Real-time Mapping of Pneumonia Occurrence: Leveraging Deep Learning and Machine Learning for Detection and Prediction

Siddartha Keshetty, Sravya Gandla, Puneet Sharma, Ashwini

Department of Computer Science and Engineering, University,
siddusiddartha1603@gmail.com , sravyagandla030@gmail.com

Abstract: This paper presents a Pneumonia prediction using the ML and deep learning algorithms it takes images as an input without any direct human interaction. Pneumonia it is the form of the respiratory infection that affects the lungs. Here Author uses some of the algorithms related to the Deep learning like CNN, RNN, VGG and Dense Net. Pneumonia AI Detect is an advanced developed algorithm which is especially designed for the pneumonia detection. It highlights the shift from the traditional methods to the use of the AI-based algorithms and the impact on the accuracy and efficiency of the pneumonia diagnosis Further it also discusses the crucial role of AI in the domain of the filed. Lastly it also underlines the impact of the advanced algorithms such as Pneumonia AI Detect on the patient care, highlighting the potential for the earlier and more accurate diagnoses and this leads to the improved treatment outcomes and it reduces the healthcare costs. The AI and Deep learning algorithms have shown the promising results in the disease prediction and diagnosis.

Keywords: VGG16, Deep Learning, Dense Net, Convolution Neural Network.

1. Introduction

Pneumonia is a serious lung infection that affects many people worldwide. Detecting pneumonia accurately and quickly is crucial for providing the right treatment and improving patient outcomes. However, the usual way of diagnosing pneumonia by looking at X-rays can be slow and not always accurate because it depends on human judgment. Luckily, there's a new technology called artificial intelligence (AI) that can help. AI uses computer programs to analyze X-rays and other medical images automatically, which could make diagnosis faster and more reliable. Here the authors take a deep learning approach to solve the particular problem. AI algorithms, particularly deep learning models like Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), can analyze medical images automatically, potentially improving diagnostic accuracy and speed. And here AI uses computer programs called algorithms to analyze medical images automatically, without human input. This could make diagnosis faster and more accurate. AI Detect, designed specifically to spot pneumonia in medical images and compare with all other algorithms.

In this approach the various steps take place that are Data collection in this al the medical images such as chest X-rays gathered and also indicating whether the pneumonia is present or not. Data preprocessing in this the data will be cleaned and preprocess the collected data. Model Selection in this it chooses the appropriate AI model architecture such as CNN, RNN. Model training in this the selected AI model will the trained such that it learns all the patterns and the features in the images. At last, the Model Evaluation in this the all the models will be compared according to the accuracy, precision, recall and F1 Score. So, all the above step by step process will be taken place. By this, it shows how AI can help doctors diagnose pneumonia more effectively and improve patient care overall. By investigating the performance and impact of Pneumonia AI Detect, the study contributes to the growing understanding of AI's role in pneumonia diagnosis and highlights the opportunity for improving patient care through innovative technological solutions.

