Medical DeepFake Prediction Using Deep Learning (False Cancerous Image Detection)

Gulam Sarwar[1], Rafid Ahmed[2], and Dr. Mohammad Monirujjaman Khan[3]

[1]Department of Electrical and Computer Engineering, North South University, Bashundhara, Dhaka-1229, Bangladesh

[*]Corresponding Author: Mohammad Monirujjaman Khan

Email: monirujjaman.khan@northsouth.edu

Abstract: The term "deep learning," which refers to the underlying artificial intelligence (AI) method, is the source of the phrase "Deepfake." Face swapping in video and digital content is accomplished utilizing deep learning algorithms, which, when provided with huge volumes of data, learn how to resolve problems for themselves. Understanding the security implications is crucial as deep learning and AI advance quickly. Attackers cannot be assumed to have limited capabilities anymore. Another type of machine learning used in the procedure is convolutional neural networks, or Sequential. Deepfake flaws are identified and fixed by GANs over a number of rounds, making it more difficult for deepfake detectors to pick them out. Cancer is the leading cause of death worldwide. Both scientists and doctors are struggling with the challenges of fighting cancer. Early cancer detection offers the best opportunity of saving many lives. Visual inspection and manual techniques are routinely used to diagnose these cancers. Manually assessing medical photographs is time-consuming and highly prone to error.

Keywords: Deepfake, Medical DeepFake, Artificial Intelligence, GAN, Sequential.

1. Introduction

Cancer is characterized by an abnormal proliferation of cells that frequently multiply uncontrollably and, in some situations, metastasis (spread). Cancer is a complex illness. It is a collection of more than 100 separate diseases. Every bodily part is susceptible to cancer, which can take many various forms and affect any tissue. The majority of malignancies are called after the kind of cell or organ in which they first appear. The new tumor that develops from a malignancy that has spread (metastasized) shares the same name as the parent (original) tumor.

Uncontrolled cell development in lung tissues is a symptom of lung cancer. This tumor has the potential to metastasize, infiltrate nearby tissue, and spread outside the lungs. Lung carcinomas, which are formed from epithelial cells, make up the great majority of primary lung malignancies. Lung cancer causes 1.3 million deaths worldwide each year and is the second most common cause of cancer-related death in women and men.

Shortness of breath, coughing (often coughing up blood), and weight loss are the most typical symptoms. 4 Both benign and malignant tumors can be removed; benign tumors do not spread to other bodily areas. Malignant tumors, on the other hand, develop rapidly and infiltrate surrounding tissues, allowing tumor cells to enter the circulation or lymphatic system and spread to new locations throughout the body. The areas of tumor growth at these distant places are known as metastases, and this process of spreading is known as metastasis. Lung cancer is one of the most challenging tumors to cure because it tends to spread, or metastasis, very early in its development. Even though lung cancer

can travel to any organ in the body, the adrenal glands, liver, brain, and bone are the most typical locations for metastasis.

In addition, metastasis from malignancies in other regions of the body frequently occurs in the lung. The cells that make up a tumor metastasis are the same cells that make up the primary tumor. For instance, metastatic prostate cancer in the lung differs from lung cancer if it spreads to the lungs through the bloodstream.

The leading cause of cancer death in both men and women worldwide is lung cancer. According to the American Cancer Society, there will be 213,380 new instances of lung cancer diagnosed in the U.S. in 2007 and 160,390 lung cancer-related fatalities.

Less than 3% of instances of lung cancer occur in those under the age of 45, making lung cancer mostly a disease of the elderly (almost 70% of those diagnosed are over 65).

In terms of the number of cancer-related deaths among women in the United States, lung cancer has also exceeded breast cancer.[1]

2. Methods and Materials

Due to the superior quality of the altered films and the simplicity of its applications for a wide range of users from professionals to beginners with varied levels of computing expertise, deepfakes have become more and more popular. These apps are primarily created utilizing deep learning techniques. Deep learning has a proven ability to represent complex, high-dimensional data. The dataset was gathered from GitHub and Kaggle and then blended to make an acceptable dataset. Both healthy and lung cancer-affected individuals with CT-GAN images were included in the collection. Sequential was used for feature extraction. The model has a flattened layer, a rectified linear unit activation function, three Conv2D layers, four two-thick layers, and one MaxPooling2D layer. the thickest layer, softmax was used as the activation function.

