ENHANCING MEDICAL IMAGE SECURITY: A PCA-DEEP LEARNING APPROACH FOR ATTACK DETECTION

N V S K Vijayalakshmi K
Research Scholar, Department of Information Technology Annamalai university,
vijayakathari@sircrrengg.ac.in

Dr. J. Sasikala
Department of Information Technology, Annamalai university, Associate Professor
sasikala.au@gmail.com

Dr. C. Shanmuganathan
Department of Computer Science and Engineering Technology, SRM university, Associate Professor
drcsnathan@gmail.com

Abstract

The integrity and authenticity of medical images are critical for accurate diagnosis and treatment planning. However, they are increasingly vulnerable to malicious attacks like tampering and manipulation. This work proposes a novel deep learning-based approach for medical image attack detection. The proposed model leverages Principal Component Analysis (PCA) as a feature extractor to capture essential information from the images while reducing dimensionality. Subsequently, two deep learning classifiers, Convolutional Neural Network (CNN) and Inception, are employed for classification. The model is trained and tested on a deepfake dataset, simulating potential attack scenarios in the medical domain. This approach aims to identify tampered medical images and enhance the security of medical image analysis. This research work investigates the effectiveness of the proposed model in distinguishing between genuine and manipulated medical images, paving the way for safeguarding the integrity of medical data and ensuring reliable decision-making in healthcare.

Keywords: PCA, CNN, tampering, Inception, CT scan, Cancer, MRI scan

1. Introduction:

Forgery attacks on medical images pose a significant threat to the integrity and reliability of diagnostic processes in healthcare. With the proliferation of digital imaging technologies, medical professionals increasingly rely on digital images for accurate diagnosis, treatment planning, and monitoring of patients. However, the ease of digital manipulation coupled with the critical nature of medical imaging makes these images vulnerable to various forms of forgery [1]. Medical image forgery encompasses a wide range of malicious activities, including but not limited to, alteration of image content, insertion of artifacts or misleading information, and creation of entirely fabricated images. These forged images can lead to misdiagnosis, inappropriate treatment decisions, and compromised patient care. The implications of forgery attacks on medical images are profound, as they undermine the trustworthiness of medical data and jeopardize patient safety. Moreover, the potential for financial fraud, legal implications, and ethical concerns further exacerbate the gravity of this issue [2]. In recent years, researchers and practitioners in the field of medical imaging have been actively exploring techniques to detect and prevent forgery attacks. Various methods, including digital watermarking, cryptographic techniques, and deep learning-based approaches, have been proposed to enhance the security and authenticity of medical images [3]. This paper aims to provide an overview of forgery attacks on medical images, examine existing detection and prevention methods, and highlight the challenges and future directions in this critical area of research [4][5].

Deep learning techniques are pivotal in detecting forgery in medical images due to their advanced capabilities and adaptability. Firstly, these techniques excel in automatic feature learning, where complex patterns and textures within images are discerned without explicit instruction [6]. This is particularly valuable in identifying subtle alterations or anomalies indicative of forgery. Secondly, deep learning models exhibit heightened sensitivity to minute variations in pixel intensities, shapes, and spatial relationships, enabling them to detect even the most inconspicuous manipulations that may elude traditional methods [7]. Furthermore, deep learning facilitates end-to-end detection pipelines, streamlining the process by directly inputting raw images and outputting predictions of authenticity [8]. This efficiency is crucial, especially in clinical settings where rapid decision-making is imperative. Additionally, deep learning enables adversarial training to fortify models against sophisticated forgery attempts by exposing them to diverse manipulation techniques during training [9]. Moreover, techniques like transfer learning and multi-modal fusion enhance forgery detection across different imaging modalities, leveraging knowledge gained from large-scale datasets and integrating features from various sources for more comprehensive assessments [10] [11]. Overall, deep learning plays a fundamental role in detecting forgery in medical images, offering robust and automated solutions to safeguard the integrity of diagnostic processes in healthcare.

The increasing reliance on medical images in various stages of healthcare, from diagnosis and treatment planning to surgical guidance and research, necessitates robust measures to ensure their integrity and authenticity. Unfortunately, the rise of sophisticated image manipulation techniques poses a significant threat to the security of medical data. Malicious actors can tamper with medical images, subtly altering crucial information, potentially leading to misdiagnosis, inappropriate treatment plans, and ultimately jeopardizing patient safety.

The overall contribution of this work has been elucidated below:

1. **Data Preprocessing:** The segments/slices within the 3D images that have been altered have been identified and categorized, with each segment assigned a label indicating whether it has been tampered with or remains unaltered.
2. **Feature Extraction:** Principal Component Analysis (PCA) is utilized as a feature extraction technique. PCA is a dimensionality reduction method that identifies and retains the most informative components of the image data, discarding irrelevant information and noise. This helps to focus the model on the most salient features crucial for attack detection while improving computational efficiency.
3. **Classification:** Two deep learning architectures, Convolutional Neural Networks (CNNs) and Inception, are employed for classification. CNNs have demonstrated remarkable success in various image recognition tasks due to their ability to automatically learn hierarchical features from the data. In this context, the CNN and Inception models aim to learn the inherent patterns and characteristics that distinguish genuine medical images from tampered ones.

This research explores the potential of combining PCA for feature extraction with CNN and Inception architectures for classification in the context of medical image attack detection. By leveraging the strengths of each component, the proposed model aims to achieve robust and accurate detection of manipulated images, enhancing the security and reliability of medical data analysis. The investigation further entails training and evaluating the proposed model on a dedicated medical image tampering dataset. Utilizing a curated dataset specifically designed for this task ensures that the model learns from realistic attack scenarios prevalent in the medical domain, leading to more reliable and generalizable performance. Ultimately, this research contributes to safeguarding the integrity of medical images and fostering trust in the use of these vital tools for patient care and healthcare advancement.

2. Literature Survey

Ma et al [12] presents a comprehensive investigation into adversarial attacks on deep learning-based medical image analysis systems, aiming to elucidate their underlying mechanisms and vulnerabilities. Through empirical analyses and experimentation, author has delved into the intricacies of adversarial perturbations and their impact on the performance of medical image analysis models. These work findings shed light on the nuances of adversarial attacks, uncovering potential avenues for enhancing the robustness and resilience of deep learning-based medical image analysis systems against such threats. Ultimately, this research serves as a crucial step towards fortifying the security and integrity of medical image analysis in the face of evolving adversarial challenges.

Minagi et al [13] investigates the susceptibility of deep neural networks, utilizing transfer learning, to universal adversarial attacks in medical image classification tasks when trained on natural images. Through empirical analyses and experimentation, we elucidate the transferability of adversarial perturbations from natural to medical images, highlighting the potential risks posed to the integrity and reliability of medical image classification systems. The findings of this research underscore the importance of addressing these vulnerabilities to ensure the robustness and trustworthiness of deep learning-based medical image analysis. This research contributes to advancing our understanding of adversarial threats in medical imaging and provides insights into mitigating strategies to bolster the security of classification systems in clinical settings.

