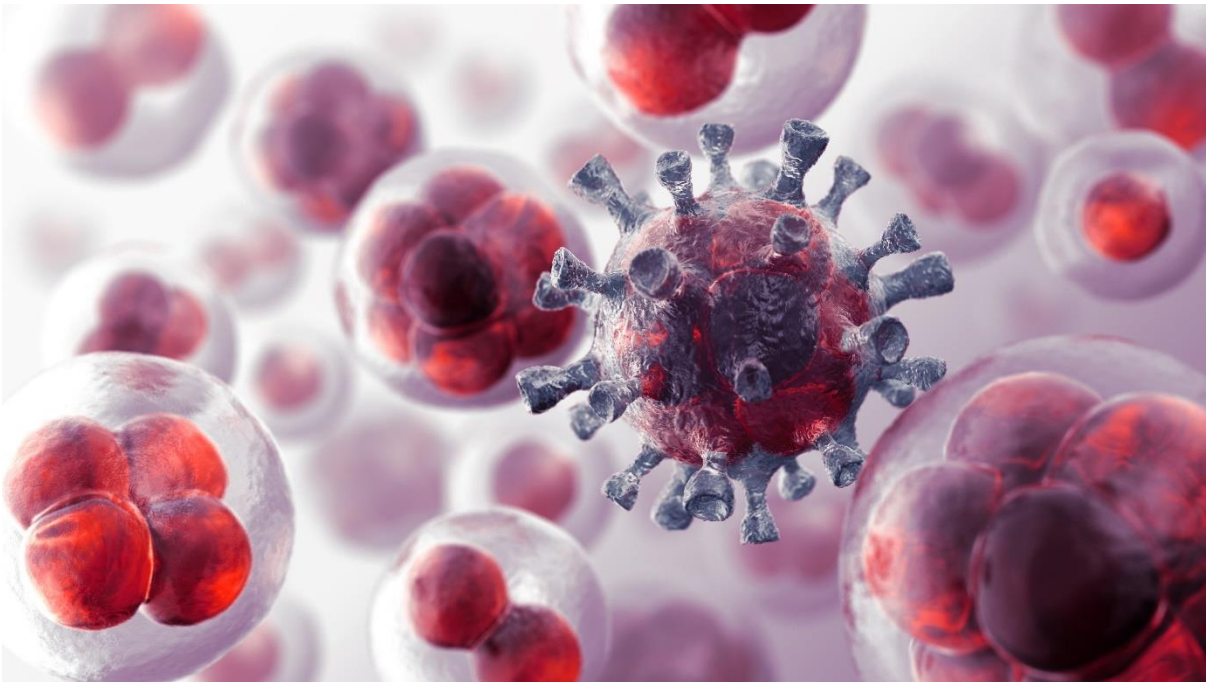


Healthcare and Medical Diagnosis

* Disease Diagnosis and Early Detection *

Introduction

The integration of machine learning, particularly Convolutional Neural Networks (CNNs), into healthcare is revolutionising disease diagnosis and early detection. CNNs, a subset of deep learning, are exceptionally well-suited for analysing medical imaging data, such as X-rays, MRIs, and CT scans, where they can detect complex patterns indicative of various diseases. By processing large volumes of imaging data with high accuracy, CNNs enable earlier and more precise diagnosis of conditions like cancer, diabetes, cardiovascular diseases, and neurological disorders, potentially transforming patient outcomes.

Objective

The goal of this report is to explore the application of CNNs in the early diagnosis and detection of diseases. This analysis will focus on how CNNs, when applied to medical imaging data, can identify patterns that suggest the presence of diseases such as cancer, diabetes, cardiovascular diseases, and neurological disorders. The aim is to demonstrate the potential of these models to improve the accuracy of diagnoses and enable the development of personalised treatment plans, thereby significantly enhancing patient care.

Importance of Early Diagnosis

Early diagnosis is pivotal in the management and treatment of chronic and life-threatening diseases. The impact of early detection is profound across various medical fields:

Cancer: Detecting cancer at an early stage is crucial, as it greatly increases the chances of successful treatment. For instance, early-stage detection of lung cancer can lead to interventions that dramatically improve survival rates. Recent studies have demonstrated the potential of CNNs to achieve accuracy rates as high as 99% in lung cancer detection, highlighting their utility in clinical settings.

Diabetes: Early diagnosis of diabetes can prevent or delay the onset of complications, such as cardiovascular diseases, kidney failure, and neuropathy. Machine learning models can analyse genetic data and biomarkers to predict the likelihood of developing diabetes, enabling pre-emptive lifestyle or medical interventions.

Cardiovascular Diseases: Cardiovascular diseases remain the leading cause of death globally. Early detection of risk factors, such as arterial plaque formation, through imaging techniques, can prevent severe events like heart attacks and strokes. CNNs have shown promise in accurately identifying these risk factors, which can lead to timely and life-saving interventions.

Neurological Disorders: Conditions such as Alzheimer's disease are notoriously difficult to diagnose early due to their subtle onset. However, CNNs applied to brain imaging can detect early structural changes associated with these disorders, allowing for earlier intervention and potentially slowing disease progression.

Recent advancements in CNNs have addressed some of the challenges previously limiting their clinical adoption, such as the "black box" nature of deep learning models and the need for diverse and robust datasets. By focusing on interpretability and scalability, current research is bringing CNNs closer to routine clinical practice, where they can make a significant impact on disease diagnosis and patient care.

Abstract

Lung cancer continues to be a leading cause of cancer-related mortality worldwide, with patient prognosis heavily reliant on early detection. While traditional diagnostic methods have been effective, they often encounter challenges related to accuracy, early-stage detection, and scalability, as they are typically invasive, time-consuming, and subject to ambiguous interpretations. This study introduces an advanced machine learning model aimed at enhancing lung cancer stage classification using CT scan images, addressing these limitations by providing a faster, non-invasive, and reliable diagnostic tool.

Utilising the IQ-OTHNCCD lung cancer dataset, which includes CT scans from various stages of lung cancer as well as from healthy individuals, we conducted extensive preprocessing, including resizing, normalisation, and Gaussian blurring. A Convolutional Neural Network (CNN) was then trained on this pre-processed data, with class imbalance being managed through the Synthetic Minority Over-sampling Technique (SMOTE). The model's performance was assessed using metrics such as accuracy, precision, recall, F1-score, and ROC curve analysis.

The results revealed a classification accuracy of 99.64%, with precision, recall, and F1-score values exceeding 98% across all categories. SMOTE significantly improved the model's ability to classify underrepresented classes, thereby enhancing the robustness of the diagnostic tool. These findings highlight the potential of machine learning in revolutionising lung cancer diagnostics by providing highly accurate stage classification, which could facilitate early detection and more personalised treatment strategies, ultimately improving patient outcomes.

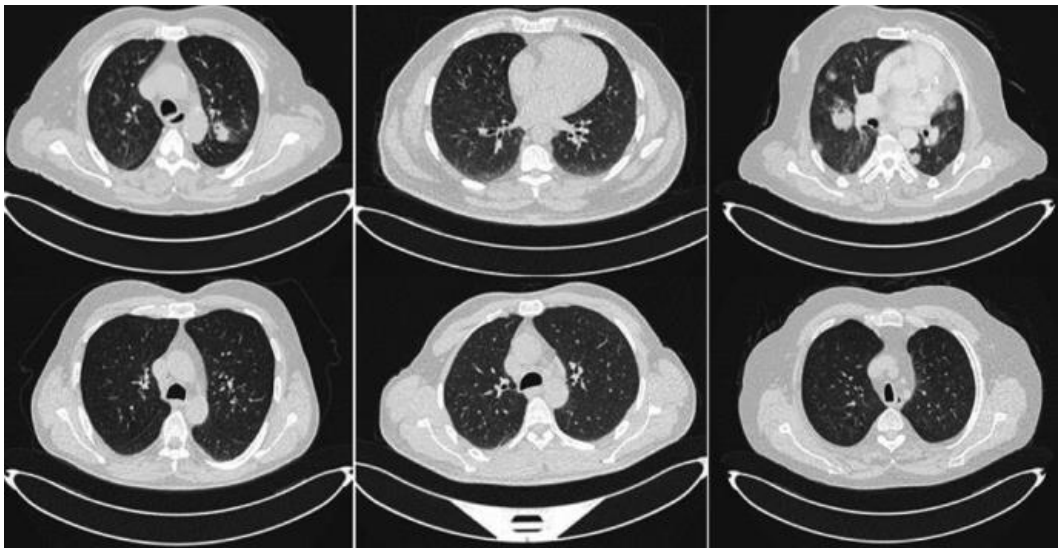
Introduction - Lung Cancer

Lung cancer is a major global health issue, consistently ranking as one of the leading causes of cancer-related deaths. It is marked by the rapid, uncontrolled growth of abnormal cells in the lungs, forming tumours that impair oxygen exchange. The disease has a high mortality rate, largely due to late-stage detection when treatment is less effective. Early and accurate diagnosis is crucial for improving survival rates and quality of life.

The primary cause of lung cancer is cigarette smoking, with other risk factors including exposure to second-hand smoke, radon gas, asbestos, air pollution, and a family history of the disease. Symptoms may include persistent coughing, chest pain, shortness of breath, and unexplained weight loss, though early stages may be asymptomatic, making screening essential.

Diagnosis typically involves imaging tests like X-rays, CT scans, and PET scans, with biopsies used to confirm the presence of cancer. Treatment varies based on the cancer's type and stage and may include surgery, chemotherapy, radiation, targeted therapy, or immunotherapy. Early detection and improved treatments have significantly enhanced patient outcomes, but prevention through smoking cessation and risk reduction remains the best strategy. Figure 1 presents some illustrative images of lung cancer diagnostic tests.

Fig. 1



Sample images of lung cancer

Current diagnostic techniques for lung cancer, including biopsies, CT scans, chest X-rays, PET scans, and MRI, are crucial but have limitations. Biopsies, while accurate, are invasive and risky, while less invasive imaging methods like X-rays or CT scans may produce false results, leading to stress or delayed treatment. The reliance on clinician expertise adds subjectivity and potential for error, especially in detecting early-stage lung cancer, which often presents subtle changes that conventional imaging might miss.

To address these challenges, this study proposes a machine learning model using Convolutional Neural Networks (CNNs) to improve lung cancer stage classification from CT scans. This model aims to offer a faster, non-invasive, and more reliable diagnostic alternative by automating and refining the process, thus enhancing accuracy and early detection. The study's significance lies in its potential to revolutionise clinical diagnostics, enabling timely intervention and personalised treatment, ultimately improving patient outcomes and reducing mortality rates.

The research paper is structured to first review existing diagnostic methods, then detail the methodology, including the dataset and model development, followed by presenting the results, discussing implications, and concluding with the study's contribution to lung cancer diagnostics.