2.1 Tools and Materials

Python is the most effective programming language for data analysis. Due to Python's wide library access, programming with deep learning-based challenges is particularly effective. To make use of a personal GPU for dataset preprocessing, large datasets and model training were handled online using Google Colab, Anaconda Navigator, and Jupyter Notebook. They were also utilized to store all data so that it could be retrieved from any GPU using GitHub. Because it offers a tracking system for collaboration and code management, GitHub is suitable for teamwork.

2.2 Dataset Description

Images of CT-GAN from two classes were part of the dataset. The CT-GAN scans in one class are of Lung Cancer patients, whereas the CT-GAN images in the other class are of healthy patients. There were two subclasses of these classes. A training set is one of them, while a validation set is the other. 2541 photos made up the dataset [2]. A proportion of 75% of the training data and 25% of the test data were used to divide the dataset in this study into training and test sets. Without gathering new data, data augmentation has been used to expand the diversity of data. The CT-GAN pictures of a healthy patient and a patient with cancer are depicted in Figs. 1 respectively.

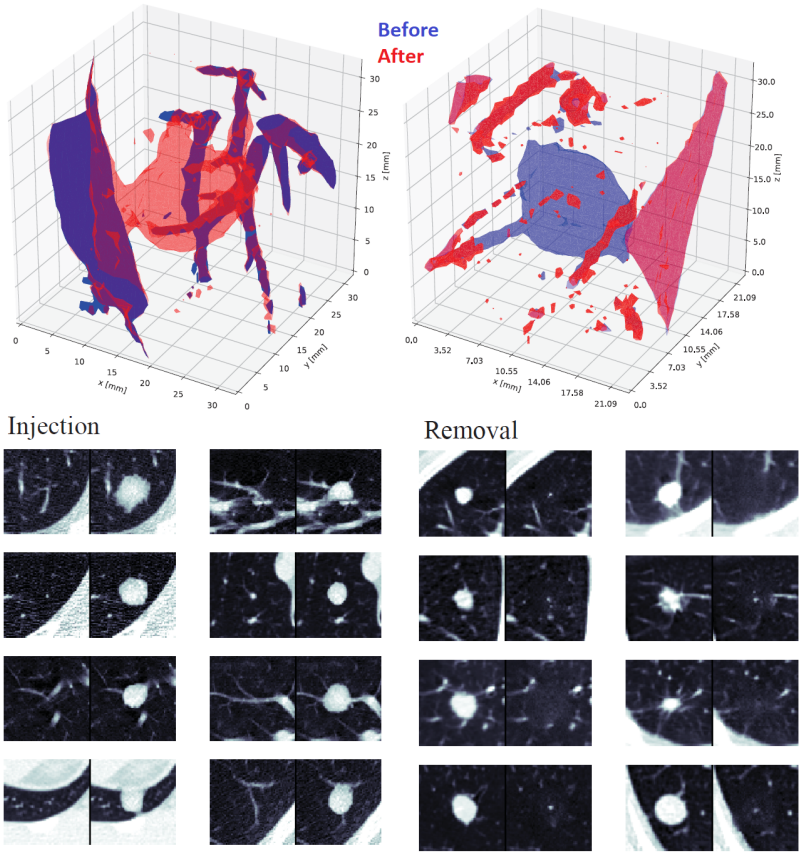

Figure 1: 3D models of injection (left) and removal (right) of a cancerous pulmonary lung nodule.

