

# Medical Diagnosis: AI-Driven Disease Exposing Using Images

MSc Research Project MSCAI1

Meenu Mohan Student ID : x23119390

School of Computing National College of Ireland

Supervisor: Kislay Raj

#### National College of Ireland Project Submission Sheet School of Computing



Student Name:	Meenu Mohan
Student ID:	x23119390
Programme:	MSCAI1
Year:	2024
Module:	MSc Research Project
Supervisor:	Kislay Raj
Submission Due Date:	20/12/2018
Project Title:	Medical Diagnosis: AI-Driven Disease Exposing Using Images
Word Count:	2605
Page Count:	13

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

<u>ALL</u> internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Meenu Mohan
Date:	11th August 2024

#### PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).Attach a Moodle submission receipt of the online project submission, to<br/>each project (including multiple copies).You must ensure that you retain a HARD COPY of the project, both for

your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only		
Signature:		
Date:		
Penalty Applied (if applicable):		

# Medical Diagnosis: AI-Driven Disease Exposing Using Images

#### Meenu Mohan x23119390

#### Abstract

The research being investigated looks into the use of AI in health evaluation, particularly for identifying covid from X-rays and cancer from CT images. I created an accurate diagnosis system using advanced deep learning techniques, including transfer learning with VGG16 architecture and explainable AI methods like Grad-CAM technology and SHAP. This study addresses two main questions:

1) Can AI-based X-ray detection outperform RT-PCR for COVID-19 detection? 2) Can the AI system accurately detect early-stage chest cancer compared to skilled radiologists?

Research results show that based on artificial intelligence approaches may enhance diagnostic accuracy and efficiency in healthcare image evaluation, potentially leading to previous disease detection and improved patient outcomes.

## 1 Introduction

Centuries-old challenges like chest cancer, as well as emerging threats like COVID-19, have had an important effect on the global medical landscape. Chest cancer is still the most usual cause of dying from cancer globally, with a projected death toll of 1.8 million in 2020 Sung et al. (2021). At the same time, the COVID-19 pandemic has impacted millions around the world, with more than 500 million reported cases by 2022 Dashboard (2022). The financial effect of these illnesses is also important. The worldwide expense of chest cancer treatment is estimated to reach 188 billion dollars in 2020 International Agency for Research on Cancer et al. (2012), while the overall expense of the COVID-19 pandemic in the United States alone might reach sixteen trillion dollars Cutler and Summers (2020).

Conventional techniques for diagnosis, while efficient, frequently encounter limitations such as deadlines, errors by humans, and, in the instance of COVID-19, potential negative results in RT-PCR tests. The aim of this research investigate how artificial intelligence might overcome these drawbacks, resulting in more accurate, efficient, and accessible diagnostic tools.

This project's main novelty is its ability to process both single and multiple images while also incorporating image quality assessment. The system begins by evaluating each image's quality. If the quality is determined to be low, the system employs enhancement techniques before continuing with the disease detection process. This method aims to increase overall accuracy while reducing errors associated with low-quality images.

#### **Research Questions:**

Is possible to estimate a specific result, and is COVID diagnosis via X-ray more accurate than RT-PCR tests?

Can this technology recognize chest tumors at its earliest stages with greater precision compared to experienced physicians?



Figure 1: Global statistics on COVID-19 and lung cancer incidence and economic impact.

## 2 Related Work

#### 2.1 AI for COVID-19 Detection

Latest study has produced promising results when employing AI to detect COVID-19 from chest X-rays. Wang et al. Wang et al. (2020) generated COVID-Net, which had 93.3 percent precision in separating COVID-19 cases from normal and non-COVID pneumonia cases. In comparison with RT-PCR, which has a stated sensitivity range of 71 percent to 98 percent Arevalo-Rodriguez et al. (2020), AI-based methods have the potential to complement standard evaluation approaches.

Furthermore, Minaee et al. (2020) conducted an in-depth investigation on deep learning for COVID-19 detection from chest X-rays, with an accuracy of 98 percent and particularity of 90 percent. Their research shows AI's potential as an efficient and accurate examination tool.



Figure 2: Comparison of AI and RT-PCR accuracy in COVID-19 detection.

#### 2.2 AI in Chest Cancer Detection

Ardila et al. (2019) demonstrated that a model using deep learning might perform on par with or greater than radiologists in lung cancer screening. The model they used achieved a coefficient intercept of 0.944 on the test set, which is comparable to radiologists's AUC of 0.957.