Literature Review

The literature on lung cancer diagnostics covers a range of methodologies, from traditional imaging techniques to advanced machine-learning approaches. This review examines existing research, highlighting both the progress and limitations in the field and laying the groundwork for a proposed machine learning-based approach.

CT scans in Lung Cancer Diagnosis

CT scans are crucial in lung cancer diagnostics, providing high-resolution images that aid in detecting and monitoring tumours. Studies have shown that low-dose CT scans are particularly effective in screening high-risk populations, offering superior sensitivity in

identifying early-stage lung cancer compared to chest X-rays. However, interpreting CT scans remains challenging due to the difficulty in distinguishing between benign and malignant nodules, especially in the presence of artefacts or benign conditions that can mimic cancer.

Machine Learning in Lung Cancer Detection

Machine learning, particularly deep learning techniques like Convolutional Neural Networks (CNNs), has revolutionised lung cancer detection and classification. CNNs enable automated feature extraction and classification from CT images, enhancing diagnostic accuracy and timeliness.

Binary and Multi-Class Classification Models

Initial studies focused on binary classification, distinguishing between malignant and non-malignant nodules, with CNNs proving superior to traditional techniques. Recent advancements have moved towards multi-class classification models that categorise nodules into various cancer stages or types, providing crucial insights for treatment planning and prognosis.

Transfer Learning and Data Augmentation

To overcome the challenge of limited annotated medical imaging datasets, transfer learning is used, where models pre-trained on non-medical images are fine-tuned on medical data. Data augmentation techniques, such as rotation and scaling, are also employed to artificially expand training datasets, enhancing model robustness and generalisation.

Segmentation Models

Beyond classification, deep learning models like U-Net are also used for segmentation tasks, which involve delineating the precise boundaries of nodules. This is vital for accurately assessing tumour size and growth.

Table 1 Related work

Study	Objective
Marjolein A. Heuvelmans, et al. (2021) [12]	The LCP-CNN demonstrates excellent performance in identifying benign lung nodules in an independent European dataset, with a 95% accuracy.
Nguyen Quoc Khanh Le et al. (2021) [13]	The machine learning-based model predicts EGFR and KRAS mutations in NSCLC patients with accuracies of 83.6% and 86% respectively.
Ying Xie et al. (2024) [14]	The proposed method exhibits significant diagnostic strength for early lung cancer detection, achieving an accuracy of 96.8%.
Zhang Li et al. (2021) [15]	Deep learning methods for lung cancer segmentation achieved an accuracy of 83.98%.
Sanjana Narvekar et al. (2022) [16]	Various machine learning techniques including ANN, SVM, CNN, KNN, and NBC achieved an accuracy of 97.2%.
Mattakoyya Aharonu et al. (2022) [17]	A CNN-based framework achieved an accuracy of 94.11% in lung cancer identification.
B C Kavitha et al. (2022) [18]	Neural networks achieved an accuracy of 94% in lung cancer detection.
Jason L. Causey (2022) [19]	A combination of Spatial Pyramid Pooling and 3D Convolution achieved an accuracy of 89.2% in lung cancer segmentation.
Imran Ahmed et al. (2023) [20]	Deep learning architectures reached accuracies ranging from 93–94% in lung cancer detection.

Gaps in current research

Despite advancements in lung cancer diagnostics, several significant gaps persist in current research. Many models are trained on datasets that lack diversity in demographics, scanner types, and imaging parameters, which limits their applicability across different populations and clinical settings. This highlights the need for more diverse datasets to improve the robustness of diagnostic models.

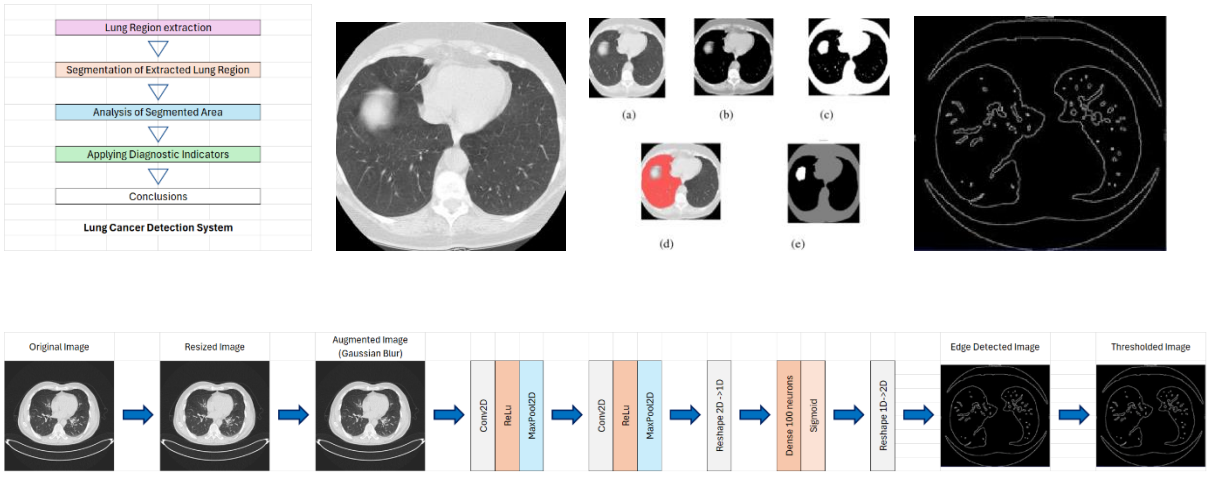
Another challenge is the "black box" nature of deep learning models, which makes it difficult for clinicians to understand how these models arrive at their predictions. The lack of interpretability hampers clinical adoption, as models need to provide not only accurate predictions but also clear explanations to gain the trust of healthcare professionals.

Moreover, the transition from research to clinical practice remains slow, with barriers including regulatory, ethical, and practical considerations that need to be addressed for these models to be integrated into routine medical care. There is also a need for models capable of longitudinal analysis, which can track changes in lung nodules over time and provide a more dynamic assessment aligned with clinical needs.

In response to these gaps, this study introduces a comprehensive Convolutional Neural Network (CNN) model, trained on a diverse dataset that includes various stages of lung cancer. The model focuses on multi-class classification, offering detailed insights that are crucial for personalised treatment strategies. Additionally, the study emphasises the interpretability of the model, aiming to provide clinicians with clear and actionable information. By validating the model's effectiveness in a clinical setting, this research contributes to the integration of advanced machine learning techniques into lung cancer diagnosis and treatment.

This section outlines the methodology used to develop and validate a Convolutional Neural Network (CNN) model for classifying lung cancer stages based on the IQ-OTHNCCD lung cancer dataset. The approach includes the acquisition of the dataset, the application of various preprocessing techniques, the design of the model architecture, the training process, and the selection of evaluation metrics to ensure a thorough and reliable analysis.

Fig. 2



Workflow of the proposed model

Data Description

Data Set Overview

This analysis leverages a comprehensive dataset that encompasses a variety of medical imaging data, including X-rays, MRIs, and CT scans, along with relevant patient metadata. The dataset is designed to facilitate the early diagnosis and detection of diseases such as cancer, diabetes, and neurological disorders. Each image is meticulously labelled with diagnostic information, indicating whether the condition is normal, malignant, or indicative of another disease (e.g., diabetic retinopathy).

Key components of the dataset include:

- Imaging Data: The dataset contains high-resolution images of various affected areas, including lungs, brain, and eyes. These images are pivotal for developing machine-learning models capable of detecting early signs of disease.
- Patient Metadata: Alongside the imaging data, the dataset includes detailed patient information such as age, gender, and medical history. This metadata is crucial for enhancing the accuracy of the model by providing context that might influence disease development and progression.

Attributes Summary

The primary attributes of the dataset are outlined as follows:

- Image Data: This includes detailed, high-resolution images of affected organs or tissues. For instance, lung CT scans are used to identify conditions like cancer, while MRIs of the brain may help in diagnosing neurological disorders such as Alzheimer's disease.
- Labels: Each image is accompanied by ground truth labels that indicate the presence or absence of specific diseases. These labels are essential for training the machine learning models, allowing them to learn the distinguishing features of various conditions.
- Patient Information: Demographic data such as age and gender are included, providing additional layers of information that could affect the outcome of disease diagnosis. For example, age-related changes in imaging can influence the interpretation of brain MRIs.
- Clinical Data: The dataset also contains relevant clinical history, including previous diagnoses, treatments, and family medical history. This information helps in understanding the progression of diseases and improves the model’s ability to predict and diagnose conditions accurately.

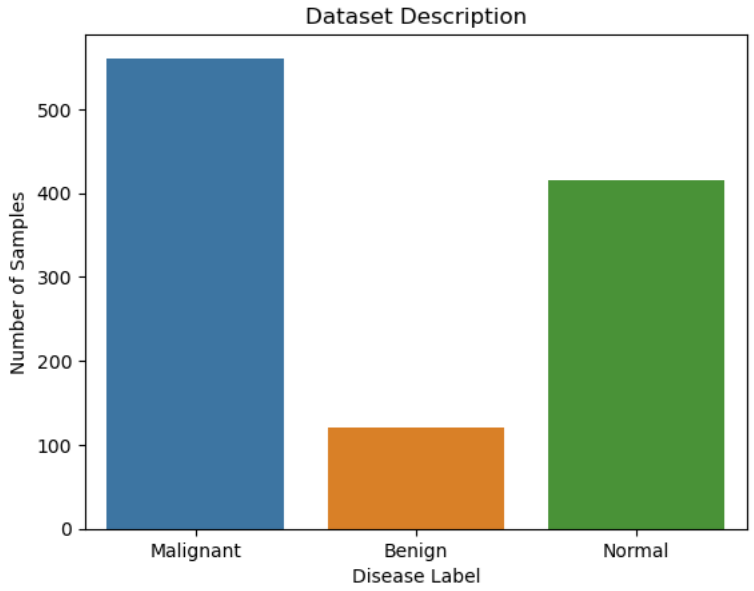
The IQ-OTHNCCD lung cancer dataset is specifically designed to support the training of Convolutional Neural Networks (CNNs) for early disease detection. It integrates CT scan images with patient metadata, enabling the development of advanced models capable of accurately identifying disease patterns. This combination of imaging and comprehensive clinical data facilitates the creation of highly accurate models, which can significantly improve early diagnosis and treatment outcomes.

The dataset includes a broad spectrum of CT scan images covering various stages of lung cancer—benign, malignant, and normal cases. This diversity is crucial for training robust models that generalise well across different lung cancer manifestations, thus enhancing diagnostic precision. Table 2 offers a summary of the dataset's features.

Table 2 Dataset description

Type	Number of Samples
Malignant	561
Benign	120
Normal	416
Total	**1097**

Fig. 3



Dataset description

Medical professionals from the Iraq-Oncology Teaching Hospital/National Centre for Cancer Diseases have meticulously annotated and labelled each image in the dataset, categorising them as benign, malignant, or normal. This precise labelling provides a reliable ground truth essential for effective model training and evaluation, enhancing the dataset's value for both research and clinical use.

