International Journal of Basic and Applied Sciences, 14 (SI-1) (2025) 279-287

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/hvw82e84
Research paper

Lung Cancer Histopathological Image Classification Using Custom CNN Model

Asiya [1] *, Dr. N. Sugitha [2]

[1] *Research Scholar, Department of CSE, Noorul Islam Centre for Higher Education, Tamil Nadu, India*
[2] *Department of ECE, Saveetha Engineering College, Thandalam, Chennai, Tamil Nadu, India*
**Corresponding author E-mail: syedasiya14@gmail.com*

Received: May 28, 2025, Accepted: May 31, 2025, Published: June 7 2025

Abstract

Lung cancer accounts for the highest death rate and disability among harmful tumors. For decades, researchers have focused on screening for lung cancer in an attempt to reduce the high associated mortality rate. Treatment options also have greatly improved in the last few years. An equally important step performed by pathologists is stang, typing and subtyping cancers which involves visual review of histopathology slides. Classifying the two most prevalent types of lung cancer, carcinomas of the squamous cell and adenocarcinoma types, needs a skilled pathologist to physically examine a material. Here, we provide Custom Developed 2D-CNN, a model for a neural network that can classify images of lung histology. Lung as well as colon cancer LC25000 histopathology picture collection and, upon which our research was based, are detailed in conjunction with the model. This dataset contains 5000 digital histopathology images labelled as benign (normal cell), adenocarcinoma (cancer cell) and squamous carcinoma (cancer cell). Various feature extraction techniques were employed to preprocess the images before feeding them to the neural network model. For data sorting, the VGG16 model as well as a customized 2D-CNN model were used. The Custom Developed 2D-CNN attained 95% accuracy. The outcomes generated by the advanced classification approach DL (deep learning)-based, show that it can distinguish between different kinds of lung cancer. Using fewer attributes and lower computational complexity than conventional approaches. Additionally, these results demonstrate that certain transformation methods increase data explanatory power as well as simplify and enhance the diagnostic process.

Keywords: *Lung Cancer Diagnosis; Histopathology Images; Deep Learning; Neural Networks.*

1. Introduction

Cancer develops when damaged DNA causes abnormal cells in different parts of the body to divide uncontrollably. Cancer has the potential to spread to many different organs and tissues. The World Health Organization has predicted that in 2019, [1] [38], In 112 countries, cancer ranks first or second among deaths in persons less than seventy years old. In addition, 134 nations have cancer as either the primary or secondary killer, according to a 2020 study by the International Agency for Research on Cancer (IARC). Over two hundred distinct cancers have been identified so far [3] [39]. Among all cancer diagnoses in the US, the study found that lung and colon cancer patients would have the highest fatality rates in 2020 [5]. The rates of colon cancer are 18.0% and those of lung cancer are 11.4%, based on GLOBOCAN 2020 estimates. In addition, by 2020, the World Health Organization predicts that 4 million individuals may be diagnosed with by lung or colon cancer. These malignancies killed over 2.7 million persons. These statistics show that lung cancer is very common and has a devastating impact over the world. And as we've seen, colon cancer is every bit as fatal as lung cancer [21].
As a result of mutations and unchecked cell proliferation, malignant tumors can form in the lungs. A lot of factors have contributed to the global rise in lung cancer cases, but two of the most important are the aging population and the prevalence of hazardous or dangerous substances [2]. The fact that symptoms won't manifest until the disease has metastasized further complicates treatment. Although non-smokers are not immune to developing lung cancer, smokers have a higher risk [4]. Most lung tumors identified through histology are carcinomas, most often lung cancer, small cell, squamous cell, and big cell types [10]. While current or former smokers are at increased risk for developing adenocarcinoma, the disease can also strike nonsmokers [7] [40] [9]. This forms on the outer lobes of the lungs and is more prevalent in younger individuals and females. It starts attacking the lungs before it spreads. A history of smoking is also associated with squamous cell carcinomas [18]. Therapy becomes more difficult for large cell carcinoma because it grows and spreads at a faster rate than small cell carcinoma, which typically grows slowly but can develop anywhere in the lung [6].
Cancer symptoms are not direct indicators of the disease, although they can help in the early detection process. Exhaustion, coughing, muscular discomfort, etc., are common symptoms that can be linked to numerous disorders. Cancer detection relies heavily on medical imaging technology. Common imaging modalities used for cancer detection include mammography, images obtained through PET scans, also CT, as well as histopathology imaging as well as ultrasonography [41], [11], [42], [13]. Clinicians rely on histopathology photos that include phenotypic information to diagnose and evaluate cancer patients. The sensitive and difficult process of manually analysing such

medical photographs requires experts [12]. Consequently, you'll need to devote a lot of time and focus [14]. Additionally, early diagnosis makes case detection much more challenging due to the symptomatology being very vague and hard help identify when the illness is in its early phases [8] [20]. Once symptoms manifest, early treatment becomes ineffective. Thanks to advancements in AI, medical image analysis methods powered by AI can now assist doctors with early diagnosis and decision-making. The references cited are [43] [15] [44].

Contribution. Making a difference. For a long time, medical issues have been addressed using conventional machine-learning methods and their associated health-related applications. The feature selection and extraction processes are unfortunately integral to these approaches. Consequently, they have problems including data loss during feature extraction and improper method selection. Since medical data are usually radiographic image data, a popular DL architecture for evaluating medical images is the CNN. However, DL ability to eliminate these limitations and provide excellent discrimination has led to its increasing popularity in medical diagnostic applications. Our contribution to this subject of detecting lung cancer using deep learning is to develop novel approaches for curating and classifying high-quality histopathology image datasets, to determine new approaches for addressing imbalanced data, to build new deep learning models that are more interpretable and explainable such that So that the images of different kinds of lung cancer can be organized, a more versatile DL framework is employed which gives a higher accuracy when compared to already existing systems. Another aim is to evolve the approaches for addressing the challenge of variability in staining and preparation techniques of the frameworks where the internal working of the classification system is unclear and to address the challenge of lack of consensus on cancer diagnosis by developing methods for incorporating uncertainty into deep learning models. Apart from this, the significant contribution is to develop new approaches for optimizing deep learning models for clinical use by addressing the challenge of transparency and accountability by developing ethical and regulatory frameworks for using deep learning models in clinical settings.