Hosny et al. (2018) established a deep learning framework for lung cancer the future that outperformed traditional radiomics approaches. Their findings show that AI has the potential for predicting outcomes for patients in along with detection.



Figure 3: AI vs Radiologist accuracy in early-stage lung cancer detection.

#### 2.3 Explainable AI in Medical Imaging

Understandable artificial intelligence techniques including Grad-CAM Selvaraju et al. (2017) and SHAP Lundberg and Lee (2017) are increasingly used to interpret deep learning algorithms in medical images. These approaches provide a visual representation of the model's decision-making procedure, resulting in greater trust and understanding in AI-assisted diagnosis.

Reves et al. (2020) utilized SHAP to interpret deep learning algorithms for COVID-19 detection, showing which features give the most to the algorithm's predictions. The study checks the transparent artificial intelligence in fostering confidence, comprehension in medical AI systems.

## 3 Methodology

The study describes the comprehensive methodology used to create and test a system based on deep learning for chest disease recognition using X-ray clips. The study included the collection of two datasets (COVID-19 and chest cancer), preprocessing, data augmentation, model development with transfer learning, training, and evaluation. Every phase was carefully planned to ensure that the results were reliable and valid, utilizing methods from science and related work.

### 3.1 Data Collection

COVID-19 Dataset: The COVID-19 dataset was sourced from Kaggle, which included chest X-ray images categorized into 'Normal,' and 'COVID-19,' The dataset aimed to provide a balanced and comprehensive set of images to develop a robust detection model.

- Source: COVID-19 Radiography Database
- Categories: COVID-19, Normal, and Pneumonia
- Total Images: 21,165
- Training Set: 15,815 images
- Validation Set: 5,350 images

Chest Cancer Dataset: The chest cancer dataset was obtained from the Kaggle repository, containing chest X-ray images labeled as 'Adenocarcinoma', 'Large cell carcinoma', 'Squamous cell carcinoma', and 'Normal.' The dataset has a diverse range of images to process and validate cancer detection model.

- Source: Chest Cancer Dataset
- Categories: Cancer and Normal
- Total Images: 8,000
- Training Set: 6,000 images
- Validation Set: 2,000 images

#### 3.2 Data Processing

Preprocessing steps were essential for getting the pictures ready for the training of neural networks. Pictures were adjusted, normalized, and augmented to enhance the system. Rescaling and Normalization: All images were resized to 224x224 pixels and rescaled to a pixel range of [0, 1] by dividing by 255. Data Augmentation: To prevent overfitting and improve generalization, data augmentation techniques were applied to the training datasets. These techniques included random rotations, width and height shifts, horizontal flips, and zoom transformations.

#### 3.3 Model Development

A pre-trained VGG16 model was used as the base for the convolutional neural network (CNN) due to its proven performance in image classification tasks. Transfer learning was employed for the knowledge gained from dataset. Base Model: The proposed VGG16 approach, trained on the data set from ImageNet, was used without a top classification layer. This model was selected for its deep architecture,

## 3.4 Model Training

The framework was constructed with the algorithm known as Adam, with dual crossentropy as the reduction operation and precision as its assessment metric. Training included blending a model with the initial data and validating it on a separate set. Training Process: To avoid overfitting, the framework was trained for fifty iterations and stopped early. Quickly stopping tracked loss of validation and reinstated the highest heavy objects noticed during training.

# 4 Design Specification

Architecture: The main architecture utilized in this study is based on the model VGG16, a well-established CNN designed for image classification tasks. VGG16 called for the deep structure, which has sixteen layers with learnable parameters. This architecture is particularly effective due to its simplicity and depth, which helps in extracting features from images at various levels of abstraction.

**Framework**: The software used makes use of TensorFlow and Keras, as which are widely used platforms for creating and instructing machine learning models. TensorFlow as is a powerful tool for defining, training, and assessing artificial intelligence designs, whereas Keras provides an API that is high-level for model construction and experimentation.

**Data Augmentation**: To enhance a model's abstraction, methods for data enhancement like rotation, dimension shifts, horizontal flips, and focusing are used. These techniques deliberately boost the diversity of the training dataset through the creation of customized versions of its initial images, thereby reducing overfitting.

**Grad-CAM**:Grad-CAM is used to make models more interpretable. Grad-CAM technology offers illustrations for CNN's predictions, demonstrating the parts of the picture that had the most impact on the choice of classification.

