Lung Cancer Detection Using Machine Learning and Deep Learning Models

1st Khattab M. Ali Alheeti
Computer Networking Systems Department, College of Computer and Information Technology
University of Anbar,
Anbar, Iraq,
co.khattab.alheeti@uoanbar.edu.iq

2nd Tabreer T. Al-Shouka
Computer Sciences Department,
College of Computer and Information Technology
University of Anbar,
Anbar, Iraq;
tab21c1002@uoanbar.edu.iq

3rd Saleem Hamad Majeed
Computer Sciences Department,
College of Computer and Information Technology
University of Anbar,
Anbar, Iraq;
saleem.hamad@uoanbar.edu.iq

4th Amjed Abbas Ahmed
Department of Computer Techniques Engineering,
Imam Al-Kadhum College (IKC),
Baghdad, Iraq,

2024 21st International Multi-Conference on Systems, Signals & Devices (SSD) | 979-8-3503-7413-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/SSD61670.2024.10549507

Abstract*— Technology and artificial intelligence play a significant role in improving healthcare and enable tasks to be automated. In addition, the diseases can be better understood and diagnosed faster, saving time and reducing costs. This study examines the impact of transfer learning models on the effectiveness of deep learning models in classifying lung cancer through the analysis of CT scan images. Additionally, it investigates the relative performance of various machine learning and deep learning models, encompassing Support Vector Machine (SVM) and convolutional neural net-works (CNN) such as Incep-tionV3, VGG16, Xception, ResNet50, and MobileNetV2, in the early detection of lung cancer based on CT scan images. The SVM model achieved an overall accuracy of 89% after preprocessing, the proposed approach was applied to five pre-trained models (ResNet50, In-ceptionV3, VGG16, Xception, MobileNetV2) using the dataset: Chest CT-Scan; Among the pre-trained CNN models, the Mo-bileNetV2 model achieved the highest accuracy of 98% and the lowest test loss, indicating it performed the best. The Xception model achieved the second-highest accuracy of 97%. The image pre-processing phase plays a significant role in im-proving system performance in terms of improving image contrast and increasing processing speed.

***Keywords—component, formatting, style, styling, insert** (key words)*

I. INTRODUCTION (*HEADING 1*)

Lung cancer (LC) is one of the most demanding conditions to cure, which has a rise in the rate of death [1]. The (IARC) draws attention to LC's prevalence and critical risk factors. Where the estimates are 2.2 million new instances of LC in 2020. IARC estimates consistently classified Considering lung cancer's prevalence, it's fair to say that it dominates the global cancer scene. even when the proportion of the cases of LC just marginally outweighed those of breast cancer[2]. Early tumor detection is typically crucial for successful therapy. Early malignancies can be effectively inspected to aid patients in a speedy recovery. The capacity of traditional medical imaging methods, like x-rays, CT scans, MRIs, etc., to detect lung cancers is quite limited. We provide a probabilistic deep 2D CNN diagnostic technique for lung cancer utilizing CNN because of its precision in image processing. With the help of our technology, CT scans are quickly evaluated, and calibrated probabilistic ratings are created that precisely reflect uncertainty [3] ,[4]. Cells that multiply unchecked can develop into cancer, a kind of non-communicable illness. Cancer, a type of hereditary illness, has an effect on each cell in a specific region in the body. The body's cells start to divide quickly when receiving therapy for cancer, and they then spread all over the place. The cancer cell interrupts the body's typical cell division process as soon as it be-gins to divide 5. The most prevalent malignancies include colon, lung, breast, prostate, and lymphoma (see Fig. 1)[5].

