International Academic Institute for Science and Technology

International Academic Journal of Science and Engineering
Vol.11, No.1, 2024, pp.351-361.

ISSN 2454-3896

International Academic Journal of Science and Engineering

www.iaiest.com

Optimization of Deep Learning Models to Predict Lung Cancer Using Chest X-Ray Images

Esraa Raheem Alzaidi [a]

[a] *College of Since, University of Al-Qadisiyah, Al-Qadisiyah, Iraq. E-mail: Esraa.alzaidi@qu.edu.iq*

Abstract

The study aims to clarify and improve the process of lung cancer diagnosis by improving deep learning techniques, especially neural network technology, given the need to improve the accuracy of lung cancer diagnosis using chest X-ray images, and the enormous potential that deep learning techniques can offer, where learning models can be trained. Deep learning on huge amounts of classified Patients and another group of people without it, as well as demographic data about patients and their medical history, then processing missing or conflicting data, converting imaging images into a format suitable for deep learning, then calculating performance metrics, evaluating the model and then improving it. The results indicated that learning techniques could be improved. Deep research has led to improved ability to predict breast cancer and provide rapid and accurate diagnosis.

Keywords: Deep Learning Models, Prediction, Diagnosis, Optimization, Neural Network.

DOI:10.71086/IAJSE/V11I1/IAJSE1140
Received: 07 May 2024/Revised: 23 May 2024/Accepted: 03 June 2024/Published: 12 June 2024

Introduction

Deep learning techniques are techniques that rely on artificial intelligence and machine learning, through which the work of the human brain is simulated through multi-layered artificial neural networks that have the ability to learn from the data that is fed to them, no matter how complex this data is and no matter how large it is (Raj et al., 2024). Through this, accurate patterns and characteristics of that data are extracted. Learning techniques have witnessed great development in the recent period thanks to the development in computing techniques and database techniques. This development has contributed to improving the accuracy and effectiveness of models and interpretation of data. In the study, data on patients with cancer will be fed into the lungs, the control group and people without lung cancer, the x-ray images of each, and information about the clinical condition and previous medical history, through which accurate patterns and characteristics of that data can be formed (Ausawalaithong et al., 2018; Balaji & Rajakumari, 2024).

Through those patterns and characteristics and comparing them with reality, it is possible to determine how to improve deep learning techniques in the process of predicting lung cancer (Vasamsetti et al., 2023), we must know how or how deep learning techniques work, and how deep learning provides innovative solutions through which we can predict lung cancer in its early stages and diagnose it early, as deep learning techniques depend. It relies heavily on image analysis, such as X-rays and CT scans, by training neural networks on a large set of annotated and illustrated images, as well as data on the clinical condition of these patients and data on their medical history, through which it is possible to discover any pattern that may be strange, no matter how accurate. In lung tissue, it can indicate the presence of a cancerous tumor (Marukatat et al., 2018). Data plays a major role in determining how to improve deep learning. Deep learning is when it analyzes clinical data, patients' medical history, and demographic characteristics of patients or non-affected people and links them with analysis of x-ray images. Certainly, the simulation process will be carried out with the utmost precision, which will ultimately lead to accurate results (Geetha, 2022; Bherje et al., 2024). It is necessary to understand the challenges and obstacles facing these technologies, so that we can deeply understand the challenges, methods and role of deep learning technologies in working and their role in predicting diseases, whether lung cancer or other diseases. We can develop specific strategies or plans to develop and improve these technologies. The most important challenges facing these technologies are the great need for large amounts of data and for this data to be of high accuracy and quality, as well as preserving patients' privacy and ensuring that these models are interpreted accurately and transparently (Harshavardhan et al., 2023). Another of the most important obstacles to these technologies is that they can sometimes be expensive, as they require modern devices that operate with high technology (Souid et al., 2023).

Improving Learning Techniques

Improving learning techniques depends on several different factors. By developing the performance of these factors, the techniques can be improved. Here are some improvements to improve these techniques:

- **Increasing the Size of the Data:** The more the size of the data increases, the more places there are to form more accurate patterns.
- **Increasing the Quality of Data:** The higher the quality of the data, the more accurate the results will be, and data deposit here means reliability and accuracy.
- **Data Diversity:** Data diversity helps control quality and control discrepancies in it, if any.
- **Engineering Technology Development:** Where new neural networks or specialized networks (such as CNN, RNN, and Transformer networks) can be developed and new models can be discovered.
- **Improving Algorithms:** Improving training algorithms such as adaptive optimization (adaptive optimization) and regression algorithms (gradient descent) that can contribute to accelerating and improving improvement learning.
- **Developing the Devices Used:** through developing central processing units and accelerating simulation and analysis processes .
- **Adapting to Different Environments:** By developing continuous learning models without forgetting previous information. Can perform engagements in dynamic and changing environments.

