Deepfake Detection for **Chest X-Ray Images**

Made By-
Shatakshi 20BAI1314
Agrim 20BAI1168
Sriganesh 20BAI1181

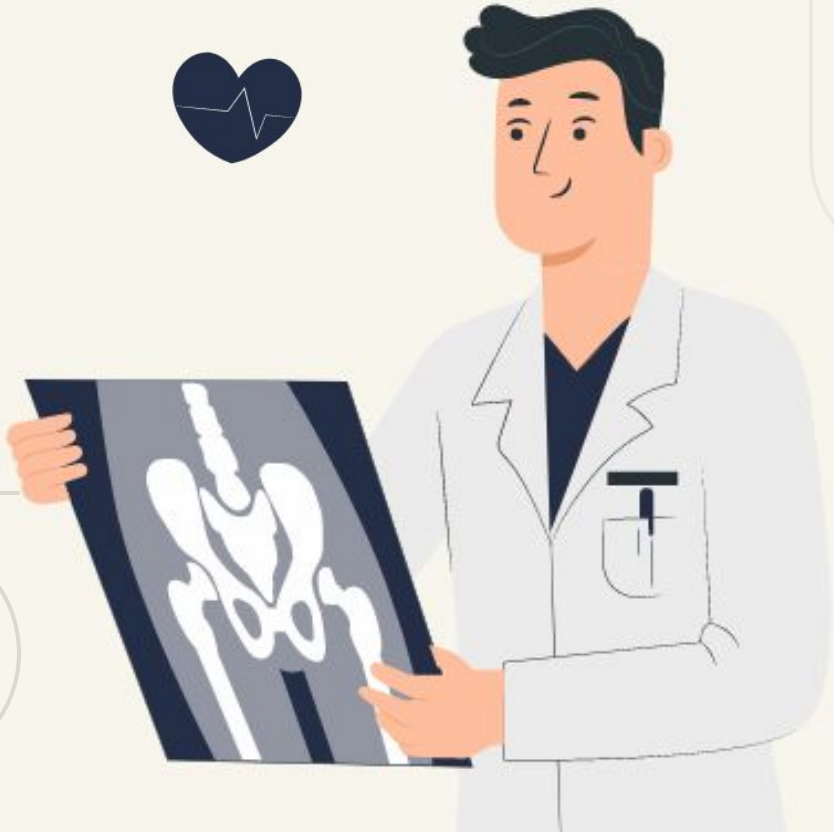


Table of contents

01 **Abstract**
Concise summary and overview of project

02 **Research gap & Novelty**
Unexplored and insufficiently addressed area

03 **Literature Survey**
Review of existing works

04 **Modules & Methodology**
Approach and techniques

05 **Results & Discussion**
The findings of our project at end

06 **Conclusion**
Summing up

Problem Statement

Deep generative networks in recent years have reinforced the need for caution while consuming various modalities of digital information. One avenue of deep fake creation is aligned with injection and removal of tumors from medical scans. Failure to detect medical deepfakes can lead to large setbacks on hospital resources or even loss of life. Therefore, the project is intends for to detect such deep fakes in chest X-ray images.

01

Abstract

Medical imaging is the non-invasive process of producing internal visuals of a body for the purpose of medical examination, analysis, and treatment. Numerous attacks on clinics and hospitals occurred in 2018, resulting in serious data breaches and delays in medical services. When an attacker has access to medical records, they are able to do far more than just demand a ransom or sell the information. One method of creating deepfakes is to inject and remove tumours from medical imaging. Failure to recognise medical deepfakes might result in significant losses of hospital resources or even death. Hence, we will attempt to build a machine learning model and train it to carry out detection between original images and deep fakes created and analyse them.

02 Research Gap & Novelty

The massive availability of applications and tools that create deep fake images and videos lead to large numbers of deep fake images and videos generated every day. It has become a great challenge for academic researchers when studying and analyzing deep fake images and videos. One of the most important challenges facing the researchers is the lack of high-quality dataset. The current deep learning methods are facing a scalability issue. On other words, the deepfake methods uses fragmented data sets to detect face swapping. However, applying their models in large scale datasets may produce unacceptable results

03
Literature Survey

1. Machine learning based medical image deepfake detection

This paper attempts to address the detection of such attacks with a structured case study. Specifically, eight different machine learning algorithms are evaluated, which include three conventional machine learning methods (Support Vector Machine, Random Forest, Decision Tree) and five deep learning models (DenseNet121, DenseNet201, ResNet50, ResNet101, VGG19) in distinguishing between tampered and untampered images. The findings of this work show near perfect accuracy in detecting instances of tumor injections and removals.