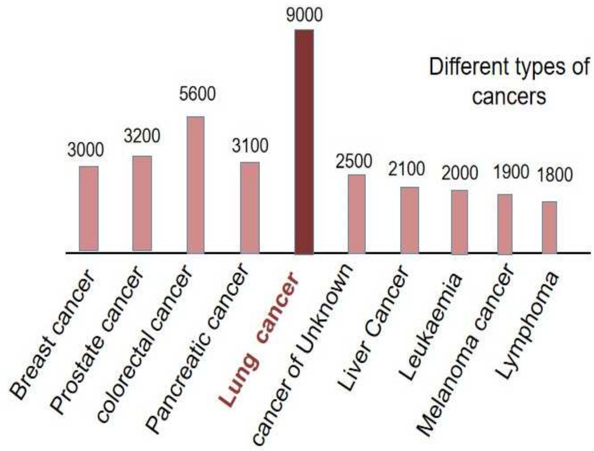
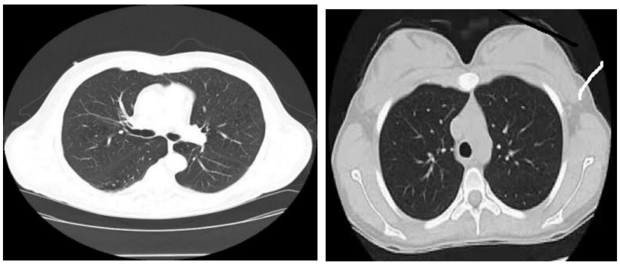


Fig. 1. displays different types of cancers

ML and DL are the applications of scientific techniques that rely on computer algorithms to draw patterns from massive datasets and refine the process over time [6],[7],[8]. Finance, insurance, social media advertising, and real-time data analysis from many resources are just a few of the many fields and applications where algorithms are often utilized [9], [10],[11]. Since patient data is typically not made available to the public, it is challenging to evaluate illness outcomes using these techniques.

Worldwide, lung cancer accounts for more deaths than any other type of cancer combined. The prognosis varies depending on the stage of lung cancer when it is discovered. In essence, there are two forms of LC: small-cell lung cancer (SCLC) and non-small-cell lung cancer (NSCLC). Smoking is strongly linked to an in-creased risk of SCLC. The majority of NSCLC cases involve younger patients and progress more slowly. "Mixed cancer" is the term used to describe the coexistence of tiny and big cells [12]. Cancer is particularly deadly because it frequently advances without showing any signs. A quarter of the population had no symptoms. Lung X-rays are typically used to determine the condition known as LC. However, if lung cancer is not found early on, it can potentially spread quickly. An early lung cancer diagnosis is now achievable because to low-dose computed tomography[13], CT scan. Early LC diagnosis is now feasible. An example of normal and abnormal LC is shown in (see Fig. 2).



(a) normal (b) abnormal

Fig. 2. displays a sample of (a) normal and (b) abnormal lung cancer

A. *Motivations and contributions*

Several methods, including MRI, isotope, X-ray, and CT, can be used to diagnose LC. The two most commonly used anatomical scanning technology in diagnosing different lung ailments are (CT) and chest X-ray radiography. Medical professionals use a computed tomography (CT) scans for a wide range of diagnostic and therapeutic purposes, including disease diagnosis and classification, immediate assessment of morphologic extents, characterization of patterning and severity, and follow-up monitoring of clinical progression and response to treatment.

Our study's main innovations and insights might be summed up as follows:

1- The main contribution is comparing LC to other cancers utilizing graphics ML and DL methods for early identification and diagnosis.

2- In Future research on LC, we also recommended considering the numerous research difficulties, unresolved problems, and prospective solutions.

3- For each dataset, a comparison table with the appropri-ate latest methods and various algorithms' outcomes is shown.

B. *CNN*

A Convolutional Neural Network is one DL technique that can use a photograph as data and assign different items and components in the photo significance (learn weights and biases) And have the ability to distinguish them easily. Compared to alternative categorization techniques, a CNN requires significantly little per-processing [14]. In contrast to fundamental approaches, where filters must be hand-engineered, CNN can pick up on these filtering' characteristics. CNN makes up three different kinds of layers (or "building blocks")

Convolution, pooling, and fully linked layers.

These characteristics are extract features: convolution and pooling. In contrast, the fully linked layer trans-forms the gathered information into the desired output, such as categorization. (see Figure 3), displays the CNN architecture.

