SID_2167132

Enhancing and Evaluating the Robustness of Transfer-Learned Models with Ensemble Learning Against Noisy Data and Adversarial Attacks in Lung Tumour Imaging

Access to the code: https://github.com/sarashahin/ML_Research

Table of Contents

1. Abstract

This research inspects the robustness of transfer learned models such as DenseNet121 and ResNet50 against altered noise and adversarial attacks(FGSM) in the diagnosis of lung tumour. One of the leading causes of cancer related deaths is lung cancer and requires advanced diagnostic tools for early detection. My study has evaluated these models under three conditions which are noisy, under adversarial attacks using the Fast Gradient Sign Method (FGSM) and with clean, unaltered test data. DenseNet121 has demonstrated high quality performance in noisy environments with a 74% accuracy suggesting better management of data annotations. Conversely, ResNet50 was more resilient against adversarial attacks showing a 72% accuracy. Both models have indicated comparable efficacy on clean data with DenseNet121 at 78% accuracy. Notably, my innovative, ensemble approach combining these architectures shows improved performance on clean data at 78% accuracy and displays increased resistance to noise with 76% accuracy. It addresses a critical gap in current research in the combined evaluation of transfer learned models under noise, adversarial attacks, and clean conditions specifically in lung tumour imaging. Additionally, the utilisation of ensemble learning methods to enhance model robustness and accuracy in detecting lung tumours under various conditions is a novel approach in this work. This research has established a baseline for clean data performance against the impact of noise and adversarial attacks.

2. Introduction

The introduction of deep learning technology in the fast developing field of medical imaging has extended new possibilities in detection and analysing different health issues, including lung tumours. Lung cancer is one of the major causes of cancer related deaths globally, thus early and accurate detection using imaging is crucial to improve patient outcomes(Lian et al., 2022). Although, the performance of deep learning models in this domain is frequently dependent on their capacity to perform accurately in the face of several issues, including noise in image data.

This study focuses on using transfer learning to adjust pre-trained neural networks, notably ResNet50 and DenseNet models, for lung tumour detection Transfer learning enables us to adapt the knowledge gained by models on large, diverse datasets to our specific problem, thus reducing the need for significant data collection in healthcare and speeding up the development of accurate diagnostic tools(Madhusudan et al., 2023).

Although, medical images regularly contain noise due to different causes, including the image processing and the inherent complexity of human tissues. Noise types such as Gaussian (representing random variation) and salt and pepper (depicting abrupt, occasional disruptions) can have a considerable impact on the clarity of medical images, putting automated systems diagnostic accuracy to the test(HM et al., 2018). To address this issue, I have investigated the adaptability of neural network models(DNN) to several common forms of noise in medical imaging. I have examined noise at varied levels and types into my lung tumour dataset to simulate real world scenarios and evaluate its impact on model performance.

Furthermore, as the reliance on automated systems for diagnosis increases, the security and reliability of these systems against adversarial attacks becomes increasingly important. Adversarial attacks use purposely prepared inputs to cause the model to make mistakes(Ghaffari et al., 2022). I used the Fast Gradient Sign Method (FGSM), a common adversarial attack approach to evaluate my models susceptibility to such attacks emphasising the relevance of robustness in medical imaging analysis.

Ensemble learning combines multiple models to improve diagnostic accuracy in medical imaging. With lung cancer's high death rate, accurate and robust diagnostic tools are crucial. This research evaluates how ensemble models perform under various data conditions(Müller et al., 2022).

In contrast, the existing research is the lack of comprehensive evaluation of transfer learned models like DenseNet121 and ResNet50 against both noisy data and adversarial attacks in lung tumour imaging while past studies have explored aspects of noise and adversarial resilience separately but my work uniquely combines these challenges to reflect real world diagnostic scenarios and highlighting the need for robust diagnostic tools in lung cancer detection.

3. Literature Review

Advancements in Lung Cancer Detection through Deep Learning

Much research has been conducted to identify and diagnose lung illnesses. Radiographic lung images can be complicated and irregular, making it difficult for doctors to diagnose problems such as lesions and vessel nodules(Eslami, M. et al., 2018). Recent innovations have investigated the application of a deep learning ensemble 2D convolutional neural network (VGG16) architecture and transfer learning from large datasets like ImageNet to improve the accuracy to identify lung cancer(Shah et al., 2023; Chaunzwa et al., 2021). These approaches have indicated significant success, the Conventional neural network(CNN) models achieved outstanding accuracy, with one reaching 94.5% based on AUC values.

Challenges of Noise and Data Quality in Medical Imaging

Earlier research, the performance of deep learning models are affected by the quality of input data. While noise in medical images derived from different sources including the image process itself and creating a considerable challenge to the accuracy of automated diagnostic systems. The concept of signal to noise ratio(SNR) was critical in evaluating and optimising the performance of imaging systems, supplying a benchmark to evaluate the impact of noise on image clarity and diagnostic reliability(Wagner et al., 1985).
To address these challenges needs innovative approaches to improve model resilience to data altered.