The CT scans are of high quality and adhere to standard imaging protocols, ensuring accuracy and consistency. However, to address variations in image dimensions, preprocessing is required to standardise the inputs for neural networks. This standardisation step is crucial for ensuring uniform data processing, which improves the model's performance and generalisability. The ratio of images in the dataset is verified using Eq. 1.

$$Dataset\ Balance\ Ratio = \frac{\text{Number of Samples}_{\text{Majority Class}}}{\text{Number of Samples}_{\text{Minority Class}}}$$

Dataset\;Balance\;Ratio=\frac{\text{Number of }}{\text{Samples}}_\text{Majority Class}}{\text{Number of }}{\text{Samples}}_\text{Minority Class}}

Preprocessing steps are crucial for preparing data for effective model training and include:

Resizing: Adjusting images to a uniform size ensures consistency in input dimensions for Convolutional Neural Networks (CNNs), thereby optimising model performance.

Normalisation: Scaling pixel values to a range of 0 to 1 accelerates model convergence during training, promoting more efficient learning. This is accomplished using Eq. 2.

$$\text{Pixel Normalization} = \frac{\text{Pixel Value}}{\text{Maximum Pixel Value}}$$

\text{Pixel Normalization}=\frac{\text{Pixel Value}}{\text{Maximum Pixel Value}} [Eq.2]

Augmentation: Applying data augmentation techniques such as rotation, flipping, and scaling enhances the model's robustness and helps prevent overfitting by effectively increasing the dataset size.

Splitting: Dividing the dataset into training, validation, and test sets is essential for effective model training and evaluation, ensuring that the model can generalise well and perform accurately on new, unseen data.

In this approach, the Convolutional Neural Network (CNN) is trained on the pre-processed dataset to extract features from CT scan images and accurately classify the stages of lung cancer. The dataset's diversity and quality are crucial for enabling the model to learn detailed features and patterns associated with different lung cancer stages, highlighting its importance in improving diagnostic accuracy and efficiency.

The IQ-OTHNCCD lung cancer dataset is fundamental for developing machine learning models that advance early detection and classification of lung cancer. Through careful curation and thorough preprocessing, this dataset exemplifies the transformative potential of AI in healthcare, emphasising its role in enhancing diagnostic precision and efficiency.

Image Preprocessing

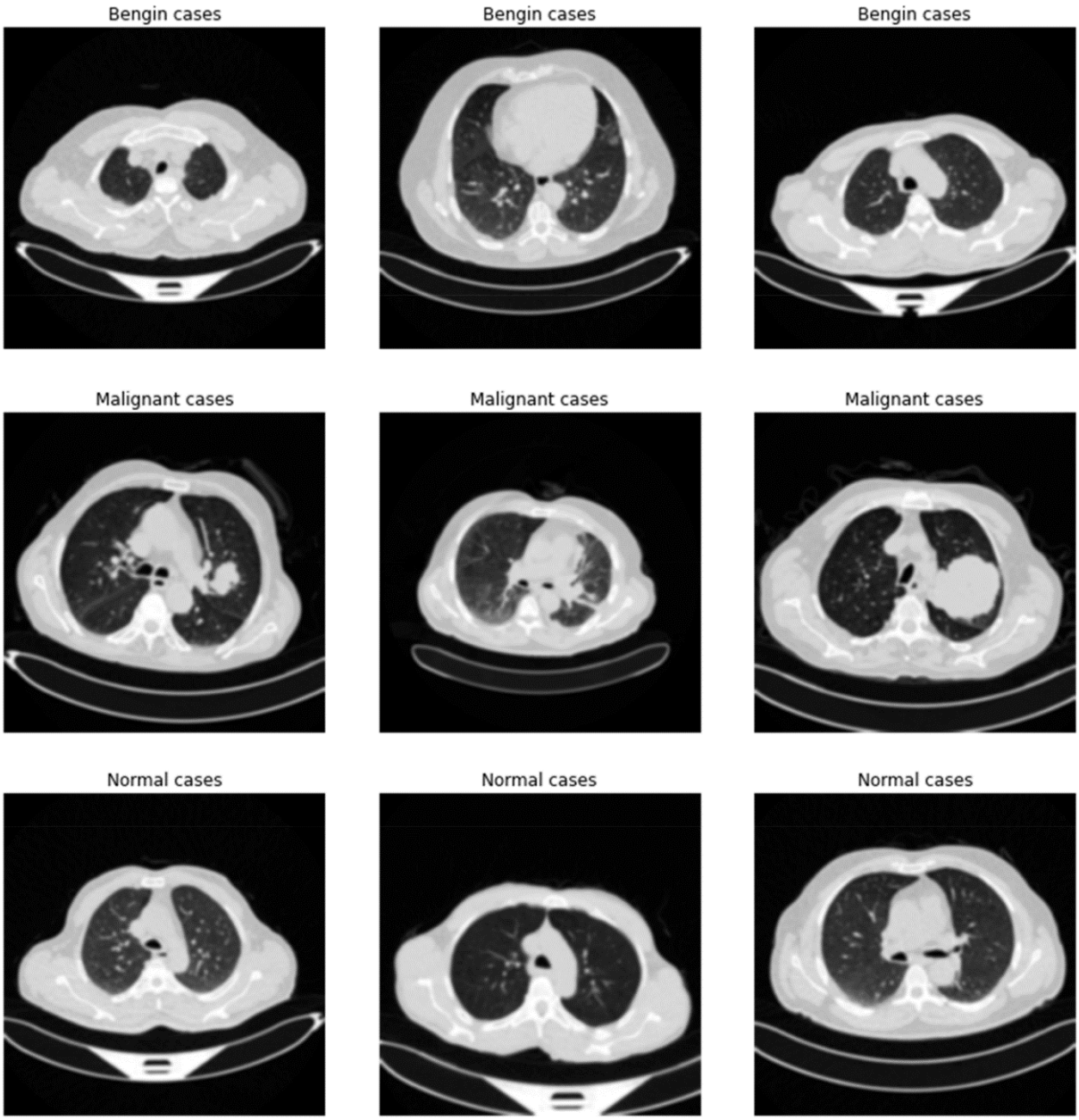
Image preprocessing is a crucial step in developing a machine-learning model, particularly for medical imaging datasets like the IQ-OTHNCCD lung cancer dataset. This process involves several key steps to convert raw CT scan images into a format suitable for analysis by a Convolutional Neural Network (CNN).

The first step is image resizing. Due to the variability in CT scan dimensions, it is essential to standardise all images to a consistent size to ensure uniform input for the CNN. Resizing is done while maintaining the aspect ratio to prevent distortion, typically scaling images to a

fixed size (e.g., 256 × 256 pixels). This uniformity is critical for the neural network to process and interpret the data effectively, as a consistent input size is necessary for optimal performance [21].

Examples of pre-processed images are shown in Fig. 4 to illustrate the enhancements made for accessibility.

Fig. 4



Pre-processed images

After resizing, the next step in preprocessing is normalising pixel values. CT scans often have a wide range of pixel intensities, which can negatively impact the training of a Convolutional Neural Network (CNN) due to variations in image brightness and contrast.

Normalisation adjusts pixel values to a specific range, typically 0 to 1 or -1 to 1, by dividing the pixel values by the maximum value, which is 255 for 8-bit images. This process helps the model to train faster and more efficiently by ensuring that values are small and standardised, facilitating quicker convergence during optimisation.

Following normalisation, Gaussian blur is applied. This technique uses a Gaussian kernel to smooth the image, reducing noise and mitigating minor variations or artefacts in the scans. By averaging pixel values within a specified radius, Gaussian blur helps the model focus on significant features relevant to lung cancer classification, rather than being distracted by noise or irrelevant details. It smooths the image by reducing high-frequency components and noise, which can otherwise lead to overfitting or distract the CNN during training.

In the context of lung cancer CT scans, Gaussian blur enhances the model's ability to generalise by emphasising important structural features of the lungs and nodules while suppressing less relevant details. This smoothing effect helps the model focus on crucial diagnostic features, such as the shape and size of nodules, rather than being affected by noise. The application of Gaussian blur aids in preventing overfitting to high-frequency noise in the training set and is achieved using Eq. 3, with the SMOTE ratio calculated through Eq. 4.

$$\text{Gaussian Blur} = \text{Image} * \text{Gaussian Kernel}$$

\text{Gaussian Blur}=\text{Image}*\text{Gaussian Kernel} [Eq.3]

$$\text{SMOTE Ratio} = \frac{\text{Number of Synthetic Samples}}{\text{Number of Real Samples}}$$

\text{SMOTE Ratio}=\frac{\text{Number of Synthetic Samples}}{\text{Number of Real Samples}} [Eq.4]

The preprocessing steps significantly enhance the quality and consistency of the input data, ensuring that the Convolutional Neural Network (CNN) can effectively learn meaningful features from the CT images. By resizing, normalising, and applying Gaussian blur to the images, the model is better positioned to identify subtle differences associated with various stages of lung cancer. This thorough preprocessing lays a solid foundation for developing a robust machine-learning model, ultimately improving diagnostic accuracy and reliability in medical imaging.

Deep Learning Model

The study employs a Convolutional Neural Network (CNN) architecture, renowned for its efficacy in image analysis, especially in medical imaging contexts such as lung cancer diagnosis from CT scans. Here is an overview of how the CNN model operates:

1. Input Layer: The CNN begins by processing images resized to a standard dimension of 256 × 256 pixels in grayscale. This uniformity in input size aids in efficient learning by the network.