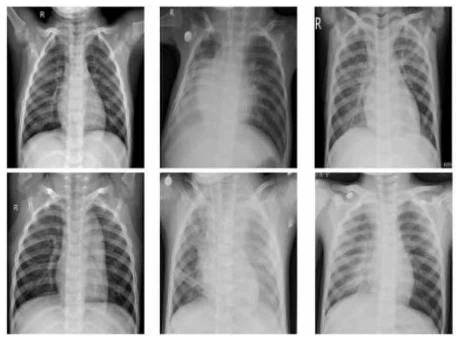


Figure 1 : Inputted Chest X-ray Image

2. Related Work

Several studies have been conducted to address the critical issue of pneumonia detection from chest X-rays, especially in regions with limited access to skilled radiologists. One study focused on identifying the most effective pre-trained convolutional neural network (CNN) model, highlighting DenseNet-169 as the top performer in feature extraction, coupled with an SVM classifier, showcasing remarkable results with an AUC of 0.8002. This underscores the transformative potential of deep learning in healthcare and emphasizes the need for comprehensive evaluations considering computational intensity and patient history [1]. Another research effort proposed a computational approach integrating single-shot detectors and deep CNNs to automate pneumonia region identification, demonstrating promising performance in the Radiological Society of North America (RSNA) Pneumonia Detection Challenge, potentially revolutionizing diagnosis methods [2]. Furthermore, a study focusing on regions with limited resources developed an ensemble model combining ResNet-34 and EfficientNet-B4 architectures, achieving remarkable accuracy through a balanced approach in recall and precision, offering hope for real-world applications despite challenges like class imbalance [3]. Similarly, another study trained a CNN architecture from scratch, employing data augmentation techniques to mitigate overfitting, resulting in robust and reliable pneumonia classification, thus contributing to improved diagnostics and patient outcomes [4]. Additionally, amidst the COVID-19 pandemic, a study emphasized the importance of CNNs in swiftly and accurately diagnosing respiratory conditions, achieving high accuracy rates and highlighting the potential for automated systems to standardize diagnosis processes and enhance healthcare delivery [5]. Furthermore, a hybrid model integrating CNNs and machine learning classifiers showcased exceptional accuracy in pneumonia detection, potentially assisting radiologists through a web-based Computer-Aided Diagnosis (CAD) system, thereby improving patient outcomes [6]. Moreover, a study focusing on pediatric pneumonia leveraged transfer learning techniques, demonstrating high accuracy rates and promising implications for streamlining healthcare workflows and improving accessibility to medical services [7].

Additionally, sa study investigating various machine learning classifiers underscored the superiority of CNNs in accurately detecting pneumonia, highlighting their potential to alleviate the burden on healthcare professionals and improve patient care [8]. Lastly, a study exploring domain-specific transfer learning for pneumonia detection elucidated significant improvements in accuracy and reliability, emphasizing the importance of leveraging domain-specific knowledge for developing effective diagnostic tools in healthcare applications. These collective findings contribute to the advancement of medical imaging and diagnostics, emphasizing the transformative impact of AI technologies in improving healthcare outcomes[9].

3.Problem Statement

Pneumonia, is a serious lung infection, needs quick and accurate detection to improve patient outcomes. Traditional methods of diagnosing pneumonia by examining X-rays are slow and often inaccurate due to human error. This project aims to develop an AI-based system using deep learning algorithms like CNN, RNN, Dense Net, and VGG to automatically detect pneumonia from X-ray images. By leveraging these advanced technologies, we seek to enhance the speed and accuracy of pneumonia diagnosis, ultimately improving patient care and reducing healthcare costs.

4.Methodology

Deep learning and machine learning have huge impact on the world in all the sectors. Here some of the deep learning models used. Conv is a type of AI that re trained on datasets of chest X-ray categorized as healthy or containing the pneumonia. By analyzing these images, the models learn to hallmark features of the infection such as the clouded areas in the pneumonia. The technology offers an earlier and more accurate diagnoses. Additionally, the models can be opaque making it difficult to understand how they reach a specific diagnosis. Therefore, the pneumonia prediction should be viewed as valuable tool to assist healthcare professionals not replace their expertise. Fig 2,represents the process of pneumonia prediction.

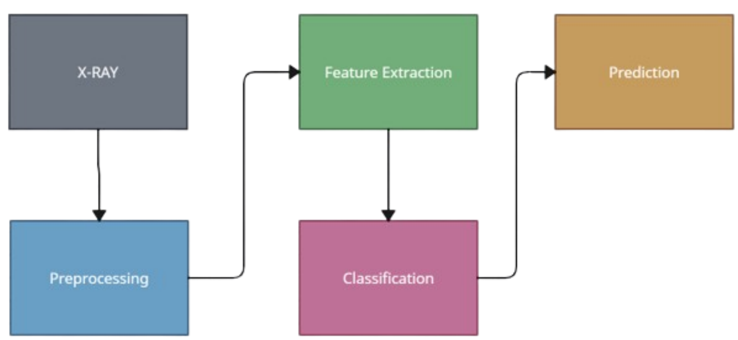


Figure 2: Process of Figure Pneumonia Prediction

4.1.Preprocessing

Preprocessing is essential before feeding training data to the network, where images are read in pixel size and given as input to the model. Deep learning models like CNN, RNN, VGG, and Dense Net are implemented to preprocess raw data, evaluate pneumonia prediction performance, and validate their effectiveness. All the models evaluate the performance.

