

# **An Experimental Study of Deep CNN Models for Medical Image Classification**

submitted by

**Priyanka Ghosh**

Class Roll No: 002210503048

Exam Roll: MCA244047

Master of Computer Application

under the supervision of

**Prof. Ram Sarkar**

in the

Department of Computer Science and Engineering

Faculty of Engineering and Technology

Jadavpur University

2024

Faculty of Engineering and Technology  
Jadavpur University

**CERTIFICATE OF RECOMMENDATION**

**To Whom It May Concern**

I hereby recommend that the project report entitled “An Experimental Study of Deep CNN Models for Medical Image Classification” has been carried out by Priyanka Ghosh (Reg. No.: 163640 of 2022-23, Roll No.: 002210503048, Exam Roll: MCA244047), under my guidance and supervision and be accepted in partial fulfilment of the requirement for the degree of Master of COMPUTER APPLICATION in DEPARTMENT of COMPUTER SCIENCE and ENGINEERING, JADAVPUR UNIVERSITY during the academic year 2023-2024.

---

Prof. Ram Sarkar  
Project Supervisor  
Dept. of Computer Science and  
Engineering  
Jadavpur University

---

Prof. (Dr.) Nandini Mukhopadhyay  
Head of the Department  
Dept. of Computer Science & Engineering  
Jadavpur University

---

Prof. Dipak Laha  
Dean, Faculty Council  
of Engineering & Technology  
Jadavpur University, Kolkata-032

Faculty of Engineering and Technology  
Jadavpur University

**CERTIFICATE OF APPROVAL**

This is to certify that the project report entitled “**An Experimental Study of Deep CNN Models for Medical Image Classification**” is a bonafide record of work carried out by PRIYANKA GHOSH, in fulfillment of the requirements for the award of the degree of Master of Computer Application in the Department of Computer Science and Engineering, Jadavpur University. It is understood that by this approval the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn therein but approve the project report only for the purpose for which it has been submitted.

---

Signature of Examiner 1

Date :

---

Signature of Examiner 2

Date :

Faculty of Engineering and Technology  
Jadavpur University

**DECLARATION OF COMPLIANCE WITH  
ACADEMIC ETHICS**

I hereby declare that this project report entitled "An Experimental Study of Deep CNN Models for Medical Image Classification" is an outcome of my genuine and independent work, as part of my Master of Computer Application program.

All information in this document has been obtained and presented in accordance with academic rules and ethical conduct.

I also declare that, as required by these rules and conduct, all materials and results that are not original to this work have been fully cited and referenced.

Priyanka Ghosh

Registration No.: 163640 of 2022-23

Examination Roll No.: MCA244047

Thesis Title :- An Experimental Study of Deep CNN Models for Medical Image Classification

---

Signature with Date

# **ACKNOWLEDGEMENT**

I want to express my profound gratitude and thanks to Prof. Ram Sarkar for his valuable guidance, suggestions and feedback throughout the working of this project. His expertise and insights have been instrumental in shaping the direction and outcome of this project.

I am also profoundly grateful to Dr. Nandini Mukhopadhyay, HOD of Department of Computer Science and Engineering, for providing the necessary resources and a conducive environment for working on this project.

I sincerely thank Prof. Mahantapas Kundu, Coordinator of the CMATER Research Lab, for ensuring all the technical and logical requirements were effectively met. My heartfelt thanks to the research scholar, Akash halder and every faculty member and scholar of the CMATER Lab for their valuable inputs, constant collaboration and willingness to share their knowledge and expertise. Their contributions have been extremely helpful and enriching.

I would also like to thank the Department of Computer Science and Engineering faculty for imparting the fundamental knowledge essential for the inception of this project.

Lastly I express my gratitude to my friends and family members for their constant support.

---

Priyanka Ghosh

Class Roll No.: 002210503048

Examination Roll No.: MCA244047

Department of Computer Science & Engineering

Jadavpur University

# *Abstract*

Deep learning, a subset of machine learning, has revolutionized artificial intelligence by enabling models to learn complex patterns directly from data. Its application in medical image analysis aids in early detection, diagnosis, and effective treatment planning, significantly improving patient outcomes. Accurate classification of medical images is crucial for timely medical interventions. This project leverages deep learning techniques to classify medical images from three datasets: lung cancer, breast cancer, and kidney cyst/tumor conditions. We employ five state-of-the-art Convolutional Neural Network (CNN) models—VGG16, Xception, EfficientNetB0, NasNetMobile, and DenseNet121—on each dataset. To ensure optimal performance, each model undergoes rigorous hyperparameter tuning and is evaluated using metrics like F1 score, accuracy, precision, and recall. After establishing baseline performances, we refine the models by incorporating the Convolutional Block Attention Module (CBAM) into the last convolutional layer of each CNN. CBAM enhances the model’s focus on relevant image parts by adaptively recalibrating feature responses, addressing challenges such as subtle differences between tissues, class imbalance, and noise. Our results show significant improvement in classification accuracy and overall performance metrics with CBAM integration. The enhanced models consistently outperform their non-attention counterparts across all datasets. The findings highlight the potential of attention mechanisms like CBAM in improving the accuracy and reliability of deep learning models for medical image analysis. This integration boosts performance and provides a more interpretable framework for understanding neural network decisions, contributing to the development of reliable diagnostic tools for better patient outcomes. By addressing inherent challenges in medical image classification and demonstrating the effectiveness of advanced deep learning techniques, this project underscores the transformative impact of deep learning in healthcare, leading to more precise and efficient diagnostic processes.

# Contents

|                                                    |    |
|----------------------------------------------------|----|
| <b>Abstract</b>                                    | i  |
| <b>Table of Contents</b>                           | ii |
| <b>List of Figures</b>                             | v  |
| <b>List of Tables</b>                              | vi |
| <b>1 Introduction</b>                              | 1  |
| 1.1 Challenges in Medical Image Processing         | 1  |
| 1.2 Applications of Medical Image Processing       | 2  |
| 1.3 Objectives                                     | 2  |
| 1.4 Scope of the Work                              | 3  |
| 1.5 Organization of the Work                       | 4  |
| <b>2 Literature Survey</b>                         | 6  |
| 2.1 IQ-OTH/NCCD Lung Cancer Dataset                | 6  |
| 2.2 Ultrasound Breast Images Dataset               | 7  |
| 2.3 Ct kidney dataset: normal-cyst-Tumor and Stone | 9  |
| <b>3 Methods and Materials</b>                     | 11 |
| 3.1 Importance of Deep Learning                    | 11 |
| 3.2 Convolutional Neural Networks                  | 12 |
| 3.3 Models Used                                    | 12 |
| 3.3.1 VGG16                                        | 13 |
| 3.3.1.1 Overview                                   | 13 |
| 3.3.1.2 Architecture                               | 14 |
| 3.3.1.3 Working Principle                          | 14 |
| 3.3.1.4 Benefits of VGG16                          | 14 |
| 3.3.1.5 Limitations of VGG16                       | 15 |
| 3.3.1.6 Discussion                                 | 15 |
| 3.3.2 Xception                                     | 15 |
| 3.3.2.1 Overview                                   | 15 |
| 3.3.2.2 Architecture                               | 16 |
| 3.3.2.3 Working Principle                          | 16 |
| 3.3.2.4 Benefits of Xception                       | 17 |
| 3.3.2.5 Limitations of Xception                    | 17 |
| 3.3.2.6 Discussion                                 | 17 |

|          |                                                        |           |
|----------|--------------------------------------------------------|-----------|
| 3.3.3    | DenseNet121                                            | 17        |
| 3.3.3.1  | Overview                                               | 18        |
| 3.3.3.2  | Architecture                                           | 18        |
| 3.3.3.3  | Working Principle                                      | 18        |
| 3.3.3.4  | Benefits of DenseNet121                                | 19        |
| 3.3.3.5  | Limitations of DenseNet121                             | 19        |
| 3.3.3.6  | Discussion                                             | 19        |
| 3.3.4    | NASNetMobile                                           | 19        |
| 3.3.4.1  | Overview                                               | 20        |
| 3.3.4.2  | Architecture                                           | 20        |
| 3.3.4.3  | Working Principle                                      | 20        |
| 3.3.4.4  | Benefits of NASNetMobile                               | 20        |
| 3.3.4.5  | Limitations of NASNetMobile                            | 21        |
| 3.3.4.6  | Discussion                                             | 21        |
| 3.3.5    | EfficientNetB0                                         | 22        |
| 3.3.5.1  | Overview                                               | 22        |
| 3.3.5.2  | Architecture                                           | 22        |
| 3.3.5.3  | Working Principle                                      | 22        |
| 3.3.5.4  | Benefits of EfficientNetB0                             | 23        |
| 3.3.5.5  | Limitations of EfficientNetB0                          | 23        |
| 3.3.5.6  | Discussion                                             | 23        |
| 3.4      | Attention Module                                       | 23        |
| 3.4.1    | Convolutional Block Attention Module                   | 24        |
| 3.4.1.1  | How CBAM Works                                         | 24        |
| 3.4.1.2  | Benefits of Using CBAM                                 | 25        |
| 3.5      | Procedure of Classification                            | 26        |
| 3.5.0.1  | Base Model Implementation                              | 26        |
| 3.5.0.2  | Enhancement with CBAM Attention                        | 26        |
| 3.5.1    | Discussion                                             | 26        |
| <b>4</b> | <b>Results and Discussion</b>                          | <b>28</b> |
| 4.1      | System configuration                                   | 28        |
| 4.2      | Evaluation Metrics                                     | 28        |
| 4.3      | Datasets                                               | 30        |
| 4.3.1    | IQ-OTH/NCCD Lung Cancer Dataset                        | 30        |
| 4.3.1.1  | Dataset Collection                                     | 31        |
| 4.3.1.2  | Value of the Data                                      | 33        |
| 4.3.2    | Ultrasound Breast Images Dataset                       | 33        |
| 4.3.2.1  | Dataset Collection                                     | 33        |
| 4.3.2.2  | Value of the Data                                      | 34        |
| 4.3.3    | Ct kidney dataset: normal-cyst-Tumor and Stone         | 36        |
| 4.3.3.1  | Dataset Collection                                     | 36        |
| 4.3.3.2  | Value of the Data                                      | 36        |
| 4.4      | Results                                                | 37        |
| 4.4.1    | Hyperparameters                                        | 37        |
| 4.4.2    | IQ-OTH/NCCD Lung Cancer Dataset Results                | 39        |
| 4.4.3    | Ultrasound Breast Images Dataset Results               | 39        |
| 4.4.4    | Ct kidney dataset: normal-cyst-Tumor and Stone Results | 40        |

|          |                                    |           |
|----------|------------------------------------|-----------|
| 4.4.5    | Confusion Matrices of the Datasets | 41        |
| 4.4.5.1  | IQ-OTH/NCCD Lung Cancer Dataset    | 41        |
| 4.4.5.2  | Ultrasound Breast Images Dataset   | 41        |
| 4.5      | Discussion                         | 42        |
| <b>5</b> | <b>Conclusions</b>                 | <b>45</b> |
| 5.1      | Limitations                        | 46        |
| 5.2      | Scope for Future Work              | 46        |
|          | <b>Bibliography</b>                | <b>48</b> |

# List of Figures

|     |                                                                                             |    |
|-----|---------------------------------------------------------------------------------------------|----|
| 3.1 | Block diagram of a basic CNN architecture [21]                                              | 13 |
| 3.2 | VGG16 Architecture [22]                                                                     | 14 |
| 3.3 | Xception Architecture [23]                                                                  | 16 |
| 3.4 | DenseNet121 Architecture [24]                                                               | 18 |
| 3.5 | NASNetMobile Architecture [25]                                                              | 21 |
| 3.6 | EfficientNetB0 Architecture [26]                                                            | 23 |
| 3.7 | Overview of CBAM [27]                                                                       | 25 |
| 3.8 | Channel Attention and Spatial Attention [27]                                                | 25 |
| 4.1 | Samples of IQ-OTH/NCCD Lung Cancer Dataset                                                  | 32 |
| 4.2 | Samples of Ultrasound Breast Images Dataset                                                 | 35 |
| 4.3 | Samples of Ct kidney dataset: normal-cyst-Tumor and Stone                                   | 38 |
| 4.4 | Confusion Matrix of the IQ-OTH/NCCD Lung Cancer Dataset using VGG16 without attention       | 41 |
| 4.5 | Confusion Matrix of the IQ-OTH/NCCD Lung Cancer Dataset using VGG16 with CBAM attention     | 42 |
| 4.6 | Confusion Matrix of theUltrasound Breast Images Dataset using Xception without attention    | 43 |
| 4.7 | Confusion Matrix of the Ultrasound Breast Images Dataset using Xception with CBAM attention | 44 |

# List of Tables

|     |                                                                       |    |
|-----|-----------------------------------------------------------------------|----|
| 4.1 | Confusion Matrix                                                      | 30 |
| 4.2 | The three classes of lung cases and the number of images in each case | 31 |
| 4.3 | The three classes of breast cases and the number of images in each    |    |
|     | case                                                                  | 34 |
| 4.4 | The Four classes of kidney cases and the number of images in each     |    |
|     | case                                                                  | 37 |
| 4.5 | Hyperparameters used in the models                                    | 39 |
| 4.6 | Performance of models on the lung cancer dataset                      | 39 |
| 4.7 | Performance of models on the breast cancer dataset                    | 40 |
| 4.8 | Performance of models on the kidney disease dataset                   | 40 |

# Introduction

In the dynamic intersection of modern healthcare and artificial intelligence, deep learning models have emerged as transformative tools, offering unparalleled capabilities in medical diagnostics. This project embarks on a pioneering journey to leverage the power of deep learning in classifying diverse medical conditions. Leveraging state-of-the-art deep learning architectures such as EfficientNetB0 [1], VGG16 [2], Xception [3], NasNetMobile [4], and DenseNet121 [5], this endeavor seeks to revolutionize the landscape of medical diagnostics.

Medical image processing is a crucial aspect of modern healthcare, leveraging advanced computational techniques to analyze and interpret complex medical images. This field plays a vital role in early disease detection, diagnosis, treatment planning, and patient monitoring. By utilizing various imaging modalities such as X-rays, MRI, CT scans, and ultrasound, medical image processing aids in enhancing the accuracy and efficiency of medical diagnoses.

## 1.1 Challenges in Medical Image Processing

Despite its immense potential, medical image processing comes with its set of challenges, including:

- **Noise and Artifacts:** Medical images often suffer from noise and artifacts due to various factors such as patient movement, equipment limitations, and imaging protocols.
- **Inter-observer Variability:** Interpretation of medical images can vary among different radiologists or clinicians, leading to inconsistencies in diagnosis.
- **Class Imbalance:** In classification tasks, certain medical conditions may be rare, resulting in class imbalance issues that can affect the performance

of machine learning models.