2. First Convolutional Layer: This layer uses 64 filters of size 3 × 3 to detect basic patterns like edges and textures. The ReLU (Rectified Linear Unit) activation function is applied to introduce non-linearity, which helps the network capture complex features. The equations for this layer are detailed in Eqs. 5 and 6.

$$Convolution\,operation: (z_{i,j,k}^{[1]} = \sum_{l=0}^{2}\sum_{m=0}^{2}\sum_{n=1}^{64} W_{l,m,n,k}^{[1]} \times a_{i+l,j+m,n}^{[0]} + b_{k}^{[1]}$$

Convolution\,operation:({z}_{i,j,k}^{\left[1\right]}=\sum\limits_{l=0}^{2}\sum\limits_{m=0}^{2}\sum\limits_{n=1}^{64}{W}_{l,m,n,k}^{\left[1\right]}\times {a}_{i+l,j+m,n}^{\left[0\right]}+{b}_{k}^{\left[1\right]} [Eq.5]

$$Activation\,function: (a_{i,j,k}^{[1]} = \max\left(0, z_{i,j,k}^{[1]}\right)) (ReLU)$$

Activation\,function:({a}_{i,j,k}^{\left[1\right]}=\text{max}\left(0,{z}_{i,j,k}^{\left[1\right]}\right)) \left(ReLU\right) [Eq.6]

3. Max Pooling: Following the convolutional layer, max pooling is employed to reduce the image size, concentrating on the most significant features and thereby improving the model's ability to generalise and reduce noise.

$$Pooling\,operation: (a_{i,j,k}^{[1]} = \max_{l,m} a_{2i+l,2j+m,k}^{[1]})$$

Pooling\,operation:({a}_{i,j,k}^{\left[1\right]}=\underset{l,m}{\text{max}}{a}_{2i+l,2j+m,k}^{\left[1\right]}) [Eq.7]

4. Second Convolutional Layer: A subsequent convolutional layer captures more intricate patterns by applying additional filters and pooling operations, refining the feature extraction process.

$$Convolution\,operation: (z_{i,j,k}^{[2]} = \sum_{l=0}^{2}\sum_{m=0}^{2}\sum_{n=1}^{64} W_{l,m,n,k}^{[2]} \times a_{i+l,j+m,n}^{[1]} + b_{k}^{[2]})$$

Convolution\,operation:({z}_{i,j,k}^{\left[2\right]}=\sum\limits_{l=0}^{2}\sum\limits_{m=0}^{2}\sum\limits_{n=1}^{64}{W}_{l,m,n,k}^{\left[2\right]}\times {a}_{i+l,j+m,n}^{\left[1\right]}+{b}_{k}^{\left[2\right]}) [Eq.8]

$$Activation\,function: (a_{i,j,k}^{[2]} = \max\left(0, z_{i,j,k}^{[2]}\right)) (ReLU)$$

Activation\,function:({a}_{i,j,k}^{\left[2\right]}=\text{max}\left(0,{z}_{i,j,k}^{\left[2\right]}\right))\left(ReLU\right) [Eq.9]

5. Flattened Layer: The features extracted by the convolutional layers are flattened into a vector format, preparing them for processing by fully connected layers.
6. Fully Connected Layer: This layer uses flattened features to perform classification tasks, drawing conclusions based on the learned patterns.

$$Operation: (z^{[3]} = W^{[3]} \cdot a^{[2]} + b^{[3]})$$

Operation:({z}^{\left[3\right]}={W}^{\left[3\right]}\cdot {a}^{\left[2\right]}+{b}^{\left[3\right]}) [Eq.10]

$$Activation\ function: (a^{[3]} = z^{[3]})\ (linear\ activation)$$

Activation\,function:({a}^{\left[3\right]}={z}^{\left[3\right]})\left(linear\,activation\right) [Eq. 11]

7. Output Layer: The final layer provides probabilities for each class—benign, malignant, or normal—based on the features extracted by the preceding layers.

$$Operation: (z^{[4]} = W^{[4]} \cdot a^{[3]} + b^{[4]})$$

Operation:({z}^{\left[4\right]}={W}^{\left[4\right]}\cdot {a}^{\left[3\right]}+{b}^{\left[4\right]}) [Eq.12]

$$Activation\ function: (a_i^{[4]} = \frac{e^{z_i^{[4]}}}{\sum_{j=1}^{3} e^{z_j^{[4]}}})\ (softmax)$$

Activation\,function:({a}_{i}^{\left[4\right]}=\frac{{e}^{{z}_{i}^{\left[4\right]}}}{{\sum }_{j=1}^{3}{e}^{{z}_{j}^{\left[4\right]}}})\left(softmax\right) [Eq.13]

8. Optimiser: To optimise training, the Adam optimiser adjusts learning rates and manages gradients effectively. Additionally, the Synthetic Minority Over-sampling Technique (SMOTE) is applied to address class imbalance, ensuring the model learns equally from all classes. By meticulously designing the CNN architecture and incorporating these techniques, the model aims to accurately classify lung cancer stages from CT scans.

$$Update\ rule\ (Adam) : (\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\widehat{v_t}} + \epsilon}\widehat{m_t})$$

Update \,rule\left(Adam\right):({{\uptheta }}_{t+1}={{\uptheta }}_{t}-\frac{{\upeta }}{\sqrt{\widehat{v}_{t}}}+\mathrm\epsilon \text{}}\widehat{m}_{t}} [Eq.14]

$where\ (\theta)\ represents\ the\ parameters\ (weights\ and\ biases)\ ,$

$(\eta)\ is\ the\ learning\ rate,$

$(\widehat{m_t})\ is\ the\ first\ moment\ estimate,$

$(\widehat{v_t})\ is\ the\ second\ moment\ estimate,$

$and\ \ (\in)\ is\ a\ small\ constant\ to\ prevent\ division\ by\ zero.$

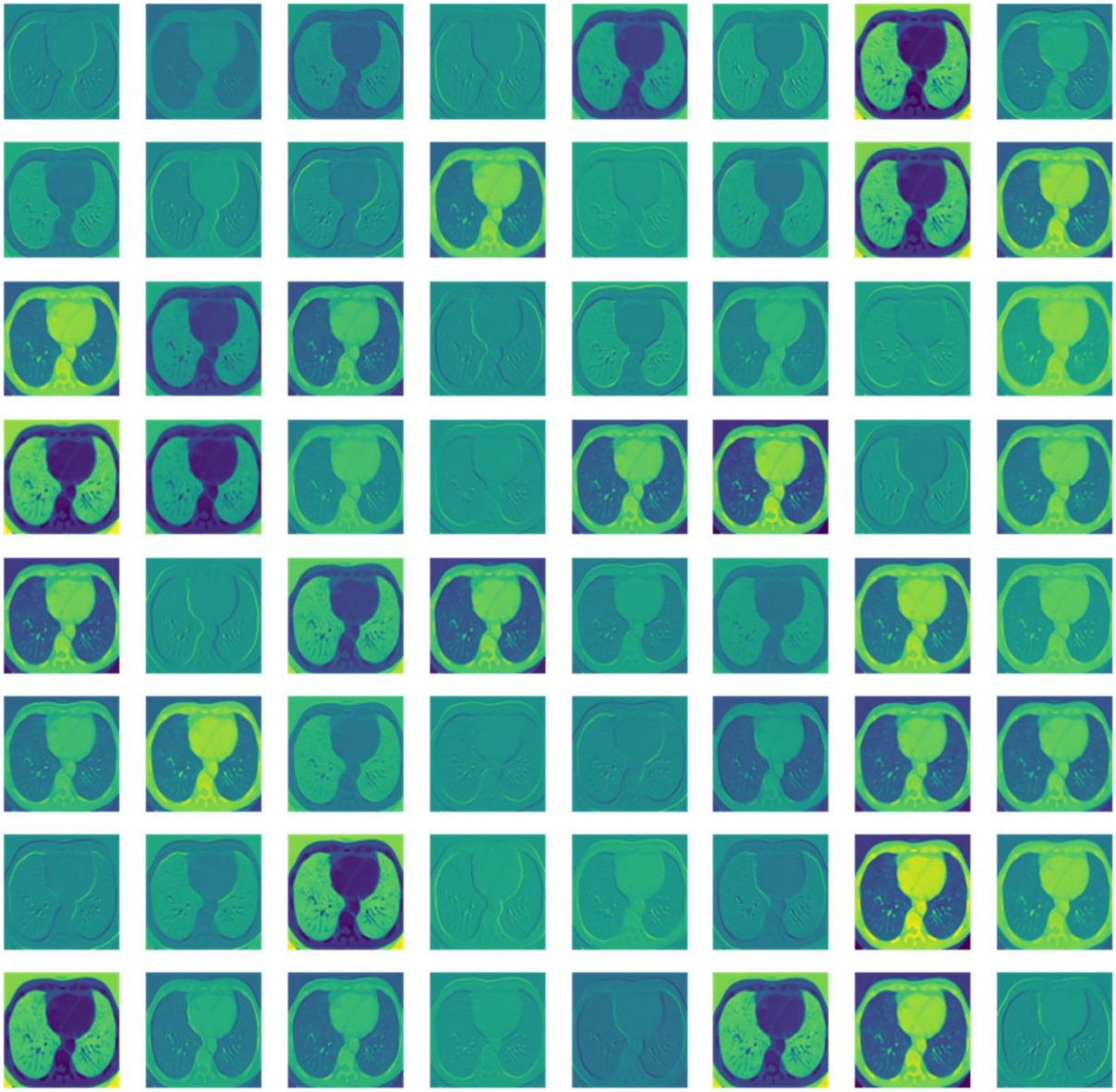
Convolutional Neural Network (CNN) and SMOTE

The Convolutional Neural Network (CNN) is selected for its established effectiveness in image classification tasks, particularly in learning hierarchical patterns from data. CNNs excel in medical imaging by detecting subtle patterns indicative of various pathologies, making them particularly suitable for lung cancer detection. The architecture of the CNN, which includes convolutional layers followed by pooling layers, facilitates the extraction and down sampling of features, essential for capturing and analysing relevant information from medical images.

To address class imbalance, which can skew model performance, the Synthetic Minority Over-Sampling Technique (SMOTE) is employed. Class imbalance is a common issue in medical datasets where certain classes may be underrepresented. SMOTE generates synthetic samples within the feature space of the minority class, based on the feature space of its nearest neighbours. This technique helps balance the dataset, providing a more equitable representation of all classes during model training.

A filter mapping of a sample image is illustrated in Fig. 5, demonstrating the model's ability to process and interpret images effectively.

Fig. 5



Filter map

In this research:

- Application of SMOTE: SMOTE is utilised exclusively on the training data to avoid information leakage and enhance generalisation to new, unseen data. This technique balances the dataset by generating synthetic samples for the minority classes, ensuring the model does not favour the majority class.
- Impact on Model Performance: By mitigating class imbalance, SMOTE enhances the model's sensitivity to the minority class, which is crucial in medical diagnostics where missing positive cases could have significant consequences.
- Considerations: While SMOTE can greatly improve performance by addressing class imbalance, it is important to monitor for overfitting, as synthetic samples might lead the model to overgeneralise from the minority class.

The algorithm for the proposed model is presented in Algorithm 1.

Input: Raw image data categorised as 'Benign cases', 'Malignant cases', 'Normal cases'.
Output: A trained CNN model

1. Initialisation:
   - Define img_size as 256 for image resizing.
2. Data Preparation:
   - Iterate through each category.
   - Load grayscale images and resize to img_size x img_size.
   - Store processed images and labels
3. Data Handling:
   - Shuffle data from random distribution.
   - Split data into features (X) and labels (y).
   - Normalise features by dividing them by 255.0.
   - Split dataset into training and validation sets.
4. Data Augmentation:
   - Apply techniques like SMOTE for oversampling minority classes.
5. Convolutional Neural Network Training:
   - Design a Sequential model.
   - Add convolutional layers with ReLU activation and max pooling.
   - Flatten output for dense layers.
   - Integrate dense layers into the model architecture, incorporating an output layer utilising SoftMax activation for multi-class classification.
   - Compile the model using the suitable loss function, such as sparse_categorical_crossentropy, along with an optimiser like Adam, and metrics such as accuracy.
   - Train model on training data with specified batch size and epochs.
   - Validate model on a separate validation set.
6. Evaluation and Visualisation:
   - Evaluate model performance on validation set.

