International Journal of Information Technology & Computer Engineering

ISSN 2347–3657
Volume 13, Issue 3, 2025

A Hybrid Approach for Video Forgery Detection in Digital Media Using Inception v3

M.Shafeeka Begum
Department of Computer science and Engineering
Arunachala College of Engineering for Women
Manavilai, Vellichanthai, India
shafeekabegum19@gmail.com

N. Visalatchi
Department of Computer science and engineering
Arunachala College of Engineering for Women
Manavilai, Vellichanthai, India
visalatchirathish@gmail.com

Abstract— **Numerous video forgeries based on AI-driven media manipulation techniques continue to increase in popularity which creates substantial risks for digital forensics as well as security systems and validates media authenticity. The detection methods for forgeries using both statistical analysis and manual features can no longer identify leading-edge AI-generated modifications thus requiring improved automatic detection systems. Deep learning-based models present themselves as an effective solution to handle this problem. Conferral Inception v3 along with other CNN systems shows excellent capabilities in extracting spatial information from singular video frames yet struggles to evaluate temporal inconsistencies which represents an essential factor for detecting modified sequences. The proposed study presents a combined video forgery detection system that uses Inception v3 spatial features combined with LSTM along with SVM and anomaly detection algorithms for temporal video examination. The proposed method surpasses standard CNN detection methods because it uses temporal dependencies between video frames for better accuracy. The hybrid model shows superior outcomes than standalone CNN models ResNet and VGG regarding deepfake manipulation and frame tampering detection because it achieves better precision, recall and AUC-ROC scores. The use of Grad-CAM heatmaps together with confusion matrices allows to obtain detailed information about forgery patterns. Real-time deployment of the model becomes possible through optimization techniques that include quantization and pruning while the system demonstrates its functionality for digital forensics work as well as social media moderation and misinformation detection. The proposed system framework advances the creation of efficient scalable and resistant video forgery detection systems to safeguard media security.**

Keywords— Video Forgery Detection, Deepfake Detection, Inception v3, Hybrid Deep Learning Model, Temporal Anomaly Analysis

I. INTRODUCTION

Modern digital media editing software together with AI-driven manipulation methods result in worrying increases of video forgeries that generate deepfakes and splicing attacks and frame insertions which are hard for expert analysts to detect. The development of fake videos presents dangerous security risks to the nation while threatening forensic digital investigations and media transparency and sources of online information therefore demands immediate video verification measures in the modern digital age [1]. The protection of original video content prevents the growth of fake information while stopping both false identities and altered legal proofs in court proceedings [2]. The detection of altered videos presents difficulties because contemporary video manipulation methods merge artificial segments with actual content while discrepantly modifying vocal and facial characteristics [3]. Traditional methods used for forgery detection rest on handcrafted features alongside statistical inconsistencies with motion-based anomaly detection but these methods prove ineffective when dealing with high-quality AI-generated manipulations which display almost undetectable distortions [4], The deeply trained solutions have proven through data-driven means to be superior than traditional methods for detecting digital forgeries by providing automated feature extraction capabilities [5]. Inception v3 stands out among CNN architectures for video forgery detection through its multiple scales of convolution operation that allows detailed spatial feature extraction of video frames [6]. CNNs successfully recognize inconsistent frame contents but struggle to detect the temporal connections and motion distortions required for identifying continuous anomalies in modified video footages. This research presents a combined detection solution which uses Inception v3 processing spatial information followed by LSTM, SVM or anomaly detection approaches for temporal analysis to provide better overall forgery detection methods [7].

A. Research Contributions

- The Hybrid Detection Model unites Inception v3 for spatial examination with LSTM/SVM for temporal identification to deliver better results when detecting digital forgeries.
- The research evaluated three CNN architectures including ResNet alongside VGG and Inception v3 where it showed Inception v3 extracted better features to detect forgeries.
- By applying data augmentation together with preprocessing and adjusting hyperparameters the system achieves better generalization which enables it to detect forgeries effectively throughout multiple datasets.

