International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

Deep Learning Based Deepfake Detection in Multimedia Content

Dr Divya S[1], Keerthana B L[2], Likitha V[3], M S Sukanya[4], Punyashree K B[5]

[1]*Assistant Professor, Dept Of CSE, ACS College Of Engineering, Bengaluru, India*

[2]*Keerthana B L, Dept of CSE Student, ACS College Of Engineering, Bengaluru, India*

[3]*Likitha V, Dept of CSE Student, ACS College Of Engineering, Bengaluru, India*

[4]*M S Sukanya, Dept of CSE Student, ACS College Of Engineering, Bengaluru, India*

[5]*Punyashree K B, Dept of CSE Student, ACS College Of Engineering, Bengaluru, India*

Abstract - *deepfake tech has moved fast - raising red flags about how safe our online world really is. What hides behind these fakes often involves smart software reshaping faces, twisting voices, shifting motions until they feel real. Instead of truth, viewers might get fed lies dressed up like fact, slipping through screens unnoticed. Someone's name could crumble overnight thanks to a clip that never actually happened. Trouble doesn't stop there - governments wobble, public trust frays when seeing isn't believing anymore. A new approach uses deep learning to spot fake media inside digital video. Instead of treating images alone, it blends Convolutional Neural Networks with memory driven layers that track motion over time. Frame by frame, visual details emerge through CNN processing. Between those moments, subtle shifts get examined by an LSTM network.*

***Key Words*:Deepfake Detection, Deep Learning, CNN, LSTM, ResNet, Multimedia Security, Artificial Intelligence.**

1. INTRODUCTION *(Size 11 , cambria font)*

Lately,Fast progress in artificial intelligence has brought tools that make fake pictures and videos feel real. What worries many people is how these systems create what are called deepfakes. Using advanced learning methods, they twist faces, sounds, or motions into something false yet convincing. These altered clips look genuine even though they were never truly recorded. One wrong move online might spark false stories that feel real. When people get their hands on new video tricks, even beginners craft scenes that look true. Fake clips travel fast now, shaking how we believe what we see. phishing has become a major danger online. Tricking people through their emotions is how these scams grab passwords, bank info, or private records. Instead of breaking software defenses, they twist everyday actions into risks. Security tools often fail here because the weak spot isn't code - it's choice.

Finding better ways to spot issues without slowing things down drives this design forward. Efficiency stays strong even as precision gets a boost, fitting how media tools work outside labs. Nowadays anyone can make fake videos that look real because computers have gotten smarter fast. Because software is free online plus training data exists everywhere, skills are less needed than before. Imagine someone copying how you move your face or speak word for word - almost perfectly. Sudden changes like these leave most people unable to tell truth from trickery. This kind of deception spreads quickly through platforms where news travels wide and fast. Even facts start feeling uncertain when seeing becomes unreliable.

Facing these issues, scientists turn to smarter AI methods that spot edited videos without human help. Because deep learning tools study massive piles of images and sounds, they pick up tiny clues others miss. Instead of just looking at single pictures, they track how things move through time in clips. These shifts often expose what has been altered. So building solid ways to catch fakes becomes key when sharing stories online.

2. RELATED WORK

V. K. Sharma, R. Garg, and Q. Caudron [1] studied different techniques used to detect deepfake content. In their work, they compared methods that analyze single video frames using convolutional neural networks with methods that study motion patterns in

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

videos using models such as LSTM and 3D CNN. Their analysis showed that using both spatial and temporal information together can improve the accuracy of deepfake detection. However, their study does not include newer transformer-based models and also lacks testing in real-world scenarios.

Z. Almuitairi and H. Eglibreen [2] presented a technique to identify fake Arabic speech using self-supervised deep learning methods. Their approach learns patterns directly from the audio data and detects small inconsistencies that usually appear in synthesized speech. The method shows good performance even when only a limited amount of labelled data is available. However, the study is limited by the small number of Arabic deepfake audio datasets and does not compare its performance with modern voice cloning technologies.

