DOI: 10.6977/IJoSI.202403_8(1).0005
A. Rajeev, Raviraj P./Int. J. Systematic Innovation, 8(1), 49-62 (2024)

Performance evaluation of deep learning models for detecting deep fakes

Aishwarya Rajeev[1,2]*, Raviraj P[3]

[1]Research Scholar, Geetha Shishu Shikshana Sangha Institute of Engineering & Technology for Women, Mysuru, Karnataka, Affiliated to VTU Belagavi, 570016.

[2]Department of CSE, Coorg Institute of Technology, Ponnampet, Karnataka, Affiliated to VTU Belagavi, 571216.

[3]Professor, Geetha Shishu Shikshana Sangha Institute of Engineering & Technology for Women, Mysuru, Karnataka, Affiliated to VTU Belagavi, 570016.

*Corresponding author mail: aishwaryarajeev@gmail.com

(Received 23 October 2023; Final version received 04 December 2023; Accepted 27 December 2023)

Abstract

The proliferation of deep fake content in multimedia has necessitated the development of robust detection mechanisms. In this study, a comparative analysis of four state-of-the-art deep learning models for detecting deep fakes is conducted: CNN+RNN, DAFDN, Hybrid Inception ResNet v2, and Xception. The evaluation focuses on their performance metrics, emphasizing accuracy as a primary measure. Through extensive experimentation and evaluation on a comprehensive dataset, the findings reveal notable distinctions among these models. The CNN+RNN architecture demonstrates a commendable accuracy of 94.8%, providing a solid baseline for comparison. Surpassing this, the DAFDN model achieves an accuracy of 97.8%, showcasing superior discriminatory capabilities in identifying manipulated content. Furthermore, the CNN model stands out with an accuracy of 98%, exhibiting remarkable effectiveness in distinguishing between genuine and deep fake media. The comparative analysis delves into the strengths and weaknesses of each model, shedding light on their respective performance levels in detecting sophisticated deep fake content. The observed accuracies underscore the nuanced differences in their architectures and training methodologies, offering insights crucial for selecting appropriate models based on specific detection requirements.

Keywords: Face Forensics, Convolutional neural network, recurrent neural network, DAFDN, Resnet v2, Xceptio

1. Introduction

Deep fake films are modified videos that use machine learning-based algorithms to swap out humans for other objects or actors in an existing image or video. Three categories of deep fake videos exist: lip-syncing, face swapping, and head puppetry. The art of head puppetry involves using a source video person's head to manipulate a video of a specific human's head and upper shoulder so that the altered person looks exactly like the source (Shad et.al.,2021). Face swapping is changing a person's face to that of another while keeping the same expression on their face. Since lip-syncing merely modifies the lip area of a video, the target person says something that isn't actually true. Although some deep fakes can be produced using classic visualization techniques or computer graphics, the most recent and widely used deep learning techniques for producing deep fake videos are auto encoders and generative adversarial networks (GAN) (Rahman et.al.,2022).

These models are used to synthesize face images of people with comparable expressions and movements based on the analysis of a person's facial emotions and movement. For deep fake technologies to train a model to create photorealistic photos and movies, a significant amount of image and video data sets are typically required. Politicians and celebrities are the first targets of profound fakes due to the sheer volume of their films and photographs that are readily available online (Nguyen et.al.,2019) In pornographic pictures and films, the heads of famous people and political figures have been replaced with deep fakes. In the first deep fake movie, a celebrity's visage was changed to that of a porn, it was released in 2017. Deep fake movies are an increasing concern to global security since they are increasingly being used to produce false speeches by world leaders (Bode 2021).

In response to this challenge, researchers have explored various deep learning architectures to enhance detection accuracy. Comparative analyses have been conducted to assess the efficacy of different models, including CNN+RNN, DAFDN (Deep Adaptive Feature Distillation Network), Hybrid Inception ResNet v2, and Xception. Each architecture brings its unique strengths in identifying subtle discrepancies and patterns within manipulated content, striving to outperform adversaries' deep fake generation techniques Verdoliva 2020). The CNN+RNN model combines Convolutional Neural Networks (CNN) for feature extraction with Recurrent Neural Networks (RNN) for temporal information processing, offering a comprehensive approach to capture both spatial and sequential patterns in videos, a common format for deep fakes (Kousik et.al.,2021). DAFDN utilizes adaptive feature distillation to distill informative features and mitigate the domain gap between real and fake videos, enhancing detection accuracy.

Meanwhile, the Hybrid Inception ResNet v2 and Xception architectures leverage the power of inception modules and efficient convolutional operations, respectively, to improve feature extraction and model robustness against increasingly sophisticated manipulations (Kamaleldin et.al.,2023). However, amidst these advancements, a notable research gap persists. Despite significant progress in deep fake detection, the adaptability of detection models to new and evolving manipulation techniques remains a challenge (Guo et.al.,2021). The rapid evolution of deep fake generation methods continuously outpaces the development of detection algorithms, leading to a need for models that can generalize across diverse types of manipulations and adapt swiftly to emerging fake media tactics (George et.al.,2023, Wang et.al.,2023).

In conclusion, while various deep learning architectures have shown promise in detecting deep fakes, the dynamic landscape of fake media creation demands continuous innovation and adaptation in detection models to effectively address the ever-evolving challenges posed by deep fakes. Closing the research gap by creating more adaptable and robust detection mechanisms stands as a critical next step in the ongoing battle against misinformation and deceptive media.

The paper is organized in a systematic manner. It begins with a thorough introduction and then quickly reviews the body of research that has already been done on face forensics. The paper's main body goes into a thorough comparison study of several false face forensics designs. The salient features of the comparison are then highlighted by a summary. The article culminates with a thorough synopsis that synthesizes the knowledge and understanding acquired during the investigation.

2. Literature survey

A literature survey on deep fake detection systems encompasses a comprehensive exploration of existing research, methodologies, and advancements in the field. This survey delves into the diverse array of approaches employed to detect and mitigate the proliferation of manipulated multimedia content, specifically focusing on deep fake videos. It encompasses an analysis of various techniques, such as machine learning algorithms, neural networks, forensic analysis, and other innovative methodologies utilized to identify and combat the rising sophistication of deep fake technology. The survey aims to synthesize the current state-of-the-art methodologies, highlight their strengths and limitations, and identify potential avenues for further research and enhancement in the realm of deep fake detection systems.