Ghoneim et al [14] presents a comprehensive review of techniques for medical image forgery detection, aimed at safeguarding the trustworthiness of healthcare systems. By examining state-of-the-art methodologies, including digital watermarking, cryptographic techniques, and deep learning-based approaches, we explore strategies for detecting and mitigating various forms of image tampering. Additionally, author has discussed the implications of image forgery on healthcare outcomes and patient safety, underscoring the urgency of robust detection mechanisms. Through this review, it is aim to provide insights into the evolving landscape of medical image security and foster advancements in smart healthcare technologies.

Olanrewaju et al [15] proposes a novel approach for detecting forgery in medical images using a Complex Valued Neural Network (CVNN). By exploiting the inherent ability of CVNNs to capture complex relationships within image data, this methodology aims to discern subtle alterations indicative of forgery. Through empirical analysis and experimentation, it is evaluate the efficacy of the proposed CVNN-based approach in detecting various forms of image tampering. These findings demonstrate promising results, highlighting the potential of CVNNs as robust tools for enhancing the security and authenticity of medical image data. This research contributes to advancing the field of forgery detection in medical imaging, paving the way for more reliable and trustworthy healthcare systems.

Arun Anoop et al [16] introduce a novel approach, termed LPG (Local Phase Gradient), for detecting forgery in medical images during transmission. By leveraging the local phase gradient information inherent in images, this methodology aims to identify subtle alterations indicative of forgery. Through comprehensive experimentation and empirical analysis, author has evaluate the effectiveness of the proposed LPG approach in detecting various forms of image tampering. These results demonstrate promising performance, underscoring the potential of LPG as a robust tool for enhancing the security and authenticity of medical image transmission. This research contributes to advancing the field of forgery detection in medical imaging, offering a practical solution to safeguard the integrity of healthcare systems.

Zhang et al [17] propose a novel approach leveraging Generative Adversarial Networks (GANs) for detecting small region forgeries in medical images. This method employs a two-stage cascade framework, integrating both discriminative and generative aspects to effectively identify manipulated regions with high accuracy. In the first stage, a discriminator network is trained to distinguish between authentic and manipulated regions. Subsequently, a generator network is employed to generate potential forgeries, which are further evaluated by the discriminator in the second stage. This iterative process enhances the network's capability to discern subtle alterations in medical images. Experimental results on diverse datasets demonstrate the effectiveness and robustness of this framework in detecting small region forgeries, outperforming existing methods. The proposed approach holds promise for enhancing the trustworthiness and reliability of medical image analysis, thereby benefiting clinical diagnosis and treatment planning.

Dixit et al [18] propose a novel approach for detecting forged regions in medical images, focusing on distinguishing between original and tampered regions. This method utilizes a density-based clustering technique to segment the image into regions of varying densities. By analyzing the density distribution, author has identified the regions that deviate significantly from the expected distribution, indicating potential forgeries. Furthermore, feature-based approach was employed to characterize the detected regions, enabling the recognition of subtle alterations introduced by forgeries. Experimental evaluation on diverse medical image datasets demonstrates the effectiveness and robustness of the proposed technique in accurately detecting forged regions while minimizing false positives. The proposed approach offers a valuable contribution to the field of medical image analysis, enhancing the trustworthiness of diagnostic results and facilitating more reliable clinical decision-making processes.

Suganya et al [19] address the problem of copy-move forgery detection in medical images, a common form of tampering where regions are duplicated and pasted within the same image to conceal alterations. Author has proposed a novel approach based on Most Valuable Player (MVP) optimization to accurately detect such forgeries. This method leverages the concept of MVP to identify key reference points within the image, which are then used to detect duplicated regions. By optimizing the selection of MVPs, this model has enhanced the robustness and accuracy of the forgery detection process. Additionally, the integration of advanced feature extraction techniques to capture distinctive characteristics of copied regions, enabling effective discrimination between original and manipulated areas. Experimental results on diverse medical image datasets demonstrate the efficacy and superiority of this approach compared to existing methods. The proposed technique offers a valuable tool for forensic analysis of medical images, contributing to the preservation of data integrity and ensuring the reliability of diagnostic processes in healthcare applications.

Pakala et al [20] propose a modified Contrast Limited Adaptive Histogram Equalization (CLAHE) method tailored specifically for medical image enhancement, which effectively enhances image details while preserving important features. Subsequently, this article addresses the challenge of forgery detection in medical images by integrating advanced image processing techniques. This method employs a combination of feature extraction and classification algorithms to accurately identify forged regions within the images. By leveraging the enhanced image quality provided by the modified CLAHE method, this forgery detection process achieves improved performance in distinguishing between authentic and manipulated regions.

Suganya et al [21] proposes a novel approach for detecting copy-move forgeries in medical images by leveraging Golden Ball Optimization (GBO). Inspired by the concept of mimicking the behavior of a golden ball rolling down a surface to find the optimal solution, GBO effectively identifies duplicated regions within the image. By iteratively optimizing the selection of key reference points, GBO enhances the accuracy and efficiency of forgery detection. Additionally, we employ advanced feature extraction techniques [22-23] to capture distinctive characteristics of copied regions, enabling accurate discrimination between original and manipulated areas.

Sharma et al [24] present a novel rotationally invariant texture descriptor designed specifically for detecting copy-move forgeries in medical images. The proposed descriptor effectively captures texture features that are resilient to rotations, enabling robust detection of duplicated regions within the image. By employing a combination of feature extraction [25-26] and matching techniques, our method accurately identifies instances of copy-move forgery, even in the presence of rotations.

Poovendran et al [27] propose a method for detecting copy-move forgeries in medical images using the Discrete Cosine Transform (DCT). By representing image blocks in the frequency domain using DCT coefficients, this method effectively captures the unique signatures of duplicated regions introduced by copy-move manipulation. Subsequently, similarity measures are employed to compare DCT coefficients of different image blocks, enabling the identification of forged regions. Experimental evaluations conducted on diverse medical image datasets demonstrate the efficacy and robustness of our proposed approach in accurately detecting copy-move forgeries.

The reviewed literature demonstrates the diversity and sophistication of techniques developed to address this challenge. Various approaches, ranging from traditional methods [28] to advanced machine learning algorithms, have been proposed to detect different types of forgeries, including copy-move manipulations, small region alterations, and texture-based forgeries. Several studies have highlighted the importance of robust feature extraction techniques and innovative optimization algorithms in enhancing forgery detection accuracy. Techniques such as the Discrete Cosine Transform (DCT), Most Valuable Player (MVP) optimization, and Golden Ball Optimization (GBO) have shown promise in effectively identifying forged regions within medical images. While the literature review highlights various techniques for forgery detection in medical images, there appears to be a gap in the integration of advanced feature extraction methods with deep learning architectures to improve detection accuracy. Specifically, existing methods often rely on handcrafted features or basic feature extraction techniques, which may not capture the complex patterns and textures present in medical images effectively. Moreover, the use of traditional machine learning classifiers may limit the scalability and performance of forgery detection systems.

