Received: 9 October 2023 | *Revised:* 6 December 2023 | *Accepted:* 21 December 2023
DOI: 10.37917/ijeee.20.2.23 *Vol. 20 | Issue 2 | December 2024*

Iraqi Journal for Electrical and Electronic Engineering
Original Article

Open Access
IJEEE University of Basrah College of Engineering

A Hybrid Lung Cancer Model for Diagnosis and Stage Classification from Computed Tomography Images

Abdalbasit Mohammed Qadir*[1], Peshraw Ahmed Abdalla[2], Dana Faiq Abd[1]
[1]Information Technology Department, University of Human Development, Sulaymaniyah, Iraq
[2]Computer Science Department, University of Halabja, Halabja, Iraq

Correspondance
*Abdalbasit Mohammed Qadir
Information Technology Department,
University of Human Development, Sulaymaniyah, Iraq
Email: abdalbasit.mohammed@uhd.edu.iq

Abstract
Detecting pulmonary cancers at early stages is difficult but crucial for patient survival. Therefore, it is essential to develop an intelligent, autonomous, and accurate lung cancer detection system that shows great reliability compared to previous systems and research. In this study, we have developed an innovative lung cancer detection system known as the Hybrid Lung Cancer Stage Classifier and Diagnosis Model (Hybrid-LCSCDM). This system simplifies the complex task of diagnosing lung cancer by categorizing patients into three classes: normal, benign, and malignant, by analyzing computed tomography (CT) scans using a two-part approach: First, feature extraction is conducted using a pre-trained model called VGG-16 for detecting key features in lung CT scans indicative of cancer. Second, these features are then classified using a machine learning technique called XGBoost, which sorts the scans into three categories. A dataset, IQ-OTH/NCCD - Lung Cancer, is used to train and evaluate the proposed model to show its effectiveness. The dataset consists of the three aforementioned classes containing 1190 images. Our suggested strategy achieved an overall accuracy of 98.54%, while the classification precision among the three classes was 98.63%. Considering the accuracy, recall, and precision as well as the F_1-score evaluation metrics, the results indicated that when using solely computed tomography scans, the proposed (Hybrid-LCSCDM) model outperforms all previously published models.

Keywords
Lung Cancer, CT Scan, Deep Learning, Diagnosis, Detection Model.

I. Introduction

Today, with the advent of deep learning techniques, the entire healthcare sector has transformed with electronic diagnosis and treatment to aid patients and medical professionals during the diagnostic and therapeutic phases of illnesses [1]. One of the most significant health issues facing the globe today is cancer [2]. The development of several genetic abnormalities and epigenetic alterations contribute to lung cancer, which causes normal cells to proliferate out of control [3]. Every year, more people worldwide pass away due to the disease. Based on the available data, it is estimated that the number of worldwide cases of cancer could reach a total of 28.40 million by the year 2040. Nevertheless, this will be made worse by the rising risk factor linked to economic expansion and globalization [4]. According to the Global Cancer Statistics 2020 [5], in a comparison between all cancer types, the greatest fatality rate belongs to lung cancer patients, with an estimated annual death rate of 1.80 million. The risk of mortality is further increased by inconsistent monitoring and care. Through diverse methods such as classification, segmentation, and detection approaches, several researchers have tackled these problems with lung cancer detection [6, 7]. Artificial intelligence has become more important in healthcare and computer vision applications in recent years because of its superior performance in prediction, detection, and suitability for categorization issues. It is necessary to overcome the constraints of the current

lung cancer detection techniques. As computer technology advances swiftly, the relevance of machine learning techniques in lung cancer diagnosis is becoming more pronounced. The exploration of diverse algorithms for lung cancer prediction is ongoing among researchers. However, studies on lung cancer prediction indicate certain limitations in these methods. For instance, selecting an appropriate kernel function for SVM poses a challenge. In the case of Naive Bayes, the requirement for a known prior probability can be a drawback; inaccuracies in the assumed prior distribution may result in suboptimal prediction outcomes [8]. Lung cancer detection systems and research must be improved in terms of accuracy in order to work properly. The development of hybrid methodologies can enhance electronic diagnosis, early-stage lung cancer detection, localization detection, tumor segmentation, and stage classification. We apply these cutting-edge techniques inside a hybrid model to use in the healthcare sector since their effectiveness inspires us. In this effort, we solved these problems of the existing scenario and created a new prototype named "Hybrid Lung Cancer Stage Classifier and Diagnosis Model" (Hybrid-LCSCDM) framework; this uses a feature extraction stage based on the VGG16 model and a classification stage based on the XGBoost algorithm to identify lung cancer as early as possible.