Organization. Built utilizing a combination of DL models and data processing methodologies, this research proposes a framework for early diagnosis of lung cancer. Lung cancer has impact on men as well as women in an equal manner. Five-class LC25000 dataset is utilized for the experiments; it includes histological pictures of cancers of the colon and the lungs, with latter group of images being utilized only for classification purposes; it comprises three classes. This idea uses both reduced processing cost and increased precision, which is a departure from previous work. In order to train the CNN that will power the classification system, images are pre-processed and feature-extracted to yield vectors or weights that the network can use to distinguish between images and provide optimal accuracy. These are the important points of the remaining parts of the paper: Research in existing domain that has been previously investigated is discussed in Section 2. Part 3 provides a synopsis of the methods and the LC25000 dataset. Section 4 offers a concise overview of CNNs, and Section 5 showcases the CNN algorithm that was specifically designed for our model. Implementation and further explanation of the CNN architecture are provided in Section 6. Everything that came out of the experiments is summed up in Section 7. In Section 8, we provide some suggestions for further research after our trial is over.

2. Literature survey

To make their finding [17] used DL and radiomics together to try to guess EGFR changes in non-small cell lung cancer. From a set of CT scans, radiomic features were taken out using the PyRadiomics package. In our case, we used a CNN model to guess the state of the EGFR mutation. Although no accuracy metrics were reported, it was noted that the radiomic features derived from CT images are useful for predicting EGFR mutations and that DL-based radiomics potential to help fight cancer caused in lungs. In [45] carried out the necessary research. To put lung nodules into groups, the main goal of this work is to use DL methods and CT images. The main thing that this study used was CT scans of lung nodules from the LIDC-IDRI collection. With the aid of a 3D CNN, the authors processed the CT scan data and performed nodule classification [16]. They achieved 89.7% accuracy and 94.4% sensitivity in the way that lung tumors are classified. To improve the way lung cancer is diagnosed and treated, the research proves that DL algorithms can identify but also accurately classify lung nodules at an early stage [22]. In the study by Amjad and colleagues [19] Based on histopathology photos, the scientists tested DL's ability to classify tumors and forecast cancer mutation status.

The researchers made use of a dataset that had been split into two parts: the training as well as validation set. All of the sections included histopathological pictures patients with lung cancer. For the purpose of lung cancer sample classification and mutation status prediction, the scientists implemented a model using a deep CNN. Overall, the study achieved 86.9% for the classification task and 78.6% for the mutation prediction task. The results indicate that deep learning-based mutation prediction and classification from histopathological images could aid significantly when dealing with lung cancer.

Research on cancer caused in lungs, identification combining PET/CT scans with histopathological images is available. The authors designed, validated, and trained their model on a dataset containing lung cancer PET/CT scans alongside histopathological images. Images were classified as benign or malignant based on features derived using CNNs, which were subsequently used for classification.

In order to summarize these approaches, focus on attributing the right ideas to the right authors as in transversal reading. They combined MRI and CT lung images into a single dataset Including features like detailed captions or rigorous proofreading demonstrates that they paid great attention to detail. The model yielded a result where sensitivity was at 94.8 %, specificity stood at 88.2 % and accuracy registered 91.3 % for the overall assessment of finding lung cancer. In MRI, with the purpose of detecting and differentiating lung cancer, as well as CT scans, the authors clearly describe how this model is adjusted. Advanced 3D CNNs were proposed as a method to frame and classify the problem posed by foetal lung MRIs and CTs. Lastly, detail which findings stand out in their results. These findings suggest that the model can timely identify and classify complex pathologies. The research found that 92.5% of the time, lung cancer diagnosis was accurate. This study lends credence to the idea that the multimodal approach based on DL can be useful in the battle against cancer of the lung.

As stated by Zhang and colleagues: For lung cancer diagnosed by CT scans, this research proposes a 3D Fast R-CNN model. The LIDC-IDRI dataset, which includes computed tomography (CT) images of lung nodules, was used for both model training and validation. 3D Faster R-CNN model differentiated between noncancerous and cancerous lung nodules. The model's accuracy rate for detecting lung cancer was 88.8 percent. The outcomes demonstrate that the suggested model is capable of autonomously processing CT scans for the purpose of detecting lung cancer. “This is the work of Francisco and collogues” (22) is left out but consider His remarks steer away from lung cancer detection, and he proposes to rather employ deep learning techniques alongside CT scans. The authors used three sets of CT scans consisting of training, validation, and testing sets, all from patients diagnosed with lung cancer. Using CNN, the authors automated the process of determining whether CT pictures were benign or malignant. The research found a 95.2% success rate in detecting lung cancer. It is prudent to delve further into DL-based lung cancer detection systems, since this type of research implies that they may be effective. The study conducted by Teramoto and colleagues [23].

A plan for using DL methods for automated lung cancer detection in CT scans is proposed in this study's hypothesis. The authors used a dataset with CT scans of people with lung cancer to train their algorithm, and then they tested it on another dataset. Researchers classified

CT scans as either benign or malignant using CNNs. The research showed that lung cancer may be detected with a 92.3% success rate. This lends credence to the idea that DL may help find lung cancer before it spreads.

In [24] suggested a DL method to assess MRI images for signs of lung cancer and to categorize them accordingly. The authors used MRI scans from people with lung cancer to train their algorithm, and then they tested it on another dataset. Authors employed a CNN model to categorize MRI scans as benign else cancerous. Detection accuracy for lung cancer was 85.7 percent. Use of the suggested DL for MRI-based lung cancer diagnosis and evaluation was successful, according to the results. Researchers [25]. This article discusses rheumatoid cancer with an emphasis on trying in order to identify lung cancer utilizing DL in conjunction with CT scans. The authors utilized a total of three sets of computed tomography scans: training, also testing as well as validation. The patients were all afflicted with lung cancer. The authors used a deep CNN that used transfer learning to pull out features that would help them classify the CT scans as either benign or cancerous. Researchers found that lung cancer could be detected 93.9% of the time, with a good level of accuracy.