Neural Network Technology

Neural network technology is one of the deep learning techniques and is one of the effective methods for predicting diseases, whether cancer, heart disease, or any other disease. Its basis is analysis of medical images such as X-rays and CT scans for early detection of signs of cancer and identification of a specific pattern that expresses growth and development. Disease. These networks work by feeding them all the medical data and images of the lungs, which

include X-ray images and CT scans, then processing these images and data, then building a neural network model such as CNN or ENN, which is among the specialized networks for image processing. Using the collected data, the model is directed to identify unknown and strange patterns. Through this model, this analysis can be evaluated. One of the most important benefits of using neural networks is the possibility of early detection of some diseases, including cancer, in the early stages, which helps in speeding up diagnosis and increasing the chances of treatment. It is very important in the prediction process, which is based on the strong results resulting from the analysis of images and data that were used. It is a technique that saves time and effort and improves the ability of doctors to diagnose cancers .

Improving Neural Network Technology

Neural network technology has been improved in several aspects. Including increased size, diversity, and accuracy of the data and improving the structure of the neural network by including new layers, changing the depth of the network, or modifying the number of units in the layers. It can also be improved by training the model for longer periods, as it allows it to learn more improved features. More complex, the other method of improvement is:

- **Improving Algorithms:** By improving algorithms to improve the performance of neural networks, such as:
  (A) Parallel Computing Algorithms.
  (B) Iterative optimization algorithms
- **Network Design Improvement Algorithms**
  By determining the appropriate structure of the neural network, improving the depth of the network, and changing the number of layers and the number of units in each layer to improve network performance Such as:

  (A) Transfer Learning.
  (B) Techniques, Regularization Techniques.
  (C) Optimization Algorithms.
- **Training speed Improvement Algorithms**
  These are algorithms that improve the speed of training, such as:

  (A) Mini-batch Gradient Descent: Reduce training time.
  (B) Momentum: Accelerates the optimization process .

Materials and Methods

In this section, individuals, devices, and tools that were used in this paper are clarified.

Study Sample

This study was conducted in one of the medical examination centers, as the center offers a comprehensive medical examination program for diseases, and chest radiography is a basic test used to help detect lung diseases that require further diagnostic workup or treatment. Specifically, the individuals underwent medical examinations for the specialist doctors, after obtaining the consent of the sick and non-sick individuals, collected all chest x-rays from the participants (n=100).

Inclusion Criteria

Inclusion criteria that were applied in the study:

(A) All individuals with posteroanterior chest radiographs.
(B) individuals between 50 and 75 years of age.
(C) C- Current or former smokers.

Exclusion Criteria

Exclusion criteria that were applied in the study:

(A) Patients whose X-ray images contain strange patterns that are not related to cancer, such as enlargements.
(B) Patients who are likely to have cancer but have not been confirmed pathologically or clinically.
(C) Individuals who have unspecified or contradictory information.

Tools and Data

When conducting the study, a set of tools and data was used. We consider a set of tools and techniques. Commonly used in such studies:

- Databases that contain images and clinical information.
- X-rays.
- **Data Processing Tools:** These are tools for processing images, analyzing and cleaning data.
- **Neural Network Technology:** It is one of the deep learning techniques Model building libraries: These are libraries that contain a wide range of tools for building deep learning models, such as PyTorch, tensorflow.
- **Evaluation and Analysis Tools:** These are tools for determining accuracy, sensitivity, and specificity, such as optical precision tools, Grad-CAM.
- Computers.

Cancer Prediction Model

Based on a neural network to predict lung cancer . the model was validated using data from a lung cancer screening trial and the results of previous patients over the previous ten years and the study sample. The inputs and outputs were as follows:

Inputs

- Clinical information (age, gender, smoking status)
- Chest radiograph
- Medical history

Outputs

Prediction Ratio: The model calculates the prediction ratio for the probability of developing lung cancer between 0% and 100%.

Statistical Analyses

We evaluated the performance of the DL model before and after improvement and validation to predict incident lung cancer for the study sample, where cancer risks were divided and determined into:

- Low risk (0% - 20%).
- Moderate risk (21% - 50%).
- High risk (51% - 80%).
- Very high risk (81% - 100%).
- **Indeterminate:** When the results are insufficient to accurately determine the degree of risk.

