iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IV **Month of publication:** April 2026

DOI: https://doi.org/10.22214/ijraset.2026.81476

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Deep Fake Detection Using CNN

Mayank Tamta, Lakshay Bhardwaj, Khushi Yadav, Janhavi Nehra, Sristhi Vashisth
Department of Computer Science and Engineering, MIET, Meerut, Uttar Pradesh, India

Abstract: ***Recent progress in convolutional neural networks has enabled automated systems to examine manipulated visual content with improved reliability. Deepfake media, generated using advanced learning techniques, poses a serious challenge to digital authenticity and public trust. This survey reviews recent deepfake generation approaches and CNN-based detection methods, focusing on commonly used datasets, model architectures, and evaluation metrics. The study highlights challenges related to generalization computational efficiency, and real-world deployment, providing insights that can support future research and practical applications in deepfake detection.***
Keywords: DeepLearning,FakeDetection,NeuralNetworks,Social Networks.

I. INTRODUCTION

Recent advances in deep learning have shown promising results in detecting such manipulations. Convolutional neural networks, in particular, are capable of identifying subtle visual inconsistencies introduced during the deepfake generation process. This paper surveys CNN-based deepfake detection techniques, emphasizing widely used datasets, model architectures such as Xception and MobileNet,and evaluation strategies that enhance detection reliability.

The rapid growth of digital media sharing platforms has increased the circulation of manipulated images and videos across the internet. Deepfaketechnologyenablesthecreationofhighly realistic synthetic media that can mislead viewers and compromise information credibility. As these manipulations become more sophisticated,ensuring the authenticity of digital content has become a critical challenge in multimedia security and digital forensics and leveraging advanced algorithms such as convolutional neural network [7]

II. LITERATURTE REVIEW

S.No	Year	Reference	CNNArchitecture	KeyFindings	Howithelpsyour project
1	2018	**MesoNet—Afchar et al.** *MesoNet: a Compact Facial Video Forgery DetectionNetwork*	Compact CNN ("Meso-4", "Meso-Inception-4") focusing on mesoscopic facial artifacts	Shallow CNNs detect blurriness and texture anomalies effectively with low computation.	Provides a lightweight baselinemodelforreal-time or small-scale deepfake detection.
2	2019	**FaceForensics++—Rösslеretal.** *FaceForensics++: Learning to Detect ManipulatedFacial Images*	EvaluatedmultipleCNNs; Xception-basedmodels achievedbestperformance	Demonstrated Xception'sstrong generalization on facialmanipulation tasks.	UseFaceForensics++for standardized training/testing and benchmarkingCNN performance.
3	2019	**XceptionNet for DeepfakeDetection** (multiple studies)	Xception (depthwise separableconvolutions) fine-tuned for binary classification	Experimentalevaluations across multiple studies indicate that Xception-based architectures	Employ Xception as a high-performing backboneforyourCNN model.
4	2020	**Celeb-DF — Li et al.** *Celeb-DF: A Large-scale ChallengingDataset for DeepFakeForensics*	CNN-based detectors (Xception, ResNet) testedforcross-dataset performance	Highlighted the challenge of realistic deepfakes and cross-datasetgeneralization.	Test your trained CNN on Celeb-DFtoevaluatereal-world robustness.

©IJRASET: All Rights are Reserved

5	2020	**DFDCDataset—Dolhanskyetal.** *DeepFake Detection ChallengeDataset*	EnsembleCNNs (Xception,EfficientNet, ResNet)	Large-scale benchmark enabling highgeneralizationfor deepfakedetection	UseDFDCfor fine-tuningorlarge-scale testingforimproved accuracy.
6	2021	Generalization& Bias Study — Mehra et al.	Generalization&Bias Study — Mehra et al.	FoundCNNmodels struggle with generalization and demographic bias	Helps in evaluating fairness and cross-domainaccuracyofyour CNN model
7	2024	**RecentSurveys/ Evaluations**	Reviews of CNN + multimodalapproaches	CNNsremaincore detectors; hybrid temporal/audio models improve performance.	Guides dataset choice, architecture selection, andperformancemetrics for your project.