2.3 Block Diagram

An image of a CANCEROUS CELL from a dataset that is split into patients with CANCER and healthy patients serves as the input in the block diagram of Fig. 3. Before fitting the model, this system underwent preprocessing like loading photographs of a certain size, splitting the dataset, and data augmentation approaches. The model was adjusted and fitted to create more precision. The confusion matrix, model loss, and model accuracy were plotted to show how loss and accuracy change over time. Finally, if the user provides an image of a patient with lung cancer as an input model, we can identify whether the image is present in the output section.

The most understandable representation of the entire system is in a block diagram. The decision-making portion of our system is crucial and significant to our inquiry. The choice is mostly driven by the model, which was trained using a sizable amount of data extracted from LUNG CANCER CELL images.

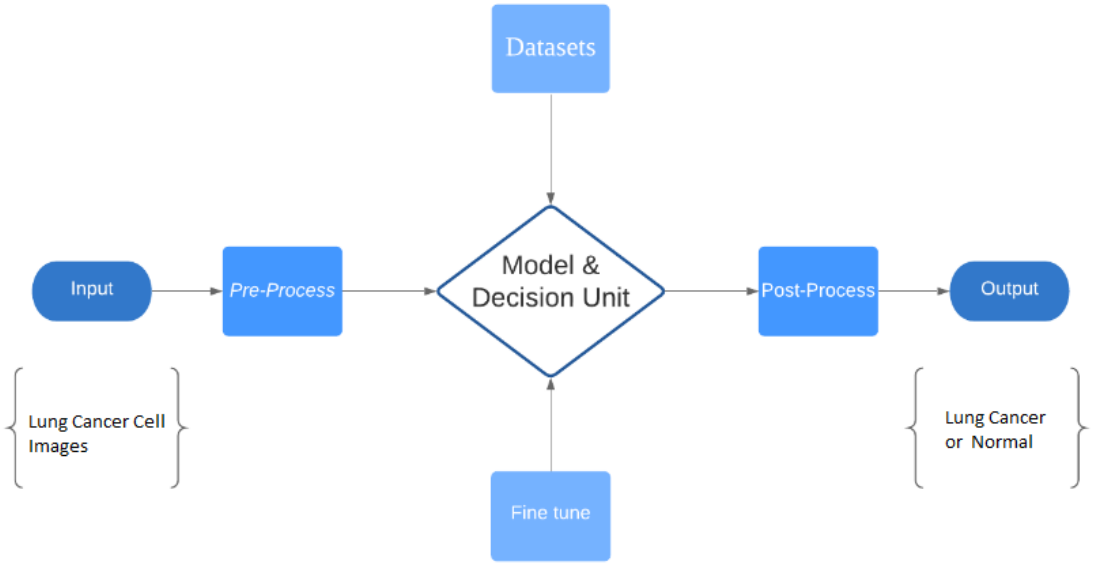

Figure 2: Block diagram of the system

2.4 System Architecture

The system architecture offers an overview of the entire system. In this design, the input is a CT-GAN image, and the output is a forecast of the image. In this case, it will predict whether CANCEROUS IMAGE will have an effect on the image. Three channels with a 224 224 input form are present. In the first two layers of the intended architecture, the filter size is 32 with padding, the kernel size is 3, and the activation function is ReLU. This is followed by the first maxpooling layer, which has a pool size of 2 and strides of 2. The aggregated features are transformed into a single column in the following flat layer. Eventually, two substantial layers appeared. ReLU performs the first one's activation function. The activation function of the least dense layer is softmax, whereas the activation function of the first layer is ReLU. Following preprocessing, the features enter the network. Figure 4 shows the building from above.

