Explainable Deep Learning Model for Pneumonia Diagnosis

[1]Gayatri Betageri, [2]Shantala Giraddhi
[1] KLE Technological University Hubbali,Karnataka,India ,
[2] KLE Technological University Hubballi,Karnataka,India
[1]betagerigayatri@gmail.com,[2] shantala@kletech.ac.in

Abstract— Pneumonia diagnosis poses a critical challenge, particularly in resource-constrained settings where access to skilled radiologists is limited. This research addresses the problem using two deep learning models: a Convolutional Neural Network (CNN) integrated with Local Interpretable Model-Agnostic Explanations (LIME) and transfer-learning-based InceptionV3 model coupled with Gradient-weighted Class Activation Mapping (Grad-CAM). The CNN model achieves an accuracy of 93.44%, offering reliable classification with localized interpretability through LIME, which highlights the specific image regions influencing predictions. Meanwhile, the InceptionV3 model outperforms with a 95% accuracy, a precision, recall, and F1-score of 97%, demonstrating its robustness for pneumonia detection. Grad-CAM enhances the interpretability of InceptionV3 by providing heatmaps that visually identify pneumonia-affected regions in chest X-rays. By combining high diagnostic accuracy with interpretable AI techniques, this study establishes that both models are highly effective for pneumonia diagnosis, with InceptionV3 and Grad-CAM being particularly well-suited for clinical adoption. This approach ensures transparency, reliability, and scalability, making it a promising solution for healthcare applications and hence enhancing the quality of life in the society.

Index Terms—Chest X-ray analysis, *CNN, Deep learning, InceptionV3, LIME*

I. INTRODUCTION

Numerous pathogens, including bacteria, viruses, and fungi, can cause pneumonia, a fatal lung disease. This illness causes the lung's alveoli, or air sacs, to become inflamed, which frequently leads to the buildup of fluid or pus. A persistent cough with mucus, fever, chills, and trouble breathing are some of the symptoms of pneumonia, which can range from minor to fatal. Serious problems are much more likely to occur in people who are already at risk, such as young children, the elderly, and those with compromised immune systems or underlying medical conditions. Notwithstanding improvements in medical care, pneumonia continues to rank among the world's leading causes of death, especially in low-resource environments where early detection and treatment can be difficult. Diagnosing pneumonia often involves a combination of clinical evaluations and diagnostic tests. In addition to performing a physical examination and evaluating the patient's symptoms, medical professionals frequently use sputum cultures, blood tests, and chest X-rays as supplementary tools. Because chest X-rays can show lung anomalies suggestive of infection, they are regarded as the gold standard for diagnosing pneumonia. However, interpreting chest X-ray images is a complex process that requires significant expertise, and misdiagnoses are not uncommon. This challenge is particularly pronounced in underdeveloped regions, where there is a shortage of skilled radiologists. Additionally, pneumonia can be classified according on the source of the infection Community acquired pneumonia (CAP) occurs outside of healthcare facilities, whereas hospital-acquired pneumonia (HAP) is contracted in medical settings and is often associated with antibiotic-resistant bacteria. Viral pneumonia, while typically less severe than bacterial forms, can still pose significant risks, as demonstrated during the COVID-19 pandemic. In order to stop the progression of pneumonia and guarantee prompt treatment, early detection is essential. This underscores the need for efficient, accurate, and accessible diagnostic tools, particularly in resource-constrained environments. Recent developments in AIML have demonstrated enormous promise for automating the processing of medical images and enhancing diagnostic precision. A specific class of DL models called CNNs has been extensively used for medical imaging tasks, such as the identification of pneumonia. These models excel in recognizing intricate patterns and abnormalities within X-ray of the chest images, often achieving diagnostic performance comparable to that of expert radiologists. However, the application of AI in clinical settings is not without challenges. One of the most significant barriers is the lack of transparency in deep learning models, often referred to as their "black box" nature. This opacity raises concerns among clinicians, who require interpretable results to validate diagnoses and ensure patient safety. To address these challenges, the field of explainable AI (XAI) has gained significant attention. XAI focuses on enhancing the transparency of AI models by offering clear and understandable explanations for their predictions. By using methods such as Grad-CAM and LIME, medical professionals can see which parts of a X-ray of the chest affected the model's judgment. By providing these visual explanations, these methods help clinicians trust and validate AI predictions more confidently. This study explores the integration of XAI methods with DL models for pneumonia detection. The study uses the InceptionV3 model combined with LIME and Grad-CAM to not only achieve high levels of diagnostic accuracy but also ensure that the decision-making process is transparent. The goal is to combine the power of automated diagnosis with clear interpretability, making it easier for AI systems to be adopted in clinical settings while

enhancing their reliability and trustworthiness.

