ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR

CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.04 (2025)

LUNG CANCER DETECTION AND CLASSIFICATION USING DEEP LEARNING: A VGG-16 BASED APPROACH

Umair Hussain[1,*] Gohar Mumtaz[1], Imran Siddiq[2] , Kashif Ali[3]
[1] Faculty of Computer Science and Information Technology,The Superior University, Lahore, Pakistan gohar.m@superior.edu.pk
[2] Department of Computer Engineering,The University of Lahore,Pakistan
imran.siddiq@dce.uol.edu.pk
[3] kashifaliakbarali8@gmail.com
[*] Corresponding Author:Umair Hussain SU92-MSCSW-F23-002@superior.edu.pk

ABSTRACT

Lung cancer is still one of the most common causes of cancer-related fatalities around the world. This is mostly because it is hard to identify early and the symptoms are hard to understand. Early detection and precise classification of lung cancer can greatly enhance patient outcomes by facilitating prompt and tailored therapy approaches.With an emphasis on computed tomography (CT) scans, this thesis provides a thorough education on the application of deep learning algorithms for the automated categorization and detection of lung cancer from medical imaging data.

The proposed methodology involves the development and training of CNN architectures tailored for image-based analysis. A carefully curated dataset of annotated lung CT images was employed to ensure model robustness and generalizability. To improve image quality and extract important characteristics, preprocessing methods such segmentation, noise reduction, and normalization were used. Various CNN models were implemented and evaluated, including transfer learning approaches using pre-trained networks to enhance performance in scenarios with limited medical data.

Efficiency was assessed using standard metrics like accuracy, recall, precision, F1-score, and area under the receiver's operating characteristic curve (AUC-ROC). The results show that the optimized deep learning model achieves a high degree of accuracy in both recognizing the stages of cancer progression and differentiating between benign and malignant nodules. The results confirm deep learning's potential as a trustworthy tool for helping radiologists make clinical decisions.

This study emphasizes the significance of incorporating advanced computational methods with medical imaging for improving diagnostic accuracy and decreasing the load on healthcare systems. The study also highlights challenges such as dataset imbalance, model interpretability, and the need for cross-institutional validation, offering recommendations for future work in this critical area of medical diagnostics.

Introduction

Worldwide, non-communicable diseases (NCDs) are regarded as the leading cause of mortality. According to Caprara (2021), the NCDs encompass a variety of illnesses, including cancer, heart disease, chronic respiratory conditions, and type 2 diabetes. Furthermore, according to estimations from the World Health Organization (WHO), NCDs will account for 71% of global fatalities and 80% of the global sickness problem in 2020 (Wang & Wang, 2020; Bigna & Noubiap, 2019). Furthermore, Wang & Wang (2020) predict that within the next ten years, the worldwide burden of NCDs would rise by 17%. An epidemiological shift from infectious to no communicable diseases was recently discovered by public health investigations. Globally, cancer has the highest rates of morbidity and death among NCDs. Globally, both in industrialized and emerging nations, the incidence and spread of cancer are increasing (Olsen, 2015). A serious public health concern that is spreading throughout the world is cancer. It is a disorder that causes uncontrollable cell and tissue division. Consequently, this results in the formation of a tumor or malignancy (Woodman et al., 2021). According to GLOBOCAN, there were 19.3 million new cases of cancer in 2020, which resulted in almost 10 million deaths worldwide. Sung and associates (2021). Lung cancer is the most common cause of death worldwide for mutually Gender and is one of the most often diagnosed cancers. Around 1.8 million citizens will die from lung cancer worldwide as a result of the 2.2 million new instance that have been detected. Mapanga and colleagues (2021); Sung and colleagues (2021). Among

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.04 (2025)

the many typical indicators and symptoms of lung cancer are fatigue, weight loss, and hemoptysis, or coughing up blood. Lung cancer is also linked to a number of risk factors, such as diet, alcohol, smoking, and air quality. Li and associates (2021). The histology of the cancer cells can differentiate between two forms of lung cancer: small-cell lung cancer (SCLC) and non-small lung cancer (NSCLC) (Woodman et al., 2021). NSCLC is believed to be the most common kind of lung cancer, accounting for 85% of instances, as opposed to SCLC, which makes for 15% of all patients (Woodman et al., 2021).

In third-world countries like Sub-Saharan Africa, where HIV has also had a disastrous effect, lung cancer has sharply increased, according to Shankar et al. (2019). Compared to other malignancies including colon cancer (65%), breast cancer (90%), and prostate cancer (99%), the overall five-year survival rate for all categories of lung cancer is lower (18%) (Woodman et al., 2021). However, in order to create innovative strategies that boost early diagnosis, assist in medical decision-making, and evaluate consequence to improve healthcare, lung cancer needs more attention from the scientific, biological, and medical fields. Identification of driver mutations and customized treatments that may be used for certain genotypes are made possible by the molecular profile of tumor tissues. Conventional chemotherapy destroys every cell without differentiating between healthy and cancerous cells. However, by focusing on particular areas and interacting with cancer driver genes, tailored treatment might stop or lessen malignant transformation. Pereira and associates (2021). A significant portion of cancer cases that are curable in wealthy countries are found in third-world countries after they have reached incurable stages due to delayed or incorrect diagnosis (Organization, 2002). For better early diagnosis and the implementation of suitable treatment plans, this has prompted scientists to assess current methods and suggest novel ways to categorize and identify lung cancer and its subtypes.