Then, Pearson tests, receiver operating characteristic curve analysis was performed, and AUROC (Area Under the Receiver Operating Characteristic Curve) values were calculated. As additional diagnostic measures, tests evaluate sensitivity, specificity, and positive . predictive value (PPV) model calibration analysis by plotting observed probabilities against predicted probabilities using the P value, and calculating significance scores and variance using ANOVA.

Results

In this part, we will present and explain the results of the study. The participating individuals were divided into two groups as follows: The first group includes patients with lung cancer, and the second group includes healthy people, with regard to demographic data as shown in Table (1).

Table 1: Shows Explains the Demographic Characteristics of the Study Sample

characteristics	Total number of participants (n = 100)		Lung cancer patients (n = 60)		Healthy people (n = 40)		F	p
	No.	%	No.	%	No.	%		
Sex								
Male	51	51	32	53	19	47.5	-	-
Female	49	49	28	47	21	52		
Smoking								
Male	66	66	52	86.6	14	35	-	-
Female	8	8	7	11.6	1	2.5		
Age (years)								0.0034
Min. – Max.	50.0 – 70.0		55.0 – 70.0		60.0 – 700		5.88	
Mean ± SD.	72.40±3.25		55.09 ±2.65		68. 03 ±2.65			

Median (IQR)	68 (71– 69)	60 (70–58)	69 (72– 69)		

IQR: Inter quartile range ***SD: Standard deviation***
F: F for One way ANOVA test
p: p-value for comparing between the two studied groups

The table shows the demographic characteristics of the study sample in terms of gender, age, and smoking. The total number of the study sample was 100 individuals, including 51 males. The percentage of females was 49% compared to 51% for males. While the percentage of females with lung cancer was 47% compared to 53% of males, while the percentage of smokers out of the total number of the study sample was 66% for males compared to 11.6% for females. The table also shows the general average age of the study sample is 68 years, the average age of lung cancer patients is 54 years, and the average age of healthy controls is 69 years. and the coefficient of variation is 5.88, the P-value is 0.034, and less than 0.05, which indicates the importance of this data and the extent to which the data has statistical significance.

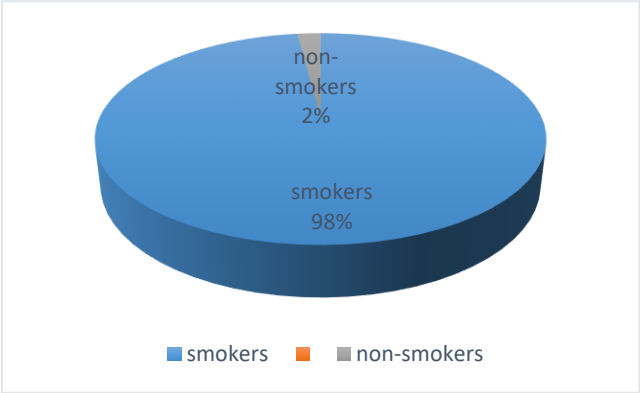


Figure 1: Shows Percentage of Lung Cancer Patients who are Smokers and Non-smokers

The table 2 shows the percentage of smokers out of the total lung cancer patients, as the percentage reached 98%.

Table 2: Lung Cancer Incidence Stratified and Risk Categories through Deep Learning Lung Cancer Risk Prediction Model, Optimization is Reported

			Model Risk Categories							
	n	Smokers	no	low	moderate	high	very high	Indeterminate	F	p-value
			risk	risk	risk	risk	risk			
		%	%	%	%	%	%	%		
pre										
cancer patients	60	86.6	0	6	50	30	11	3	5.88	0.034
Healthy people	40	11.6	100	0	0	0	0	0		
Sig. bet. Grps.			p1=0.038		p2=0.039			p3<0.001		
Post										
cancer patients	60	86.6	0	5	51	31	13	0	12.5	<0.001
Healthy people	40	11.6	99	1	0	0	0	0		
Sig. bet. Grps.			p1<0.001		p2<0.001			p3<0.001		

IQR: Inter quartile range ***SD: Standard deviation*** ***t: Paired t-test***
F: F for One way ANOVA test, pairwise comparison bet. Each 2 groups were done using Post Hoc Test
p: p value for comparing between the two studied groups

The table shows the extent to which the accuracy of the deep learning model was verified in predicting the risks of lung cancer before and after the improvement processes on deep learning techniques, specifically the neural network technique. It is clear from the table that before performing the improvement process, the risk prediction rates were as clear as they were for cancer patients who had a rate For those with a low risk, it was 6%, for those who had a medium risk, it was 50%, for those with a high risk, it was 30%, for those who had a very high risk, it was 11%, while he was unable to determine 3% to determine the risk, For healthy people,There were no risks, as the percentage of no risks was 100%, the value of the variance was 5.88, and the value of the PS value was.034, which means that the data is of great importance and has statistical significance, and after performing the improvement

processes, it was noted that they had a percentage Risks: The appearance of healthy people has a low-risk rate. It was also noted that the p-value decreased to less than one in a thousand and the variance value increased to 12.5, which means an improvement in the accuracy of the data and the results obtained.