- **Anatomical Variability:** Human anatomy exhibits significant variability across individuals, making it challenging to develop universal models that generalize well across diverse patient populations.

## 1.2 Applications of Medical Image Processing

Medical image processing finds its applications in various domains, including:

- **Disease Diagnosis and Detection:** Automated analysis of medical images aids in the early detection and diagnosis of diseases such as cancer, cardiovascular disorders, and neurological conditions.
- **Treatment Planning:** Medical imaging assists clinicians in planning surgical interventions, radiation therapy, and other treatment modalities by providing detailed anatomical information.
- **Image-Guided Interventions:** Real-time imaging techniques enable precise localization of pathological sites during minimally invasive procedures, improving procedural outcomes and patient safety.
- **Disease Progression Monitoring:** Longitudinal analysis of medical images facilitates monitoring disease progression, assessing treatment response, and predicting patient outcomes.

In this project, we aim to address the challenges of medical image processing by employing deep learning techniques for the classification of lung cancer, breast cancer, and kidney stone tumor conditions.

## 1.3 Objectives

In pursuit of advancing medical diagnostics through the lens of artificial intelligence, this project embarks on a multifaceted exploration of deep learning methodologies applied to diverse medical datasets. The overarching objective is to enhance diagnostic capabilities, optimize model performance, and facilitate seamless integration of AI-driven technologies into clinical workflows. With a focus on model efficacy, interpretability, and scalability, this endeavor aims to unlock transformative advancements in healthcare delivery and improve patient outcomes.

The primary objective of this project is to develop and evaluate a sophisticated medical image classification system using advanced Convolutional Neural

Network (CNN) architectures—VGG16, DenseNet121, EfficientNetB0, NasNet-Mobile, and Xception. The project focuses on classifying three distinct medical image datasets, each pertaining to different medical conditions: lung cancer, breast cancer, and kidney stone tumors. After obtaining the initial classification results from these CNN models, the project aims to enhance their performance by integrating Convolutional Block Attention Module (CBAM) attention mechanisms. This enhancement involves extracting features from the last convolutional layer of each model, applying CBAM attention, followed by Global Average Pooling and dense layers to improve the classification accuracy. The objective is to assess the impact of CBAM attention on the performance of these CNN models, ultimately contributing to the development of more precise and reliable diagnostic tools in medical imaging.

## 1.4 Scope of the Work

The scope of this ambitious endeavor is expansive and multifaceted, representing a comprehensive exploration of deep learning methodologies applied to three distinct and clinically relevant medical datasets:

1. **Lung Cancer Dataset:** Comprehensive analysis of diverse lung cancer subtypes and pathological conditions.
2. **Kidney Tumor Stone Dataset:** Investigation of renal abnormalities and stone formations for diagnostic purposes.
3. **Breast Cancer Dataset:** Examination of breast lesions and tumor characteristics to enhance diagnostic accuracy.

Through meticulous analysis and model development, this project aims to cultivate robust classification systems with the capacity to discern subtle nuances and intricate patterns within medical images. By harnessing the predictive prowess of advanced neural architectures, including but not limited to EfficientNetB0, VGG16, Xception, NasNetMobile, and DenseNet121, the project endeavors to push the boundaries of medical diagnostics. The ultimate goal is to develop classification frameworks that not only achieve superior accuracy in identifying and categorizing pathological features but also demonstrate adaptability and generalizability across diverse medical contexts.

Moreover, beyond immediate clinical applications, this endeavor aspires to make significant contributions to the burgeoning field of computational healthcare:

4. Elucidating the synergies between artificial intelligence and medical science to drive transformative advancements in healthcare delivery.

5. Facilitating early detection, personalized treatment strategies, and improved patient outcomes through the integration of data-driven insights and advanced technologies.

This multidimensional exploration not only aims to enhance diagnostic accuracy but also seeks to redefine the paradigm of medical diagnostics in the digital era. Through interdisciplinary collaboration, rigorous experimentation, and knowledge dissemination, this project endeavors to lay the groundwork for the realization of precision medicine, where data-driven insights and advanced technologies converge to revolutionize clinical practice and usher in a new era of patient-centric care.

Furthermore, the project's scope extends beyond mere model development, encompassing rigorous validation, optimization, and deployment processes to ensure the scalability, efficiency, and real-world applicability of the developed classification systems. By integrating seamlessly into clinical workflows, these systems have the potential to empower healthcare professionals with actionable insights, streamline decision-making processes, and ultimately improve patient outcomes on a global scale.

In essence, this project represents a holistic and forward-thinking approach to medical diagnostics, leveraging cutting-edge technologies, interdisciplinary collaboration, and a deep understanding of clinical needs to drive transformative change in healthcare delivery. Through relentless innovation and a steadfast commitment to excellence, this endeavor seeks to shape the future of medicine and pave the way for a more equitable, efficient, and patient-centric healthcare system.

Through the comprehensive exploration of different deep learning models applied to diverse medical datasets, this project endeavors to unlock new frontiers in medical diagnostics. By elucidating the strengths and limitations of each model and fostering interdisciplinary collaboration between AI researchers and healthcare practitioners, this endeavor aims to pave the way for precision medicine in the digital age, where accurate and timely diagnosis empowers clinicians to deliver personalized and effective patient care.

## 1.5 Organization of the Work

The remainder of this project unfolds in the following manner:

- **Chapter 2:** This chapter embarks on a captivating journey through a comprehensive literature survey. This exploration delves deep into the realm of existing research endeavors, unveiling a tapestry of methodologies, chal-

lenges, and groundbreaking advancements in the captivating field of medical image classification.

- **Chapter 3:** This one illuminates the path where novel methodologies and materials take center stage. Here, innovative approaches and techniques are unveiled, each meticulously crafted to navigate the intricate landscape of medical image classification, shedding light on the complexities inherent in this vital domain.
- **Chapter 4:** This chapter offers a vivid presentation of the results gleaned from intensive experiments conducted on three distinct datasets. This chapter not only showcases performance metrics and quantitative evaluations but also embarks on critical discussions, unraveling the implications of the findings and their profound impact on the realm of medical image classification.
- **Chapter 5:** As the journey reaches its culmination, Chapter 5 beckons, offering a compelling conclusion to the narrative. Here, essential conclusions drawn from the study come to life, offering profound insights into future research directions. This concluding chapter serves as a beacon of knowledge, synthesizing key findings, unveiling overarching themes, and underscoring the monumental contributions made to the field of medical image classification.

# Literature Survey

This section provides an overview of relevant studies and research papers that have explored the classification of medical image datasets similar to those used in this project. By reviewing existing literature, insights can be gained into the methodologies, techniques, and performance metrics employed by researchers in similar domains.

## 2.1 IQ-OTH/NCCD Lung Cancer Dataset

Lung cancer detection using deep learning techniques has become increasingly crucial for early diagnosis and improved patient outcomes. This subsection presents a review of studies focused on the classification of lung cancer datasets. By analyzing previous research efforts, insights can be gained into the challenges and opportunities inherent in classifying lung cancer images, as well as the effectiveness of various deep learning approaches in this context.

Padmanandam et al. [6] propose a method called RN-PDMP, which combines a Cross Entropy Focal Loss-based preprocessing model for addressing class imbalance, Census Transformed Residual Neural Network-based feature extraction, and a Partial Derivative Multilayer Perceptron Deep Learning-based Classifier for classification. Our approach achieves a precision of 99% and recall of 97%, outperforming existing methods in terms of recall for small samples and true positive pulmonary nodules detection.

Mishra et al. [7] propose three models: synthetic minority oversampling technique (SMOTE) analysis, class-weighted approach, and data augmentation on CNN models. ResNet50, AlexNet, and DenseNet-121 are utilized to classify lung CT images for cancer detection. Evaluation is conducted on the IQ-OTH/NCCD dataset, with preprocessing applied to different CNN models. DenseNet-121 achieves 99.4% accuracy with SMOTE analysis, outperforming ResNet50 and AlexNet with

accuracies of 86.05% and 96.5%, respectively. The proposed method surpasses state-of-the-art approaches, highlighting the effectiveness of DenseNet-121 with SMOTE analysis in lung cancer detection.

Yan et al. [8] propose a method on lung cancer detection using Convolutional Neural Networks (CNNs) applied to CT images. Three CNN-based models, namely ResNet50, AlexNet, and DenseNet-121, are employed for classification. The proposed approach addresses class imbalance through synthetic minority oversampling technique (SMOTE) analysis, class-weighted approach, and data augmentation. Evaluation is conducted on the IQ-OTH/NCCD dataset, with DenseNet-121 achieving the highest accuracy of 99.4% using SMOTE analysis.

Parvathy et al. [9] present a lung cancer detection model comprising three phases: preprocessing, segmentation, and detection. Chest CT images undergo preprocessing with image denoising using a General Adversarial Network (GAN) and contrast enhancement via adaptive histogram equalization. Active contour algorithm segments tumor-affected regions, followed by detection using a 3DCNN model. Evaluation metrics demonstrate the model's effectiveness, achieving an accuracy of 98%, precision of 97.2%, recall of 98.5%, error of 2%, specificity of 97.7%, and NPV of 97.1%. The approach facilitates early-stage lung cancer identification and preventative measures.

Charnsripinyo et al. [10] present lung cancer classification using machine learning techniques applied to CT-scan images. The dataset, comprising 1096 images, was preprocessed and balanced, with features extracted using multiple image filters. Experimental cases were conducted using decision tree, random forest, support vector machine, and K-nearest neighbor techniques, evaluated for binary and multi-class classification. Performance metrics including confusion matrix, accuracy, recall, precision, and F1-score were considered. Random forest and K-nearest neighbor techniques achieved high performance, with accuracies of 99.73% for 2-class classification and 98.69% for 3-class classification.

## 2.2 Ultrasound Breast Images Dataset

The classification of breast cancer datasets is explored in this subsection through a review of relevant literature. By synthesizing findings from previous studies, a comprehensive understanding can be developed regarding the challenges associated with breast cancer image classification and the state-of-the-art techniques employed to achieve accurate and reliable classification results.

Hossain et al. [11] propose a novel method for classifying breast ultrasound images into benign and malignant tumors to aid in early breast cancer diagnosis. The approach involves preprocessing the images to improve quality and remove

noise, followed by segmentation of the breast tumor region using an optimized UNet architecture. A fine-tuned CNN model, coupled with an attention unit, is then utilized for tumor classification. The proposed RKONet-13 model achieves an accuracy of 98.41%, outperforming transfer learning models and demonstrating stability through k-fold cross-validation. Additionally, the model exhibits robustness in classifying breast cancer on mammogram images, achieving a performance of 96.21%. Overall, the study highlights the effectiveness of a comprehensive approach combining image preprocessing, segmentation, and classification for breast cancer diagnosis using ultrasound images.

Uysal et al. [12] address the challenge of breast cancer detection by employing artificial intelligence (AI) for ultrasound analysis, aiming to enhance the efficacy and objectivity of computer-aided diagnosis (CAD). The dataset comprises 780 images, with data augmentation and preprocessing techniques applied. Three models—ResNet50, ResNeXt50, and VGG16—are utilized for classification into benign, malignant, and normal samples. Results demonstrate promising accuracy, with ResNeXt50 achieving the highest accuracy of 85.83%. The findings suggest that AI-driven diagnosis could alleviate the burden on medical practitioners and enhance diagnostic accuracy, offering potential benefits for the field of medicine. Further research and data expansion are recommended to enhance model performance.

Balasubramaniam et al. [13] propose the application of LeNet, a classic convolutional neural network (CNN) architecture, for breast cancer data analysis. By utilizing corrected Rectified Linear Unit (ReLU) activation and batch normalization techniques, LeNet demonstrates improved performance in extracting discriminative features and classifying malignant and benign tumors. The proposed classifier achieves an accuracy of 89.91%, outperforming benchmarking deep learning models. The findings suggest that LeNet, enhanced with corrected ReLU and batch normalization, offers a promising approach for breast cancer detection and diagnosis, with potential benefits for improving patient outcomes.

Pourasad et al. [14] propose a novel method for identifying and diagnosing breast tumors from ultrasound images. Six techniques are employed for image detection and segmentation, followed by feature extraction using the fractal method. Classification techniques including k-nearest neighbor, support vector machine, decision tree, and Naïve Bayes are utilized, alongside a convolutional neural network (CNN) architecture for direct breast cancer classification. The CNN model achieves 99.8% accuracy on the training set and 88.5% sensitivity on the test set. Additionally, a second CNN model accurately identifies tumor location, with 92% of images showing high performance. Post-processing morphological operations enable tumor location and volume identification. These findings offer potential

for patient monitoring and early intervention to prevent disease progression.

Roni et al. [15] address the need for accurate and efficient breast cancer classification models by proposing a methodology to identify the optimal convolutional neural network (CNN) architecture. Eight CNN models are evaluated for binary breast cancer classification using ultrasound images. A generative adversarial network is employed to enhance algorithm precision due to limited data availability. Statistical analyses are conducted to identify the best-performing CNN model for early breast cancer detection. Xception achieves 95% accuracy in both normalized and denormalized conditions, outperforming other models. Its comprehensive feature extraction technique contributes to its superior performance, making it a promising candidate for breast cancer diagnosis.

### 2.3 Ct kidney dataset: normal-cyst-Tumor and Stone

In this subsection, we review research papers that have addressed the classification of kidney tumor stone datasets. By examining previous methodologies and experimental results, valuable insights can be obtained into the complexities of classifying kidney tumor stone images and the efficacy of different classification models in this domain.

Mehedi et al. [16] propose a segmentation-based kidney tumor classification method using deep neural networks (DNNs). Kidneys are segmented using UNet and SegNet, followed by classification using MobileNetV2, VGG16, and InceptionV3 models. The "CT KIDNEY DATASET: Normal-Cyst-Tumor and Stone" dataset is utilized for model training. The classification models achieve accuracies of 95.29%, 99.48%, and 97.38% on the test set, with VGG16 exhibiting the highest accuracy, sensitivity, and specificity. The explainable AI method GradCam is applied to interpret model results.

Bingol et al. [17] propose a hybrid model for classifying CT images containing stones, tumors, and cysts. Feature maps from specific layers of the model are optimized and combined using the Relief method, followed by classification using a wide neural network classifier. The proposed model achieves a significantly higher accuracy of 99.37% compared to pre-trained models, demonstrating promising performance in kidney CT image classification.