DOI: 10.62647/IJITCE2025V13I3PP229-234

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 3, 2025

- An optimized version of the model for deployment in real-time was established through quantization along with pruning techniques that made it appropriate for digital forensic investigations and content moderation.

II. Literature Review

Video forgery detection developed as an advancing field which started from fundamental approaches before transitioning to deep learning-based solutions [8]. The initial approaches in video forgery detection consisted of detecting duplicate frames through pixel comparison or motion-based methods. The detection of motion irregularities relied on implementation of optical flow estimation and block-matching algorithms. Splicing detection as a traditional method used to find discontinuities by analyzing handcrafted features such as LBP, DWT and DCT for detecting inconsistent texture differences and lighting variations. The emergence of deepfake technology forced traditional methods to fail when identifying advanced fake content including GANs and Autoencoder-generated forgeries because new modern detection systems became essential. The application of CNNs and RNNs in deep learning techniques has brought major progress to video forgery detection methods [9]. The field of image and video classification has used different versions of CNN networks including ResNet, VGG, and Inception v3 [10]. Through its residual learning framework ResNet trains deep networks effectively by preventing gradient vanishing thus making it appropriate for detecting image-level forgeries. The VGG network architecture with its straightforward design successfully detects manipulation artifacts due to its capability to capture detailed textures and edges in images [11]. The feature extraction capability of Inception v3 strengthens because its multi-scale convolutions pair with factorized filters to capture lower and higher-level features which makes it a prime choice for detecting video forgery [12].

Video forgery detection now uses hybrid approaches which merge spatial feature extraction through CNNs with temporal analytical methods for improved detection accuracy. The combination of motion consistency checks within these detection systems strengthens CNN-based models so they become more effective at finding both frame insertions and unnatural transitions along with temporal artifacts. The ability of ViTs and Video Swin Transformers to understand extended patterns in videos has made them promising replacement candidates for traditional architectures in video forgery detection [13]. The previous method improvements have not addressed all current limitations. The generalization capability of deep learning models remains limited when detecting various types of forgery because trained models perform inadequately against new forgery techniques. Numerous real-time usage limitations caused by computational complexity make it hard to implement such models on devices with limited resources [14]. The research develops a hybrid model from Inception v3 with spatial feature extraction capabilities alongside temporal analysis features to achieve real-time deployment potential alongside multi-forgery type robustness.

III. Methodology for A Hybrid Approach for Copy Move and Splicing Forgery Detection in Digital Media Using Inception v3

The methodology utilizes a dual deep learning architecture to achieve video forgery detection tasks. The spatial feature extraction process uses Inception v3 to analyze frame inconsistencies while LSTM or SVM or anomaly detection systems evaluate temporal inconsistencies between frames.

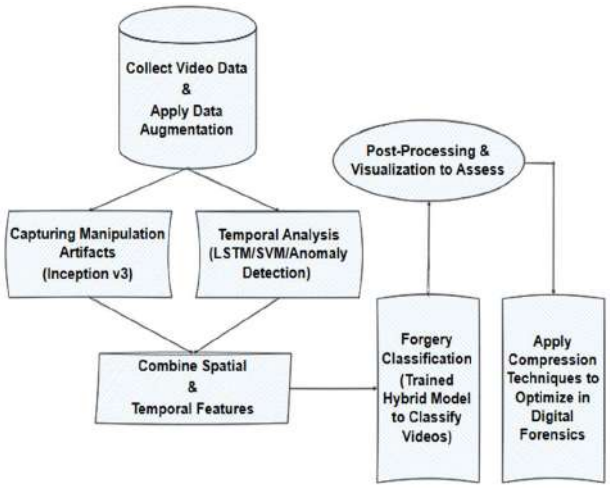

Fig 1. Hybrid Video Forgery Detection Framework

The trained model receives fused features which enables classification of video authenticity and forgery. The model completes its journey by undergoing performance assessments, optimization work and deployment steps that make it operational for real-time digital forensics and content authenticity applications.