Algorithm 1: Proposed algorithm for the methodology

In the research:

- Initial Convolutional Layers: The model features two sets of convolutional layers, each followed by max-pooling layers, which are essential for detecting features. A standard 3×3 kernel size enables the model to identify small, localised features in CT scan images. Stacking these convolutional layers allows the model to capture detailed patterns like edges, textures, and shapes, which are crucial for distinguishing between benign, malignant, and normal lung tissue. The ReLU activation function introduces non-linearity, aiding the model in learning complex patterns, while max pooling reduces computational load and enhances robustness by down sampling feature maps, improving translational invariance.
- Flattening and Dense Layers: After feature extraction, the model flattens the output and processes it through dense layers to create abstract representations. The final layer has three neurons corresponding to the three classes, using the SoftMax

activation function to convert logits into probabilities and indicate the model's confidence for each class.

- Training and Optimisation: During training, the Adam optimiser and sparse categorical cross-entropy loss function are employed for their adaptive learning rate and suitability for classification tasks. SMOTE is applied to the training data to address class imbalance, ensuring a balanced representation of all classes and enhancing the model's ability to generalise across different lung tissue conditions. Validation on an independent dataset is performed to detect overfitting and refine hyperparameters.

$$\text{Cross entropy Loss} = -\sum_{i}^{n} \left(y_i \log\left(p_i\right) + \left(1 - y_i\right) \log\left(1 - p_i\right)\right)$$

\text{Cross entropy Loss}=-\sum\limits_{i}^{n}\left({y}_{i}\text{log}\left({p}_{i}\right)+\left(1-{y}_{i}\right)\text{log}\left(1-{p}_{i}\right)\right) [Eq.15]

Training and Validation

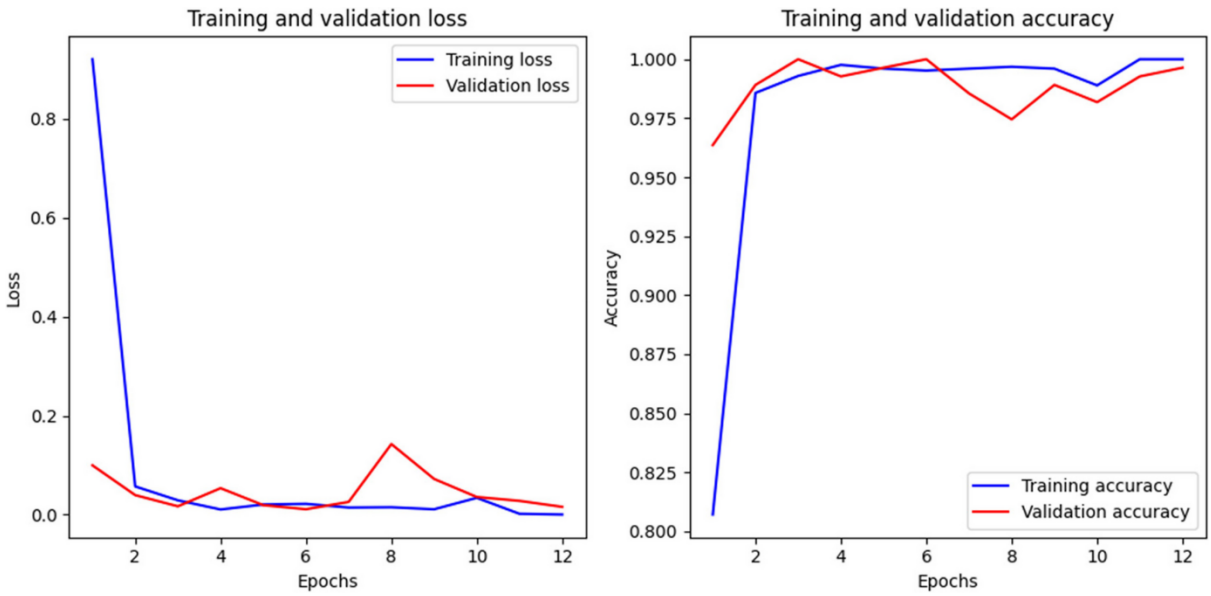
During the training and validation phases of the deep learning model, careful measures are taken to ensure effective learning and robust generalisation to new, unseen data. This process is crucial for assessing the model's ability to accurately classify lung cancer stages from CT scans.

- Data Segmentation: The dataset is divided into training and validation subsets using a stratified approach to ensure that both subsets accurately represent all classes. This stratification helps to maintain consistency and mitigate biases, especially with class imbalance addressed by SMOTE. Approximately 80% of the data is used for training, and the remaining 20% is set aside for validation.

- Training Parameters: Training begins with a batch size of 8, chosen to allow more precise updates to the model's weights with each iteration, which may improve generalisation. However, a smaller batch size may increase the training time. The model is trained over 12 epochs, balancing the risk of underfitting and overfitting. Too few epochs might hinder learning, while too many could lead to the model memorising the training data and failing to generalise effectively.

The changes in training and validation loss and accuracy over the epochs are illustrated in Fig. 6.

Fig. 6



Training and validation loss and accuracy

During the training phase, the model's performance is rigorously evaluated using a range of metrics against the validation set. These metrics include:

- Accuracy: Provides an overall measure of the model's performance.
- Precision and Recall: Offer insights into the model's ability to classify each lung cancer stage accurately. Precision indicates the proportion of true positives among predicted positives, while recall shows the proportion of true positives among actual positives. These are crucial in medical diagnostics, where both false negatives and false positives can have significant consequences.
- F1-Score: Balances precision and recall, providing a single metric that reflects the model's overall performance in terms of both precision and recall.

Additionally, performance is assessed using a confusion matrix and ROC curves:

- Confusion Matrix: Shows the model's true positives, false positives, false negatives, and true negatives, offering a detailed view of classification performance.
- ROC Curves and AUC: Help evaluate the model's ability to distinguish between classes at various thresholds, with the AUC representing the model's overall ability to discriminate.

Hyperparameter Tuning:

To enhance the CNN model's performance, several hyperparameters were finely tuned:

- Learning Rate: Adjusted to ensure efficient convergence of the loss function.
- Batch Size: Chosen to balance training time and gradient stability.
- Number of Filters and Filter Size: Explored to identify the optimal configuration for feature extraction.

- Dropout Rate: Optimised to prevent overfitting and improve model robustness.

A grid search strategy was employed to systematically evaluate different combinations of these hyperparameters using cross-validation, leading to the selection of the most effective settings. The chosen hyperparameters were validated with a separate validation set, ensuring robustness and reliability.

Iterative Refinement:

The training and validation process is iterative, with adjustments made to the model's architecture, hyperparameters, or training methods based on validation results. This iterative approach continues until the model achieves an optimal balance of accuracy, generalisability, and robustness, ensuring its effectiveness for clinical applications in lung cancer stage classification.

Statistical Methods

In analysing the IQ-OTH/NCCD lung cancer dataset, a range of statistical and machine learning techniques were utilised to thoroughly evaluate the data, focusing primarily on classification metrics to gauge the performance of predictive models.

- Confusion Matrix: This tool provides a visual representation of the model's performance by displaying the counts of true positives, true negatives, false positives, and false negatives. It is essential for understanding the model's classification accuracy and identifying areas where misclassifications occur.

- Accuracy: Accuracy measures the proportion of correctly predicted observations out of the total observations. It is calculated by dividing the number of correct predictions by the total number of observations, as shown in Eq. 16. While accuracy is a useful metric, it can be misleading in cases where class distributions are imbalanced. Thus, it is important to consider additional metrics to gain a more comprehensive view of model performance.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \times 100\%$$

\text{Accuracy}=\frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}\times 100\% [Eq.16]

- Precision (Positive Predictive Value): Precision was used to evaluate the accuracy of positive predictions. It is calculated as the ratio of true positives to the sum of true positives and false positives. This metric is particularly important in situations where the consequences of false positives are significant. The precision is determined using Eq. 17.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

\text{Precision}=\frac{\text{True Positives}}{\text{True Positives}+\text{False Positives}} [Eq.17]

- Recall (Sensitivity or True Positive Rate): Recall measures the model's effectiveness in identifying positive instances. It is calculated as the ratio of true positives to the sum of true positives and false negatives. This metric is crucial in medical diagnostics, where missing a positive case can have serious implications. Recall is determined using Eq. 18.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

\text{Recall}=\frac{\text{True Positives}}{\text{True Positives}+\text{False Negatives}} [Eq.18]

- F1-Score: The F1-score, representing the harmonic mean of precision and recall, was employed to balance these two metrics. It is especially useful in cases of class imbalance. It offers a more reliable measure than accuracy, particularly when false negatives and false positives have varying levels of importance. The F1-score is calculated using Eq. 19.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

\text{F1-Score}=2\times \frac{\text{Precision}\times \text{Recall}}{\text{Precision}+\text{Recall}} [Eq.19]

- Cohen's Kappa: Cohen's Kappa statistic was used to evaluate the agreement between the observed and predicted classifications, factoring in the agreement that might occur by chance. This metric provides a more detailed insight into the model's performance, especially in the context of imbalanced datasets. It is calculated using Eq. 20.

$$\text{Cohen's Kappa} = \frac{p_o - p_e}{1 - p_e}$$

\text{Cohen's Kappa}=\frac{{p}_{o}-{p}_{e}}{1-{p}_{e}} [Eq.20]

- Mean Squared Error (MSE) and Root Mean Squared Error (RMSE): MSE (Mean Squared Error) and RMSE (Root Mean Squared Error) were calculated to assess the average squared difference and the square root of the average squared differences,

respectively, between predicted and actual classification categories. These metrics are crucial for evaluating the variance in prediction errors. MSE and RMSE are computed using Eqs. 21 and 22, respectively.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2$$

\text{MSE}=\frac{1}{n}{\sum }_{i=1}^{n}{\left({y}_{i}-\widehat{{y}_{i}}\right)}^{2} [Eq.21]

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}}\,(22)$$

\text{RMSE}=\sqrt{\text{MSE}}\left(22\right) [Eq.22]

- Mean Absolute Error (MAE): The Mean Absolute Error (MAE) calculates the average size of the errors in a prediction set, ignoring whether the errors are positive or negative. As a linear metric, it treats all errors equally, assigning the same weight to each difference. This is determined by using Eq. 23.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \widehat{y_i}|$$

\text{MAE}=\frac{1}{n}{\sum }_{i=1}^{n}\left|{y}_{i}-\widehat{{y}_{i}}\right| [Eq.23]