A. Main Contributions

We suggested an intelligent diagnosis module named (Hybrid-LCSCDM) to identify and classify lung cancers at an early stage. The proposed model was employed based on a hybrid approach for both stage classification and feature extraction, which reduced detection time complexity and increased detection rate accuracy.

1. A hybrid system is presented for impactful analysis.
2. This model's dependence is between image-quantitative and clinical features; for a more precise computer-aided diagnosis, it also offers a novel feature representation.
3. This technique employs VGG16 for feature extraction and an XGBoost classifier for lung cancer classification and diagnosis in an efficient manner.
4. The Lung Cancer Dataset (IQ-OTH/NCCD) was used to assist us in carrying out the tests in this study.
5. The suggested Hybrid-LCSCDM has been used to test the effectiveness of a novel technique also presented in this research.

The organization of the paper is as follows: Section II. delves into a comprehensive literature review, where we explore and synthesize relevant previous. Section III. , titled 'Methodology,' details the methodologies and approaches we employed in our research. In Section IV. , we present our results, and finally, Section V. is dedicated to the discussion.

II. Literature Review

Sharma et al. examined different kinds of colon and lung (3 malignant and two benign) tissues by creating a diagnosis system utilizing the histopathological dataset. According to the data, the proposed model in this research can successfully identify malignant tissues up to 96.33 percent of the time [9]. The success of the CAD system relies on the extraction of relevant texture features for distinguishing between cancerous and noncancerous nodules, and the authors mention area, calcification, shape, size, and contrast enhancement as features. The appropriateness and sufficiency of these features in capturing the variations in lung nodules need careful consideration.

For the purpose of assisting radiologists, Masood et al. devised a method for diagnosing lung nodules based on 3D-DCNN employing computer-aided diagnosis assistance systems. In this work, a computer-aided diagnostic (CAD) model was validated and trained on the LIDC-IDRI, ANODE09, and LUNA16 datasets [10]. A deep CNN-based binary classifier was developed by Kalaivani et al. through the utilization of the DenseNet model, with the aim of identifying patients with either benign or aggressive lung cancer [11], as the researchers used 201 lung scans in a dataset in this study, 85% of the dataset was used to train the model while using the remaining of the images for the testing phase, test results showed that the suggested approach achieved 90.85% accuracy on the mentioned dataset.

To classify various cancer types using genetic data with reference to tumor RNA sequences, ElNabi et al. provided a special optimized deep-learning model, namely a convolutional neural network (CNN) and decision tree (BPS, O-DT) while making use of the particle swarm optimization. This study addressed the performance criteria, including recall, F_1-score, and precision [11].

Qin et al. explained how to collect, analyze, and fuse multitype interdependent characteristics to determine EGFR mutation status utilizing computer-assisted diagnostics. In this study, the CNN-RNN architecture is employed to provide a novel hybrid network paradigm. Using CNN, quantitative image characteristics are retrieved, and a model is created to show how different types of features relate to one another [12].

To detect lung cancer, Joshua et al. proposed an unsupervised learning model for 3D CNN [13]. An improved gradient activation function in the 3D CNN binary classifier model increases the visibility of lung tumors. The suggested AlexNet detection method employing the LUNA dataset is contrasted with a 2D CNN learning classifier that is already in use. The proposed model is useless since there was not insufficient data

for training, with only 10% of the training dataset utilized.
Chaunzwa et al. constructed a model for the purpose of identifying lung cancer patients with early-stage adenocarcinoma (ADC) and squamous cell carcinoma (SCC) via a supervised CNN detection system. Utilizing real-time non-SCLC from patients who were impacted in the early stages and were collected at Massachusetts General Hospital, CNN has been tested [14]. About 311 data phases have been gathered, and they are all present in the database. They created CNN, a learning detection system with a 71 percent AUC detection rate, which was inadequate.
Chaturvedi et al. evaluated the methods of the most recent studies conducted on detecting and classifying lung cancer. Super Bowl Dataset 2016, LUNA 16, and standard datasets LIDC-IDRI are accustomed to supervised learning algorithms like SVM, CNN, and KNN. According to the authors of the article, these algorithms are often used in the identification of diseases and in CT data [15]. Naik et al. [16] provided a detailed description of a pulmonary nodule classification system utilizing a fractal network. The Fractalnet model was employed on the LUNA16 dataset for training and validating the system's performance, resulting in an accuracy of 94.7%. Nasser et al. [17] presented a LungNet-SVM approach for the efficient segmentation and classification of pulmonary nodules in CT images into just two classes. LungNet-SVM is a modified iteration of the AlexNet architecture, and a support vector machine (SVM) algorithm is employed as a classifier. During the training and validation phases, the model considers three different input image sizes (16×16, 32×32, *and* 48×48) and undergoes optimization using three different optimizers—Adam, RMSprop, and SGD—in order to fine-tune the model for optimal accuracy. The experimental results reveal that the LungNet-SVM model, particularly when utilizing the SGD optimizer, attains the highest accuracy when operating on 48×48 input image sizes with 97.64% accuracy.
The current approaches and datasets utilized in various algorithms are summarized in Table I.