The Threat of Adversarial Attacks and the Need for Robust Defences

Further study, the sensitivity of CNN models to adversarial perturbations represent a considerable threat to reliability of medical imaging diagnostics. Studies have highlighted the susceptibility of deep learning models to adversarial perturbations, which could lead to misdiagnosis in clinical practice(Apostolidis et al., 2021; Junhao et al., 2023). The advance of adversarial attacks requires the development of robust defence mechanisms. The research has shown the effect of adversarial images on the classification accuracies of DL models trained to classify malignant tumours using three popular oncologic imaging modalities. The computed tomography (CT) model was trained to identify cancerous lung nodules.
Oncologic images were unstable to modest pixel-level alterations. Most oncologic images were misclassified due to a pixel-level disturbance of 0.004 (for pixels normalised to the range of 0 to 1) (CT 25.6%, MRI 6.4% accuracy). Adversarial training increased the stability and robustness of DL models trained on oncologic images compared to naive models ([CT 67.7% v 26.9%], mammography [63.4% vs 27.7%], and MRI [87.2% vs 24.3%])(Joel et al., 2022).

Research Gap

However notable progress was made by applying deep learning to lung cancer detection and also addressing individual challenges for example noise and adversarial attacks, there was a critical gap in evaluating and enhancing the robustness of these models against both types of attacks. This gap underlines research that not only advances the state of the art in lung cancer detection but also ensures the reliability and security of these systems in real world diagnostic scenarios.

4. Methods

4.1. Pre-trained approaches

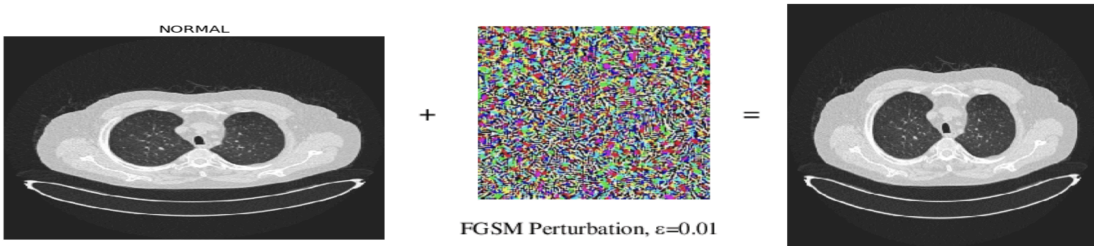
In this study, I have implemented a transfer learning technique using the ResNet50 and DenseNet121 models for lung tumour, a powerful deep learning framework pre-trained on the ImageNet dataset(Pan et al., 2020).The model architecture was built with weights from ImageNet, skipping the top layer to tailor the network to my models. Key to my methodology is the introduction of noises and adversarial attacks through augmentation to test the models resilience against real world variability in medical images.

4.1.1. Preparing and Preprocessing for deep learning models

Divided the dataset into two distinct groups, clean data and noise augmented data. This division was implemented to evaluate the robustness of the model against variations and disturbances in the imaging data. The dataset, comprising 1100 training images and 400 test images across three classes (normal, benign, malignant)
Signal to noise ratio (SNR) as the foundation for my noise introduction strategy. SNR, expressed in decibels (dB) shows the desired level of signal relative to the background noise. It is calculated as the ratio of signal power to noise power, represented in mathematically as SNRdB = 10 $\log_{10}$(P noise/P signal)
The SNR adjusts from a logarithmic scale to a linear scale, enabling direct manipulation of image data(Pizurica et al., 2006).
I have designed a range of Signal to Noise Ratios (SNR) from 28 to 40 to simulate different levels of noise within the images.

For Gaussian and Salt & pepper noises, Gaussian noise specified by a normal distribution around a mean (usually zero), the variance(σ^2) is adjusted according to the desired SNR level. Gaussian noise is computed as Inoisy = I+N(0, σ^2),where I is the original image and N(0, σ^2) represents the Gaussian noise.
Salt and pepper noise is another common artefact in digital imaging and imitated by randomly altering pixels to maximum (salt) or minimum (pepper) intensity values(Gonzalez et al., 2018).

Preprocess these noise models throughout a custom function that randomly applies either Gaussian or Salt and Pepper noise to the input images. Additionally, the FGSM attack function and the adversarial training generate methods designed to make the models more robust particularly against adversarial attacks.
The FGSM function is one of the Adversarial examples in my code. It first calculates the gradient of the loss which measures prediction error with respect to the input image. This gradient shows that the direction in which a small change will increase the loss the most. I then adjust the image slightly in this direction and control it by a small factor epsilon which is about 0.01 to create the adversarial image. The idea is that training the model with these slightly altered images will make it more robust against such attacks in real world scenarios.