III. METHODOLOGY

A. Data Collection andPreprocessing

The deepfake detection process follows a structuredpipelinedesignedtoanalyzeboth authentic and manipulated video content. Adiversesetofrealandforgedvideo samples is selected to improve the model's ability to handle variations in facial appearance and recording conditions. [3]

This dataset should includebothauthentic and manipulated videos, incorporating a varioustypesofconditions,suchas indoor,outdoor environments, varying lighting conditions, diverse facialexpressions, and different camera qualities. Ensuring diversity in the datasethelpsmodelsgeneralizebetter,enablingthemto effectively identify deepfakesacrossvariedreal-world situations and mitigate potential biases that couldcompromisethe reliability of the detection system. Labels mustaccurately differencebetweenrealanddeepfake videos.

Themainaimoftheworkwastodetermine if videos were genuine or if they were created through [2] deepfake technologies. Therefore,itisapparentthatforthissystem towork,theinputshouldsolelyconsistofa video clip. However, based on the fact that input for deep learning models is images, the conversion from video input to model inputshouldbeconducted.Thisisachieved throughapreprocessingmodule.Thevideos may have more elements than just a face. Each frame of the video is not restricted to the individual's face only, since most of the video frame consists of the person's body parts as well as the background area of the picture.Besides,theseuncorrelatedfeatures may hinder the model's training using face wrapping arttifacts. [8]The face area is the focal point, while the image gets input as facial data from the image by the pre- processing module. This pre-processing module further consists of three discrete stages: grabbing frames from a video feed, detectingfacesinsuchframesoneatatime, and storing these areas where there are faces in the form of pictures. Each of these procedures is expounded on here below.

We first started capturing the video input into frames. Each video is converted into a sequenceofimageframesusingOpenCVto enable image-based processing suitable for CNN models. Since deepfake manipulation mainlyaffectsfacialregions,facedetectionis applied to isolate facial areas from each frame,reducingtheinfluenceofbackground information on model training.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

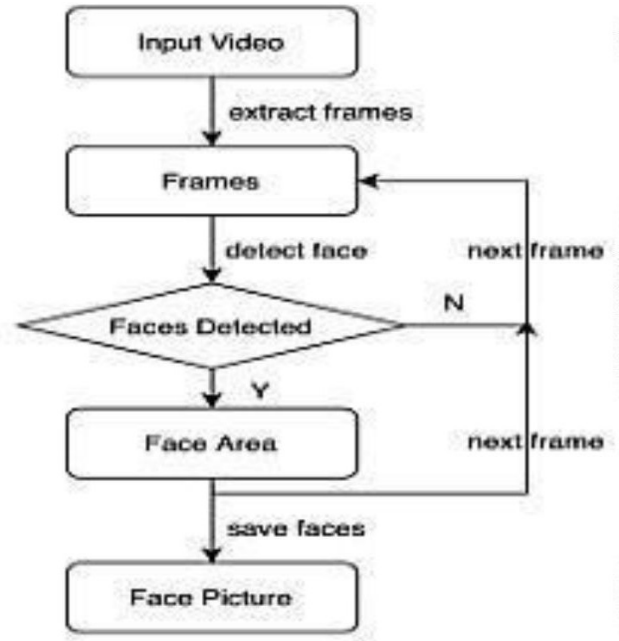


Fig3.1.FlowChartof DetectingDeep Fake

Inaddition,consideringbotharesimilar, puttingalloftheminatrainingsetwill maketrainingineffectiveduetotheirhigh similarity, and this may also result in other problems, including overtraining. The videos used here had 30 frames per second for each clip selected in this study. Upon assessment, it became apparent that one option has to choose an image interval of four.

The second step was to detect faces thatareavailableinanimageand automatically tag them. Facial regionsaredetectedusingapre-trainedHaar-basedclassifier,which enables efficient localization of faces within each frame. To avoid incorrect detections,thelargestdetectedfacial region is selected for further processing,ensuringaccuratefacial representation. Classifie was then selected out of a number of classifiers that weretesteddueto itsprecisereawhereafacecanbe found, which means it createsthebest area concerning face localization. However, there were a fewnon-face selectionsmade.Forinstance, after conductingsome experiments .

The third step was to store the detected face area as a new image. Prior to storage, all facial photos had to be uniformly resized. While using the Xception model, the size of the picture was supposed to be 299*299, whereas with MobileNet it is 224*224.

B. Model Selection

Currently, there exist various deep learning models and frameworks. In this paper, Xception and MobileNet were selected as the models due to the reasons below. To begin with, Xception has high performance givenitsperformancebenchmarkingatthe FaceForensics testing environment.