2. DeepFake knee osteoarthritis X-rays from generative adversarial neural networks deceive medical experts and offer augmentation potential to automatic classification

This study presents generative adversarial neural networks capable of generating realistic images of knee joint X-rays with varying osteoarthritis severity. It offers 320,000 synthetic (DeepFake) X-ray images from training with 5,556 real images. the models are validated regarding medical accuracy with 15 medical experts and for augmentation effects with an osteoarthritis severity classification task. Survey of 30 real and 30 DeepFake images for medical experts is devised. The result showed that on average, more DeepFakes were mistaken for real than the reverse. The result signified sufficient DeepFake realism for deceiving the medical experts.

3. Deepfakes on Retinal Images using GAN

At present, data driven approaches to classifying medical images are prevalent. However, most medical data is inaccessible to general researchers due to standard consent forms that restrict research to medical journals or education. Our study focuses on GANs, which can create artificial fundus images that can be indistinguishable from actual fundus images. Before using these fake images, it is essential to investigate privacy concerns and hallucinations thoroughly. As well as, reviewing the current applications and limitations of GANs is very important.

4. MeDiFakeD: Medical Deepfake Detection using Convolutional Reservoir Networks

Medical Deepfake pertains to application of AI-triggered deepfake technology on to medical modalities like Computed Tomography (CT) scan, X-Ray, Ultrasound etc. The tampering attacks, involve either insertion or removal of certain disease conditions, tumors in/from the modality under analysis. This paper implements and demonstrates a practical, lightweight technique which aims to accelerate deepfake detection for biomedical imagery by detecting malignant tumors injected in modalities of healthy patients. The developed technique makes use of convolutional reservoir networks (CoRN), which enable ensemble feature extraction and results in improved classification metrics.

5. Comparative Analysis of State-of-the-Art Deep Learning Models for Detecting COVID-19 Lung Infection from Chest X-Ray Images

This paper comprehensively evaluates the applicability of the recent top ten state-of-the-art Deep Convolutional Neural Networks (CNNs) for automatically detecting COVID-19 infection using chest X-ray images. Moreover, it provides a comparative analysis of these models in terms of accuracy. This study identifies the effective methodologies to control and prevent infectious respiratory diseases. Our trained models have demonstrated outstanding results in classifying the COVID-19 infected chest x-rays.

6. Deep Learning Model for Computer-Aided Diagnosis of Urolithiasis Detection from Kidney–Ureter–Bladder Images

Kidney–ureter–bladder (KUB) imaging is a radiological examination with a low cost, low radiation, and convenience. Although emergency room clinicians can arrange KUB images easily as a firstline examination for patients with suspicious urolithiasis, interpreting the KUB images correctly is difficult for inexperienced clinicians. Recently, AI-based computer-aided diagnosis (CAD) systems have been developed to help clinicians who are not experts make correct diagnoses for further treatment more effectively. In this study, we proposed a CAD system for KUB imaging based on a deep learning model designed to help first-line emergency room clinicians diagnose urolithiasis accurately.