Our results show that the suggested DL method could be used to find and diagnose lung cancer early on. This is what [26] did. As an example, this piece talks about how deep learning is being used to create programs that can look at CT scans and find lung cancer. After that, these programs can look at the data on their next run. The authors trained their models with CT scans of people who had lung cancer. Then they were split into sets that would be used for testing, sets that would be used for training, and sets that would be used for proof. Using a deep CNN model with transfer learning, the authors split the CT images into two groups based on whether they showed healthy tissue or cancerous tissue. Up to 93.9% of the time, lung cancer can happen, according to researchers. In their experiment, they show that the suggested method using DL could be used to find and identify lung cancer early on. Fig. 1 shows the different ways that lung cancer can be explained that are linked to the table 1.

Table 1: Various Research on Lung Cancer

S. No	Dataset Used	Feature Extraction Technique	Model Used	Results
[27]	NSCLC dataset (The Non-Small Cell Lung Cancer dataset)	DL-based feature extraction using CNNs	Ensemble of Deep Learning models consisting of multiple CNNs	The overall segmentation accuracy was 88.8 percent using the suggested strategy.
[28]	LIDC-IDRI dataset	Transfer Learning-based feature extraction using pre-trained CNN models	Deep Learning-based model consisting of multiple CNNs and a fully connected layer	With 85.7% overall categorization accuracy, the suggested strategy was successful.
[29]	NSCLC dataset	CNN-based feature extraction using deep learning techniques	CNN-based model with multiple hidden layers	An total accuracy of 88.2% was reached by the suggested strategy in terms of classification.
[30]	NSCLC dataset	DL-based feature extraction using CNNs	CNN-based model with saliency maps	Overall, the segmentation accuracy was 88.9% and the classification accuracy was 90.5% using the suggested strategy.
[31]	LIDC-IDRI dataset	Transfer Learning-based feature extraction using pre-trained CNN models	CNN-based model with multiple hidden layers	In terms of overall categorization accuracy, the suggested technique reached 88.5%.
[32]	NSCLC dataset	CNN-based feature extraction using DL techniques	CNN-based model with multiple hidden layers	Overall, the segmentation accuracy was 89.7 percent and the classification accuracy was 91.6 percent using the suggested method.
[33]	LIDC-IDRI dataset	Transfer Learning-based feature extraction using pre-trained CNN models	CNN-based model with multiple hidden layers	The overall categorization accuracy was 88.4 percent using the suggested strategy.
[34]	NSCLC dataset	CNN-based feature extraction using DL techniques	CNN-based model with multiple hidden layers	In terms of overall categorization accuracy, the suggested technique reached 90.5%.
[35]	NSCLC dataset	Statistical feature extraction using the Gray Level Co-occurrence Matrix (GLCM)	Two-stage approach using Support Vector Machines (SVMs)	With 85.3% overall categorization accuracy, the suggested strategy was a success.
[36]	Various datasets used across different studies	A variety of feature extraction techniques were used across studies, including CNN-based and statistical feature extraction methods	A variety of models were used across studies, including CNN-based models and SVMs	Results from a variety of investigations are summarized in the review, which includes accuracy percentages between 76% and 97%.

Challenges

- Data of sufficient quality is hard to come by: amassing a large and diverse collection of high-quality histopathological pictures for rarer subtypes of lung cancer is challenging. As a result, recognition of some subtypes by deep learning algorithms may not be as proficient.
- Difficulty in identifying cancerous regions: Especially for those untrained in the field, the histopathology images can be quite intricate. Locating the malignant portions of the image tends to be difficult and many pathologists may not consistently agree on the interpretation.
- Variability in staining and preparation techniques: Different laboratories as well as different technicians have their own way of preparing and staining the images and histopathology is no exception.
- Time and resource-intensive annotation process: Histopathology image annotation at the level of detail required for detecting lung cancer is a highly specialized task that entails a significant time investment, especially when expert clinicians are needed to provide gold-standard annotations.
- Lack of transparency and accountability: The lack detection technologies of sufficient ethical and regulatory conventions poses difficulties in the guarantee of transparency and accountability during their construction and utilization.
- Issues concerning accessibility to computing resources: Not all experts and doctors have the resources to keep up and run the huge computer models that are needed to diagnose lung cancer. DL methods for finding lung cancer may not be as easy to get or use because of this.

3. Dataset

Following a concise summary of the dataset, all approaches for data pre-treatment are detailed. In this study [37] is consulted. LC25000 Dataset has micrographs of the stomach and lungs. The dataset is split into five separate groups, and each has 5,000 samples: colorectal adenocarcinomas, colon cancer that is not cancerous, lung cancer that is not cancerous, and colon cancer that is not cancerous. However, for the purposes of our categorization, we will only be looking at the lung cancer classes. Images from different lung cancer categories are shown in Fig. 1. Transformed from 1024 × 768 pixels into 768 x 768 pixels squares, the initial dataset merely had 750 lung photos and 500 colon photographs. Lastly, the dataset was expanded to 25000 photos using Augmenter's rotation and flip features.

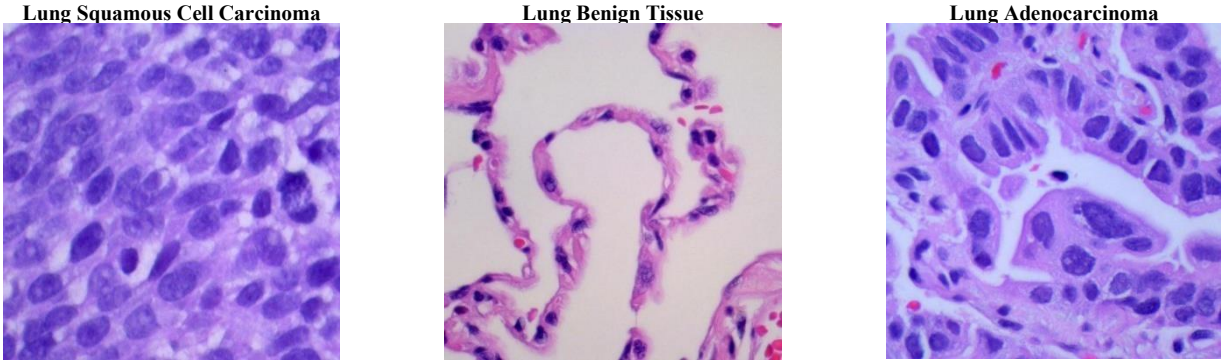

Fig. 1: Lung Cancer Image Samples from the Dataset.