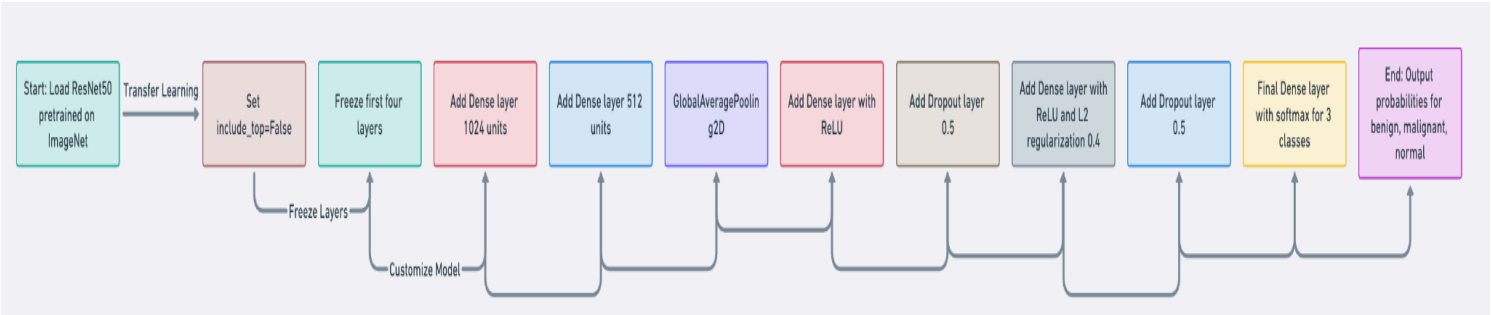


(Figure_1)

4.1.2. Pre-trained ResNet50 with Gaussian and Salt & pepper

The model begins by loading the ResNet50 model pretrained on the ImageNet dataset. Transfer learning utilised the rich feature representations that ResNet50 has learned from a huge and mixed set of images(Zhang et al., 2016). By setting include_top=False, excluding the top fully connected layer. Then proceed to freeze four initial layers of the base model by setting their trainable attribute to True. To customise the model I have added two new dense layers(1024 and 512) on top of the base ResNet50 model. The GlobalAveragePooling2D layer used the feature maps from the convolutional layers into a single 1D vector per map, reducing the number of parameters and helping to prevent overfitting. This is followed by fully connected (Dense) layers with ReLU activation and two dropout layers(0.5) in between to introduce non linearity and further prevent overfitting with regularisation L2(0.4) in between layers. The final Dense layer uses a softmax activation function to output probabilities across the three classes as benign, malignant, and normal(Figure_2).

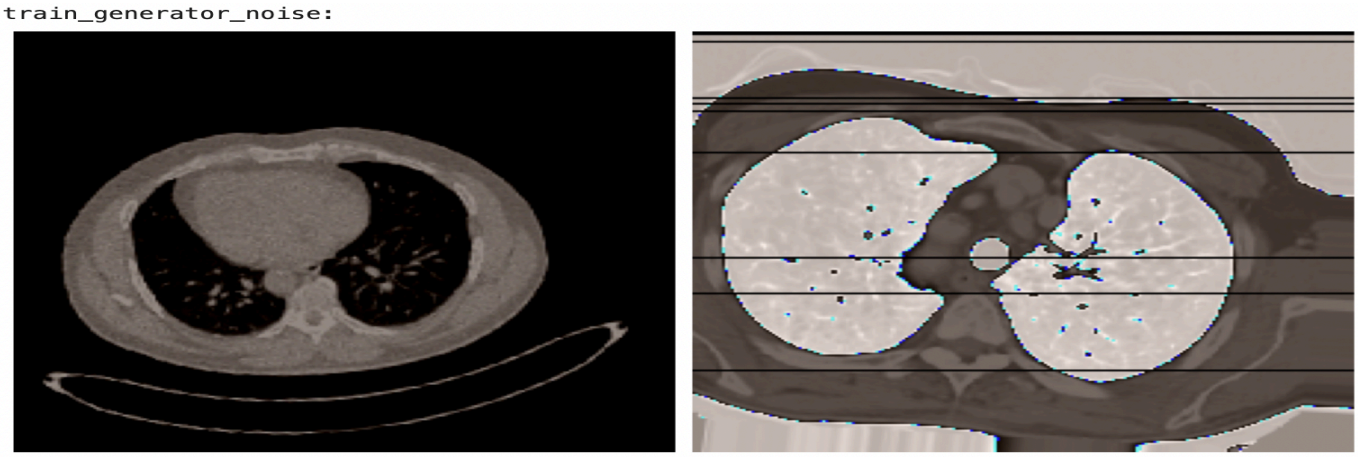
Architecture Design for ResNet50



(Figure_2)

Next step, by calculating signal-to-noise ratio (SNR) mentioned in 3.1.1, added Gaussian, and salt-and-pepper noises to images. These functions artificially introduce noise into training images(Figure_3).