Researchers for example can use FaceForensics as a benchmarking platform totesttheirmodels.Asitcomeswitha detailedmanual,Xceptionisbetterthan otherson4differentdatasetsondifferent reasearches[6]

Xception has a similar architecture to MobileNet, which makes it the reason why itwasselected.Thetwo modelsarebuilton CNN with the use of both depthwise and pointwise convolutional layers. Unlike MobileNets, which has fewer features in order to make the model more efficient.

C. Xception for Deepfake Detection

Xception is anarchitecture for a convolutionalneuralnetworkthatisfamous foritshandlingofcomplicatedmodels without sacrificing resources. It acts like a useful tool that can notice tiny things, so it becomes easy to figure out whether anything has been tampered with inavideo ornot.Usingtransfertrainingtechniques makes it possible to make Xception better usingcustomizedaspectsofthedeepfake detection data, makingit perfect at distinguishingbetweengenuinepictures and fake ones.

©IJRASET: All Rights are Reserved

D. MobileNet for Deepfake Detection

MobileNet is optimized for computationsavinguses,henceitslightness. MobileNet can still detect deep fakes despite being small. One way this can be done is to use MobileNet on deepfake detection datasets for later tuning in real- time or for edgebased systems due to their limitations in computations

E. Feature Extraction

Particularly in videos, feature extraction is key because of the significant role that temporal information plays. It is very important to extract features that capture motion patterns,consistency overtime, and spatialrelationshipsbetweenframesfor effective deepfake detection. After extracting the features, the model is trained over the preprocessed data set.

F. Model Training

Model training includes dividing the preprocessed dataset into training, validation, and test subsets. This process often employs transfer learning, where the model is initialized using weights from a pre-trained network and subsequently fine- tuned for the specific task of deepfake detection.

G. ValidationandEvaluation

While still in the training process, it's important to see how the model performs using the validation set in order to avoid overfitting. Tune the hyperparameters (learning rate, batch_size and optimizer) such that the model performs best. Check how successful your model is under differentconditionswhentestedusingtester sets. These can be accuracy (measure of correct predictions), precision (how much usefulinformationyougetfromwhatyou're looking at), recall (how many people who shouldhavehadsomethingdoneknewabout it),F1-score(balancebetweenprecisionand recall), and the ROC curve. The ROC and IEEE does not plot the accuracy or error of a binary classifier but displays the separation between classes.[14]

It illustrates the connection between sensitivity and specificity. Evaluate the trainedmodelintermsoftesting;evaluateit according to its classification of class instancesintopositiveandnegativeclasses. Anomaly refers to the feature space of instances that are unusual. The separated classes are linearly separable.

Productiondescribestasksthatarebroken down into smaller components and then arranged in a sequence. Machine learning departsfrom traditionalstatisticalmethods.

IV. RESEARCH GAPS

1) Scope and Purpose: The literaturereviewsystematicallyexamines112studies on deepfake detection from 2018 to 2020. The aim is to classify and evaluate various deepfake detection methods, tools, and datasets.
2) Deepfake Techniques: Research is categorized into four main detection techniques
3) Deep Learning: The most widely used and effective, accounting for 77% of studies.
4) Machine Learning: Focuses on traditional techniques like decision trees and support vector machines.- Statistical Methods: Utilize statistical patterns for deepfake detection.
5) Blockchain: Provides decentralized verification of authenticity, though less common.
6) Datasets: Popular datasets include FaceForensics++, Celeb-DF, and DFDC, widely used for training and testing.
7) Evaluation: The studies predominantlyuse accuracy and AUC (area under the curve) for performance measurement, with deep learning methods consistently outperforming others in accuracy and robustness.
8) Detection: The paper highlights for standardized evaluation frameworks to ensurecomparabilityandfairnessamong variousdetectionmodels.
9) GeneralizationAcrossDatasets: Many deep fake detection models perform wellon specific datasets but struggle with unseen data or deep fakes generated by different methods. Research is needed to develop models that generalize effectively across diverse deep fake types and sources.
10) Real-Time Detection: Current detection methods often lack the efficiency needed for real-time analysis, which is crucial for applications like social media monitoring. Exploring lightweight and fast deeplearningmodelscouldbridgethisgap.
11) Robustness Against Advanced **Techniques:** As deep fake generation methods evolve, detection models need to adapt to new forms of manipulation. Researchintoadversarialtrainingandmore adaptable architectures can make models more resilient against evolving deep fake techniques.
12) Computational Efficiency: Deeplearning-based detection methods typically require high computational resources, limiting their

practical deployment, especially on edge devices. There's a need for models that balance performance with computational efficiency, potentially throughmodelcompressionoroptimization techniques.