III. Methodology

The presented model consists of two major components: a deep transferring learning model, namely "VGG-16", is the first component that is used as a feature extractor, and the second component is a machine learning algorithm, namely "XGboost classifier," the proposed model was trained and tested on the IQ-OTH/NCCD dataset.

A. Performance Metrics

We used several performance metrics to estimate our model (Hybrid-LCSCDM), including:

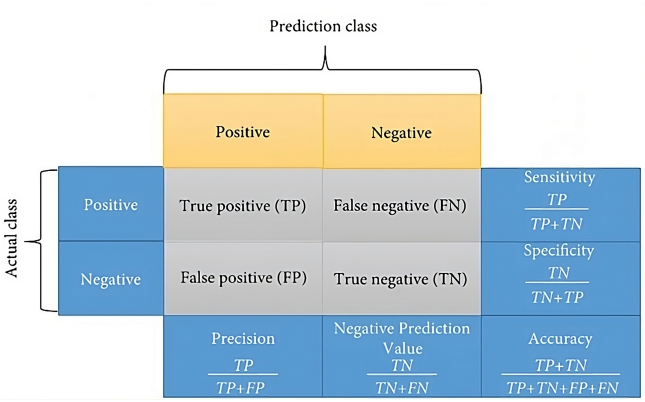

Fig. 1. Performance measures that are used to validate the suggested hybrid-LCSCDM model.

1) Confusion Matrix

A confusion matrix displays the model's predictions. It is intended to display instances in which the model properly and erroneously worked on the data. Fig. 1 shows the components of a confusion matrix; it provides details on the anticipated performance indicators for confirming the validity of the suggested Hybrid-LCSCDM model and shows four findings. The first two are the true positive (TP) and true negative (TN), as these are the correctly predicted predictions by the proposed model. The other two are false positive (FP) and false negative (FN), which are the predictions that our proposed model failed to correctly predict. The rows reflect the possible classifications, while the accurate categorization of the data is represented as columns [19].

2) Accuracy

One of the most popular and widely used metrics in machine learning, this metric shows the proportion of the correct predictions with respect to the overall data [19].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

3) Precision

Precision indicates the ratio of accurately anticipated outcomes compared to the total of the correctly predicted observations [19] and is described as:

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

4) Recall

The recall is used to calculate the proportion of properly predicted positive outcomes relative to the total number of outcomes in a given class [19] and may be defined as follows:

TABLE I.
SOME OF THE CURRENT APPROACHES AND DATASETS UTILIZED IN VARIOUS ALGORITHMS FOR THE DETECTION OF LUNG CANCER

Study	Year	Methodology	Dataset(s)	Performance	Analysis
[9]	2011	Algorithms for unsupervised learning	LIDC-IDRI	Acc. (94.3%)	Focus has been placed on a tissue-based categorization approach
[10]	2018	3D-CNN model	BR-Dataset	Acc. (92.65%)	Additional trending algorithms might improve the accuracy, precision, and recall
[18]	2020	DenseNet classifier	LIDC-IDRI	Acc. (90.85%)	Simply utilized a small database
[11]	2020	BPSO-DT	LUNA	Acc. (88.25%)	Other trending algorithms can increase the accuracy
[12]	2020	RNN, CNN	LUNA16, LIDC-IDRI, and ANODE09	Acc.(91%) AUC.(0.78)	The proposed model, compared to the other approaches, achieved a lower accuracy rate
[13]	2021	3D CNN-AlexNet	LUNA	Acc. (89%)	With 10% of the data evaluated, the AlexNet model's flaw was shown to be ineffective for real-time medical evaluation
[14]	2021	CNN	LUAD	Acc. (71%)	This model's limitations are not concentrated on the segmentation and preprocessing that increase the accuracy of the model
[15]	2021	KNN, SVM classifiers	LUNA 16, LIDC-IDRI	Acc. (91%)	The complexity of elapsed time is considerable
[16]	2022	SqueezeNet + ResNet	LUNA16	Acc. (94.87%)	
[17]	2023	Modified CNN + SVM	LUNA16	Acc. (97.64%)	The proposed LungNet-SVM classifies lung cancer into only two categories: benign and malignant

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

5) F_1 Score

The F_1 score is the weighted average of accuracy and recall, resulting in a value between 0 and 1. F_1 score is considered a superior performance statistic than accuracy [19]and is defined as follows:

$$F_1 score = \frac{2 \times (recall * precision)}{recall + precision} \tag{4}$$

It should be mentioned that the distribution of the data is the determining factor in the process of selecting the metrics of F_1 score or accuracy. In cases when the classes are very imbalanced, the F_1 score tends to be an appropriate option over accuracy since the majority of real-world classification problems, there is an uneven distribution of classes. While accuracy is employed in cases where the distribution of the class is comparable, the disadvantage is that this metric does not consider the ratio of the distribution between the classes, and this can affect the obtained results, leading to incorrect conclusions [20].

