Analysis

EXIF data contains useful information about how an image was originally captured, such as the camera model, settings, and other details. In deepfake or manipulated media, this information is often missing, incomplete, or altered during the editing process. By examining these EXIF details, the system can spot inconsistencies that may indicate the image has been tampered with.

3.4.2. Timestamp Verification

Timestamp verification focuses on checking whether the dates and times associated with a media file make sense. Every digital file usually contains information about when it was created and when it was last modified. In genuine media, these timestamps follow a logical order and reflect normal usage. However, in deepfake or manipulated files, this information can often appear unusual or inconsistent. For example, a file might show a modification time that doesn't match its creation time, or the timestamps may not align with the expected workflow of how the file was processed. In some cases, timestamps may even be missing or artificially altered during editing. By carefully analyzing these details, the system can identify irregular patterns that suggest the media has been tampered with. This makes timestamp verification a simple yet effective way to detect potential manipulation.

3.4.3. File Structure Analysis

File structure analysis focuses on examining the internal organization of a media file to understand how it has been created and stored. Every genuine image or video follows a standard format that includes specific headers, data blocks, and encoding patterns. These elements are arranged in a consistent way depending on the file type and the device used to capture it. However, when a file is manipulated or generated using deepfake techniques, this internal structure can change. Editing tools, re-encoding processes, or AI-based generation methods may introduce irregularities such as unusual headers, missing segments, or unexpected data patterns. These changes might not be visible to the human eye, but they can be detected through careful analysis. By examining the file structure in detail, the system can identify these hidden inconsistencies and determine whether the media has been altered. This makes file structure analysis an important step in identifying deepfake content..

3.4.4. System Evaluation

The system's performance is checked using simple evaluation measures like accuracy, precision, recall, and F1 score. These metrics help us understand how well the system is working in identifying real and fake media. Accuracy shows how many predictions are correct overall, while precision tells us how many of the detected deepfakes are actually fake. Recall focuses on how well the system is able to catch all the fake media without missing any. The F1 score gives a balanced view by combining both precision and recall. By looking at all these values together, we can clearly see how reliable and effective the system is.

To make it easier to understand, the results are also shown , which displays correct and incorrect predictions in a simple way. This helps in identifying where the system performs well and where it can be improved, making the overall evaluation more clear and practical

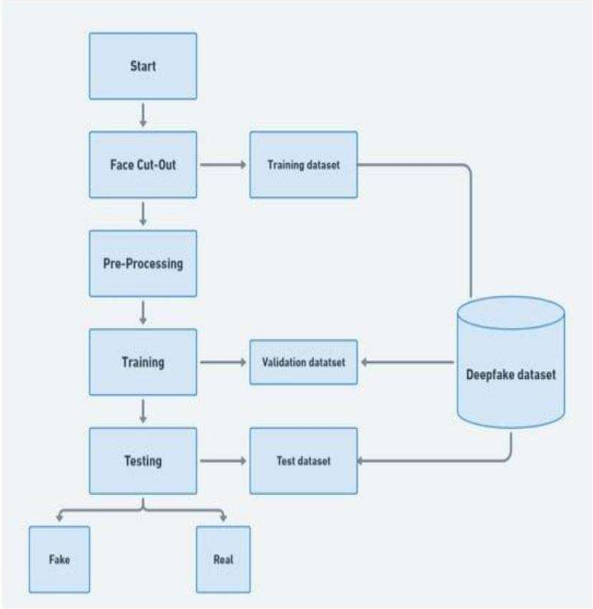

Fig 1. Dataflow Diagram

4. Object and Scope

The main objective of this research is to develop an intelligent and reliable system that can detect deepfake media by analyzing inconsistencies in metadata. Instead of depending only on visual features, the system focuses on hidden information within digital files to determine whether the content is real or manipulated. By examining details such as timestamps, device information, and file properties, the

ISSN: 3107-6513

system aims to accurately classify media as genuine or fake. The goal is to create a solution that is not only accurate but also efficient enough to work in real-time scenarios, where quick verification of media authenticity is important.