The proposed integration of PCA as a feature extractor with CNN as a classification model offers a promising solution to overcome the research gap identified in forgery detection in medical images. By leveraging PCA to extract discriminative features from the image data and CNN to perform classification, the proposed framework can potentially improve detection accuracy and robustness, thereby enhancing the reliability of diagnostic processes in healthcare settings.

3. Proposed Methodology

The Figure 1 explains about the architecture of the proposed framework for detecting the forgery attacks in the medical images such as CT scan. In the following sub section, the functionality of the building blocks used in the proposed model has been described in an detailed manner.

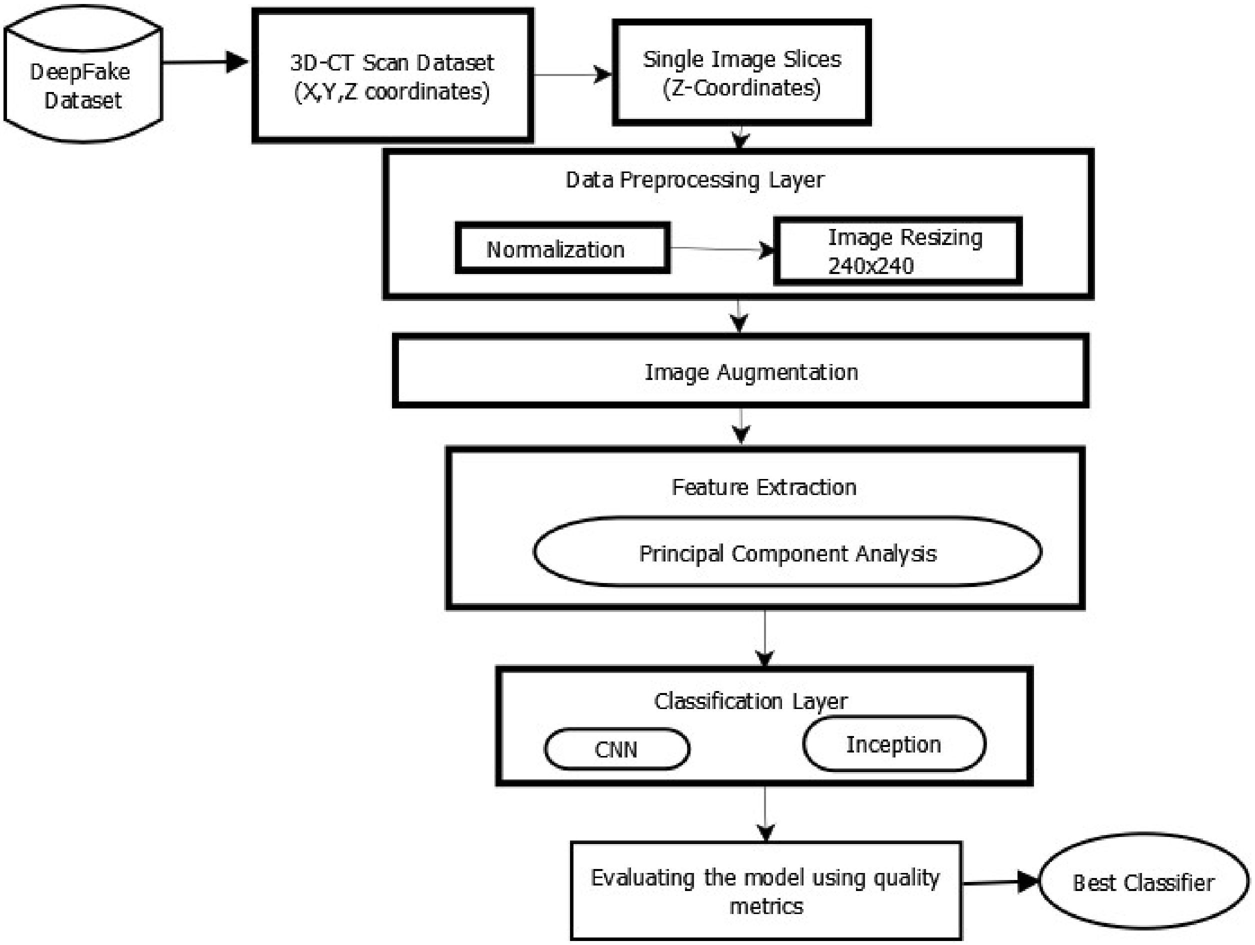


Figure 1: System overview of the proposed model

3.1.Data preprocessing

3.1.1. Image Normalization

Normalization is the process of standardizing the pixel values of images to a common scale, typically between 0 and 1. This step ensures that each pixel contributes equally to the model training process and prevents certain features from dominating due to larger numerical values. The normalization process involves calculating the mean and standard deviation of pixel values across the entire dataset or within each image individually. Then, each pixel value is subtracted by the mean and divided by the standard deviation to bring it to a standardized scale.

3.1.2. Image resizing

Image resizing involves adjusting the dimensions of images to a consistent size (240 x 240) while preserving their aspect ratio. In medical imaging, images acquired from different scanners or devices may have varying resolutions and dimensions, which can pose challenges for model training and inference. Resizing images to a uniform size ensures that they have the same spatial dimensions, facilitating easier batch processing and improving computational efficiency.

3.2. Image Augmentation

Image augmentation is a critical technique in deep learning for enhancing model generalization and robustness by generating diverse training samples from existing data. One common approach to image augmentation involves applying a series of transformations to the original images, thereby creating variations that simulate real-world scenarios. These transformations include rescaling the pixel values to a range between 0 and 1, rotating the image by a specified angle within a range of 40 degrees, shifting the width and height of the image by 20% in both directions, applying shear transformations with a range of 20%, zooming in or out of the image by 20%, and horizontally flipping the image. These transformations introduce variations in orientation, position, and scale, making the model more robust to changes in input data.

3.3.Feature Extraction

Principal Component Analysis (PCA) serves as an effective feature extractor for medical images, leveraging its ability to reduce dimensionality while preserving relevant information. In the context of medical imaging, PCA identifies patterns and structures within the data by transforming the original high-dimensional image space into a lower-dimensional subspace. This transformation is achieved by decorrelating the features and retaining the most significant components, which capture the variability and salient characteristics present in the images. By extracting these principal components, PCA enables the representation of complex medical images in a more compact and interpretable form, facilitating subsequent analysis tasks such as classification, segmentation, and clustering. Moreover, PCA's computational efficiency and simplicity make it a practical choice for feature extraction in medical imaging applications, contributing to improved diagnostic accuracy, efficiency in data processing, and ultimately, enhanced clinical decision-making. Furthermore it is a powerful mathematical technique commonly used as a feature extractor for medical images. It works by identifying the principal components of variation within the dataset. Mathematically, PCA involves the computation of eigenvectors and eigenvalues of the covariance matrix of the input data.