13) Explainability:Deeplearningmodels, especially CNNs, making it difficult the decision-making processes. Future work could focus on making these models more interpretable, aiding trust and transparency in automated detection.
14) Privacy Ethical Considerations: Implementing deep fake detection involves analyzing potentially sensitive user data. Research is needed to ensure detection systems respect privacy and ethical considerations, such as minimizing data collection or ensuring user consent.
15) AttackResistance:Deeplearningmodels may be susceptible to adversarial attack,whereminoralterationsininputdata can fool the model. Exploring defense mechanisms to harden models against such attacks is essential for reliable detection.
16) Forward Progression: To from a given dataset, while the calculate the output values of a neural network,theinputdatamustpassthrough the network sequentially, progressing from one layer to the next, starting from the input layer and ending at the output layer. This step-by-step process of transmitting inputs through the network togenerateoutputsisreferredtoasforward propagation.
17) Backpropogation:Eachhiddenlayer in a neural network receives inputs fromthe previous layer, processes them through an activation function, and then generates the predicted output values. Backpropagation is a reverse process that focuses on adjusting the weights tooptimize the network's ability to accurately predict outputs. This process involves calculating the gradient of the error and using stochastic gradient descent (SGD) to minimize the cost function, thereby reducing prediction errors and improving the model's performance.
18) FaceManipulation:particularlyin face manipulation, pose a significant threat bydistortingtheoriginalfactsindigital images.As technology continues to evolve, ithasbecomeincreasinglyimportanttoemployitverifytheoriginalityofvideos and identify manipulated information. Thesedetectionmethodsarecrucialin ensuringthecredibilityandintegrityofvisual media in today's digital landscape. Deeplearningtechniques,suchasGAN, are based on the principles of autoencoders and decoders to aid in identifying fake images or videos. These networks leverage the generator and discriminator framework to identify and differentiate between real and manipulated content.
19) **GAN:**GenerativeAdversarialNetworks are made up of two neural networks:ageneratorandadiscriminator. Thegeneratorproducessyntheticimages discriminator analyzes these images to determine theirauthenticity differentiating between real and fake videos
20) Media Filter : Amediafilteristhen used to remove unwanted noise from the video. To improve image quality, bicubic interpolation is applied to increase pixel density,followed bybicubictransformation to enhance overall clarity. Yadav and Salmani (2019) discussed the principles behind deepfake techniques, including face image swapping, with an emphasis on achieving high precision (Maheswaran et al., 2018).
21) DBNmodel : The existing approaches face several challenges, including inefficiencies in detecting deepfake images, high error rates, long processing times, and inaccuracies in data assessment. The FF-LBPH-DBN model specifically addresses these issues by focusingonminimizingcomputationalcosts and applying various metrological parameters in a more efficient manner.

V. ENHANCEMENT

As deepfake technology continues to advance, it increasingly threatens the authenticityandreliabilityofdigitalmedia. This growing concern presents numerous challenges in the deepfake detection, particularly in the application of deep learning, where ongoing development and advancements are necessary to keep pace with evolving manipulation techniques. One area of focus should be on improving theinterpretationandinterpretationofdeep learning models for research. Building descriptive artificial intelligence (XAI) models that provide information about the featuresandpatternsusedfordetectioncan increase transparency and build trust in these systems.[4] IEEE (Institute of Electrical and Electriconics engineers) makes techniques like visualization, feature mapping, and model interpretation can provide insight into the innerworkings of deep learning models, enhancing the understanding and application of these processes. Anotherimportantareaforimprovementis thequalityofthemodelandtheavailability of training materials. The effectiveness of deepfake detection models depends significantly on the quality and accessibility of training data. Researchers should focus on creating and expanding a good data repository with a variety of in- depth measurements, including the most accurate control methods. Collaborating with media organizations, government agencies, and other stakeholders can help collect and manage this data, ensuring detection models are trained on the most important and cutting-edge data. For more complex business processes, the detection model needs to be updated and modified accordingly. Researchers should focus on developing flexible and adaptive deep learning models capable of quickly detecting and addressing emerging deepfake threats.

