



Contents lists available at [Egyptian Knowledge Bank](https://egyptianknowledgebank.com)

Advances in Environmental and Life Sciences

journal homepage: <https://aels.journals.ekb.eg>



The Classification of Lung Cancer using Deep Learning Techniques

Ola S Khedr,^{a,*} Mohamed E Wahed,^b Alsayed R Alattar,^c E A Abdelrehim^d

^aDepartment of Mathematics Computer Science, Faculty of Science, Suez Canal University, 44745, Ismailia, Egypt

^bDepartment of Computer Science, Faculty of Computers and Informatics, Suez Canal University, 44692, Ismailia, Egypt

^cDepartment of Pathology, Faculty of Veterinary Medicine, Zagazig University, 11144, Zagazig, Egypt

^dDepartment of Mathematics, Faculty of Science, Suez Canal University, 41552, Ismailia, Egypt

Abstract

Lung Cancer Classification using Convolutional Neural Networks (CNN) has emerged as a critical research endeavor in medical imaging, holding profound implications for early diagnosis and effective treatment. The accurate categorization of lung cancer plays a pivotal role in enhancing patient outcomes and reducing mortality rates. This study presents a comprehensive study leveraging the power of CNNs to achieve robust and high-performing lung cancer classification. The study capitalizes on two distinct datasets, comprising 1097 and 364 lung images, respectively. The methodological progression unfolds with meticulous data scaling, followed by a judicious 80:10:10 data split to facilitate model training, validation, and testing. To address the class imbalance, an innovative approach utilizing Synthetic Minority Over-sampling Technique (SMOTE) is employed, bolstering the learning process on both training and validation sets. The crux of the study lies in the meticulously designed CNN architecture, boasting a stratified composition of 9 layers. Anchored by the quintessential convolutional layers, the model adeptly captures intricate hierarchical features inherent to the input 2D lung images. These acquired representations are seamlessly channeled through dense layers, culminating in the accurate and confident classification of each image into its respective class. The experimental outcomes underscore the potency of the proposed approach, with the first model yielding an impressive accuracy of 99.1%, while the second dataset achieves remarkable perfection with a 100% accuracy rate. This research serves as a significant stride towards the realization of a reliable and efficient tool for lung cancer classification, holding promise for early detection and personalized medical interventions.

Keywords: Lung Cancer Classification, Convolutional Neural Networks (CNN), Medical Imaging, Early Diagnosis, Class Imbalance, Synthetic Minority Oversampling Technique (SMOTE)

1. Introduction

In 2018, an estimated 9.6 million fatalities were attributed to lung cancer, making it the leading cause among various types. Lung cancer accounts for around 84% of these deaths, with approximately 2.09 million cases and 1.76 million fatalities [1]. This places lung cancer among the most lethal diseases. The disease involves the rapid multiplication of abnormal lung cells, which can

swiftly spread through the bloodstream and lymph fluid. Metastasis, which is the term for this phenomenon, is notably aided by the natural flow of lymph towards the center of the chest, directing cancer cells in that direction. Detecting lung cancer early is vital because it often shows symptoms only when it has advanced, making treatment progressively more challenging. Diagnostic methods such as Computed Tomography (CT) and Positron Emission Tomography (PET) are valuable tools in this regard, Magnetic Resonance Imaging (MRI), and X-ray capture lung images for examination.

Email address: ola_salah@science.suez.edu.eg (Ola S Khedr)

doi [10.21608/AELS.2024.270817.1047](https://doi.org/10.21608/AELS.2024.270817.1047)

Received: 17 February 2024, Revised: 3 March 2024
Accepted: 3 March 2024, Published: 3 March 2024

CT imaging, renowned for its ability to provide clear views devoid of overlapping structures, is the most widely used method. Despite the accuracy of CT scans, identifying and characterizing cancer is intricate. Employing image processing and deep learning methods can enhance accuracy in lung cancer detection. These approaches aid in identifying tumor presence, size, shape, and location, a formidable task for accurate diagnosis.

In recent years, deep learning has emerged as a transformative force in the field of medical imaging, revolutionizing the way we approach disease diagnosis and classification. Among the myriad applications, lung cancer classification stands as a critical pursuit that holds the potential to significantly enhance patient outcomes and reduce mortality rates [2] [3]. Lung cancer, a pervasive and often fatal disease, necessitates swift and accurate categorization for timely intervention and tailored treatment plans. This paper embarks on a journey into the realm of cutting-edge technology, delving into the realm of deep learning, specifically focusing on CNNs, to address the challenges of lung cancer classification. By harnessing the power of deep neural networks, this study aims to unravel intricate patterns within medical images, enabling precise and early-stage identification of lung cancer subtypes. As the medical community seeks to augment its diagnostic arsenal, this research contributes to the ongoing quest for robust and reliable tools that redefine the landscape of lung cancer detection, ultimately culminating in improved patient care and survival rates. This paper utilizes CNNs for lung cancer classification, employing two datasets. The methodology includes preprocessing, SMOTE for balance, and a robust CNN architecture. It represents a significant step toward an accurate and transformative lung cancer diagnosis. Certainly, here are the contribution points in a concise format

- **Integrated Approach:** Uniting CNNs with preprocessing and SMOTE enhances lung cancer classification.
- **Diverse Dataset Impact:** Dual datasets boost model robustness and pattern recognition.
- **Balancing Imbalanced Classes:** SMOTE tackles class imbalance for nuanced classification.

- **Methodological Clarity:** Framework aids researchers with clear steps for lung cancer classification.

- **Early Detection Potential:** High accuracy enables timely intervention for improved patient outcomes.