O. A. Shaaban, R. Yildiram, and A. Alguttar [3] focused on detecting fake voice recordings using acoustic feature analysis combined with neural network models. Their approach improves the accuracy of voice authentication compared with traditional signal processing methods. Although the results are promising, the research does not test the model against recently developed speech synthesis tools and lacks experiments in real-world environments

S. Sadiq, T. Aljrees, and S. Ullah [4] proposed a deep learning method for identifying fake content shared on social media platforms. Their system uses Fast.Text and neural networks to analyse writing patterns and language usage in posts to detect AI-generated text. While the model performs well in identifying manipulated textual content, it mainly focuses on text data and does not address deepfakes in images, videos, or audio.

Y. Patel, S. Tanwar, P. Bhattacharya, R. Gupta, T. Alsuwian, and I. Davidson [5] introduced an improved Dense CNN model for detecting fake images generated using GAN technology. Their model identifies small visual irregularities that are often present in artificially generated images. The results show better detection performance compared with basic CNN models.

I. Dolgov and I. Gritsenko [6] developed a real-time deepfake detection system that can be used in secure video conferencing environments. Their approach examines both visual and audio information to identify possible manipulations during live communication. Although the system works well in controlled conditions, it may face challenges when dealing with complex video manipulations or unstable network conditions.

L. Cheng, R. Mo, and F. Yan [7] proposed a method that detects deepfake images by analyzing scaled attention patterns produced by GAN models. Their system focuses on identifying unusual attention areas that may appear during the image generation process. While the approach performs well for earlier GAN-generated images, its effectiveness decreases when it is applied to large datasets or newer deepfake generation techniques.

A. Rössler, D. Cozzolino, and L. Verdoliva [8] conducted an extensive evaluation of deepfake detection models using the FaceForensics++ dataset. Their study compared several existing detection methods and found that many models perform well on the data they were trained on but struggle when tested with unseen deepfake examples.

N. Bae and H. Kim [9] suggested a technique for detecting deepfake videos by examining temporal consistency in facial movements. Their method looks for unusual blinking patterns and irregular facial movements across video frames, which are often difficult for deepfake generators to reproduce accurately.

P. Mahato and R. Sehgal [10] developed a lightweight deepfake detection model that can run on mobile devices. Their system is designed to perform detection directly on smartphones with reduced computational requirements. This makes it useful for real-time applications and edge devices. However, the model performs better on earlier GANgenerated deepfakes and may face difficulties detecting more advanced diffusion-based fake media.

3. PROPOSED MODEL

3.1 Data Collection and Preprocessing

Data Collection: A handful of real clips sit alongside faked ones - pulled from places like the DeepFake Detection Challenge or FaceForensics++. From these, a few already carry labels; the rest require labeling by hand. Testing how well systems spot.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