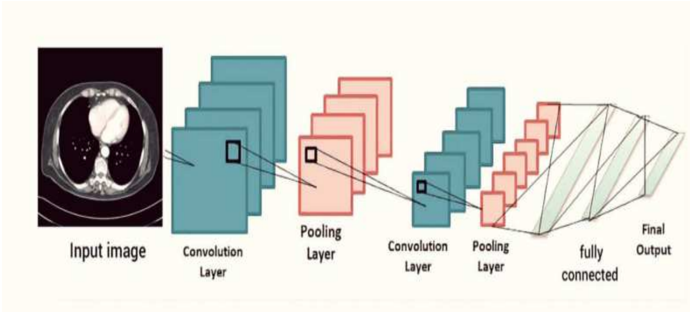


Fig. 3. CNN architecture

C. *Transfer Learning*

Transfer learning is an ML technique that customizes a pre-existing model for deployment on a related task. Pre-trained models are typically trained on vast datasets for specific applications, such as image classification or natural language processing [15]. And their learned features and parameters are then transferred to a new model for a different task. Transfer learning is beneficial when the new task involves a small dataset or requires fast training. Taking advantage of the pre-trained model's knowledge, the new model can frequently achieve superior accuracy and demand less training data than creating a new model from the beginning. There are different types of transfer learning, one of which is fine-tuning. In fine-tuning, the pre-trained model's parameters are updated to adjust for the new task. Another type is feature extraction, where the learned features of the pre-trained model are utilized as input to a new model that is trained for the new task.

D. *ImageNet-Trained Models*

ImageNet-trained models are deep (CNNs) trained on the Image-Net dataset, comprising over a million labeled im-ages from 1,000 object categories. The Image-Net dataset is a popular benchmark for image classification tasks in computer vision. ImageNet-trained models have shown exceptional performance in various computer vision tasks and have become the norm for comparing the efficiency of different computer vision models across research and industry. Typically, ImageNet-trained models consist of numerous layers, ranging from tens to hundreds of layers and millions of parameters. These models utilize convolution, pooling, and non-linear activation functions to ex-tract features from images and acquire intricate representations of objects. Alex Net, InceptionV3, VGG16, Mo-bileNetV2, ResNet, and Xception are well-known ImageNet-trained models. These models have undergone training on the ImageNet dataset using various architectures and techniques, resulting in high accuracy in image classification tasks.

II. Literature review and related

Analyzing the efficacy of three distinct CNN structures using the same parameters reveals the effectiveness of different CNN layer settings. Silva et al. [16] employ the CNN method with the LIDC-IDRI data set. Results show that Precision is 82.3% and Attentiveness is 79.4%. accuracy 83.8%

With the LIDC-IDRI dataset, Zhao et al. [17] used a CNN. This work further investigated the CNN model's performance by changing the learning rate, the size of the kernel, and other factors. LeNet and ALexNet were combined in a hybrid CNN with layer and parameter values, and their accuracy is 88.1%.

The segmentation of pancreatic cysts from an ab-dominal Yuyin Zhou1 et al. [18] focuses on CT images. Because of the low border contrast and the wide range of geometric characteristics, the pancreas is challenging to segment. Without including a human during the testing phase, the approach achieves a Dice Sorensen co-efficient (DSC) of 63.44%, which is appropriate for clinical applications.

Yu Gu et al. [19] created a novel technique using a 3D deep CNN and Algorithms for Multi-Scale detection. To recognize lung nodules to help the radiologists correctly diagnose LC. By using this method, the radiologist can obtain a second opinion on the success of the nodule identification procedure. Additionally, 3D CNN uses spatially rich contextual data to give additional exclusionary characteristics compared to 2D CNN. Multi-scale cube prediction and cube clustering techniques were also used to find the tiny nodules. LUNA 16 datasets are used to assess the proposed approach. The competitive performance indicator of 0.7967 indicated that 87.94% of the techniques were sensitive and were more encouraging.

Hongtao Xie et al. [20] developed a framework for the identification of lung nodules in two dimensions using convolution neural networks to assist in the process of analyzing CT data. Two region-proposal networks, in combination with a de-convolutional layer, are utilized to modify the Faster- RCNN for the first identification of potential nodules. The following step is training three models for three various types of sections. Then, to separate the actual clusters from the candidate clusters, false-positive results reduction is performed using a boost structure built on a 2DCNN. The recommended method's sensitivity was 86,42

The study by M. Praveena et al. [21] advised the classification and forecasting of lateral thorax X-rays to be performed utilizing the Mobile-Net V2 modified technique. Rapid reading, recognition, and consideration of CT images are required. The NIH Chest/Xray 14 database was used to obtain the results, which contained a wide variety of ranges, an Auc of 0.81, and the Accuracy is 90%. Resampling the database significantly improves the model's performance, according to our findings. The objective is to develop a method it might be learned, along with changed components that could be used in smaller IoT devices and require less computing power.