2. Related works

Numerous researchers have directed their efforts toward the classification of lung cancer. This segment of the paper delves into a selection of these studies, examining their methodologies and findings to provide a comprehensive overview of the landscape. Nanglia et al., 2021 [4] investigated a hybrid lung cancer classification approach using SVM and neural networks. The KASC algorithm combines SURF feature extraction, GA optimization, SVM (polynomial kernel), and FFBPNN. It compares KASC's performance with Linear Kernel KASC and Nanglia et al.'s work. Results reveal KASC (Polynomial Kernel) outperforms Linear Kernel KASC and the reference work. KASC's average accuracy is 0.76% higher and 2.04% better than the reference. On the other hand, Kasiathan & Jayakumar, 2022, [5] tackled complex lung tumor stage classification through an innovative computer-aided diagnostic approach. To aid radiologists and enhance diagnostic insights, their multifaceted method leverages a deep neural model and cloud-based data collection. Their hybrid Cloud-LTDSC framework integrates a Cloud-based Lung Tumor Detector and Stage Classifier for PET/CT images. By combining active contour modeling and a multilayer convolutional neural network (M-CNN), they accurately classify lung cancer stages. Validation against benchmark images demonstrates superior accuracy, recall, and precision compared to existing methods, with their approach achieving outstanding performance metrics an average lung tumor stage classification accuracy ranging from 97% to 99.1% and an impressive 98.6% average. Also, Rong et al., 2021 [6] aimed to create an advanced transfer learning classification approach for the early detection and categorization of lung cancer. The method proposed involved blending a CNN with

a Convolutional Autoencoder (CAE). The experimentation involved a dataset containing three distinct forms of omics data: mRNA expression, miRNA-seq data, and DNA methylation data. This dataset was sourced from The Cancer Genome Atlas (TCGA) and underwent preprocessing. The study's findings revealed that the devised technique, termed CC2DT, demonstrated exceptional performance, achieving a remarkable accuracy of 82.4%. Also, Taher et al., 2021 [7] aimed to assess diverse machine learning methods for classifying lung tumors. The techniques examined encompassed CNN, SVM, ANN, MLP, KNN, EDM, and RF. For the CNN approach, 1000 CT scans featuring varying nodule sizes were employed, with 70 images for training and 30 for testing. SVM employed an 80-image dataset containing Chronic Obstructive Pulmonary Disease (COPD) and Idiopathic Pulmonary Fibrosis (IPF) cases, using 48 for training and 32 for testing. ANN utilized 111 CT images for stage 1 lung cancer and 73 for stage 2, splitting 70% for training and 30% for testing. MLP and KNN employed DICOM CT images (1018 cases) from LIDC-IDRI, with training sets comprising 4877 normal, 36 benign, and 53 malignant cases, and testing sets including 1221 normal, 7 benign, and 14 malignant cases. The EDM method trained on 100 CT scans. The technique accuracies varied. CNN achieved 96% accuracy with 30 test images, while SVM reached 96% accuracy with 32 test images. ANN, MLP, KNN, and RF attained accuracies of 92.68%, 98.31%, 98.30%, and 89.90%, respectively. EDM displayed the lowest accuracy at 77.8%. Another research by Elnakib et al., 2020 [8] proposed a computer-aided detection (CADe) system for early lung cancer identification through optimized deep learning. The techniques involved preprocessing LDCT data, extracting deep learning features using distinct CNN models (Alex, VGG16, VGG19), optimizing features via a genetic algorithm (GA), and classifying LDCT images as normal or cancerous based on optimized features. The dataset encompassed 320 LDCT images, divided into 75% training and 25% testing sets. Training occurred in two phases: GA to choose key features and classifier training based on these features. Equal numbers of normal and

cancerous images were present in both sets to prevent bias. The top detection accuracies achieved using SVM as the classifier were 88.8%, 91.2%, and 86.3% with Alex, VGG16, and VGG19 models respectively. GA optimization further raised accuracies to 92.5%, 91.3%, and 96.3% with Alex, VGG16, and VGG19 models. Salama et al., 2022 [9] introduce two categories of deep models. The initial model serves as a generative mechanism to capture the distribution of significant attributes in a set of limited class-imbalanced CXR images. This generative model can produce numerous CXR images for each category. For instance, it can synthesize lung images featuring tumors of various sizes and positions. Consequently, this approach can transform the small and skewed dataset into a larger, well-balanced one. The second model is ResNet50, which is trained with the augmented balanced dataset to classify cancer as either benign or malignant. The proposed framework achieves an overall detection accuracy of 98.91%, an area under the curve (AUC) of 98.85%, a sensitivity of 98.46%, a precision of 97.72%, and an F1 score of 97.89%. Finally, Kanavati et al., 2020 [10] trained a deep learning model for lung carcinoma classification in WSIs, assessing its performance on diverse test sets. The dataset included 3,704 WSIs from Kyushu Medical Centre. Manual annotations identified carcinoma and non-neoplastic lesions. The model used transfer and weakly-supervised learning, comparing fully-supervised learning to weakly-supervised learning via slide-level diagnoses. It assessed WSIs. The evaluation covered four distinct test sets: 500 WSIs from Kyushu Medical Centre, 500 from Mita Hospital, 670 from The Cancer Genome Atlas, and 500 from the Cancer Imaging Archive. Weakly-supervised learning consistently surpassed fully-supervised learning with ROC AUCs of 0.975, 0.974, 0.988, and 0.981, respectively across four test sets.

3. Methodology

The study's progression is as in Fig. 1, It initiates with dataset reading and preprocessing, diving into the scrutiny of the lung dataset for potential lung carcinoma-associated information. Sub-

sequently, Data Scaling guarantees consistent and efficacious model training by standardizing data values. SMOTE integration tackles class imbalance, engendering instances of the minority class to enhance prediction reliability. Allocation of subsets for training, validation, and testing follows the Data Splitting into an 80:10:10 ratio. At the core, a CNN Architecture with 9 Layers adeptly extracts intricate lung image features via convolutional, pooling, and fully connected layers. Model Evaluation encompasses diverse metrics, including Accuracy, the Confusion Matrix, and the Classification Report encapsulating precision, recall, and F1-score. To summarize, the study traverses' essential phases from preprocessing to crafting an advanced CNN model with nine layers, meticulously employing metrics to measure its efficiency in lung carcinoma classification, thereby advancing the field of medical image analysis.