Motivation

The utility of deep learning extends beyond image classification; recent research highlights its importance in enhancing IoT security protocols (Khan et al., 2023), improving the overall landscape of AI in cybersecurity (Khan et al., 2024), and applying neural networks for anomaly detection in sensitive IoHT environments (Arif et al., 2023).Researchers and healthcare professionals have paid little attention to cancer, despite the continent's higher death rate. Additionally, 57% of all new cancer occurrences globally occur in low-income countries, a number that has grown due to a lack of knowledge, a lack of preventative measures, and longer life expectancies. Hamdi and associates (2021). Reported by Mapanga et al. (2021), lung cancer is the most prevalent kind of cancer and is associated with significant ratio of morbidity and death, particularly if discovered at an advanced stage.

A three-dimensional image of the chest may be created by using computed tomography to take pictures of the lungs in different dimensions.

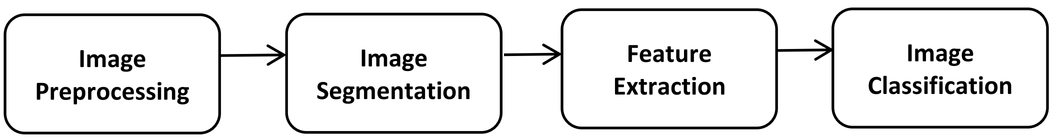


Figure 1: Computer-Aided Diagnosis (CAD)

Lung Nodule Classification and Diagnosis

Picture quality (such as signal to noise ratio and resolution) alter based on the CT scanner and other real-world gaining settings, even with the use of proven screening and diagnostic methodologies. Many outdated methods (such support vector machines) primarily depend on the feature manufacturing, which comprises data pretreatment, modification, and manually designed feature designs, to classify discriminative features. Lung nodule representations are frequently extracted using intensity, morphological, and texture data.

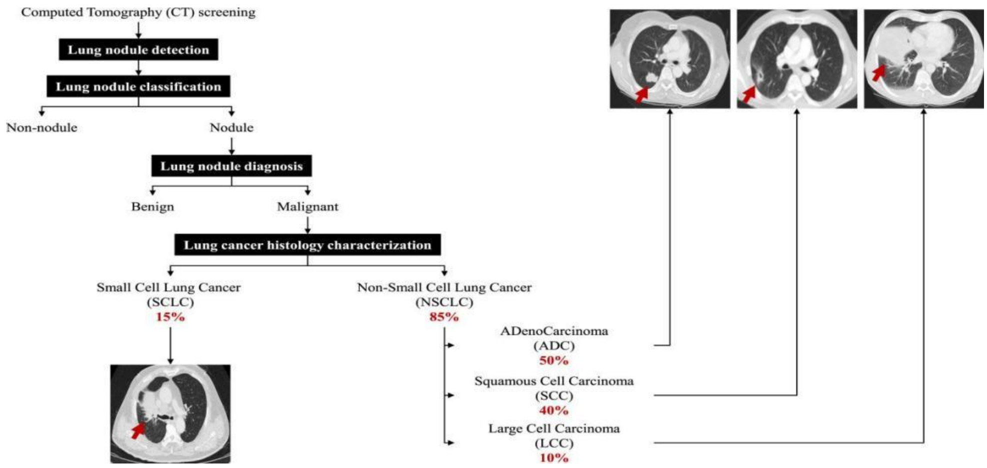


Figure: 2. Non-invasive lung cancer computer-aided diagnosis [10]
Source:https://www.sciencedirect.com/science/article/pii/S0010482522004747#:~:text=After%20being%20detected%2C%20lung%20nodules,85%25%20of%20all%20lung%20cancers.

Cancer Progression Estimation

The speed at which lung cancer may be successfully detected by imaging is determined by the "Mean Soujourn Time" (MST), which gauges how rapidly the disease growths from a pre-clinical to an observable clinical stage. Consequently, MST is widely used [15] in population screening to assess the level of over diagnosis and establish the optimal interval between screenings. Patients having greater MSTs (i.e., at reduced risk of cancer) ought to have lengthier screening intermissions since MST varies based on imaging modalities and patient cohorts [16]. Second, in clinical practice, missing or insufficient observations are frequent. For example, some patients could travel to another hospital where data is not shared, or they might miss a planned screening exam. Third, there is often an uneven gap between screening tests [17]. Fourth, it might be challenging to estimate some observed illness states because of their limited sample size. For instance, if an early-stage malignancy is found, patients are often removed from further monitoring and receive an intervention [18].

a) Image Pre-Processing

CT scans cannot be used directly in a CAD framework. Before being used, they Have to be thoroughly pre-processed. To eliminate noise and prepare photos for usage, a variety of pre-processing techniques are applied. This contributes to improving the accuracy and overall system performance.