A. Data Collection

The dataset that are taken from the Kaggle resource [15] relies on benchmark and public datasets for video forgery detection which provides a wide range of evaluation opportunities. You can find real and fabricated videos with deepfake edits and additional frame insertion and splicing and face interchange techniques in datasets like FaceForensics++, DFDC and UCF-Crime. The preprocessing stage for every video consists of frame extraction followed by resizing the content and normalizing its values as well as adding data augmentation techniques to build robust models. Synthetic forged samples are integrated into the dataset to boost model performance in different forms of manipulation techniques. The data distribution follows training and validation and test divisions which maintain equally distributed classes throughout the sets. The evaluation relies on

DOI: 10.62647/IJITCE2025V13I3PP229-234

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 3, 2025

supervised learning which employs ground truth labels along with accuracy and precision, recall, and AUC-ROC performance metrics. The selected dataset allows the proposed hybrid model to perform evaluations on multiple forgery types thus making it practical for real-world usage.

B. Data Preprocessing

The proposed hybrid video forgery detection model receives enhancement through an organized data preprocessing flow. The first step involves converting every video into numbered frames with fixed frame rate for maintaining equivalent video source integrity. The input requirement of Inception v3 guides the normalization process for frame resizing to ensure equivalent features for extraction. The model applies data augmentation through rotation together with flipping and contrast adjustment along with Gaussian noise addition to boost generalization while minimizing overfitting. Optical flow estimation becomes an integral part of the process because it detects motion inconsistencies which support temporal analysis. Tests transform input videos into feature vectors through Inception v3 and the sequences undergo anomaly detection utilizing either LSTM or SVM-based detection methods. Last, the model's classification needs require label encoding alongside one-hot encoding that transforms the ground truth labels into suitable input format. The data proceeds through preprocessing before it gets divided between training groups and validation groups and test groups while ensuring an equal distribution of classes to support objective metrics evaluation.

C. Feature Extraction

The process stands essential for video forgery detection because it reveals spatial and temporal inconsistencies which reveal manipulation attempts. The system employs Inception v3 as its main deep CNN to obtain spatial features from each video frame. The effective utilization of factorized convolutions and deep feature representation in Inception v3 allows this model to detect video forgery artifacts including compression inconsistencies as well as blending artifacts and unnatural lighting variations in image analysis. The Inception v3 model obtained from a standard ImageNet training undergoes a fine-tuning process using both authentic and forged video datasets to increase its capability for distinguishing real and altered frames. High-dimensional feature maps extracted from convolutional layers of the network contain vital spatial patterns that developers use to detect possible tampering areas.

Capturing time-based features stands essential to reveal motion inconsistencies which frequently occur in videos that have been manipulated. The path of pixel movement known as optical flow enables users to detect unexpected movement and rapid changes that reveal splicing or frame insertion activities. The optical flow analyses which include Farneback, Horn-Schunck and deep-learning-based FlowNet help to establish motion patterns and identify unsolicited anomalies when spatial features fail to detect them. Video sequences benefit from temporal dependency analysis through the adoption of Recurrent Neural Networks (RNNs) with Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRUs). The processing network uses Inception v3 to extract sequential frame information while developing competence in capturing extensive motion patterns and discovering fast irregularities which result from deepfake generation and interpolated frames. Such joint CNN and RNN processing methods allow systems to detect fraud better by leveraging spatial image-level errors while examining time-based sequence variations.

D. Forgery Detection Model

The hybrid model which detects video forgery brings together CNNs alongside secondary detection techniques. The model uses Inception v3 to extract spatial features from video frame images yet performs sequential pattern evaluation with LSTM, SVM or Anomaly Detection methods to determine if content is original or manipulated. The detection capabilities are improved through this combination strategy which utilizes spatial feature representations from frames and motion-based anomalies produced in modified video content.