(Feng Ding et al 2020) The author created a creative framework as a digital forensics tool to protect end users. It is built on deep learning and uses categorization to find assaults. The suggested model's data collecting effectiveness, resilience, and detection performance are all improved when compared to the traditional approaches, which are supported by our experiments. Additionally, our suggested approach makes use of 5G HetNets to allow high-quality real-time forensics services on edge consumer devices (ECE), such as smartphones and laptops, which has significant practical implications. Additionally, certain conversations are held to describe potential risks in the future.

(Nickson M. Karie et al 2019) The DLCF Framework, developed as a result of this research, provides a general framework for converting DL cognitive computing techniques into Cyber Forensics (CF). By imitating human decision-making in neural networks, DL uses a variety of machine learning techniques to address problems. These considerations suggest that DL has the potential to both offer forensic investigators options while also having the potential to drastically impact the field of CF in a number of ways. Examples of such remedies include minimizing prejudice in forensic investigations, contesting the admissibility of certain types of evidence in court proceedings or other civil hearings, and many others.

(Akash Chintha et al 2020) this paper offered straightforward yet effective digital forensic techniques for spoof audio and deep fake image detection. The methods combine bidirectional recurrent structures, entropy-based cost functions, and convolutional latent representations to extract semantically rich information from recordings. They are shown using the Face Forensics++, Celeb-DF, and ASVSpoof 2019 Logical Access video datasets and audio datasets, setting new standards in every category. To show generalization to other domains and learn more about the efficacy of the new designs, extensive investigations are carried out.

(Bin Wu et al 2023) to extract relevant and unusual phrases from local areas, the author presented a brand-new framework called FPCNet. For the purpose of identifying face forgery films, this system employs CNN, LSTM, CGLoss, and adaptive feature fusion. In experiments, the suggested technique's detection speed reaches 420 FPS, and the auc scores on the Raw CelebDF, Raw Face Forensics++, F2F, and NT datasets achieve the best results of 99.7%, 99.9%, 94.7%, and 82.0%, respectively. The experimental findings show that the suggested framework outperforms existing frame-level approaches in terms of time economy while also boosting detection performance.

(Ahmed Sedik et al 2022) in this study, a cyber-facial spoofing assault was combined with a deep learning methodology for video face forensic recognition utilizing convolutional neural networks (CNN) and convolutional long short-term memories (ConvLSTM). Simulation findings showed that the ConvLSTM with CNN methodology gave improved classification results in comparison to other conventional strategies. with an accuracy of 99% and up to 95%. In each technique, the classification function was handled by the SoftMax layer.

(Jiahui Wu et al 2023) remote photoplethysmography (rPPG) technique collects heartbeat signals from video recordings by analyzing the small variations in skin color induced by cardiac activity. This is a strong biological signal for deep fake detection since it develops distinct rhythmic patterns in response to various manipulation approaches. To capture both spatial and temporal differences, a two-stage network made up of a Temporal Transformer and a Mask-Guided Local Attention module (MLA) is proposed. The effectiveness of our method in comparison to all existing rPPG-based methods has been thoroughly tested on the Face Forensics ++ and Celeb-DF datasets. The proposed method's usefulness is also demonstrated through visualization.

(Davide Coccomini et al 2022) Since most algorithms are becoming more adept at creating realistic human faces, the author focused on video deep fake detection on faces in this work. We particularly combine different types of Vision Transformers with an Efficient Net B0 convolutional network used as a feature extractor, and the results are comparable to some more recent methods that also use Vision Transformers. Unlike current methodologies, the author does not employ distillation or collective approaches. Additionally, we offer a fundamental inference procedure based on a straightforward voting system for addressing several faces in a single video clip. The top model scored an F1 score of 88.0% and an AUC of 0.951 on the Deep Fake Detection Challenge (DFDC), which is very close to the state-of-the-art.

Aishwarya Rajeev et al [18] Numerous techniques, including Random Forest, Multilayer Perceptron (MLP), and Convolutional Recurrent Neural Networks (CRNN), are employed in this study to execute various kinds of forensic investigation. Also employed is image fusion, which may combine many photos to create a single image with more information and extract characteristics from the original images. According to this study's findings, the random forest has a 98.02 percent accuracy rate when it comes to producing the best results for network forensic investigation. The paper seeks to present an extensive summary of the work that has been done over the past few years to analyze current techniques and techniques for video source authentication using machine learning.

3. Face Forensics

Face forensics, commonly referred to as facial recognition forensics, is the use of forensic techniques and technology to analyze and study face photographs or videos with the aim of establishing identity, verifying a person's identity, or obtaining evidence. To extract and analyze face characteristics and patterns in order to infer or draw conclusions entails using a variety of techniques, algorithms, and tools.

Face forensics may be used in a variety of fields, including biometrics, digital forensics, law enforcement, and security. The following are some typical uses for facial forensics:

- ***Facial Identification:*** Face forensics is frequently used to identify people in surveillance footage, pictures, or videos. To uncover probable matches, face recognition algorithms compare the subject's facial traits with a database of well-known people (Aishwarya et.al.,2023).
- ***Face Authentication:*** It includes matching a person's face traits with their stored biometric information to confirm their identification. Mobile devices, access control systems, and other security applications all make use of this technology (Xiao et.al.,2019).
- ***Facial Image Analysis:*** To extract information from photos or videos or to spot modifications, forensic professionals employ face image analysis tools. This might involve analyzing facial expressions, locating locations or characteristics, and determining the veracity or integrity of a picture, among other things (Ahmadi et.al.,2021).
- ***Facial Age Progression/Regression:*** Face forensics methods can be applied to a person's face to imitate the aging or de-aging of their face based on their present or former look. Investigating missing persons or identifying people in unresolved instances may benefit from this (Ross et.al.,2020).
- ***Facial Emotion Analysis:*** In order to identify emotional states like happiness, sorrow, rage, or surprise, face recognition algorithms analyze facial expressions. Fields like psychology, market research, or human-computer interface may find a use for this (Chandaliya et.al.,2022).
- ***Facial Image Retrieval:*** Face forensics may help in searching through massive databases of pictures or videos based on particular features or traits of the face. Criminal investigations or the identification of people of interest may benefit from this (Ivanova et.al.,2020).

It's crucial to recognize that the application of face forensics involves issues of privacy and ethics. Discussions about regulation and protecting personal privacy have arisen in response to the potential for misuse or abuse of face recognition technology.