b) Image Segmentation

picture division is the process of dividing one picture into many parts. Finding boundaries in a picture is the main purpose of image segmentation. As segmentation lowers the picture complexity, the analysis procedure gets simpler [19].

c) Feature Extraction

In order to facilitate processing and classify our raw data into manageable classes, feature extraction aims to minimize the number of dimensions in our data. Numerous variables that demand computer resources to process and provide conclusions are characteristics of massive amounts of data. Techniques for feature extraction focus on making the data simpler without sacrificing any information. These methods are in charge of selecting and combining the characteristics in order to reduce the volume of data [20].

d) Image Classification

Image classification is a fundamental activity that seeks to understand a picture as a whole. In order to identify the image, the goal is to give it a unique name. Images in which just one item is visible and being inspected are typically referred to as image classification. Object identification, on the other hand, looks at more realistic situations when an image may contain many things and calls for both categorization and localization tasks [21].

Problem Statement

Due to the complexity of CT scan designs, lung cancer prediction is still difficult despite a large number of studies on CNN model designs for medical image processing. Feature representation, proper architecture, parameter selection, and determining the ideal weight-bias combination are among the performance challenges that deep learning (DL) models confront. Inspired by its encouraging outcomes in improving CNN models for breast cancer image classification, the Ebola Optimization Search Algorithm (EOSA) is specifically discussed.

Research Questions

1. How effective are deep learning models in classifying and detecting different types of lung cancer?
2. What are the main obstacles to employing deep learning for medical imaging-based lung cancer detection?
3. How do the execution metrics of deep learning models compare to conventional techniques in lung cancer classification and detection?
4. How might the interpretability of deep learning models for lung cancer diagnosis be enhanced by incorporating explainable AI techniques?

Research Objectives

The expansive nature of modern research is evident across all disciplines, with significant findings ranging from technical advancements in medical imaging and deep learning to comprehensive sociological studies on cultural empowerment and economic agency within diaspora communities (Tauseef, Jamal, & Tauseef, 2025; Tauseef, Jamal, & Tabasam, 2025).The aim of this endeavor is to improve the deep learning model for lung cancer diagnosis using a metaheuristic approach. The proposed approach is intended to assist physicians in the early diagnosis of the disease, enhancing decision-making and enabling appropriate treatment.

Literature Review

Convolutional neural networks have become the most popular method for analyzing medical images in some last years because they outperform algorithms that rely on manually generated features. Successful deployment of deep learning models, such as the VGG-16 based approach, requires addressing critical factors beyond technical performance, including the commitment to ethical analytics and digital transparency (Hassaan et al., 2023), realizing AI's potential for sustained competitive advantage (Akbar et al., 2023), managing its impact on workforce development (Jamshaid et al., 2024), and ensuring effective educational governance that builds community trust ("AI for Inclusive Educational Governance," 2024).

CNNs have been employed at the cell level in digital pathology jobs to identify cell nuclei and mitosis(Sardar et al., 2024). The first challenge involving WSIs to identify lymph node metastases of breast cancer was CAMELYON 16. Deeper and more potent CNN architectures, such as Google Net, VGG-Net, and ResNet, may be trained because of the challenge's extensive annotated training set.

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.04 (2025)

Related Works

To tackle the segmentation challenge, a number of methods have been devised. Most of these techniques belong to one of five fundamental categories: threshold-based, learning-based, region-growing, edge-detection, or deformable border. In the following, we concisely look over these groups.

Table 1: Comparison Table Of Different Study

Sr.No	Reference	Dataset Used	Methodology	Results
1	Maja Stella et al. [22]	ACDC LUNGH	VGG16, ResNet50, CNN	Accuracy 97.9 93
2	Janee Alam et al. [23]	UCI ML database	W Transform, GLCM	Identification- 97 Cancer Prediction- 87
3	M. B. Rodrigues, et al. [24]	LIDC-IDRI	Laplace, Gaussian & Sobel filtering, multilayer perceptron, SVM, KNN, SCM Mean HU	Accuracy (SCM Mean HU) - 96.70
4	Yutong Xie et al.[25]	LIDC-IDRI	KBC DL, U-Net, 3D-GLCM-SVM	Accuracy-91.60% Sensitivity-86.52% Specificity-94% AUC-95.70%
5	Qi Dou et al. [26]	LUNA16, Kaggle Data science Bowl 2021	3D CNN,	Sensitivity-87% Specificity- 99.1%
6	H. Xie et al. [27]	LUNA16(Testing),	2DCNN R-CNN(Detection Of Nodules)	AUC-0.954
7	X. Li et al. [28]	LIDC-IDRI, General Hospital Of Guangzhou Military Command	Anisotropic nonlinear diffusion filter, Random Walker (RW), RF, GLCM, LBP, Gabor Filter	Sensitivity- 0.92 Specificity-0.83 Accuracy-0.90 AUC-0.95
8	Liu, S. V M. et al. [29]	(TCIA)	SVM or XGBoost	AUC-0.850 Accuracy-0.797
9	Xiao, Z et al. [30]	3D-UNet CT	images LUNA16	DSC: 95.30%
10	Jalali, Y et al. [31]	Modified U-Net, replaced with a pre-trained ResNet-34 network (Res BCDU-Net)	CT images LIDC-IDRI	Accuracy: 97.58%