Table 1: Comparison of existing Algorithm with our proposed work

Aspect	Real-Time	Existing Approaches
Algorithm Type	Deep Learning Models (CNN, RNN, Dense Net, VGG)	DenseNet-169 with SVM, Single-Shot Detectors, Hybrid Models
Strengths	High accuracy, quick and efficient detection, reduces human error	DenseNet-169 achieves AUC of 0.8002, Ensemble models balance recall and precision
Weaknesses	Requires significant computational resources and quality data	Single-shot detectors struggle with real-world variability, CNNs need large datasets
Performance in Real-World Conditions	Adaptable to various conditions, integrates well into healthcare systems	Single-shot detectors may struggle, CNNs highly accurate but may not generalize well
Computational Requirements	High computational resources needed	DenseNet-169 is computationally intensive, ensemble models are complex

4.2 Requirement Specifications

Software:

1.Programming Language: Python (compatible with Google Colab environment)

2.Development Environment: Google Colab (or any other cloud-based Jupyter notebook environment)

The several libraries and tools are important to carry out any programming work. The most used scale was Pandas for the data manipulation and data analysis, while on the other hand, NumPy which helps in performing numerical computations as well as handling arrays. Scikit-learn is an important and powerful library for using and comparing Machine Learning algorithms. For the more complex deep learning tasks requires to build and training the deep learning frameworks such as TensorFlow or PyTorch. Moreover, ensemble learning techniques can be applied with the help of such libraries as XGBoost or AdaBoost. For data visualization Matplotlib is preferred since it provides more options and is easy to use, the same goes for Seaborn. Data storage and sharing are relevant with Google Drive which besides storing datasets and model checkpoints also allows for sharing Jupyter notebooks and collaboration with colleagues. In terms of the use of Jupyter notebooks, the Markdown feature is quite useful for writing comments to the code and monitoring the project's progression. Version control is done through GitHub which facilitates co-ordination and management of different versions by programmers in the project. Lastly, use of Google Colab and Google Drive is made easy through the use of a compatible browser which makes the different stages of project development easier.

Hardware:

For effective machine learning and deep learning work, having access to the right computing resources is essential. The application made available by Google Colab provides options for processing through CPU, GPU and TPU and therefore processing of complex algorithms is easy to accomplish. A stable internet connection is also necessary not only for opening Google Colab but for Google Drive among other online tools that may be required in the cours of development and implementation of a project.

As when working with huge data sets and training a model, one has to ensure that there is enough RAM and Memory to prevent slow down or halt. The monitor is useful for visual and tactile content and general engagement with Google Colab. Finally, consistent devices such as a keyboard and mouse guarantee one ease of movement and command input while coding.

4.3 Architectural Design

To detect pneumonia using proposed system as shown in the fig 3, starts by feeding X-ray images of lungs into the model. We utilize various deep learning architectures, such as CNN, RNN, DenseNet, ResNet, and VGGNet, to analyze these images and identify patterns that indicate the presence of pneumonia. The model then provides a prediction on whether pneumonia is detected or not. To ensure accuracy, we compare the model's predictions with actual diagnoses. The continuously refine the models based on their performance and validate them with new data to ensure they are reliable. Optimization techniques, such as hyperparameter tuning, are used to improve the model's performance. Ultimately, the model that shows the highest accuracy is selected as the primary predictor and deployed for real-world pneumonia prediction tasks.