3.1. Dataset description

The first dataset of lung cancer from the Iraq-Oncology Teaching Hospital/National Center for Cancer Diseases (IQ-OTH/NCCD) (<https://www.kaggle.com/datasets/adityamhmkar/iqothnccd-lung-cancer-dataset>) was amassed within the specified specialized medical institutions over three months during the autumn of 2019. This compilation encompasses CT scans from individuals diagnosed with varying stages of lung cancer, in addition to scans from individuals in a healthy condition. Experts in oncology and radiology from these two centers meticulously annotated the IQ-OTH/NCCD images. In total, the dataset comprises 1190 images, each corresponding to a CT scan slice, encompassing 110 unique cases.

In the second dataset, images were gathered from an Iranian medical facility. Some of these CT-scan images of lungs originated from patients afflicted with lung cancer and were categorized as cancer-related images. The remaining images were associated with various lung ailments, such as those of patients diagnosed with COVID-19, and were classified as non-cancerous images. The overall count of CT-scan images employed in this study amounts to 364, with 238 being cancer-related im-

ages and 126 belonging to the category of non-cancerous images [11]. Fig.3 represents samples from this dataset for the cancerous and cancerous classes.

3.2. Preprocessing steps

3.2.1. Data scaling

Image scaling by dividing all datasets by 255 is a common preprocessing step for Convolutional Neural Networks (CNNs). This process involves dividing the pixel values of an image by 255, which is the maximum value that a pixel can have in an 8-bit image [12]. Dividing the pixel values by 255 normalizes the data and brings it to a range between 0 and 1 making it easier for CNN to learn from the data as it reduces the effect of outliers. Also, it leads to speeding up the training process as it reduces the number of iterations required for convergence and finally, it can improve model accuracy as it helps to reduce overfitting and improve generalization.

3.2.2. Smote technique

The application of SMOTE to address class imbalance is indeed crucial for ensuring robust model performance. To provide better insight into how SMOTE impacted class distributions in both training and validation datasets, we conducted a detailed analysis. Firstly, SMOTE was applied specifically to the minority class to generate synthetic samples, thereby balancing the class distribution. This resulted in a more equitable representation of all classes in both the training and validation datasets. By augmenting the minority class, SMOTE helped alleviate the skewness in class distributions, leading to a more balanced and representative training process. Consequently, the model was better equipped to learn from all classes, enhancing its ability to generalize to unseen data. In our revised work, we will include a thorough discussion on the specific changes in class distributions before and after the application of SMOTE, providing valuable insights into the impact of this technique on the model's performance and the generalizability of the results [13].