In the paper of Chintakayala Tejaswini et al. [22], they created a method to detect lung cancer. It employs feature extraction and ML methods to put it in another way, employing CNN architecture, which is a DL algorithm that assists in LC diagnosis. SVM and CNN approaches are more accessible than any other Machine-Learning/Deep-Learning algorithms for the LC dataset un-der consideration since both deal with binary categorization, which states whether or not LC exists in an individual's body.

Naive Bayes, Support Vector Machines, Decision Trees, and Logistic Regression were among the classification methods that Seema Babusing Rathod et al. [23] used to investigate how well they could predict lung cancer. The primary purpose of this research is to examine how well classification algorithms perform in the early detection of lung cancer. In the United States, smoking causes roughly 80% of all cases of fatal lung cancer. This investigation has been fruitful in identifying cases of lung cancer. The AUC and accuracy of the auxiliary apparatus are both at 94%.

III. Research proposed

A. Dataset

In my study, I will focus on the most common used datasets, LIDC-IDRI and LUNA16.

The most well-known publicly accessible LIDC-IDRI dataset includes 1018 DICOM photos weighing roughly 124 GB, which require a significant amount of learning and preparation materials and, based on the system, may take several days. These datasets also comprise diagnostic and (CT) scans for LC, along with labeled annotated tumors. It is made up of further than a thousand DICOM images; The DICOM files feature a header that provides the patient's id information in addition to other scan-related information like the slice thickness. The length of the pictures is (z, 512, 512), where z is the amount of CT scan segments, which differ based on the screen's quality [24].

The 888 CTs in the LUNA16 collection are the only ones with homogeneous segments and so well pictures.

Tumor classification was performed using collections of labeled tumors from skilled physicians and test dataset nodules that were acquired after mapping from the LUNA16 dataset. These mappings alone were chosen as the regression Tumor coefficients.

Chest CT-Scan images dataset The Kaggle dataset comprises of 1,000 chest CT scans image, the dataset consists of various images that are stored in the JPEG and PNG format. The total download size of the dataset is approximately 124 MB, as the images are stored in their original uncompressed format. The image formats are converted into JPEG and the image Dimension into (371 x 251) to unify the images' dimensions as a result of the requirement for a substantial dataset in deep learning, the dataset was augmented from 1000 to 10,009 images in the initial scenario, divided to 5003 for normal, and 5006 for abnormal. In the following scenario, the dataset was further expanded to 15,001 images then divided to 7500 for normal, and 7501 for abnormal.

B. Image Preprocessing

In the world of digital imagery, various image processing techniques are employed to enhance and optimize visual content. These techniques play a crucial role in manipulating and improving the quality, appearance, and compatibility of images. Among these techniques, image format, resizing, brightness adjustment, image exposure, image sharpness are key aspects that contribute to achieving desired results. Figure.4 shows a comparison of unprocessed and processed image sample.