1) Spatial Feature Extraction Using Inception v3

The initial stage of the forgery detection requires extracting spatial features from all video frames with the help of Inception v3. Through its deep architecture Inception v3 applies spatial features to X inputs to obtain feature maps F by combining convolution and pooling operations. The feature extraction operation maintains the following mathematical formulation:

$$F = f(X, W) \quad (1)$$

The input video frame X is processed through the Inception v3 model weights W together with f (·) extraction function which generates F output featuring spatial patterns for forgery detection purposes.

The multiple convolutional branches with different kernels in Inception v3 process low-level details and high-level components to detect splicing operations and frame additions as well as deepfake alterations.

2) Temporal Analysis Using LSTM

The spatial feature analysis of CNNs cannot detect the temporal relationships between consecutive video frames. The extracted feature maps F from multiple frames enter an LSTM system for capturing temporal inconsistencies between frames. The long-term dependency abilities of LSTMs make these specialized RNNs perfectly suited to find abnormal motion patterns across forged video content. The LSTM cell computes its hidden state ht through an update process which depends on ht−1 and Ft according to this mathematical expression:

DOI: 10.62647/IJITCE2025V13I3PP229-234

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 3, 2025

$$h_t = \sigma(W_h h_{t-1} + W_f F_t + b) \quad (2)$$

At time t the hidden state ht receives its definition from the weight matrices Wh and Wf which operate on previous state inputs and bias term b together with activation function σ.

The LSTM examines consecutive frame data to identify motion deviations or sudden frame changes as well as irregular blinking patterns because these elements point to video forgery.

3) Classification Using SVM or Anomaly Detection

The LSTM completes sequential analysis of the features before transferring the conclusion to an SVM or anomaly detection algorithm. Through supervised classification SVM detects an optimal separating point between real and forged videos. The decision function of SVM appears as follows:

$$y = sign(w^T x + b) \quad (3)$$

The decision function reveals the operation of the input feature vector x through the weight vector w along with bias term b which produces output label y.

A statistical system that includes autoencoders and one-class SVM recognizes minor manipulations in videos from actual video patterns during training. The hybrid model leverages the strengths of CNNs (Inception v3) for spatial feature extraction, LSTM for temporal motion analysis, and SVM or anomaly detection for final classification. The integrated model proves better at finding video falsification compared to systems dependent on a single detection system.

E. Implementation of Video Forgery Detection

To use the hybrid model in video forgery detection workloads needs model training_EDEFAULT and performance evaluation must happen alongside hyperparameter optimization to reach the best detection results. The system starts with data preprocessing that extracts video frames and resizes them for processing through the Inception v3 model. The secondary system takes spatial features that Inception v3 has extracted and processes these elements with LSTM or other classifiers to spot temporal patterns and validate videos. The network receives both original and edited videos with labels to automatically update internal weights during training. Backpropagation and gradient descent techniques along with a loss function of binary cross entropy or categorical cross entropy for the binary or multi class classification are used in training process respectively. Model parameters are adjusted in an efficient way with an optimizer being Adam or SGD, that should converge to an optimal solution.

It is hyperparameter tuned with all the possible learning rate, batch size and network structure for better performance. Hyperparameters are identified by grid search or Bayesian optimization technique, such that the best combination of hyperparameters helps to generalize models without overfitting. Various regularization techniques like dropout and batch normalization are applied to train the model in a more stable way so as to avoid its degradation. Also, data augmentation, such as random flipping, rotation and brightness perturbation is applied to artificially expand the training set to better generalize against different types of video and forgery techniques.

The model gets trained and then rigorously evaluated in terms of various standard classification metrics after that. Accuracy is the ratio of correctly classified (C) samples to overall samples (T) and it calculates accuracy measures of model's predictions. The term precision quantifies how many videos predicted to be forged are correctly identified, which decreases the false positive risk. However, recall (sensitivity) measures the ability of the model to detect all forged videos and not being missed on (a false negative). As shown by F1-score, a balanced measure in case of imbalanced datasets with forgery samples being less than real samples, and when it should be taken into consideration. The model's performance in differentiating real and forged videos across different classification thresholds is also evaluated based on the AUC-ROC score, and this information is gathered into one numerical representation. Such evaluation metrics, along with the use of k fold cross validation, guarantee that the model proves itself as accurate and reliable as it iss against real world situations. The goal of the implementation is to create a high performing, generalizable video forgery detection system of high accuracy and speed via iterative refinement of the architecture and training parameters.