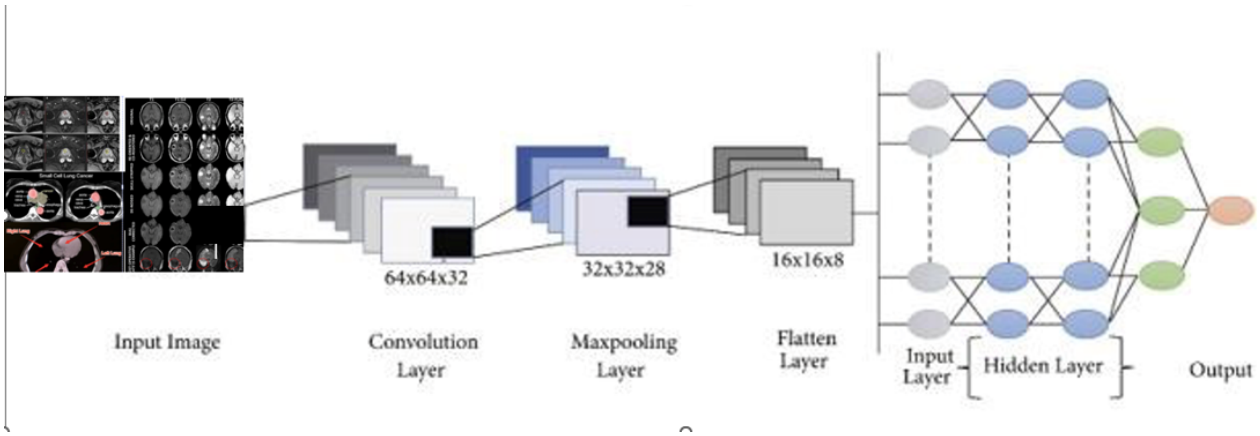

Figure 3: System architecture

2.4.1 Convolutional Layer

The convolutional layer is the base layer of Sequential. The design attributes are decided by this. The input image is given a filter in this layer. The function map is created by convolutional processing the output of the same filters.

A convolution operation multiplies weight sets with the input. An array of input data is combined with a two-dimensional collection of weights to create a filter. A dot product, which produces a single value, is created by multiplying an input and filter patch of equal size. This product is applied between the filter-sized patch of the input and the filter. The filter is smaller but the same size as the input. A filter is used to multiply the input from several places. The filter is designed as a special way to carefully scan the entire image and identify certain components. Assume that the NN input is VRAB, where A is the total number of features, B is the total number of input frequency bands, and A is the number of features that each input frequency band represents. B stands for the size of the filter bank function vector in the context of filter bank features. Assume that v = [v1 v2... vB], where vB is the vector representing the band b function. Calculations for the activations of the convolution layer comprise

$$h_{j,k} = \theta\left(\sum_{b=1}^{s} w_{b,j}^{T} v_{b+k-1} + a_j\right), \quad (1)$$

where aj is the bias of the jth feature map, wb,j is the weight vector for the jth filter's bth band, hj,k is the output of the jth feature map's convolution layer for the kth convolution layer band, s denotes the filter scale, and (x) is the activation function [3].

2.4.2 Pooling Layer

The pooling layer condenses the existence of features by allowing downsampling. It often comes after a convolution layer and has some spatial invariance. The two popular pooling strategies known as average pooling and max pooling, respectively, describe the average presence of a function and the most activated existence of a function [4].

The pooling layer actually strips the photos of pointless elements and turns them into literate visuals. The layer continuously averages the value of the current view when utilizing average pooling. When using maxpooling, the layer continually selects the top value from the active view of the filter. Fewer output neurons are produced as a result of the max-pooling method, which selects only the maximum value using the matrix size specified in each feature map. Consequently, even when the image dramatically reduces, the scenario itself remains the same size. To prevent overfitting, a dropout layer is used, and a pooling layer is essential for reducing the number of feature mappings and network parameters.

One can compute the activation of max pooling as follows:

$$p_{j,m} = max_{k=1}^{r}\left(h_{j,(m-1)(n+k)}\right), \quad (2)$$

Where $p_{j,m}$ is the effectiveness of the jth function map pooling layer and the mth pooling layer band, n is the subsampling factor, r is the pooling scale, which is the number of bands to be pooled together, and n is the subsampling factor.

2.4.3 Flatten Layer

The flattened layer is used to create a single, long, thin one-dimensional feature and to convert matrix data into a one-dimensional array for use in the fully linked layer. It is an option to flatten vectors. The final classification model, also known as a fully connected layer, is then connected to the single vector [5]. All of the pixel data is connected by a single set of fully connected layers. Layer flattening and layer connection are Sequential's last processes. To be prepared for the next completely linked layer of photo classification, it is converted into a one-dimensional array.