- Image Format: An image format refers to the structure and organization used to store and represent digital images. It determines how the image data is stored, compressed, and encoded. Different image formats have varying characteristics, such as file size, image quality, and compatibility with different software and devices. Some common image formats include (JPEG, PNG, GIF, TIFF, and BMP), These are just a few examples of image formats, and there are many other formats available, each with its specific features and use cases. The choice of image format depends on factors like intended use, image complexity, desired quality, and compatibility requirements. different image formats have varying levels of compatibility with different devices and software. Choosing the appropriate image format ensures that the image can be easily viewed and accessed across different platforms.
- Image Resizing: Image resizing refers to the process of changing the dimensions (width and height) of a digital image. It involves scaling the image either up (enlarging) or down (reducing) while maintaining its aspect ratio. it can be performed using image editing software, online tools, or programming libraries. When resizing images, it's important to maintain the aspect ratio to avoid distorting the image. This means that the width and height are scaled proportionally to preserve the original im-age's shape. resizing an image allows it to fit specific dimensions or aspect ratios, making it suitable for various display or printing purposes. It provides flexibility in adjusting the image's size to meet specific requirements. also can improve performance when displaying them on websites or applications. Smaller image sizes result in faster loading times, reducing bandwidth usage and improving user experience.
- Brightness adjustment: Refers to the process of modifying the overall brightness level of an image. it involves increasing or decreasing the intensity of the lightness values across the image to make it brighter or darker, respectively. also can be per-formed using various image editing software, ranging from basic photo editors to advanced tools like Adobe Photoshop. Typically, the adjustment is made by manipulating the image's lightness values or by using a brightness/contrast slider that allows for real-time preview and adjustment. it's important to exercise caution when adjusting brightness, as extreme adjustments can result in loss of detail or introduce artifacts in the image. adjusting the bright-ness of an image can improve its visibility by making dark areas brighter and enhancing details. This is particularly useful in images with poor lighting conditions or low contrast. modifying the brightness can also be used creatively to achieve desired effects, such as creating a bright and vibrant atmosphere or emphasizing certain elements in the image.
- Image exposure: Image exposure refers to the amount of light that reaches the image sensor or film during the process of capturing a photograph. It directly affects the brightness and tonal range of an image. Proper exposure is crucial in photography to achieve the desired level of brightness, contrast, and detail in the captured image. in post-processing, exposure adjustments can be made using image editing software. This involves increasing or decreasing the brightness and tonal values to correct exposure-related issues or achieve desired artistic effects. adjusting image exposure can help correct underexposed or overexposed images. Balancing the expo-sure levels enhances the visibility of details, im-proves color accuracy, and creates a more visually appealing image. By modifying the exposure, the photographer or editor can selectively enhance or reduce the brightness of highlights and shadows, al-lowing for better tonal balance and dynamic range in the image.
- Image sharpness: Image sharpness refers to the clarity and crispness of details in a digital image. A sharp image exhibits well-defined edges, clear textures, and fine details, while a blurry image lacks sharpness and appears soft or out of focus. In post-processing, image sharpness can be enhanced or adjusted using various software tools. Sharpening algorithms are applied

selectively to enhance the edges and details in an image while minimizing the amplification of noise or artifacts. This can help im-prove the perceived sharpness of an image without sacrificing overall quality. However, it's worth noting that while sharpness is generally desired in im-ages, certain artistic effects or intentional blurring techniques may be used to convey a specific mood or style. increasing image sharpness enhances the level of detail and definition, resulting in a clearer and more visually appealing image. This is particularly beneficial for images that appear blurry or lack sharpness due to various factors like camera focus or image compression. Enhancing sharpness can help bring out specific details in an image, making them more prominent and noticeable. This is advantageous in photography, product imaging, or any situation where emphasizing fine details is important.

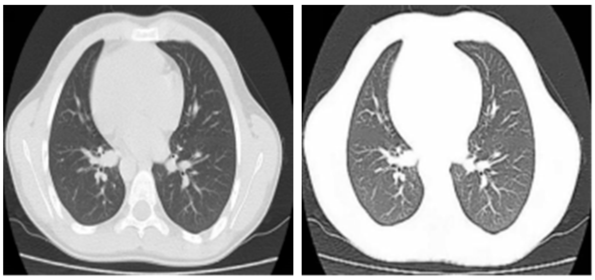


image before processing image after processing

Fig. 4. compares image samples taken before and after processing.

C. Support Vector Machine Results

The SVM algorithm was applied to classify medical images. The accuracy was extracted to evaluate the algorithm's efficiency; Table 1 shows that the SVM model after preprocessing has a higher test accuracy and overall accuracy than the SVM model before preprocessing. This suggests that the preprocessing steps have improved the model's generalization ability to new, un-seen data. The overall accuracy of the SVM model after preprocessing is higher, indicating that the preprocessing steps have improved the model's performance on the entire dataset.

The Table 1 indicates that preprocessing, the SVM model achieved a validation accuracy of 0.7179 and a test accuracy of 0.8287, resulting in an overall accuracy of 0.83. However, after applying the preprocessing techniques, the SVM model's performance in-creased slightly, with a validation accuracy of 0.6762 and a test accuracy of 0.8941, resulting in an overall accuracy of 0.89. This suggests that the preprocessing techniques helped to improve the SVM model's performance on the test dataset.

TABLE I. EFFECTIVENESS OF PREPROCESSING ON SVM MODEL PERFORMANCE.

Model	Validation Accuracy	Test Accuracy	Accuracy
SVM before Preprocessing	0.7179	0.8287	0.83
SVM after Preprocessing	0.6762	0.8941	0.89

D. Transfer Learning Model Results

This section evaluates the transfer learning performance Models for Lung cancer.