- Receiver Operating Characteristic (ROC) Curve and Area Under the Curve (AUC): The ROC curve visually represents the model's diagnostic performance by plotting the true positive rate against the false positive rate across different threshold levels. The AUC (Area Under the Curve) provides a single value summarising the model's overall performance across all classification thresholds. This is calculated using Eq. 24.

$$\mathrm{AUC} = \int_0^1 \mathrm{ROC\ Curve}\,(t)\,dt$$

\text{AUC}={\int }_{0}^{1}\text{ROC Curve}\left(t\right)dt [Eq.24]

- F2-score: The F2-score was calculated to place greater emphasis on recall over precision, which is particularly useful in scenarios where failing to identify positive cases is more harmful than generating false positives. This metric is determined using Eq. 25.

$$\text{F2-Score} = \left(1 + 2^2\right) \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\left(2^2 \times \mathrm{Precision}\right) + \mathrm{Recall}}$$

$$\text{F2-Score}=\left(1+{2}^{2}\right)\times \frac{\text{Precision}\times \text{Recall}}{\left({2}^{2}\times \text{Precision}\right)+\text{Recall}}$$ [Eq.25]

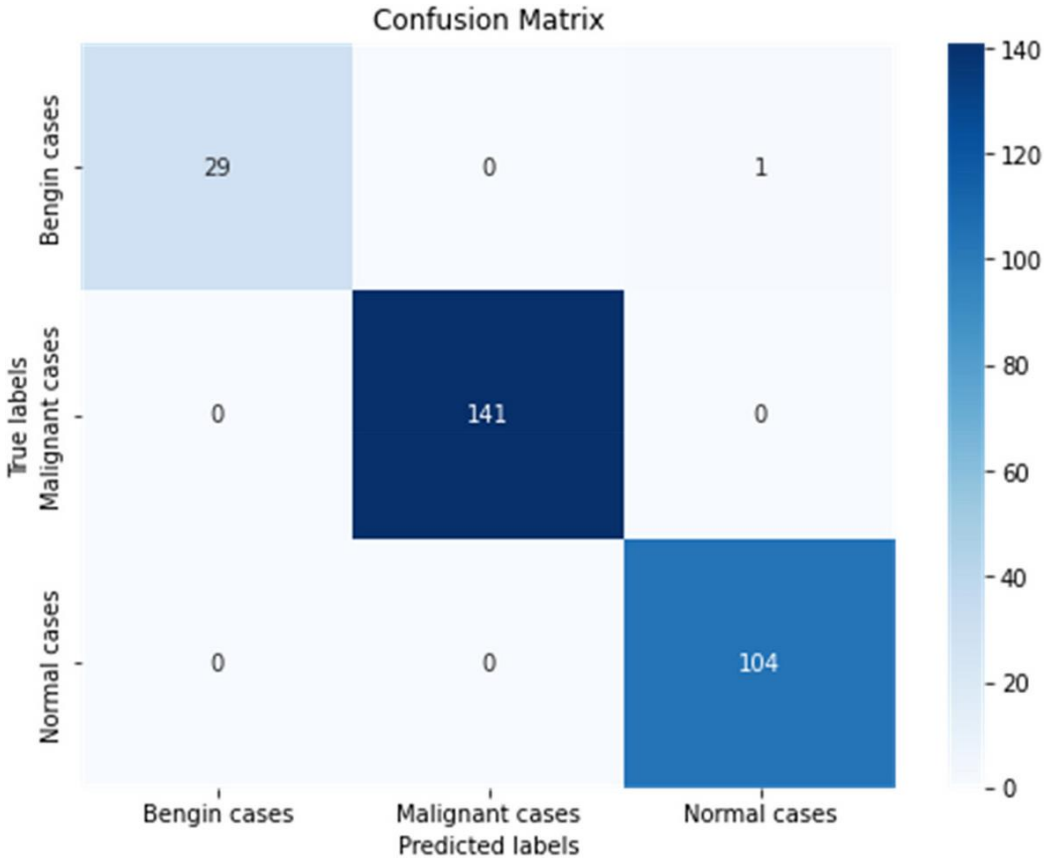
These statistical methods and metrics offered a comprehensive assessment of the model's performance, ensuring a thorough analysis of its predictive accuracy and reliability in classifying cases within the IQ-OTH/NCCD lung cancer dataset.

Results

The results from evaluating the IQ-OTH/NCCD lung cancer dataset using the predictive model provided comprehensive insights across various statistical metrics, highlighting the model's effectiveness in classifying lung cancer stages. A detailed analysis of each metric is as follows:

- Confusion Matrix: The confusion matrix offered an in-depth view of the model's classification performance, showing a high number of true positives and true negatives, indicating accurate predictions. The minimal instances of false positives and false negatives further emphasized the model's accuracy in distinguishing between benign, malignant, and normal cases. This is illustrated in Fig. 7.

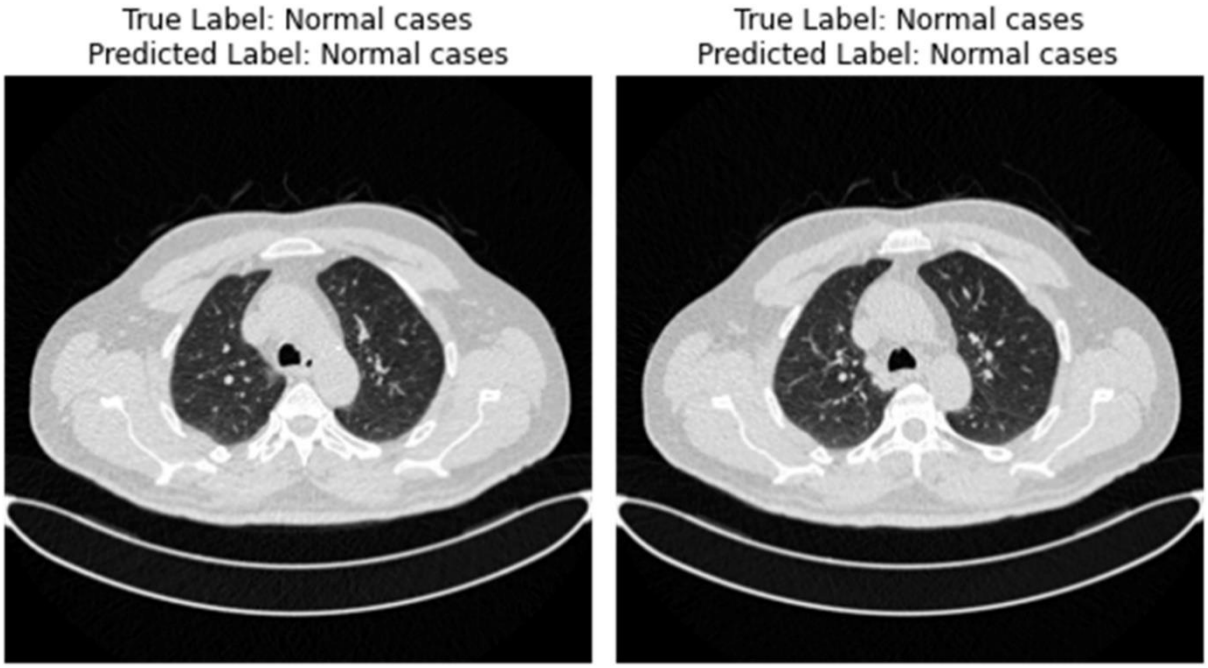
Fig. 7



Confusion matrix

- Accuracy: The model achieved an impressive overall accuracy of 99.64%, demonstrating its strong ability to correctly identify and classify instances within the dataset. This high accuracy underscores the model's reliability and effectiveness in clinical diagnostic settings, providing a solid foundation for further validation and potential clinical use. Fig. 8 visually presents the correctly classified instances, offering additional insight into the model's performance.

Fig. 8



Correctly classified instances

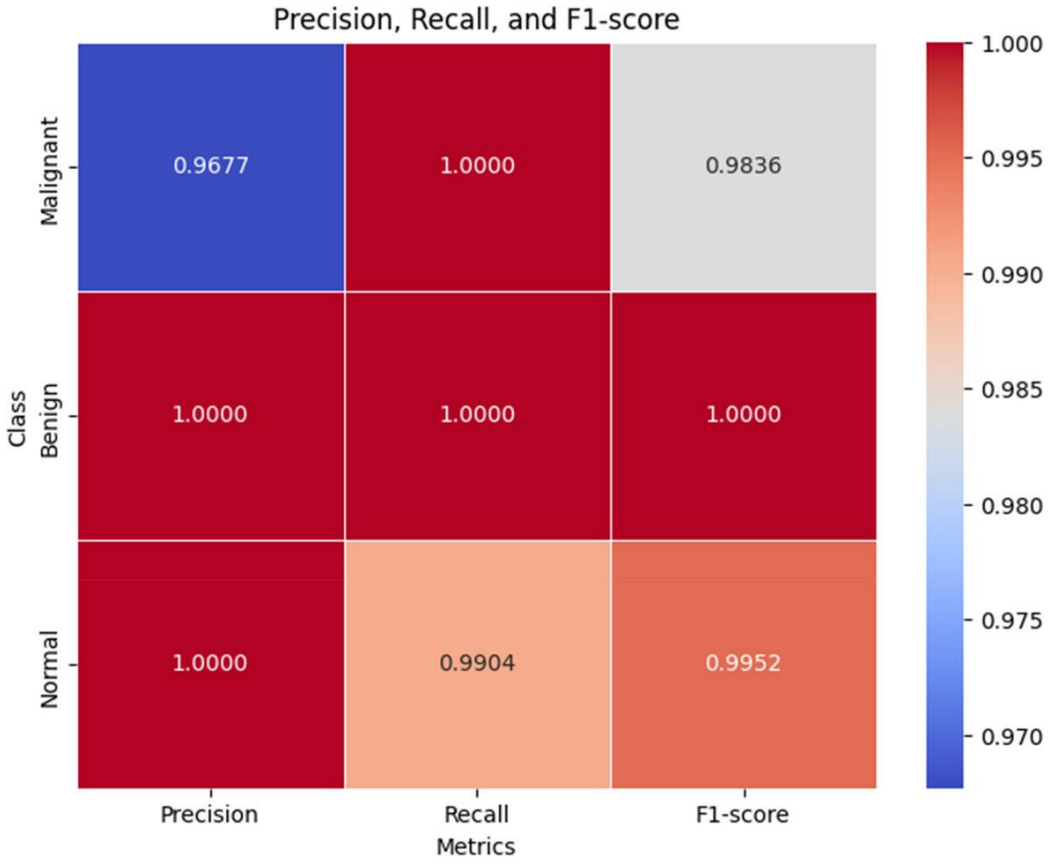
- Precision: The precision metric revealed a strong performance by the model, achieving 96.77% for benign cases. This indicates a high likelihood that cases predicted as benign are truly benign. Additionally, the model achieved 100% precision for both malignant and normal cases, highlighting its exceptional accuracy in predicting these categories without false positives.
- Recall: The recall scores were impressive, with the model reaching 100% for both benign and malignant cases, and 99.04% for normal cases. These results reflect the model's excellent sensitivity in detecting true positive cases, which is crucial in medical diagnostics to avoid missing any positive instances.
- F1-score: The F1-scores, which balance precision and recall, were 98.36% for benign cases, 100% for malignant cases, and 99.52% for normal cases. These scores demonstrate the model's effective performance in maintaining accuracy while minimizing false negatives. For a detailed view of the classification metrics, Table 3 presents a comprehensive statistical summary.