The model is given data that has already been handled in the form of "augmented inputs. "To begin, we split the data for each class into a training set with 4,500 points and a test set with 5,000 points, using the provided random sample method. Another change was the randomization of some of the photographs and the scaling to 150x150 pixels for the others. Following the shear and zoom changes, the images are normalized.

4. Methodology

The three steps in the given approach are: First, histopathology images are pre-processed. Then, features are extracted and a model is trained. Finally, classification is performed. Initially, the sizes of the histopathological images are analyzed, and then these images are augmented to balance the samples. The image data generator is then used to divide the photos into sets, and the size is set. Lastly, the picture data generator with the DL model is used to pull out in-depth features. In the end, these traits are used to train different deep learning models, such as CNN. As shown in Fig.2, the suggested method has three steps.

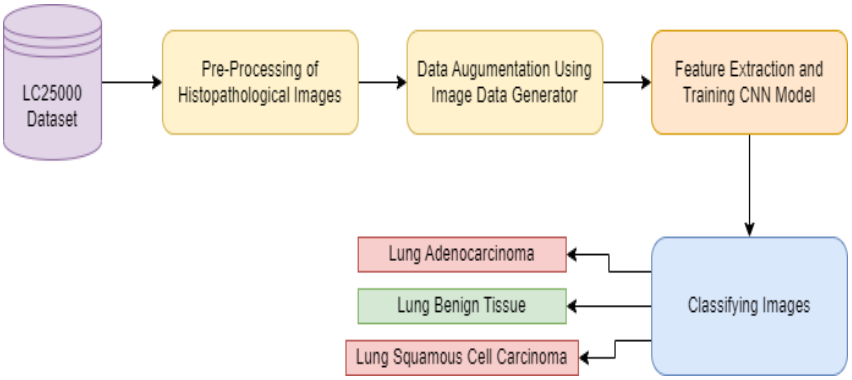

Fig. 2: Stages of the Proposed Approach.

In order to initiate the training process, a certain pixel size is sent into the input layer of every CNN. Consequently, the size of the input layers in the DL models is first adjusted to fit the histological dimensions of lung as well as colon cancer images. For better training results from deep learning models with less overfitting, an augmentation strategy is obviously necessary. One other benefit of augmentation is that Increases the number of training pictures in a dataset, which helps training models learn better. This research uses several methods to improve things, including x and y choice scaling, flipping x and y orientation, translating x and y orientation by an angle range, and shearing x and y direction within a range. In this case, an Image data creator is used to get the features of the picture. Every so often, this type of generator takes in the original data, changes it randomly, and gives back the result that only has the changed data. There is no info. You can also use the Keras Image Data Generator class to give the model more data. A picture data generator can flip data horizontally, change its size, shear it, translate it, rotate it, and do other odd things with it. The Keras image data maker makes groups of data from tensor pictures that can be added to real-time data sets. We can look through the data in groups with Keras' picture data creator. There are several methods and inputs in the picture data generator class that let you change how the data is made. Now, get data from the pictures with the picture data maker. What the CNN DL model does when it gets the data is usually change the picture data so that the enhanced image is in the form of a matrix. In the end, this data matrix picture tells the DL models what to do.

5. Custom CNN algorithm

Think about a few settings to get CNN model to segregate lung cancer pictures. Following an explanation of how CNNs function, we will apply the specific implementation to the lung cancer dataset. To begin with, let us consider the raw image dataset of lung cancer images as which consists of three classes which are determined as A = {A1, A2, A3}. This dataset is first pre-processed, and the images present in it are sorted according to the classes such that the images belonging to the same class are together. Next, using the Image Data Generator from the Keras library, the images are scaled down and re-sampled using various parameters such as shear range, zoom range, and rotation range. Finally, before generating the data, the data generator will split it into two sets: training and validation. A number of variables, including the size of the target, the batch, the color mode, regulate this procedure. In this case, we will refer to both the training dataset as A' and the validation dataset as A". The Figure and table below represent the same, where Fig. 3 represents the dataset and its classification, and Table 2 represents the algorithm of ImageDataGenerator for the lung cancer dataset.

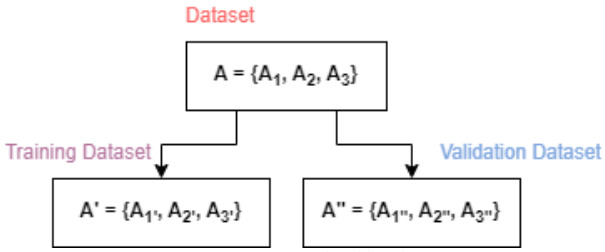

Fig. 3: Dataset Classification.

Table 2: Image Augmentation Algorithm

Algorithm 1 - Image Data Generator for raw Lung cancer images
1: INPUT: A = {A_1, A_2, A_3}
2: repeat
3: repeat
4: Obtain a single image from all the images of class A_
5: rescale image (1./255)
6: Augment the image (150 x 150)
7: until: All images in class A_
8: until: All classes in A i.e., A = {A_1, A_2, A_3}
9: OUTPUT: Augmented Images ~A

The above method might improve raw images by working on both the training set and the validation set of data from the lung cancer classes raw image dataset.
The next step is to use the image data generator to resize and add to the images. Then, the images are fed into the DL model, which is CNN. The images are in the form of a matrix, so the conversion happens in numbers. Each number represents a pixel's value, and since the image is in RGB format, each pixel's channel value will always be three. There is an added picture that is 150 x 150 pixels. This image is fed into the neural network's convolutional layers. The pre-processed and enhanced image data will be called ~A' for the training dataset and ~A" for the validation dataset. The convolutional matrix will be called W' and W", and the features extracted from the images will be called PW' and PW" for the training dataset and validation dataset, respectively. The process for getting the features out of photos using convolutional layers is shown in Table 3.