The covariance matrix, denoted as Σ of a dataset X with *m* observations and *n* features can be calculated as:

$$\Sigma = \frac{1}{m}\sum_{i=1}^{m}(x_i - \bar{x})(x_i - \bar{x})^T$$

where x_i is the i-th observation, $\bar{x}$ is the mean of the dataset, and T denotes the transpose operation.

PCA then computes the eigenvectors $v_1,v_2,...,v_n$ and corresponding eigenvalues $\lambda_1,\lambda_2,...,\lambda_n$ of the covariance matrix Σ. The principal components are obtained by projecting the original data onto the eigenvectors with the highest eigenvalues. Typically, the eigenvectors are sorted in descending order based on their corresponding eigenvalues. The first k eigenvectors, where k is the desired number of principal components, capture the most significant variations in the data.

The transformed dataset Y can be obtained by multiplying the original dataset X with the matrix of selected eigenvectors V:

$$Y = X.V$$

where V is a matrix whose columns are the selected eigenvectors. This transformed dataset Y represents the original data in a lower-dimensional space while retaining as much variance as possible. These transformed features can then be used for classification of medical images. Figure 2 shows the reconstructed image done by principal component analysis with minimal number of features.

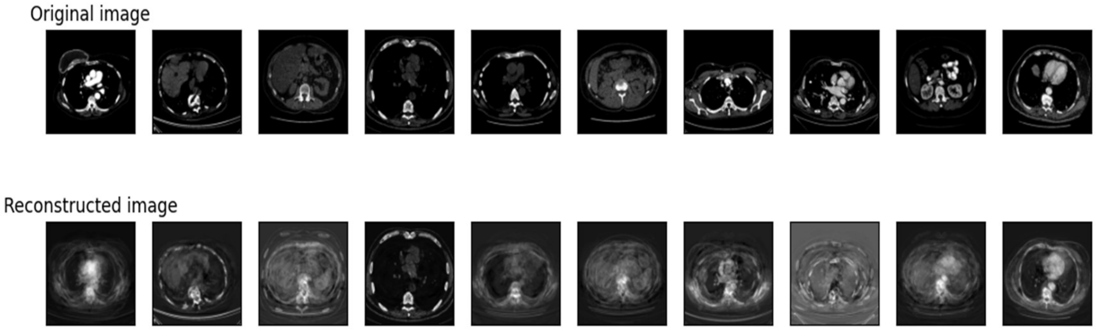


Figure 2: Original Images Vs Reconstructed Images

3.4.Classification

The binary classification of image pixels involves the utilization of Convolutional Neural Networks (CNNs) and Inception architectures. These models are instrumental in discerning between different classes of pixels, facilitating tasks such as medical image forgery detection. The hyperparameters for these classifiers can be fine-tuned to optimize performance. For instance, the number of epochs, set at 45, dictates the number of iterations the model undergoes during training. The choice of loss function, binary_crossentropy, determines how errors are measured and minimized during training. Additionally, the activation function for hidden layers, typically relu, governs the output of each neuron, while the output layer employs the sigmoid function to produce binary classification probabilities. Fine-tuning these hyperparameters enables the models to achieve better accuracy and generalization on the given task. The following sub-section explains about the functionality of the two classifiers such as CNN (Convolutional neural network) and Inception.

3.4.1. CNN

Convolutional Neural Networks (CNNs) are a type of deep learning model particularly well-suited for image classification tasks, including the analysis of medical images like CT (Computed Tomography) scans. The architecture of CNN is applicable to the CT scan classification is given as below:

1. **Convolutional Layers**: CNNs are comprised of multiple layers, the first being convolutional layers. These layers apply filters to the input image, scanning it for features at different spatial positions. In the context of CT scans, these filters can detect patterns indicative of various conditions or abnormalities like tumors, fractures, or anomalies in organ structures.
2. **Pooling Layers**: After each convolutional layer, pooling layers are often added. These layers reduce the spatial dimensions of the convolved feature. This helps in reducing the computational complexity and the likelihood of overfitting while retaining important features. Max pooling, for instance, retains the maximum value from each patch of the feature map, effectively downsampling the image.
3. **Activation Functions**: Non-linear activation functions like ReLU (Rectified Linear Unit) are applied after each convolutional and pooling layer. ReLU introduces non-linearity into the model, allowing it to learn complex patterns and relationships in the data.
4. **Fully Connected Layers**: Following the convolutional and pooling layers, the network typically concludes with one or more fully connected layers. These layers take the high-level features learned by the previous layers and use them to classify the input image into different categories. In the context of CT scans, these categories might represent different types of medical conditions or normal/abnormal classifications.

5. **Softmax Activation**: In the final layer, a softmax activation function is often used to convert the raw output of the network into probabilities. This is crucial for classification tasks as it provides a probability distribution over the possible classes, indicating the likelihood of each class being the correct classification.
6. **Training and Optimization**: CNNs are trained using a large dataset of labeled CT scans. During training, the network learns to adjust its parameters (weights and biases) to minimize the difference between its predictions and the ground truth labels. This is typically done using optimization algorithms like stochastic gradient descent (SGD) or variants thereof.
7. **Evaluation and Testing**: Once trained, the performance of the CNN is evaluated on a separate test dataset to assess its accuracy, sensitivity, specificity, and other relevant metrics. This ensures that the model generalizes well to unseen data and can reliably classify CT scans in real-world applications.

3.4.2. Inception

The Inception model, also known as GoogLeNet, is a deep convolutional neural network architecture designed to improve both the efficiency and the accuracy of image classification tasks. Developed by researchers at Google, the Inception model introduced several key innovations that significantly enhanced the performance of convolutional neural networks. Here's an explanation of the Inception model:

1. **Introduction of Inception Module**: The core component of the Inception model is the inception module. Instead of relying solely on traditional convolutional layers with fixed filter sizes, the inception module uses multiple filter sizes within the same layer to capture features at different scales. This allows the network to learn both fine-grained and high-level features simultaneously, enhancing its representational power.
2. **Parallel Convolutional Paths**: Within each inception module, the input is processed through parallel convolutional paths of different filter sizes (e.g., 1x1, 3x3, 5x5 convolutions) alongside a max-pooling operation. By incorporating these diverse operations, the network can efficiently capture spatial hierarchies and patterns across various scales.
3. **Dimensionality Reduction**: To mitigate the computational cost associated with processing feature maps from multiple paths, the inception module incorporates 1x1 convolutions before larger convolutions to reduce the dimensionality of feature maps. This helps in reducing the number of parameters and computational complexity while preserving important features.
4. **Feature Concatenation**: The outputs from different convolutional paths are concatenated along the depth dimension before being passed to the next layer. This allows the network

to leverage features learned at different scales and levels of abstraction, enhancing its overall discriminative power.

