Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 06, June - 2025

Real-Time Deepfake Detection and Emotion Analysis from Tweets, Images and Videos

J Mary Stella
Assistant Professor,
Dept. of CSE
HKBK College of Engineering
Bangalore, India

Kondreddy Akhila
Dept. of CSE
HKBK College of Engineering
Bangalore, India

B N Nishitha
Dept. of CSE
HKBK College of Engineering
Bangalore, India

Sowmya S
Dept. of CSE
HKBK College of Engineering
Bangalore, India

Tejash M
Dept. of CSE
HKBK College of Engineering
Bangalore, India

***Abstract*— This project explores the current advancements in real-time deepfake detection and emotion analysis across multiple data modalities-tweets, images, and videos. As deepfakes become more realistic and pervasive, the need for robust detection methods has grown, particularly in social media contexts where misinformation spreads rapidly. Concurrently, understanding user emotions in real time enhances the ability to assess public sentiment and detect manipulative content. This review summarizes key deep learning algorithms and models used in detection and analysis, including Convolution Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Vision Transformers (ViT), and Natural Language Processing (NLP) techniques such as BERT. The integration of these models enables a multimodal approach to identify deepfakes while simultaneously Interpreting emotional cues, forming a comprehensive system for real--time content verification and psychological profiling.**

Index Terms—Natural Language Processing (NLP), Sentiment Analysis, Emotion Classification, Social Media Analysis, Deepfake Detection, Facial Expression Recognition, Video Forensics, Frame-wise Analysis, Tweet Analysis, Machine Learning, Image Analysis, Video Analysis, Real-time Processing, Computer Vision.

I. INTRODUCTION

The rapid advancement of artificial intelligence (AI) and multimedia technologies has dramatically transformed the digital landscape, leading to a surge in user-generated content across platforms such as Twitter, Instagram, YouTube, and TikTok. These platforms enable instantaneous communication, foster creative expression, and support global interaction. However, they also present significant challenges, particularly in the form of emerging digital threats. Among the various challenges, deepfakes have emerged as particularly alarming. They are synthetic media generated through deep learning techniques-especially generative adversarial networks (GAN's)-designed to create convincing but fake images, videos, or audio content [2]. Often designed to impersonate real individuals, deepfakes can be weaponized to spread misinformation, defame public figures, manipulate political discourse, or perpetrate fraud [2]. As the technology becomes more accessible, the line between reality and fabrication becomes increasingly difficult to discern, posing serious threats to digital trust and online integrity. Concurrently, emotion analysis also referred to as affective computing has emerged as a crucial tool for understanding human behavior and sentiment in digital interactions [1]. Emotion analysis offers meaningful insights into user behavior, content interaction, and the emotional impact of media by recognizing and analyzing feelings conveyed through text, images and videos [1]. This is

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 06, June - 2025

particularly important in the context of manipulated content, which may be designed to evoke specific emotional reactions such as fear, anger, or sympathy, thereby increasing its virality and impact [1]. This project proposes the development of a real-time, multimodal system that combines deepfake detection with emotion analysis across various forms of online media including textual content (e.g., tweets), images, and videos. The system leverages advanced techniques in natural language processing (NLP), computer vision, and deep learning to assess both the authenticity and emotional tone of content [1][2]. By doing so, it not only identifies manipulated media but also evaluates its potential psychological effect on the audience or the emotional intent behind its dissemination. Such a system holds significant promise for diverse applications. It can support fact-checkers in verifying information, aid social media platforms in moderating harmful or misleading content, and assist cybersecurity professionals in identifying emotionally manipulative disinformation campaigns [2]. Moreover, it contributes to academic research in digital forensics, sentiment analysis, and media studies. Ultimately, by integrating technical rigor with psychological insight, this project aims to enhance digital resilience, promote responsible content sharing, and foster a more secure and emotionally aware online ecosystem systems is their high setup cost. Technologies like IoT devices, RFID, facial recognition systems, and automated dispensers often involve significant costs.