Methodology

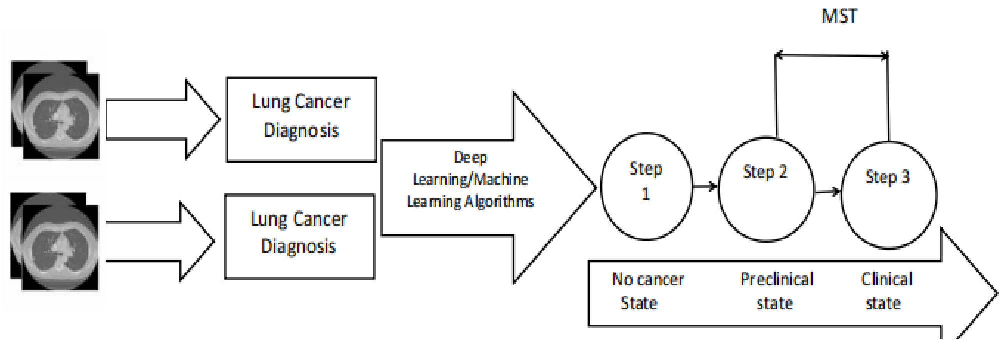
This chapter outlines the process for classifying and detecting lung cancer using deep learning approaches, mainly (CNNs). Deep learning models called **CNNs** are made to evaluate

ISSN E: 3006-1466
ISSN P: 3006-1458
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW

organized, grid-like input, like pictures. Convolutional, pooling, and completely linked layers are among its many layers. CNNs' hierarchical feature removal skills make them extremely successful for applications like object identification, picture segmentation, and image classification. It describes the model architecture, training procedure, evaluation metrics, implementation specifics, dataset selection, and preprocessing methods.

To tackle some of the aforementioned issues, this study focuses on three areas:

1) Segmenting the lung
2) Classifying and diagnosing nodules
3) Modeling the course of lung cancer



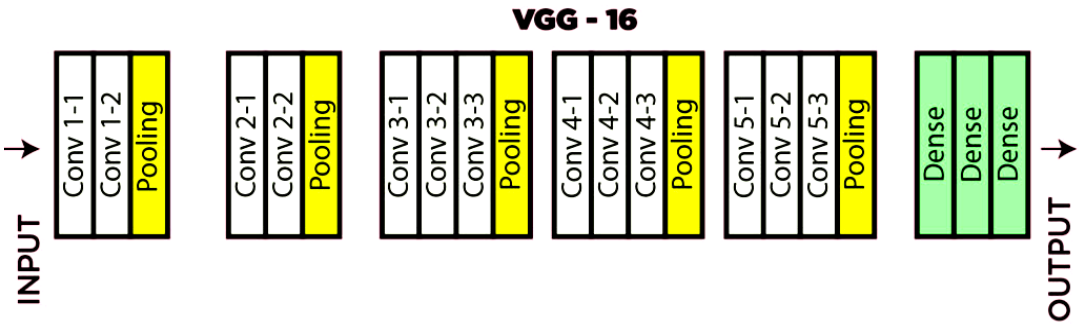


Figure 3: Architecture Diagram

Research Design

A deep learning-based method is employed Using medical images to identify and categorize lung cancer, particularly CT scans and X-rays. The research follows an experimental design, utilizing supervised learning with labeled medical imaging data. Its depth is noteworthy due to its 16 layers, which include 3 completely connected layers and 13 convolutional layers. Aside from its exceptional performance on many computer vision tasks, such as identifying objects and picture classification, VGG-16 is renowned for its efficacy and ease of use. Max-pooling layers follow a stack of convolutional layers with increasingly deeper depths in the model's architecture.

Figure:4 VGG-16 Architecture Map

Data Collection and Preprocessing

Data Source

Source : https://www.kaggle.com/datasets/andrewmvd/lung-and-colon-cancer histopathological-images

This collection contains 25,000 histopathological images divided into 5 classes. Each image has a dimension of 768 by 768 pixels and is stored as a jpeg file. 750 total photos of lung tissue (250 benign lung tissue, 250 lung cancer, and 250 lung small cell carcinoma) and 500 total images of colon tissue (250 benign colon tissue and 250 colon adenocarcinomas) were generated from an initial collection of HIPAA-compliant and validated sources. Using the Augmenter package, these were still converted to 25,000.

Five classifications, each with five thousand photos, make up the dataset:

a. Adenocarcinoma of the lung
b. Lung benign tissue
c. lung squamous cell cancer
d. colon adenocarcinoma
e. colon benign tissue

Data Analysis

The Lung and Colon Cancer Histopathological Image Dataset (LC25000), a substantial collection of histopathological images, was developed to enable the construction and assessment of ML models for the detection and categorization of cancer. The collection contains all 25,000 color JPEG images, each with a pixel size of 768 by 768.