The last step, training the model involves callbacks to optimise the learning process. I used early stopping to prevent overfitting by stopping the training process if the validation loss fails to improve after a certain number of epochs. Additionally, a learning rate scheduler dynamically adjusts the learning rate based on the performance on the validation set, ensuring that the model converges to the best solution efficiently.



(Figure_3)Gausain, Salt & prepare noises

4.1.3. Pre-trained DenseNet121 with Gaussian and Salt & pepper

The model begins by loading the DenseNet121 model pretrained on the ImageNet dataset. Transfer learning utilised the rich feature representations that DenseNet121 has learned from a huge and mixed set of images(Huang et al., 2017). Firstly, I have adjusted the model to freeze the four initial layers trainable. I have added a GlobalAveragePooling2D layer, which simplifies the feature maps into a single vector that means reducing the complexity and the chance of overfitting.
Following this, I incorporated Dense layers with 1024 and 512 neurons to build more capacity to learn from the rich feature set with ReLU activation to incorporate non linear.
Dropout layers in between with a rate of 0.5 which are placed to prevent overfitting by randomly omitting some of the neurons during training. Regularisation was also applied using L2(0.4)to further prevent overfitting. The final Dense layer uses a softmax activation function to output probabilities across the three classes as benign, malignant, and normal.

Next step, by calculating signal-to-noise ratio (SNR) mentioned in 3.1.1, added Gaussian, and salt-and-pepper noises to images. These functions artificially introduce noise into training images(Figure_3).

4.1.4. Pre-trained ResNet50 and DenseNet121 with clean image data and Adversarial attack FGSM

The process is mentioned in 3.1.2 except applying Adversarial attack FGSM on clean images without any noises.
The FGSM function generator creates batches during model training. It uses the FGSM function to alter clean images slightly and make them adversarial.

For each batch(32) of images from the training data then applies the FGSM technique to create adversarial examples.
After that the model trains on these adversarial examples and learns to recognize and correctly classify on clean images.

4.1.5. Pre-trained ResNet50 and DenseNet121 with noisy image data and Adversarial attack FGSM

The process is mentioned in 3.1.2 on noisy image data.

I have trained my model with an adversarial_training_generator function that not only applies the noise but also develops adversarial examples. FGSM adversarial example is altered slightly images which are designed to fool the model by providing a robust training cycle.

4.1.6. Ensemble Learning

The research used pre-trained DenseNet121 and ResNet50 models also evaluating their performance individually and as an ensemble. Models were evaluated on clean, noisy, and adversarially altered and clean test data.

5. Evaluation

5.1. Dataset overview

The dataset is splitted into two training and test sets. The training set includes 1100 images while the test set contains 400 images from Kaggle. These images are labelled as lung tumours into three categories normal, benign and malignant.

Dataset Summary Table

Attribute	Description
Total Images	1,500 (1,100 in train set, 400 in test set)
Classes	Normal, Benign, Malignant
Training Set	1,100 images
Test Set	400 images
Noise Addition	Gaussian noise, Salt-and-Pepper noise
Adversarial Attack	FGSM method applied to both clean and noisy data
Class Weights	Used to balance classes in the training process
Objective	To evaluate the robustness of pre-trained models against noise and adversarial attacks in medical imaging

One of the challenges I have faced with this dataset was the imbalance among classes. To address this imbalance, I calculated class weights that adjust the importance of each class during training.This would verify that the model pays equal attention to all classes(Figure_4).