Face forensics, or face manipulation detection, is an important area of research in computer vision and deep learning. Various methods have been developed to detect manipulated or fake faces using deep learning concepts. Here are a few different techniques for face forensics:

3.1. Facial Expression Analysis

Facial expression analysis is an intriguing area that focuses on comprehending and analyzing the emotions expressed via facial expressions. Researchers and practitioners can explore the intricate world of human emotions by analyzing the minute movements, configurations, and dynamics of the face (Sikkandar et.al.,2020). In this procedure, essential face traits including the position of the eyebrows, the state of the eyes, the shape of the lips, and more are extracted from facial photos or videos. Then, these characteristics are examined using a range of techniques, including machine learning and computer vision algorithms, to categorize and interpret emotions including happiness, grief, rage, surprise, fear, and disgust. (Keshari et.al.,2019). Applications for facial expression analysis may be found in a variety of industries, including psychology, HCI, market research, and even the professional diagnosis of mental health issues. Facial expression analysis helps us better comprehend non-verbal communication and human emotions by utilizing ongoing technological and algorithmic breakthroughs (Hussain et.al.,2020).

3.2. Facial Landmark Detection

To locate and analyze important spots or landmarks on the face, a major approach used in face forensics is known as facial landmark detection. These landmarks, including the corners of the mouth, nose, and eyes, offer geometric information and serve as starting points for further study. Researchers may learn important details about the structure, position, and emotions of the face by precisely detecting and tracking facial landmarks. These insights are crucial for spotting and analyzing possible modifications (Ashwin et.al.,2019). Facial landmark detection is essential in face forensics for determining the veracity and accuracy of a face picture. Disparities or inconsistencies brought on by manipulations, such as face swapping or morphing, can be found by analyzing the locations, configurations, and motions of landmarks. When landmark spatial connections differ from what is expected, it may be a sign that the picture has been altered or tampered with. Additionally, facial landmark identification can help pinpoint regions of interest for later research. For instance, by identifying the eye landmarks, researchers may concentrate on analyzing eye-related alterations, such as changing the color of the eyes or adding digital contact lenses. Similar to this, recognizing mouth landmarks can assist in spotting possible lip-syncing or speech manipulation. It's crucial to remember that face forensics facial landmark detection might be difficult. It could be delicate to changes in facial expression, occlusions, or head posture (Agbolade et.al.,2019). Accurate landmark detection may also be hampered by the presence of cosmetics, accessories, or facial hair. As a result, to manage these complexity levels and guarantee correct outcomes, strong and precise algorithms are required (Bozkir et.al.,2023). In conclusion, facial landmark detection is an important method in face forensics that provides geometric data and helps spot any alterations. Researchers may improve their study of facial integrity by utilizing precise landmark detection, making a contribution to the fields of digital forensics and biometrics as well as assuring the reliability of face-based authentication systems.

3.3. Face Swapping Detection

Face forensics experts use the important technology of "face swapping detection" to spot instances of faces being switched or replaced in photos or videos. This method seeks to spot visual tricks when one person's face is digitally swapped out for another, which frequently produces believable but misleading results. Face swapping detection is essential for maintaining the authenticity and integrity of visual material in face forensics. Face swapping detection algorithms can spot obvious evidence of manipulation by analyzing a variety of visual signals and attributes, including facial landmarks, textures, lighting, and consistency in facial emotions (Dargan et.al.,2020). The geometric alignment and arrangement of facial landmarks before and after the swap is one typical method utilized in face swapping detection. Key facial features like the eyes, nose, and mouth may be precisely detected and tracked, making it possible to see any differences or irregularities in their locations. Face swapping may be present if there are significant differences in the way these landmarks are spaced out from one another. The uniformity and naturalness of the face textures in the swapped region may also be examined using texture analysis tools (Verdoliva 2020). Anomalies can point to the presence of face alteration, such as artificial blending or variances in lighting.

Due to improvements in face manipulation methods and the possibility for perfect blending, face swapping detection can be difficult. Face swapping algorithms based on deep learning can provide extremely convincing results that are hard to spot. In order to improve face swapping detection techniques' accuracy and sturdiness, it is essential to conduct continuing research and make improvements in the fields of deep learning, computer vision, and forensic analysis (Hashmi et.al.,2020). In order to detect instances of face replacement or swapping, face forensics must employ a fundamental method called face swapping detection. It contributes to the creation of trustworthy digital media and raises the trustworthiness of face-based authentication systems by assessing facial landmarks, textures, and other visual clues to assure the integrity and authenticity of visual material.

3.4. Deep Fake Detection

A key method in face forensics for identifying and detecting heavily altered or artificial face photos and films is deep fake detection. The term "deep fake" refers to media that has been purposefully made using deep learning algorithms. Often, this involves swapping out a person's face for another, creating very life-like and deceptive visual results. Deep fake identification in face forensics is essential for reducing the dangers that might result from the illicit exploitation of altered material. To analyze and examine the veracity of face material, deep fake detection algorithms use a variety of methodologies. Examining visual artifacts, inconsistencies, and abnormalities that are often included throughout the deep fake creation process is involved in these procedures (Deshmukh et.al.,2020). Analyzing minute artifacts and flaws that result from the synthesis process is a typical strategy in deep fake detection. Deep fakes frequently display uneven mixing, irregular lighting, and differences in the resolution and texture of the face. Algorithms can recognize these red flags and differentiate deep fakes apart from real face material by utilizing deep learning models and computer vision techniques (Awotunde et.al.,2022).

Using sophisticated machine learning models to discover and identify patterns particular to deep fakes is an alternative strategy. These models can spot statistical discrepancies and distinctive traits related to deep fake manipulation by training on a sizable dataset of both genuine and deep fake samples. With this method, key characteristics are frequently extracted and the legitimacy of the information is categorized using convolutional neural networks (CNNs) or recurrent neural networks (RNNs). However, due to the quick advancement of deep fake-generating techniques as well as the introduction of advanced adversarial strategies, deep fake detection approaches confront difficulties. Deep fakes produced by adversarial networks may be extremely convincing and difficult to differentiate from legitimate information, making detection more difficult. In order to keep ahead of developing methods and guarantee the integrity of digital material, ongoing research, and breakthroughs in deep fake detection are essential (Byrnes et.al.,2021).

In order to recognize fabricated or altered face photos and videos, deep fake detection is an essential approach in face forensics. Deep fake detection approaches help protect against the possible abuse of deep fakes and improve the credibility of digital material by analyzing visual artifacts, and inconsistencies, and utilizing machine learning models.