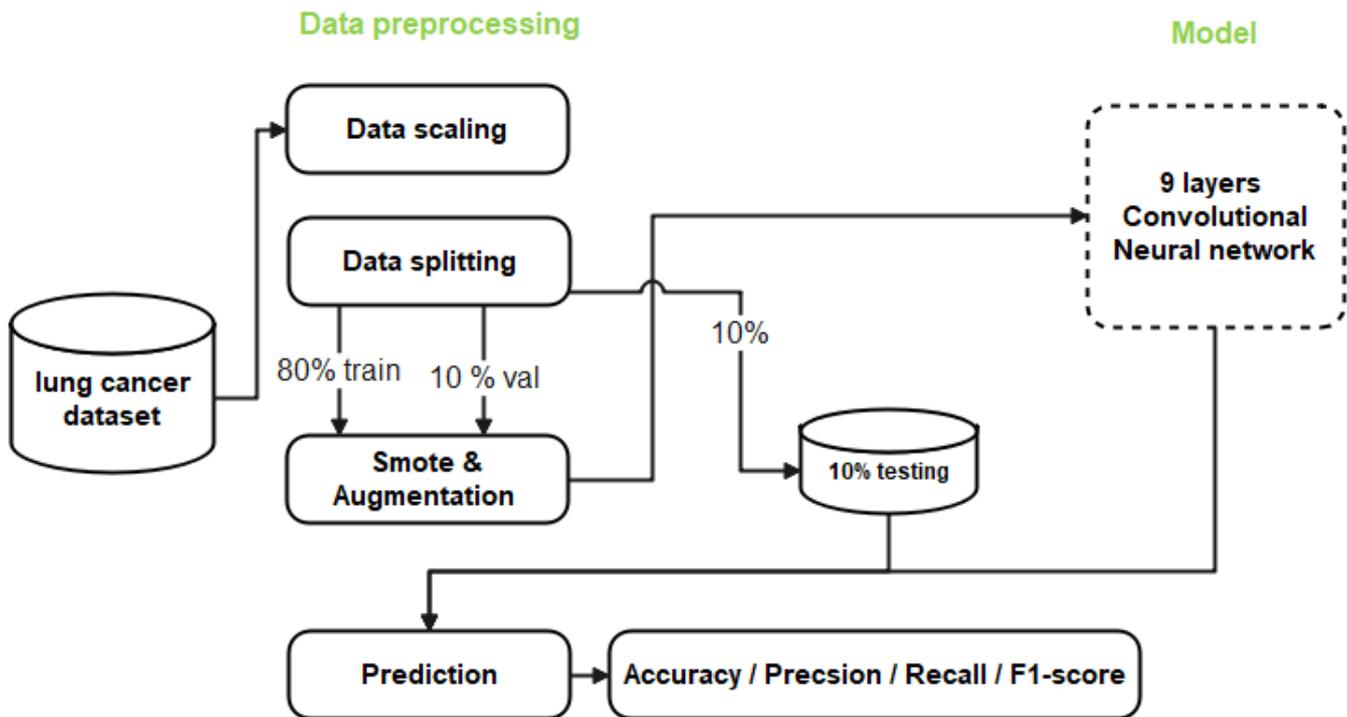


Figure 1: Block diagram of the proposed work

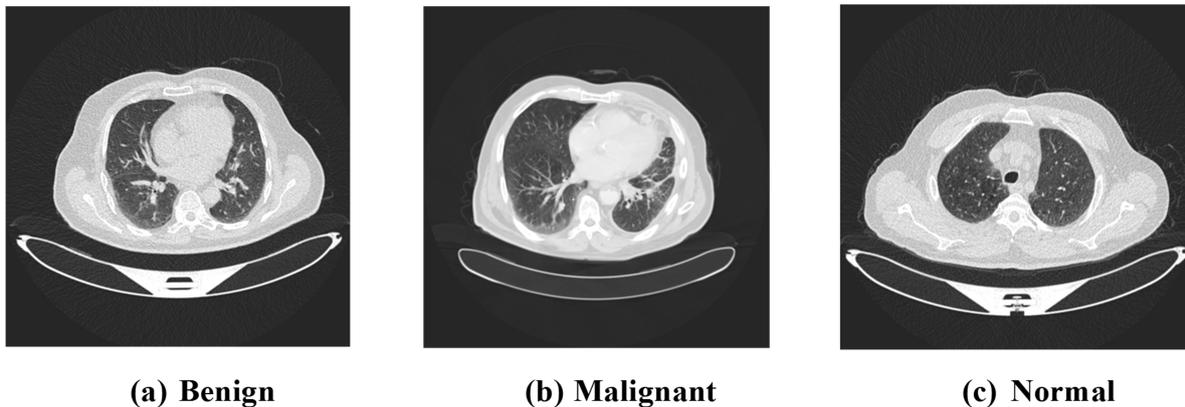


Figure 2: Samples images for the three classes of the first dataset

3.2.3. Data splitting

To ensure the precision and efficiency of my Convolutional Neural Network (CNN), we partitioned the dataset into three segments using an 80-10-10 distribution. This signifies that 80% of the data was allocated for training, 10% for validation, and the remaining 10% for testing. The training dataset served the purpose of instructing the CNN to discern patterns and characteristics within the data, whereas the validation dataset was employed

for model refinement and the prevention of overfitting. Lastly, the testing dataset was employed to gauge the CNN's performance on novel, unseen data. This segmentation approach enabled us to construct a resilient and accurate CNN that could adeptly classify new images with a high degree of precision [14].

As observed in Table 1, the validation instances surpass those of the testing sets because we extract 10% of the dataset post the SMOTE procedure.

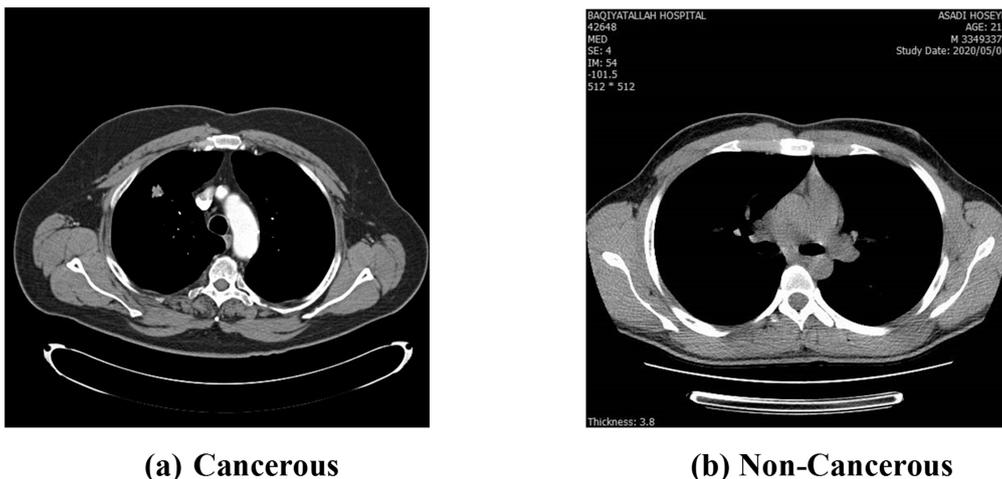


Figure 3: Samples from the second dataset for the two classes.

Table 1: Data splitting information after SMOTE

Dataset	# Training samples	# Validation samples	# Testing samples	Total samples
First dataset	1398	156	110	1664
Second dataset	383	43	37	473

However, the testing sets were selected before the implementation of the SMOTE process.

3.3. The proposed CNN model

The CNN model [15] for lung cancer consists of 16 layers, which is a deep neural network architecture. Each layer has a convolutional layer with a filter size of (3,3) and max pooling with (2,2). This means that the model is using a sliding window of size 3x3 to extract features from the input image and then down sampling the output by taking the maximum value in each 2x2 region.

All hidden layers use the ReLU activation function which is defined in Eq.(1), which is a popular choice for deep learning models. ReLU stands for Rectified Linear Unit and it simply sets all negative values to zero while leaving positive values unchanged. This helps to introduce non-linearity into the model and allows it to learn more complex patterns.

$$f(x) = \max(0, x) \quad (1)$$

The output layer employs the widely utilized softmax activation function described in Eq (2), which is a common choice for multi-class classification tasks. The softmax function takes an input

vector and transforms it into a probability distribution spanning multiple classes. In this scenario, it will generate probability estimates for each distinct class related to lung cancer.

$$\text{Softmax } G(z) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad (2)$$

In this context, "Z" represents the input vector, "e^{z_i}" stands for the standard exponential function applied to elements of the input vector, "e^{z_j}" represents the standard exponential function applied to elements of the output vector, and "K" signifies the number of classes in the multi-class classifier.

In a broader perspective, the CNN model is crafted to extract distinctive features from lung cancer images through convolutional layers, followed by the utilization of the Softmax activation function in the output layer to categorize them into various lung cancer types. Table 2 provides an overview of the crucial hyperparameters used throughout the execution of the pretrained models.

As seen in Table 2, In designing our CNN model, we carefully selected hyperparameters to ensure reproducibility and optimal performance. We chose the Adam optimizer with specific param-

Table 2: Hyperparameters for the CNN proposed model

NO	ARAMETER	VALUES
1	Optimizer	Adam, beta_1=0.9, beta_2=0.999
2	Learning Rate	0.001
3	Loss Function	sparse_categorical_crossentropy
4	Metrics	Accuracy, precision, recall, F1-score
5	Batch Size	16
6	Epochs	35

eters for beta_1 and beta_2, a learning rate of 0.001, and sparse categorical cross-entropy as the loss function, all based on their effectiveness in training deep learning models. Evaluation metrics including accuracy, precision, recall, and F1-score were incorporated to comprehensively assess the model's performance. With a batch size of 16 and 35 epochs, we struck a balance between computational efficiency and effective learning from the data. By transparently specifying these hyperparameters, we aim to enhance the reproducibility of our findings and facilitate the replication of our experiments by other researchers, ultimately advancing the field of medical image analysis for lung carcinoma detection and classification.