Dataset Composition:

5 different classes are used to organize the dataset, each containing 5,000 images:

1. **Colon Adenocarcinoma (colon_aca):** Images depicting malignant glandular tumors originating in the colon.
2. **Benign Colonic Tissue (colon_n):** Images of non-cancerous colon tissues.
3. **Lung Adenocarcinoma (lung_aca):** Images showing malignant tumors with glandular differentiation in the lung.
4. **Lung Squamous Cell Carcinoma (lung_scc):** Images of malignant tumors arising from the squamous epithelium of the lung.
5. **Benign Lung Tissue (lung_n):** Images of non-cancerous lung tissues.

Data Sample:

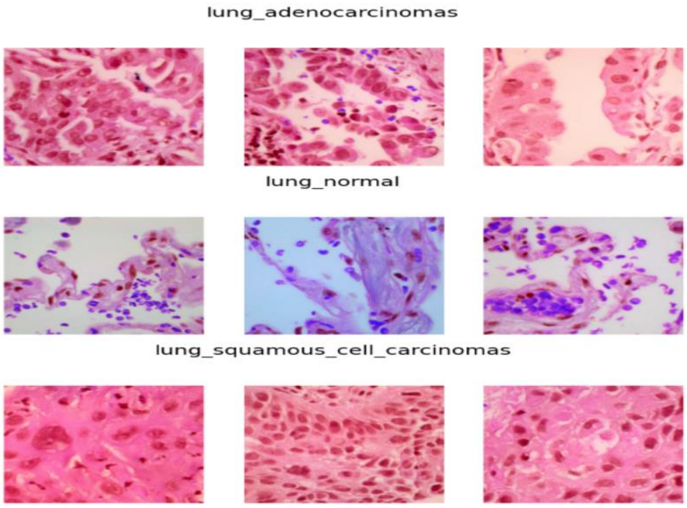


Figure:5 Data Sample

VGG-16 Model

VGG-16 follows a **sequential architecture** where multiple small convolutional layers (3x3 filters) are piled on top of one another, followed by fully connected layers.

- **Total Layers:** 16 weight layers (13 convolutional layers + 3 fully connected layers)
- **Input Image Size:** 224 × 224 × 3 (RGB image)

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR

- **Convolutional Layers:** Uses **3×3 filters** with a stride of 1 and padding of 1 (to preserve spatial dimensions)
- **Pooling Layers:** Uses **2×2 max pooling** with a stride of 2 (to reduce dimensions)
- **Fully Connected Layers:** 3 layers at the end, including a softmax classifier
- **Activation Function: ReLU** for non-linearity
- **Number of Parameters: 138 million**
- **Final Output:** 1000 classes (for ImageNet classification)

Layer-Wise Breakdown

Layer Type	Filter Size	Number of Filters	Activation	Output Shape
Input Layer	-	-	-	(224, 224, 3)
Conv1_1	3.0×3.0	64.0	ReLU	(224, 224, 64)
Conv1_2	3.0×3.0	64.0	ReLU	(224, 224, 64)
Max Pooling	2.0×2.0	-	-	(112, 112, 64)
Conv2_1	3.0×3.0	128.0	ReLU	(112, 112, 128)
Conv2_2	3.0×3.0	128.0	ReLU	(112, 112, 128)
Max Pooling	2.0×2.0	-	-	(56, 56, 128)
Conv3_1	3.0×3.0	256.0	ReLU	(56, 56, 256)
Conv3_2	3.0×3.0	256.0	ReLU	(56, 56, 256)
Conv3_3	3.0×3.0	256.0	ReLU	(56, 56, 256)
Max Pooling	2.0×2.0	-	-	(28, 28, 256)
Conv4_1	3.0×3.0	512.0	ReLU	(28, 28, 512)
Conv4_2	3.0×3.0	512.0	ReLU	(28, 28, 512)
Conv4_3	3.0×3.0	512.0	ReLU	(28, 28, 512)
Max Pooling	2.0×2.0	-	-	(14, 14, 512)
Conv5_1	3.0×3.0	512.0	ReLU	(14, 14, 512)
Conv5_2	3.0×3.0	512.0	ReLU	(14, 14, 512)
Conv5_3	3.0×3.0	512.0	ReLU	(14, 14, 512)
Max Pooling	2.0×2.0	-	-	(7, 7, 512)
Fully Connected 1	-	4096	ReLU	(1, 4096)
Fully Connected 2	-	4096	ReLU	(1, 4096)
Fully Connected 3	-	1000	Softmax	(1, 1000)

Table :2 Layer-Wise Breakdown

Precision: Indicates the proportion of projected positive cases that were true.

$$\frac{\text{(TPTP+FP)}}{\text{frac\{TP\}\{TP + Fp\}TP}} + \text{FpTP}$$

False positives are reduced with greater accuracy.

Recall: Indicates the proportion of true positive cases that were accurately anticipated.

$$\frac{\text{(TPTP+FN)}}{\text{frac\{TP\}\{TP + FN\}TP}} + \text{FNTP}$$

There are fewer false negatives when recall is higher.