2.4.4 Fully Connected Layer

Fully linked layers, which Sequentials mostly use, have shown to be particularly useful for computer vision picture recognition and categorization. Convolution and pooling are the first steps of the Sequential method, which separates the image into features and analyzes them separately [6].

In a completely connected layer, all input is flattened and connected to every neuron. The ReLU activation function is a completely linked layer that is often used. The softmax activation function was used to forecast the output images in the last layer of the fully linked layer. The convolutional neural network's architecture includes a fully linked layer. These are the final and most significant layers of the convolutional neural network.

2.4.5 Pretrained Models

Data is one of the most essential components of deep learning systems, and the lack of medical data or datasets is one of the biggest obstacles for academics in the field of medical research. Labeling and data analysis take a lot of effort and money. The benefit of transfer learning is that it does not require vast datasets. The calculations get cheaper and simpler. Transfer learning is a technique that applies the pretrained model, which was trained on a big dataset, to the new model that needs to be trained, incorporating fresh data that is comparatively smaller than required. This procedure started the Sequential's training for a specific job using a small dataset, incorporating a sizable dataset that had already been trained in the pretrained models [7].

Three Sequential-based pretrained models were used in this work to classify CT-GAN images. The models in use include MobileNet V2, VGG16, and InceptionV3. CT-GAN images were divided into two categories. One is a SARS-CoV-2 patient, and the other is healthy. The transfer learning technique, which was employed in this study, is efficient in terms of training time and can work with little to no data by leveraging ImageNet data. In Fig. 5, which displays the symmetric system architecture of the transfer-learning technique, the symmetric system design is shown.

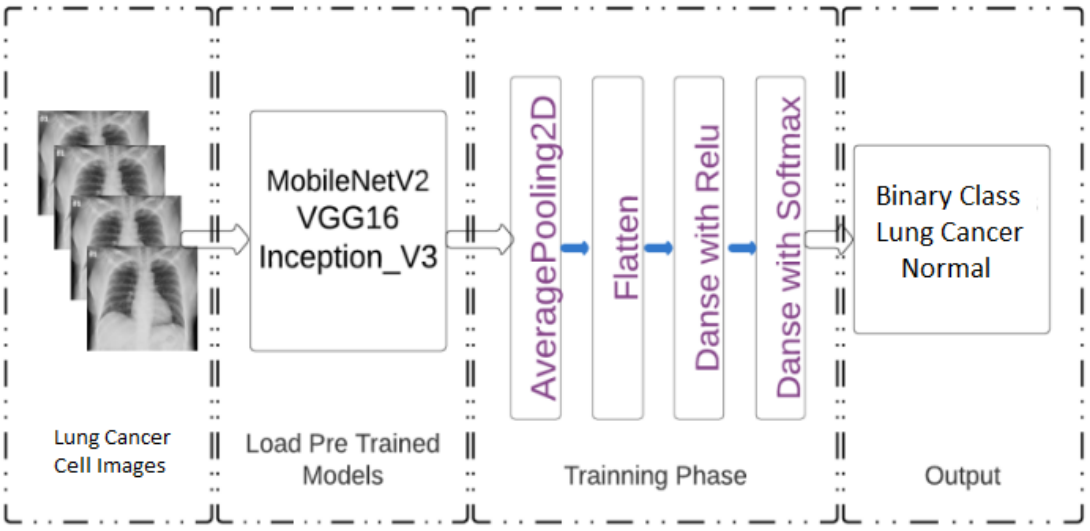

Figure 4: System architecture of the pre-trained model

As shown in Fig. 5, the system architecture is made up of four main parts. The first component is the CT-GAN images, while the second is loading a pretrained model. The second section loads three pre-trained models. In the third stage, the loaded pretrained models were updated using the following layers, as shown in Fig. 5. The final output section will include the results as CANCEROUS -infected and healthy patients.