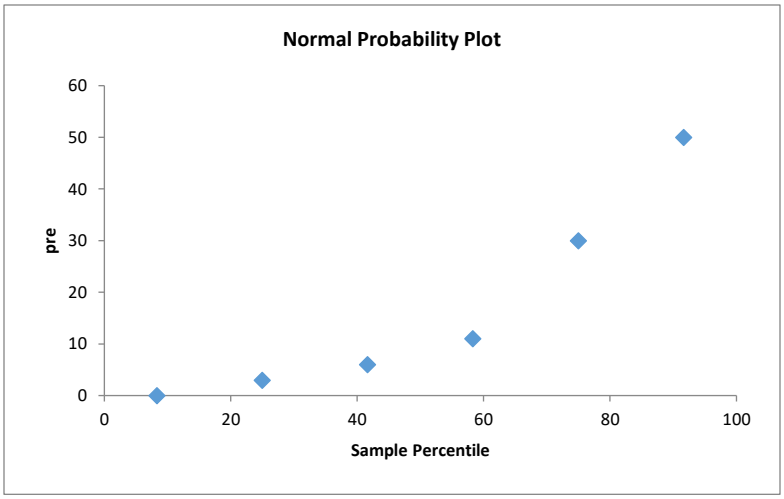


Figure 2: Shows the Normal Probability Plot for the Model Before Improvement

Figure (2) shows a normal distribution chart for the neural network model before improvements were made. It is clear from the figure that the shape takes the shape of a bell, and this indicates that the distribution of the data is normally distributed.

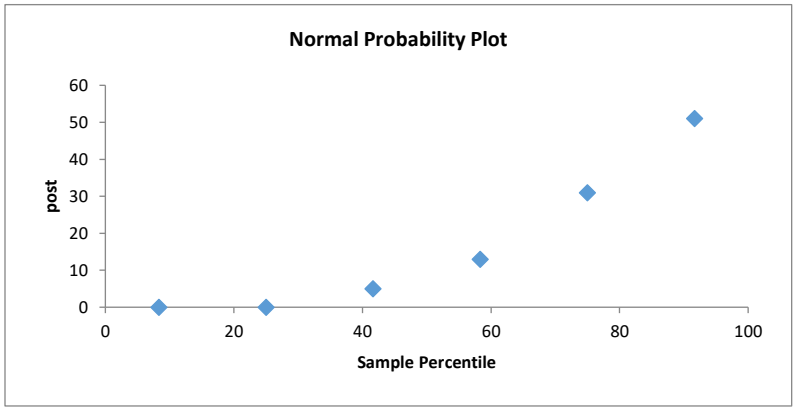


Figure 3: Shows the Normal Probability Plot for the Model After Improvement

Figure (3) shows the normal distribution chart for the neural network model, after making improvements. It is clear from the figure that the shape takes the shape of a bell, and this indicates that the distribution of the data is normally distributed. It also indicates that the data are all normally distributed, as the sample mean is 50 and the sample standard deviation is 10. This analysis is based on the visual information in the image.

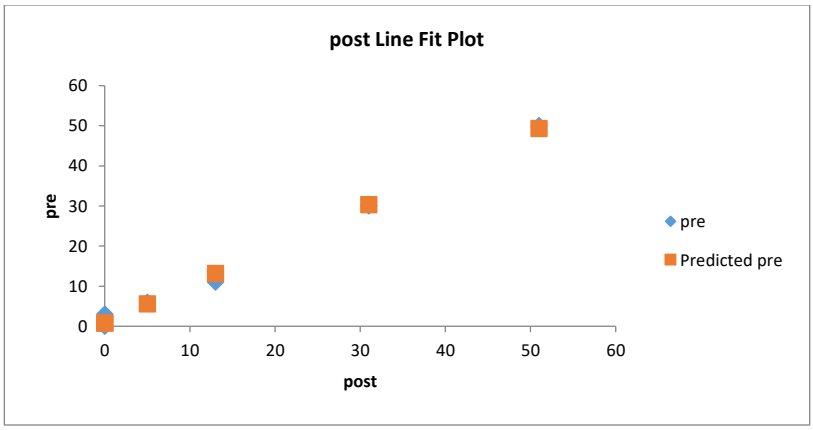


Figure 4: Shows the Normal Probe Fit Plot for the Model After and Before Improvement

Figure No. (4) indicates a linear chart showing the relationship between the values of risk rates before improving the model and after improving the model, where each data point in the chart represents one value for each variable. Number 2, the blue trend line shows the data points.