Table 3 Classification report

	Precision	Recall	F1-score
Malignant	0.9677	1	0.9836
Benign	1	1	1
Normal	1	0.9904	0.9952

Based on Table 3, Fig. 9 provides a heatmap to offer a clearer visual representation of the classification metrics and performance details.

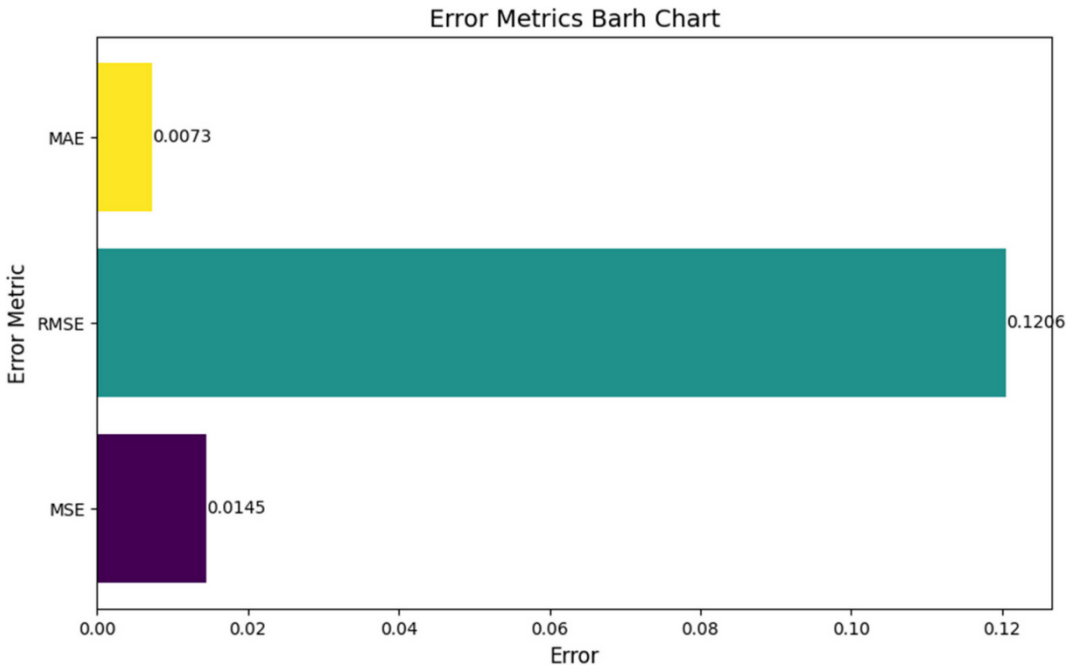
Fig. 9



Classification report

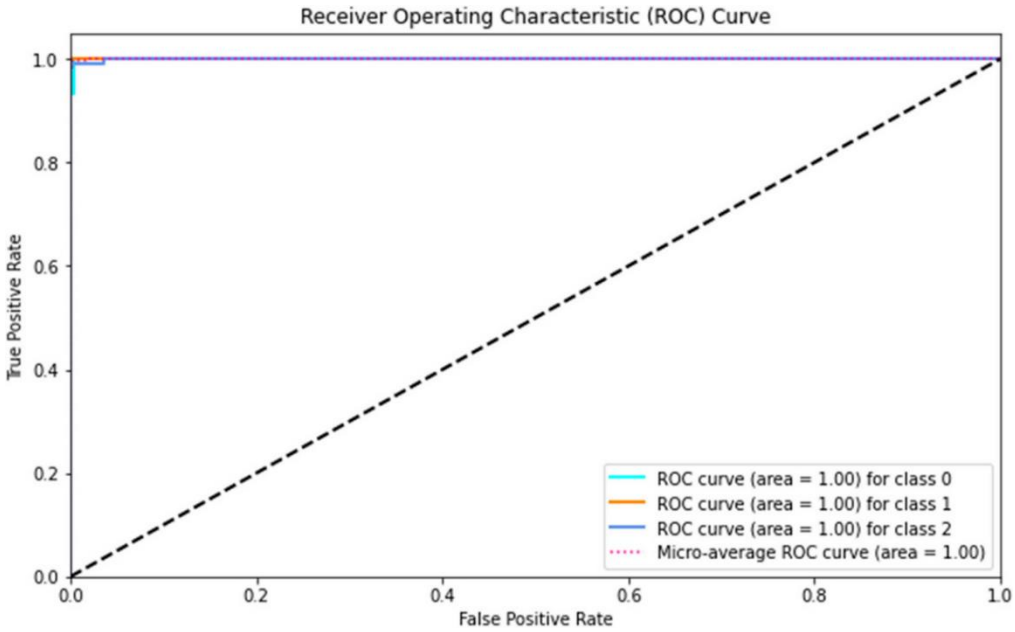
- Cohen's Kappa: The model achieved a Cohen's Kappa score of 0.9938, reflecting near-perfect agreement with actual classifications and indicating a high level of consistency beyond random chance.
- Mean Squared Error (MSE) and Root Mean Squared Error (RMSE): The MSE was 0.0145, and the RMSE was 0.1206, suggesting minimal prediction error variance and indicating that the model's predictions closely align with actual values.
- Mean Absolute Error (MAE): With an MAE of 0.0073, the model demonstrated a minimal average prediction error, highlighting its high accuracy. A bar chart visualising these error metrics is provided in Fig. 10.

Fig. 10



- Receiver Operating Characteristic (ROC) Curve and Area Under the Curve (AUC): The ROC curves and AUC values were exemplary, with AUCs of 1.00 for malignant, benign, and normal cases. This perfect score highlights the model's excellent ability to distinguish between different classes at various thresholds. The ROC-AUC curve is shown in Fig. 11.

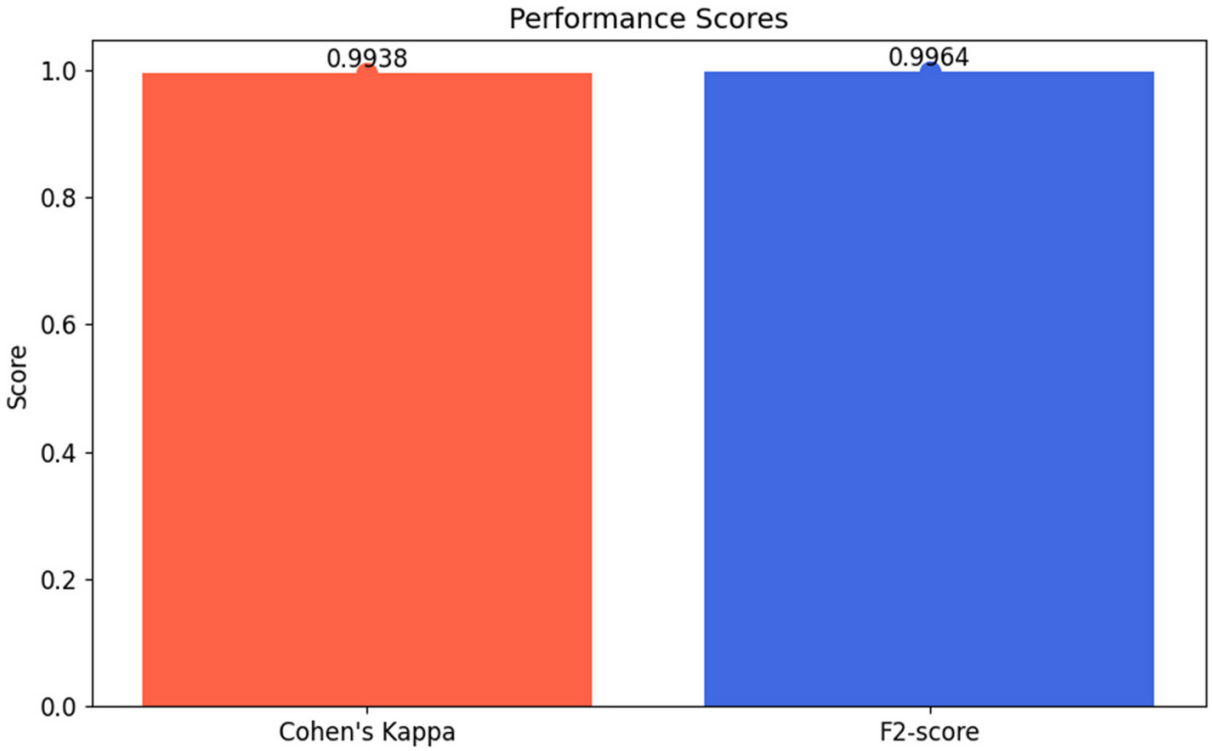
Fig. 11



ROC curve

- F2-score: The model achieved an F2-score of 0.9964, emphasising its strong capability in identifying positive cases. This is crucial in medical diagnostics where accurate detection of conditions is vital. The performance score is visually represented in Fig. 12.

Fig. 12



Performance scores

The detailed results across these metrics offer a thorough evaluation of the model's performance, showcasing its precision, reliability, and robustness in classifying lung cancer stages from the IQ-OTH/NCCD dataset. These findings highlight the model's potential as a valuable diagnostic tool, warranting further exploration and consideration for clinical application.

Discussion

The analysis of the IQ-OTH/NCCD lung cancer dataset using the model demonstrates exceptional performance in medical image classification. With an accuracy of 99.64% and impressive precision and recall metrics across benign, malignant, and normal categories, the model proves to be a highly dependable diagnostic tool. These results are noteworthy not only for their high scores but also for the model’s ability to effectively differentiate between benign and malignant cases, which is crucial for patient management and treatment strategies.

The high F1-score highlights the model’s balanced approach to precision and recall, thus reducing the likelihood of misdiagnosis. The F2-score, which prioritises recall, is particularly significant in the medical field where missing a positive case (false negative) can have

serious repercussions. A comparison of the proposed model with baseline models is presented in Table 4.