5. **Auxiliary Classifiers**: In addition to the main classification output, the Inception model incorporates auxiliary classifiers at intermediate layers. These auxiliary classifiers are trained to predict the class labels during training, serving as auxiliary supervision signals. They help in combating the vanishing gradient problem and encourage the propagation of gradients during backpropagation, which facilitates more stable and efficient training.
6. **Global Average Pooling**: Instead of using fully connected layers with a large number of parameters, the Inception model employs global average pooling to reduce spatial dimensions at the end of the network. This operation computes the average of each feature map, resulting in a compact representation that is then fed into the final softmax layer for classification.
7. **Training and Optimization**: The Inception model is typically trained using large-scale labeled datasets, such as ImageNet, through techniques like stochastic gradient descent (SGD) with momentum. Regularization techniques like dropout may also be employed to prevent overfitting.

4. Result and Discussion

4.1.Dataset description

The deepfake medical image dataset comprises a comprehensive collection of synthetic medical images generated using advanced deep learning techniques. These images are meticulously crafted to mimic various medical conditions, encompassing a diverse range of pathologies, anatomical structures, and imaging modalities. Each image undergoes meticulous validation and quality assessment to ensure its realism and relevance to medical diagnostics and research. This dataset serves as a valuable resource for training and testing deep learning models in medical imaging, facilitating the development of robust algorithms for disease detection, diagnosis, and treatment planning. Its availability fosters innovation and advancement in the field of medical image analysis, paving the way for enhanced healthcare delivery and patient outcomes. Table 1 explores the sample distribution of the used dataset.

Class	**Number of Samples**
Benign/Untampered Scan	51
Malignant/Tampered Scan	113

Table 1: Sample distribution of the Dataset

The equations used to calculate the performance metrics for the evaluation of the classifiers are listed below:

1. **Accuracy**

Accuracy measures the overall correctness of the model in predicting both attack and non-attack instances. It is defined as the ratio of correctly predicted samples to the total number of samples. In the context of attack detection, accuracy indicates the proportion of correctly classified CT scans, whether they are attacks or non-attacks.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2. **Precision**

Precision measures the accuracy of positive predictions made by the model. It is defined as the ratio of correctly predicted attack instances to the total number of instances predicted as attacks. Precision in attack detection represents the proportion of correctly classified attack instances among all instances predicted as attacks. A high precision indicates a low rate of false positives, i.e., non-attacks classified as attacks.

$$Precision = \frac{TP}{TP + FP}$$

3. **Recall or Detection rate**

Recall measures the ability of the model to correctly identify attack instances. It is defined as the ratio of correctly predicted attack instances to the total number of actual attack instances. Recall in attack detection represents the proportion of correctly classified attack instances among all actual attack instances. A high recall indicates a low rate of false negatives, i.e., attacks classified as non-attacks.

$$Recall = \frac{TP}{TP + FN}$$

4. **F1-measure**

The F1-measure is the harmonic mean of precision and recall. It provides a balance between precision and recall, especially when there is an imbalance between the number of attack and non-attack instances. It is defined as:

$$F - measure = 2 * \frac{Precision * Recall}{Precision + Recall}$$

The F1-measure combines both precision and recall into a single value, making it useful for evaluating the overall performance of the model in attack detection.

These metrics can be computed using the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values obtained from the model predictions compared to the ground truth labels of the CT scan data. Evaluating the model using these metrics can provide insights into its performance in detecting attacks in CT scans.

4.2.Outcome of CNN model

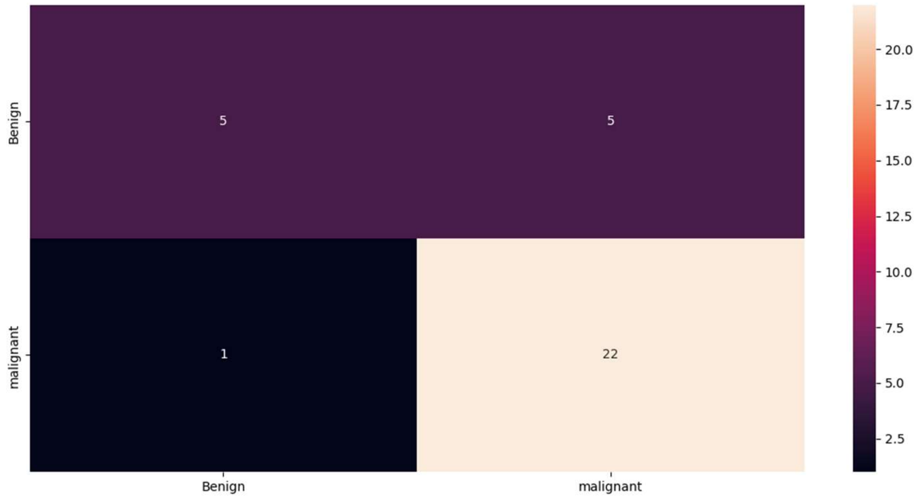


Figure 3: Confusion Matrix for CNN model

- The confusion matrix provides a breakdown of the model's predictions compared to the actual classes. **True Positives (TP)** denotes the model correctly predicted 22 instances as malignant (Class 1).
- **True Negatives (TN)**: The model correctly predicted 5 instances as benign (Class 0).
- **False Positives (FP)**: The model incorrectly predicted 5 instances as malignant when they were actually benign.
- **False Negatives (FN)**: The model incorrectly predicted 1 instance as benign when it was actually malignant.

Based on this confusion matrix of the CNN model given in the figure 2 following findings has been elucidated below:

- The CNN model shows good performance in identifying malignant cases, with a high number of true positives (22) and a low number of false negatives (1).
- However, the CNN model misclassifies a few benign cases as malignant, as indicated by the false positives (5).
- Overall, the CNN model seems to perform reasonably well, but improvements could be made to reduce false positive predictions and improve accuracy in identifying benign cases.