Hossain et al. [18] focus on the early detection of chronic kidney disease using CT images through the application of deep neural networks (DNNs). Five CNN classification methods, including watershed segmentation, are employed to classify four types of kidney CT images: cyst, normal, stone, and tumor. The segmentation of CT images is performed using the watershed algorithm, followed by training various classification networks such as EAnet, ResNet50, VGG19, InceptionV3,

and SqueezeNet. The models are trained on the CT Kidney Normal Cyst Tumor and Stone dataset from Kaggle. Results show that EANet, SqueezeNet, VGG19, InceptionV3, and ResNet50 achieve accuracies of 83.6%, 97.3%, 99.9%, 98.8%, and 87.9% respectively, on the test set. The modified VGG19 model demonstrates the highest sensitivity, specificity, and overall accuracy among the models evaluated.

Ripa et al. [19] focuses on the Deep learning CNN models such as Keras, VGG16, SVM, U-NET, and Water-Shed to classify images from the dataset. Initial experiments show promising results, with 100% accuracy achieved by the Keras model and 50% accuracy by the VGG16 model. The models are trained on images representing four classes: normal, cyst, stone, and tumor. Evaluation of the models against input photos demonstrates their ability to accurately determine the stage of kidney illness. Further refinement of the models could potentially facilitate the diagnosis and treatment of kidney diseases by the medical community.

Majid et al. [20] explore the use of machine learning (ML) and Fine-tuned Transfer Learning (TL) for the detection of kidney tumors (KT) in CT images. Different image dimensions and feature engineering techniques, such as the Grey Level Co-occurrence Matrix (GLCM), are investigated. Various ML models, including Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting (GB), and Light Gradient Boosting Model (LGBM), are employed for KT detection. Additionally, fine-tuned ResNet-101 and DenseNet-121 models are evaluated. Results show that ResNet-101 and DenseNet-121 achieve the highest accuracy and outperform other models, demonstrating the potential of ML approaches for efficient classification of KT in CT kidney scans.

# Methods and Materials

In this comprehensive chapter, we embark on a detailed exploration of the methods and materials employed in the classification of three vital medical image datasets: lung, breast, and kidney. Our endeavor is propelled by the utilization of sophisticated convolutional neural network (CNN) architectures, namely VGG16, Xception, DenseNet121, NASNetMobile, and EfficientNetB0, meticulously crafted to discern intricate patterns and extract salient features from medical images with unparalleled precision and accuracy. Each CNN model boasts unique capabilities in feature extraction and pattern recognition, enriching our analysis of diverse pathological conditions encapsulated within the datasets. Through this meticulously crafted chapter, we aspire to offer a comprehensive elucidation of the methodologies harnessed and the datasets utilized, laying a robust foundation for a profound understanding of our classification approach.

## 3.1 Importance of Deep Learning

Deep learning, a subset of machine learning, has emerged as a cornerstone of AI-driven advancements in various domains, including healthcare. Unlike traditional machine learning approaches, deep learning models are capable of automatically learning hierarchical representations of data, thereby enabling more nuanced and accurate predictions. In medical imaging, where subtle features can have significant diagnostic implications, the ability of deep learning models to extract complex patterns from raw data is particularly valuable. By leveraging large datasets and powerful computational resources, deep learning has demonstrated remarkable success in tasks such as image classification, object detection, and segmentation, revolutionizing the field of medical diagnostics.

## 3.2 Convolutional Neural Networks

CNNs are a class of deep learning models specifically designed for processing structured grid-like data, such as images. Inspired by the hierarchical organization of the visual cortex in the human brain, CNN have revolutionized the field of computer vision and image analysis.

At the core of CNN are convolutional layers, which apply a series of learnable filters across the input image, capturing local spatial patterns and features. Through the process of feature extraction and hierarchical representation learning, CNN can discern intricate patterns and features within images, without the need for explicit feature engineering.

CNN also employ pooling layers to downsample the extracted features, reducing computational complexity and enhancing translational invariance. This enables CNN to robustly classify images despite variations in orientation, scale, and position.

One of the key advantages of CNN is their ability to automatically learn hierarchical representations of data through the training process. By iteratively adjusting the parameters of the network using back propagation and gradient descent, CNN can adaptively learn to recognize complex patterns and structures within images.

The utilization of CNN has led to significant advancements in various applications, including image classification, object detection, segmentation, and even natural language processing tasks. Their versatility, scalability, and robustness make them indispensable tools in the field of artificial intelligence, driving innovations in healthcare, autonomous vehicles, robotics, and beyond.

In summary, CNN represents a powerful paradigm for automated feature learning and image analysis, offering unparalleled capabilities in pattern recognition and classification. Through continuous research and development, CNN continues to shape the future of artificial intelligence and revolutionize the way we perceive and interact with visual data. [Figure 3.1](#) represents the CNN architecture.

## 3.3 Models Used

In this project, we have utilized several state-of-the-art CNN models to classify medical images of different body organs: lung, breast, and kidney. The models employed for this classification task include VGG16, Xception, DenseNet121, NASNetMobile, and EfficientNetB0. Each of these models has its unique architecture and advantages, which make them suitable for different types of image analysis tasks. Below, we provide detailed descriptions of these models.

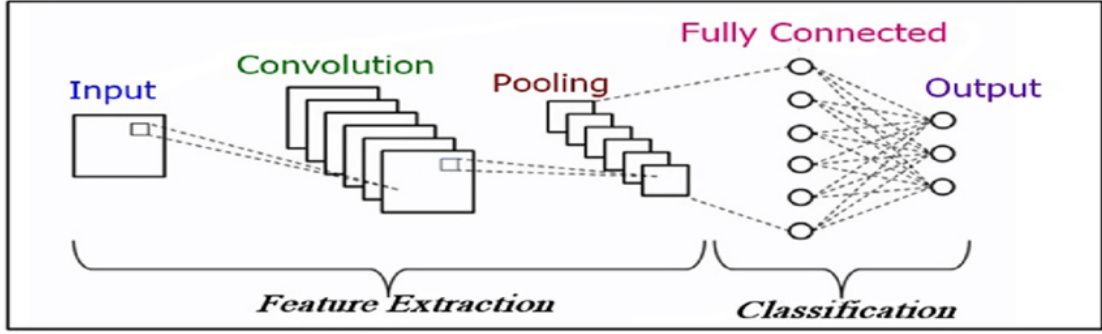


Figure 3.1: Block diagram of a basic CNN architecture [21]

### 3.3.1 VGG16

The VGG16 model [2], developed by Karen Simonyan and Andrew Zisserman from the University of Oxford, is a deep convolutional neural network that is well-known for its simplicity and effectiveness in image recognition tasks. The architecture consists of 16 weight layers, including 13 convolutional layers and 3 fully connected layers. This model significantly contributed to the success of deep learning in visual recognition tasks, as demonstrated by its performance in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2014.

In this project, the VGG16 model was utilized for image classification across different datasets, including lung, breast, and kidney images. The process involved using the base VGG16 model with some hyperparameter tuning and then enhancing it by adding a Convolutional Block Attention Module (CBAM) to improve performance further.

Below, we outline the details of how the VGG16 model was used, the benefits of using this model, the process of hyperparameter tuning, and the integration of the CBAM attention mechanism.

#### 3.3.1.1 Overview

The VGG16 model consists of 13 convolutional layers and 3 fully connected layers, making it a total of 16 weight layers. The key features of the VGG16 architecture include:

- **Convolutional Layers:** Small 3x3 receptive fields, which enable the capture of intricate patterns in the images.
- **Max-Pooling Layers:** 2x2 max-pooling layers that reduce the spatial dimensions of the feature maps.
- **Fully Connected Layers:** Three fully connected layers at the end of the network, including the final softmax layer for classification.

### 3.3.1.2 Architecture

The VGG16 architecture follows a simple pattern of stacking multiple convolutional layers followed by max-pooling layers. The convolutional layers use small 3x3 filters, which enables the network to capture intricate patterns in the images. The max-pooling layers reduce the spatial dimensions of the feature maps, leading to translation invariance. The architecture has been explain through the following [Figure 3.2](#).

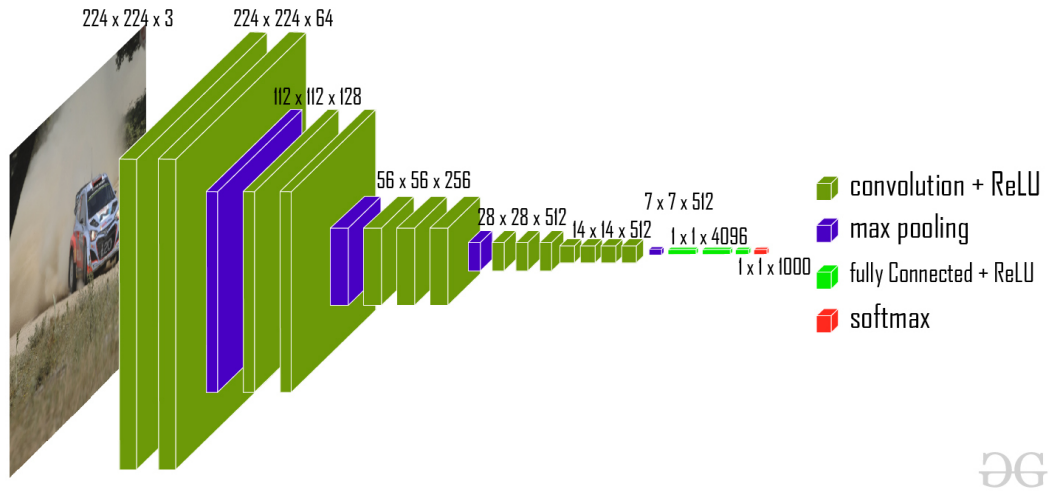


Figure 3.2: VGG16 Architecture [\[22\]](#)

### 3.3.1.3 Working Principle

During the forward pass, an input image is fed into the VGG16 network. The image undergoes a series of convolutions and pooling operations, which extract and transform features at different spatial hierarchies. The output of the convolutional layers is a set of feature maps capturing increasingly abstract representations of the input image.

### 3.3.1.4 Benefits of VGG16

- **Simplicity:** The VGG16 architecture is straightforward, making it easy to implement and understand.
- **Transfer Learning:** The pre-trained weights from ImageNet provide a strong starting point, which can be fine-tuned for specific tasks.
- **Performance:** Despite its simplicity, VGG16 achieves high performance in various image classification tasks.
- **CBAM Integration:** Adding attention mechanisms like CBAM can further improve the model's ability to focus on relevant features in the images.

### 3.3.1.5 Limitations of VGG16

- **Slow Training:** Training the VGG16 model is very slow. The original VGG model required 2-3 weeks of training on an Nvidia Titan GPU.
- **Large Model Size:** The size of the VGG16 model's pre-trained ImageNet weights is 528 MB. This large size consumes significant disk space and bandwidth, making it inefficient in terms of storage and data transfer.
- **High Parameter Count:** With 138 million parameters, VGG16 is prone to the exploding gradients problem, which can hinder the training process and make it challenging to optimize.

### 3.3.1.6 Discussion

The VGG16 model, with its deep architecture and pre-trained weights, proved to be effective for image classification tasks in this project. The addition of the CBAM attention mechanism further enhanced the model's performance, demonstrating the benefits of attention in deep learning models.

## 3.3.2 Xception

The Xception model<sup>[3]</sup>, developed by François Chollet, is an extension of the Inception architecture that replaces the standard Inception modules with depthwise separable convolutions. The name "Xception" stands for "Extreme Inception." This architecture is highly effective in image classification tasks and has been demonstrated to outperform Inception models on several benchmarks.

In this project, the Xception model was utilized for image classification across different datasets, including lung, breast, and kidney images. The process involved using the base Xception model with some hyperparameter tuning and then enhancing it with advanced techniques to further improve performance.

Below, we outline the details of how the Xception model was used, the benefits of using this model, the process of hyperparameter tuning, and its application in our project.

### 3.3.2.1 Overview

The Xception model is built on the idea of depthwise separable convolutions, which decouple the process of learning spatial features and channel-wise features. The key features of the Xception architecture include:



#### 3.3.2.4 Benefits of Xception

- **Efficiency:** Depthwise separable convolutions reduce the computational cost and the number of parameters.
- **Performance:** Xception has demonstrated superior performance on several image classification benchmarks compared to traditional Inception models.
- **Flexibility:** The architecture can be easily fine-tuned and adapted for various image classification tasks.

#### 3.3.2.5 Limitations of Xception

- **Complexity:** While more efficient than traditional convolutional layers, depthwise separable convolutions are still computationally intensive and may require substantial hardware resources.
- **Parameter Sensitivity:** The performance of the Xception model can be highly sensitive to hyperparameter choices, necessitating careful tuning.
- **Large Model Size:** The pre-trained model can be quite large, which may pose challenges in environments with limited storage or bandwidth.

#### 3.3.2.6 Discussion

The Xception model, with its depthwise separable convolution architecture and pre-trained weights, proved to be effective for image classification tasks in this project. Through hyperparameter tuning and various enhancement techniques, the model's performance was further improved, demonstrating its flexibility and robustness in deep learning applications.

### 3.3.3 DenseNet121

The DenseNet121 model<sup>[5]</sup>, developed by Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Weinberger, is a deep convolutional neural network known for its dense connectivity pattern. Unlike traditional convolutional networks, DenseNet connects each layer to every other layer in a feed-forward fashion. This architecture promotes feature reuse and improves gradient flow, leading to better performance and efficiency.

In this project, the DenseNet121 model was utilized for image classification across different datasets, including lung, breast, and kidney images. The process involved using the base DenseNet121 model with some hyperparameter tuning and

then enhancing it by adding a Convolutional Block Attention Module (CBAM) to improve performance further.

Below, we outline the details of how the DenseNet121 model was used, the benefits of using this model, the process of hyperparameter tuning, and the integration of the CBAM attention mechanism.

### 3.3.3.1 Overview

The DenseNet121 model consists of dense blocks, each containing multiple convolutional layers. The key features of the DenseNet121 architecture include:

- **Dense Connections:** Each layer receives inputs from all preceding layers, enhancing feature propagation and reuse.
- **Bottleneck Layers:**  $1 \times 1$  convolutions used to reduce the number of input feature maps, improving computational efficiency.
- **Transition Layers:** Layers that perform down-sampling between dense blocks to control the complexity of the network.

### 3.3.3.2 Architecture

The DenseNet121 architecture follows a pattern of dense connectivity, where each layer is connected to every other layer in a feed-forward manner. The architecture has been explained through [Figure 3.4](#).