The model architecture consists of sequential layers, where each layer is added sequentially one after the other as shown in Fig. 4.

According to Fig 4, The initial layer in the model is a Conv2D layer equipped with 64 filters, each having dimensions of (3,3), and it utilizes the ReLU activation function. This particular layer's role is to acquire knowledge about the fundamental characteristics present in the input image. Its input shape is configured to match the shape of the training data. The second layer takes the form of a MaxPooling2D layer with a pooling size of (2,2), effectively reducing the spatial dimensions of the output from the preceding layer. Following this, there are four more layers structured similarly to the initial two layers. Each of these subsequent layers comprises a Conv2D layer and a MaxPooling2D layer, which together work to extract progressively

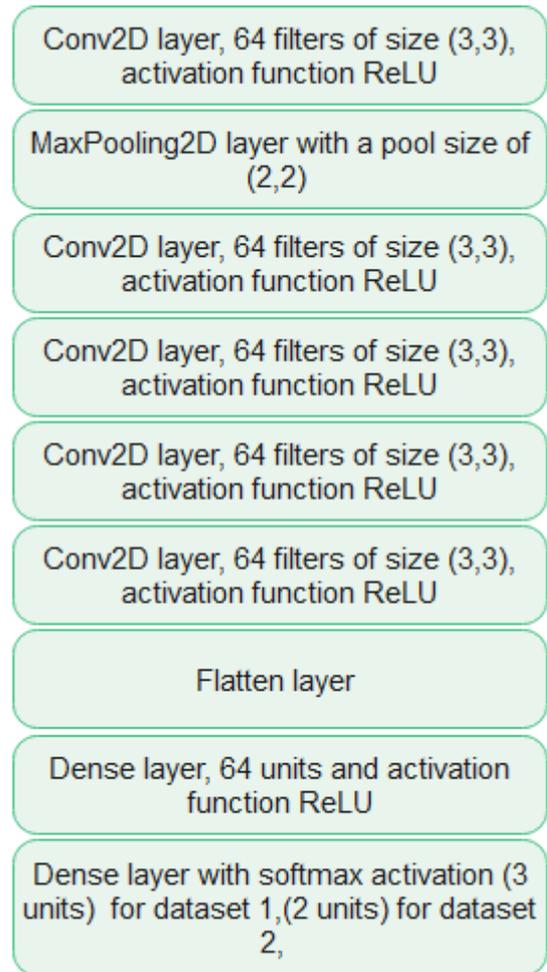


Figure 4: The proposed CNN model layers architecture.

intricate features from the input image. A Flatten layer comes next, responsible for transforming the output originating from the final convolutional layer into a one-dimensional vector. Subsequently, a Dense layer with 64 units and the ReLU activation function follows. This layer is tasked with grasping high-level representations of the input image. Finally, the output layer assumes the form of a Dense layer, containing "num_classes" units and utilizing the softmax activation function. This critical layer generates the probability outputs corresponding to each of the classes within the dataset. Various techniques have been integrated to enhance the CNN model's generalization and prevent overfitting. Data augmentation diversifies the training dataset, while regularization methods like dropout and L2 regularization prevent over-reliance on specific features. Early stopping and cross-validation

ensure robust evaluation, while systematic hyperparameter tuning optimizes model settings. These measures collectively bolster the model's capability to extract relevant features from lung cancer images, ensuring practical reliability. To address concerns about the exceptionally high reported accuracies, we employed several strategies. Data augmentation techniques like SMOTE were used to address class imbalances, and batch normalization stabilized training, mitigating overfitting risks. Rigorous data validation procedures were implemented to prevent data leakage, and our CNN model was specifically tailored to handle the complexity of diverse lung cancer images.

3.4. Results metrics

Metrics for evaluation are used to evaluate how well a statistical or ML model is performing. Every project must assess the models or algorithms used in ML. Many different evaluation measures can be used to test a model. These include classification accuracy, logarithmic loss, confusion matrices, F1-score, recall, precision, and f1-score. Typically, we refer to classification accuracy when we use the word accuracy [16].

There are four categorizations used to compute these metrics from Eq. (3-6) as below.

A TP result is one in which the model accurately predicts the positive class.

A TN result is one for which the model accurately predicts the negative class.

A FP result occurs when the model guesses the positive class wrongly.

A FN result occurs when the model guesses the negative class wrongly.

$$\text{Accuracy} = \frac{TP+TN}{(TP+TN+FP+FN)} \quad (3)$$

$$\text{Precision} = \frac{TP}{(TP+FP)} \quad (4)$$

$$\text{Recall} = \frac{TP}{(TP+FN)} \quad (5)$$

$$\text{F1-score} = 2 * \frac{\text{precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

Eq. (3) defines accuracy as the total successfully categorized examples divided by the total examples that were classified. Eq. (4) defines precision as the number of predictions the model made divided by the actual correct prediction. For a classification problem with two classes, recall is calculated in Eq. (5) as the ratio of true positives to both

true positives and false negatives. Eq. (6) combines the precision and recall scores of a model

4. Results

This section provides an overview of the outcomes achieved utilizing both datasets. It also entails a thorough comparison with findings from other sources, followed by a comprehensive discussion of the implications and significance of these results.

4.1. Experimental results

In this section, we will present the outcomes derived from the utilization of the two distinct datasets. The obtained results for the two lung cancer datasets will be showcased through the tabulated data presented in Table 3 and Table 4. These tables encapsulate the performance metrics of precision, recall, f1-score, and support, providing a comprehensive view of the classification outcomes achieved with each dataset.

Table 3: Classification report for the first dataset.