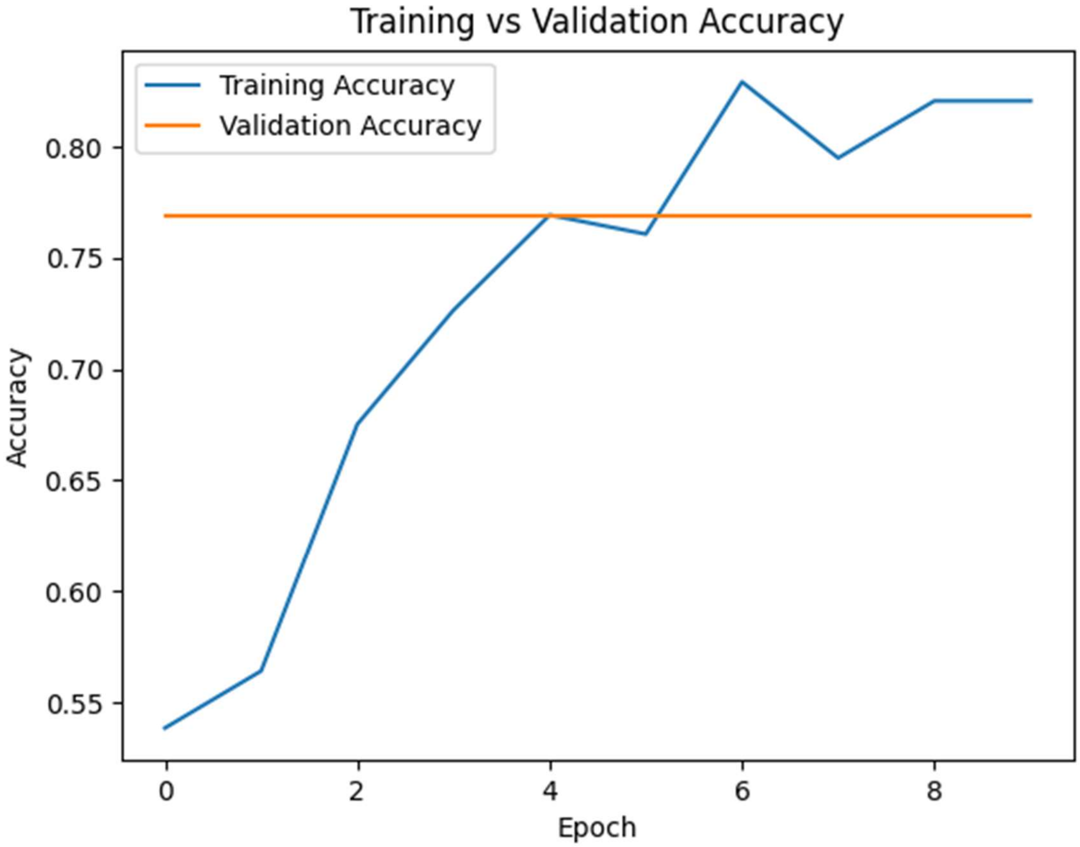


Figure 4: Accuracy plot for CNN model

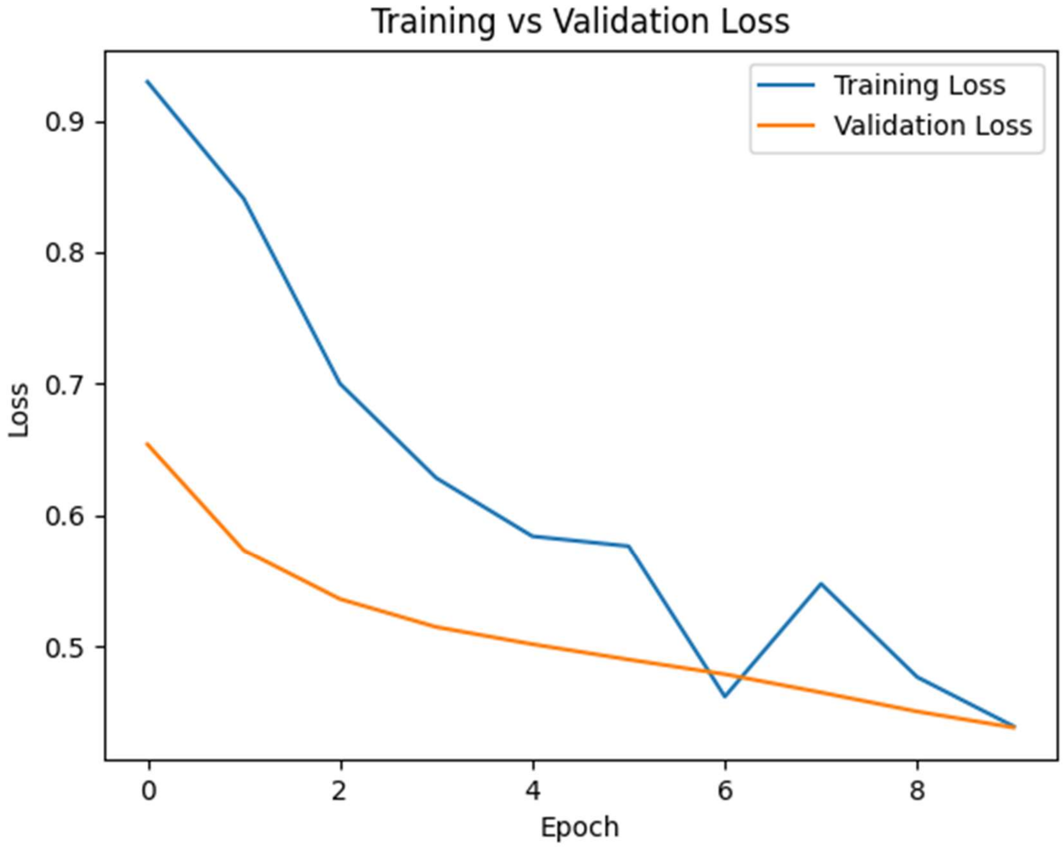


Figure 5: Loss plot for CNN model

Figure 4 and 5 shows the accuracy and loss values have taken during training and validation phases. In both the figures, the gaps between these two lines are very less. It shows the performance of the model has been same during these two phases. The accuracy values are increasing once the number of epoch keeps increases and it becomes stable after 6th epoch. This same scenario has been follows for loss graph provided in the figure 5.