B. Dataset

The source of the dataset used in this research is the National Center for Cancer Diseases/IQ-OTH (IQ-OTH/NCCD) [21]; the data was collected over three months in the autumn of 2019. It includes CT images of individuals who are healthy as well as patients with varying stages of lung cancer.
Radiologists and oncologists at these two institutions annotated IQ-OTH/NCCD images. The entire collection consists of 1190 images corresponding to CT scans. These images were obtained from 110 patients. Originally, Siemens SOMATOM was employed as the scanner, and the original format of CT

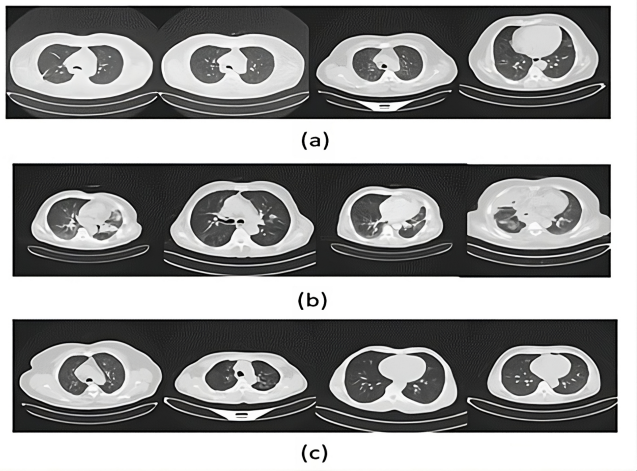

Fig. 2. Samples of the dataset: (A) Benign, (B) Malignant, and (C) Normal.

images was the DICOM format. CT procedure consists of 120 kV, Window sizes of 350 to 1200 HU, slices of 1 mm of thickness, and window centers ranging from 50 to 600 for reading with full inspiration, with breath-hold. Before the analysis process, all photos were undergone de-identification. Each scan includes a number of slices. The number of these slices ranges between 80 and 200, and each of them is represented with a picture of the patient's chest shown from different perspectives and various sides. Each of the benign patients had a different age, gender, geographical location, educational level, and living condition. Some of these patients were workers and farmers, while another portion of them were employees of the Transport and Oil Ministry of Iraq. Mostly living in the provinces of Babylon, Wasit, Salahuddin, Diyala, and Baghdad. Fig. 2 shows a sampling of our datasets. The dataset was divided into three categories: malignant (40), benign (15), and normal (55).

Benign lesions, such as benign tumors or cysts, are noncancerous but may resemble cancer on scans. They typically have well-defined borders and uniform appearances, unlike malignant tumors, which are often irregular and rapidly changing. A common benign lung lesion is a hamartoma, characterized by its slow growth and well-circumscribed nature [22], while normal lung imaging shows healthy lung tissue with no abnormalities. These images are essential for comparison to identify pathological changes in subsequent scans [23].

C. Preprocessing

In this study, two major preprocessing steps are utilized. First, the dataset images came with an original size of 512x512 pixels; the size of these images was reduced to 224x224, which is the default input size of the VGG16 network. Second, all of the pixels were normalized to the range of 0 and 1 in order to reduce the computational complexity of the image recognition process. In our study, we addressed the challenge of class imbalance in our dataset through the application of Stratified 5-fold Cross-Validation. This technique is particularly beneficial for imbalanced datasets as it ensures that each fold maintains a consistent ratio of the different class labels, closely mirroring the original dataset's class distribution. By employing this method, we aimed to achieve a more reliable and unbiased evaluation of our model's performance. Each of the five folds was used once as a validation set, while the remaining four folds served as the training set. This approach not only allowed for a thorough assessment of the model across different subsets of the data but also contributed to mitigating the potential bias that could arise from the uneven class distribution. The consistent representation of classes in each fold helped in validating the robustness and generalizability of our model, ensuring that the validation metrics were reflective of its performance across the diverse class spectrum [24].