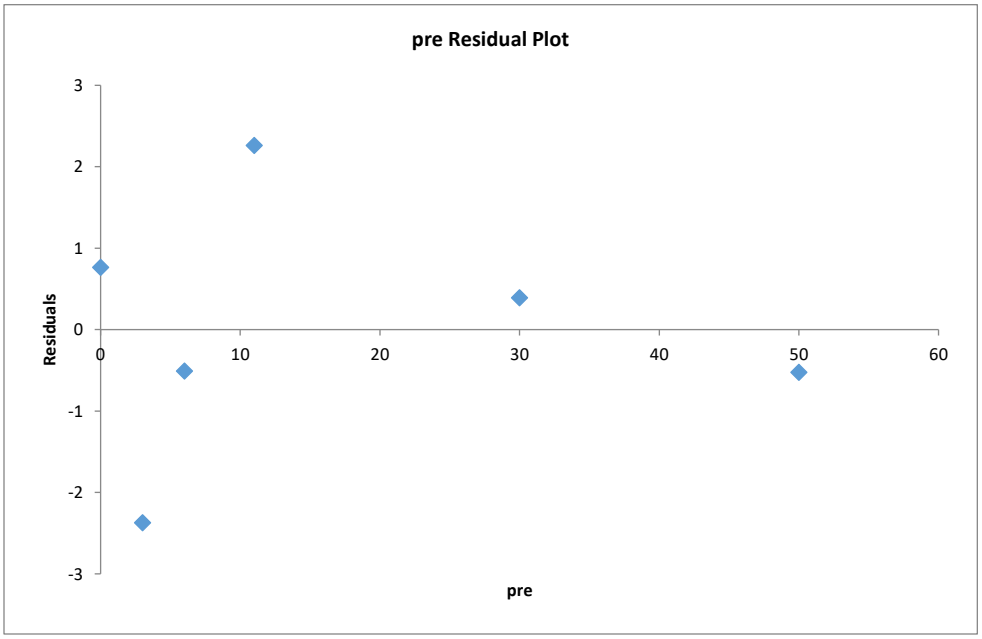


Figure 5: Shows the Residual Before Improvement

Figure (5) graphically shows a chart of the remaining values of the expected risk values before improvement, where the Y-axis represents the remaining values, and the X-axis represents the values that were improved for the expected risks. It is clear from the graph that there are slight differences between the expected values and the resulting values, which means an improvement in the results.

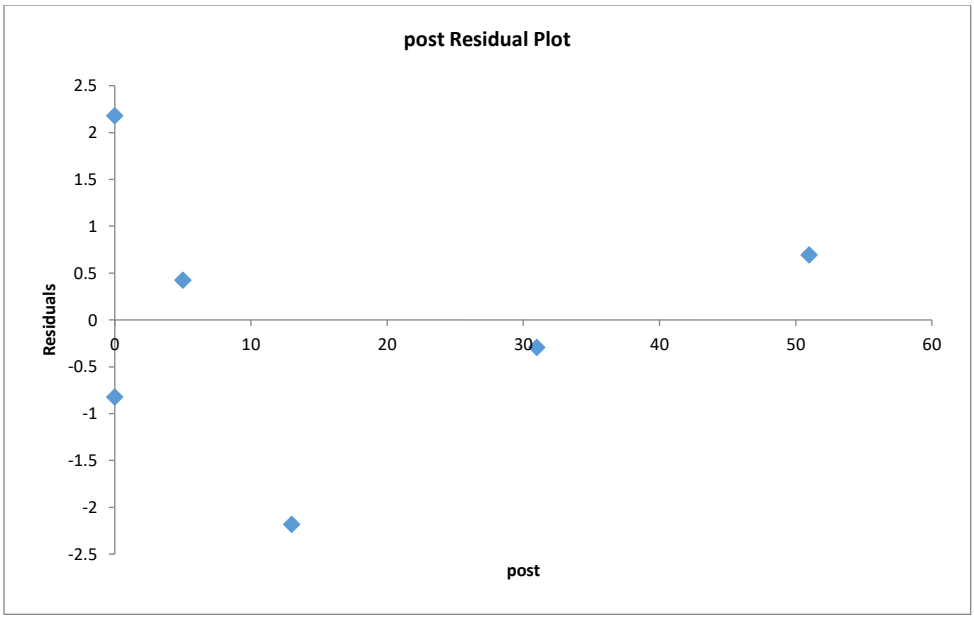


Figure 6: Shows the Residual After Improvement

Figure (6) graphically shows a plot of the residual values for the predicted risk values after improvement, where the Y-axis represents the residual values, and The X-axis is the values that are optimized for the expected risks. It is clear from the chart that there are small differences between the expected values and the resulting values, which means an improvement in the results.