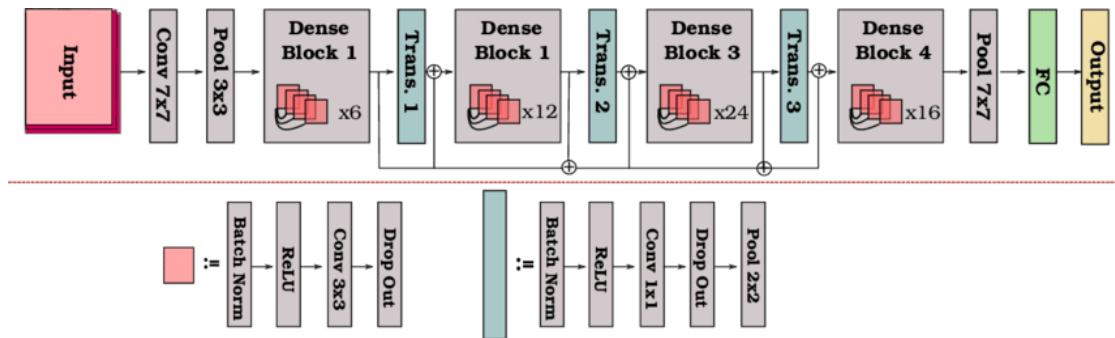


Figure 3.4: DenseNet121 Architecture [\[24\]](#)

### 3.3.3.3 Working Principle

During the forward pass, an input image is fed into the DenseNet121 network. The image undergoes a series of convolutions and pooling operations within dense blocks. Each dense block consists of several layers, each of which receives inputs from all previous layers in the block. This design promotes feature reuse and improves the gradient flow, leading to better learning and performance.

#### 3.3.3.4 Benefits of DenseNet121

- **Efficient Gradient Flow:** Dense connections ensure efficient gradient flow, making it easier to train deep networks.
- **Feature Reuse:** Each layer has direct access to the gradients from the loss function and the original input, leading to better feature reuse and learning.
- **Reduced Parameters:** Despite its deep architecture, DenseNet121 uses fewer parameters than traditional networks due to its efficient connectivity pattern.
- **CBAM Integration:** Adding attention mechanisms like CBAM can further improve the model's ability to focus on relevant features in the images.

#### 3.3.3.5 Limitations of DenseNet121

- **Memory Consumption:** The dense connections can lead to increased memory consumption, which can be a limitation for very deep networks.
- **Complexity:** The dense connectivity pattern, while beneficial, adds complexity to the model architecture.

#### 3.3.3.6 Discussion

The DenseNet121 model, with its efficient architecture and pre-trained weights, proved to be effective for image classification tasks in this project. The addition of the CBAM attention mechanism further enhanced the model's performance, demonstrating the benefits of attention in deep learning models.

### 3.3.4 NASNetMobile

The NASNetMobile model<sup>[4]</sup>, developed by Barret Zoph, Vijay Vasudevan, Jonathon Shlens, and Quoc V. Le from Google Research, is a convolutional neural network designed through neural architecture search (NAS). NASNetMobile is a smaller version of the NASNet model, optimized for mobile and embedded applications. It strikes a balance between high accuracy and computational efficiency, making it suitable for deployment on resource-constrained devices.

In this project, the NASNetMobile model was utilized for image classification across different datasets, including lung, breast, and kidney images. The process involved using the base NASNetMobile model with some hyperparameter tuning and then enhancing it by adding a Convolutional Block Attention Module (CBAM) to improve performance further.

Below, we outline the details of how the NASNetMobile model was used, the benefits of using this model, the process of hyperparameter tuning, and the integration of the CBAM attention mechanism.

#### 3.3.4.1 Overview

The NASNetMobile model is designed using an automated neural architecture search method, which optimizes the architecture for specific tasks. The key features of the NASNetMobile architecture include:

- **Cell-Based Design:** The network is composed of repeated "normal" and "reduction" cells, which were discovered through NAS.
- **Efficient Convolutions:** Uses separable convolutions to reduce the number of parameters and computational cost.
- **Scalability:** Designed to be scalable and adaptable to different computational budgets.

#### 3.3.4.2 Architecture

The NASNetMobile architecture is built from two types of cells discovered via neural architecture search: normal cells, which maintain the spatial dimensions of the input, and reduction cells, which reduce the spatial dimensions. The architecture has been explained through [Figure 3.5](#).

#### 3.3.4.3 Working Principle

During the forward pass, an input image is processed through a series of normal and reduction cells. Normal cells keep the spatial dimensions of the input feature maps constant, while reduction cells decrease the spatial dimensions. This hierarchical structure allows the network to capture features at multiple scales and resolutions, facilitating robust image classification.

#### 3.3.4.4 Benefits of NASNetMobile

- **Automated Architecture Design:** Neural architecture search optimizes the network structure for high performance.
- **Efficiency:** Uses separable convolutions to achieve high accuracy with fewer parameters and reduced computational cost.
- **Adaptability:** Scalable design allows the model to be adapted for different computational budgets and deployment scenarios.

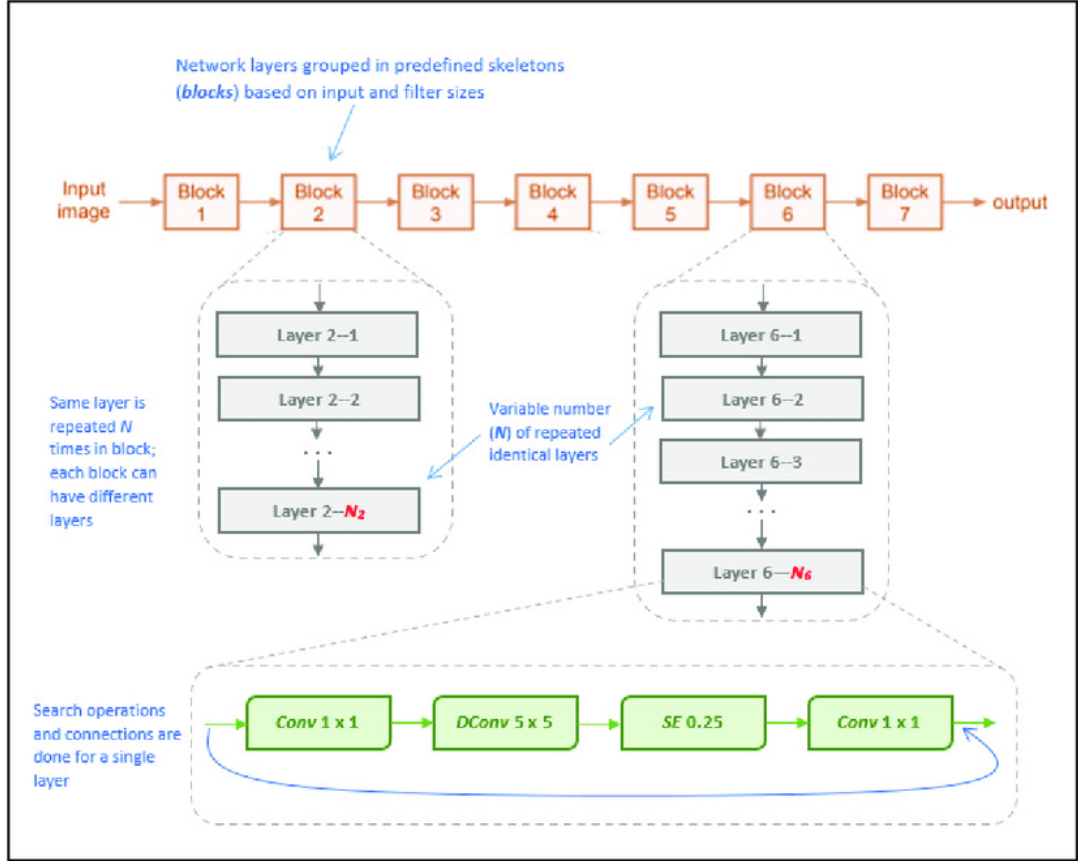


Figure 3.5: NASNetMobile Architecture [25]

- **CBAM Integration:** Adding attention mechanisms like CBAM can further improve the model's ability to focus on relevant features in the images.

#### 3.3.4.5 Limitations of NASNetMobile

- **Complexity of Architecture Search:** Neural architecture search is computationally expensive and requires significant resources.
- **Memory Consumption:** Despite being optimized for mobile, NASNetMobile can still have relatively high memory requirements compared to simpler models.

#### 3.3.4.6 Discussion

The NASNetMobile model, with its efficient architecture designed through neural architecture search, proved to be effective for image classification tasks in this project. The addition of the CBAM attention mechanism further enhanced the model's performance, demonstrating the benefits of attention in deep learning models.

### 3.3.5 EfficientNetB0

The EfficientNetB0 model<sup>[1]</sup>, developed by Mingxing Tan and Quoc V. Le from Google Research, is part of the EfficientNet family, which is known for its efficient scaling approach. EfficientNetB0 is designed to achieve high accuracy while maintaining low computational cost, making it suitable for a wide range of applications, including mobile and embedded devices.

In this project, the EfficientNetB0 model was utilized for image classification across different datasets, including lung, breast, and kidney images. The process involved using the base EfficientNetB0 model with some hyperparameter tuning and then enhancing it by adding a Convolutional Block Attention Module (CBAM) to improve performance further.

Below, we outline the details of how the EfficientNetB0 model was used, the benefits of using this model, the process of hyperparameter tuning, and the integration of the CBAM attention mechanism.

#### 3.3.5.1 Overview

EfficientNetB0 is part of a family of models that scale depth, width, and resolution in a principled manner. The key features of the EfficientNetB0 architecture include:

- **Compound Scaling:** Efficiently scales network dimensions (depth, width, resolution) using a compound coefficient.
- **Mobile Inverted Bottleneck Convolutions:** Uses a combination of depthwise separable convolutions and squeeze-and-excitation optimization.
- **Efficient Building Blocks:** Incorporates blocks that optimize both accuracy and efficiency.

#### 3.3.5.2 Architecture

The EfficientNetB0 architecture follows a systematic scaling strategy, balancing network depth, width, and resolution. The architecture has been explained through [Figure 3.6](#).

#### 3.3.5.3 Working Principle

During the forward pass, an input image is processed through a series of mobile inverted bottleneck blocks, each containing depthwise separable convolutions and squeeze-and-excitation layers. These blocks effectively capture and enhance features while maintaining computational efficiency.

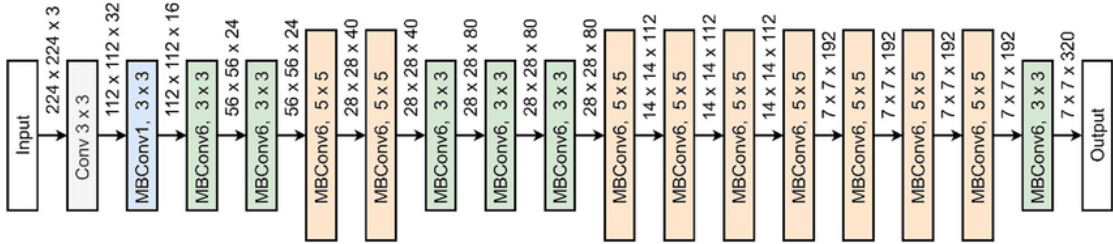


Figure 3.6: EfficientNetB0 Architecture [26]

#### 3.3.5.4 Benefits of EfficientNetB0

- **Efficiency:** The compound scaling method allows EfficientNetB0 to achieve high performance with fewer parameters and lower computational cost.
- **Versatility:** Suitable for deployment on various platforms, including mobile and embedded devices, due to its efficient design.
- **Performance:** Achieves state-of-the-art accuracy on multiple benchmarks with significantly fewer resources compared to other models.
- **CBAM Integration:** Adding attention mechanisms like CBAM can further improve the model's ability to focus on relevant features in the images.

### 3.3.5.5 Limitations of EfficientNetB0

- **Complexity of Training:** Training EfficientNet models from scratch requires significant computational resources.
- **Sensitivity to Hyperparameters:** The performance of EfficientNetB0 can be sensitive to the choice of hyperparameters, requiring careful tuning.

### 3.3.5.6 Discussion

The EfficientNetB0 model, with its efficient architecture and pre-trained weights, proved to be effective for image classification tasks in this project. The addition of the CBAM attention mechanism further enhanced the model's performance, demonstrating the benefits of attention in deep learning models.

### 3.4 Attention Module

Attention mechanisms have become a vital component in modern deep learning architectures, especially in tasks involving sequential data and image processing. Attention allows models to focus on specific parts of the input data, dynamically highlighting relevant features while ignoring less important ones. This capability

significantly enhances the performance of models by improving their ability to capture intricate patterns and relationships within the data.

### 3.4.1 Convolutional Block Attention Module

The Convolutional Block Attention Module (CBAM) is an attention mechanism designed to improve the representational power of convolutional neural networks. CBAM operates by sequentially applying channel and spatial attention modules, which enables the model to focus on important features across both dimensions.

Given an intermediate feature map  $F \in \mathbb{R}^{C \times H \times W}$  as input, CBAM sequentially infers a 1D channel attention map  $M_c \in \mathbb{R}^{C \times 1 \times 1}$  and a 2D spatial attention map  $M_s \in \mathbb{R}^{1 \times H \times W}$ . The overall attention process can be summarized as:

$$F' = M_c(F) \otimes F, \quad (3.1)$$

$$F'' = M_s(F') \otimes F', \quad (3.2)$$

where  $\otimes$  denotes element-wise multiplication. During multiplication, the attention values are broadcasted (copied) accordingly: channel attention values are broadcasted along the spatial dimension, and vice versa.  $F''$  is the final refined output. [Figure 3.8](#) depicts the computation process of each attention map. The following describes the details of each attention module.

#### 3.4.1.1 How CBAM Works

CBAM enhances the learning process by refining feature maps through two distinct attention operations:

- **Channel Attention Module:** This module emphasizes informative channels while suppressing less useful ones. It does so by leveraging both global average pooling and global max pooling to aggregate information across spatial dimensions, followed by a shared network that generates attention maps for channels.
- **Spatial Attention Module:** This module focuses on significant spatial locations. It uses average pooling and max pooling operations to create a pooled feature map, which is then convolved to produce a spatial attention map.

The combined effect of these modules allows the network to enhance its focus on the most pertinent features, thereby improving overall performance in tasks

like image classification and object detection. Figure 3.7 and Figure 3.8 illustrate it better.