Table 3: Feature Extraction Algorithm

Algorithm 2–Feature Extraction with convolutional layers
1: INPUT: ~A' = {~$A_{1'}$,~ $A_{2'}$,~ $A_{3'}$} or ~A" = {~$A_{1"}$,~ $A_{2"}$,~ $A_{3"}$}
2:Initialize the convolutional kernel matrices W' and W'' with random values
3: repeat
4: Obtain a single image from all the images of ~A' or ~A"
5: Apply convolutional operation (f', k'), (f", k'')
6: Apply Max pooling operation (p)
7: Update the W' and W" matrix with the updated weights after the above two operations.
8: until: All images in ~A' or ~A"
9: OUTPUT: Updated Weight matrices i.e., PW' and PW"

It can be used on both training data and validation data that are separate from the augmented pictures of lung cancer classes obtaining the extracted features in the matrices; these matrices are further minimized by applying more convolutional and max pooling operations by using the same algorithm hence, obtaining the final weight matrices which are flattened down for the classification purpose. The equation for convolutional and pooling operations for the lung cancer images is determined below:

$$PW'[m, n] = (\sim A' * k')[m, n] = \sum_j \sum_k k'[j, k] \sim A'[m - j, n - k]$$

$$PW''[m, n] = (\sim A'' * k'')[m, n] = \sum_j \sum_k k''[j, k] \sim A''[m - j, n - k]$$

In the above equations, PW' determines the weight matrices of the training data, whereas PW" determines the weight matrices of validation data provided to the dense network for classification. The other parameters, such as [m, n], determine the pixel range in an image to which the convolutional and pooling operations use ~A' and ~A" (images) and k,' and k' (starting kernel matrices) to get the feature map matrices. After obtaining the weight matrices by applying the convolutional and pooling operations on each image, the matrices are fed to the dense layer, which performs the lung cancer classification. In the next step, A dense neural network is trained using the classified characteristics, which comprises neurons that help classify the images according to the classes specified. Let us consider the initial input to the dense layers

as the transformed version PW', which is denoted as ~PW'; These are the matrix values that have been smoothed; they are the input neurons for the dense layer. The dense layer is made up of weights. which are randomly initialized and updated accordingly by forward and backward propagation in numerous iterations, let us consider this randomly initialized weights as S'. These weights are forwarded to the following dense layer and updated. Finally, the weights determine each class's probability from where the class's maximum probability is chosen to classify that image. The algorithm for this process is specified in Table 4.

Table 4: Training and Classification Algorithm

Algorithm 3–Training and classification with dense layers
1: INPUT: PW' or PW"
2: Transforming the matrices i.e., Flattening them into one-dimensional Neurons
3: Initializing the weight matrices randomly for the dense layers S'
4: repeat
5: Apply X→ (DL1)
6: ApplyY→ (DL2)
7: ApplyY←, X← and update values(~DL2 & ~DL1)
8: Update the S' and S" weights
9: until: Iteration Convergence
10: OUTPUT: Trained model (M) and final updated weights (~S')

The above algorithm works for training data which gets transformed into an input layer for the dense layers. The training process occurs accordingly; To ensure that the model's training was successful, iteratively classifying the values from the validation set is employed. The output of the above algorithm will give the thoroughly trained model M and the final updated weights ~S'. These weights determine the probability of each class when an input image is given.

6. Implementation

Classifying images, especially microscopic ones like those used in histopathology, is challenging because of the high convolution of inter-intraclass interactions. The existence of similar morphological and structural textures indicates that the underlying structures are complex and interconnected. The histopathological images displayed in Fig. 1 showcase some of the most complex textures ever seen. The ability of DL to learn features immediately from input is one reason for its popularity. Eliminating the need for laborious feature extraction methods. Discovering abstract-level features and then delving down into the feature map DL has many benefits, one of which is the ability to extract structural semantics. Classification and diagnosis of medical images have been greatly improved in recent years by using DL and, more specifically, CNN. CNN may learn feature mappings from an input data flow by combining a supervised classifier with its many trainable layers. One type of input data stream is signal data, which includes things like audio, video, images, and time series. Think about a coloured picture, a 3D tensor feature map where each colour channel is a 2D tensor. The convolution, max pooling, and dense or fully linked layers make up a CNN. There is more than one way to construct these layers in order to build a CNN. In Fig. 4, we can see a CNN in action.

CNN design would be incomplete without the convolution layer. Convolution layers take the input signal as well as weights and use a dot product to find the output for each area. The input vector is twisted along a set of weights called the kernel or filters. Although tiny, each filter accounts for the full range of the input volume. Three by three, five by five, and eight by eight are the usual dimensions of a filter for visual inputs. All of the geometrical structures in a picture can be learned by filters because these weights are shared across neurons. The stride measures the separation between the uses of these filters. The image is subjected to overlapping convolutions when the hyper parameter stride surpasses the filter size.

In order to down sample the image along the volume component, it is common practice to insert a pooling between two convolution layers. In order to shrink the representation's physical footprint over time, this is essential. So, to manage over-fitting, reduce the amount of parameters and calculations the network needs. Bypassing activation, pooling resizes images along height and width. The max pooling method outperformed its competitors, according to the researchers. It gives you a window to pick the neighbor with the highest number from an input patch. An activated layer that is eventually linked or dense forms a full connection between activations of inputs and their activations. After then, sequential bias offset and matrix multiplication are used in the calculation. Data categorization results, probability densities, or logit values are stored in the last dense layer.

6.1. Custom developed CNN model for lung cancer detection

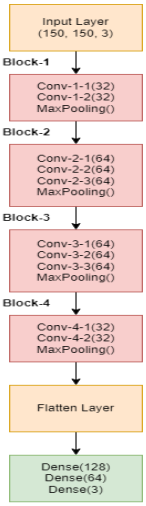

Fig. 4: Lung Cancer Image Classification Architecture.

7. Result analysis

Fig. 5, 6, 8, 9, 10, and 11 show the results for both the VGG16 model and the custom-developed CNN model, which were used for the analysis VGG16 Model. Comparison of Standard Model and Proposed Model shown in Table 5.