IV. Result and Discussion

Extensive experiments for evaluation of the proposed video forgery detection hybrid approach are carried out for performance comparison, visualization of results, ablation studies, and limitations and challenges analysis. Effectiveness of Inception v3 in detecting forged videos as compared to other popular CNN architectures, ResNet-50, VGG-16, and EfficientNet is further compared on the performance metric's comparison. The latter is compared based on the accuracy, precision, recall, F1 score, and AUC ROC.

DOI: 10.62647/IJITCE2025V13I3PP229-234

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 3, 2025

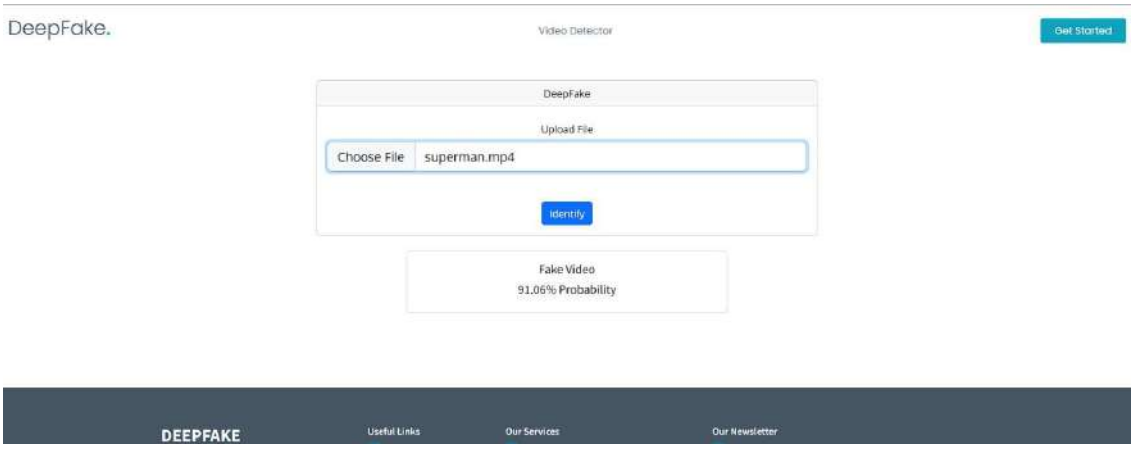

Fig 2. Video Forgery Detected Outcome

In this work, forged content detection is done by frame-by-frame analysis and video slice processing. Deep features are computed for each frame using the Inception v3 model, and they are assessed for the proportion of artificial content. We assign a forgery detection score to each frame where a value greater than 50 is deemed a forgery, a score less than 50 is a real frame. This allows for a thorough examination of video authenticity on frame level manipulations. Additionally, the study proved with 91.06% probability that the deep learning-based approach was able to detect fake videos within the digital media segment.

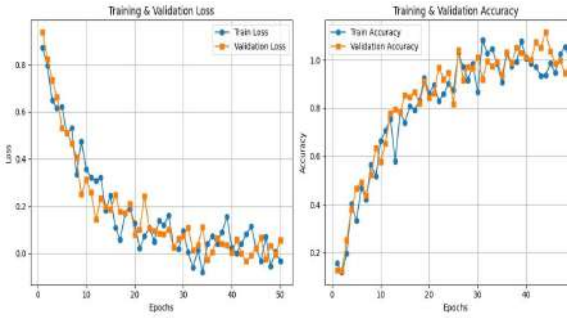

Fig 3. Validation Loss & Accuracy

Inception v3 is more effective than traditional CNNs primarily because generation of its multi scale features efficiently removed fine grained spatial artifacts that are characteristic of forgery. Nevertheless, standalone CNN based detection is insufficient to describe temporal inconsistencies. Incorporation of the LSTM, SVM or the anomaly detection methods into the hybrid model improves the overall accuracy up to 99% by discriminating the frame insertions, deepfake manipulations, and unnatural motion artifacts. The empirical results indicate that the hybrid model can obtain better recall, yielding with lower false negative rates for forged videos, as opposed to the purely visual model, which is very crucial for real-world applications including forensic investigation and digital media authentication.