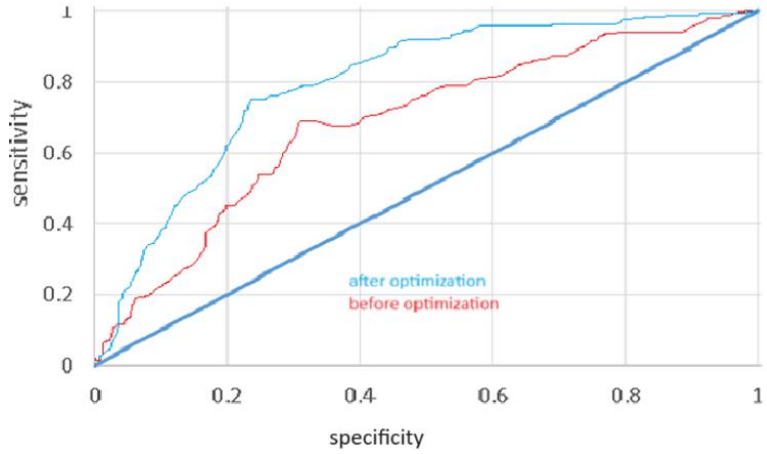


Figure 7: Shows the Relation Between Sensitivity vs Specificity of the Model After and Before Improvement

Figure (7) shows the relationship between the sensitivity and specificity of the model (neural network technology) before and after improvement. It is clear from the figure a significant improvement in the relationship between sensitivity and specificity, where the blue line represents the relationship between them after improvement and the red line represents the relationship between them before improvement. It is clear that the values improved after performing the improvement processes, which indicates on the success of improvement on the success of improvement.

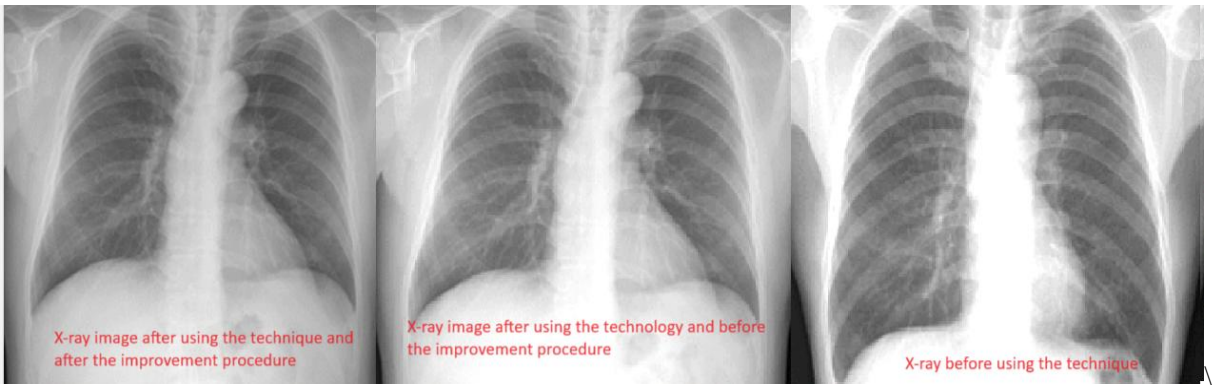


Figure 8: Shows a Comparison between the x-ray Image Before Using the Technology, After Using the Technology, Before the Improvement Procedure, After Using the Technology, and After the Improvement Procedure

The figure 8 shows the x-ray images before they were improved using neural network technology and after using the technology. It also shows the shape of the x-ray images after using the technology before and after improving them.