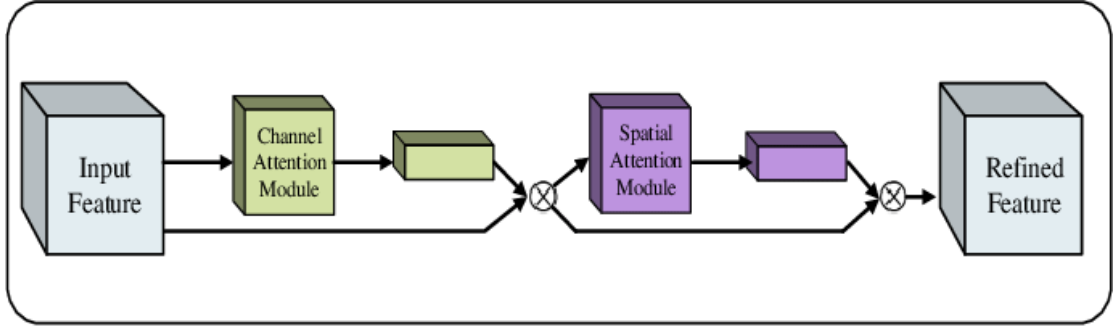


Figure 3.7: Overview of CBAM [27]

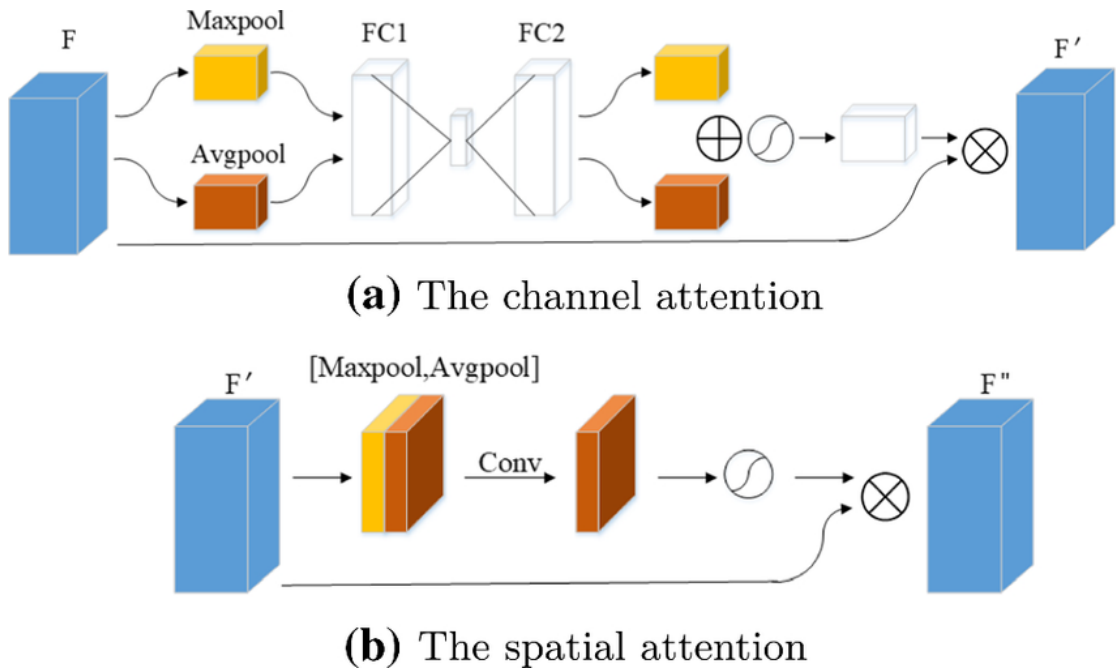


Figure 3.8: Channel Attention and Spatial Attention [27]

#### 3.4.1.2 Benefits of Using CBAM

- **Enhanced Feature Representation:** By focusing on important features both channel-wise and spatially, CBAM improves the model's ability to capture relevant information.
- **Versatility:** CBAM can be easily integrated into any convolutional neural network architecture, making it a flexible choice for various applications.
- **Improved Performance:** Models with CBAM generally exhibit better performance due to the refined attention mechanism that helps in focusing on crucial parts of the input data.

## 3.5 Procedure of Classification

In this project, we employed a systematic approach to classify medical images using various Convolutional Neural Network (CNN) models. The following steps outline the procedure followed for each of the five CNN models used in the classification tasks:

### 3.5.0.1 Base Model Implementation

- **Model Initialization:** Each CNN model was initialized with pre-trained weights from the ImageNet dataset to leverage the benefits of transfer learning.
- **Hyperparameter Tuning:** Key hyperparameters, including learning rate, batch size, and optimizer type, were tuned to achieve optimal performance on the classification tasks.
- **Evaluation:** The performance of the base CNN models was evaluated on the lung, breast, and kidney datasets, and the results were stored for comparison.

### 3.5.0.2 Enhancement with CBAM Attention

- **Extraction of Last Convolutional Layer:** The last convolutional layer of each CNN model was extracted.
- **Integration of CBAM:** A Convolutional Block Attention Module (CBAM) was added to the extracted convolutional layer to enhance the model's attention mechanism.
- **Global Average Pooling:** Applied global average pooling to reduce the spatial dimensions of the feature maps.
- **Dense Layers:** Added dense layers to build the final classification model.
- **Evaluation:** The enhanced CNN models with CBAM were evaluated on the same datasets, and the results were compared with the base models.

### 3.5.1 Discussion

The classification procedure outlined above was systematically applied to each of the five CNN models to ensure a robust and comprehensive evaluation. By comparing the performance of the base models with their CBAM-enhanced counterparts, we were able to determine the effectiveness of integrating attention mechanisms in improving classification accuracy. The results from these evaluations

provide valuable insights into the potential of advanced deep learning techniques in medical image classification.

Each of these models brings unique strengths to the table, contributing to the overall effectiveness of our classification tasks on lung, breast, and kidney datasets. By leveraging these diverse architectures, we aim to achieve robust and accurate classification results across different medical image datasets. In conclusion, the combination of these advanced CNN architectures provides a comprehensive approach to medical image classification. By analyzing the performance of VGG16, Xception, DenseNet121, NASNetMobile, and EfficientNetB0, we can determine the most effective model for each specific dataset. This comparative study not only enhances our understanding of the capabilities of each model but also contributes to the development of more accurate and efficient diagnostic tools in medical imaging.

# Results and Discussion

In this section, we present and analyze the performance results of the CNN models on three distinct medical image datasets: lung cancer, breast cancer, and kidney disease. We evaluate the effectiveness of the baseline CNN models and their enhanced versions with the Convolutional Block Attention Module (CBAM). The evaluation metrics used to assess the models include Accuracy, Precision, Recall, and F1 Score. These metrics provide a comprehensive view of the models' capabilities in correctly classifying medical images and handling class imbalances. Additionally, we discuss the impact of hyperparameter tuning on the models' performance, illustrating how careful adjustment of parameters such as learning rate, number of epochs, and optimizer settings can lead to significant improvements. Through detailed comparison and analysis, we aim to highlight the benefits of incorporating attention mechanisms and optimized hyperparameters in CNN models for medical image classification.

## 4.1 System configuration

A Jupyter Notebook with a 12 GB NVIDIA Tesla T4 GPU in Google's collaborative environment is used to run the whole set of experiments. Tensorflow, Keras, Matplotlib, Scikit, Numpy, and Pandas are the main open-source modules that have been used for implementing the suggested method in Python.

## 4.2 Evaluation Metrics

Evaluation metrics are essential tools for assessing the performance of machine learning models, particularly in classification tasks. They provide quantitative measures that help in understanding how well the model is performing and in comparing different models. Key evaluation metrics include Accuracy, Precision,

Recall, and F1 Score. These metrics give insights into different aspects of model performance, such as the proportion of correct predictions, the balance between sensitivity and specificity, and the trade-off between precision and recall.

**Accuracy:** The percentage of events that are correctly predicted over all instances is known as accuracy. It is employed to assess how well a model performs in terms of producing accurate predictions. It is determined by dividing the total number of predictions made by the model by the number of predictions that were accurate. Mathematically, Equation 4.1 represents the accuracy.

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \quad (4.1)$$

**Precision:** The proportion of accurate positive predictions amongst all positive predictions made by a model is known as its precision. It is determined by dividing  $T_P$  by the product of  $T_P$  and  $F_P$ . Mathematically, precision can be represented as in Equation 4.2.

$$Precision = \frac{T_P}{T_P + F_P} \quad (4.2)$$

**Recall:** Recall is the percentage of actual positive events that the model correctly identifies. It is the proportion of  $T_P$  to  $T_P$  plus  $F_N$  as shown in Equation 4.3. A higher recall demonstrates that the model is better at identifying positive examples, whereas a lower recall raises the possibility that some positive cases may have been missed by the model.

$$Recall = \frac{T_P}{T_P + F_N} \quad (4.3)$$

**F1 Score** The F1 Score is a harmonic mean of Precision and Recall, providing a single metric that balances both concerns. It is particularly useful in scenarios where the class distribution is imbalanced. The F1 Score is defined in the Equation 4.4:

$$F1 \text{ Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4.4)$$

**Confusion Matrix** A confusion matrix is a table used to describe the performance of a classification model by comparing the predicted labels with the true labels. It provides a comprehensive view of the model's performance by showing the counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The confusion matrix is structured as follows in the following Table 4.1:

Table 4.1: Confusion Matrix

|                        | <b>Predicted Positive</b> | <b>Predicted Negative</b> |
|------------------------|---------------------------|---------------------------|
| <b>Actual Positive</b> | TP                        | FN                        |
| <b>Actual Negative</b> | FP                        | TN                        |

The confusion matrix helps in understanding not just the accuracy of a model but also the types of errors it makes. This detailed breakdown is invaluable for diagnosing issues such as class imbalance or model bias, and for guiding further improvements in model training and tuning.

### 4.3 Datasets

In this section, we meticulously dissect and elucidate the intricacies of the diverse datasets that constitute the cornerstone of our classification endeavor. The lung cancer dataset, a compendium of high-resolution CT scans meticulously curated from renowned medical repositories, encompasses a plethora of pathological manifestations, including diverse tumor subtypes and pulmonary abnormalities. Furthermore, the breast cancer dataset, meticulously collated from esteemed medical archives, encapsulates a rich tapestry of mammographic images portraying benign and malignant lesions, facilitating the nuanced exploration of breast pathology. Lastly, the kidney tumor dataset, sourced from reputable medical databases, offers a comprehensive array of CT scans depicting renal anomalies ranging from cysts and stones to malignant tumors, enabling a profound examination of renal pathology. Through meticulous curation, preprocessing, and annotation, these datasets serve as invaluable assets in our pursuit of accurate and robust classification methodologies.

#### 4.3.1 IQ-OTH/NCCD Lung Cancer Dataset

Lung cancer detection is a critical aspect of reducing mortality rates associated with this aggressive disease. Early and accurate detection significantly improves the chances of successful treatment and long-term survival. Various imaging techniques, such as chest X-rays, CT scans, and PET scans, are employed to identify lung cancer at its earliest stages. Recently, advancements in machine learning and deep learning have shown promising results in enhancing the accuracy and efficiency of lung cancer detection. Convolutional Neural Networks (CNNs) are particularly effective in analyzing medical images to distinguish between normal,

benign, and malignant lung tissues, thereby aiding radiologists in making precise diagnoses and developing appropriate treatment plans.

#### 4.3.1.1 Dataset Collection

The Iraq-Oncology Teaching Hospital/National Center for Cancer Diseases (IQ-OTH/NCCD) lung cancer dataset [28] was collected from various specialist hospitals over a period of three months in the fall of 2019. This dataset includes CT scans of patients diagnosed with lung cancer at different stages, as well as scans of healthy subjects. The cases in the dataset are categorized into three classes: normal, benign, and malignant. Specifically, there are 40 cases diagnosed as malignant, 15 cases diagnosed as benign, and 55 cases classified as normal.

The CT scans were originally collected in DICOM format using a SOMATOM scanner from Siemens. The scanning protocol included settings of 120 kV and a slice thickness of 1 mm. The window width ranged from 350 to 1200 HU, and the window center ranged from 50 to 600 HU, with images being read during breath-hold at full inspiration. All images were de-identified before analysis to ensure patient privacy. Written consent was waived by the institutional oversight and review board of the participating medical centers, which approved the study.

Each scan contains several slices, with the number of slices ranging from 80 to 200 per scan. Each slice represents an image of the human chest from different sides and angles. The 110 cases included in the dataset vary in gender, age, educational attainment, area of residence, and living status. The subjects include employees of the Iraqi Ministries of Transport and Oil, as well as farmers and gainers. Most subjects come from the central region of Iraq, particularly the provinces of Baghdad, Wasit, Diyala, Salahuddin, and Babylon. The following Table 4.2 shows the number of images for each case. The data samples are illustrated in Figure 4.1.

Table 4.2: The three classes of lung cases and the number of images in each case

| Case      | Number of Images |
|-----------|------------------|
| Benign    | 120              |
| Malignant | 561              |
| Normal    | 416              |
| Total     | 1097             |