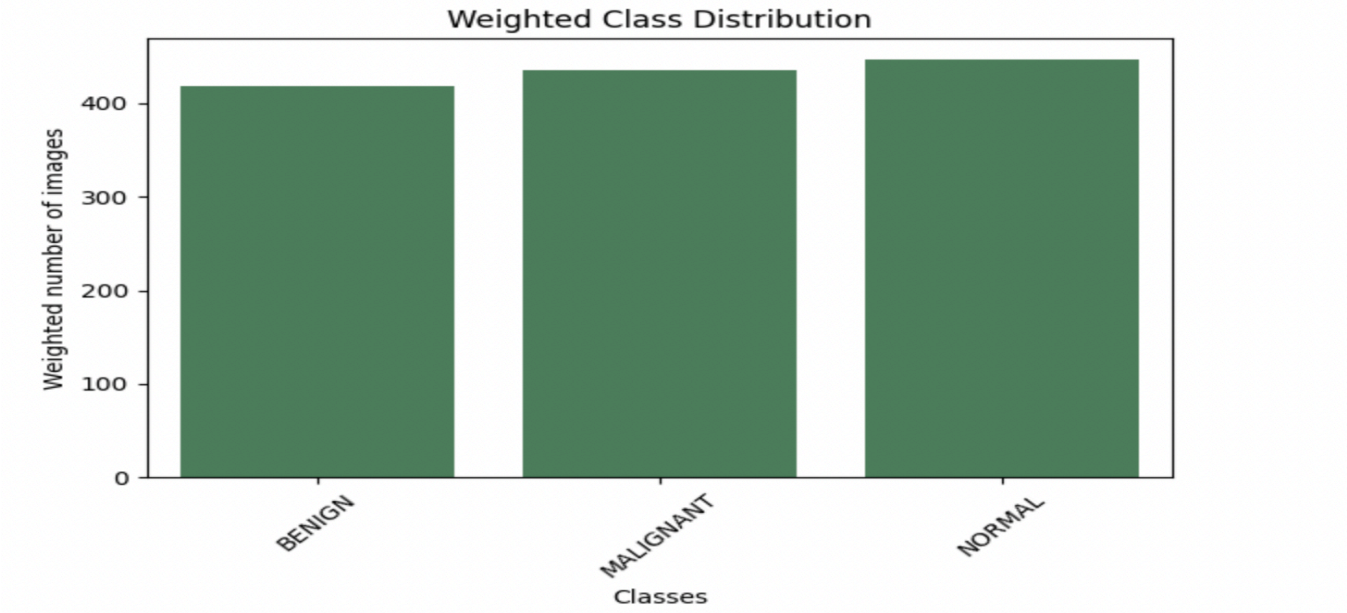


Figure_4

5.2. Results

This section presents the model results by showing the accuracy, recall, precision, Area Under the Curve(AUC), loss and F1-score for each class. To compare the performance of the models.

5.2.1. ResNet50

ResNet50 on Validation Clean dataset as baseline

Accuracy started from 58.18% in the first epoch then it increased to 99.77% by the end which indicates the model is able to correctly identify the classes of the images.

Precision calculates the accuracy of positive predictions. By starting from 61.34% then reached around 100% showing that not mislabeling negative cases as positive.

Recall demonstrates the model able to find all the positive cases. It began at 49.54% and has improved by the end showing the model improved to detect the most actual positives.

Loss metric has decreased significantly from 755.8 to around 0.12 considering the model can learn efficiency.The AUC score is 1.0. This shows that model performance is differentiating between the image classes.

Metric	Training	Validation
Accuracy	99.77%	99.77%
Precision	99.77%	99.77%
Recall	99.54%	99.54%
AUC	1.0000	1.0000
Loss	0.1201	0.1306

ResNet50 on Validation Noise dataset

I have applied two main types of noise which are Gaussian and Salt & Pepper over a range of SNR values from 28 to 40 dB. This table shows how the model validation accuracy is varied with the different types and levels of noise. For instance, it performed very well with a salt_pepper noise and SNR of 40.0 reaching a validation accuracy of 0.9884, and even better with gaussian noise at the same SNR level, reaching an accuracy of 99%.

	Epoch	SNR	Noise_Type	Validation_Accuracy
0	1	40.0	gaussian	0.939954
1	2	30.0	gaussian	0.960739
2	3	30.0	gaussian	0.963049
3	4	32.5	gaussian	0.819861
4	5	32.5	salt_pepper	0.930716
5	6	30.0	salt_pepper	0.939954
6	7	30.0	gaussian	0.981524
7	8	35.0	gaussian	0.930716
8	9	37.5	salt_pepper	0.972286
9	10	35.0	salt_pepper	0.928406
10	11	35.0	gaussian	0.919169
11	12	35.0	salt_pepper	0.969977
12	13	35.0	gaussian	0.969977
13	14	35.0	gaussian	0.824480
14	15	32.5	gaussian	0.916859
15	16	37.5	gaussian	0.960739
16	17	35.0	salt_pepper	0.933025
17	18	35.0	salt_pepper	0.969977
18	19	40.0	gaussian	0.757506
19	20	30.0	salt_pepper	0.939954
20	21	40.0	salt_pepper	0.988453
21	22	40.0	gaussian	0.993072

ResNet50 on Test Clean dataset as baseline

The table has highlighted the model performing very well in detecting malignant cases with a precision and recall of 96% and 100%.

```
Evaluation on Clean Test Data:
10/10 [==============================] - 28s 3s/step
              precision    recall  f1-score   support

      benign       0.71      0.65      0.68       100
   malignant       0.96      1.00      0.98       100
      normal       0.66      0.69      0.68       100

    accuracy                           0.78       300
   macro avg       0.78      0.78      0.78       300
weighted avg       0.78      0.78      0.78       300

Accuracy: 78.00%
```

Evaluation on Clean Test Data with Adversarial Examples FGSM

I have evaluated the model with adversarial attack FGSM modified images which are designed to trick it, the model was correct about 71%. Shows that model with FGSM accuracy dropped from 78% to 71%.

```
Evaluation on Clean Test Data with Adversarial Examples:
10/10 [==============================] - 2s 84ms/step
              precision    recall  f1-score   support

      benign       0.56      0.70      0.62       100
   malignant       0.99      1.00      1.00       100
      normal       0.59      0.43      0.50       100

    accuracy                           0.71       300
   macro avg       0.71      0.71      0.70       300
weighted avg       0.71      0.71      0.70       300

Accuracy: 0.71
```

ResNet50 on Test Noise dataset

The model was evaluated on images with artificially added salt and pepper and Gaussian noise, it achieved an accuracy of 70.67%. This means that even with the altered images the model could correctly identify around 71% of them as their true classes.