3.5. Deep Texture Analysis

Deep texture analysis is a crucial tool in face forensics. It focuses on analyzing the complex patterns, specifics, and irregularities contained in the texture of a face using deep learning models and cutting-edge computer vision techniques. It seeks to identify minute texture variations that may be a sign of face switching, digital retouching, or other types of manipulation. Convolutional neural networks (CNNs) trained to identify and extract pertinent characteristics from facial textures are one popular method in deep texture analysis. Algorithms can spot differences or anomalies that may indicate tampering by analyzing the consistency and coherence of texture patterns across various face areas (Chen et.al.,2022).

However, when dealing with alterations in lighting, picture quality, or the presence of cosmetics and accessories that might affect the texture look, deep texture analysis may run into difficulties. The identification method is made more difficult by the fact that very advanced modification techniques may produce deep fakes with convincing texture features. To keep up with changing manipulation techniques and boost the precision and resilience of deep texture analysis approaches, ongoing research and development are crucial (Xi et.al.,2020).

In conclusion, deep texture analysis is an important method in face forensics for identifying and analyzing textural elements in facial photographs. It adds to the identification and detection of manipulated or altered faces by utilizing deep learning models and studying texture patterns and irregularities, improving the integrity and reliability of face-based forensic investigation and digital media authentication.

Table 1. Comparison table of different face forensics techniques using deep learning concepts

Ref.	Technique	Description	Advantages	Limitations
Jeong et.al.,2020	**Facial Expression Analysis**	Analyses facial expressions to infer emotions or intentions	Provides insights into the emotional state of a person	Difficulty in accurately interpreting complex or subtle facial expressions
Zhu et.al.,2019	**Facial Landmark Detection**	Identifies key facial landmarks for further analysis	Provides geometric information about the face	Sensitive to occlusions or variations in facial expressions
Zhang et.al.,2019	**Face Swapping Detection**	Identifies instances where faces have been swapped or replaced	Effective in detecting face replacement or swapping	Limited to face swapping techniques
Zhao et.al.,2021	**Deep Fake Detection**	Detects manipulated faces using deep learning models	Effective against deep fake videos	Limited to specific types of manipulations (e.g., deep fake videos)
Bonomi et.al.,2021	**Deep Texture Analysis**	Analyses textural features within the face using	Effective in detecting inconsistencies or	May struggle with subtle manipulations or

		deep learning models	manipulations in texture patterns	variations in image quality

4. Analysis On Face Forensics

Face forensics analysis is a broad discipline that includes a range of methods and tools for analyzing and interpreting facial characteristics in the settings of criminal investigations and legal evidence. It entails using face recognition algorithms to match and identify people according to their facial features. Experts in face forensics also use picture authentication techniques to establish the veracity and integrity of facial images, such as examining any indications of alteration or digital interference. In order to understand emotions, believability, and deceit, they also analyse the facial expressions shown in photos or movies. Techniques for estimating a person's age and predicting their anticipated appearance at various ages are used, as well as methods for comparing facial traits to databases of missing people or suspects. In order to confirm identities or determine whether persons in visual evidence are the same, facial comparison and superimposition techniques are used. In addition, forensic anthropology employs facial reconstruction methods to recreate the look of unidentified human remains' faces, aiding in the identification process. But it's important to recognize the possible drawbacks of face forensics, such as changes in picture quality, lighting, position, and the existence of barriers or disguises, which may affect the precision of studies. Therefore, face forensics analysis should be performed by qualified experts who take these elements into account and use caution when making interpretations.

In this study, we concentrate on the face forensics research conducted by several researchers, compare those works, evaluate the findings, and, after comparing all the works, focus on the shortcomings of prior research and utilize those shortcomings as the foundation for our future work.

4.1. Deep Fake Video Detection Using CNN and RCNN :

(Ashifur Rahman et al 2022) deep fake videos that are convincing and increasing quickly in popularity can now fool even experienced professionals. The political, social, and personal spheres are all affected significantly by these profoundly false videos. In high quality and lengthy video data, modern machine learning experiments provide demonstrable success in spotting fraudulent movies, however, this performance is not demonstrated in low resolution and brief video clips. In this study, the authors created a model using convolutional neural networks (CNN) and recurrent neural networks (RNN) that shows mentionable accuracy in detecting fraudulent films in low-resolution and short-duration video data. In our experiment, the author employed the Kaggle Deep Fake Detection Challenge (DFDC) dataset and the Face Forensics++ dataset. When it came to identifying false videos, the model performed 94.8% accurately for the RCNN model and 94.2% accurately for the CNN model. The author compared the performance of our models to cutting-edge techniques and evaluated our models using several performance measures. Comparable performance is shown by the model. The mathematical equation for RNN has been shown in equations 1 and 2

Recurrent Neural Networks (RNNs) are a class of neural networks specifically designed to work with sequence data. They are defined by the recurrence of neural network modules, allowing them to maintain a memory of previous inputs to make decisions about the current input. Mathematically, an RNN can be represented through its forward pass equations.

Let's denote:

- x_t as the input at time step t
- h_t as the hidden state at time step t
- y_t as the output at time step t

The basic equations for a simple RNN are as follows:

$$h_t = \sigma(W_{ih} \cdot x_t + W_{hh} \cdot h_{t-1} + b_h) \tag{1}$$

$$y_t = \text{softmax}\left(W_{hy} \cdot h_t + b_y\right) \tag{2}$$

Where:

- W_{ih} is the weight matrix for input to hidden connections.
- W_{hh} is the weight matrix for hidden to hidden connections.
- W_{hy} is the weight matrix for hidden to output connections.
- b_h and b_y are the bias terms for the hidden and output layers, respectively.
- σ is an activation function, commonly the hyperbolic tangent (tanh) or the Rectified Linear Unit (ReLU).

4.1.1. Data Set:

This study utilized BlazeFace, MTCNN, and face recognition DL libraries to extract faces swiftly. BlazeFace and face recognition are particularly efficient for processing numerous images. Combining these three DL libraries enhances the accuracy of face detection. They stored face images in JPEG format, sized at 224 x 224 resolution. The dataset was split into training, validation, and test sets, containing 162,174 images in total. Specifically, 112,378 were allocated for training, 24,898 for validation, and another 24,898 for testing, maintaining a 70:15:15 ratio respectively. Both the real and fake image classes were equally represented across all sets.