04

Modules & Methodology

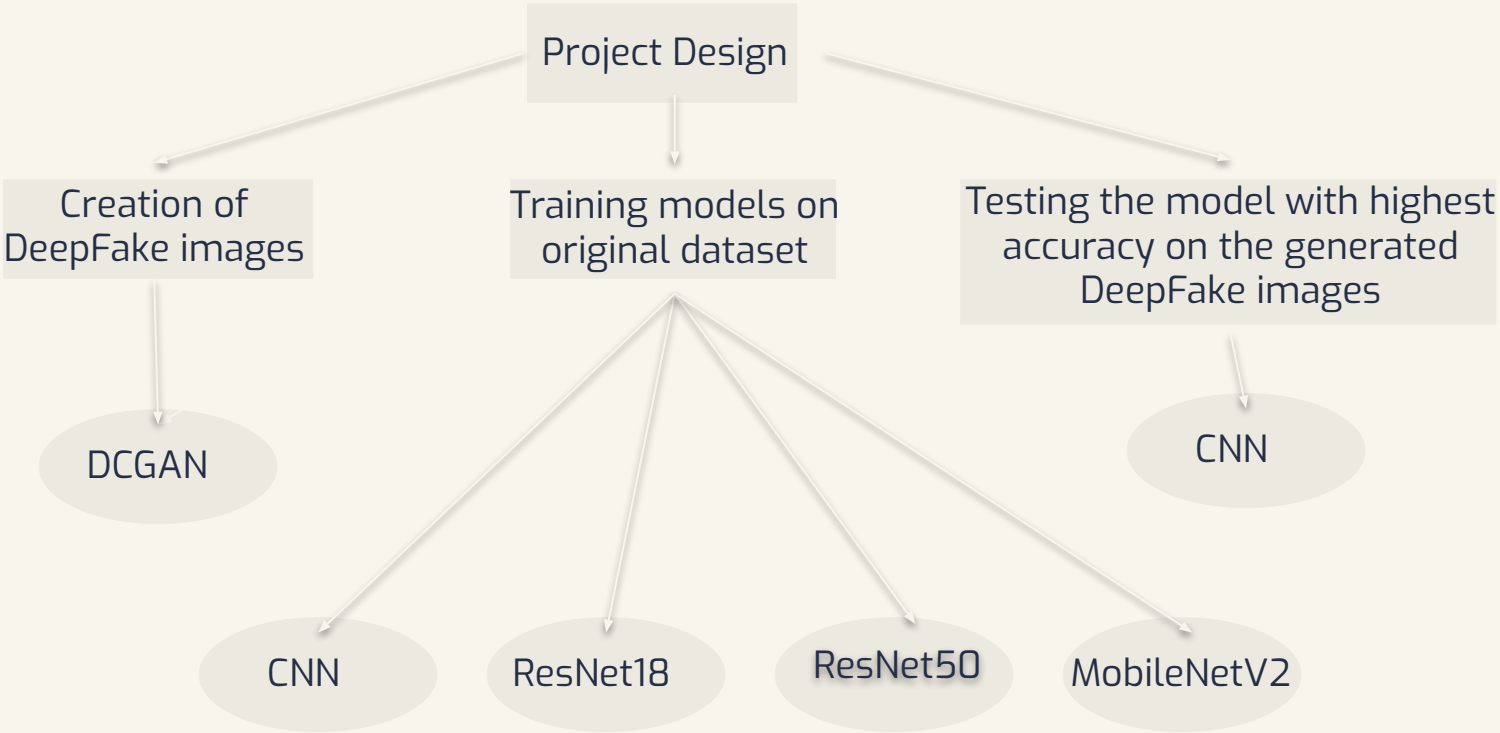
Self contained components & units in the project.

We illustrate the difference between deepfake images and original ones by training a CNN model on both - the original and the deepfake datasets. Hence, the initial part of our project would include the training of an appropriate CNN model on the original dataset.

For the initial training, we plan to train a variant ResNet model on the Chest X-Ray dataset obtained from Kaggle. In ResNets, the technique of skip connections is used. The skip connection connects activations functions of a layer to further layers by skipping some layers in between, and this structure forms a residual block.ResNets are made by stacking these blocks together.

For the creation of deepfakes, we plan to use a Generative Adversarial Network (GAN). GANs are used for teaching a deep learning model to generate new data from that same distribution of training data. They are made up of two different models - a generator and a discriminator.

Design


Project Design
Creation of DeepFake images
Training models on original dataset
Testing the model with highest accuracy on the generated DeepFake images
DCGAN
CNN
CNN
ResNet18
ResNet50
MobileNetV2

05
Results & Discussions

Among the 4 models we trained on the original dataset, our custom made CNN model gave the highest accuracy of around 99.6%. The same model, when trained on a dataset of 500 deepfakes generated by our DCGAN model from the original dataset, gives an accuracy of 50%. Thus, our hypothesis that the accuracy of the model drops steeply when the learning from a set of original models is transferred to a deepfake image was found to be true for the given use case.

06 Conclusion

Building a Convolutional Neural Network (CNN) model from scratch for detecting chest-CT deepfake images is a complex task that requires a deep understanding of both computer vision and deep learning techniques. However, it is possible to create a robust and accurate model by following a series of steps.

First, we collected a dataset of chest-CT images. The dataset was properly curated and labeled to ensure that the model is trained on high-quality data. Next, the CNN model architecture was designed, which involves selecting appropriate layers, filter sizes, activation functions, and pooling methods. This was done based on the specific requirements of the deepfake detection task and the size and complexity of the dataset.

Once the model architecture was finalized, the model was trained using the dataset. The model was validated and tested during the training process to ensure that it is learning the features of the images and is not overfitting or underfitting. Finally, the trained model was evaluated using our test dataset of deepfake images. The model's performance metrics was analyzed to determine its effectiveness in detecting chest-CT deepfake images.

In conclusion, building a CNN model from scratch for detecting chest-CT deepfake images is a challenging task, but we were able to build a model which helped achieve our goal.

References

1. https://www.sciencedirect.com/science/article/pii/S2666827022000263

2. https://www.nature.com/articles/s41598-022-23081-4

3. https://thesai.org/Downloads/Volume13No8/Paper_80-Deepfakes_on_Retinal_Images_using_GAN.pdf

4. https://ieeexplore.ieee.org/abstract/document/9938172

5. https://arxiv.org/abs/2208.01637

6. https://www.sciencedirect.com/science/article/pii/S1746809422002944

Thank You!