# 5 Implementation

The software is capable of handling both single and multiple image inputs. For single image prediction, the function 'predict single image' was added, which takes an individual image through the whole process of quality assessment, enhancement, and disease detection. The 'predict multiple images' function was designed to simultaneously process a set of pictures while offering general statistics along with performance metrics for every image in the set.

Dataset Preparation: The dataset was prepared by organizing images into distinct directories for each class. The training and validation datasets were processed using ImageDataGenerator from Keras, which applied real-time data augmentation to increase model robustness.

Model Development: The model known as VGG16 has been loaded with Image Net pretrained weights, removing the most layers that were fully linked. Custom layers that were completely linked have been placed on the top, comprising a global average pooling layer and dense ones, to tailor the model to the classification task at hand.

Grad-CAM Implementation: To visualize model predictions, Grad-CAM was implemented. This involved extracting gradients of the predicted class concerning the last con-

volutional layer and generating heatmaps that highlight important areas of the image contributing to the model's decision.



Figure 4: Output of single image.



Figure 5: Output of the multiple images.

#### Tools and Languages:

- Languages: Python
- Libraries: TensorFlow, Keras, NumPy, Matplotlib, OpenCV
- Tools: Visual Studio Code
- Model: A trained VGG16-based model adapted for the specific image classification task.
- Visualizations: Grad-CAM heatmaps illustrating the areas of the images that influenced the model's predictions.
- Logs and Metrics: Training and validation loss and accuracy metrics for evaluating model performance.

# 6 Evaluation

This section presents an in-depth evaluation of the research's conclusions and main results, in addition to the consequences for both scholars and practitioners. Only the results that are most appropriately promoting the topic of the study and its goals will be provided. Provide a comprehensive examination of the outcomes. Use statistical techniques for assessing and evaluating testing findings and their significance levels. To present the outcomes, employ visual tools such as charts, graphs, and plots.

\_\_\_\_\_

#### 6.1 Case 1: Lung Cancer Detection

The lung cancer detection model has been assessed based on its accuracy and reliability in recognizing cancerous versus non-cancerous images. The model achieved an impressive 85 percent accuracy on the validation dataset, indicating a strong ability to correctly categorize images. Precision, an important metric for ensuring that identifications are accurate, was 87 percent, indicating that the model has become reliable for identifying true positive cases of lung cancer. Meanwhile, the recall rate was 82 percent, indicating that the model correctly identified a significant portion of actual cancer cases but missed others. The F1 Score and recall, was 84.5 percent, indicating overall effective performance.

A confusion matrix has been employed during the evaluation process to visually represent the true positives, negatives, false positives, and false negatives. This matrix was very useful in understanding the model's error distribution. A classification report also included detailed recall, precision and F1 Scores for each class, which aided in the evaluation of model performance. The ROC curve and AUC were utilized to evaluate the compromises between specificity and sensitivity, allowing for a deeper comprehension of the model's testing abilities. Visual aids like the confusion matrix plot and ROC curve were critical in communicating the findings. The confusion matrix heatmap revealed the model's strengths and weaknesses in image classification, while the ROC curve depicted the balance of true positive and false positive rates, providing insight into the model's overall diagnostic performance.

1/1	<b>0s</b> 304ms/step	
1/1	0s 162ms/step	
Image: 000139 (6).png		
Predicted Class: adenocarcinoma		
Confidence: 0.85		
Original Image Quality: high		
File Size: 117.27 KB		
Dimensions: 498x362		
Image quality is high. S	kipping Grad-CAM, SHAP, and LIME explanations.	

Figure 6: Cancer Detection.

#### 6.2 Case 2: COVID-19 Detection

The COVID-19 detection model was evaluated for its ability to classify X-ray images as COVID-19 positive or negative. The model achieved a high accuracy of 90 percent on the validation set, indicating excellent image classification performance. Precision for COVID-19 detection was 89 percent, indicating that the model was accurate in predicting positive cases. The recall was slightly higher at 91 percent, indicating that the model accurately identified the majority of COVID-19 cases. The F1 Score was 90 percent, indicating a well-balanced performance in both precision and recall.

To get the system's performance, I used a Precision-Recall Curve, which highlighted the trade-off between precision and recall, which is especially useful when dealing with potentially imbalanced datasets. The ROC Curve and its corresponding AUC metric were also used to evaluate the model's ability to accurately distinguish between positive and negative cases.

Visualizations were critical in interpreting the results. The Precision-Recall Curve provided a detailed view of the model's performance at different threshold settings, whereas the ROC Curve depicted the trade-offs between true positive and false positive rates, allowing for a more complete understanding of the model's diagnostic accuracy.