4.1.2. Roc Curve:

A Receiver Operating Characteristic (ROC) curve represents the performance of a binary classification model graphically. At different threshold settings, it compares the true positive rate (sensitivity) to the false positive rate (1 - specificity). The schematics of ROC curve is shown in Fig.1

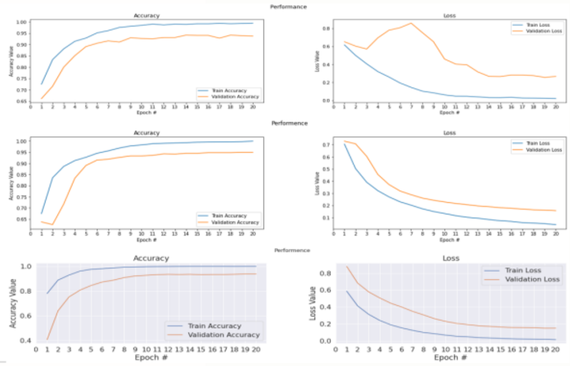

Fig. 1. ROC curve of all CNN models [38]

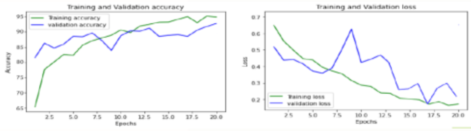

Fig. 2.

ROC curve of CRNN model [38]

4.1.3. Result and Discussion:

Table 2. Result of CNN and RCNN model [38]

System Name	Architecture	Accuracy	Precision
CNN	InceptionRes-NetV2	93.75%	98.0%
	Mobile Net	94.2%	99.0%
	DenseNet121	93.86%	98.0%
RCNN	CNN+RNN	94.8%	94.4%

4.2. Dual Attention Network Approaches to Face Forgery Video Detection [44]:

(YI-XIANG LUO et al 2022)in this study, a Forgery Feature Attention Module (FFAM) and a Spatial Reduction Attention Block (SRAB) were integrated into the backbone network to construct a Dual Attention Forgery Detection Network (DAFDN). The two attention processes that have been presented are embedded by DAFDN, it additionally makes it possible for the convolution neural network to extract odd traces from the warped images. This study compares the effectiveness of the proposed DAFDN with other techniques using two benchmark datasets, DFDC and Face Forensics++. The results show that the proposed DAFDN approach achieves AUC values of 0.911 and 0.945, respectively, in the DFDC and Face Forensics++ datasets. These outcomes surpass those of earlier developed techniques like XceptionNet and Efficient Net-related techniques.

The whole process can be expressed as follows.

$$\begin{aligned} M_{sr}(F) &= \sigma\left(f^{7\times7}([f^{1\times1}(F); \text{MaxPool}\ (F)])\right) \\ &= \sigma\left(f^{7\times7}([F_c; F_{\max}])\right) \end{aligned} \tag{3}$$

Where σ denotes the sigmoid function; $f^{7\times7}$ and $f^{1\times1}$ represent that they were calculated via convolution, and the superscript stands in for the kernel size.

4.2.1. Data Set:

In order to assess how well DAFDN performs in identifying deep fake movies, Deep fake Detection Challenge (DFDC) and Face Forensics++ (FF++) are used as two benchmark datasets. A first-generation deep fake dataset, the FFCC dataset contains 1000 YouTube raw video sequences. The front of the face is not obscured in any of the movies because they were all manually chosen, making it possible for forgery techniques to produce lifelike forgeries. Two methods—classical computer graphics and deep learning—

can be used to generate counterfeit videos in the order they appear in the film.

4.2.2. Roc Curve:

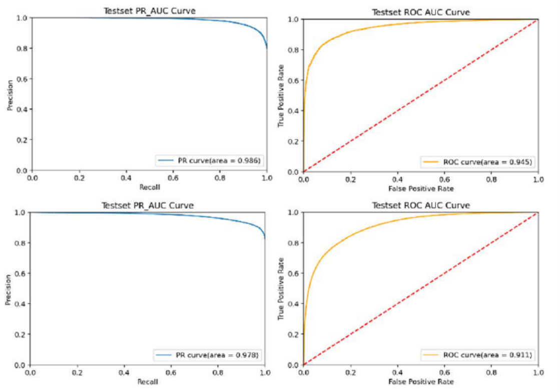

Fig. 3. PR curve and ROC curve of DAFDN on FFCC and DFDC [44].

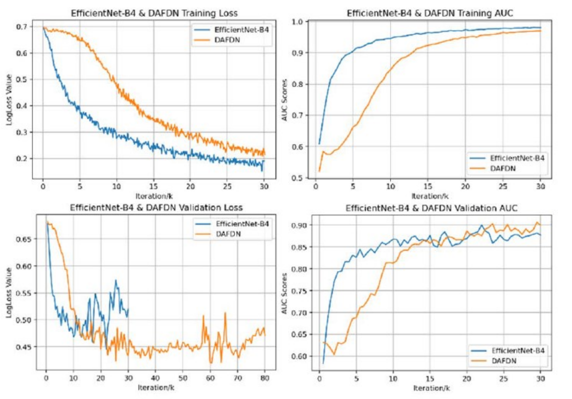

Fig. 4. Loss and AUC of EfficientNet-B4 and DAFDN [44].

4.2.3. Results of Analysis:

DAFDN received PR-AUC scores of 0.978 and 0.911 with DFDC, as well as 0.978 and 0.945 with FFCC. Both the FF++ and DFDC datasets show strong performance with DAFDN. Given that the test set is composed of data that have never been used for model training, the model's outstanding performance illustrates its great generalizability.

Table 3. Performance by DAFDN [39]

System Name	Architecture	Accuracy	Precision
DAFDN	DAFDN+ DFDC	97.8%	91.1%
DAFDN	DAFDN+ FF++	97.8%	94.5%

4.3. Deepfake Video Detection Through the Use of a Hybrid CNN Deep Learning Model :

(Sumaiya Thaseen Ikram et al 2023)with the use of numerous software programs and cutting-edge AI (Artificial Intelligence) technology, a number of fake films and images are created in the current era, leaving behind certain telltale evidence of manipulation. Videos may be used in a variety of unethical ways to intimidate, quarrel, or frighten others. Make sure that no fraudulent videos are produced using such techniques. Deep Fake is the name of an AI-based method for creating synthetic human photographs. They are produced by mixing and overlaying pre-existing videos over the original videos. In order to extract frame-level characteristics, a method that combines InceptionResnet v2 and Xception is built in this research. For experimental analysis, the DFDC deep fake detection challenge on Kaggle is used. The accuracy and training time of these deep learning-based algorithms are increased by using this dataset for both training and testing. The following results were obtained: accuracy 0.985, recall 0.96, f1-score 0.98, and support 0.968.