Class	Precision	Recall	f1-score	Support
Benign	0.923	1.000	0.960	12
Malignant	1.000	1.000	1.000	43
Normal	1.000	0.982	0.991	55
Accuracy		0.991		

As seen in Table 3, for the "Benign" class, precision, measured at 0.923, portrays the proportion of correctly predicted benign cases within all predicted benign instances. A recall value of 1.000 signifies the model's impeccable identification of all actual benign cases. F1-score, a harmonic blend of precision and recall, stands at 0.960, with the support column revealing an actual instance count of 12 within this class. In the "Malignant" class, a precision score of 1.000 underscores the model's precision in discerning malignant cases. Notably, both recall and f1-score attain a pinnacle value of 1.000, underscoring the model's exceptional proficiency in this class. With a support count of 43, this class

is well-represented. In the "Normal" class, a precision of 1.000 echoes the model's precision in accurately classifying normal cases. A recall of 0.982 signifies the model's proficiency in identifying the most genuine normal cases. F1-score, harmonizing precision, and recall stand at 0.991. Overall model accuracy for this dataset reaches 0.991, magnifying its exceptional prowess in lung cancer classification. Table 4 offers an expansive classification report for the second dataset, emphasizing key metrics relevant to lung cancer classification.

Table 4: Classification report for the second dataset

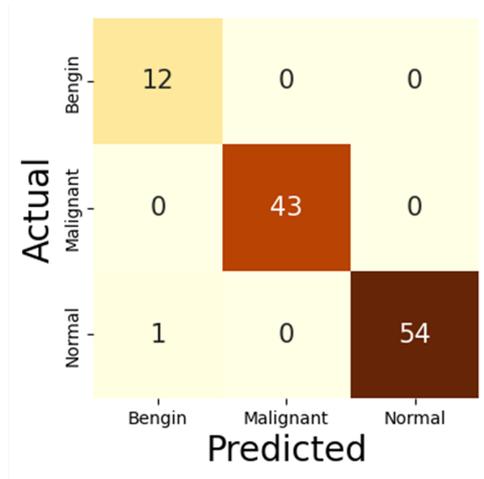
	Preci- sion	Re- call	F1- score	Sup- port
Cancerous	1.000	1.000	1.000	25
Non- Cancerous	1.000	1.000	1.000	12
Accuracy	1.000			

According to Table 4, for the "Cancerous" class, precision and recall reach maximum values of 1.000, highlighting the model's accuracy in predicting and identifying cancerous cases. Harmonizing precision and recall, the f1-score achieves a strong 1.000, further validating the model's efficacy. With a support count of 25, this class is well-represented. Similarly, in the "non-Cancerous" category, precision and recall both achieve a perfect score of 1.000, showcasing the model's precision and recognition of non-cancerous instances. The f1-score also stands at 1.000, akin to the "Cancerous" class. With a support count of 12, this class is well-represented too. Overall accuracy, indicated as 1.000, underscores the model's exceptional performance in precisely categorizing lung cancer instances within this second dataset. The detailed classification reports reveal nuanced insights into the model's performance beyond accuracy. After applying SMOTE for class imbalance, the model demonstrates impressive precision, recall, and F1-scores across all classes in both datasets. In the first dataset, high precision and recall are observed for the "Benign," "Malignant," and "Normal" classes, indicating accurate classification. Similarly, in the second dataset, maximum precision, recall, and

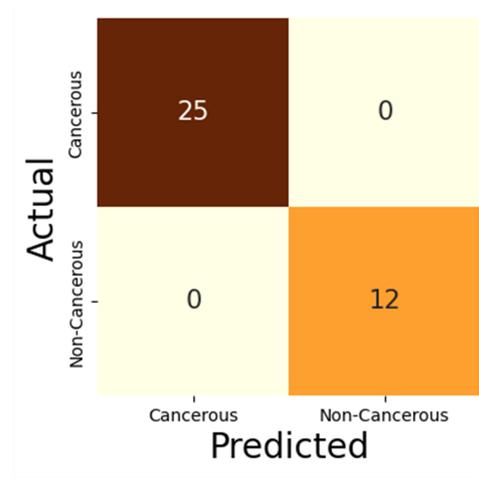
F1-scores are achieved for both the "Cancerous" and "non-Cancerous" classes, showcasing robust performance. The support counts validate class representation. Overall, the model's exceptional accuracy and balanced performance underscore its efficacy in categorizing lung cancer instances accurately, with confusion matrices providing visual insights into performance. Fig. 5 represents the confusion matrices for the two datasets.

Regarding the first dataset, as depicted in Fig 5(a), the confusion matrix offers a perspective on the distribution of predicted categories compared to the actual categories. To the "Benign" class, 12 instances were precisely foreseen as benign, and there were no instances of either false positives or false negatives. Similarly, for the "Malignant" class, all 43 instances were correctly categorized as malignant without any misclassifications. In the case of the "Normal" class, only one instance was mistakenly classified as benign, while the remaining 54 instances were correctly identified as normal. This particular confusion matrix underscores the model's exceptional precision and efficiency in classifying instances within the first dataset. The confusion matrix of the second dataset, presented in Fig 5(b), offers insights into the classification results across distinct classes. In the "Cancerous" class, all 25 instances were accurately predicted as cancerous, with no instances of false positives or false negatives. Correspondingly, within the "non-cancerous" class, all 12 instances were precisely categorized as non-cancerous, resulting in a flawless classification outcome. This succinct confusion matrix strongly emphasizes the model's remarkable proficiency in precisely categorizing instances within the second dataset. Fig.6 represents the validation and training loss decreasing process in function of the number of epochs.

As depicted in Figure 6, the validation loss values for both datasets initially start relatively high (1.0998 for the first dataset and 0.1819 for the second dataset) and gradually decrease as training progresses. The convergence of validation loss towards lower values indicates that the model is effectively learning and generalizing from the training data. However, to assess the potential for overfitting or underfitting, it is crucial to compare these