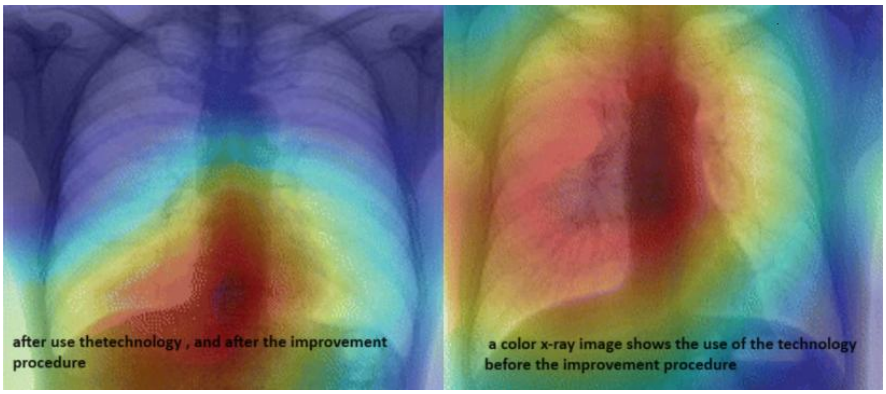


Figure 9: A Comparison between a Color x-ray Image Shows the Use of the Technology Before the Improvement Procedure, After Use, and After the Improvement Procedure

The figure 9 shows two X-ray images: one taken before the model was optimized and one after. The former resembles a faded painting, lacking contrast and color, making it difficult to distinguish details. In contrast, the second X-ray resembles a high-resolution masterpiece, full of vibrant colors and fine detail.

Discussion

Table (1) indicates that is average age of the study sample is 68 years, the average age of lung cancer patients is 54 years, and the average age of healthy controls is 69 years. The coefficient of variation is 5.88, the P- value is It is 0.034, which is less than 0.05, which indicates the importance of this data and the extent to which the data has statistical significance. Figure (1) indicates that is the percentage of smokers out of the total number of patients with lung cancer, which amounts to 98% of the total number of patients with lung cancer.

Table (2) summarizes the effectiveness of a deep learning model, specifically a neural network, in predicting lung cancer risk before and after improvement processes. It reveals that prior to improvements, the model achieved the following risk prediction rates: 6% for low risk, 50% for medium risk, 30% for high risk, and 11% for very high risk. The model was unable to determine risk in 3% of cases. Importantly, it perfectly identified healthy individuals (100% no-risk). Statistical analysis confirms the data's significance with a p-value of 0.034 and a variance of 5.88. Following the improvements, the model displayed a key change: it identified a low risk of cancer in some previously categorized healthy individuals. Additionally, the p-value dropped significantly (below 0.001), indicating a stronger statistical association. The variance also increased to 12.5, suggesting a potential improvement in data accuracy and the resulting predictions.

Figure (2) indicates that the data are all normally distributed, as the sample mean is 50 and the sample standard deviation is 10. This analysis is based on the visual information in the image which refers that a normal distribution chart for the neural network model before improvements were made. Figure (3) indicates that the data are all normally distributed, as the sample mean is 50 and the sample standard deviation is 10. This analysis is based on the visual information in the image which refers that a normal distribution chart for the neural network model after improvements were made.

Figure (4) indicates a linear chart showing the relationship between the values of risk rates before improving the model and after improving the model, where each data point in the chart represents one value for each variable. It also becomes clear that the relationship between the two variables is linear and strong, and this is normal. They are almost the same data, but with an improvement in the value of the data on the risk values that were predicted.

Figure (5) graphically shows a plot of the residual values for the predicted risk values, where the Y axis represents the residual values, the residuals, and the There are no clear patterns in the distribution of points, which indicates that the basic assumptions of linear regression have been disappeared. It is therefore concluded that the model is very good for describing the relationship between the predicted values before optimization and the residuals, and that the data was not distorted and did not require any transformation before or after performing the transformation. Figure (6) graphically shows a plot of the residual values for the predicted risk values, where the Y axis represents the residual values, the residuals, and the There are no clear patterns in the distribution of points, which indicates that the basic assumptions of linear regression have been disappeared. It is therefore concluded that the model is very good for describing the relationship between the predicted values after optimization and the residuals, and that the data was not distorted and did not require any transformation before or after performing the transformation.

Figure (7) shows how the model (powered by neural networks) got significantly better at identifying true positives and negatives. The blue line represents how well it performs after improvement, while the red line shows its performance before. As you can see, the blue line is much closer to the ideal corner, indicating a clear improvement! This means the model can now more accurately distinguish between positive and negative cases.

Figure (8) shows the x-ray images before they were improved using neural network technology and after using the technology as well. It also shows the shape of the x-ray images after using the technology before and after improving them. It becomes clear as soon as the eye sees the extent of the discrepancy and difference in the accuracy and clarity of the existing image. This shows the importance of using neural networks and the importance of improving them to obtain It has good results that can predict the extent of the risk of developing cancer. Also, Figure (9) shows color X-ray images before improving the model after improving the model, and the extent of the difference in contrast, color tone, and image clarity becomes clear. This transformation highlights the success of model improvements. Enhanced X-ray images reveal the extent of the improvement.