On numerous assignments and seat stamps across a variety of model sizes, MobileNetV2 enhances the state-of-the-art performance of adaptable models. MobileNetV2 functions as a series of n repeated layers in each line [8]. MobileNet factors the normal form into depthwise convolution using depthwise separable. This suggests a pointwise convolution, often known as a depth of 11 [8]. Another pretrained model employed was InceptionV3. The maximum number of pooling layers is often present. The ability of VGG16 to extract features at low levels with the aid of a small kernel makes it quite useful as well. A compact kernel can effectively extract the features from CT-GAN images [9]. This work used VGG16 with the proper layer addition for the outcome due to the limited dataset [10].

3 Result and Analysis

Our model provided 85.60% accuracy and 82% validation accuracy in the 10th epoch after training it with the train generator, validation generator, step per epoch=8, and 10 epochs. The accuracy of the training was fairly low in the first few epochs, starting at 80%, then changed to 85.600% after the 10th epoch. After the tenth epoch, the validation accuracy decreased to 0.9844 from a starting point of 93%. The train loss was 4% and the validation loss was 6% for VGG16, which has a train accuracy of 98% and a validation accuracy of 98%. This study discovered 88% training accuracy and 91% validation accuracy for ResNet50. According to the study, training losses were 29% and validation losses were 21%. Table 1 provides a history of the four models' accuracy and losses.

Table 1: Accuracy and loss of the model

Model	Accuracy	Validation Accuracy	Loss	Validation Loss
Sequential model	85.60%	82%	0.9999%	1.6%

3.1 Model Accuracy

The accuracy history plot shows that the train's accuracy increased dramatically after each epoch. The accuracy increased with each succeeding epoch after beginning at 77% in the first. The validation accuracy of the model was 94% and increased up to the final epoch. The model accuracy plot displays a line for test accuracy that is always between 94% and 98% accurate and an expanding line for training accuracy. Figures 5 show the model accuracy and model loss, respectively.

Figure 5: (a) Model accuracy and **(b)** Model loss

3.2 Confusion Matrix

The systems showed a confusion matrix, where the columns represented actual values and the rows represented expected values. A classification model's confusion matrix summarizes the results of the predictions. The confusion matrix is used to sum up and break down correct and incorrect predictions by class, and Eqs. (3), (4), (5), and (6) are used to create the forn n matrices FP, FN, TP, and TN [11].

$$TP_i = a_{ii} \tag{3}$$

$$FP_i = \sum_{j=1, j\neq i}^{n} a_{ji} \tag{4}$$

$$FN_i = \sum_{j=1, j\neq i}^{n} a_{ij} \tag{5}$$

$$TN_i = \sum_{j=1, j\neq i}^{n} \sum_{k=1, k\neq i}^{n} a_{jk} \tag{6}$$

When analyzing errors, three terms are crucial. Predictions, information, and features like these. A confusion matrix can be used to do prediction-based error analysis, which can be represented by the proportion of true positives, true negatives, false positives, and false negatives. The type and size of the data are also crucial for error analysis. The training and test sets may have a significant impact on

the outcomes, so splitting the data appropriately for trains and tests is important for error analysis. Features are essential to error analysis. To cut down on errors, regularization and feature engineering were also used.