Table 2 presents the classification report for transfer learning models and compares their performance in precision, recall, F1-score, support, test loss, test accuracy, and accuracy. The classification report shows the performance of each model in two classes, "Normal" and "Abnormal." The test loss represents the difference between the predicted and actual values for the test set. In contrast, the test accuracy measures the percentage of correctly classified instances in the test set. The accuracy column represents the overall accuracy of the models. Upon examining the data provided in the table, it can be observed that all models performed well in terms of accuracy. The accuracies ranged from (0.94 for Resnet to 0.98 for MobileNetV2).

TABLE II. CLASSIFICATION REPORT FOR TRANSFER LEARNING MODELS: COMPARISON

Model		Precision	Recall	F1-Score	Support	Accuracy
Resnet	Normal	0.96	0.92	0.94	1125	0.94
	Abnormal	0.92	0.96	0.94	1126	
Inception V3	Normal	0.95	0.97	0.96	1125	0.96
	Abnormal	0.97	0.95	0.96	1126	
VGG 16	Normal	0.97	0.96	0.97	1125	0.97
	Abnormal	0.96	0.97	0.97	1126	
Xception	Normal	0.97	0.97	0.97	1125	0.97
	Abnormal	0.97	0.97	0.97	1126	
Mobile Net V2	Normal	0.98	0.98	0.98	1125	0.98
	Abnormal	0.98	0.98	0.98	1126	

E. Experimental Results

The following experimental results present the performance evaluation of a classification model Mo-bilenetV2 across ten different experiments. The aim of these experiments was to assess the accuracy and effectiveness of the model in classifying data into positive and negative categories. The experiments recorded the number of TP, FN, FP, and TN for each run, along with the corresponding accuracy. Table 3 illustrate evaluation of classification performance for the MobilenetV2 mod-el.

TABLE III. EXPERIMENTAL RESULTS OF A CLASSIFICATION MODEL MOBILENETV2

Exp.	TP	FN	FP	TN	Accuracy
1	1101	24	24	1102	98%
2	1090	35	27	1099	97%
3	1097	28	20	1106	98%
4	1092	33	14	1112	98%
5	1096	29	26	1100	98%
6	1096	29	29	1097	97%
7	1099	26	25	1101	98%
8	1085	40	12	1114	98%
9	1099	26	26	1100	98%
10	1096	29	23	1103	98%

Based on the provided experimental results, we can observe the following:

- Overall High Accuracy: The experiments consistently achieved high accuracy levels, ranging from 97% to 98%. This indicates that the models performed well in classifying the data.
- Consistency in True Positives and True Negatives: The number of true positives (TP) and true negatives (TN) remained relatively stable across the experiments, with small variations. This suggests that the models were able to correctly identify the positive and negative instances consistently.
- False Positives and False Negatives: The number of false positives (FP) and false negatives (FN) also showed minor fluctuations across the experiments. While the values were generally low, it is important to consider these errors as they can have different implications depending on the application.
- Comparative Analysis: Experiment 2 and Experiment 6 had slightly lower accuracy compared to the other experiments, with accuracies of 97%.
- High True Positive and True Negative Rates: Across all experiments, the models consistently achieved high true positive and true negative rates, indicating their effectiveness in correctly identifying positive and negative instances.

The results suggest that the models performed well and achieved high accuracy in classifying the data.

F. Discussion

Over the past few years, morbidity and deaths have been rising globally because of LC, posing a severe risk to human life. Find a reliable technique of diagnostic that will allow you to identify patients early, give you a foundation for creating treatment strategies, and help you avoid and manage LC. According to investiga-tions, CNN is used in line with a number of "tumor seg-ment nets," such as the "VGG 16" and "Distended con-volution. The tumor segment process enables scientists to examine the specific site of tumor growth. CNN is among the most widely utilized tools for detecting LC. Due to its superior image processing, rapid output pro-duction, cluster extraction, resolution image, and other qualities, this methodology is the most successful detec-tion method. Since CNN has so many advantages over other ML algorithms, the healthcare sectors presently use it extensively.

In our study, we noticed the use of LIDC-IDRI as data in most of the research, and CNN was considered one of the best DL algorithms specialized in classifying and discovering diseases through images. The emphasis on performance is distinct, and deep CNN-based CT and MRI exams have a great practical significance for LC medical prognosis [25].