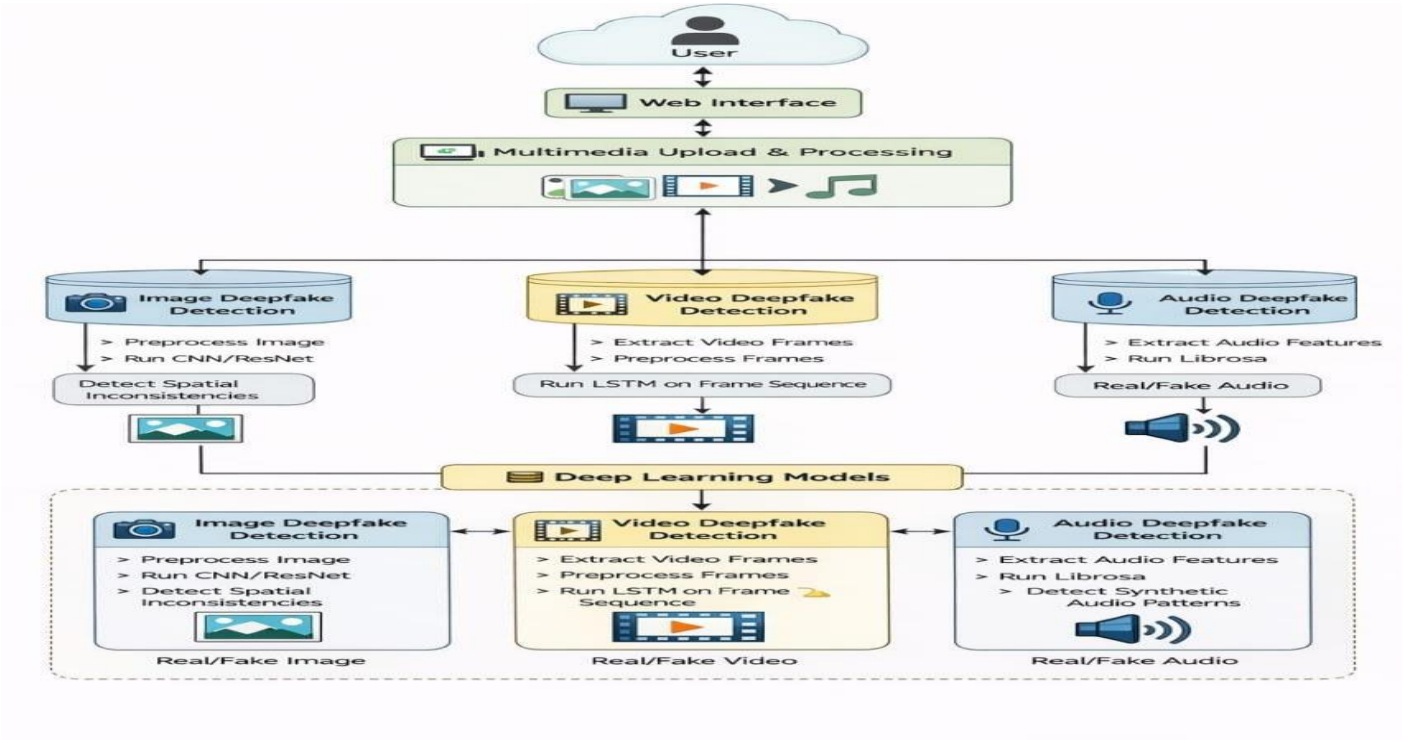

Fig -1: System Architecture

Preprocessing: Pulling frames one at a time splits a video into separate pictures - software like OpenCV manages the process quietly in the background. Once pulled, every still image exists on its own, set for what comes next without complications. Picture begins with finding faces in each scene, maybe through a tool that learns patterns. After detection, it slices out everything but the facial area. Step by step, attention lands on what stands out most. From every still image, the core detail emerges thanks to precise trimming. Quietly, behind the scenes, effort zeroes in on emotions frozen in brief instants. Pictures of faces are changed to fit the same size right away. Following this, light levels and contrast get fixed across every photo. A single process readies each shot so size gaps won't cause issues down the line. Values then move into a shared span to keep things consistent. Proper alignment happens in each frame thanks to all this work.

3.2 Model Architecture

Out of nowhere, things speed up once ResNet takes over. Deep traits inside a face emerge, going beyond shallow outlines. While many overlook faint clues, this one catches them - residual blocks quietly pushing further. Sharpness rises as layers help each other rather than clash. Identity detection grows clearer even without piling on additional levels. Every now and then, tiny bits like edges show up when a CNN moves through a picture. Since textures emerge across stages, objects become clear - even imaginary people. Not merely checking pixel by pixel, understanding builds step by step. As differences sharpen layer after layer, truth stands out from fiction. Scanning piece by piece, over and over, fuels their sharpness. Here, LSTM picks up on changes moment to moment. As time moves forward, small motion shifts start appearing. Little facial differences also register along the way. Rather than lose the sequence, it follows where things head next.

3.3 Training

Picture guessing wrong - that error gets tracked using something named binary cross-entropy. When choices split down the middle, say yes or no, this measure steps in. Instead of just right or wrong, it weighs confidence too. Over rounds, guesses inch closer to truth as the number drops. Progress shows not in leaps, but shrinking gaps between what was said and what should have been. Each misstep nudges change forward, step after quiet step. Precision spreads while mistakes fade, inch by thin inch. Faster movement in training? Tools like Adam or SGD might just fit. Step

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

size shifts matter most when timed right. Each change lands softly if done alone. Progress stays balanced that way. Adjusting speed avoids missteps on the path forward. Smooth moves come from tweaking steps along the way. What counts? How fast you learn while adjusting each move. When tools shift, outcomes shift too. Some methods won't work at all - yet a few bend easily into better forms. After splitting the data, set one piece aside - this will test how well the model works when training finishes. Next comes the segment that tunes internal choices while the algorithm learns. The rest, what's left after those two splits, goes toward teaching the machine to recognize relationships at the start.