D. Feature Extraction using VGG-16

Initially, the first VGG network design was proposed in a study by Simonyan and Zisserman [25]. Two variations of the model were created, which are a 16-layer (VGG-16) as well as a 19-layer (VGG-19) network, with the aim of submitting these models and participating in the ImageNet challenge in 2014, where the competition was won by the team of Visual Geometry Group (VGG), achieving a second place in the localization track while taking first place in the classification track. The architecture of VGG16 is composed of five convolutional layer blocks followed by three layers fully linked. Convolutional layers make sure that each of the activation maps stays at the same spatial dimensions compared to the layer, which is below it, through the use of 3x3 kernels and a padding and stride of 1.

After each convolution, a rectified linear unit (ReLU) activation and max pooling operation are immediately performed to reduce the spatial dimension. In the activation map, each spatial dimension from the previous layer is cut in half when using 22 kernels through the max pooling layers while using two strides and without padding. The final layer is the softmax layer, which consists of 1,000 completely linked layers and is then employed after two fully connected layers containing 4,096 units with a ReLu activation function.

At the same time, one of the drawbacks of the VGG6 model is that it is expensive to analyze, and it requires a lot of processing power in terms of memories and parameters. The model contains almost 138 million parameters. In the proposed model, an XGBoost classifier replaces the fully connected layers, in which approximately 123 million parameters reside. These layers are considered the bulk of these parameters; this

drastically lowers the number of required parameters.
The XGBoost classifier was added to replace the top layer of the previously mentioned network, namely VGG16. This study uses a CT scan image dataset that is labeled. Therefore, inductive transfer learning is the type of learning that is used in this context since the ImageNet weights are used by the VGG16 model. The VGG16 model's weight, which was used to classify 1000 pictures out of a dataset of 14 million, is transferred and utilized to conduct feature extraction on the CT image dataset. The VGG16 model feature extraction findings are sent to the XGBoost algorithm [26, 27].

E. Classification using XGBoost classifier

Extreme Gradient boosting, or XGboost, uses a decision tree technique with gradient boosting. XGboost is built for speed and efficiency. Its engineering objective is to push the computational resource limitations considering enhanced tree algorithms. Python, C++, Java, R, and other programming languages are just a few of the interfaces available for XGboost. In this study, the Python interface was used. The approach optimizes computing speed and memory use, while the term "Boosting" is a strategy that uses an ensemble technique in which new models are introduced to older models to rectify their flaws. Models are successively introduced until no more improvements are possible. A well-known example of boosting is the AdaBoost algorithm, which weights difficult-to-predict data points, while Gradient boosting is a technique that works on creating new models to forecast mistakes or the residuals of previous models and merging them to produce the final prediction. While adding new models to minimize loss, a gradient descent approach is employed, which the name Gradient boosting comes from. With each additional model, the accuracy of the prediction increases.
The aim of the model of gradient descent for a boosted tree is derived from (5) using Taylor expansion. Equation (6) includes the regularization, whereas (7) accounts for the tree's generalization. The regularization aim of XGboost is to choose a model with simple prediction functions [28].

$$\omega_j^* = -\frac{A_j}{B_j+\lambda} \tag{5}$$

$$f(\text{" obj "}) = -\frac{1}{2}\sum_{j=1}^{T}\frac{A_j^2}{B_j+\lambda} + \gamma T \tag{6}$$

$$\text{"Gain"} = \frac{1}{2}\left[\frac{A_L^2}{B_L+\lambda} + \frac{A_R^2}{B_R+\lambda} + \frac{(A_L+A_R)^2}{B_L+B_R+\lambda}\right] - \gamma \tag{7}$$

Fig. 3 demonstrates the suggested model's design. Mainly, The VGG-16 model was employed for feature extraction, whereas the XGboost model was applied for the calcification phase in both the training and testing stages.

TABLE II.
ANALYZING PERFORMANCE MEASURES IN COMPARISON TO OTHER TECHNIQUES

Study	Technique	Dataset	Accuracy%
The proposed method	Hybrid-LCSCDM	IQ-OTHNCCD	98.54
AL-Huseiny et al. [29]	GoogLeNet	IQ-OTHNCCD	94.38
Kareem et al. [30]	SVM	IQ-OTHNCCD	89.88
Al-Yasriy et al. [31]	AlexNet	IQ-OTHNCCD	93.54

IV. RESULTS AND DISCUSSION

In this part, the performance of the suggested (Hybrid-LCSCDM) is evaluated, and experimental findings are shown. We employed a transfer learning method using VGG16 as the foundation model to extract features from CT scan data, which were then used to train an XGBoost classifier.
For the testing and evaluation phases of the proposed hybrid model, the publicly available (IQ-OTH/NCCD) dataset was used. For efficient testing of the model, the Our research experiments were conducted on a computer with an Intel Core i7-13700F CPU, NVIDIA RTX 4060 GPU, and 64 GB of RAM, providing the necessary computational power and efficiency for our data-intensive tasks.
In Table II, in terms of Recall and accuracy, we contrasted our performance measures with the outcomes of existing models. In Table II, the evaluation metrics are shown; the Hybrid-LCSCDM (VGG-16-XGboost) model has achieved an accuracy of 98.5%, which is higher than the compared models related to lung cancer detection. In Fig. 4, we have provided