The system was implemented using the Colaboratory platform, a cloud-based development environment that provides access to Python programming language with version 3.10.11, TensorFlow library with version 2.12.0, and Keras library with version 2.12.0. The training stage of the system was executed on a GPU provided by Google Cloud, which is available in Colaboratory as a free resource to accelerate machine learning work-loads. The available GPU memory in Colab is typically around 12GB to 16GB, and the use of GPUs can significantly speed up machine learning tasks, especially for deep learning models that involve large datasets and complex architectures.

G. Compassion Between Proposed Algorithm and Pre-vious Studies

This section compares the results of this work with earlier studies. In Table 4, the performance of the suggested system can be observed in comparison to previous re-search. This section presents a comparative analysis be-tween the findings of the present study and those of previous research.

TABLE IV. PRESENTS A COMPARATIVE ANALYSIS OF THE PROPOSED SYSTEM WITH OTHER RELEVANT STUDIES.

Authors	Algorithms	Dataset	Accuracy
Muntasir Mamun 2023	CNN with Inception V3, Xception, and ResNet-50	CT scan images from the Chest CT-Scan	92%
Zebel-E-N. Akhand 2023	CNN model that combines ResNet101 and InceptionV3 models	CT scan images from the Chest CT-Scan	93%
Sajad Dadgar 2023	CNN with transfer learning models ResNetV2	CT scan images from the Chest CT-Scan	91%

Proposed work	CNN with transfer learning models MobileNetV2	CT scan images from the Chest CT-Scan	98%

CONCLUSION

In conclusion, we have looked into lung tumor identification using CNN to explain the current advances and forthcoming issues. The study was picked as one of the most recent studies and published till 2022 in journals with Sober magazine indexes. It has been highlighted that the bulk, with+ just a small number of them using their own datasets and Datasets from LUNA, LIDC, and CT scan images from the Chest CT-Scan were utilized in the research for the recognition of LC. In spite of the fact that many of the current methods for identifying lung nodules have produced findings with high sensitivity or low FP, it is obvious from an analysis of them that a number of problems remain. As a result, there is still a need for the best way to identify LC as soon as feasible. Due to this reason, this research article study will be beneficial to scholars and experts to lower the number of lung cancer patients.

The pre-trained models were performed; the result shows that the MobileNetV2 model has the highest scores of Accuracy 0.98, and Test Loss of 0.0610, which indicates that it served the best overall; the Xception model is ranked second with a score of Accuracy of 0.97, followed by InceptionV3 with a score of Accuracy 0.96, The VGG16 model has a score of Accuracy of 0.96, and finally, the Resnet model has the lowest score Accuracy of, 0.94. The chapter comprehensively com-pares several machine learning models and demonstrates the importance of preprocessing and transfer learning in improving model performance.

REFERENCES

[1] A. A. Nafea, M. S. Ibrahim, M. M. Shwaysh, K. Abdul-Kadhim, H. R. Almamoori, and M. M. AL-Ani, "A Deep Learning Algorithm for Lung Cancer Detection Using EfficientNet-B3," *Wasit J. Comput. Math. Sci.*, vol. 2, no. 4, pp. 68–76, 2023.

[2] S. Metintaş, "Epidemiology of Lung Cancer," in *Airway diseases*, Springer, 2023, pp. 1–45.

[3] A. R. Hmeed, S. A. Aliesawi, and W. M. Jasim, "Enhancement of the U-net architecture for MRI brain tumor segmentation," in *Next Generation of Internet of Things: Proceedings of ICNGIoT 2021*, 2021, pp. 353–367.

[4] A. A. Shah, H. A. M. Malik, A. Muhammad, A. Alourani, and Z. A. Butt, "Deep learning ensemble 2D CNN approach towards the detection of lung cancer," *Sci. Rep.*, vol. 13, no. 1, p. 2987, 2023.

[5] U. Pastorino *et al.*, "Prolonged lung cancer screening reduced 10-year mortality in the MILD trial: new confirmation of lung cancer screening efficacy," *Ann. Oncol.*, vol. 30, no. 7, pp. 1162–1169, 2019.

[6] S. Aliesawi, C. C. Tsimenidis, B. S. Sharif, and M. Johnston, "Efficient channel estimation for chip multiuser detection on underwater acoustic channels," in *2010 7th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP 2010)*, 2010, pp. 173–177.