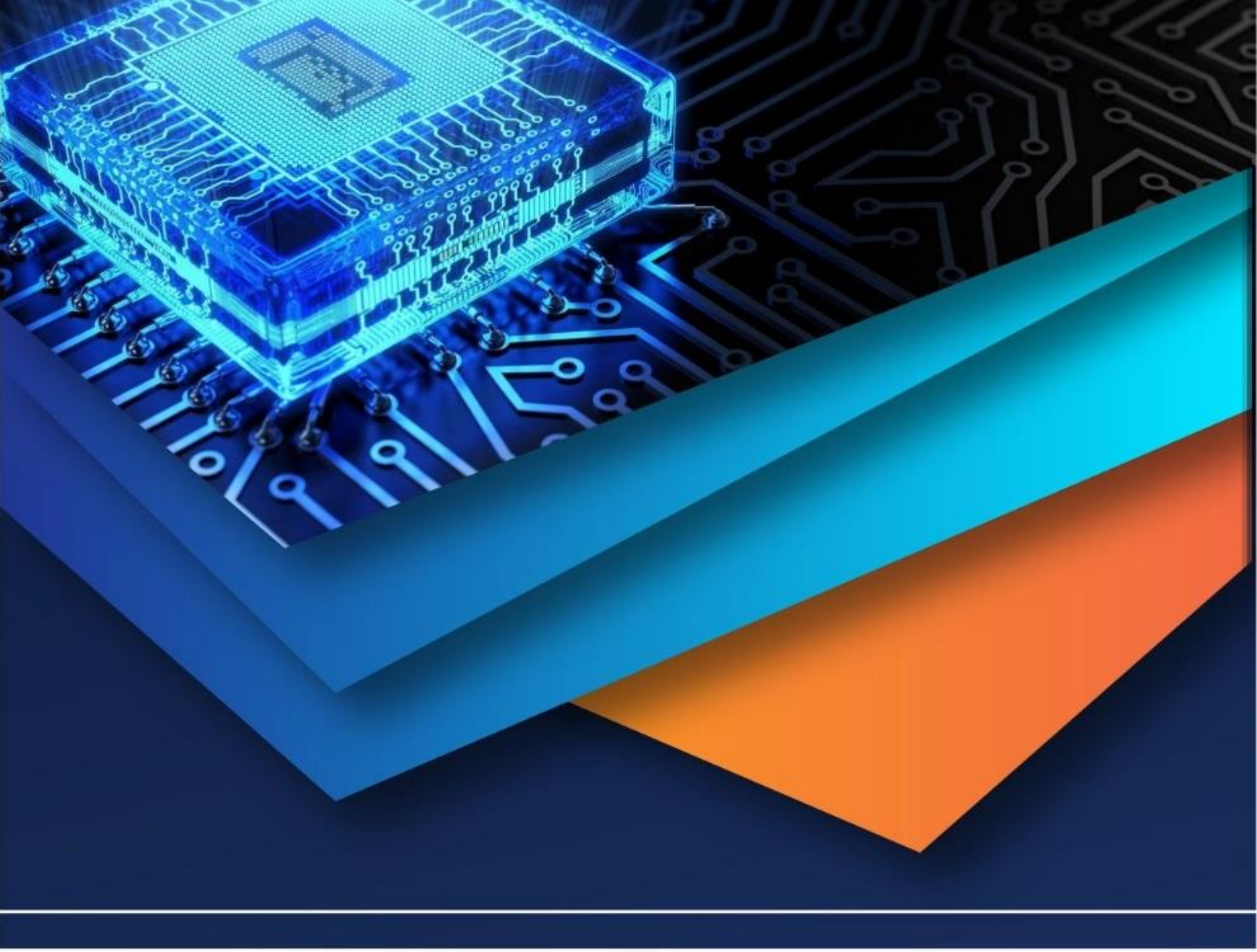
This will include techniques such as resilience training, adaptivelearning, and continuous learning thatcan help findmodels that are ahead of the norm and control their impact on deep tech exchanges. Various formats, such as image, audio, and text files, Researchers should focus on developing more efficient deep learning architectures and using data from different sources to increase discovery accuracy. Additionally, investigating common communication techniques can reveal effective patterns acrossdifferentmediaandtypesofleaders, thereby improving their effectiveness and efficiency in the actual world. More importantly, it is very much important to use these systems at scale and efficiently. Researchers should investigate strategiesto optimize deep learning models, such as model compression, quantization, and the hardware acceleration, to provide insight into deep learning across resource constraints such as edge devices and exposing deepfake videos through face warping artifact detection [15] To ensure deep solutions meet real-world needs and requirements, researchers need to develop partnerships with experts and industry stakeholders.This collaborationcancreate a framework and guidance for the responsible use and application of deep research tools, ultimately increasing their impactandsocialbenefits.Wecancontinueto develop and deliver better, more transparent, and more flexible solutions to the growing threat of media and informationmisuseinthisnewdigitalage.