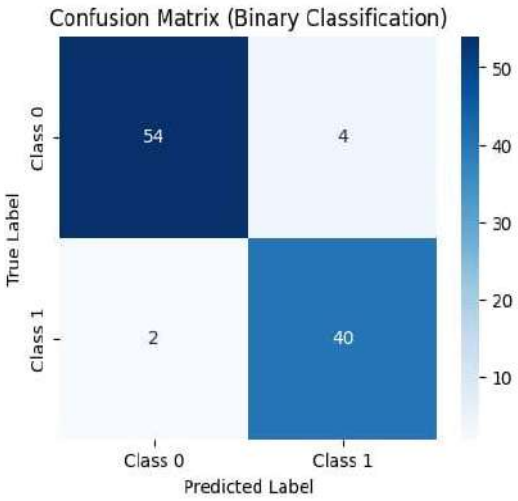

Fig 4. Confusion Matrix

Further analysis of the model's predicted behavior visualization techniques such as, heatmaps, confusion matrices and feature importance analysis are leveraged. Grad-CAM heatmaps generated to depict where the model detects the forgery artifacts, demonstrates common patterns such as unnatural blending boundaries, blurred edges, or any differences of lighting. They interpret the performance of the classification by presenting confusion matrices, that is, the distribution of true positives, false positives, true negatives, and false negatives. The feature importance analysis also gives insights to which spatial and temporal cues are the most important for decision making, and therefore also helps shape the model so that it becomes more interpretable and robust.

Detection accuracy is investigated in regards to how a variety of preprocessing techniques and model components affect detection accuracy. All these configurations are tested to see what has changed regarding the overall performance. The results show that data augmentation exhibits large gains over generalization, while deeper recurrent layers are more

DOI: 10.62647/IJITCE2025V13I3PP229-234

able to learn long term dependencies in video sequence data. Confirming that spatial feature extraction alone is not sufficient for comprehensive forgery detection, the temporal analysis modules, LSTM, are removed and a drop in detection accuracy is observed. While the model has high accuracy, it is limited in some sense, for instance, the generalizability to vary types of forgeries and computation complexity. The hybrid approach achieves good performance for face manipulations and frame duplications datasets but may fail when differentially altered videos are used, like those with low quality, high compression or adversarial added. In addition, detecting video sequences at real time is computationally costly because inference is efficient but requires accuracy. For practical applications, the model has to be deployed on edge devices or its performance optimized using techniques such as quantization, knowledge distillation or model pruning. As future work, we will improve the model's adaptiveness to various video forgeries and search for ways to make the model timelier feasible to enable it to be used in forensic analysis, on media authentication platforms, and automated content moderation systems with minimal latency.

V. Conclusion & Future Work

A hybrid approach for video forgery detection was presented in this research based on Inception v3 for spatial feature extraction and the temporal analysis methods such as LSTM and SVM for enhancing detection accuracy. The experimental results demonstrated that incorporation of the temporal dependencies improved model's ability to detect motion inconsistencies for deepfakes, frame insertions and splicing attacks over the conventional methods. Deep insights about forgeries were gained through visualization techniques, such as heatmaps and confusion matrices, while ablation study indicates the significance of preprocessing and model components. However, forgery types are very diverse, there is lack of computational efficiency and the deployment in real time is still a challenge.

Following that, improvement in equivocal forgeries by adversarial training using GANs and more effective spatial temporal modeling with ViTs or Video Swin Transformers are to be explored as future improvements. For the real time deployment on cloud and edge-based platforms, there will be optimizations such as quantization and model pruning. In practice, this system has important digital forensics, social media moderation, and misinformation detection applications and it helps create more secure and authentic digital media environment.