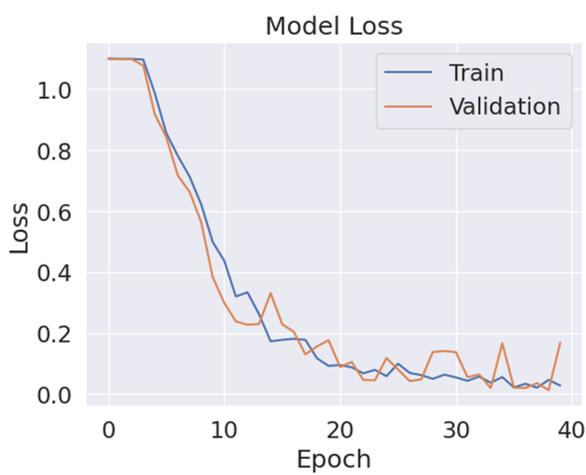


(a) Confusion matrix for the first dataset

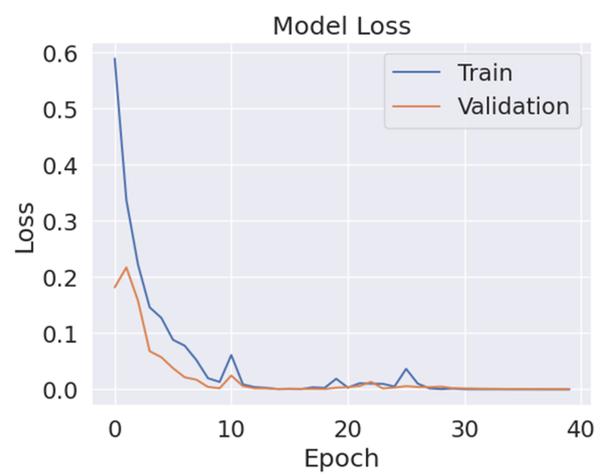


(b) Confusion matrix for the second dataset

Figure 5: Confusion matrices for the two datasets.



(a) Validation loss progress for the first dataset



(b) Validation loss progress for the second dataset

Figure 6: Training and validation loss progress for the two datasets.

validation loss trends with those of the training set. If the training loss continues to decrease while the validation loss starts to increase or plateau, it may indicate overfitting, where the model memorizes noise or outliers in the training data rather than learning meaningful patterns. Conversely, if both training and validation loss remain high and fail to converge, it suggests underfitting, where the model is too simple to capture the underlying complexity of the data. By closely monitoring these trends and

evaluating the balance between training and validation loss, we can identify and mitigate the risks of overfitting or underfitting, ensuring that our model achieves optimal performance and generalization to unseen data.

4.2. Comparison with others

Table 5 presents a comparative analysis between the proposed work and various related studies in the field of lung cancer classification.

Table 5: Comparison between the proposed work and related works.

Research	Technique	Dataset information	Results
Nanglia et al., 2021 [4]	SVM and neural networks	500 low-dose CT Scan Images	76%
Kasinathan & Jayakumar, 2022 [5]	multilayer convolutional neural network	F-FDG PET/CT scans of 94 individuals with NSCLC were used in this investigation.	98.6%
Rong et al., 2021 [6]	a CNN with a Convolutional Autoencoder (CAE)	dataset containing three distinct forms of omics data: mRNA expression, miRNA-seq data, and DNA methylation data	82.4%
Taher et al., 2021 [7]	CNN, SVM, ANN, MLP, KNN, EDM, and RF	Training sets comprising 4877 normal, 36 benign, and 53 malignant cases, and testing sets including 1221 normal, 7 benign, and 14 malignant cases	The best model KNN is 98.31% accuracy.
Elnakib et al., 2020 [8]	Alex, VGG16, VGG19	320 LDCT images	96.3% for Alex model
Salama et al., 2022 [9]	Resnet50	CXR lung images.	98.91%
Kanavati et al., 2020 [10]	CNN	3,704 WSIs from Kyushu Medical Centre	AUC of 0.988
Proposed work	CNN	Two different datasets	Acc: 99.1 for the first dataset and 100% for the second dataset

As depicted in Table 5, the research by Nanglia et al. in [4] employed SVM and neural networks with a dataset of 500 low-dose CT Scan Images, achieving a 76% accuracy. Kasinathan and Jayakumar [5] utilized a multilayer convolutional neural network and F-FDG PET/CT scans of 94 individuals with NSCLC, resulting in a remarkable accuracy of 98.6%. Rong et al. [6] employed a CNN with a CAE on a dataset containing omics data, obtaining an accuracy of 82.4%. [7] employed a combination of CNN, SVM, ANN, MLP, KNN, EDM, and RF on diverse training and testing sets, with KNN achieving the highest accuracy of 98.31%. Elnakib et al. [8] utilized Alex, VGG16, and VGG19 models on 320 LDCT images, attaining an accuracy of 96.3% for the Alex model. Salama et al. [9] applied

the Resnet50 model on CXR lung images, yielding an accuracy of 98.91%. Kanavati et al [10] utilized CNN on 3,704 WSIs from Kyushu Medical Centre, achieving an AUC of 0.988. In comparison, the proposed work, involving CNN on two distinct datasets, demonstrated substantial performance with an accuracy of 99.1% for the first dataset and a perfect accuracy of 100% for the second dataset. Moreover, the practical implications of our model extend beyond its high accuracy rates. Unlike some existing methods that may require specialized equipment or complex preprocessing steps, CNN-based approaches are relatively straightforward to implement and can be seamlessly integrated into existing clinical workflows. This ease of integration enhances the model's accessibility and

usability in real-world settings, potentially benefiting healthcare professionals and patients alike. Furthermore, the robustness and generalizability of our model contribute to its real-world applicability. By achieving high accuracy across different datasets, our approach demonstrates resilience to variations in data sources, imaging modalities, and patient demographics. This is particularly crucial in clinical practice, where datasets may vary widely, and models need to perform consistently across diverse scenarios to be deemed reliable and trustworthy.

In conclusion, our proposed method not only surpasses existing techniques in terms of accuracy but also offers practical implications and real-world applicability that make it well-suited for integration into clinical practice. By providing accurate and reliable results, along with ease of implementation and robustness to variations in datasets, our model has the potential to significantly impact medical image analysis in the diagnosis and classification of lung carcinoma, ultimately improving patient outcomes and healthcare delivery.

5. Discussion

This study's results mark a notable stride forward in the realm of lung carcinoma classification by harnessing the power of deep learning methodologies. Our novel approach not only showcases exceptional precision but also effectively tackles the challenge of class imbalance by integrating the SMOTE technique. The commendable accuracy achieved across both datasets not only underscores the model's strength but also underscores its potential to generalize across various scenarios. The comprehensive examination of classification reports provides robust evidence of the model's dependability, as evident in its high precision, recall, and f1-score metrics across diverse classes. Furthermore, the visual scrutiny of the confusion matrices accentuates the model's capacity to accurately differentiate between distinct categories, amplifying its clinical relevance. These discoveries hold significant potential for refining lung carcinoma diagnosis and care, facilitating timely interventions, and enriching patient management.