F1-score: The harmonic mean of recall and accuracy, which balances the 2 measures.

Formula: 2×Recall × Precision Recall2 + Precision \times \frac {Accuracy \times Memory' {Accuracy + Memory} 2×Recall+Precision Accuracy × Memory

When there is an imbalance in class, it helps.

Support: How many real instances there are every class inside the test dataset.

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.04 (2025)

Results and Discussion

Class-wise Analysis

Class	Precision	Recall	F1-score	Support
0	0.95	0.95	0.95	1037
1	1.00	1.00	1.00	970
2	0.95	0.95	0.95	993

Table :3 Class-wise Analysis

Interpretation

- **Class 0 (1037 instances)**:
  - Precision = **0.95** → 95% of predictions for class **0** were accurate.
  - Recall = **0.95** → 95% of actual class **0** instances were appropriately recognized.
  - F1-score = **0.95** → Good balance of exactness and recollection.
- **Class 1 (970 instances)**:
  - Precision = **1.00** → All predictions for class **1** were correct.
  - Recall = **1.00** → No false negatives, meaning every actual instance of **1** was classified correctly.
  - F1-score = **1.00** → Perfect performance for this class.
- **Class 2 (993 instances)**:
  - Precision = **0.95** → 95% of predictions for class **2** were accurate.
  - Recall = **0.95** → 95% of actual class **2** instances were accurately classified.
  - F1-score = **0.95** → Strong performance.

Overall Model Performance

Metric	Value	Support
Accuracy	0.97	3000
Macro Avg	0.97	3000
Weighted Avg	0.97	3000

Table 4: Overall Model Performance

Confusion Matrix:

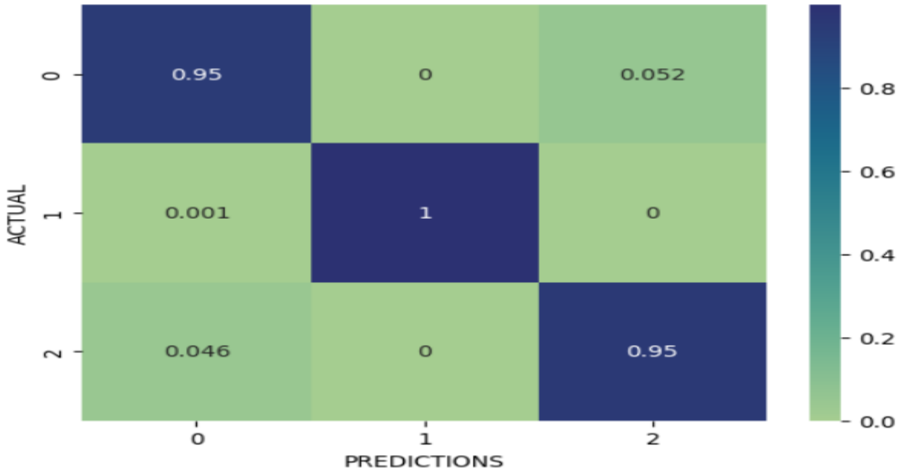


Figure 6: Confusion Matrix

Breakdown of Each Class

ISSN E: 3006-1466
ISSN P: 3006-1458
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW

Real Class → Predicted Class	0	1	2
0 (True class 0) → Predicted as **0**	**0.95**	0	0.052
1 (True class 1) → Predicted as **1**	0.001	**1.00**	0
2 (True class 2) → Predicted as **2**	0.046	0	**0.95**

Table 5: Breakdown of Each Class

Interpretation:

- **Class 0:**
  - 95% of instances were correctly classified as **0**.
  - **5.2% were misclassified as Class 2**.
- **Class 1:**
  - **Perfect classification! (100%)** No errors for this class.
- **Class 2:**
  - 95% of instances were correctly classified as **2**.
  - **4.6% were misclassified as Class 0**.

Key Insights

High Accuracy (Consistent with the Classification Report)

Class 1 Has Perfect Prediction (1.00 Precision & Recall)

Misclassification Occurs Between Classes 0 & 2

- Some examples from **Class 0 were misclassified as Class 2**.
- Some examples from **Class 2 were misclassified as Class 0**.

Conclusion

Since lung cancer is still one of the most common and deadly diseases in the world, early and accurate identification is essential to improving patient survival rates. The past few years, deep learning methods like CNNs have shown a lot of promise for automating the detection and classification of lung cancer. By leveraging CNN-based architectures, medical imaging data like Computed Tomography (CT) scans and X-rays can be analyzed with high precision, significantly aiding radiologists and clinicians in their diagnostic processes. This research highlights the effectiveness of deep learning prototypes in lung cancer classification by employing CNNs for feature extraction and pattern recognition. The need for sizable, annotated datasets to guarantee model generalizability is one significant problem. Deep learning models are data-intensive, and the limited availability of high-quality labeled medical images can hinder their performance.