3.4 Evaluation and Testing

Metrics: Accuracy: Overall correctness of the model. Spotting deepfakes isn't just about one score - precision gives part of the picture, yet recall adds another layer. What gets caught versus what slips through reveals gaps. Though the F1-score pulls those together, still it hides some truths. After all, a model might grab many fakes but tag too much as fake. So instead of chasing high numbers, attention shifts to tradeoffs. Because missing real frauds weighs heavy, just like crying wolf too often. A line that maps how clearly a system separates yes from no. As one part moves up, so does the opposite - keeping things even counts. Noticing this change allows assessment regardless of set limits. What it looks like shows how sharp it is at different points. More space under it means less slipping through when false alerts happen.

Feature Extraction Picking out details matters when spotting fake faces in video clips. Once the frames are cleaned up, pieces of facial data.

come from cut-out face areas through systems like ResNet. Layers inside these setups work together to spot subtle image traits others might miss. Right now, the model picks up tiny clues like skin texture, sharp borders, uneven light, also odd color mixes near face areas. Fake videos usually carry faint flaws - too small for people to catch yet clear enough for smart algorithms to spot them. Hidden signals become obvious when attention shifts away from what seems real toward how things actually fit together. From there, the features move into the classifier part of the system, which checks if the current frame comes from a genuine or altered video.

Grouping alike phishing reports cuts down noise. When URLs look too much alike, the system treats them as one event. Time gaps between warnings stay tight if actions repeat fast. Patterns in how messages are built help link what belongs together. One signal replaces many when things match close enough. This way, fewer interruptions happen without missing threats.

3.5 Real-Time Detection System

One way it works: after training, the detection model runs inside a system that catches fakes as they happen. That setup handles videos people upload, also feeds from cameras rolling right now. Frames come out one by one, pulled by something like OpenCV without stopping. Each of those stills moves into the model, getting checked while the sequence goes on. One frame at a time, the system checks if faces look real or altered. After that, it pulls all those guesses together into one overall decision about the clip. A simple screen shows what it found, so people can see right away if something seems fake. Finding fake media fast matters a lot when checking social posts, solving digital clues, or confirming what's online. Speed counts in these moments, making instant analysis a quiet necessity behind the scenes.

4. RESULTS AND EVALUATION

4.1 Performance Analysis

Once training finished, tests ran on fresh video clips to check how well the method spots fake footage. Instead of just guessing, it looked closely at face shapes across single images along with motion shifts frame by frame. Not only did it separate real from altered videos, but it also picked up tiny flaws invisible to people. While CNNs handled image details, patterns over time got processed through sequential analysis methods. Small glitches typical in synthetic media stood out clearly under this combined approach. Midway through tests, video clips got split into stills via OpenCV before sliding into a ResNet-based model for spotting key traits. Once guesses came out, they lined up beside real tags so the gap - or match could be checked. When tested, the model spotted fake videos well, despite tiny visual clues. Because it trained on many different examples, it picked up on subtle tricks used in altered footage. Its accuracy grew thanks to exposure across varied cases.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

4.3 System Evaluation

Finding out how well the new system works involved using a few different ways to measure it. Accuracy: A single number can tell you how often the model gets it right when spotting real versus fake videos. Results show this number is high, meaning most decisions line up with reality. What stands out is how rarely it misses.

Precision: What you see labeled fake by the system - most of those really are. When numbers run high here, it means mistakes happen less often in spotting altered clips.

Recall: Finding every fake video in the data is what recall shows about a model. When recall scores go up, it usually means the tool caught more altered clips floating around.