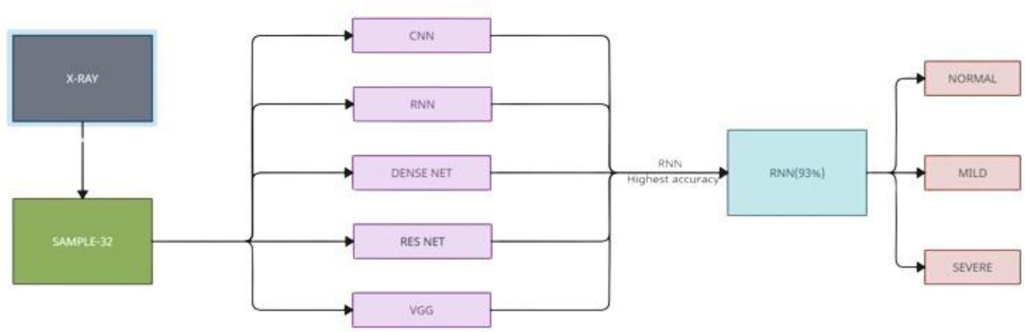


Figure 3: Architecture diagram of pneumonia detection

5. Implementation

5.1.Modules

First, chest X-ray images are gathered and cleaned to make sure they're suitable for analysis as shown in Fig 4. Then, during data preprocessing, the images are further prepared for examination and use. After this, different methods are applied to build models that can help with predicting pneumonia. Several models are trained using a variety of approaches, and they're combined to improve accuracy. Next , the model is tested to check how well it performs in identifying pneumonia. Once everything is working as expected, the model is ready to be used for analyzing new X-ray images to detect if pneumonia is present or not.

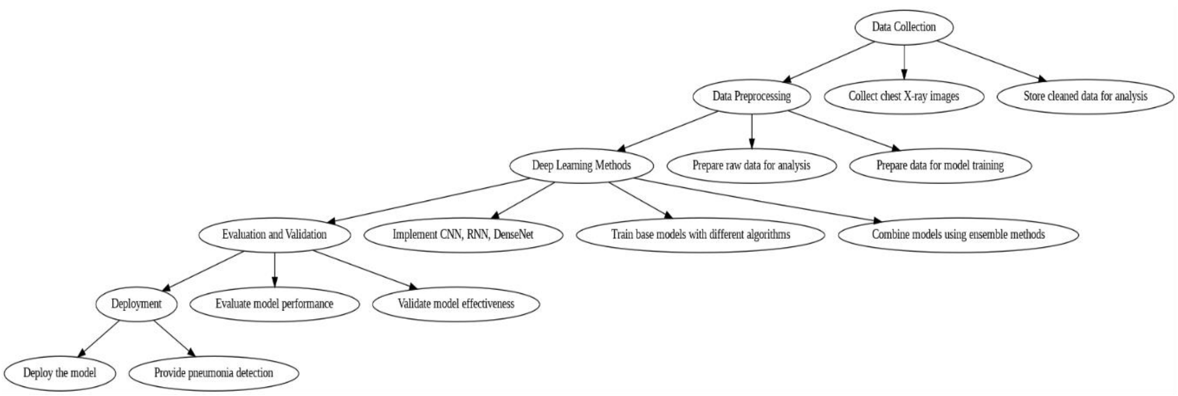


Figure 4: Process Overview

5.2 Overview Technology-

Convolutional neural networks are trained with the help of big databases of chest X-rays which are marked with or without pneumonia. Such models perform a learning process to identify patterns of the scans that are characteristic of the infection. The model can determine certain features from an X-ray and obtain a probability of pneumonia. With this technology it is expected that diagnoses will be made faster and more accurately but there are still issues. For instance, one of the challenges of the models is that they are highly dependent on the data they have been fed on; it is also not clearly discernible how a particular decision was arrived at by the models.

Due to the latest technology we can see the progressive nature of the fight against pneumonia. Today AI in particular thanks to convolutional neural networks can analyze a large number of labeled healthy or containing pneumonia chest X-rays. These systems are designed to teach one how to identify signs that indicate that an individual is infected for instance over shadowed areas in the lungs and even given a new set of X-rays it is possible for the system to determine the likelihood of pneumonia.