TABLE III.
EVALUATION METRICS FOR THE HYBRID-LCSCDM MODEL

Fold	Accuracy	Precision	Recall	F1-Score
1	97.73%	0.9818	0.9306	0.9519
2	99.10%	0.9821	0.9821	0.9821
3	98.17%	0.9853	0.9543	0.9684
4	99.09%	0.9930	0.9821	0.9874
5	98.64%	0.9891	0.9682	0.9780
Average	98.54%	0.9863	0.9635	0.9736

a visual representation of the results obtained by the Hybrid-LCSCDM model. This visualization aims to offer a clear and

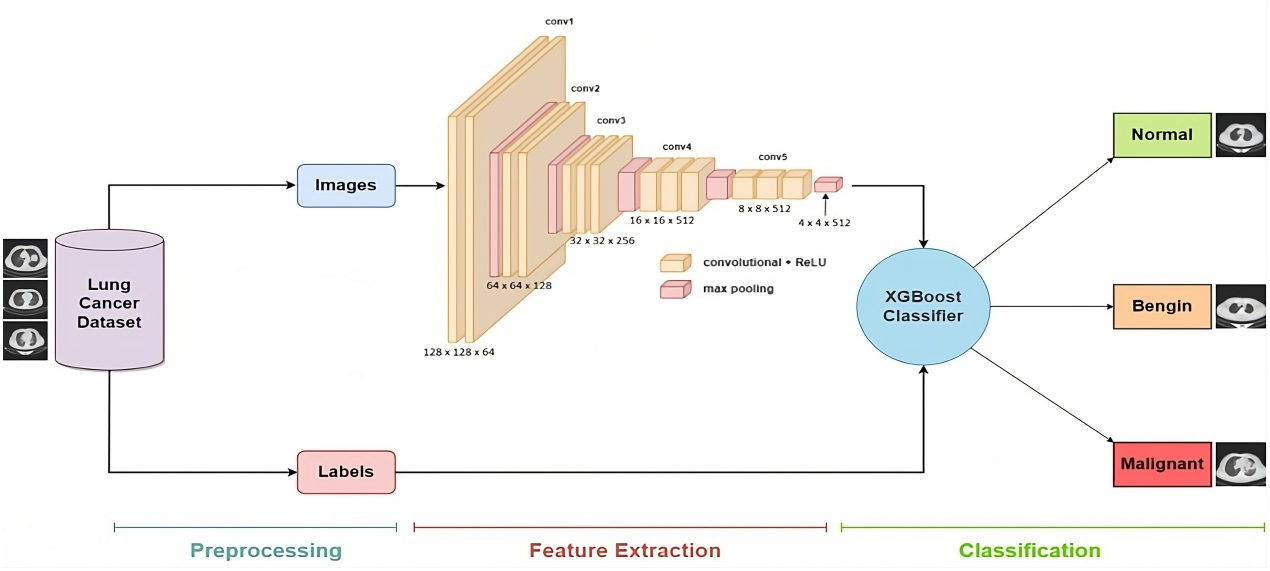

Fig. 3. Proposed model architecture.

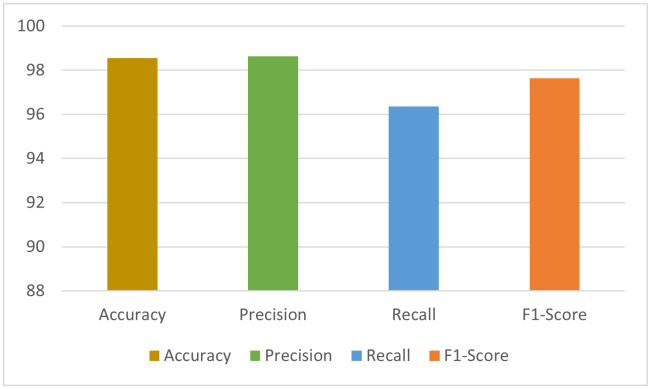

Fig. 4. Average metrics results comparison of the proposed model.

comprehensive understanding of the model's performance across various evaluation metrics, including precision, recall, and F_1 score. The goal of this study was to use CT scans to identify and categorize lung cancer, which is proven to be one of the main causes of death worldwide. The outcomes showed that the suggested model could accurately predict CT images with the necessary sensitivity.