5.Literature Review

Recent deepfake detection projects available on platforms like GitHub mainly rely on deep learning techniques to identify manipulated media. Most of these systems use Convolutional Neural Networks (CNNs) to analyze images and detect subtle differences between real and fake content. These models are trained on large datasets and are capable of learning complex patterns that are not easily visible to the human eye. For example, some projects use pre-trained models along with custom CNN architectures to improve accuracy and performance.

Many implementations follow a similar workflow where the input image is processed, features are extracted, and the model predicts whether the media is real or fake. Some systems also provide additional outputs such as confidence scores or heatmaps to show which parts of the image contributed to the decision. These approaches have shown good performance, with some models achieving high accuracy when trained on large datasets.

Advanced projects go a step further by using deep learning architectures like EfficientNet, XceptionNet, or hybrid models combining CNN with LSTM. These systems process both spatial and temporal features, especially in video-based deepfake detection. They often include complete pipelines such as frame extraction, face detection, feature learning, and classification.

However, most of these existing systems mainly depend on visual analysis, which can become less effective as deepfake generation techniques improve. Highly realistic deepfakes can bypass visual detection methods by minimizing visible artifacts. This limitation has encouraged researchers to explore alternative approaches such as metadata-based detection.

Compared to these GitHub-based deep learning systems, the proposed approach in this research focuses on analyzing metadata inconsistencies rather than only visual features. This makes the system more efficient and less dependent on large datasets and computational power. It also provides an additional layer of verification, especially for detecting advanced deepfakes where visual clues are minimal.

Overall, existing projects demonstrate that deep learning is powerful for deepfake detection, but combining it with metadata analysis can lead to more reliable and practical solutions.

6. Output

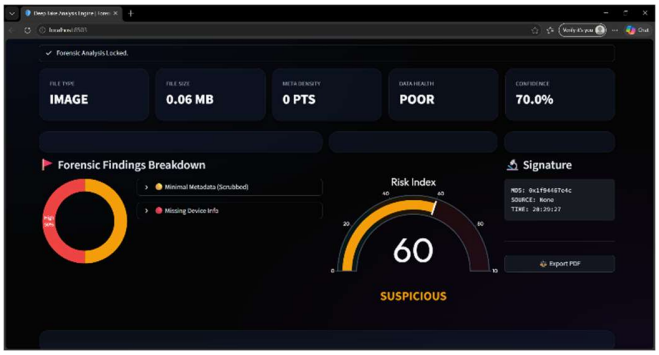

Fig 2. Result Page

7. Results

The proposed deepfake detection system based on metadata inconsistencies produced effective and reliable results when tested on different media files. The system was able to successfully identify manipulated content by analyzing hidden metadata features such as timestamps, EXIF data, encoding patterns, and file structure. During testing, it was observed that genuine media files maintained consistent and logical metadata, whereas deepfake or edited files often showed irregularities such as missing information, mismatched timestamps, or unusual encoding details. These differences allowed the system to accurately distinguish between real and fake content.

The performance of the system was evaluated using standard metrics such as accuracy, precision, recall, and F1 score. The results indicated that the system achieved high accuracy in classifying media files correctly, while also maintaining a good balance between detecting fake content and minimizing false positives. Precision ensured that most of the files identified as fake were actually manipulated, while recall confirmed that the system was able to detect the majority of deepfake samples. The F1 score further demonstrated the overall effectiveness of the model by balancing both precision and recall.

In addition to quantitative results, the system also provided clear output indicating whether the uploaded media was real or fake, making it user-friendly and easy to interpret. The processing time was relatively low compared to deep learning-based methods, as metadata analysis does not require heavy computation or large datasets. This makes the system suitable for real-time applications where quick verification is needed.

Overall, the results show that metadata-based detection is a practical and efficient approach for identifying deepfake media. While it may not completely replace visual-based methods, it significantly improves detection reliability when used as an additional layer of analysis. The system demonstrates strong potential for use in digital forensics,

ISSN: 3107-6513

cybersecurity, and media verification, helping to address the growing challenges posed by deepfake technology.