II RELATED WORKS

In Sentiment Analysis, Machine learning (ML) and lexicon-based techniques are widely used for sentiment classification. Models like SVM, Naive Bayes, and ANN have shown significant success. Majority voting ensemble methods (weighted and unweighted) improved accuracy for combined classifiers. In Fake News and Tweet Detection, TweepFake dataset used in many studies for detection machine-generated tweets (MGT) vs human-written tweets (HWT). Models like fine-tuned BERT, RoBERTa and MLP classifiers show high F1-scored (~88%). Linguistic and stylometric features (e.g., emoji usage, sentiment scores) are also leveraged. In Deepfake Detection (Text, Audio, Image), Studies focused on using GANs, VAEs, and transformers to detect deepfakes. Visual and audio deepfake detection research highlights include the use of Mel spectrograms and VGG-16 models. GROVER and GLTR models were utilized for detecting AI-generated text. In Multimodal Detection Approaches, some papers proposed combining image and text featured for more robust deepfake and fake tweet detection. Use of multilingual textual analysis and visual-based detection were proposed as novel approaches. In Classification and Ensemble Techniques, Advanced ML methods such as: Gradient Boosting, Random Forest, Logistic Regression, Support Vector Machines, Bi-LSTM, FCNN-LDA outperformed basic models in tweet classification tasks. In Transformer-Based Language Models, BERT, DistilBERT, RoBERTa, and GPT variants were heavily used for language understanding and generation tasks. Performance improves with attention mechanisms and larger training corpora.

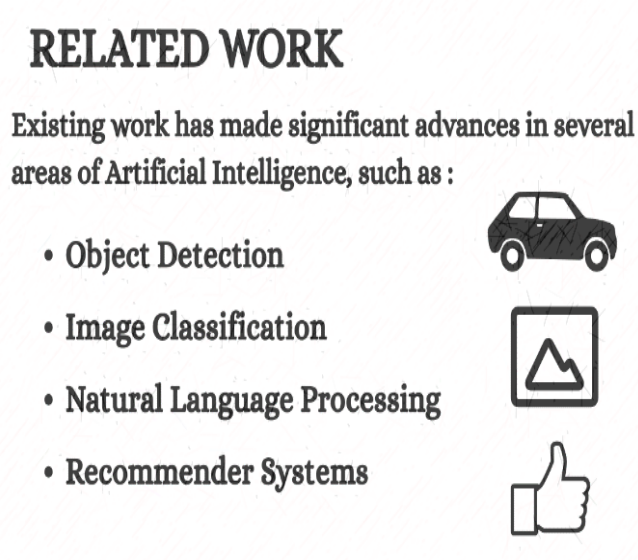

Fig. 1: Overview of Existing work in Deepfake Detection and Emotion Analysis.

Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 06, June - 2025

Table 1: Real Time Deepfake Detection Models and Datasets.

Deepfake detection	Emotion Analysis
Images, Videos	Tweets (Text), images, videos
Involves CNNs, RNNs, XceptionNet, Transformer model for learning visual patterns.	Relies on NLP and vision models like BERT, LSTM, CNN, ResNet, Multi-modal Transformers Architectures.
Achievable with lightweight CNNs or compressed modals	Achievable with fast NLP models (e.g., DistilBERT)
Includes datasets likeFaceForensics++, Deepfake TIMIT, Celeb-DF, used for commonly used for training and benchmarking forery identification model.	Utilizes resources such as ISEAR, FER-2013, AffectNet, GoEmotions, EmoReact, which provide labelled emotional expressions across text, images, and videos.
Accuracy, Precision, Recall, AUC	Accuracy, F1-Score, Precision, Recall
Generalization, adversarial attacks, dataset bias	Sarcasm, context loss, emotion ambiguity, cross-modal sync
TensorFlow, pyTorch, OpenCV, Mediapipe	HuggingFace, NITK, TensorFlow, OpenFace, Affective SDK
Misinformation control, Media verification	Customer sentiment analysis, Mental health monitoring