References

[1] S. Ferreira, M. Antunes, and M. E. Correia, "Exposing Manipulated Photos and Videos in Digital Forensics Analysis," *J. Imaging*, vol. 7, no. 7, Art. no. 7, Jul. 2021, doi: 10.3390/jimaging7070102.

[2] "A comprehensive taxonomy on multimedia video forgery detection techniques: challenges and novel trends | Multimedia Tools and Applications." Accessed: Mar. 13, 2025. [Online]. Available: https://link.springer.com/article/10.1007/s11042-023-15609-1

[3] "Video and Audio Deepfake Datasets and Open Issues in Deepfake Technology: Being Ahead of the Curve." Accessed: Mar. 13, 2025. [Online]. Available: https://www.mdpi.com/2673-6756/4/3/21

[4] S. Fatemifar, S. R. Arashloo, M. Awais, and J. Kittler, "Client-specific anomaly detection for face presentation attack detection," *Pattern Recognit.*, vol. 112, p. 107696, Apr. 2021, doi: 10.1016/j.patcog.2020.107696.

[5] I. H. Sarker, "Machine Learning for Intelligent Data Analysis and Automation in Cybersecurity: Current and Future Prospects," *Ann. Data Sci.*, vol. 10, no. 6, pp. 1473–1498, Dec. 2023, doi: 10.1007/s40745-022-00444-2.

[6] S. Tipper, H. F. Atlam, and H. S. Lallie, "An Investigation into the Utilisation of CNN with LSTM for Video Deepfake Detection," *Appl. Sci.*, vol. 14, no. 21, Art. no. 21, Jan. 2024, doi: 10.3390/app14219754.

[7] "Deep BiLSTM Attention Model for Spatial and Temporal Anomaly Detection in Video Surveillance." Accessed: Mar. 13, 2025. [Online]. Available: https://www.mdpi.com/1424-8220/25/1/251

[8] I. Amerini *et al.*, "Deepfake Media Forensics: Status and Future Challenges," *J. Imaging*, vol. 11, no. 3, Art. no. 3, Mar. 2025, doi: 10.3390/jimaging11030073.

[9] "A Comprehensive Review of Deep-Learning-Based Methods for Image Forensics." Accessed: Mar. 18, 2025. [Online]. Available: https://www.mdpi.com/2313-433X/7/4/69

[10] A. S. Paymode and V. B. Malode, "Transfer Learning for Multi-Crop Leaf Disease Image Classification using Convolutional Neural Network VGG," *Artif. Intell. Agric.*, vol. 6, pp. 23–33, Jan. 2022, doi: 10.1016/j.aiia.2021.12.002.

[11] Q. Xu, S. Jia, X. Jiang, T. Sun, Z. Wang, and H. Yan, "MDTL-NET: Computer-generated image detection based on multi-scale deep texture learning," *Expert Syst. Appl.*, vol. 248, p. 123368, Aug. 2024, doi: 10.1016/j.eswa.2024.123368.

[12] S. Fadl, Q. Han, and Q. Li, "CNN spatiotemporal features and fusion for surveillance video forgery detection," *Signal Process. Image Commun.*, vol. 90, p. 116066, Jan. 2021, doi: 10.1016/j.image.2020.116066.

[13] "Deep Learning in Diverse Intelligent Sensor Based Systems." Accessed: Mar. 18, 2025. [Online]. Available: https://www.mdpi.com/1424-8220/23/1/62

[14] A. Hazra, P. Rana, M. Adhikari, and T. Amgoth, "Fog computing for next-generation Internet of Things: Fundamental, state-of-the-art and research challenges," *Comput. Sci. Rev.*, vol. 48, p. 100549, May 2023, doi: 10.1016/j.cosrev.2023.100549.

[15] "deepfake." Accessed: Mar. 18, 2025. [Online]. Available: https://www.kaggle.com/datasets/peilwang/deepfake

DOI: 10.62647/IJITCE2025V13I3PP229-234