VI. CONCLUSION

Inconclusion,theproliferationofdeepfakes presents quite significant challenge, exacerbatedbytheaccessibility oftoolsfor creating and distributing fake images and videos on social media platforms. Deep learning methods have emerged as a promisingsolutionfordetectingdeepfakes, with various techniques developed for image and video detection. This paper provides a survey of current applications andtoolsformakingdeepfakesanddetailed scrutiny for deepfake detection methods based on images and videos. We discussed their architectures, tools, and performance andhighlightedpubliclyaccessibledatasets used for training and evaluation. Although great strides have been recorded in the detection of deepfakes through deep learning,muchremainstobedoneas faras improving the quality of these gifs is concerned. This makes it hard for the current techniques because they can never bewithoutchallengedue totheprogressive nature of their construction. This is why moreresearch needs to focus on enhancing deep fake detection algorithm performance with respect to identifying model architecturesthatworkbestoralternatively deploying such architectures within social networking sites so that they can help dampen the effects emanating from these deepfakes. Moving forward, addressing these challenges and advancing research in these deepfake detection will be crucial to safeguarding against the harmful effects of altered media and preserving trust and integrity in digital content.

REFERNCES

[1] S. Senhadji and R. A. San Ahmed presentedamethodfordetectingfakenews by utilizing Naïve Bayes and Long Short- Term Memory (LSTM) algorithms in the IAES International Journal of Artificial Intelligence,Volume11,Issue2,2022,pages 748-754.
[2] K. N. Ramadhani and R. Munir conducted a comparative study on various methodsfordeepfakevideodetection, International Conference on Information and Communications Technology, November2020,pages394-399.
[3] D.Pan,L.Sun,R.Wang,X.Zhang, and R. O. Sinnott explored deepfake detection through deep learning techniques at the IEEE/ACM International Conference on Big Data Computing, Applications, and Technologies,December2020,pages134-
[4] 143.
[5] A.A. Maksutov, V. O. Morozov, A. A. Lavrenov, and A. S. Smirnov discussed machine learning methods for detecting deepfakes at the IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, January 2020, pages408-411.
[6] T.T. Nguyen, Q. V. Nguyen, and D. T. Nguyen provided a comprehensive survey on deep learning methods for the creation and detection of deepfakes, published in ComputerVisionandImageUnderstanding, Volume223,2022,pages1-19.
[7] A.O.KwokandS.G.Kohexaminedthe social construction of technology perspectiveindeepfakeresearch,published in Current Issues in Tourism, Volume 24, Issue13,2020,pages1798-1802.
[8] M.Westerlundreviewedtheemergence of deepfake technology in the Technology InnovationManagementReview,Volume9, Issue11,2019,pages40-53.
[9] Y.LiandS.Lyupresentedamethodfor exposingdeepfakevideosbydetectingface warpingartifacts, arXiv:1811.00656,2018.
[10] A. M. Almars reviewed deepfake detection techniques using deep learning in the Journal of Computer and Communications,Volume9,Issue5,2021, pages 20-35.
[11] N. S. Ivanov, Arzhakova, and V. G. Ivanenkoexploredthecombinationofdeep learningandsuper-resolutionalgorithmsfor deepfake detection at the IEEE Conference ofRussianYoungResearchersinElectrical andElectronicEngineering,January2020, publishedintheproceedingsofthe3rd pages 326-328.
[12] K.Zhu,B.Wu,andB.Wangproposed a deepfake detection method using clustering-based embedding regularization attheIEEE5thInternationalConferenceon Data Science in Cyberspace, July 2020, pages 257-264.
[13] F. Matern, C. Riess, and M. Stamminger exploited visual artifacts to expose deepfakes and face manipulations, presented at the IEEE Winter Applications of Computer Vision Workshops, January 2019, pages 83-92.
[14] E. Sabir, J. Cheng, A. Jaiswal, W. AbdElmageed, I. Masi, and P. Natarajan introduced recurrent convolutional strategiesfordetectingfacemanipulationin videosinInterfaces(GUI),Volume3,Issue 1, 2019, pages 80-87.
[15] D. Güera and E. J. Delp employed recurrent neural networks for deepfake video detection at the 15th IEEE International Conference on Advanced Video and Signal-Based Surveillance, November2018,pages1-6.
[16] Y. Li and S. Lyu discussed exposing deepfake videos through face warping artifactdetection, arXiv:1811.00656, 2018.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)