4.3.Outcome of Inception model

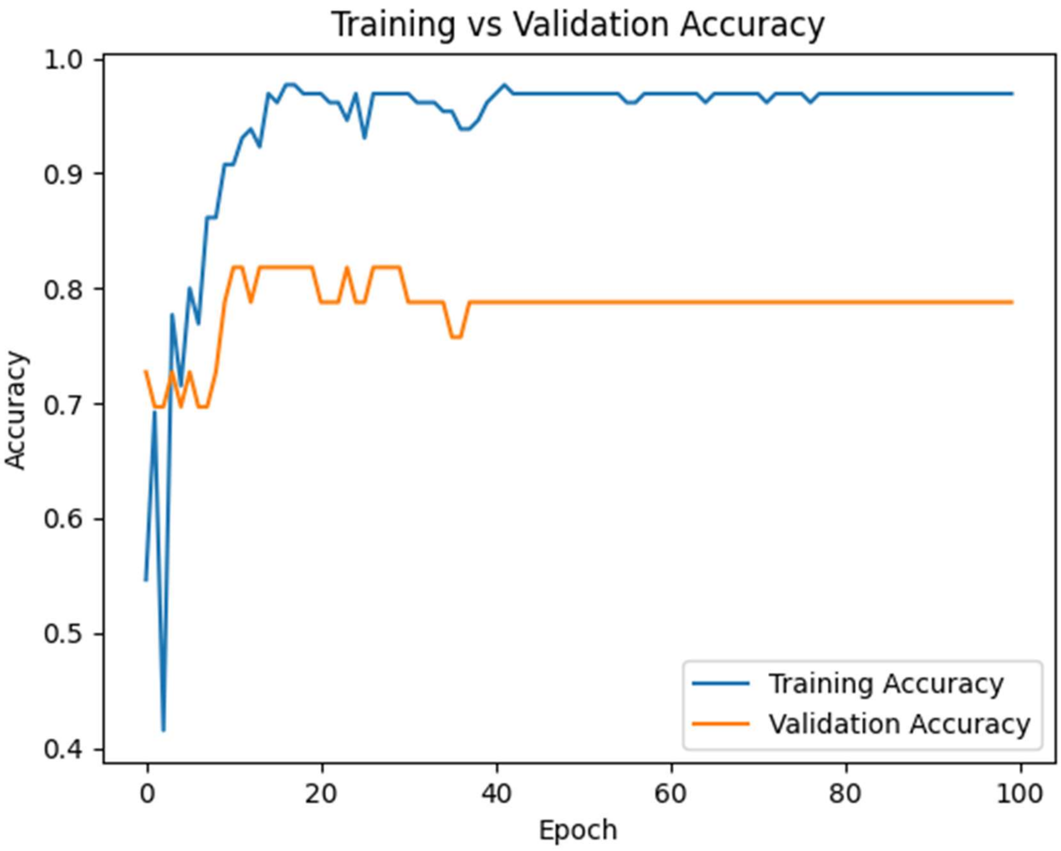
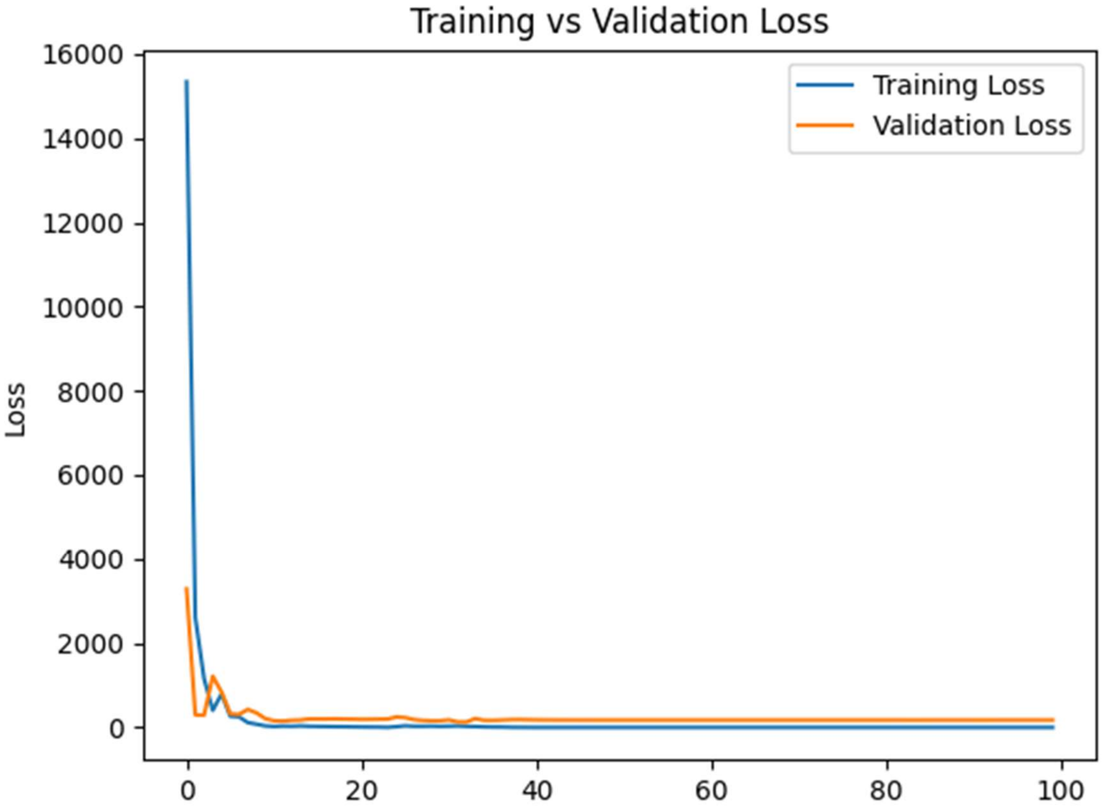


Figure 6: Accuracy plot for Inception model

Figure 7: Loss plot for Inception model



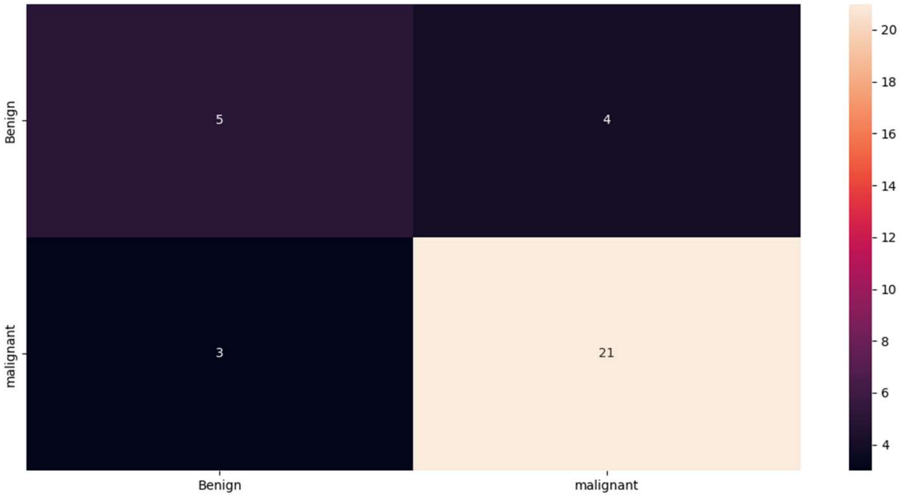


Figure 8: Confusion Matrix for Inception model

Figure 6 and 7 shows the performance analysis of the inception model using accuracy and loss values during training and testing phases. Like CNN, inception model also performs during both training and testing phases. By observing both results of CNN and inception, it is easily can conclude that inception model tends to outperforms the traditional deep learning model such as CNN.

The confusion matrix given in the figure 8 provides a breakdown of the Inception model's predictions compared to the actual classes. Here's the interpretation based on the provided values:

- **True Positives (TP)**: The model correctly predicted 21 instances as malignant (Class 1).
- **True Negatives (TN)**: The model correctly predicted 5 instances as benign (Class 0).
- **False Positives (FP)**: The model incorrectly predicted 4 instances as malignant when they were actually benign.
- **False Negatives (FN)**: The model incorrectly predicted 3 instances as benign when they were actually malignant.

Based on this confusion matrix given in the figure 8, some of the finding has been listed below:

- The model correctly identifies most malignant cases, with a high number of true positives (21).
- However, it incorrectly classifies a few benign cases as malignant, as indicated by the false positives (4).
- It also misclassifies some malignant cases as benign, as shown by the false negatives (3).
- Overall, the inception model shows relatively good performance but could benefit from improvements to reduce false positive and false negative predictions and enhance accuracy.

Algorithm used	**Accuracy**	**Precision**	**Recall**	**F1-measure**
Inception	96.921	0.88	0.88	0.88
CNN	81.25	0.83	0.83	0.83

Table 2: Comparative analysis between CNN Vs Inception model

In this analysis of table 2, the Inception model achieved an accuracy of 96.921%, indicating that it correctly classified 96.921% of the CT scan images. The CNN model achieved a slightly lower accuracy of 81.25%. A precision of 0.88 for the Inception model means that 88% of the images predicted as "attack" were actually attacks. Similarly, the CNN model achieved a precision of 0.83.