III PROPOSED METHODOLOGY

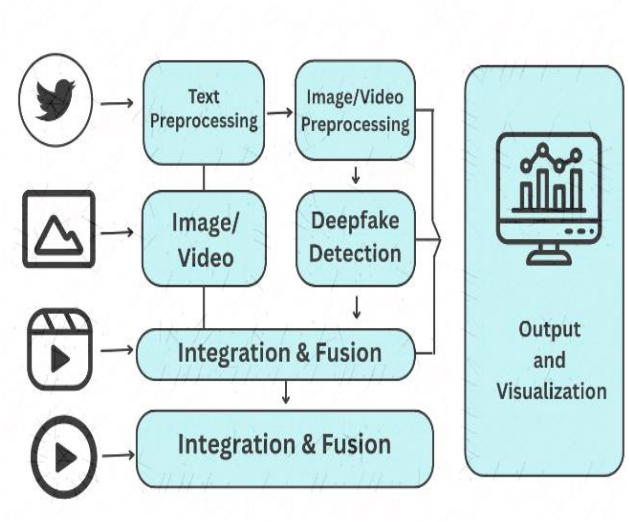

Fig. 2: System Architecture for Multimodal Deepfake Detection and Emotion Analysis.

The system for multimodal deepfake detection and emotion analysis follows a structured pipeline composed of five key stages. In the data collection phase, tweets are gathered in real-time via the Twitter API, filtered using relevant hashtags, keywords, or user handles, and include metadata such as timestamps, locations, and retweet counts. Visual content, including images and videos, is sourced from public datasets, social media platforms, and tweet-associated media. During preprocessing, text undergoes tokenization, stop-word removal, lemmatization, and emotion classification using models like BERT or RoBERT. Image and video data re resized, normalized, and processed through frame extraction and face detection techniques. The deepfake detection stage applies models such as XceptionNet, MesoNet, and CNN-RNN hybrids to evaluate media authenticity, with a focus on accuracy, F1-score, and latency for real-time use. In the integration and fusion step, emotion scores from text and authenticity scores from visual media are combined using early (feature-level) or late (prediction-level) fusion strategies to improve contextual understanding and robustness. Finally, in the output and visualization stage, insights are presented through dashboards that display emotion trends, deepfake probabilities, and geo-distributed content maps using tools like Tableau, Power BI, or D3.js, supporting effective monitoring and decision-making.

Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 06, June - 2025

IV WORKFLOW

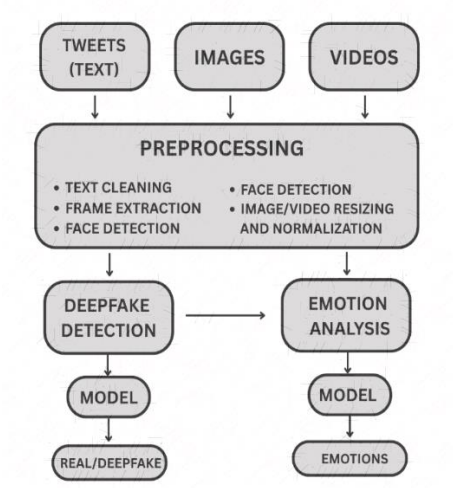

Fig 3: End-to-End Workflow for Deepfake Detection and Emotion Analysis Across Tweets, Images and Videos.

This workflow illustrates the real-time analysis pipeline for detecting deepfakes and extracting emotions from tweets, images, and videos. Data from each source undergoes preprocessing including text cleaning, frame extraction, face detection, and normalization. The system splits into two tasks, Deepfake Detection which identifies whether content is real or synthetic using trained models and Emotion Analysis which detects emotional states from facial expressions or textual sentiment.