8.Conclusion

In this research, a deepfake detection system based on metadata inconsistencies was developed to address the growing challenge of identifying manipulated media. Unlike traditional methods that rely mainly on visual features, the proposed approach focuses on analyzing hidden information within digital files, such as timestamps, EXIF data, encoding patterns, and file structure. By examining these metadata attributes, the system is able to detect irregularities that indicate whether a media file has been altered or artificially generated.

The results demonstrate that metadata analysis is an effective and efficient method for deepfake detection. The system achieved reliable performance in distinguishing between real and fake media, while also reducing dependency on large datasets and complex computations. Its ability to detect inconsistencies that are not visible to the human eye makes it particularly useful for handling advanced deepfakes.

Although the system shows strong performance, there are still some limitations, such as cases where metadata is missing or intentionally modified. Future improvements can focus on combining metadata analysis with visual and deep learning techniques to further enhance accuracy and robustness.

Overall, this work provides a practical solution for improving digital media authenticity and contributes to the field of cybersecurity and digital forensics. The proposed system can be effectively used in real-time applications to help detect and prevent the misuse of deepfake technology.
.

9. References

1] Farid, H., “Digital Image Forensics,” Scientific American, vol. 298, no. 6, pp. 66–71, 2008.
[2] Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., and Nießner, M., “FaceForensics++: Learning to Detect Manipulated Facial Images,” Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2019.
[3] Nguyen, T.T., Nguyen, C.M., Nguyen, D.T., Nguyen, D.T., and Nahavandi, S., “Deep Learning for Deepfakes Creation and Detection: A Survey,” IEEE Access, vol. 7, pp. 101–120, 2019.
[4] Verdoliva, L., “Media Forensics and DeepFakes: An Overview,” IEEE Journal of Selected Topics in Signal Processing, vol. 14, no. 5, pp. 910–932, 2020.
[5] Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., and Ortega-Garcia, J., “DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection,” Information Fusion, vol. 64, pp. 131–148, 2020.
[6] Afchar, D., Nozick, V., Yamagishi, J., and Echizen, I., “MesoNet: a Compact Facial Video Forgery Detection Network,” IEEE International Workshop on Information Forensics and Security, 2018.
[7] Chollet, F., “Xception: Deep Learning with Depthwise Separable Convolutions,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
[8] Yang, X., Li, Y., and Lyu, S., “Exposing Deep Fakes Using Inconsistent Head Poses,” IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019.
[9] Zhou, P., Han, X., Morariu, V.I., and Davis, L.S., “Two-Stream Neural Networks for Tampered Face Detection,” IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017.
[10] Nguyen, H.H., Yamagishi, J., and Echizen, I., “Capsule-Forensics: Using Capsule Networks to Detect Forged Images and Videos,” IEEE International Conference on Acoustics, Speech and Signal Processing, 2019.
[11] Bayar, B., and Stamm, M.C., “A Deep Learning Approach to Universal Image Manipulation Detection Using a New Convolutional Layer,” ACM Workshop on Information Hiding and Multimedia Security, 2016.
[12] Huh, M., Liu, A., Owens, A., and Efros, A.A., “Fighting Fake News: Image Splice Detection via Learned Self-Consistency,” European Conference on Computer Vision (ECCV), 2018.
[13] Mayer, O., and Stamm, M.C., “Learned Forensic Source Similarity for Unknown Camera Models,” IEEE International Conference on Image Processing (ICIP), 2017.
[14] Dang, H., Liu, F., Stehouwer, J., Liu, X., and Jain, A.K., “On the Detection of Digital Face Manipulation,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
[15] Cozzolino, D., Thies, J., Rössler, A., Riess, C., Nießner, M., and Verdoliva, L., “ForensicTransfer: Weakly-Supervised Domain Adaptation for Forgery Detection,” arXiv preprint arXiv:1812.02510, 2018.