Table 4 Comparison with existing studies

Study	Objective	Accuracy
Asghar Ali Shah et al. (2023) [23]	Convolutional Neural Networks (CNNs) and Ensemble	95%
Mohammad A. Alzubaidi et al. (2021) [24]	SVM with HOG features	88%
Dimitrios Mathios et al. (2021) [25]	Cell-free DNA fragmentomes	94%
Shahid Mehmood et al. (2022) [26]	Transfer learning	98.40%
Elias Dritsas, Maria Trigka (2022) [27]	Rotation Forest model	97.10%
Mehedi Masud et al. (2021) [28]	Deep learning and digital image processing	96.33%
Iftkhar Naseer et al. (2023) [29]	Modified U-Net Based Lobe Segmentation and detection	97.70%
Bharathy S, Pavithra R, Akshaya. B (2022) [30]	Random Forest algorithm	88.50%
Gopi Kasinathan and Selvakumar Jayakumar (2022) [31]	Hybrid technique for PET/CT images	98.60%
Das, S., et al. (2023) [32]	CNN and Inception V3	93.44%
Tasnim, Nowshin, et al. (2024) [33]	CNN, Resnet50, and InceptionV3	98%
Safta, Wiem, and Ahmed Shaffie (2024) [34]	Integration of 3D-Local Octal Pattern (LOP) descriptor, 3D-Convolutional Neural Network (CNN), and geometric feature analysis	97.84%
Khaliq, Kiran, et al. (2023) [35]	Transfer learning with Densely Connected Convolutional Networks (DenseNet-121)	99%
Nigudgi, Surekha, and Channappa Bhyri. (2023) [36]	Transfer learning with hybrid model (AlexNet, VGG, GoogleNet)	97%
Proposed Model	Double Layered CNN with Advanced Image Processing	99.64%

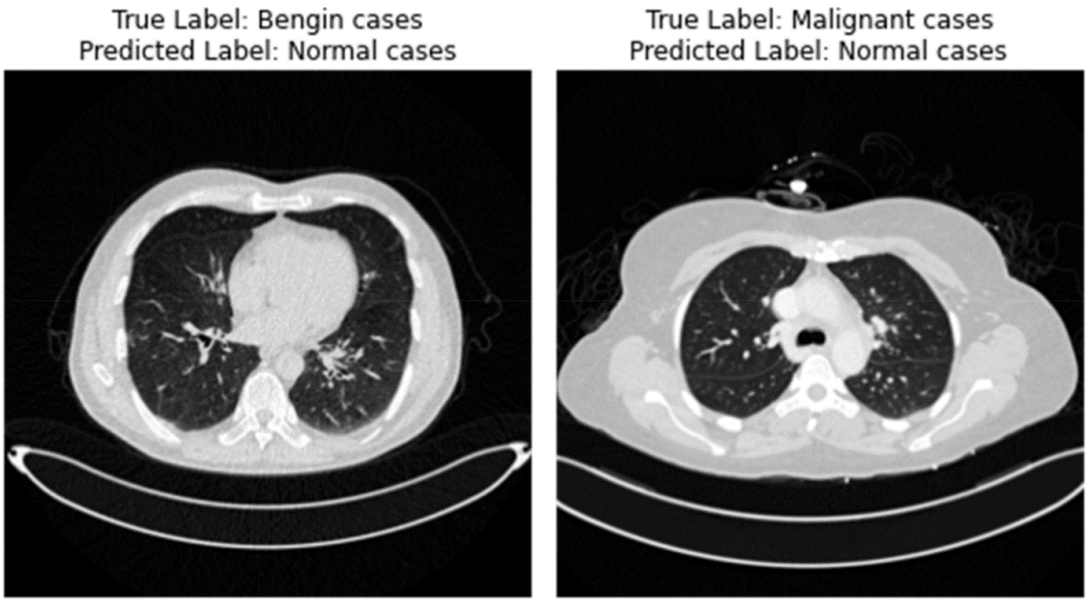
In the field of lung cancer detection, many existing models predominantly address binary classification, often overlooking the crucial distinction between benign and malignant cases [37]. The model's tri-classification capability represents a significant advancement, offering a more detailed diagnostic tool compared to traditional binary classifiers. When compared to existing methods, the performance of the model highlights its superior detection capabilities, potentially providing a more nuanced and informative diagnostic approach.

Integrating such a high-performing model into clinical practice could transform lung cancer diagnostics [22, 38]. It has the potential to enhance radiologists' efficiency, reduce diagnostic time, and increase throughput. Accurate classification of lung nodules as benign, malignant, or normal could significantly decrease unnecessary interventions, reducing patient exposure to invasive procedures and associated risks. Additionally, the model could streamline patient

management, facilitating prompt treatment for malignant cases and appropriate follow-up for benign conditions [39, 40].

However, the study has some limitations. The model was trained on data from a specific demographic and geographic area, which may affect its generalisability to broader populations. Furthermore, the model's performance in a controlled study setting might not fully reflect the variability encountered in real-world clinical environments. The opaque nature of deep learning models also poses a challenge in clinical contexts, where understanding the rationale behind a diagnosis is as important as the diagnosis itself [41]. Fig. 13 illustrates some instances of misclassification to provide clarity on these issues.

Fig. 13



Misclassified instances

The analysis of the IQ-OTH/NCCD lung cancer dataset demonstrates the impressive performance of the CNN model in medical image classification. Achieving an accuracy of 99.64% with exceptional precision and recall across benign, malignant, and normal categories, the model proves to be a highly reliable diagnostic tool. The high F1-scores further emphasise the model's balanced approach to precision and recall, which is crucial for minimising misdiagnoses. The model's strong emphasis on recall, as reflected in the F2-score, is particularly relevant in medical diagnostics, where failing to detect a condition can have significant consequences. A comparative analysis of the model with baseline models is presented in Table 4.

The model's ability to handle tri-classification—distinguishing between benign, malignant, and normal cases—sets it apart from existing binary classifiers, offering a more detailed diagnostic perspective. When compared to current methods, the model's performance highlights its advanced detection capabilities and suggests a more nuanced diagnostic approach.

Incorporating this model into clinical practice could revolutionise lung cancer diagnostics. It has the potential to enhance radiologists' efficiency by reducing diagnostic time and

increasing throughput. Accurate classification of lung nodules could minimise unnecessary interventions, reducing patient exposure to invasive procedures and associated risks. Additionally, it could streamline patient management, ensuring timely treatment for malignant cases and appropriate follow-up for benign conditions.

However, several limitations must be addressed. The dataset used primarily represents a specific demographic and geographic area, which may limit the model's generalizability to broader populations. Variations in demographics, such as age, ethnicity, and underlying health conditions, can affect lung cancer presentation in CT scans. Future research should expand the dataset to include a diverse range of CT images from various populations and regions. Collaborations with international medical institutions and the use of publicly available imaging repositories could facilitate this expansion. Additionally, employing advanced data augmentation techniques to simulate demographic variations could further enhance the dataset's diversity.

Sensitivity Analysis of Precision, Recall, and F1-Score

We performed a sensitivity analysis to evaluate the CNN model's performance regarding precision, recall, and F1-score. Adjusting classification thresholds impacted the model's false positive and false negative rates. Increased precision led to fewer false positives but a higher risk of false negatives, requiring a balance in medical diagnostics. Conversely, enhanced recall improved true positive detection but also increased false positives. The F1-score, as a harmonic mean of precision and recall, highlighted the balance between these metrics. Optimising for a high F1-score emphasised the model's overall effectiveness in balancing precision and recall, crucial for reliable lung cancer diagnosis.

Regulatory Considerations for Clinical Application

Implementing machine learning models in clinical settings involves navigating complex regulatory requirements to ensure patient safety, data security, and efficacy. Key regulatory hurdles include obtaining approval from medical device regulatory bodies like the U.S. FDA or the European Medicines Agency. These agencies require extensive validation studies to demonstrate the model's accuracy and reliability across diverse datasets and clinical scenarios.

Regulatory guidelines also demand interpretability and transparency from machine learning models. Clinicians need to understand the model's decision-making process to integrate it effectively into clinical workflows. This requirement poses a challenge for deep learning models, often viewed as "black boxes." Developing methods for model explainability, such as feature importance analysis or visual explanations, is crucial for meeting regulatory standards.

Data privacy and security are significant concerns, with regulations like GDPR and HIPAA governing patient data protection. Ensuring anonymisation, secure storage, and ethical use of patient data is essential, involving robust encryption, access controls, and audit trails.

Post-market surveillance is also vital for regulatory compliance, requiring ongoing monitoring of the model's performance, identifying potential biases, and updating the model as needed. Establishing a framework for continuous evaluation and improvement is crucial for maintaining efficacy and safety.

Addressing these regulatory challenges requires collaboration between developers, healthcare providers, and regulatory bodies. By adhering to these frameworks, we can

successfully integrate advanced diagnostic tools into healthcare, improving patient outcomes and advancing medical diagnostics.

Future Research Directions

Future research should focus on validating the model across various populations and healthcare settings to assess its universality and robustness. Integrating multimodal data, such as patient history and genetic information, could enhance diagnostic precision. Improving the interpretability of deep learning models could facilitate their integration into clinical decision-making processes. Additionally, prospective studies evaluating the model's impact on clinical outcomes, patient satisfaction, and healthcare efficiency would provide valuable insights into its practical benefits and areas for improvement.

Conclusion

This study conducted a thorough examination of the IQ-OTH/NCCD lung cancer dataset using an advanced machine learning model, which exhibited outstanding performance in classifying lung cancer stages. The model achieved an impressive accuracy rate of 99.64% and demonstrated remarkable precision and recall across benign, malignant, and normal case classifications. The balanced F1-score and high emphasis on recall in the F2-score further underscore the model's diagnostic accuracy and sensitivity. These findings represent a significant advancement in the model's ability to discern between subtle stages of lung cancer, offering a crucial tool for early and precise diagnosis.

The impact of these results on lung cancer diagnostics is profound. The model's exceptional accuracy in classifying lung cancer stages promises to enhance diagnostic procedures, improving both accuracy and efficiency in detecting lung cancer. This advancement could lead to earlier treatment interventions, potentially improving patient outcomes and survival rates. Furthermore, the model's capacity to distinguish between benign and malignant nodules may reduce the need for unnecessary invasive procedures, thereby decreasing patient risk and lowering healthcare costs.

Future research should prioritise the external validation of the model to confirm its effectiveness across varied populations and clinical environments. Enhancing the model's interpretability is also essential for clinical adoption, as understanding the rationale behind diagnostic decisions is crucial in medical practice. Integrating the model with other diagnostic data and clinical workflows could further increase its utility and impact.

Prospective studies are necessary to assess the model's real-world clinical impact, particularly its potential to improve patient outcomes, streamline diagnostic processes, and reduce healthcare expenditures. Additionally, exploring the model's adaptability to other types of cancers or medical imaging modalities represents an exciting direction for future research.

This study underscores the transformative potential of sophisticated machine-learning models in lung cancer diagnostics. By providing a more precise and nuanced approach to detecting and classifying lung cancer, these models hold the promise of advancing patient care and outcomes in oncology. The continued development and integration of such models into clinical practice are expected to drive significant progress in the field of cancer diagnosis and treatment.