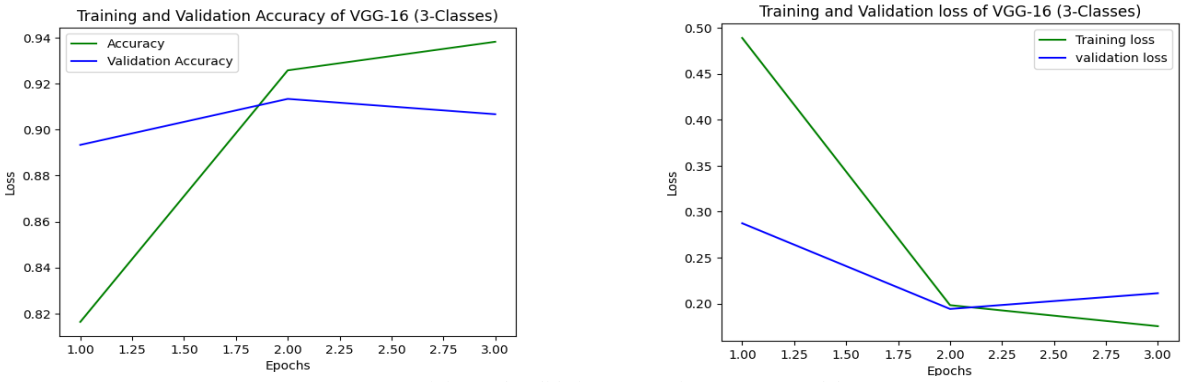

Fig. 5: Training and Validation Curves for VGG-16 Model.

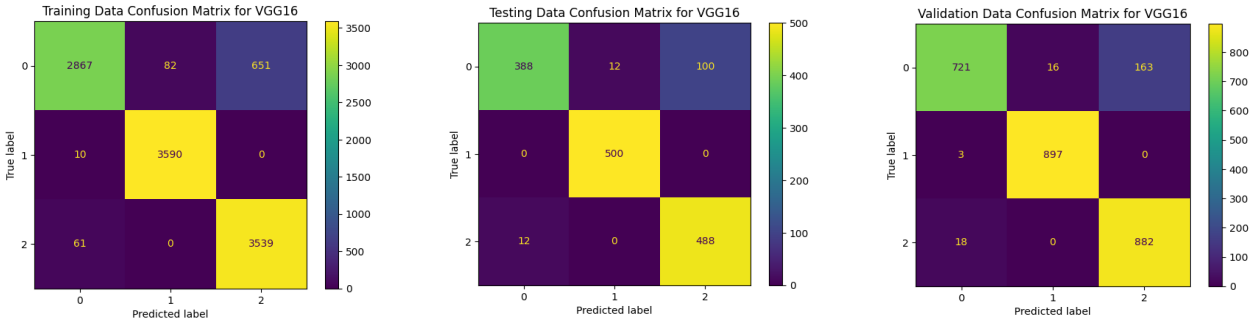

Fig. 6: Confusion Matrices for VGG-16 Model.

Table 5: Comparison of Standard Model and Proposed Model

	VGG16	Custom CNN
Training Accuracy	0.9255555	0.95916666
Testing Accuracy	0.9173333	0.952
Validation Accuracy	0.92592592	0.95592

To categorize lung cancer images, the VGG-16 model was used, and the outcomes are shown in the figures above. For a total of twenty iterations, the VGG-16 model got an accuracy of six throughout training and validation and the Fig. 7 represents the data from training, also testing, as well as validation that is confused. Last but not least, the precision of the sets is shown in Fig. 8. In general, this model achieves an accuracy of 92%.

Custom - CNN Model:

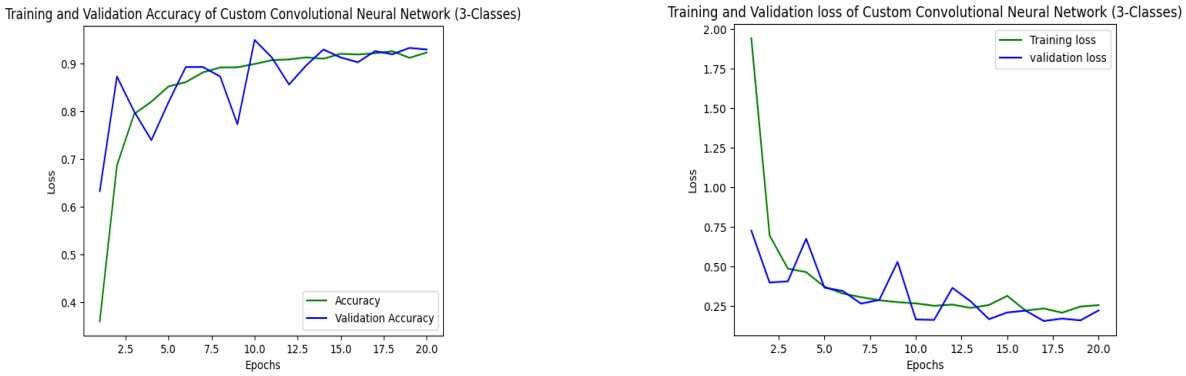

Fig. 9: Training and Validation Curves for Custom-CNN Model.

Fig. 10: Confusion Matrices for Custom-CNN Model.

Fig. 9 and 10 show the data that was used for training, testing, and validation, along with their own confusion matrix. These figures also show the results of the Custom-CNN model's ability to classify images of lung cancer. Last but not least, Fig. 11 shows how true the sets are. This model is right 96% of the time, on average. When it comes to the Lung cancer histopathology image collection, the Custom built CNN model is more accurate than the VGG-16 model. We can see this by comparing the two models. The Custom developed-CNN model, on the other hand, has better classification measures, and there isn't much difference in accuracy between the two.

8. Conclusion

The LC25000 Lung Cancer Image dataset was the subject of a series of deep learning studies detailed in this article. Using a CNN architecture originally developed for object color picture recognition, we successfully adapted it to classify lung histopathology images. We were also able to handle high-resolution textured images without downsampling them by suggesting a training and evaluation method for training the CNN architecture. We compared custom-built CNNs to deep CNN models that used transfer learning and conventional ML models that used traditional texture descriptors on the LC25000 Lung Cancer Images dataset. The results showed that custom-developed CNNs performed better. Hyperparameter tuning and alternative CNN designs could be the subject of future studies. The methods for creating interclass images for different types utilizing neural style transfer are also presented. To further aid in the visualization and comprehension of mutations across several ontologies, generative models could be employed to generate histopathological images.

References

[1] Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, *71*(3), 209-249. https://doi.org/10.3322/caac.21660.

[2] Ziwei, M., & Han, L. L. (2023). Scientometric Review of Sustainable Land Use and Management Research. *Aquatic Ecosystems and Environmental Frontiers*, *1*(1), 21-24.

[3] International Agency for Research on Cancer. https://www.iarc.who.int.