Table Summary for Noisy Test Data:

Noise Type	SNR	Accuracy (%)	Precision	Recall	AUC
Salt & Pepper	35.0	70.67	0.7143	0.7	0.8294
Salt & Pepper	35.0	70.67	0.7143	0.7	0.8294
Gaussian	40.0	70.67	0.7143	0.7	0.8294
Gaussian	30.0	70.67	0.7143	0.7	0.8294
...	...	...	...	...	...
Salt & Pepper	35.0	70.67	0.7143	0.7	0.8294

Evaluation on Noise Test Data with Adversarial Examples FGSM

By evaluating the model on adversarial test data which has achieved an accuracy of 72.33% with a precision of 72.58% and a recall of 72%. This indicates that the model is stable in detecting the correct labels even when the data is perturbation of predictions. The AUC (Area Under the Curve) score of 0.8329 shows a good ability to recognise between the classes.

```
10/10 [==============================] - 1s 93ms/step - loss: 1.2466 - accuracy: 0.7233 - precision: 0.7258 - recall: 0.7233 - auc: 0.8329
     noise_type   snr  Accuracy  Precision    Recall       AUC
0   salt_pepper  31.0  0.723333   0.725753  0.723333  0.832861
1      gaussian  34.0  0.723333   0.725753  0.723333  0.832861
2      gaussian  28.0  0.723333   0.725753  0.723333  0.832861
3   salt_pepper  40.0  0.723333   0.725753  0.723333  0.832861
4   salt_pepper  37.0  0.723333   0.725753  0.723333  0.832861
..          ...   ...       ...        ...       ...       ...
327 salt_pepper  37.0  0.723333   0.725753  0.723333  0.832861
328    gaussian  40.0  0.723333   0.725753  0.723333  0.832861
329 salt_pepper  28.0  0.723333   0.725753  0.723333  0.832861
330    gaussian  28.0  0.723333   0.725753  0.723333  0.832861
331 salt_pepper  34.0  0.723333   0.725753  0.723333  0.832861
```

Comparing the performances of two models on noisy and adversarial test data shows that the model on adversarial data has slightly better accuracy, precision, recall, and AUC values than on noisy data.

5.2.2. DenseNet121

DenseNet121 on Validation Clean dataset as baseline

At the start, the accuracy of the model DenseNet121 was lower but as it learned from more data the performance improved, achieving high precision and recall rates.

Epoch	Training Accuracy	Validation Accuracy	Precision	Recall	AUC
1	54.61%	45.03%	58.35%	47.12%	0.73
2	71.20%	77.37%	74.97%	65.90%	0.89
5	90.90%	90.07%	92.22%	88.82%	0.98
10	98.50%	99.08%	98.72%	97.93%	0.99
70 (End)	99.54%	98.61%	99.54%	99.54%	0.99

DenseNet121 on Validation Noise dataset

The DenseNet 121 model shows it can accurately recognize images even when they are noisy. It does best when the noise is not too high or too low like with SNR 40 db but it might struggle a bit more with certain types of noise like Salt & Pepper with SNR 34db.

Epoch	SNR	Noise Type	Validation Accuracy
1	28.0	Gaussian	89.15%
7	40.0	Salt & Pepper	95.15%
15	34.0	Salt & Pepper	68.36%
26	40.0	Gaussian	100%
25	34.0	Salt & Pepper	99.77%

DenseNet121 on Test Clean dataset

The model identifies malignant cases with high precision 97% and recall 98% indicating it's highly reliable for detecting this category.

For benign and normal cases, the model shows balanced performance with precision and recall around 71% and 67%.
The accuracy across all categories averages to 78%.

```
Evaluation on Clean Test Data:
10/10 [==============================] - 3s 159ms/step
              precision    recall  f1-score   support

      benign       0.71      0.65      0.68       100
   malignant       0.97      0.98      0.98       100
      normal       0.67      0.72      0.69       100

    accuracy                           0.78       300
   macro avg       0.78      0.78      0.78       300
weighted avg       0.78      0.78      0.78       300

Accuracy: 78.33%
```