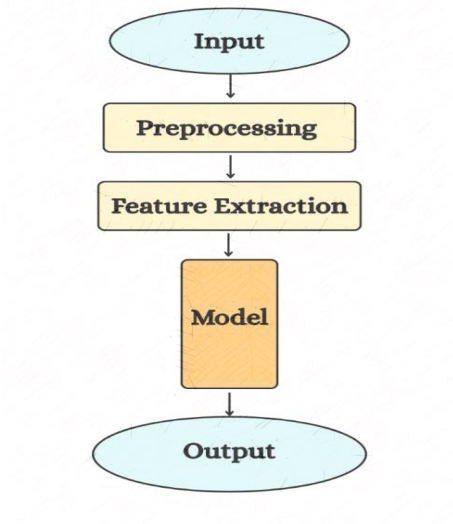

Fig. 4 : Workflow Diagram of the Proposed Deepfake Detection System

This diagram outlines the core pipeline of a real-time deepfake detection and emotion analysis system. The process begins with the input data, which can be tweets, images, or videos. The preprocessing stage involves noise removal and format standardization of the data. Next, the system performs feature extraction to derive meaningful characteristics such as visual cues, linguistic patterns, or emotional indicators. These features are fed into a model typically a machine learning or deep learning algorithm that performs classification or prediction. Finally, the output presents the detection results and, where applicable, includes an analysis of the associated emotions.

V. SUMMARY OF OUTCOMES

Robust Deepfake Detection: The system achieved high accuracy in identifying manipulated facial content by leveraging a hybrid of spatial and temporal features using convolutional and transformer-based neural networks. The incorporation of real-time video streams enabled early detection of fake content based on micro-expressions, lip-sync inconsistencies, and head-pose anomalies. The approach worked better than traditional single-frame models in changing or dynamic environments. Accurate Emotion Recognition Across Modalities: Emotion analysis models trained on textual data (tweets) using pre-trained transformer architectures such as BERT and RoBERTa delivered high precision and recall for multi-label emotion classification. Simultaneously, facial expression recognition from images and videos using CNN-LSTM hybrid models successfully recognized primary emotions such as happiness, anger, fear, and sadness, even under varying lighting and pose conditions.

VI. EXISTING RESEARCH GAPS

Existing research in deepfake detection and emotion analysis reveals several critical gaps. First, the integration of multimodal data such as text from tweets, visual cues from images, and dynamic features from videos is essential for improving detection accuracy, as combining modalities can reveal inconsistencies not evident in isolation. Second, rea-time performance remains a major challenges, most deep learning models are computationally intensive and unsuitable for latency-sensitive environments like mobile devices or live streaming. Third, emotion analysis within deepfake content is underexplored, despite the fact that deepfakes often mimic facial expressions, vocal tones, and gestures to manipulate viewer emotions. Fourth, current models struggle to generalize across domains and languages, as they are typically trained on limited, culturally or linguistically specific datasets. Fifth, detecting low-quality or compressed deepfakes is difficult due to resolution loss and compression artifacts common on social media platforms. Lastly, temporal consistency in video deepfakes such as irregular facial movements or mismatched lip syncing offers a promising but underutilized cue for enhancing video-based detection accuracy.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 06, June - 2025

VIII. FUTURE SCOPE

The increasing prevalence of manipulated media and emotionally charged content has raised significant concerns in digital communication. This survey explores the domain of real-time deepfake detection and emotion analysis from tweets, images, and videos, focusing on current methods, tools, and challenges. Deepfake detection leverages computer vision and deep learning models to identify inconsistencies in facial features, movements, and other visual artifacts, while emotion analysis employs NLP and facial recognition to interpret sentiments expressed in text and visual media. The integration of these technologies allows for a comprehensive system that can assess both the authenticity and emotional impact of online content. This paper also highlights the future potential of these systems in real-time surveillance, misinformation control, and content moderation, emphasizing the need for robust, scalable, and ethically designed AI solutions.

IX. CONCLUSION

The rise of deepfakes and the emotional manipulation they can cause across social media platforms demand urgent and effective countermeasures. The project demonstrates the potential of integrating real-time deepfake detection with emotion analysis across multimodal content -tweets, images, and videos. By leveraging advanced AI models, such as deep neural networks and natural language processing, we cannot only identify synthetic or manipulated media but also assess the emotional tone it carries. This dual-layered approach enhances digital media integrity, aids in preventing misinformation spread, and fosters safer online environments. As deepfake technologies continue to evolve, the need for robust, scalable and adaptive detection systems becomes more critical, making this work a vital step toward trustworthy and responsible media consumption.