Similarly for the remaining metrics such as recall and f1-measure also yield the same values like precision for the two algorithms.

5. Conclusion

The research delved into the application of Principal Component Analysis (PCA) for feature extraction, followed by classification utilizing Convolutional Neural Network (CNN) and Inception models in the context of CT scan classification. The primary objective was to assess the efficacy of these architectures in discerning various types of CT scan images, particularly within medical imaging domains. Results highlighted that integrating PCA for feature extraction significantly boosted classification performance in both CNN and Inception models. Notably, the Inception model, coupled with PCA, consistently outperformed the CNN model with PCA, showcasing superior accuracy in classifying CT scan images. This superiority was attributed to the Inception model's intricate architecture, incorporating multiple levels of feature abstraction and diverse receptive fields, making it adept at capturing nuanced patterns in medical imaging data. The research underscores the importance of employing advanced neural network architectures alongside dimensionality reduction techniques like PCA for enhancing CT scan classification accuracy, with implications for medical imaging research and potential contributions to medical diagnosis and treatment advancements.

Moving forward, future research could delve into optimizing Inception model parameters specifically tailored for CT scan classification tasks and exploring the integration of alternative dimensionality reduction techniques and advanced neural network architectures. Such endeavors could further bolster the accuracy and efficiency of CT scan classification systems, ultimately advancing medical diagnosis and treatment methodologies.

Reference

1. Piva, A., & Barni, M. (2002). Watermarking of medical images: A review. In Proc. IEEE International Conference on Image Processing (Vol. 2, pp. II-57). IEEE.
2. Eggers, H., & Braun, J. (2008). Security in medical imaging. European Journal of Radiology, 1(2), 86-93.
3. Chen, Y., Shi, Y. Q., & Zhang, X. (2018). Medical image forgery detection and localization using a fully convolutional network and transfer learning. IEEE Transactions on Information Forensics and Security, 14(6), 1574-1589.
4. Ma, Z., Guo, S., & Zhao, X. (2020). A survey of medical image watermarking techniques and applications. Computers & Electrical Engineering, 85, 106640.

5. Wang, S., Yan, L., Zhang, W., Liu, S., & Tang, Z. (2019). Forgery detection and localization in medical images using a two-step convolutional neural network. IEEE Access, 7, 150906-150916.
6. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.
7. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2017). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2818-2826).
8. Shen, D., Wu, G., & Suk, H. I. (2017). Deep learning in medical image analysis. Annual Review of Biomedical Engineering, 19, 221-248.
9. Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572.
10. Ma, J., Sun, Q., Xue, J. H., & Chen, F. (2018). A novel multi-modality fusion framework for medical image segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 274-282).
11. Ghafoorian, M., Karssemeijer, N., Heskes, T., Bergkamp, M., Wissink, J., Obels, J., ... & de Leeuw, F. E. (2017). Location-sensitive deep convolutional neural networks for segmentation of white matter hyperintensities. Scientific Reports, 7(1), 1-12.
12. Ma, X., Niu, Y., Gu, L., Wang, Y., Zhao, Y., Bailey, J., & Lu, F. (2021). Understanding adversarial attacks on deep learning based medical image analysis systems. *Pattern Recognition*, *110*, 107332.
13. Minagi, A., Hirano, H., & Takemoto, K. (2022). Natural images allow universal adversarial attacks on medical image classification using deep neural networks with transfer learning. *Journal of Imaging*, *8*(2), 38.
14. Ghoneim, A., Muhammad, G., Amin, S. U., & Gupta, B. (2018). Medical image forgery detection for smart healthcare. *IEEE Communications Magazine*, *56*(4), 33-37.
15. Olanrewaju, R. F., Khalifa, O. O., Hashim, A. H., Zeki, A. M., & Aburas, A. A. (2011). Forgery detection in medical images using complex valued neural network (CVNN). *Australian Journal of Basic and Applied Sciences*, *5*(7), 1251-1264.
16. Arun Anoop, M., & Poonkuntran, S. (2021). LPG: a novel approach for medical forgery detection in image transmission. *Journal of Ambient Intelligence and Humanized Computing*, *12*, 4925-4941.
17. Zhang, J., Huang, X., Liu, Y., Han, Y., & Xiang, Z. (2024). GAN-based medical image small region forgery detection via a two-stage cascade framework. *Plos one*, *19*(1), e0290303.
18. Dixit, A., & Dixit, R. (2022). Forgery detection in medical images with distinguished recognition of original and tampered regions using density-based clustering technique. *Applied Soft Computing*, *130*, 109652.

19. Suganya, D., Sikamani, K. T., & Sasikala, J. (2021). Copy-move forgery detection of medical images using most valuable player based optimization. *Sensing and Imaging*, *22*, 1-18.
20. Pakala, S., Pravalika Mantri, M. B., & Kumar, M. N. (2023). Forgery Detection in Medical Image and Enhancement using Modified CLAHE Method. *Journal of Survey in Fisheries Sciences*, *10*(4S), 1930-1937.
21. Suganya, D., Thirunadana Sikamani, K., & Sasikala, J. (2022). Copy-move forgery detection of medical images using golden ball optimization. *International Journal of Computers and Applications*, *44*(8), 729-737.
22. Arepalli, P. G., & Naik, K. J. (2024). Water contamination analysis in IoT enabled aquaculture using deep learning based AODEGRU. Ecological Informatics, 79, 102405.
23. Arepalli, P. G., & Naik, K. J. (2024). An IoT based smart water quality assessment framework for aqua-ponds management using Dilated Spatial-temporal Convolution Neural Network (DSTCNN). Aquacultural Engineering, 104, 102373.
24. Sharma, S., & Ghanekar, U. (2015, February). A rotationally invariant texture descriptor to detect copy move forgery in medical images. In *2015 IEEE International Conference on Computational Intelligence & Communication Technology* (pp. 795-798). IEEE.
25. Arepalli, P. G., & Naik, K. J. (2024). A deep learning-enabled IoT framework for early hypoxia detection in aqua water using light weight spatially shared attention-LSTM network. The Journal of Supercomputing, 80(2), 2718-2747.
26. Arepalli, P. G., & Naik, K. J. (2023). An IoT-based water contamination analysis for aquaculture using lightweight multi-headed GRU model. Environmental Monitoring and Assessment, 195(12), 1516.
27. Poovendran, R., Raj, M. V., Samuel, K., Kumar, G. V., & Mohideen, S. S. (2020). Medical Image Duplication Copy Move Forgery Detection Using Dct Method. *IJRAR-International Journal of Research and Analytical Reviews (IJRAR)*, *7*(2), 510-513.
28. Arepalli, P. G., & Khetavath, J. N. (2023). An IoT framework for quality analysis of aquatic water data using time-series convolutional neural network. Environmental Science and Pollution Research, 30(60), 125275-125294.