Conclusion

One of the most important conclusions that were drawn through this study is that deep learning techniques are very effective tools for predicting diseases, especially cancer and heart disease, or more precisely all diseases that depend on imaging and x-rays, regardless of the type of imaging. On the other hand, although these tools It is effective and important, but there are some challenges facing these technologies, including that they require a large volume of data and great diversity in order for the prediction or prediction models to be accurate, and it is also a fairly expensive technology, both in terms of devices and in terms of the software used to operate those devices. One of the most important conclusions is the possibility of improving deep learning models, especially neural network models, and thus improving the accuracy and quality of images so that we can deduce or predict the severity of the disease, whether it is cancer or other diseases. We also notice from the tables that the percentage of lung cancer among smokers has reached more than ninety-eight percent, and this is a serious indicator of the extent and magnitude of the harms of this bad habit, which leads to death and shortened life expectancy among smokers .

Another of the most important conclusions that can be reached through this study is the extent of the importance of the factors influencing deep learning technology and taking them into consideration carefully while not neglecting any influencing factor of these factors so that we can obtain the best results. Among the most important conclusions is also that neural networks and deep learning techniques it can analyze images and identify specific patterns that indicate the presence of cancer within a few seconds or minutes, while traditional methods take much longer than this. Therefore, it saved valuable time and effort for doctors and paved the way for them towards speedy diagnosis and towards speedy treatment, as well as learning networks. Deep is an effective tool for evaluating the patient's response to treatment and not just predicting the disease. It can also predict the results of treatment with high

accuracy . All of this helps doctors provide a climate for developing methods for treating predicted diseases. It also provides a new opportunity for scientific research in the field of cancer, as it contributes to understanding the causes of cancer and the reasons that may lead to its development. Therefore, it is possible to confirm the effectiveness of regeneration so that we can develop strategies to prevent diseases, and then the other stage, which is the treatment stage. Neural networks can also be used to develop new models to test new drugs and treatments, which leads to improving the effectiveness of treatment. It also opens new horizons for personalized medicine, and through deep learning networks, medical resources can be improved by identifying patients most at risk of developing cancer and directing them to the personalized examinations and treatment necessary to improve results .

References

Raj, S., Kavitha, J. C., & Subitha, D. (2024, November). An Ensemble Model based on Deep Learning for Lung Disease Prediction using Chest XRay Images. *In 2024 8th International Conference on Electronics, Communication and Aerospace Technology (ICECA)* (pp. 1633-1640). IEEE.

Ausawalaithong, W., Thirach, A., Marukatat, S., & Wilaiprasitporn, T. (2018, November). Automatic lung cancer prediction from chest X-ray images using the deep learning approach. *In 2018 11th biomedical engineering international conference (BMEiCON)* (pp. 1-5). IEEE.

Vasamsetti, S., Chemboli, V., Shreyas, G. S. S., & Thota, S. (2023, April). Comparative Performance Analysis of Deep Learning Models for Lung Disease Prediction Using Chest X-Ray Images. *In 2023 International Conference on Inventive Computation Technologies (ICICT)* (pp. 159-165). IEEE.

Marukatat, S., Ausawalaithong, W., Thirach, A., & Wilaiprasitporn, T. (2018, November). Automatic lung cancer prediction from chest x-ray images using the deep learning approach. *In 2018 11th Biomedical Engineering International Conference (BMEiCON).* IEEE.

Geetha, N. (2022). Deep Separable Convolution Network for Prediction of Lung Diseases from X-rays. International *Journal of Advanced Computer Science and Applications, 13(6).*

Harshavardhan, P. S., Kumar, N., & Akhil, M. (2023, July). Multiple Lung Disease Prediction Using Deep Learning. *In 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-7). IEEE.

Balaji, A., & Rajakumari, S. B. (2024, October). Performance Evaluation of Deep Learning Models for the Classification of Lung Diseases in X-Ray Images. *In International Conference on Advancements in Smart Computing and Information Security* (pp. 302-312). Cham: Springer Nature Switzerland.

Bherje, A., Jidge, A., Roy, C., Hulke, A., Aswathy, M. A., Yadav, V., & Veenamol, K. V. (2024, March). Design of deep learning-based approach to predict lung cancer on CT scan images. *In 2024 5th International Conference on Innovative Trends in Information Technology (ICITIIT)* (pp. 1-5). IEEE.

Souid, A., Hamroun, M., Soufiene, B. O., & Sakli, H. (2023). Efficient and Fast Lung Disease Predictor Model. *In Machine Learning and Deep Learning Techniques for Medical Image Recognition* (pp. 90-102). CRC Press.