In conclusion, CNN-based deep learning models offer a promising and efficient approach for lung cancer classification and detection. Their ability to analyze complex medical imaging data with high precision can revolutionize early cancer diagnosis, leading to timely treatment and improved patient outcomes. Nevertheless, issues with model interpretability, data quality, and clinical application need to be fixed if AI-driven lung cancer detection is to reach its full potential in the healthcare industry.

Future Recommendation

Future research in lung cancer detection using CNNs should focus on optimizing architectures, enhancing dataset diversity, and incorporating multi-modal data fusion, added genetic and clinical data, to improve diagnostic accuracy further. Additionally, Thorough validation is necessary before integrating deep learning models into practical healthcare settings. regulatory approvals, and collaboration between AI researchers and medical experts.

1. Enhancing Dataset Quality and Diversity

- **Larger and More Diverse Datasets:** To expand the generalization of deep learning models, future studies should focus on curating large-scale datasets with diverse lung cancer cases from different populations and medical institutions.

2. Addressing Class Imbalance Issues

- **Cost-Sensitive Learning:** Assigning higher weights to underrepresented classes (such as malignant cases) can help reduce bias in model predictions.

3. Improving Model Interpretability and Explain ability

- **Explainable AI (XAI) Integration:** Incorporating techniques such as Grad-CAM, SHAP, and LIME can help medical professionals consider the decision-making process of CNN models.

4. Optimizing CNN Architectures for Better Performance

- **Hybrid Deep Learning Models:** Combining CNNs with other deep learning approaches (e.g., Transformer models, Capsule Networks, or Recurrent Neural Networks) increase the accuracy of feature extraction and categorization.

5. Incorporating Multi-Modal Data Fusion

- **Integrating Clinical and Genetic Data:** Combining medical imaging with genetic, histopathological, and clinical data can provide a much holistic understanding of lung cancer progression and improve classification accuracy.

References

1) Nageswaran, S., Arunkumar, G., Bisht, A. K., Mewada, S., Kumar, J. S., Jawarneh, M., & Asenso, E. (2022). [Retracted] Lung Cancer Classification and Prediction Using Machine Learning and Image Processing. BioMed research international, 2022(1), 1755460.
2) Chaturvedi, P., Jhamb, A., Vanani, M., & Nemade, V. (2021, March). Prediction and classification of lung cancer using machine learning techniques. In IOP conference series: materials science and engineering (Vol. 1099, No. 1, p. 012059). IOP Publishing.
3) Kasinathan, G., & Jayakumar, S. (2022). Cloud-Based Lung Tumor Detection and Stage Classification Using Deep Learning Techniques. Biomed research international, 2022(1), 4185835.
4) 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.
5) Hassaan, A., Jamshaid, M. M., Siddique, M. N., Akbar, Z., & Niaz, S. (2023). ETHICAL ANALYTICS & DIGITAL TRANSFORMATION IN THE AGE OF AI: EMBEDDING PRIVACY, FAIRNESS, AND TRANSPARENCY TO DRIVE INNOVATION AND STAKEHOLDER TRUST. Contemporary Journal of Social Science Review, 1(04), 1-18.
6) Muhammad Mudaber Jamshaid, Ahmed Hassaan, Zeeshan Akbar, Muhammad Nouman Siddique, & Sikander Niaz. (2024). IMPACT OF ARTIFICIAL INTELLIGENCE ON WORKFORCE DEVELOPMENT: ADAPTING SKILLS, TRAINING MODELS, AND EMPLOYEE WELL-BEING FOR THE FUTURE OF WORK. Spectrum of Engineering Sciences, 2(1). Retrieved from https://thesesjournal.com/index.php/1/article/view/1417
7) Akbar, Z., Hassaan, A., Jamshaid, M. M., Siddique, M. N., & Niaz, S. (2023). Leveraging Data and Artificial Intelligence for Sustained Competitive Advantage in Firms and Organizations. Journal of Innovative Computing and Emerging Technologies, 3(1).