Overall, these evaluations provide insights into the models' capabilities and limitations, laying the groundwork for a better understanding of their practical applications and potential improvements.



Figure 7: Covid-19 Detection.

## 6.3 Image Quality Assessment and Enhancement

One of the most significant challenges encountered during the project's development was the presence of false positives and negatives, which were caused primarily by lowquality images. To address this, I added an image quality assessment step before running the main prediction process. Images are rated as 'high' or 'low' quality based on file size, sharpness, and contrast. For images classified as 'low' quality, I use a series of enhancement techniques:

- 1. Image sharpening
- 2. Contrast adjustment
- 3. Super-resolution techniques
- 4. Denoising methods

After enhancement, the images are reassessed for quality. This process significantly reduced the number of false positives and negatives, which improved the models' overall accuracy.

To better comprehend and analyze the model's decisions, I used three explainable AI

techniques.

- 1. Grad-CAM: This technique highlights the image regions that had the greatest influence on the model's decision-making.
- 2. SHAP values help us understand how each feature (or, in this case, each part of the image) contributes to the overall prediction.
- 3. LIME: LIME provides local explanations for individual predictions, allowing us to understand why the model chose a specific decision for a given image.



Figure 8: SHAP Generation.



Figure 9: Lime Generation.



Figure 10: After enhancement low changed to high.

### 6.4 Discussion

The experiments with VGG16-based models for chest cancer and COVID-19 detection produced encouraging results, with high accuracy and precision metrics. The models demonstrated the ability to effectively distinguish between different types of medical images, utilizing transfer learning to improve performance. The chest cancer model achieved an accuracy of 85 percent on the validation set, while the COVID-19 model achieved 89 percent. These findings are consistent with recent research demonstrating the effective-ness of deep learning models in medical image classification.

However, while the models performed admirably, several areas require careful examination. The VGG16 architecture, while effective, may not be suitable for all types of medical images. More advanced architectures, such as EfficientNet or ResNet, could benefit the models by improving their performance and generalization capabilities. Furthermore, the use of binary classification may not capture the full complexity of the problem, particularly when multiple disease stages or conditions exist.

Several design aspects of the experiment:

- 1. Data Imbalance: Despite efforts to augment the dataset, the distribution of classes might still affect model performance. The chest cancer dataset, for example, had a higher number of images for some classes compared to others. This imbalance could lead to biased results, and techniques such as class weighting or SMOTE (Synthetic Minority Over-sampling Technique) might improve model robustness.
- 2. Data Quality and Diversity: The quality of the images used for training and validation plays a crucial role in model accuracy. Inconsistent image resolutions, varying lighting conditions, or the presence of artifacts could impact model performance. Expanding the dataset with more diverse images and incorporating real-world variations could address these issues.
- 3. Hyperparameter Tuning: While the models were trained with default hyperparameters, there is potential for further optimization. Exploring different learning rates, batch sizes, and other hyperparameters through systematic grid search or Bayesian optimization could enhance performance.

4. Model Interpretability: Although Grad-CAM provided valuable insights into model decision-making, further efforts to improve model interpretability are needed. Techniques such as SHAP could offer more detailed explanations of feature contributions.

# 7 Conclusion and Future Work

In the research, I wanted to create and test AI models for predicting chest cancer and COVID-19 in medical clips. The goals included building a strong model with transfer learning and evaluating its performance using various metrics. The models achieved high accuracy, demonstrating their suitability for practical use in medical diagnostics.

The research's key findings highlight the effectiveness of VGG16-based models for medical image classification. The models were effective at distinguishing between different disease states, with promising results for both chest cancer and COVID-19 detection. The use of image quality assessment and enhancement techniques proved critical in increasing the accuracy of the created models. By addressing the issue of poor image quality, I was able to significantly reduce false positives and negatives. Future research should concentrate on a few key areas:

- 1. Exploring Advanced Architectures: More sophisticated models such as EfficientNet, DenseNet, and transformers may provide better performance and generalization.
- 2. Improving Data Quality and Diversity: By expanding the dataset with more diverse and high-quality images, as well as using data augmentation techniques, model robustness can be improved.
- 3. Model Interpretability: Using advanced interpretability techniques like SHAP or LIME can provide more insight into model decision-making processes.
- 4. Multi-class and multi-stage classification: Creating models that can handle multiple classifications or disease stages can provide a more complete diagnostic tool.
- 5. Clinical Integration: Investigating the integration of these models into clinical workflows, such as creating user-friendly interfaces and validating the models in realworld scenarios, can increase their practical utility.
- 6. Optimizing Image Enhancement: Additional research into advanced image enhancement techniques has the potential to significantly improve the quality of low-resolution or noisy medical images.
- 7. Automated Quality Threshold Adjustment: Creating a system that can automatically adjust quality assessment thresholds based on the unique characteristics of various datasets or imaging equipment could improve our model's adaptability.
- 8. Expanding Explainable AI Applications: While we've used Grad-CAM, SHAP, and LIME, there's still room to investigate other explainable AI techniques or create novel methods specifically for medical imaging tasks.

I would like to report that the AI system successfully addressed both research questions. By incorporating image quality assessment and enhancement techniques, I have developed an effective system that not only detects diseases but also reduces false negatives and positives. The key innovation of this project is its approach to image quality. Unlike many other projects that only focus on disease detection, our system first assesses the quality of the input images. When low-quality images are detected, it employs several. enhancement techniques before proceeding with the disease detection process. This critical step addresses an ongoing problem in medical imaging AI: the occurrence of false negatives and false positives due to poor image quality.

Results show that this method significantly improves the accuracy and reliability of disease detection. The system achieved 90 percent accuracy in COVID-19 detection, which is comparable to or better than many RT-PCR tests, particularly given the rapid turnaround time of AI-based analysis. In the case of early-stage chest cancer detection, our system's 85 percent accuracy is comparable to that of skilled radiologists, with the added benefit of consistency and scaling. The addition of AI techniques improves the system's usefulness in clinical settings. These techniques provide deeper into the AI's decisionmaking process, increasing trust and allowing for expert validation of the results.

In conclusion, this study has produced a system that not only meets but exceeds the original research objectives. By improving image quality before analysis, I have created a more reliable disease detection tool that effectively reduces false negatives and positives. This approach is a significant step forward in the field of medical imaging AI, addressing critical limitations in existing systems and paving the way for more accurate, reliable AI-assisted diagnoses in healthcare.

Future research will concentrate on improving these techniques, broadening the range of detectable diseases, and integrating this system into clinical workflows. The success of this project opens up exciting opportunities to improve early disease detection and patient outcomes using advanced AI technologies in medical imaging.

## References

- Ardila, D., Kiraly, A. P., Bharadwaj, S., Choi, B., Reicher, J. J., Peng, L., Tse, D., Etemadi, M., Ye, W., Corrado, G. et al. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography, *Nature medicine* 25(6): 954–961.
- Arevalo-Rodriguez, I., Buitrago-Garcia, D., Simancas-Racines, D., Zambrano-Achig, P., Del Campo, R., Ciapponi, A., Sued, O., Martinez-Garcia, L., Rutjes, A. W., Low, N. et al. (2020). False-negative results of initial rt-pcr assays for covid-19: a systematic review, *PloS one* 15(12): e0242958.
- Cutler, D. M. and Summers, L. H. (2020). The covid-19 pandemic and the 16trillionvirus, Jama 324(15): 1495 -1496.
- Dashboard, W. (2022). Who coronavirus (covid-19) dashboard, World Health Organization. Available from: https://covid19. who. int.
- Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H. and Aerts, H. J. (2018). Artificial intelligence in radiology, *Nature Reviews Cancer* 18(8): 500–510.
- International Agency for Research on Cancer, W. et al. (2012). Cervical cancer estimated incidence, mortality and prevalence worldwide in 2012, *World Health Organization*.

- Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions, Advances in neural information processing systems **30**.
- Minaee, S., Kafieh, R., Sonka, M., Yazdani, S. and Soufi, G. J. (2020). Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning, *Medical image analysis* 65: 101794.
- Reyes, M., Meier, R., Pereira, S., Silva, C. A., Dahlweid, F.-M., Tengg-Kobligk, H. v., Summers, R. M. and Wiest, R. (2020). On the interpretability of artificial intelligence in radiology: challenges and opportunities, *Radiology: artificial intelligence* **2**(3): e190043.
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization, *Proceedings of the IEEE international conference on computer vision*, pp. 618–626.
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A. and Bray, F. (2021). Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA: a cancer journal for clinicians* **71**(3): 209– 249.
- Wang, L., Lin, Z. Q. and Wong, A. (2020). Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images, *Scientific reports* 10(1): 19549.