Deep learning is now often employed in a variety of machine vision applications, including segmentation, object recognition, and image categorization [32]. Using deep neural networks in databases, it is possible to classify images with high accuracy. Machine vision and deep learning have been used to automatically diagnose cancer in a number of works. In this domain, the outcomes of these study efforts are commendable. For the purpose of detecting and classifying lung cancer in this investigation, we created a hybrid model. When the proposed model's performance is compared to that of earlier research for the same purpose, it outperforms all of the work that has been done. The suggested system was trained using two techniques: XGBoost classifier for classification and VGG16 for feature extraction. The findings indicated that the hybrid approach-based proposed solution had the highest overall accuracy value at 98.54 percent.

V. CONCLUSION

In conclusion, this research concentrated on developing a Hybrid Lung Cancer Stage Classifier and Diagnosis Model (Hybrid-LCSCDM) that comprises detection modules and stage classifiers with the aim of achieving a reduction in the rate of death caused by lung cancer. Lung cancer, as a fatal disease, is caused by the unlimited distribution of cells in the lung tissue, which is the main cause of this type of cancer. Early detection of lung cancer is vital since it may improve patient survival. This paper presents a hybrid transfer learning and machine learning model that works on computed tomography images in order to detect lung cancer.

The proposed Hybrid-LCSCDM module utilizes neural networks using a hybrid learning approach as a predictor and stage classifier. The process of classification includes feature extraction using a pre-trained model, namely VGG16, preceded by the CT image classification through the use of the XGBoost algorithm. The suggested model more accurately diagnoses lung cancer via computed tomography images compared to the previous studies. The findings of the performance evaluation show the accuracy of the model as high as 98.54%;

such results show that this hybrid model is useful in lung cancer detection and can be employed in real-world cases. Future improvements to our Hybrid-LCSCDM model could involve expanding the dataset for broader coverage of lung cancer cases integrating advanced deep learning techniques for improved accuracy. These steps would enhance the model's effectiveness and its applicability in diverse clinical settings.

Conflict of Interest

The authors have no conflict of relevant interest to this article.

References

[1] P. Abdalla, A. Mohammed, O. Rasheed, S. Karim, B. Mohammed, and K. Ghafoor, "Transfer learning models comparison for detecting and diagnosing skin cancer," *Acta Informatica Malaysia*, vol. 7, no. 1, pp. 1–7, 2022.

[2] P. Abdalla, A. Qadir, O. Rashid, K. Rawf, A. Abdulrahman, and B. Mohammed, "Deep transfer learning networks for brain tumor detection: The effect of mri patient image augmentation methods," *International Journal of Electronics and Communications*, vol. 75, p. 83, 2021.

[3] S. Yadav, M. Zain, P. Sahai, S. Porwal, and V. Chauhan, "Challenges encountered in cancer care and management during covid-19 in south asian countries," *Asian pacific journal of cancer care*, vol. 5, no. S1, pp. 101–107, 2020.

[4] M. Bicakci, O. Ayyildiz, Z. Aydin, A. Basturk, S. Karacavus, and B. Yilmaz, "Metabolic imaging based subclassification of lung cancer," *IEEE Access*, vol. 8, pp. 218470–218476, 2020.

[5] H. Sung *et al.*, "Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 71, no. 3, pp. 209–249, 2021.

[6] Z. Xiao, B. Liu, L. Geng, F. Zhang, and Y. Liu, "Segmentation of lung nodules using improved 3d-unet neural network," *Symmetry*, vol. 12, no. 11, p. 1787, 2020.

[7] M. Savic, Y. Ma, G. Ramponi, W. Du, and Y. Peng, "Lung nodule segmentation with a region-based fast marching method," *Sensors*, vol. 21, no. 5, p. 1908, 2021.

[8] S. Liu and W. Yao, "Prediction of lung cancer using gene expression and deep learning with kl divergence gene selection," *BMC bioinformatics*, vol. 23, no. 1, p. 175, 2022.

[9] D. Sharma and G. Jindal, "Computer aided diagnosis system for detection of lung cancer in ct scan images," *International Journal of Computer Electrical Engineering*, vol. 3, no. 5, p. 714, 2011.

[10] A. Masood and et al., "Computer-assisted decision support system in pulmonary cancer detection and stage classification on ct images," *Journal of biomedical informatics*, vol. 79, pp. 117–128, 2018.

[11] M. L. R. AbdelNabi, M. W. Jasim, H. M. El-Bakry, M. H. N. Taha, and N. E. M. Khalifa, "Breast and colon cancer classification from gene expression profiles using data mining techniques," *Symmetry*, vol. 12, no. 3, p. 408, 2020.