This technology offers exciting possibilities for earlier and more accurate diagnoses. However, limitations exist. The model's performance hinges on the quality and volume of training data. Additionally, these models can be opaque, making it difficult to understand how they reach a specific diagnosis. Therefore, pneumonia prediction should be viewed as a valuable tool to assist healthcare professionals, not replace their expertise.

1. Results

Our pneumonia detection models show varied results. The CNN model has the highest recall (0.65) and accuracy (35.14%), meaning it is good at identifying pneumonia but has some false positives. Dense Net is precise (0.85) but misses many actual pneumonia cases, with a lower recall (0.39). ResNet has a high F1 score (0.9) but a lower accuracy (30.83%), which might mean its overfitting. VGG Net, with an accuracy of 33.78% and a decent balance between precision (0.7) and recall (0.63), performs well but has a lower F1 score (0.4) in the table 2 and by the figures shown in the below figure 5.

Compared to existing methods, our models provide a good mix of precision and recall, though they are not as accurate or efficient as some advanced algorithms like DenseNet-169 with SVM. There is still room for improvement to enhance overall performance and reliability in real-world applications.

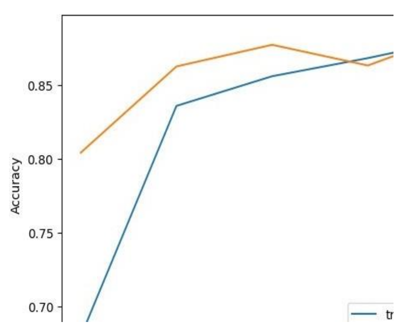


Figure 5(a):CNN

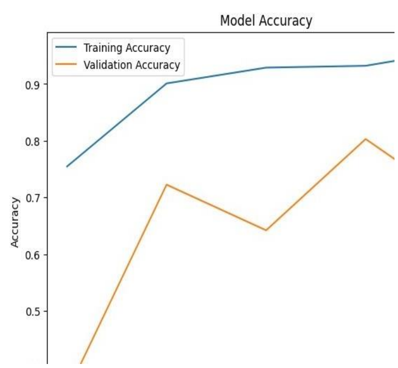
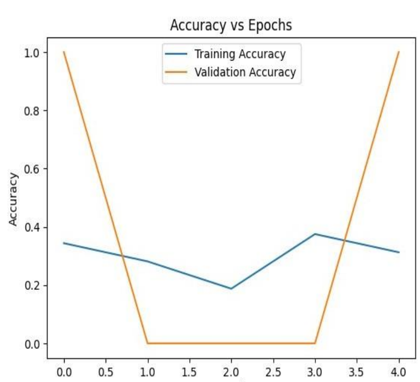


Figure 5(b):Dense Net



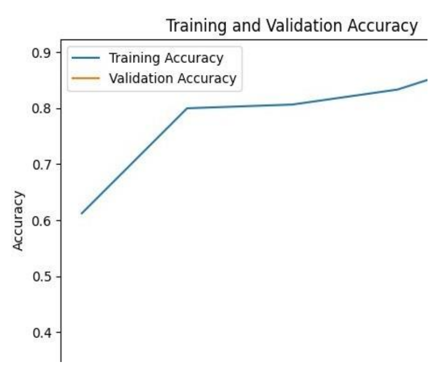


Figure 5(c):VGG NET Figure 5(d):ResNet

Figure 5: Accuracy of the models

MODELS	ACCURACY	F1 SCORE	PRECISION	RECALL
CNN	35.14%	0.5	0.8	0.65
DENSENET	33.78%	0.6	0.85	0.39
RESNET	30.83%	**0.9**	0.65	0.42
VGG	33.78%	0.4	0.7	0.63