4.3.1. Data Set:

The DFDC dataset serves as the foundation for experiments, distinguishing itself from other deep fake datasets by collecting over 100,000 clips involving 3,426 paid actors. Unlike many other datasets that use non-consensual footage shot in controlled environments, the DFDC dataset stands out for its diverse collection of face swap videos sourced from various algorithms, including Deep fake, GAN-based, and non-learned methods. Notably, each of the 100,000 forged videos in this dataset presents a unique target/source switch, showcasing a wide array of scenarios spanning indoor and outdoor settings with different lighting conditions. Despite disruptions, DF-1.0 encompasses 1,000 distinct forged videos, contributing to the dataset's comprehensive nature and its status as the largest publicly available face swap video dataset.

4.3.2. Roc Curve:

The ROC curve for Sumaiya's suggested study is presented in Figure. 5. The genuine positive rate vs. false positive rate for various parameter cut-off points is plotted in this graph.

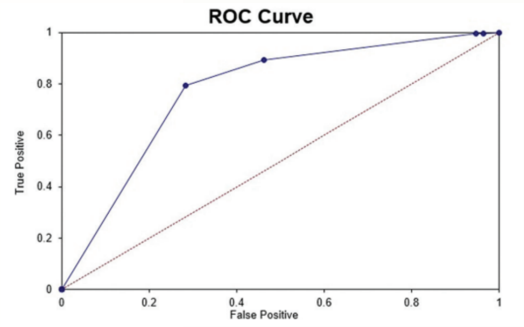

Figure. 5. ROC Curve for Training and Validation [45]

4.3.3. Result and Discussion

Table 4. Performance by CNN based face detection

System Name	Architecture	Accuracy	Precision	Recall	F1-Score
CNN	Inception	96%	97%	94%	93%
CNN	Xception	93%	98%	98%	91%
CNN	Hybrid Inception-Resnet v2 and Xception	98%	99%	97%	98%

5. Comparison of Previous Research:

Comparing previous deep learning-based face forensics results can provide insights into the developments and performance obtained by various models and methodologies.

Table 5. Comparative analysis

Author	System	Architecture	Accuracy	Precision
Ashifur Rahman et al [38]	**RCNN**	CNN+RNN	94.8%	94.4%
YI-XIANG LUO et al [39]	**DAFDN**	DAFDN+ FF++	97.8%	94.5%
Sumaiya Thaseen Ikram et al [40]	**CNN**	Hybrid Inception Resnet v2 and Xception	98%	99%

The table outlines a comparison among three distinct systems—RCNN, DAFDN, and CNN—focusing on their structure, precision, and accuracy. RCNN merges Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), achieving 94.8% accuracy and 94.4% precision. In contrast, DAFDN integrates DAFDN and FF++ elements, displaying exceptional performance with 97.8% accuracy and 94.5% precision. The CNN system combines Inception Resnet v2 and Xception models, yielding a remarkable 98% accuracy and an impressive precision of 99%. These statistics highlight how each system's design effectively handles tasks, offering valuable insights into their capabilities within neural networks.

Table 6. Models significance and limitation

Model	Advantages	Limitations
RCNN	Handles varied facial poses and orientations well	Computationally intensive
	Good at detecting faces even in cluttered backgrounds	Requires large datasets for training
	Can identify faces across different scales	May struggle with low-resolution or distorted images
DAFDN	Focuses on fine-grained facial features	Limited to frontal or near-frontal face orientations
	Robust against some common deepfake manipulations	May not generalize well to diverse deepfake variations
	Efficient and faster than some other deepfake models	Can be susceptible to adversarial attacks
CNN-based	Flexible architecture adaptable to different tasks	Performance highly dependent on data quality and quantity
	Can learn intricate patterns and features effectively	Prone to overfitting without proper regularization
	Handles varying lighting conditions well	Limited by training data availability and diversity

This study examined three existing deep learning methods that outperformed other existing techniques. This research assessed CNN+RNN, DAFDN, Hybrid Inception Resnet v2, and Xception, and the outcomes showed that the Inception Xception model outperforms RCNN and Dual Attention Network in terms of performance. Combining three models, such as Inception Net, ResNet, and Xception Net, generates a more complicated structure and increases computation time when compared to others. There are a few limitations in the Inception and Xception models since both are computationally costly due to their deep and complex topologies. They have a huge number of parameters and actions, making them longer to train and infer than simpler models. This computational complexity may restrict their usefulness in resource-constrained contexts or real-time applications where low latency is critical.

Pre-trained models obtained from extensive datasets (such as Image Net) are used as a starting point in transfer learning, and can help Xception Net. You may fine-tune Xception Net on face forensics datasets, which are often smaller, by exploiting the learned characteristics from these models. Furthermore, regularization approaches can aid in the prevention of over fitting and the improvement of generalization. Methods like dropout and batch normalization can help to regularize the model and limit the danger of over fitting on the training data. These methods motivate the model to learn more robust and generalizable features.

6. Future Scope:

Deep fake detection architecture holds great potential in the future, as advanced countermeasures are needed against the rise of sophisticated AI techniques that generate hyper-realistic manipulated content. Key players in this evolving landscape include Convolutional Neural Networks (CNNs), Region-based Convolutional Neural Networks (RCNNs), and the Domain Adaptive Few-Shot Detection Network (DAFDN). CNNs have been at the forefront of deep fake detection, but challenges remain in enhancing their robustness against evolving deep fake techniques. Future research may focus on improving interpretability, incorporating attention mechanisms, and exploring novel architectures to detect subtle artifacts and inconsistencies introduced by deep fake algorithms. RCNNs capture spatial relationships in images, but adapting to real-time processing and handling video streams efficiently remains a challenge. DAFDN, designed to adapt to domain shifts and few-shot scenarios, presents a promising direction for future research. However, challenges persist, such as adversarial attacks and ethical considerations surrounding privacy and consent in the deployment of deep fake detection technologies. In conclusion, the future of deep fake detection architecture is poised for continued innovation and refinement, with researchers navigating the evolving landscape of deep fake generation techniques, enhancing detection networks' speed and efficiency, and addressing ethical considerations to ensure responsible deployment in real-world scenarios.