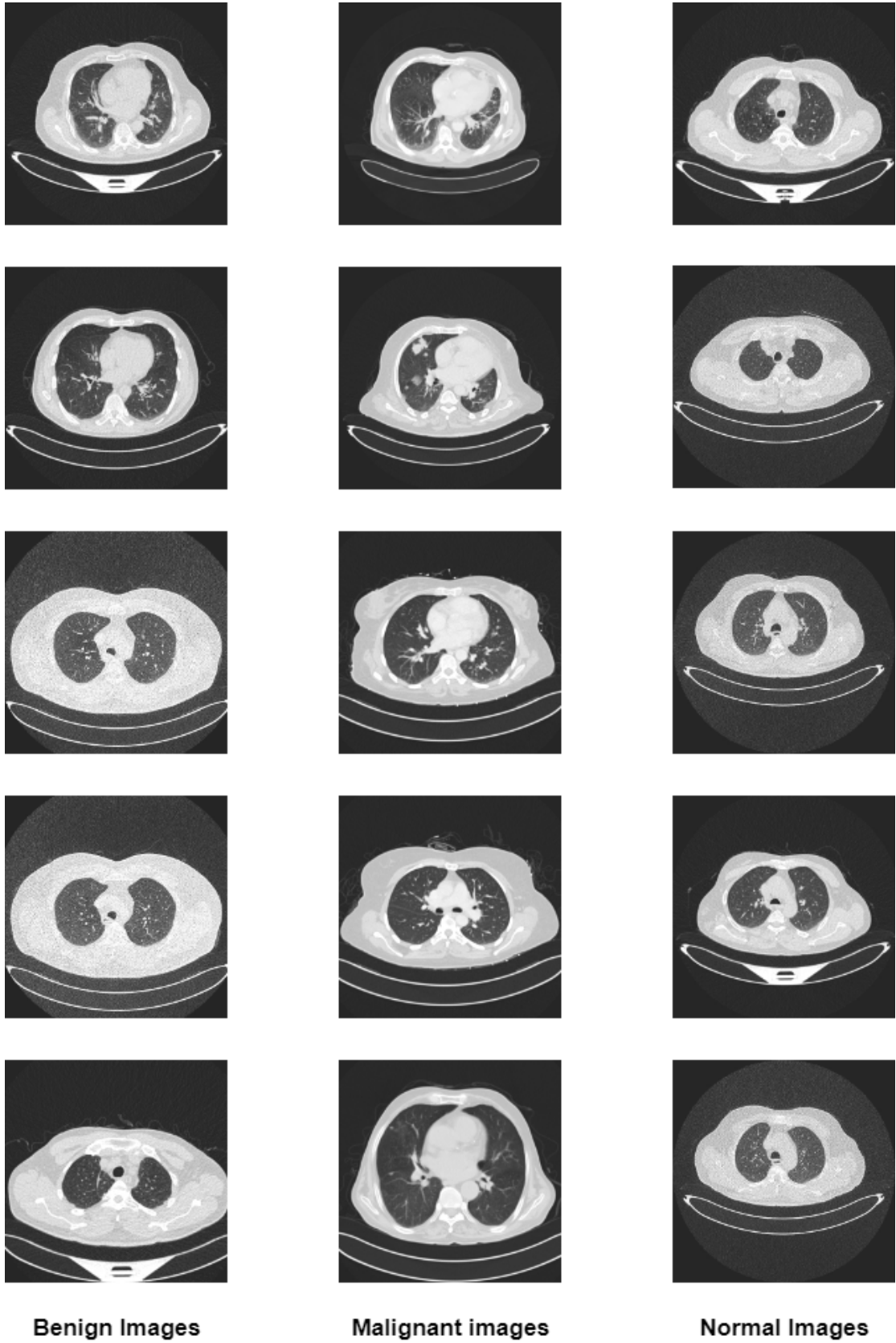


Figure 4.1: Samples of IQ-OTH/NCCD Lung Cancer Dataset

#### 4.3.1.2 Value of the Data

- **Importance of Early Detection:** The dataset plays a crucial role in facilitating early detection of lung cancer, which is essential for improving patient outcomes and reducing mortality rates associated with this aggressive disease.
- **Training and Evaluation Resource:** Researchers and medical professionals can utilize this dataset as a valuable resource for training and evaluating machine learning models and algorithms designed for lung cancer detection and classification.
- **Diversity of Lung Cancer Cases:** The dataset contains a diverse range of lung cancer cases, including different stages of malignancy, benign conditions, and normal cases, providing a comprehensive representation of lung cancer pathology.
- **Real-World Representation:** Reflecting real-world clinical scenarios, the dataset encompasses scans from various patients with varying demographics, including gender, age, educational background, and area of residence, contributing to the robustness and generalizability of machine learning models.
- **Contribution to Research Advancements:** By sharing this dataset, researchers contribute to the advancement of medical research in the field of lung cancer detection and diagnosis, fostering collaboration and innovation in developing effective screening and diagnostic tools.

### 4.3.2 Ultrasound Breast Images Dataset

Breast cancer is one of the most common causes of death among women worldwide. Early detection helps in reducing the number of early deaths. The data reviews the medical images of breast cancer using ultrasound scan. Breast Ultrasound Dataset is categorized into three classes: normal, benign, and malignant images. Breast ultrasound images can produce great results in classification, detection, and segmentation of breast cancer when combined with machine learning.

#### 4.3.2.1 Dataset Collection

The breast-ultrasound-images-dataset<sup>[29]</sup> collected at baseline include breast ultrasound images among women in ages between 25 and 75 years old. This data was collected in 2018. The number of patients is 600 female patients. The dataset consists of 780 images with an average image size of 500\*500 pixels. The

images are in PNG format. The ground truth images are presented with original images. The images are categorized into three classes, which are normal, benign, and malignant.

Ultrasound (US) images are generally in grayscale. They were collected and stored in a DICOM format at Baheya hospital. The consumed time used to collect and annotate the images is about one year. US dataset is categorized into three classes: normal, benign, and malignant. At the beginning, the number of images collected was 1100. After performing preprocessing to the dataset, the number of images was reduced to 780 images. The original images contain unimportant information not used for mass classification. Moreover, they may affect the output results of the training process. The instruments used in the scanning process are LOGIQ E9 ultrasound system and LOGIQ E9 Agile ultrasound system. These instruments are usually used in top-notch imaging for radiology, cardiac and vascular application. They produce image resolution of 1280\*1024. The transducers are 1–5 MHz on ML6-15-D Matrix linear probe. The number of images in each class is shown in [Table 4.3](#). The data samples are illustrated in [Figure 4.2](#).

Table 4.3: The three classes of breast cases and the number of images in each case

| Case      | Number of Images |
|-----------|------------------|
| Benign    | 487              |
| Malignant | 210              |
| Normal    | 133              |
| Total     | 780              |

#### 4.3.2.2 Value of the Data

- **Primary Diagnostic Tool:** Ultrasound scanning represents a primary diagnostic tool for breast cancer examination and early detection, offering a safe alternative to other radiological imaging techniques.
- **Resource for Machine Learning:** The Breast Ultrasound dataset serves as a valuable resource for training machine learning models geared towards the classification, detection, and segmentation of early signs of breast cancer masses or micro-calcifications.
- **Research Utility:** Researchers interested in the classification, detection, and segmentation of breast cancer can leverage this dataset of breast ultrasound images, either independently or in combination with other datasets, to conduct in-depth analyses and derive valuable insights.

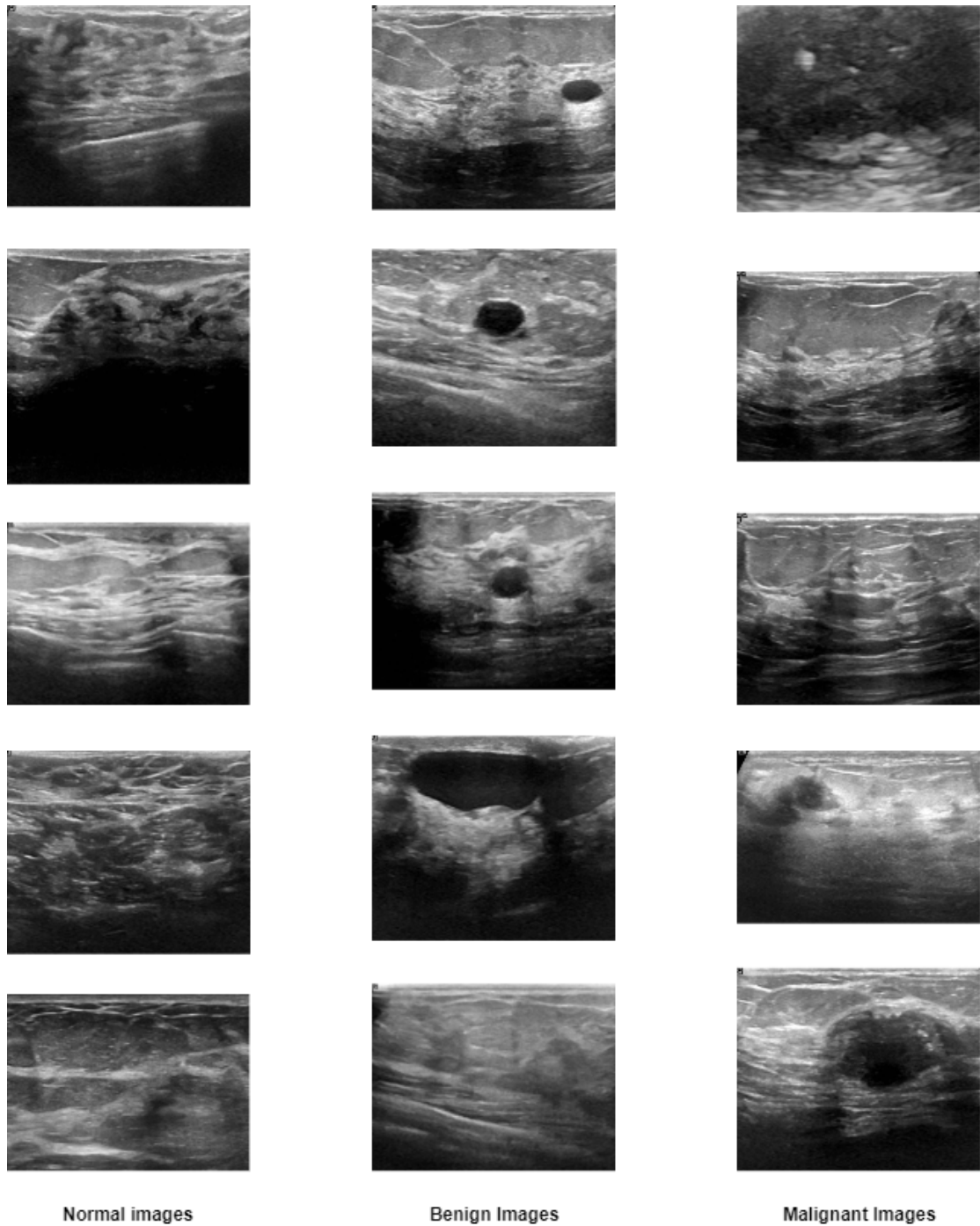


Figure 4.2: Samples of Ultrasound Breast Images Dataset

- **Comprehensive Nature:** With its comprehensive nature, encompassing breast cancer states ranging from normal to benign and malignant, this dataset offers a diverse range of scenarios for analysis and model development.
- **Pioneering Resource:** This dataset stands as a pioneering resource in its domain, representing the first publicly available breast ultrasound dataset to the best of our knowledge.

### 4.3.3 Ct kidney dataset: normal-cyst-Tumor and Stone

Kidney stones and tumors are serious conditions that can significantly affect a patient's health if not detected and treated early. Kidney stones are hard deposits made of minerals and salts that form inside the kidneys, causing severe pain, blood in urine, and potential kidney damage. Tumors in the kidneys can be either benign or malignant, with the latter posing a serious threat as it can lead to kidney cancer if not diagnosed and treated promptly.

Detecting kidney stones and tumors typically involves imaging techniques such as CT scans, which provide detailed pictures of the kidneys and surrounding structures. These images allow radiologists to identify the presence, size, and location of stones and tumors. Machine learning and deep learning techniques have been increasingly employed to assist in the automated detection and classification of these conditions, improving diagnostic accuracy and enabling quicker intervention. The development and application of such advanced methods are crucial for enhancing patient outcomes and advancing medical research in nephrology.

#### 4.3.3.1 Dataset Collection

The dataset [30] was collected from PACS (Picture archiving and communication system) from different hospitals in Dhaka, Bangladesh where patients were already diagnosed with having a kidney tumor, cyst, normal or stone findings. Both the Coronal and Axial cuts were selected from both contrast and non-contrast studies with protocol for the whole abdomen and urogram. The Dicom study was then carefully selected, one diagnosis at a time, and from those the researchers created a batch of Dicom images of the region of interest for each radiological finding. Following that, they excluded each patient's information and meta data from the Dicom images and converted the Dicom images to a lossless jpg image format. After the conversion, each image finding was again verified by a radiologist and a medical technologist to reconfirm the correctness of the data.

This dataset contains 12,446 unique data within it in which the cyst contains 3,709, normal 5,077, stone 1,377, and tumor 2,283 which is shown in the following Table 4.4. The data samples are illustrated in Figure 4.3.

#### 4.3.3.2 Value of the Data

- **Improves Diagnostic Accuracy:** The dataset can be used to train machine learning models to improve the accuracy of kidney disease diagnosis.
- **Supports Medical Research:** Provides a comprehensive set of labeled

Table 4.4: The Four classes of kidney cases and the number of images in each case

| Case   | Number of Images |
|--------|------------------|
| cyst   | 3,709            |
| normal | 5,077            |
| stone  | 1,377            |
| tumor  | 2,283            |
| total  | 12,446           |

data that can be used for research in medical imaging and diagnostic techniques.

- **Facilitates Early Detection:** Helps in the early detection of kidney conditions, potentially leading to better patient outcomes.
- **Enables AI Development:** Supports the development and validation of AI algorithms for automated kidney disease detection and classification.
- **Publicly Available Resource:** Being one of the few datasets with diverse kidney conditions, it offers a valuable resource for the medical imaging community.

## 4.4 Results

This section presents the results of our experiments on three medical image datasets: lung cancer, breast cancer, and kidney disease. We evaluate each of the five CNN models with and without the integration of the Convolutional Block Attention Module (CBAM). The performance metrics considered are Accuracy, Precision, Recall, and F1 Score.

### 4.4.1 Hyperparameters

Hyperparameter tuning is a crucial process in the development and optimization of machine learning models, particularly deep learning architectures such as Convolutional Neural Networks (CNNs). Hyperparameters are the configuration settings used to structure and train the model, as opposed to the model parameters which are learned from the training data. Common hyperparameters include the learning rate, number of epochs, batch size, optimizer type, and loss function. Proper tuning of these hyperparameters is essential because they have a significant impact on the model's performance, influencing how well the model learns from the

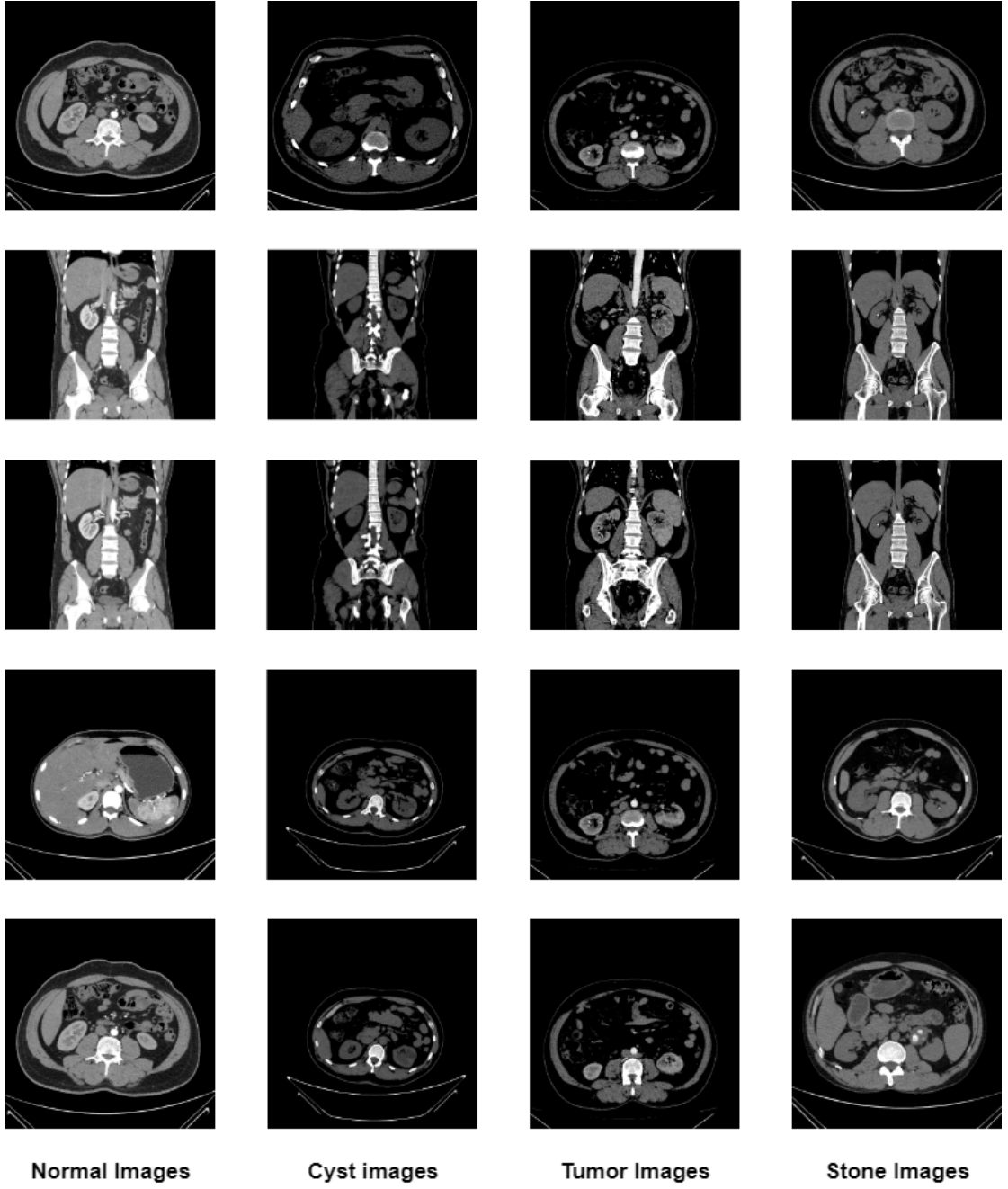


Figure 4.3: Samples of Ct kidney dataset: normal-cyst-Tumor and Stone

data and generalizes to unseen data. Hyperparameter tuning involves systematically testing different combinations of hyperparameters and selecting the set that yields the best performance according to predefined metrics, such as accuracy, precision, recall, and F1 score. This process can be computationally intensive, but it is critical for achieving optimal model performance, preventing overfitting, and ensuring that the model can effectively capture the underlying patterns in the data. [Table 4.5](#) shows the hyperparameters which were tuned for all models to achieve optimal performance:

Table 4.5: Hyperparameters used in the models

| Hyperparameter   | Value              |
|------------------|--------------------|
| Number of Epochs | 20                 |
| Loss Function    | Cross-Entropy Loss |
| Optimizer        | Adam               |
| Learning Rate    | 0.001              |

#### 4.4.2 IQ-OTH/NCCD Lung Cancer Dataset Results

The performance of each model on the lung cancer dataset is summarized in Table 4.6. The models are evaluated based on the Accuracy, Precision, Recall, and F1 Score metrics.

Table 4.6: Performance of models on the lung cancer dataset

| Model          | Attention Used | Accuracy | Precision | Recall | F1 Score |
|----------------|----------------|----------|-----------|--------|----------|
| VGG16          | No             | 0.94     | 0.97      | 0.97   | 0.97     |
| VGG16          | CBAM           | 0.98     | 1.00      | 0.98   | 0.98     |
| Xception       | No             | 0.85     | 0.84      | 0.86   | 0.85     |
| Xception       | CBAM           | 0.89     | 0.88      | 0.90   | 0.89     |
| EfficientNetB0 | No             | 0.92     | 0.91      | 0.93   | 0.92     |
| EfficientNetB0 | CBAM           | 0.95     | 0.94      | 0.96   | 0.95     |
| NasNetMobile   | No             | 0.93     | 0.92      | 0.94   | 0.93     |
| NasNetMobile   | CBAM           | 0.96     | 0.95      | 0.97   | 0.96     |
| DenseNet121    | No             | 0.81     | 0.80      | 0.82   | 0.81     |
| DenseNet121    | CBAM           | 0.86     | 0.85      | 0.87   | 0.86     |

#### 4.4.3 Ultrasound Breast Images Dataset Results

The performance of each model on the breast cancer dataset is summarized in Table 4.7. The models are evaluated based on the Accuracy, Precision, Recall, and F1 Score metrics.

Table 4.7: Performance of models on the breast cancer dataset

| Model          | Attention Used | Accuracy | Precision | Recall | F1 Score |
|----------------|----------------|----------|-----------|--------|----------|
| VGG16          | No             | 0.82     | 0.93      | 0.95   | 0.94     |
| VGG16          | CBAM           | 0.86     | 0.97      | 0.99   | 0.98     |
| Xception       | No             | 0.92     | 0.98      | 0.98   | 0.98     |
| Xception       | CBAM           | 0.98     | 0.99      | 0.99   | 0.99     |
| EfficientNetB0 | No             | 0.92     | 0.91      | 0.93   | 0.92     |
| EfficientNetB0 | CBAM           | 0.96     | 0.95      | 0.97   | 0.96     |
| NasNetMobile   | No             | 0.93     | 0.92      | 0.94   | 0.93     |
| NasNetMobile   | CBAM           | 0.97     | 0.96      | 0.98   | 0.97     |
| DenseNet121    | No             | 0.85     | 0.84      | 0.86   | 0.85     |
| DenseNet121    | CBAM           | 0.88     | 0.87      | 0.89   | 0.88     |

#### 4.4.4 Ct kidney dataset: normal-cyst-Tumor and Stone Results

The performance of each model on the kidney disease dataset is summarized in [Table 4.8](#). The models are evaluated based on the Accuracy, Precision, Recall, and F1 Score metrics.

Table 4.8: Performance of models on the kidney disease dataset

| Model          | Attention Used | Accuracy | Precision | Recall | F1 Score |
|----------------|----------------|----------|-----------|--------|----------|
| VGG16          | No             | 0.91     | 0.93      | 0.95   | 0.94     |
| VGG16          | CBAM           | 0.93     | 0.96      | 0.98   | 0.97     |
| Xception       | No             | 0.85     | 0.84      | 0.86   | 0.85     |
| Xception       | CBAM           | 0.89     | 0.88      | 0.90   | 0.89     |
| EfficientNetB0 | No             | 0.92     | 0.91      | 0.93   | 0.92     |
| EfficientNetB0 | CBAM           | 0.96     | 0.95      | 0.97   | 0.96     |
| NasNetMobile   | No             | 0.91     | 0.90      | 0.92   | 0.91     |
| NasNetMobile   | CBAM           | 0.95     | 0.94      | 0.96   | 0.95     |
| DenseNet121    | No             | 0.83     | 0.82      | 0.84   | 0.83     |
| DenseNet121    | CBAM           | 0.86     | 0.85      | 0.87   | 0.86     |

#### 4.4.5 Confusion Matrices of the Datasets

Here we provide the confusion metrics of two datasets for the models which has shown best results. The analysis of confusion matrices provides deeper insights into the performance of the classification models beyond simple accuracy metrics. These matrices offer a detailed breakdown of the classification results, showcasing the true positive, true negative, false positive, and false negative predictions for each class.

##### 4.4.5.1 IQ-OTH/NCCD Lung Cancer Dataset

Here are the confusion matrices for the IQ-OTH/NCCD Lung Cancer Dataset on VGG16 model. [Figure 4.4](#) shows the performance of VGG16 on this dataset without using attention and [Figure 4.5](#) shows the performance of VGG16 on this dataset using CBAM attention.

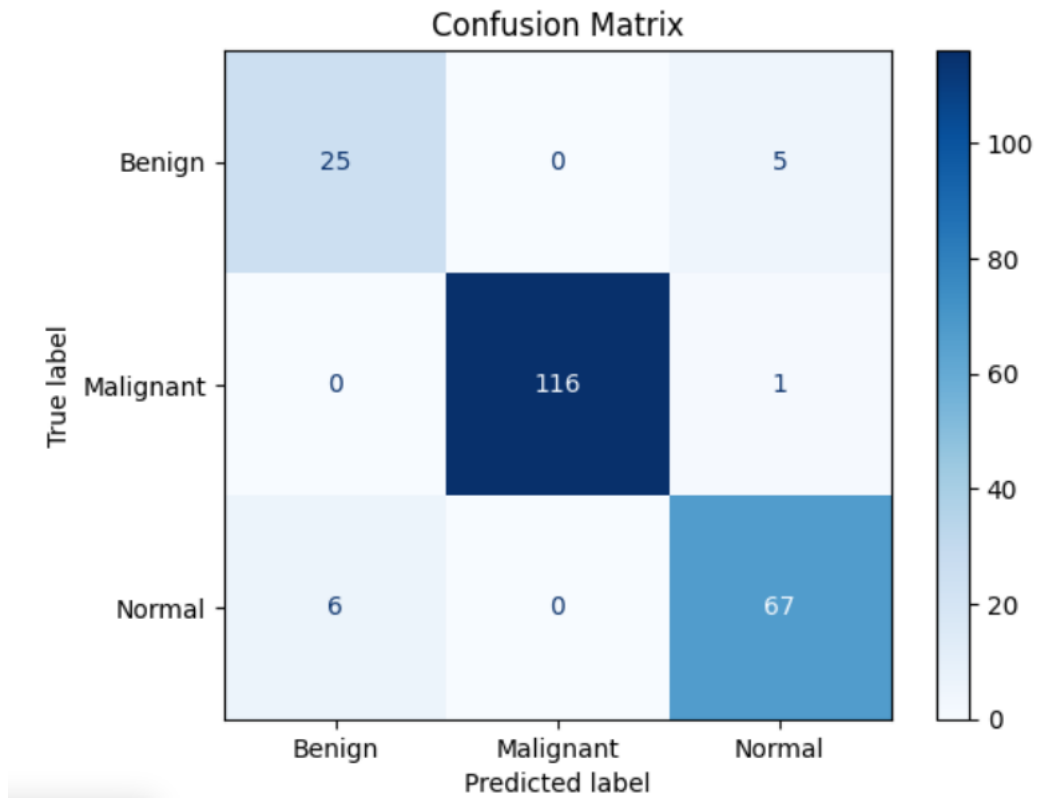


Figure 4.4: Confusion Matrix of the IQ-OTH/NCCD Lung Cancer Dataset using VGG16 without attention

##### 4.4.5.2 Ultrasound Breast Images Dataset

Here are the confusion matrices for the Ultrasound Breast Images Dataset on Xception model. [Figure 4.6](#) shows the performance of Xception on this dataset

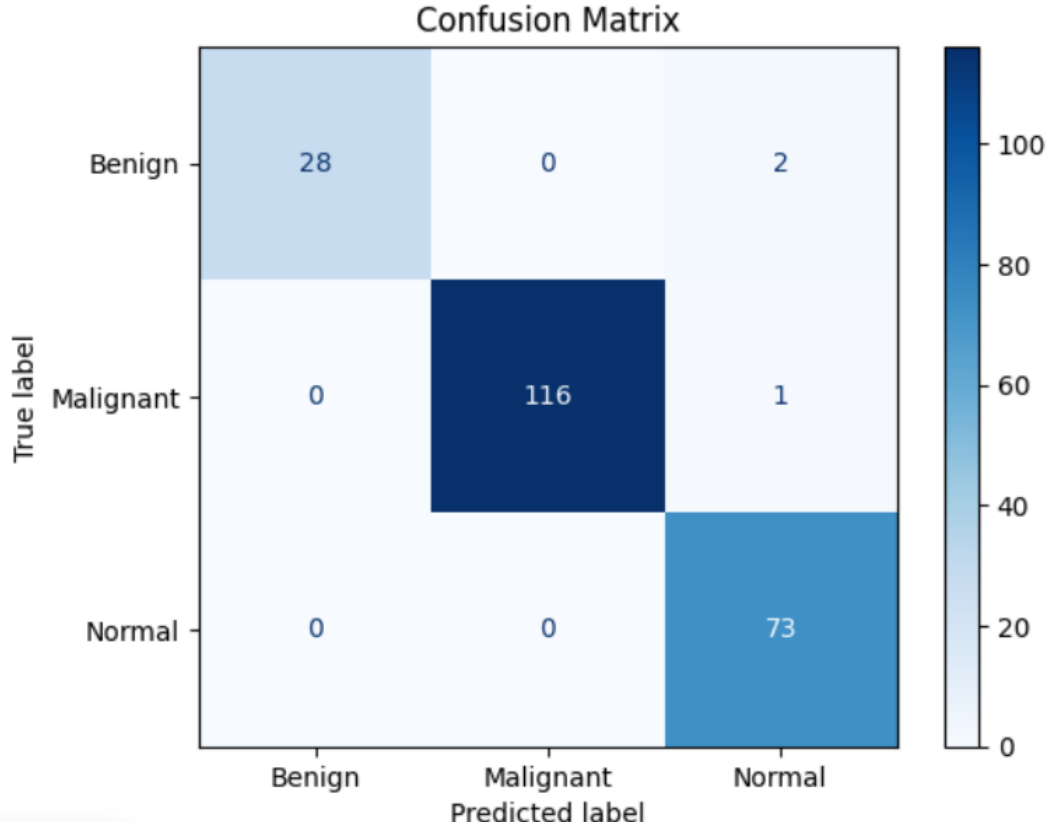


Figure 4.5: Confusion Matrix of the IQ-OTH/NCCD Lung Cancer Dataset using VGG16 with CBAM attention

without using attention and [Figure 4.7](#) shows the performance of xception on this dataset using CBAM attention.