8) AI FOR INCLUSIVE EDUCATIONAL GOVERNANCE AND DIGITAL EQUITY EXAMINING THE IMPACT OF AI ADOPTION AND OPEN DATA ON COMMUNITY TRUST AND POLICY EFFECTIVENESS. (2024). Contemporary Journal of Social Science Review, 2(04), 2557-2567. https://doi.org/10.63878/cjssr.v2i04.1502
9) Tauseef, F., Jamal, A., & Tauseef, F. (2025). Empowerment through culture: Identity formation among South Asian women in the U.S. diaspora. Al-Aasar, 2(4), 347–355. https://doi.org/10.63878/aaj996
10) Tauseef, F., Jamal, A., & Tabasam, A. H. (2025). Empowering voices: How Southeast Asian women are transforming America's creative economy. Social Science Review Archives, 3(3), 2441–2448.
11) Kumar, D., Wong, A., & Clausi, D. A. (2015, June). Lung nodule classification using deep features in CT images. In 2015 12th conference on computer and robot vision (pp. 133-138). IEEE.
12) Sardar, Mahnoor, Muhammad Majid Niazi, and Fawad Nasim. "Ensemble deep learning methods for detecting skin cancer." Bulletin of Business and Economics (BBE) 13, no. 1 (2024).
13) Dey, R., Lu, Z., & Hong, Y. (2018, April). Diagnostic classification of lung nodules using 3D neural networks. In 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018) (pp. 774-778). IEEE.
14) Khan, Muhammad Ismaeel, Hassan Tahir, Md Ismail Jobiullah, Ali Raza A. Khan, Sakera Begum, and Ihtasham Hafeez. "Enhancing IoT Security: A Lightweight Cloning Approach for RFID/NFC Access Control Systems." Cuestiones de Fisioterapia 52, no. 2 (2023): 231-248.
15) Khan, Muhammad Ismaeel, Aftab Arif, and Ali Raza A. Khan. "The Most Recent Advances and Uses of AI in Cybersecurity." BULLET: Jurnal Multidisiplin Ilmu 3, no. 4 (2024): 566-578.
16) Arif, Aftab, Fadia Shah, Muhammad Ismaeel Khan, Ali Raza A. Khan, Aftab Hussain Tabasam, and Abdul Latif. 2023. "Anomaly Detection in IoHT Using Deep Learning: Enhancing Wearable Medical Device Security." Migration Letters 20 (S12): 1992–2006.
17) Lakshmanaprabu, S. K., Mohanty, S. N., Shankar, K., Arunkumar, N., & Ramirez, G. (2019). Optimal deep learning model for classification of lung cancer on CT images. Future Generation Computer Systems, 92, 374-382.
18) Sun, W., Zheng, B., & Qian, W. (2016, March). Computer aided lung cancer diagnosis with deep learning algorithms. In Medical imaging 2016: computer-aided diagnosis (Vol. 9785, pp. 241-248). SPIE.
19) Wu, J., Zan, X., Gao, L., Zhao, J., Fan, J., Shi, H., ... & Xie, X. (2019). A machine learning method for identifying lung cancer based on routine blood indices: qualitative feasibility study. JMIR medical informatics, 7(3), e13476.
20) Tran, G. S., Nghiem, T. P., Nguyen, V. T., Luong, C. M., & Burie, J. C. (2019). Improving accuracy of lung nodule classification using deep learning with focal loss. Journal of healthcare engineering, 2019(1), 5156416.
21) Khan, A. A., Arslan, M., Tanzil, A., Bhatty, R. A., Khalid, M. A. U., & Khan, A. H. (2024). Classification of colon cancer using deep learning techniques on histopathological images. Migration Letters, 21(S11), 449-463.
22) Šarić, Matko, et al. "CNN-based method for lung cancer detection in whole slide histopathology images." 2019 4th International Conference on Smart and Sustainable Technologies (SpliTech). IEEE, 2019.

23) Alam, Janee, Sabrina Alam, and Alamgir Hossan. "Multi-stage lung cancer detection and prediction using multi-class svm classifie." 2018 International conference on computer, communication, chemical, material and electronic engineering (IC4ME2). IEEE, 2018.
24) Rodrigues, M. B., Da Nobrega, R. V. M., Alves, S. S. A., Reboucas Filho, P. P., Duarte, J. B. F., Sangaiah, A. K., & De Albuquerque, V. H. C. (2018). Health of things algorithms for malignancy level classification of lung nodules. IEEE Access, 6, 18592-18601.
25) Xie, Y., & Li, Q. (2022). A review of deep learning methods for compressed sensing image reconstruction and its medical applications. Electronics, 11(4), 586.
26) 26) Dou, Q., So, T. Y., Jiang, M., Liu, Q., Vardhanabhuti, V., Kaissis, G., ... & Heng, P. A. (2021). Federated deep learning for detecting COVID-19 lung abnormalities in CT: a privacy-preserving multinational validation study. NPJ digital medicine, 4(1), 60.
27) Zhao, Y. M., Shang, Y. M., Song, W. B., Li, Q. Q., Xie, H., Xu, Q. F., ... & Xu, A. G. (2020). Follow-up study of the pulmonary function and related physiological characteristics of COVID-19 survivors three months after recovery. EClinicalMedicine, 25.
28) Li, X., Lovell, J. F., Yoon, J., & Chen, X. (2020). Clinical development and potential of photothermal and photodynamic therapies for cancer. Nature reviews Clinical oncology, 17(11), 657-674.
29) Liu, S. V., Reck, M., Mansfield, A. S., Mok, T., Scherpereel, A., Reinmuth, N., ... & Horn, L. (2021). Updated overall survival and PD-L1 subgroup analysis of patients with extensive-stage small-cell lung cancer treated with atezolizumab, carboplatin, and etoposide (IMpower133). Journal of Clinical Oncology, 39(6), 619-630.
30) Xiao, Z., Liu, B., Geng, L., Zhang, F., & Liu, Y. (2020). Segmentation of lung nodules using improved 3D-UNet neural network. Symmetry, 12(11), 1787.
31) Jalali, Y., Fateh, M., Rezvani, M., Abolghasemi, V., & Anisi, M. H. (2021). ResBCDU-Net: a deep learning framework for lung CT image segmentation. Sensors, 21(1), 268