Evaluation on Clean Test Data with Adversarial Examples FGSM

The model identifies malignant cases with a precision of 99% and a recall of 96% showing that it is still reliable for detecting under conditions even with adversarial examples like FGSM.
The model identifying benign cases has decreased with a precision of 56% but a higher recall of 80% showing it can still recognize most benign cases but with more false positives.
The identifying normal cases has notably falled with a precision of 66% and a recall of 39% indicating a number of normal cases are incorrectly classified.

```
Evaluation on Clean Test Data with Adversarial Examples:
10/10 [==============================] - 2s 52ms/step
              precision    recall  f1-score   support

      benign       0.56      0.80      0.66       100
   malignant       0.99      0.96      0.97       100
      normal       0.66      0.39      0.49       100

    accuracy                           0.72       300
   macro avg       0.74      0.72      0.71       300
weighted avg       0.74      0.72      0.71       300

Accuracy: 0.7166666666666667
```

DenseNet on Test Noise dataset

The DenseNet 121 model was tested on noisy test data and identified an overall accuracy of 74.33%. This demonstrates the model is relatively robust in handling images with noise to maintain a decent level of precision and recall across different SNR (Signal to Noise Ratio) levels.

SNR (dB)	Noise Type	Accuracy	Precision	Recall	AUC
34	Gaussian	74.33%	74.15%	72.67%	0.878
31	Salt & Pepper	74.33%	74.15%	72.67%	0.878
40	Salt & Pepper	74.33%	74.15%	72.67%	0.878
...	...	...	...	...	...
37	Gaussian	74.33%	74.15%	72.67%	0.878

Evaluation on Noise Test Data with Adversarial FGSM

The performance of model DenseNet 121 on adversarial test data indicates that it has an accuracy of 66% showing that while the model identifies the correct classes in altered images that the efficacy is reduced compared to clean images.

SNR (dB)	Noise Type	Accuracy (%)	Precision	Recall	AUC
28	Salt & Pepper	66	65.76	64.67	0.806
31	Salt & Pepper	66	65.76	64.67	0.806
40	Gaussian	66	65.76	64.67	0.806
37	Gaussian	66	65.76	64.67	0.806

5.2.3. Ensemble Learning

I have utilised an ensemble of eight pre-trained models, four each from ResNet and DenseNet architectures which trained on clean and noisy data including adversarial examples generated among Fast Gradient Sign Method (FGSM). Ensemble predictions were calculated by averaging the output probabilities from all models.

Condition	Model	Accuracy	Precision	Recall
Clean Data	Ensemble	78.00%	0.80	0.78
Noisy Data	Ensemble	76.33%	0.76	0.76

5.2.4. Comparison of the best models

The structure of architecture of these DenseNet121 and ResNet50 such as how layers are connected and how data pass through the model could affect how they respond to noise and adversarial examples like Fast Gradient Sign Method FGSM.

Both models are excellent on clean data(baseline) conditions that show with high accuracy. DenseNet121 performs slightly better in noisy conditions which indicates better on noise resilience while ResNet50 has an advantage against adversarial attacks.

Summary Table of Models Comparison

Model	Test Type	Accuracy (%)	Precision	Recall	AUC	Notes
DenseNet121	Noisy	74.33	0.741	0.727	0.878	Performed well under noise.
DenseNet121	Adversarial	66.00	0.658	0.647	0.806	Struggled with adversarial attacks.
DenseNet121	Clean Data	78.33	N/A	N/A	N/A	Highest accuracy on clean data.
DenseNet121	Clean Data with Adversarial Examples	71.67	N/A	N/A	N/A	Decreased performance under adversarial examples.
ResNet50	Noisy	70.67	0.714	0.700	0.829	Lower performance under noise compared to DenseNet121.
ResNet50	Adversarial	72.33	0.726	0.723	0.833	Better performance on adversarial data compared to DenseNet121.
ResNet50	Clean Data	78.00	N/A	N/A	N/A	Comparable to DenseNet121 on clean data.
ResNet50	Clean Data with Adversarial Examples	71.00	N/A	N/A	N/A	Slightly lower than DenseNet121 under adversarial examples.
Ensemble	**Clean Data**	**78.00**	**0.80**	**0.78**	**N/A**	**Best performance on clean data.**
Ensemble	**Noisy Data**	**76.33**	**0.76**	**0.76**	**N/A**	**Improved resilience to noise.**

6. Conclusions

One of the worst cancers, lung cancer claims 10 million lives each year while with current medical research like chest MRI lung nodule diagnosis is critical. Therefore, automated decision support systems are essential for lung cancer diagnosis(Ted et al., 2022).
This research presents a comprehensive comparison of pre-trained multimodal deep learning methods for detailed classification of lung tumour.

Based on DenseNet121 and ResNet50 models under different testing conditions such as noisy, adversarial, and clean data.

DenseNet121 indicates a slightly higher resilience to noisy data and making it a preferable choice when dealing with images that might contain various types of noise. Its architecture allows it to maintain higher accuracy in such scenarios.
On the other hand, ResNet50 shows better performance against adversarial attacks(FGSM).
Both models perform equally well with clean data that showcasing their strong capabilities in ideal conditions without noise or intentional attacks.
Both models are highly effective with their strengths in various aspects of data robustness and accuracy.
The ensemble learning approach was able to predict across various models contributes to its robustness against different data challenges.