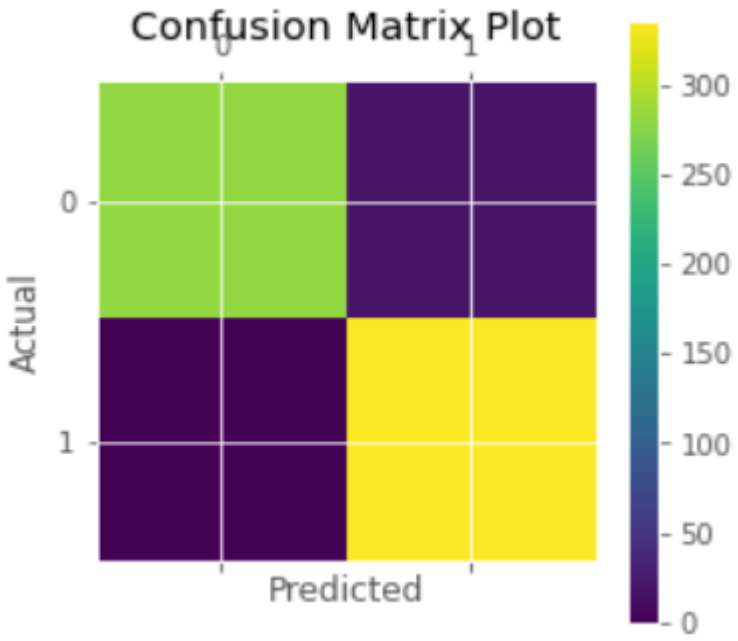

Figure 6: Confusion Matrix

3.3 Model Evaluation

The models' performance analysis is evaluated based on their accuracy, precision, recall, and F1-score. The performance of the proposed model was assessed using the terms true positive (TP), false positive (FP), true negative (TN), and false negative (FN). The frequency at which the impacted photos can be precisely distinguished from all other images is known as recall, also known as sensitivity. Precision is the opposite of recall. By combining the precision and recall measures, the F1-score illustrates the frequency with which the predicted value is accurate. It is frequently referred to as the harmonic mean of p and r in mathematics. There are given the equations below.

Matrix analysis can be used to evaluate a system's performance both before and after it has been modelled. An indicator of how well a model or system performs should be calculated based on how frequently the model predicts the actual outcome. The accuracy is computed using Eqs. 7 and 8 of the mathematical formulas [12].

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

(8)

$$accuracy = \frac{correctpredictions}{totalnumberofexample}.$$

Recall, also known as sensitivity, is the frequency with which the true value can be identified among a set of all possible values. With the use of the expression in Eq. (9) [13], recall may be calculated.

$$re - call = \frac{TP}{TP+FN}. \quad (9)$$

Precision is the quantity of accurate identifications. You may calculate the number of times the model's positive forecast was accurate using the mathematical formula Eq. This is more relevant to the model's positive identification (10).

$$precision = \frac{TP}{TP+FP}. \quad (10)$$

A single matrix that characterizes precision and recall, the F1-score, can be used to summarize the classifier's performance for both recall and precision. In other mathematical terms, it is also referred to as a harmonic means of recall and precision. Using Eq, the F1-score is computed (11).

$$f1 - score = \frac{2pr}{p+r}. \quad (11)$$

TP stands for true positive, FP for false positive, and FN for false negative in Equations 3 and 4. In Eq. 10, precision and recall are denoted by the letters p and r, respectively. Table 2 provides model evaluation results for our customized Sequential models, MobileNetV2, VGG16, and InceptionV3, which include precision, recall, and F1-score. It demonstrates that MobileNetV2 and InceptionV3 have superior precision, recall, and F1-score ratings to the other models. Table 1 displays how well InceptionV3 performs in comparison to other models and how accurately it produced higher findings.

Table 2: Model Evaluation

Model	State	Precision	Recall	F1-score
Sequential model	Normal	0.86	1.00	0.97

4. Conclusion

Deepfake technology has the power to cast doubt on the veracity of human vision. In the near future, it might lead to significant moral and legal issues. In this work, we present a constructive view of deepfake technology, arguing that rather than being an ethical problem, it may serve to protect privacy and address some ethical issues. By using this technique, medical research could be enhanced and medical video data sharing encouraged. Precision markerless movement tracking is just one example of how it might increase human knowledge.

In this study, we suggest a privacy protection pipeline that preserves original keypoint data. The fundamental concept is to swap the faces, while keeping the body keypoints and even the facial keypoints the same, to de-identify the person. We investigate the dependability of privacy protection, the constancy of essential information, and the permanence of face alteration. The effectiveness of the suggested pipeline in maintaining keypoint position, its resistance to attacks, and its adaptability to various users and recording environments have all been demonstrated.