[4] Shimazu, S. (2024). Intelligent, Sustainable Supply Chain Management: A Configurational Strategy to Improve Ecological Sustainability through Digitization. *Global Perspectives in Management, 2*(3), 44-53.

[5] Siegel, R. L., Miller, K. D., & Jemal, A. (2018). Cancer statistics, 2018. *CA: a cancer journal for clinicians*, *68*(1), 7-30. https://doi.org/10.3322/caac.21442.

[6] Dhamala, K. (2024). Pharmacist-Delivered Interventions on Pain Management: Review and Cluster-Randomized Trial. *Clinical Journal for Medicine, Health and Pharmacy*, *2*(4), 11-20.

[7] Asuntha, A., & Srinivasan, A. (2020). Deep learning for lung Cancer detection and classification. *Multimedia Tools and Applications*, *79*(11), 7731-7762. https://doi.org/10.1007/s11042-019-08394-3.

[8] World Health Organization (WHO). Cancer. https://www.who.int/news-room/fact-sheets/detail/cancer.

[9] Tekade, R., & Rajeswari, K. (2018, August). Lung cancer detection and classification using deep learning. In *2018 fourth international conference on computing communication control and automation (ICCUBEA)* (pp. 1-5). IEEE. https://doi.org/10.1109/ICCUBEA.2018.8697352.

[10] Aghababaei, F., Jouki, M., & Mooraki, N. (2024). Evaluating the quality of fried whiteleg shrimp (Litopenaeus vannamei) fillets coated with quince seed gum containing encapsulated cinnamon extract. *International Journal of Aquatic Research and Environmental Studies, 4*(2), 99-115. https://doi.org/10.70102/IJARES/V4I2/7.

[11] Shakeel, P. M., Burhanuddin, M. A., & Desa, M. I. (2019). Lung cancer detection from CT image using improved profuse clustering and deep learning instantaneously trained neural networks. *Measurement*, *145*, 702-712. https://doi.org/10.1016/j.measurement.2019.05.027.

[12] Palash, P. S., & Dhurvey, P. (2024). Analysis of Flyash Aggregate Behavior in Geopolymer Concrete Beams Using Method of Initial Functions (Mathematical Programming). *Archives for Technical Sciences, 2*(31), 168–174. https://doi.org/10.70102/afts.2024.1631.168.

[13] Pradhan, K., & Chawla, P. (2020). Medical Internet of things using machine learning algorithms for lung cancer detection. *Journal of Management Analytics*, *7*(4), 591-623. https://doi.org/10.1080/23270012.2020.1811789.

[14] Mohandas, R., Veena, S., Kirubasri, G., Thusnavis Bella Mary, I.., & Udayakumar, R. (2024). Federated Learning with Homomorphic Encryption for Ensuring Privacy in Medical Data. *Indian Journal of Information Sources and Services, 14*(2), 17–23. https://doi.org/10.51983/ijiss-2024.14.2.03.

[15] Sori, W. J., Feng, J., Godana, A. W., Liu, S., & Gelmecha, D. J. (2021). DFD-Net: lung cancer detection from denoised CT scan image using deep learning. *Frontiers of Computer Science*, *15*, 1-13. https://doi.org/10.1007/s11704-020-9050-z.

[16] Zhang, J., & Song, X. (2024). The AI-assisted Traditional Design Methods for the Construction Sustainability: A Case Study of the Lisu Ethnic Minority Village. *Natural and Engineering Sciences, 9*(2), 213-233. https://doi.org/10.28978/nesciences.1569562.

[17] Wang, G., Wang, B., Wang, Z., Li, W., Xiu, J., Liu, Z., & Han, M. (2021). Radiomics signature of brain metastasis: prediction of EGFR mutation status. *European radiology*, *31*, 4538-4547. https://doi.org/10.1007/s00330-020-07614-x.

[18] Liu, D., Camp, L.J., Wang, X., & Wang, L. (2010). Using Budget-Based Access Control to Manage Operational Risks Caused by Insiders. *Journal of Wireless Mobile Networks, Ubiquitous Computing and Dependable Applications, 1*(1), 29-45.

[19] Rehman, A., Kashif, M., Abunadi, I., & Ayesha, N. (2021, April). Lung cancer detection and classification from chest CT scans using machine learning techniques. In *2021 1st international conference on artificial intelligence and data analytics (CAIDA)* (pp. 101-104). IEEE. https://doi.org/10.1109/CAIDA51941.2021.9425269.

[20] Gu, Y., Chi, J., Liu, J., Yang, L., Zhang, B., Yu, D., & Lu, X. (2021). A survey of computer-aided diagnosis of lung nodules from CT scans using deep learning. *Computers in biology and medicine*, *137*, 104806. https://doi.org/10.1016/j.compbiomed.2021.104806.

[21] Zhang, G., Jiang, S., Yang, Z., Gong, L., Ma, X., Zhou, Z., & Liu, Q. (2018). Automatic nodule detection for lung cancer in CT images: A review. *Computers in biology and medicine*, *103*, 287-300. https://doi.org/10.1016/j.compbiomed.2018.10.033.

[22] Carrillo-Perez, F., Morales, J. C., Castillo-Secilla, D., Gevaert, O., Rojas, I., & Herrera, L. J. (2022). Machine-learning-based late fusion on multi-omics and multi-scale data for non-small-cell lung cancer diagnosis. *Journal of Personalized Medicine*, *12*(4), 601. https://doi.org/10.3390/jpm12040601.

[23] Teramoto, A., Tsukamoto, T., Kiriyama, Y., & Fujita, H. (2017). Automated classification of lung cancer types from cytological images using deep convolutional neural networks. *BioMed research international*, *2017*(1), 4067832. https://doi.org/10.1155/2017/4067832.

[24] Sang, J., Alam, M. S., & Xiang, H. (2019, May). Automated detection and classification for early stage lung cancer on CT images using deep learning. In *Pattern recognition and tracking XXX* (Vol. 10995, pp. 200-207). SPIE. https://doi.org/10.1117/12.2520333.

[25] Ibrahim, A. O., Shamsuddin, S. M., Abraham, A., & Qasem, S. N. (2019). Adaptive memetic method of multi-objective genetic evolutionary algorithm for backpropagation neural network. *Neural Computing and Applications*, *31*(9), 4945-4962. https://doi.org/10.1007/s00521-018-03990-0.