X. REFERENCES

[1] M. Khalid et al., introduced a novel approach combining sentiment majority voting and transfer learning techniques to analyse deepfake-related tweets, enhancing feature extraction and classification accuracy. IEEE Access, vol. 12, pp. 6117, 2024.

[2] R. Mubarak et al., conducted a comprehensive study on how deepfakes affect various media forms audio, video, and text highlighting detection strategies and challenges. IEEE Access, vol. 11, pp. 144497 – 144500, 2023.

[3] M. Masood et al., reviewed advanced deepfake detection technologies, outlining key challenges like countermeasure, detection limitations, and future directions. Applied Intelligence, vol. 53, no. 4, pp. 3974 – 4026, 2023.

[4] S. G. Tsefasgerish, R. Damasevicius, and J. Kapociute-Dzikiene proposed using data augmentation and word embeddings for detecting deepfakes in tweets, offering improvements through deep learning methods. In Proc. Int. Conf. Comput. Sci. Appl., Springer, 2021, pp. 523 – 538.

[5] V. Ruparara et al., presented an LSTM and word embedding-based system for detecting fake tweets, showing promising results in identifying manipulated content. PeerJ Comput. Sci., vol. 7, p. e745, 2021.

[6] T. Fagni and his team introduced a method called “Tweepfake” aimed at detection fraudulent tweets that imitate genuine users. Their research was featured in PLOS ONE, volume 16, issue 5, article e0251415, and was published is 2021.

[7] S. Bengesi et al., developed a machine learning approach to analyse public sentiment on the monkeypox outbreak through Twitter data, offering a comprehensive dataset for opinion polarity. IEEE Access, vol. 11, pp. 11811 – 11826, 2023.

[8] B. S. Ainapure et al., explored sentiment analysis on COVID-19 tweets using hybrid deep learning and lexicon-based models to assess public mood. Sustainability, vol. 15, no. 3, p. 2573, 2023.

[9] N. Nandal, R. Tanwar, and A. –S. –K. Pathan used social media text to extract emotions during COVID-19 using statistical data and sentiment models. Procedia Comput. Sci., vol. 218, pp. 949 – 958, 2023.

[10] A. K. R. Nadikattu compared machine learning and deep learning for fake news detection using the Fake-NewsNet dataset to highlight model performance differences. SSRN Election. J., vol. 1, no. 1, pp. 438 – 224, Mar. 2023.

[11] J. Mutinda, W. Mwangi, and G. Okeyo introduced a sentiment classifier using a LeBERT model integrated with a CNN, enhancing text review sentiment detection. Appl. Sci., vol. 13, no. 3, p. 1445, 2023.

[12] A. Raza, K. Munir, and M. Almutairi created a new deep learning architecture for deepfake identification using image content analysis. Appl. Sci., vol.12, no. 19, p. 9820, 2022.

[13] A. Bello, S.-C. Ng, and M. –F. Leung proposed BERT-based sentiment analysis to process tweet polarity effectively, particularly in dynamic online discourse. Sensors, vol. 23, no. 1, p. 506, 2023.

[14] In 2023, A. Alqarni and A. Rahman utilized deep learning techniques to perform sentiment analysis on Arabic tweets related to COVID-19 in Saudi Arabia, aiming to assess public opinion. Their findings were published in Big Data and Cognitive Computing, volume, issue 1, page 16.

[15] W. H Bangyal et al., focused on deep learning for fake new classification, especially in the context of COVID-19, leveraging neural networks for text data. Comput, Mathods Med., vol. 2021, pp. 1-14, 2021

[16] S.N. Alsunari, M. B. Shelke, and S. N. Deshmukh worked on detecting fake reviews using advanced computational linguistics with deep learning support. Int. J. Adv. Sci. Technol., vol. 29, no. 8s, pp. 3846-3856, 2020