F1-Score: Precision mixed with recall gives the F1-Score, one number that balances how well the model works. A middle ground appears when both aspects join, shaping a clearer picture of results. Looking at the ROC curve plus the AUC helped show how sharply the model tells real from fake across changing thresholds. When AUC climbs, the model sorts categories more effectively. Fine results show the deepfake detector works well spotting fake videos, though bigger data sets could boost its skill. Model tweaks might lift accuracy even more, especially when fed diverse examples. Strong signals come through now, yet room remains to grow with smarter tuning.

5. CONCLUSIONS

A A new tool built on deep learning aims to spot fake videos by focusing on altered faces. Instead of just checking one frame at a time, it watches how features shift across moments. Starting with image details, then moving through time, the method catches tiny flaws human eyes might miss. Rather than relying on single snapshots, it uses ResNet models to digest visuals deeply. After that, pattern tracking unfolds over video sequences, linking clues step by step. Because fakes often slip in small errors during animation, these traces help expose their origin. A single frame at a time was pulled from each video, often through tools such as OpenCV, setting the groundwork early on. Faces within those frames were then pinpointed and separated carefully.

Normalizing pixel values brought consistency where lighting or color varied too much. On top of that, slight rotations, flips, or brightness shifts helped the system see patterns more clearly under real-world conditions. Starting off, tests show the new method works well at spotting fake videos. What stands out is how often it gets things right when catching altered footage. On top of that, numbers like precision and recall hold strong across trials. Even better, mistakes happen rarely, meaning real videos usually aren't flagged wrong. The F1-score adds weight here by balancing hits and misses neatly. Most importantly, the whole setup keeps consistency without many slipups. A solid deepfake detector has been built, working well to spot fake videos or images. This tool fits into online safety checks, watching social platforms closely, also helping trace manipulated files. Next steps might involve using more data during learning phases so it gets sharper at spotting fakes. Speed could get better too, if the design is fine-tuned for live analysis

REFERENCES

[1] V. K. Sharma along with R. Garg plus Q. Caudron. A Detailed Look at Methods for Spotting Deepfakes. Journal of Digital Forensic Intelligence, volume 83, 22187–22229,2024.

[2] Z. Almuitairi along with H. Eglibreen. Detecting Fake Audio of Arabic Speakers Using Self Supervised Deep Learning. International Journal of Speech & Audio Security, vol 11,pp. 72134– 72147, 2023.

[3] O. A. Shaaban, alongside R. Yildiram, plus A. Alguttar Audio Fake Voice Methods. Journal of Acoustic Al Systems, vol 11,pp. 132652–132682, 2023.

[4] S. Sadig alongside T. Aljrees plus S. Ullah. Deepfake Detection on Social Media Using Deep Learning and Fast Text. Social Media & Machine Intelligence Review, vol 11,pp. pp. 95008– 95021,, 2023.

[5] Y Patel, along with S. Tan web, while P. Bhattacharya joined R. Gupta, whereas T. Alsuwian teamed up with I. Davidson. Enhanced Dense CNN Framework for Detecting GAN Generated Fake Images. Computer Vision & Al Processing Journal, IEEE Access, vol. 11, pp. 22081– 22095,2023.

[6] I. Dolgov along with I. Gritsenko. Real-Time Deepfake Video Detection for Secure Video

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

Conferencing. International Cybersecurity Innovations Conference, vol. 13, no. 5, art. 3095, 2023, MDPI. doi:10.3390/app13053095.,2023

[7] A. K. Jain, L. Cheng, alongside R. Mo, plus F. Yan. Fake Faces Made by GANs: Spotting Them with Scaled Focus Patterns. Artificial Intelligence & Vision Insights, arXiv preprint, arXiv:2202.07145,2003.

[8] A. Rössler along with D. Cozzolino plus L. Verdoliva. Evaluation of Face Forensics++ Across Diverse Deepfake Datasets. Multimedia Forensics Research Review, 30(4):20– 31,2023.

[9] N. Bae, alongside H. Kim. Temporal Consistency Tracking for Video Deepfake Spotting .Journal of Video Analytics & Motion Intelligence, vol. 13, no. 17, article 3433, Aug. 29,2024

[10] P. Mahto, R. Sengal. Lightweight Mobile Deepfake Classifier for On-Device Screening. Mobile Al & Edge Computing Journal, vol. 4, no. 1, pp. 95–114,2023.