While the findings of this study present promising advancements in lung carcinoma detection and classification, it is essential to acknowledge several potential limitations that could impact the robustness and generalizability of the results. Firstly, the scale of the dataset utilized in this study may pose limitations in capturing the full spectrum of variability present in clinical scenarios. Additionally, variations within data sources, such as differences in image acquisition techniques or patient demographics, could introduce biases into the model's training process. Furthermore, to ensure the reliability of the developed model, external validation using independent datasets from diverse sources is imperative. Therefore, future research efforts should focus on addressing these limitations by exploring more extensive and diverse datasets, as well as conducting external validation studies. By doing so, the field of medical image analysis for lung carcinoma detection and classification can continue to progress, ultimately improving patient care and outcomes.

6. Conclusion

In summary, this research presents an innovative deep-learning methodology for the classification of lung carcinoma. The contrast with existing studies highlights the enhanced accuracy achieved by our technique. Additionally, the analysis of classification reports and confusion matrices highlights its effectiveness in terms of precision, recall, and f1-score across various categories. This study not only contributes to the realm of medical image analysis but also lays the groundwork for future investigations and practical implementations focused on refining lung carcinoma diagnosis and treatment using sophisticated machine learning methods. In addition to theoretical advancements, bridging the gap between lung cancer classification and practical application requires clinical validation and future research focus. Clinical validation involves rigorous testing in real-world settings, collaborating with medical professionals to ensure the model's efficacy. Future research should aim to enhance interpretability, integrate multimodal data fusion, and develop transfer learning approaches for im-

proved generalizability. By emphasizing clinical validation and outlining future directions, we can ensure theoretical developments translate effectively into clinical practice, benefiting patients and healthcare providers.

7. Conflicts of Interest

There are no material conflicts of interest that the authors need to report.

Availability of dataset

We employed a pair of publicly accessible datasets, both of which can be accessed online. The initial dataset can be found at:

<https://www.kaggle.com/datasets/adityamahimkar/iqothnccd-lung-cancer-dataset>

while the second dataset is accessible through:
<https://data.mendeley.com/datasets/p2r42nm2ty/1>

Contributions

Ola S. Khedr and Mohamed E. Wahed wrote the main manuscript text and made the practical part, Al-Sayed R. Al-Attar Prepared the Image dataset and E. A. Abdel-Rehim reviewed the manuscript.

References

- [1] S. Ali, S. Asad, Z. Asghar, A. Ali, D. Kim, End-to-End 2D Convolutional Neural Network Architecture for Lung Nodule Identification and Abnormal Detection in Cloud, *Comput. Mater. Contin* 75 (2023) 461–475.
- [2] Z. Ahmed, K. Mohamed, S. Zeeshan, X. Q. Dong, Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine (2020) 2020–2020[link].
URL <https://doi.org/10.1093/database/baaa010>
- [3] A. Tufail, Y. K. Ma, M. K. A. Kaabar, F. Martínez, A. R. Junejo, I. Ullah, R. Khan, Deep Learning in Cancer Diagnosis and Prognosis Prediction: A Minireview on Challenges, Recent Trends, and Future Directions, *Comput. Math. Methods Med* (2021) 2021–2021.
- [4] P. Nanglia, S. Kumar, A. N. Mahajan, P. Singh, D. Rathee, A hybrid algorithm for lung cancer classification using SVM and Neural Networks 335–341 [link].
URL <https://doi.org/10.1016/j.ict.2020.06.007>
- [5] G. Kasinathan, S. Jayakumar, Cloud-Based Lung Tumor Detection and Stage Classification Using Deep Learning Techniques, *Biomed Res. Int* (2022) 2022–2022.
- [6] Z. H. U. Rong, D. A. I. Lingyun, L. I. U. Jinxing, G. U. O. Ying, Diagnostic Classification of Lung Cancer Using Deep Transfer Learning Technology and Multi-Omics Data, *Chinese J. Electron* 30 (2021) 843–852.
- [7] F. Taher, N. Prakash, A. Shaffie, A. Soliman, A. El-Baz, An Overview of Lung Cancer Classification Algorithms and their Performances, *IAENG Int. J. Comput. Sci* 48 (2021).
- [8] A. Elnakib, H. M. Amer, F. E. Z. Abou-Chadi, Early lung cancer detection using deep learning optimization, *Int. J. Online Biomed. Eng* 16 (2020) 82–94.
- [9] W. M. Salama, A. Shokry, M. H. Aly, A generalized framework for lung Cancer classification based on deep generative models, *Multimed. Tools Appl* 81 (2022) 32705–32722.
- [10] F. Kanavati, G. Toyokawa, S. Momosaki, M. Rambeau, Y. Kozuma, F. Shoji, K. Yamazaki, S. Takeo, O. Iizuka, M. Tsuneki, Weakly-supervised learning for lung carcinoma classification using deep learning, *Sci. Rep* 10 (2020) 1–11.
- [11] N. Maleki, CT-Scan images, Mendeley Data (2020) 2013–2013[link].
URL <https://doi.org/10.17632/p2r42nm2ty.1>
- [12] S. Bhanja, A. Das, Impact of Data Normalization on Deep Neural Network for Time Series Forecasting (2018). [link].
URL <http://arxiv.org/abs/1812.05519>
- [13] A. Fernández, S. García, F. Herrera, N. V. Chawla, SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary, *J. Artif. Intell. Res* 61 (2018) 863–905.
- [14] R. R. Picard, K. N. Berk, Data splitting, *Am. Stat* 44 (1990) 140–147.
- [15] S. Albawi, T. A. Mohammed, Understanding of a Convolutional Neural Network, *Icet.* (2017).
- [16] M. Hossin, M. N. B. Sulaiman, A Review On Evaluation Metrics For Data Classification Evaluations, *Int. J. Data Min. Knowl. Manag. Process* 5 (2015) 1–11.