[26] Rahman, T., Chowdhury, M. E., Khandakar, A., Islam, K. R., Islam, K. F., Mahbub, Z. B., & Kashem, S. (2020). Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray. *Applied Sciences*, *10*(9), 3233. https://doi.org/10.3390/app10093233.

[27] El-Ghany, S. A., Azad, M., & Elmogy, M. (2023). Robustness fine-tuning deep learning model for cancers diagnosis based on histopathology image analysis. *Diagnostics*, *13*(4), 699. https://doi.org/10.3390/diagnostics13040699.

[28] Bębas, E., Borowska, M., Derlatka, M., Oczeretko, E., Hładuński, M., Szumowski, P., & Mojsak, M. (2021). Machine-learning-based classification of the histological subtype of non-small-cell lung cancer using MRI texture analysis. *Biomedical Signal Processing and Control*, *66*, 102446. https://doi.org/10.1016/j.bspc.2021.102446.

[29] Chaunzwa, T. L., Hosny, A., Xu, Y., Shafer, A., Diao, N., Lanuti, M., & Aerts, H. J. (2021). Deep learning classification of lung cancer histology using CT images. *Scientific reports*, *11*(1), 1-12. https://doi.org/10.1038/s41598-021-84630-x.

[30] Chowkkar, M. (2020). *Breast Cancer Detection from Histopathological Images using Deep Learning and Transfer Learning* (Doctoral dissertation, Dublin, National College of Ireland).

[31] Talo, M. (2019). Automated classification of histopathology images using transfer learning. *Artificial intelligence in medicine*, *101*, 101743. https://doi.org/10.1016/j.artmed.2019.101743.

[32] Sivaranjith, C., & Subramani, M. (2013). Development of reversible programmable gate array. International Journal of Communication and Computer Technologies, 1(2), 72-78. https://doi.org/10.31838/ijccts/01.02.01.

[33] Surendar, A. (2024). Emerging trends in renewable energy technologies: An in-depth analysis. Innovative Reviews in Engineering and Science, 1(1), 6-10. https://doi.org/10.31838/INES/01.01.02.

[34] Hatuwal, B. K., & Thapa, H. C. (2020). Lung cancer detection using convolutional neural network on histopathological images. *International Journal of Computer Trends and Technology, 68*(10), 21-24. https://doi.org/10.14445/22312803/IJCTT-V68I10P104.

[35] Liu, Y., Li, A., Liu, J., Meng, G., & Wang, M. (2021). TSDLPP: a novel two-stage deep learning framework for prognosis prediction based on whole slide histopathological images. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *19*(4), 2523-2532. https://doi.org/10.1109/TCBB.2021.3080295.

[36] Civit-Masot, J., Bañuls-Beaterio, A., Domínguez-Morales, M., Rivas-Perez, M., Muñoz-Saavedra, L., & Corral, J. M. R. (2022). Non-small cell lung cancer diagnosis aid with histopathological images using Explainable Deep Learning techniques. *Computer Methods and Programs in Biomedicine*, *226*, 107108. https://doi.org/10.1016/j.cmpb.2022.107108.

[37] Borkowski, A. A., Bui, M. M., Thomas, L. B., Wilson, C. P., DeLand, L. A., & Mastorides, S. M. (2021). Lc25000 lung and colon histopathological image dataset. *arXiv*.

[38] World Health Organization, D. M. (2019, October). *Global Health Estimates: Life expectancy and leading causes of death and disability*.

[39] Yadav, A. R., & Mohite, S. K. (2020). Cancer-A silent killer: An overview. *Asian Journal of Pharmaceutical Research*, *10*(3), 213-216. https://doi.org/10.5958/2231-5691.2020.00036.2.

[40] Bhatia, S., Sinha, Y., & Goel, L. (2019). Lung cancer detection: a deep learning approach. In *Soft Computing for Problem Solving: SocProS 2017, Volume 2* (pp. 699-705). Springer Singapore. https://doi.org/10.1007/978-981-13-1595-4_55.

[41] Radhika, P. R., Nair, R. A., & Veena, G. (2019, February). A comparative study of lung cancer detection using machine learning algorithms. In *2019 IEEE international conference on electrical, computer and communication technologies (ICECCT)* (pp. 1-4). IEEE. https://doi.org/10.1109/ICECCT.2019.8869001.

[42] Rahane, W., Dalvi, H., Magar, Y., Kalane, A., & Jondhale, S. (2018, March). Lung cancer detection using image processing and machine learning healthcare. In *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)* (pp. 1-5). IEEE. https://doi.org/10.1109/ICCTCT.2018.8551008.

[43] Das, S., & Majumder, S. (2020, November). Lung cancer detection using deep learning network: A comparative analysis. In *2020 Fifth international conference on research in computational intelligence and communication networks (ICRCICN)* (pp. 30-35). IEEE. https://doi.org/10.1109/ICRCICN50933.2020.9296197.

[44] Jakimovski, G., & Davcev, D. (2018, September). Lung cancer medical image recognition using Deep Neural Networks. In *2018 Thirteenth International Conference on Digital Information Management (ICDIM)* (pp. 1-5). IEEE. https://doi.org/10.1109/ICDIM.2018.8847136.

[45] Chen, L., Gu, D., Chen, Y., Shao, Y., Cao, X., Liu, G., & Shen, D. (2021). An artificial-intelligence lung imaging analysis system (ALIAS) for population-based nodule computing in CT scans. *Computerized medical imaging and graphics*, *89*, 101899. https://doi.org/10.1016/j.compmedimag.2021.101899.

[46] Toha, A., Ahmad, H., & Lee, X. (2025). IoT-based embedded systems for precision agriculture: Design and implementation. SCCTS Journal of Embedded Systems Design and Applications, 2(2), 21–29.

[47] Muralidharan, J. (2024). Advancements in 5G technology: Challenges and opportunities in communication networks. Progress in Electronics and Communication Engineering, 1(1), 1–6. https://doi.org/10.31838/ECE/01.01.01.

[48] Velliangiri, A. (2025). Bioenergy from Agricultural Waste: Optimizing Biomass Supply Chains for Rural Electrification. National Journal of Renewable Energy Systems and Innovation, 18-26.