For Future work, consider experiments with hybrid models that combine the strengths of DenseNet121 noise resistance with ResNet50 adversarial robustness.

In conclusion, this study has contributed valuable intuition into the optimisation of deep learning models for lung tumour detection and indicating the robustness against both image imperfections and adversarial attacks. The combined of these findings into clinical practice to enhance the early detection of lung tumour.

Access to the code: https://github.com/sarashahin/ML_Research

7. References

1- Shah AA, Malik HAM, Muhammad A, Alourani A, Butt ZA. Deep learning ensemble 2D CNN approach towards the detection of lung cancer. Sci Rep. 2023 Feb 20;13(1):2987. doi: 10.1038/s41598-023-29656-z. PMID: 36807576; PMCID: PMC9941084.
2- Chaunzwa, T.L., Hosny, A., Xu, Y. et al. Deep learning classification of lung cancer histology using CT images. Sci Rep 11, 5471 (2021). https://doi.org/10.1038/s41598-021-84630-x.
3- Wagner RF, Brown DG. Unified snr analysis of medical imaging systems. Phys Med Biol. 1985 Jun;30(6):489-518. doi: 10.1088/0031-9155/30/6/001. PMID: 29081545; PMCID: PMC5658075.
4- Eslami, M. et al.,'A new formulation to reduce the number of variables and constraints to expedite SCUC in bulky power systems', Proceedings of the National Academy of Sciences, India Section A: Physical Sciences, 89(2), pp. 311–321, 2018.
5- Apostolidis, Kyriakos D., and George A. Papakostas. 2021. "A Survey on Adversarial Deep Learning Robustness in Medical Image Analysis" Electronics 10, no. 17: 2132. https://doi.org/10.3390/electronics10172132.
6- Medical Artificial Intelligence and Automation Laboratory, Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, TX, 75390, USA.
7- Junhao Dong, Junxi Chen, Xiaohua XieB, Jianhuang Lai, Senior Member, IEEE, and Hao ChenB, Senior Member,2023, IEEEhttps://doi.org/10.48550/arXiv.2303.14133.
8- Joel MZ, Umrao S, Chang E, Choi R, Yang DX, Duncan JS, Omuro A, Herbst R, Krumholz HM, Aneja S. Using Adversarial Images to Assess the Robustness of Deep Learning Models Trained on Diagnostic Images in Oncology. JCO Clin Cancer Inform. 2022 Feb;6:e2100170. doi: 10.1200/CCI.21.00170. PMID: 35271304; PMCID: PMC8932490.
9- Lian J, Long Y, Huang F, Ng KS, Lee FMY, Lam DCL, Fang BXL, Dou Q, Vardhanabhuti V. Imaging-Based Deep Graph Neural Networks for Survival Analysis in Early Stage Lung Cancer Using CT: A Multicenter Study. Front Oncol. 2022 Jul 13;12:868186. doi: 10.3389/fonc.2022.868186. PMID: 35936706; PMCID: PMC9351205.
10- Madhusudan G Lanjewar, Kamini G Panchbhai, Panem Charanarur, Lung cancer detection from CT scans using modified DenseNet with feature selection methods and ML classifiers,Expert Systems with Applications,Volume 224,2023,https://doi.org/10.1016/j.eswa.2023.119961.
11- Ali HM (2018) MRI Medical Image Denoising by Fundamental Filters. High-Resolution Neuroimaging - Basic Physical Principles and Clinical Applications. InTech. Available at: http://dx.doi.org/10.5772/intechopen.72427.
12- Ghaffari Laleh, N., Truhn, D., Veldhuizen, G.P. et al. Adversarial attacks and adversarial robustness in computational pathology. Nat Commun 13, 5711 (2022). https://doi.org/10.1038/s41467-022-33266-0.

13- Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. IEEE Transactions on knowledge and data engineering, 22(10), 1345-1359.
14- Pizurica, A., & Philips, W. (2006). Estimating the probability of the presence of a signal of interest in multiresolution single-sensor surveillance systems. IEEE Transactions on Image Processing, 15(3), 654-665.
15- Gonzalez, R. C., & Woods, R. E. (2018). Digital Image Processing (4th ed.). Pearson. [ISBN: 9780133356724].
16- He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
17- Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
18- Ted Hubbard(2022). Deep transfer learning approach for lung tumour detection with resilience testing under suboptimal conditions.
19- D. Müller, I. Soto-Rey and F. Kramer, "An Analysis on Ensemble Learning Optimized Medical Image Classification With Deep Convolutional Neural Networks," in IEEE Access, vol. 10, pp. 66467-66480, 2022, doi: 10.1109/ACCESS.2022.3182399.