[7] A. D. Sallibi and K. M. A. Alheeti, "Detection of Autism Spectrum Disorder in Children Using Efficient Pre-Processing and Machine Learning Techniques," in *2023 International Conference on Decision Aid Sciences and Applications (DASA)*, 2023, pp. 505–514.

[8] A. A. Nafea, "Artificial Neural Network and Latent Semantic Analysis for Adverse Drug Reaction Detection Abstract :," *J Comput Sci.*, no. May, 2023.

[9] A. A. Nafea, N. Omar, and M. M. AL-Ani, "Adverse Drug Reaction Detection Using Latent Semantic Analysis," *J. Comput. Sci.*, vol. 17, no. 10, pp. 960–970, 2021.

[10] S. A. Aliesawi, D. S. Alani, and A. M. Awad, "Secure image transmission over wireless network," *Int. J. Eng. Technol.*, vol. 7, no. 4, pp. 2758–2764, 2018.

[11] A. D. Sallibi and K. M. A. Alheeti, "Machine Learning Methods to Detect Autism Among Children," in *2023 Al-Sadiq International Conference on Communication and Information Technology (AICCIT)*, 2023, pp. 224–228.

[12] L. Hussain, S. Rathore, A. A. Abbasi, and S. Saeed, "Automated lung cancer detection based on multimodal features extracting strategy using machine learning techniques," in *Medical imaging 2019: physics of medical imaging*, 2019, vol. 10948, pp. 919–925.

[13] Ö. Günaydin, M. Günay, and Ö. Şengel, "Comparison of lung cancer detection algorithms," in *2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)*, 2019, pp. 1–4.

[14] A. K. Kareem, M. M. Al-ani, and A. A. Nafea, "Detection of Autism Spectrum Disorder Using A 1-Dimensional Convolutional Neural Network," *J Comput Sci.*, vol. 20, pp. 1182–1193, 2023.

[15] A. A. Nafea, M. Mishlish, A. M. S. Shaban, M. M. AL-Ani, K. M. A. Alheeti, and H. J. Mohammed, "Enhancing Student's Performance Classification Using Ensemble Modeling," *Iraqi J. Comput. Sci. Math.*, vol. 4, no. 4, pp. 204–214, 2023.

[16] G. da Silva, A. Silva, A. de Paiva, and M. Gattass, "Classification of malignancy of lung nodules in CT images using convolutional neural network," in *Anais do XVI Workshop de Informática Médica*, 2016, pp. 2481–2489.

[17] X. Zhao, L. Liu, S. Qi, Y. Teng, J. Li, and W. Qian, "Agile convolutional neural network for pulmonary nodule classification using CT images," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 13, pp. 585–595, 2018.

[18] Y. Zhou, L. Xie, E. K. Fishman, and A. L. Yuille, "Deep supervision for pancreatic cyst segmentation in abdominal CT scans," in *International conference on medical image computing and computer-assisted intervention*, 2017, pp. 222–230.

[19] Y. Yang *et al.*, "Deep learning aided decision support for pulmonary nodules diagnosing: a review," *J. Thorac. Dis.*, vol. 10, no. Suppl 7, p. S867, 2018.

[20] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights Imaging*, vol. 9, pp. 611–629, 2018.

[21] M. Praveena, A. Ravi, T. Srikanth, B. H. Praveen, B. S. Krishna, and A. S. Mallik, "Lung cancer detection using deep learning approach CNN," in *2022 7th International Conference on Communication and Electronics Systems (ICCES)*, 2022, pp. 1418–1423.

[22] C. Tejaswini, P. Nagabushanam, P. Rajasegaran, P. R. Johnson, and S. Radha, "CNN Architecture for Lung Cancer Detection," in *2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT)*, 2022, pp. 346–350.

[23] S. B. Rathod and L. Ragha, "Analysis of CT Scan Lung Cancer Images using Machine Learning Algorithms," in *2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2022, pp. 1273–1279.

[24] J. Sang, M. S. Alam, and H. Xiang, "Automated detection and classification for early stage lung cancer on CT images using deep learning," in *Pattern recognition and tracking XXX*, 2019, vol. 10995, pp. 200–207.

[25] P. C. Jacobs *et al.*, "Coronary artery calcium can predict all-cause mortality and cardiovascular events on low-dose CT screening for lung cancer.," 2012.