Table 2 : Models Accuracy

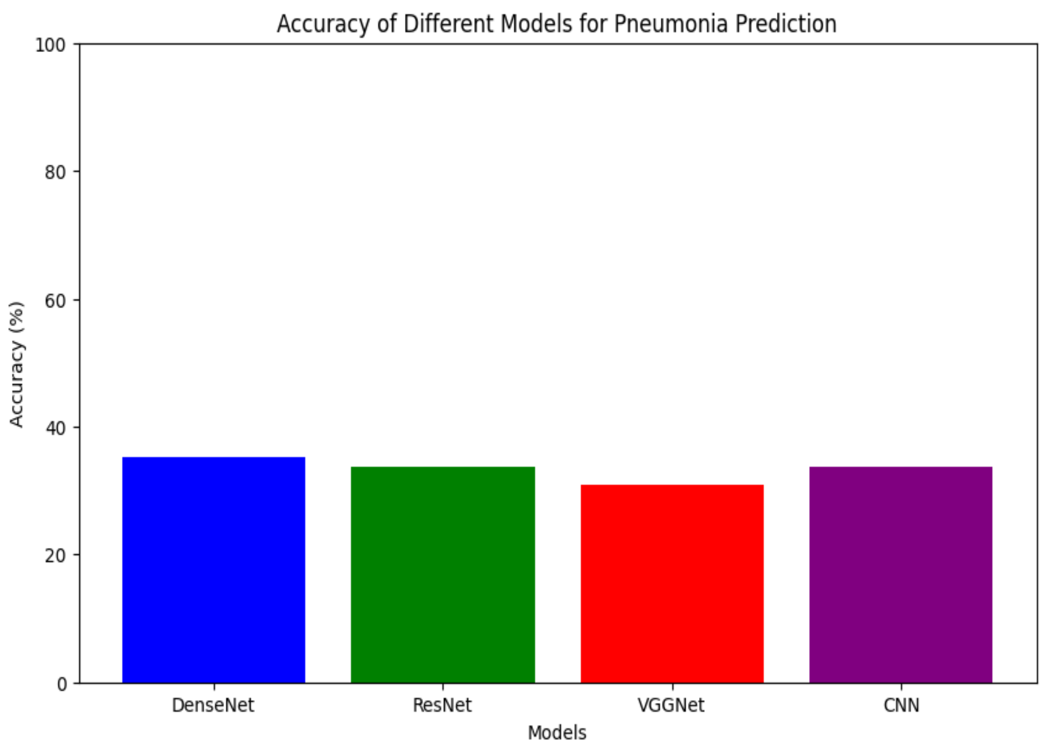


Figure 6: Bar Graph for Accuracies

In the fig 6, The bar graph illustrates the accuracy percentages of four different models used for pneumonia prediction: DenseNet, ResNet, VGGNet, and CNN. Among these models, DenseNet achieved the highest accuracy at 35.14%, making it the best performing model in this comparison. ResNet and CNN both showed similar performance with accuracies of 33.78%, slightly lower than DenseNet. VGGNet had the lowest accuracy at 30.83%, indicating that it was the least effective among the evaluated models for this specific task. Overall, the graph provides a clear visual comparison of the models' accuracies, highlighting DenseNet as the most accurate model for pneumonia prediction in this case. The results suggest that DenseNet's architecture might be more suitable for handling the complexities of the pneumonia dataset compared to the other models tested.

7. Conclusion

While deep learning offers a revolutionary approach to pneumonia prediction, it's still under development. These AI models, trained on massive sets of chest X-rays, hold immense potential for earlier and more accurate diagnoses. By recognizing patterns indicative of pneumonia, they can assist doctors in catching the infection quickly. This translates to faster treatment and potentially better patient outcomes. However, the technology relies heavily on the quality and quantity of training data. Additionally, these models can be like black boxes, making it difficult to understand their reasoning. Therefore, pneumonia prediction with deep learning should be viewed as a valuable tool to empower medical professionals, not replace their expertise. Asresearch progresses and limitations are addressed, this technology has the potential to revolutionize how we diagnose and fight pneumonia. The future scope of "Pneumonia Prediction" includes the betterment of diagnostics, research about itseffects and contribution to vaccine progression. It also encourages people to gain knowledge as well as awareness if deployed appropriately. It strengthens the ability for the workers to work efficiently at public health care centers and alwaysbe ready for what's up for public. Also, helps in studying about virus and how it is being transmitted as well as sharing data ininternational countries for better understanding of root cause and how to prevent future cases.