7. Conclusion:

Only a limited number of individuals working in law enforcement, intelligence, and private investigations had any practical use for multimedia forensics. fourteen years ago. Both offense and defense had an artisanal feel and needed meticulous effort and commitment. Artificial intelligence has primarily modified these rules. Today, it appears that high-quality imitations are made on a production line, requiring extraordinary efforts from scientists and decision-makers. In actuality, today's multimedia forensics is fully developed, important organizations are supporting significant research projects, and experts from other fields are actively contributing with quick developments in concepts and techniques. This analysis will look at three studies that looked into CNN+RNN, DAFDN, Hybrid Inception Resnet v2, and Xception in relation to Face Forensics. According to the results, the Inception Xception model performs better than RCNN and Dual Attention Network.

Reference:

Shad, H. S., Rizvee, M. M., Roza, N. T., Hoq, S. M., Monirujjaman Khan, M., Singh, A. &amp; Bourouis, S. (2021). Comparative analysis of deepfake image detection method using convolutional neural network. Computational Intelligence and Neuroscience, 2021.

Rahman, A., Islam, M. M., Moon, M. J., Tasnim, T., Siddique, N., Shahiduzzaman, M. & Ahmed, S. (2022). A qualitative survey on deep learning based deep fake video creation and detection method, Aust. J. Eng. Innov. Technol, 4(1), 13-26.

Nguyen, T. T., Nguyen, C. M., Nguyen, D. T., Nguyen, D. T. & Nahavandi, S. (2019). Deep learning for deepfakes creation and detection, arXiv preprint arXiv:1909.11573, 1(2), 2.

Bode, L. (2021). Deepfaking Keanu: YouTube deepfakes, platform visual effects, and the complexity of reception, Convergence, 27(4), 919-934.

Verdoliva, L. (2020). Media forensics and deepfakes: an overview, IEEE Journal of Selected Topics in Signal Processing, 14(5), 910-932.

Kousik, N., Natarajan, Y., Raja, R. A., Kallam, S., Patan, R., & Gandomi, A. H. (2021). Improved salient object detection using hybrid Convolution Recurrent Neural Network. Expert Systems with Applications, 166, 114064.

Kamaleldin, M. G. M., Abu-Bakar, S. A. & Sheikh, U. U. (2023). Transfer Learning Models for CNN Fusion with Fisher Vector for Codebook Optimization of Foreground Features, IEEE Access.

Guo, Z., Yang, G., Chen, J. & Sun, X. (2021). Fake face detection via adaptive manipulation traces extraction network, Computer Vision and Image Understanding, 204, 103170.

George, A. S. & George, A. H. (2023). Deepfakes: The Evolution of Hyper realistic Media Manipulation, Partners Universal Innovative Research Publication, 1(2), 58-74.

Wang, T., Zhang, Y., Qi, S., Zhao, R., Xia, Z. & Weng, J. (2023). Security and privacy on generative data in aigc: A survey, arXiv preprint arXiv:2309.09435.

Ding, F., Zhu, G., Alazab, M., Li, X. & Yu, K. (2020). Deep-learning-empowered digital forensics for edge consumer electronics in 5G HetNets, IEEE consumer electronics magazine, 11(2), 42-50.

Karie, N. M., Kebande, V. R., & Venter, H. S. (2019). Diverging deep learning cognitive computing techniques into cyber forensics, Forensic Science International: Synergy, 1, 61-67.

Chintha, A., Thai, B., Sohrawardi, S. J., Bhatt, K., Hickerson, A., Wright, M., & Ptucha, R. (2020). Recurrent convolutional structures for audio spoof and video deepfake detection, IEEE Journal of Selected Topics in Signal Processing, 14(5), 1024-1037.

Wu, B., Su, L., Chen, D., & Cheng, Y. (2023). FPC-Net: Learning to detect face forgery by adaptive feature fusion of patch correlation with CG-Loss, IET Computer Vision, 17(3), 330-340.

Sedik, A., Faragallah, O. S., El-sayed, H. S., El-Banby, G. M., El-Samie, F. E. A., Khalaf, A. A. & El-Shafai, W. (2022). An efficient cybersecurity framework for facial video forensics detection based on multimodal deep learning, Neural Computing and Applications, 1-18.

Wu, J., Zhu, Y., Jiang, X., Liu, Y., & Lin, J. (2023). Local attention and long-distance interaction of rPPG for deepfake detection, The Visual Computer, 1-12.

Coccomini, D. A., Messina, N., Gennaro, C. & Falchi, F. (2022, May). Combining efficientnet and vision transformers for video deepfake detection, In Image Analysis and Processing–ICIAP 2022: 21st International Conference, Lecce, Italy, May 23–27, 2022, Proceedings, Part III, Cham: Springer International Publishing, 219-229.

Aishwarya Rajeev, A., & Raviraj, P. (2023). An Insightful Analysis of Digital Forensics Effects on Networks and Multimedia Applications, SN Computer Science, 4(2), 186.

Xiao, J., Li, S., & Xu, Q. (2019). Video-based evidence analysis and extraction in digital forensic investigation, IEEE Access, 7, 55432-55442.

Ahmadi, F., Gupta, G., Zahra, S. R., Baglat, P. & Thakur, P. (2021, March). Multi-factor biometric authentication approach for fog computing to ensure security perspective, In 2021 8th international conference on computing for sustainable global development (INDIACom), IEEE, 172-176.

Ross, A., Banerjee, S., & Chowdhury, A. (2020). Security in smart cities: A brief review of digital forensic schemes for biometric data, Pattern Recognition Letters, 138, 346-354.

Chandaliya, P. K., & Nain, N. (2022). ChildGAN: Face aging and rejuvenation to find missing children, Pattern Recognition, 129, 108761.

Ivanova, E., & Borzunov, G. (2020). Optimization of machine learning algorithm of emotion recognition in terms of human facial expressions, Procedia Computer Science, 169, 244-248.

Sikkandar, H., & Thiyagarajan, R. (2020). Soft biometrics-based face image retrieval using improved grey wolf optimisation, IET Image Processing, 14(3), 451-461.