[12] R. Qin and et al., "Fine-grained lung cancer classification from pet and ct images based on multidimensional attention mechanism," *Complexity*, vol. 2020, 2020.

[13] E. S. J. Neal, D. Bhattacharyya, M. Chakkravarthy, and Y.-C. Byun, "3d cnn with visual insights for early detection of lung cancer using gradient-weighted class activation," *Journal of Healthcare Engineering*, vol. 2021, 2021.

[14] T. L. Chaunzwa and et al., "Deep learning classification of lung cancer histology using ct images," *Scientific reports*, vol. 11, no. 1, pp. 1–12, 2021.

[15] P. Chaturvedi, A. Jhamb, M. Vanani, and V. Nemade, "Prediction and classification of lung cancer using machine learning techniques," in *IOP Conference Series: Materials Science and Engineering*, vol. 1099, p. 012059, IOP Publishing, 2021.

[16] S. Dodia, A. Basava, and M. P. Anand, "A novel receptive field-regularized v-net and nodule classification network for lung nodule detection," *International Journal of Imaging Systems and Technology*, vol. 32, no. 1, pp. 88–101, 2022.

[17] I. Naseer, T. Masood, S. Akram, A. Jaffar, M. Rashid, and M. A. Iqbal, "Lung cancer detection using modified alexnet architecture and support vector machine," *Computers, Materials & Continua*, vol. 74, no. 1, 2023.

[18] N. Kalaivani, N. Manimaran, S. Sophia, and D. Devi, "Deep learning based lung cancer detection and classification," in *IOP conference series: materials science and engineering*, vol. 994, p. 012026, IOP Publishing, 2020.

[19] M. Grandini, E. Bagli, and G. Visani, "Metrics for multi-class classification: an overview," *arXiv preprint arXiv:2008.05756*, 2020.

[20] A. Churcher and et al., "An experimental analysis of attack classification using machine learning in iot networks," *Sensors*, vol. 21, no. 2, p. 446, 2021.

[21] H. Alyasriy and A. Muayed, "The iq-othnccd lung cancer dataset," *Mendeley Data*, vol. 1, p. 2020, 2021.

[22] C. I. Henschke, D. F. Yankelevitz, R. Mirtcheva, G. McGuinness, D. McCauley, and O. S. Miettinen, "Ct screening for lung cancer: frequency and significance of part-solid and nonsolid nodules," *American Journal of Roentgenology*, vol. 178, no. 5, pp. 1053–1057, 2002.

[23] R. Smith-Bindman and et al., "Trends in use of medical imaging in us health care systems and in ontario, canada, 2000-2016," *Jama*, vol. 322, no. 9, pp. 843–856, 2019.

[24] S. Prusty, S. Patnaik, and S. K. Dash, "Skcv: Stratified k-fold cross-validation on ml classifiers for predicting cervical cancer," *Frontiers in Nanotechnology*, vol. 4, p. 972421, 2022.

[25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:*, 2014.

[26] J. Pardede, B. Sitohang, S. Akbar, and M. L. Khodra, "Implementation of transfer learning using vgg16 on fruit ripeness detection," *Int. J. Intell. Syst. Appl*, vol. 13, no. 2, pp. 52–61, 2021.

[27] A. M. Qadir, P. A. Abdalla, and M. I. Ghareb, "Malaria parasite identification from red blood cell images using transfer learning models," *Passer Journal of Basic and Applied Sciences*, vol. 4, no. Special issue, pp. 63–79, 2022.

[28] R. Kumar and S. Geetha, "Malware classification using xgboost-gradient boosted decision tree," *Adv. Sci. Technol. Eng. Syst*, vol. 5, pp. 536–549, 2020.

[29] M. S. Al-Huseiny and A. S. Sajit, "Transfer learning with googlenet for detection of lung cancer," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 22, no. 2, pp. 1078–1086, 2021.

[30] H. F. Kareem, M. S. AL-Husieny, F. Y. Mohsen, E. A. Khalil, and Z. S. Hassan, "Evaluation of svm performance in the detection of lung cancer in marked ct scan dataset," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 21, no. 3, p. 1731, 2021.

[31] H. F. Al-Yasriy, M. S. Al-Husieny, F. Y. Mohsen, E. A. Khalil, and Z. S. Hassan, "Diagnosis of lung cancer based on ct scans using cnn," in *IOP Conference Series: Materials Science and Engineering*, vol. 928, p. 022035, IOP Publishing, 2020.

[32] Y. S. Ismael, M. Y. Shakor, and P. A. Abdalla, "Deep learning based real-time face recognition system," *NeuroQuantology*, vol. 20, no. 6, pp. 7355–7366, 2022.