For this approach to be more widely applicable, additional enhancements are required. First of all, it mainly relies on face detection, which would be useless if it missed even one single face. This occurs when a face is too small for face detectors to detect it but still recognizable to humans. This problem can be fixed by manually searching missed frames for valuable medical videos. The approach may also produce monster faces in multi-person recordings when multiple persons appear in the same image. Additionally, this technique performs poorer on profiles, which harms the continuity of facial data. Additionally, personal information could unintentionally leak due to dress or hairstyle. further anonymization

This is the first work that, to the best of our knowledge, applies deep-fake technology to keypoint invariant de-identification and shows how a face-swapping strategy can enable privacy-preserving data exchange for highly valuable medical films and photos.

Data Availability Statement: You can access the information used to support the study's findings for free at - https://www.kaggle.com/datasets/ymirsky/medical-deepfakes-lung-cancer?select=Response+EXP1+-+AI_patients.csv

Funding Statement: The author(s) received no specific funding for this study.

Conflicts of Interest: The authors would like to confirm there are no conflicts of interest regarding the study.

References

[1] Firaol Lemessa Kitila, “A Brief Review on Lung Cancer” .CODEN (USA)-IJPRUR, e-ISSN: 2348-6465 .://www.researchgate.net/publication/351065253

[2]T. Rahaman, “COVID-19 radiography database,” *Kaggle*, 2020. [Online]. Available: https://www.kaggle.com/tawsifurrahman/covid19-radiography-database

[3] O. A. Hamid, L. Deng, D. Yu, "Exploring convolutional neural network structures and optimization techniques for speech recognition,”*ISCA,*Vol. 11, pp. 73-5, 2013. Available: https://www.microsoft.com/en-us/research/publication/exploring-convolutional-neural-network-structures-and-optimization-techniques-for-speech-recognition/

[4] J. Brownlee, "A gentle introduction to pooling layers for convolutional neural networks", *Machine Learning Mastery*, 2021. Available: https://machinelearningmastery.com/pooling-layers-for-convolutional-neural-networks/

[5] J. Jeong, "The most intuitive and easiest guide for CNN," *Medium*, 2021. Available: https://towardsdatascience.com/the-most-intuitive-and-easiest-guide-for-convolutional-neural-network-3607be47480

[6] S. Saha, "A comprehensive guide to convolutional neural networks—the ELI5 way," *Medium*, 2021.

Available:https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

[7] I. Apostolopoulos and T. Mpesiana, "Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks," *Physical and Engineering Sciences in Medicine*, vol. 43, no. 2, pp. 635-640, 2020. [Online]. Available: 10.1007/s13246-020-00865-4

[8] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and C. L. Chieh, "MobileNetV2: inverted residuals and linear bottlenecks," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 4510-4520, 2018.

[9] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang *et al.,*"MobileNets: efficient convolutional neural networks for mobile vision applications,"*arXiv,*pp.1-7, 2017. Available: https://arxiv.org/abs/1704.04861

[10] C. Sitaula, and M. B. Hossain, "Attention-based VGG-16 model for COVID-19 chest X-ray image classification," *Springer Science+Business Media, LLC, part of Springer Nature,* vol. 51, no.5, pp. 2850-2863, 2020.

[11] K. Simonyanand A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv* pp.1-8*,* 2015. Available: https://arxiv.org/abs/1409.1556v4

[12] M. S. Junayed et al., "AcneNet - A Deep CNN Based Classification Approach for Acne Classes," *2019 12th International Conference on Information & Communication Technology and System (ICTS)*, 2019, pp. 203-208.

[13] P. D. Ailab, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness& correlation," *Machine Learning Technologies,*vol.2. pp. 37-63, 2011..Available: http://www.bioinfo.in/contents.php?id=51

[14] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and f-score, with implication for evaluation," *Lecture Notes in Computer Science*, pp. 345-359, 2005. Available: 10.1007/978-3-540-31865-1_25