## 4.5 Discussion

In this project, we evaluated the performance of various deep learning models, specifically VGG16, Xception, EfficientNetB0, NasNetMobile, and DenseNet121, on different medical image datasets, including lung cancer, breast cancer, and kidney disease, focusing on the impact of the Convolutional Block Attention Module (CBAM) on classification accuracy. Across all datasets, the integration of CBAM consistently improved model performance. For the lung cancer dataset, VGG16 with CBAM achieved the highest accuracy of 98%, significantly outperforming its non-CBAM counterpart at 95%. Similar trends were observed in the breast cancer dataset, where VGG16 and EfficientNetB0 with CBAM reached 98% and 97% accuracy, respectively. NasNetMobile also benefited greatly from CBAM, achieving 96% accuracy in both lung and breast cancer datasets. For kidney disease classification, VGG16 again led with 97% accuracy using CBAM, compared to 94% without, while EfficientNetB0 and NasNetMobile with CBAM achieved 96% ac-

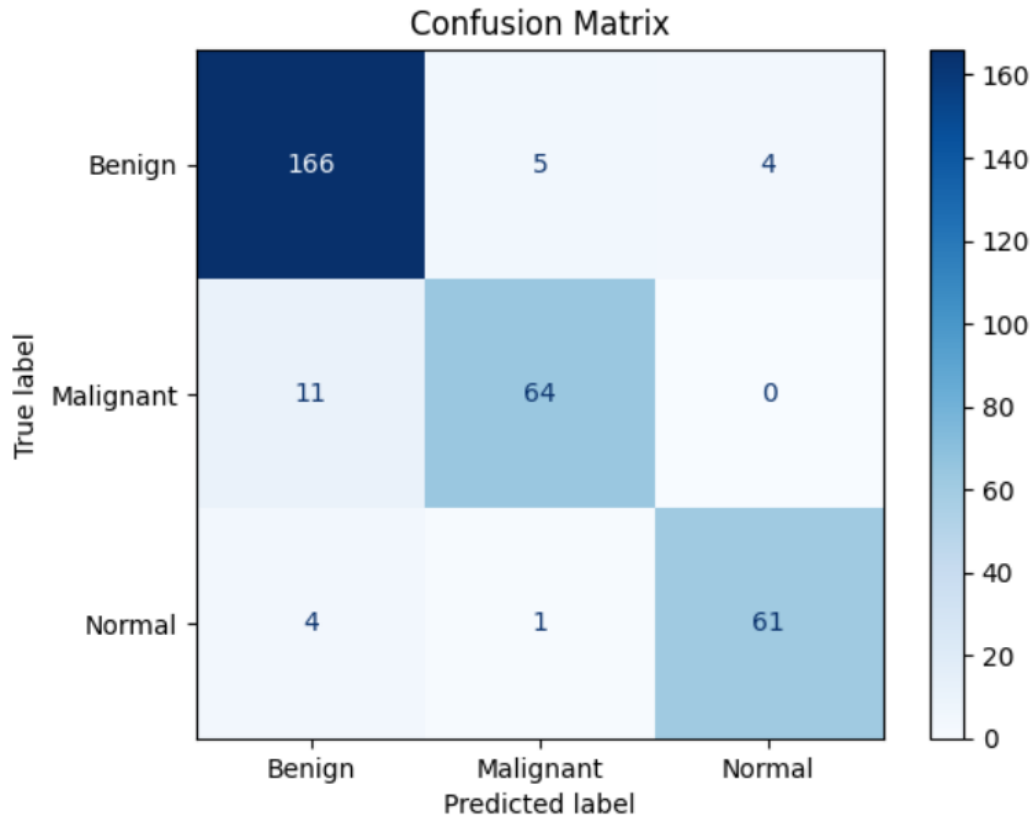


Figure 4.6: Confusion Matrix of the Ultrasound Breast Images Dataset using Xception without attention

curacy. DenseNet121, although not the top performer, still showed improvements with CBAM. Overall, VGG16 with CBAM emerged as the most effective model across all datasets, underscoring the significant role of attention mechanisms in enhancing the feature extraction and classification capabilities of deep learning models in medical image analysis.

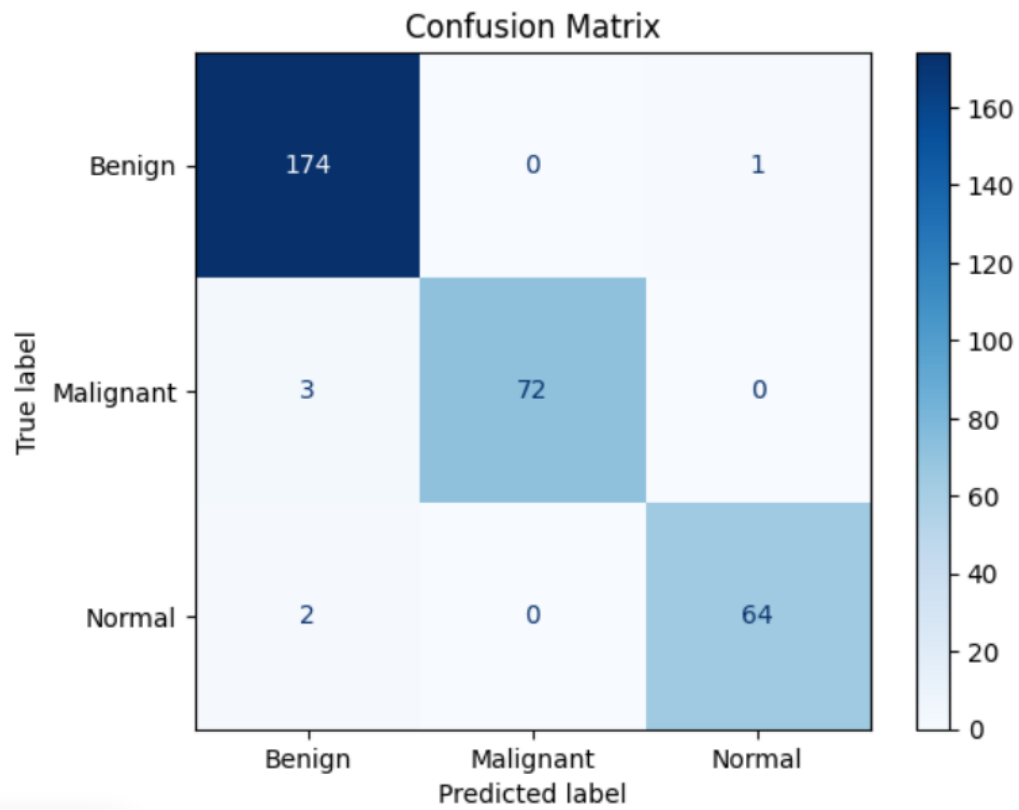


Figure 4.7: Confusion Matrix of the Ultrasound Breast Images Dataset using Xception with CBAM attention

# Conclusions

In this study, we embarked on a comprehensive exploration of convolutional neural network (CNN) models for the classification of medical images, focusing on datasets related to lung cancer, breast cancer, and kidney disease. Through rigorous experimentation, we evaluated the performance of both baseline CNN models and their enhanced counterparts with the integration of the Convolutional Block Attention Module (CBAM). Our results demonstrate the efficacy of attention mechanisms in improving classification accuracy across all three datasets. Moreover, we showcased the importance of hyperparameter tuning in fine-tuning model performance, emphasizing the significance of parameters such as learning rate, optimizer type, and number of epochs in achieving optimal results.

This study comprehensively evaluated the performance of various deep learning models in classifying different medical image datasets, namely lung cancer, breast cancer, and kidney disease, while also investigating the impact of incorporating the Convolutional Block Attention Module (CBAM). The results demonstrated consistent improvements in classification accuracy across all datasets when utilizing CBAM, with VGG16 notably emerging as the top-performing model. These findings underscore the importance of attention mechanisms in enhancing feature extraction and classification accuracy in medical image analysis tasks. Moving forward, this research provides valuable insights into the selection of appropriate deep learning architectures and attention mechanisms for effectively tackling medical image classification challenges, thereby contributing to advancements in clinical diagnosis and treatment.

Overall, this study contributes to the growing body of research in medical image classification, providing insights into the effectiveness of deep learning techniques and highlighting avenues for future research and development.

## 5.1 Limitations

While our study provides valuable insights into the application of CNN models for medical image classification, it is not without its limitations. Here are some key limitations to consider:

- **Limited Dataset Size:** The performance of our models may be influenced by the size and diversity of the datasets used for training and evaluation. Larger and more diverse datasets could provide a better representation of the variability present in clinical data and lead to more robust models.
- **Single-Task Classification:** Our models were trained and evaluated for specific classification tasks related to lung cancer, breast cancer, and kidney disease. Extending the scope to multi-task or multi-class classification could provide a more comprehensive understanding of the models' capabilities and their generalizability across different medical conditions.
- **Evaluation Metrics:** While we employed commonly used evaluation metrics such as Accuracy, Precision, Recall, and F1 Score, there may be other metrics like AUC score, Kappa statistics that are relevant to specific clinical contexts or objectives. Also results using cross-validation schemes would give more insights on the datasets. Exploring additional evaluation metrics and their implications for clinical decision-making could enhance the interpretability of our results.

Addressing these limitations in future research endeavors could further enhance the validity and applicability of CNN models for medical image classification tasks.

## 5.2 Scope for Future Work

Despite the advancements made in this study, there are several avenues for future research and improvement in the field of medical image classification. Here are three key points to consider:

1. **Exploration of Advanced Attention Mechanisms:** While CBAM has shown promise in enhancing classification performance, further exploration of advanced attention mechanisms could lead to even more significant improvements. Techniques such as self-attention mechanisms and transformer-based architectures warrant investigation for their potential to capture long-range dependencies in medical images.

2. **Integration of Multimodal Data:** Incorporating additional modalities, such as textual reports or patient metadata, could enrich the information available to the classification models. Fusion techniques that effectively combine information from multiple modalities have the potential to enhance diagnostic accuracy and enable more comprehensive patient assessments.
3. **Transfer Learning and Domain Adaptation:** Leveraging pre-trained models and domain adaptation techniques could facilitate the development of robust classification models, particularly in scenarios with limited annotated data. Fine-tuning pre-trained models on medical image datasets and exploring techniques for domain adaptation could improve model generalization and performance across diverse clinical settings.

# Bibliography

- [1] M. Tan and Q. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” in *International conference on machine learning*, PMLR, 2019, pp. 6105–6114.
- [2] S. Karen, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv: 1409.1556*, 2014.
- [3] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.
- [4] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, “Learning transferable architectures for scalable image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8697–8710.
- [5] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [6] K. Padmanandam, R. Khilar, N. Pitla, R. K. Dhanaraj, D. Pamucar, and B. Balusamy, “Residual partial derivative multilayer perceptron deep learning based classifier for lung cancer detection,” 2023.
- [7] O. Mishra, D. Parashar, A. Kukker, *et al.*, “Imbalanced class problem analysis for lung cancer detection using convolutional neural networks,” in *Data Analytics for Intelligent Systems: Techniques and solutions*, IOP Publishing Bristol, UK, 2024, pp. 6–1.
- [8] C. Yan and N. Razmjoooy, “Optimal lung cancer detection based on cnn optimized and improved snake optimization algorithm,” *Biomedical Signal Processing and Control*, vol. 86, p. 105319, 2023.
- [9] C. Parvathy and J. Jayan, “Lung cancer prediction in chest ct using an active contour based segmentation and 3dcnn,” *Multimedia Tools and Applications*, pp. 1–25, 2023.

- 
- [10] T. Charnsripinyo and N. Wattanapongsakorn, "Multi-class classification of lung cancer using machine learning and image processing," in *2023 IEEE International Conference on Computing (ICOCO)*, IEEE, 2023, pp. 176–181.
  - [11] S. Hossain, S. Azam, S. Montaha, *et al.*, "Automated breast tumor ultrasound image segmentation with hybrid unet and classification using fine-tuned cnn model," *Heliyon*, vol. 9, no. 11, 2023.
  - [12] F. Uysal and M. M. Köse, "Classification of breast cancer ultrasound images with deep learning-based models," *Engineering Proceedings*, vol. 31, no. 1, p. 8, 2022.
  - [13] S. Balasubramaniam, Y. Velmurugan, D. Jaganathan, and S. Dhanasekaran, "A modified lenet cnn for breast cancer diagnosis in ultrasound images," *Diagnostics*, vol. 13, no. 17, p. 2746, 2023.
  - [14] Y. Pourasad, E. Zarouri, M. Salemizadeh Parizi, and A. Salih Mohammed, "Presentation of novel architecture for diagnosis and identifying breast cancer location based on ultrasound images using machine learning," *Diagnostics*, vol. 11, no. 10, p. 1870, 2021.
  - [15] N. A. Roni, M. S. Hossain, M. B. Hossain, M. I. A. Efati, and M. A. Yousuf, "Deep convolutional comparison architecture for breast cancer binary classification," in *International Conference on Machine Intelligence and Emerging Technologies*, Springer, 2022, pp. 187–200.
  - [16] M. H. K. Mehedi, E. Haque, S. Y. Radin, M. A. U. Rahman, M. T. Reza, and M. G. R. Alam, "Kidney tumor segmentation and classification using deep neural network on ct images," in *2022 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, IEEE, 2022, pp. 1–7.
  - [17] H. Bingol, M. Yildirim, K. Yildirim, and B. Alatas, "Automatic classification of kidney ct images with relief based novel hybrid deep model," *PeerJ Computer Science*, vol. 9, e1717, 2023.
  - [18] M. S. Hossain, S. Hassan, M. Al-Amin, R. Hossain, *et al.*, "Kidney disease detection and classification from ct images using watershed segmentation and deep learning," Ph.D. dissertation, Brac University, 2022.
  - [19] T. A. Ripa, N. Faiyaz, M. Hassan, *et al.*, "A computer assisted detection framework of kidney diseases based on cnn model," in *International Conference on Intelligent Computing & Optimization*, Springer, 2023, pp. 217–230.

- 
- [20] M. Majid, Y. Gulzar, S. Ayoub, *et al.*, “Enhanced transfer learning strategies for effective kidney tumor classification with ct imaging,” *Int J Adv Comput Sci Appl*, vol. 14, p. 2023, 2023.
  - [21] S. Ghosh, S. Banerjee, S. Das, *et al.*, “Evaluation and optimization of biomedical image-based deep convolutional neural network model for covid-19 status classification,” *Applied Sciences*, vol. 12, no. 21, p. 10 787, 2022.
  - [22] *Vgg-16 / cnn model*, <https://www.geeksforgeeks.org/vgg-16-cnn-model/>.
  - [23] K. Srinivasan, L. Garg, D. Datta, *et al.*, “Performance comparison of deep cnn models for detecting driver’s distraction,” *CMC-Computers, Materials & Continua*, vol. 68, no. 3, pp. 4109–4124, 2021.
  - [24] N. Radwan, “Leveraging sparse and dense features for reliable state estimation in urban environments,” Ph.D. dissertation, University of Freiburg, Freiburg im Breisgau, Germany, 2019.
  - [25] M. Reda, R. Suwwan, S. Alkafri, Y. Rashed, and T. Shanableh, “Agroid: A mobile app system for visual classification of plant species and diseases using deep learning and tensorflow lite,” in *Informatics*, MDPI, vol. 9, 2022, p. 55.
  - [26] F. J. P. Montalbo, “Automated diagnosis of diverse coffee leaf images through a stage-wise aggregated triple deep convolutional neural network,” *Machine Vision and Applications*, vol. 33, no. 1, p. 19, 2022.
  - [27] H. Li, W. Hu, Y. Zang, and S. Zhao, “Action recognition based on attention mechanism and depthwise separable residual module,” *Signal, Image and Video Processing*, vol. 17, no. 1, pp. 57–65, 2023.
  - [28] H. F. Kareem, M. S. AL-Husieny, F. Y. Mohsen, E. A. Khalil, and Z. S. Hassan, “Evaluation of svm performance in the detection of lung cancer in marked ct scan dataset,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 21, no. 3, p. 1731, 2021.
  - [29] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, “Dataset of breast ultrasound images,” *Data in brief*, vol. 28, p. 104 863, 2020.
  - [30] M. N. Islam, M. Hasan, M. K. Hossain, M. G. R. Alam, M. Z. Uddin, and A. Soyulu, “Vision transformer and explainable transfer learning models for auto detection of kidney cyst, stone and tumor from ct-radiography,” *Scientific Reports*, vol. 12, no. 1, pp. 1–14, 2022.