Keshari, T., & Palaniswamy, S. (2019, July). Emotion recognition using feature-level fusion of facial expressions and body gestures, In 2019 international conference on communication and electronics systems (ICCES), IEEE, 1184-1189.

Hussain, S. A. & Al Balushi, A. S. A. (2020). A real time face emotion classification and recognition using deep learning model, In Journal of physics: Conference series, IOP Publishing, 1432(1), 012087.

Ashwin, T. S. & Guddeti, R. M. R. (2019). Unobtrusive behavioral analysis of students in classroom environment using non-verbal cues, IEEE Access, 7, 150693-150709.

Agbolade, O., Nazri, A., Yaakob, R., Ghani, A. A. & Cheah, Y. K. (2019). 3-Dimensional facial expression recognition in human using multi-points warping, BMC bioinformatics, 20(1), 1-15.

Bozkir, E., Özdel, S., Wang, M., David-John, B., Gao, H., Butler, K. & Kasneci, E. (2023). Eye-tracked Virtual Reality: A Comprehensive Survey on Methods and Privacy Challenges. arXiv preprint arXiv:2305.14080.

Dargan, S. & Kumar, M. (2020). A comprehensive survey on the biometric recognition systems based on physiological and behavioral modalities, Expert Systems with Applications, 143, 113114.

Verdoliva, L. (2020). Media forensics and deepfakes: an overview, IEEE Journal of Selected Topics in Signal Processing, 14(5), 910-932.

Hashmi, M. F., Ashish, B. K. K., Keskar, A. G., Bokde, N. D., Yoon, J. H. & Geem, Z. W. (2020). An exploratory analysis on visual counterfeits using conv-lstm hybrid architecture, IEEE Access, 8, 101293-101308.

Deshmukh, A. & Wankhade, S. B. (2020). Deepfake Detection Approaches Using Deep Learning: A Systematic Review, Intelligent Computing and Networking: Proceedings of IC-ICN 2020, 293-302.

Awotunde, J. B., Jimoh, R. G., Imoize, A. L., Abdulrazaq, A. T., Li, C. T. & Lee, C. C. (2022). An Enhanced Deep Learning-Based DeepFake Video Detection and Classification System, Electronics, 12(1), 87.

Byrnes, O., La, W., Wang, H., Ma, C., Xue, M., & Wu, Q. (2021). Data hiding with deep learning: A survey unifying digital watermarking and steganography, arXiv preprint arXiv:2107.09287.

Chen, L., Zhang, Y., Song, Y., Liu, L. & Wang, J. (2022). Self-supervised learning of adversarial example: Towards good generalizations for deepfake detection, In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 18710-18719.

Xi, Z., Niu, Y., Chen, J., Kan, X. & Liu, H. (2020). Facial expression recognition of industrial internet of things by parallel neural networks ecombining texture features, IEEE Transactions on Industrial Informatics, 17(4), 2784-2793.

Jeong, D., Kim, B. G., & Dong, S. Y. (2020). Deep joint spatiotemporal network (DJSTN) for efficient facial expression recognition, Sensors, 20(7), 1936.

Zhu, M., Shi, D., Zheng, M., & Sadiq, M. (2019). Robust facial landmark detection via occlusion-adaptive deep networks, In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 3486-3496.

Zhang, W., & Zhao, C. (2019, November). Exposing face-swap images based on deep learning and ELA detection, In Proceedings, MDPI, 46(1), 29.

Zhao, H., Zhou, W., Chen, D., Wei, T., Zhang, W., & Yu, N. (2021). Multi-attentional deepfake detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2185-2194.

Bonomi, M., Pasquini, C., & Boato, G. (2021). Dynamic texture analysis for detecting fake faces in video sequences, Journal of Visual Communication and Image Representation, 79, 103239.

Rahman, A. (2022). Deepfake Video Detection Using CNN and RCNN (Doctoral dissertation, Bangladesh University of Business and Technology).

Luo, Y. X. & Chen, J. L. (2022). Dual Attention Network Approaches to Face Forgery Video Detection. IEEE Access, 10, 110754-110760.

Ikram, S. T., Chambial, S., & Sood, D. (2023). A performance enhancement of deep fake video detection through the use of a hybrid CNN Deep learning model. International journal of electrical and computer engineering systems, 14(2), 169-178.

AUTHOR BIOGRAPHIES

Aishwarya Rajeev is currently working as an Assistant Professor & Head in the Department of Artificial Intelligence and Data Science at Coorg Institute of Technology, Ponnampet, Karnataka. She holds an M.E in Computer Science and Engineering with first rank and gold medal from Mahendra Engineering College, Affiliated to Anna University, Chennai, India in 2015. She also holds a MBA in IT & Systems from ICFAI University, Tripura, India in 2012. She received B. Tech Degree in Information Technology from Cochin University of Science and Technology, India in 2008. She has 15.6 years of experience in teaching has published papers, and attended various conferences and workshops. She is also a life member of professional bodies like ISTE, IE, and IAENG. She serves as Editorial Board Member and Reviewer of 2 International Journals. She can be contacted at email: aishwaryarajeev@gmail.com.

P. Raviraj completed his doctorate degree in Computer Science and Engineering in the area of Image Processing. He holds the position of Director-IQAC and Professor & Head in the Department of Computer Science & Engineering at GSSS Institute of Engineering & Technology for Women, Mysore, Karnataka. He has 19 years of teaching and research experience. He has published more than 94 papers in International journals and conferences. Five research scholars have completed their Ph.D. under his guidance at various universities. At present, he has guiding Ph.D. research scholars in the areas of Image Processing, Pervasive & Cloud computing, Bio-Inspired Algorithms Robotics etc. He has received the project grant Rs.5 Lakhs from the VGST, Govt. of Karnataka for the “Underwater Robotic Fish for Surveillance and Pollution monitoring”. He filed the patent entitled “An effective ROI based Hybrid Progressive Medical Image Transmission and Reconstruction” in the year 2021. He is serving as a Ph.D thesis Adjudicator, Doctor Committee member and Subject expert for various universities. He has served as a chairperson and keynote speaker at many National and International Conferences. He serves as an Editorial Board Member and Reviewer for more than 15 International Journals. He is also a life member of professional bodies like ISTE, CSI etc. He has received awards and recognitions such as ‘Rhastriya Gaurav Award-2015’, ‘Shri P.K.Das Memorial Best Faculty Award-2012’, ‘Young Achiever Award-2016’, ‘Best Circuit Faculty Finalist Award-2017’ in his credit.