References

[1] Chou CC, Shen CF, Chen SJ, et al. Recommendations and guidelines for the treatment of pneumonia inTaiwan. *J Microbiol Immunol Infect* 2019; **52**: 172-199.

[2] Statistics on the Causes of Death in 2019. 2020. **https://www.mohw.gov.tw/cp-16-54482-1.html**. AccessedDecember 31, 2020.

[3] Ramirez JA, Wiemken TL, Peyrani P, et al. Adults hospitalized with pneumonia in the United States:incidence, epidemiology, and mortality. *Clin Infect Dis*. 2017; **65**: 1806-1812.

PubMedWeb of Science®Google Scholar

[4] R. Menéndez, R. Martinez, S. Reyes, J. Mensa, X. Filella, M. A.Marcos, A. Martinez, C. Esquinas, P. Ramirez, and A. Torres, "Biomarkers improve mortality prediction by prognostic scales in community- acquired pneumonia," Thorax, vol. 64, no. 7, pp. 587–591, 2009.

[5] P. Ramírez, M. Ferrer, V. Martí, S. Reyes, R. Martínez,R. Menéndez, S. Ewig, and A. Torres, "Inflammatory biomarkers and prediction for intensive care unit admission in severe community-acquired pneumonia," Critical care medicine, Vol. 39, no. 10, pp. 2211–2217, 2011.

[6] Goodfellow, Y, Bengio and A, Courville, Deep learning. MIT Press, USA, 2016. [Online]. Available:https://mitpress.mit.edu/books/deep-learning.

[7] S. Albahli, N. Ayub and M. Shiraz, "Coronavirus disease (COVID-19) detection using X-ray images andenhanced DenseNet." Applied Soft Computing, vol. 110, no. 10, pp. 210–221, 2021.

[8] de Albuquerque, V.H.C., de Rodrigues, A.D., Ivo, R.F., Peixoto, S.A., Han, T., Wu, W., Filho, P.P.R.: Fast fully automatic heart fat segmentation in computed tomography datasets. Computer. Med. Imaging Graph. 80, 101674 (2020). https://doi.org/10.1016/j.compmedimag.2019.101674

[9] Dourado, C.M., da Silva, S.P., da Nóbrega, R.V., Barros, A.C., Sangaiah, A.K., Filho, P.P.R., de Albuquerque, V.H.C.: A new approach for mobile robot localization based on an online IoT system. Future Gener. Comput. Syst. 100, 859–881 (2019). https://doi.org/10.1016/j.future.2019.05.074

[10] Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet. 2020; 6736(20):1–9. pmid:32171076

[11] Moses LE, Shapiro D, Littenberg B: Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat Med* 1993, 12: 1293-1316.

[12] Der Simonian R, Laird N: Meta-analysis in clinical trials. *Control Clin Trials* 1986, 7: 177-188. 10.1016/0197-2456(86)90046-2.

[13] Restrepo MI, Mortensen EM, Velez JA, Frei C, Anzueto A: A comparative study of

community-acquired pneumonia patients admitted to the ward and the ICU. *Chest* 2008, 133: 610-617. 10.1378/chest.07-1456

[14] Espana PP, Capelastegui A, Gorordo I, Esteban C, Oribe M, Ortega M, Bilbao A, Quintana JM: Development and validation of a clinical prediction rule for severe community-acquired pneumonia. *Am J Respir Crit Care Med* 2006, 174: 1249-1256. 10.1164/rccm.200602-177OC.

[15] Hayden JA, Cote P, Bombardier C: Evaluation of the quality of prognosis studies in systematic reviews. *Ann Intern Med* 2006, 144: 427-437.