II. LITERATURE SURVEY

A. BACKGROUND STUDY

The main techniques for initially diagnosing pneumonia are imaging testing and medical history. Examples include X-rays [6], MRI [5], CT [4], and other imaging methods. X-ray of the chest are commonly used in clinical settings due to their low cost, and doctors usually use them for manual diagnosis [7]. However, manual diagnosis has a wide range of accuracy rates and a significant degree of subjectivity because 3 it takes a high level of professionalism and clinical experience, and because people are prone to visual fatigue [8]. Therefore, computer-aided diagnostics can speed up the diagnosis of pneumonia. DL is being used in the medical industry as a result of the quick advancements in computers.

Rajpurkar et al. [1] developed CheXNeXt. It is a convolutional neural network. It detected fourteen different pathologies, that includes pulmonary masses, pneumonia, nod ules and pleural effusion. It utilizes the front view of chest radiographs. A convolutional neural network with 121 layers is used in this architecture. In this study, the F1 metric approach was used to compare the performance of a 121 layer CNN and a radiologist. He used 20% of the images in the dataset for validation, 10% for testing, and 70% are used for training. Then this model could score an F1 score of 0.435. The accuracy of radiologists was 0.385.

Four models were created by Nada M. Elshennawy et al. [2] as part of a framework for pneumonia identification. Using techniques like image rotation, skewing, and shifting, he exploited the Kaggle dataset to expand the dataset's size. After resizing and normalizing then extract the features using four models. The four models used are MobileNet, ResNet152, CNN, and LSTM-CNN. It is shown that ResNet152 of accuracy 99.22%, MobileNet of 96.48%, CNN of 92.19%, and LSTM-CNN of 91.80%.

Chouhan V et al. [3] introduced a DL model for diagnosing pneumonia based on the transfer learning concept. Trans-fer learning requires a CNN model that has been trained on a large picture dataset, such ImageNet. The target pneumonia detection task is manifested using the pre-trained model rather than beginning the CNN from scratch. With an accuracy of 98.97%, the model can more precisely identify relevant features in chest X-ray images associated with pneumonia using the pre-training data.

Tohidul Islam et al. [10] introduced a model for detecting pneumonia by using deep transfer learning from X-ray of the chest. This paper first identified the DL models which are most effective for pneumonia identification from the X-ray images. From this selected two networks have more accuracy and sensitivity for feature extraction. In the next step combined two sets of features and it is used as input to the conventional classifier. By doing this, actually combining the advantages of both traditional and DL methods with an accuracy of 89.93%.

Researchers from a wide range of fields have effectively applied deep learning models to the medical field as a result of deep learning's rapid advancement [9].Hua et al. [10] used convolutional neural network models with Deep Belief Networks (DBN) to improve the lung disease detection system and raise the diagnosis accuracy. To tackle the problem of insufficient medical data, Shin et al. [11] developed a transfer learning method that may be used for deep learning with small samples and reduce the significant amount of training data required for deep learning models. In order to achieve autonomous illness identification, Zhu et al. [12] developed the Deep Lung system, which uses a three-dimensional convolutional network model to detect and classify lung nodules.

B. MOTIVATION

Pneumonia remains a leading cause of death worldwide, especially among vulnerable groups like children, the elderly, and those with weakened immune systems. Diagnosing pneumonia is challenging due to the complexity of interpreting chest X-rays and the shortage of skilled radiologists, particularly in under-resourced areas. While CNNs offer high diagnostic accuracy, their lack of transparency hinders their clinical adoption. This research aims to enhance AI-driven pneumonia detection by combining accuracy with explainability, improving trust and enabling faster, more reliable diagnoses to ultimately reduce mortality and improve patient outcomes.

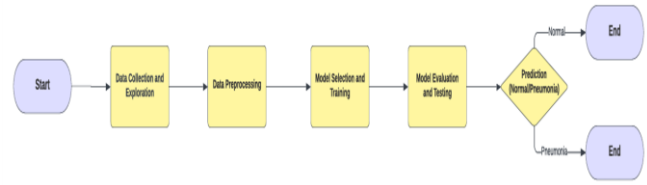
III. PROPOSED METHODOLOGY

The systematic approach used to create and verify a deep learning model combined with explainable AI (XAI) methods for pneumonia detection is described in this section. The process comprises several key steps, as described below:

IV. DATASET

This study uses a dataset of 5,856 chest X-ray images from pediatric patients aged 1 to 5, focusing on anterior-posterior views. The images include both normal and pneumonia cases and are divided into training, testing, and validation sets. The dataset, about 1.24 GB in size, is labeled with disease type (NORMAL, BACTERIA, VIRUS) and patient details. It was carefully validated to ensure high-quality data for training and evaluating deep learning models aimed at improving pneumonia diagnosis.

A. MODEL SELECTION AND CUSTOMIZATION



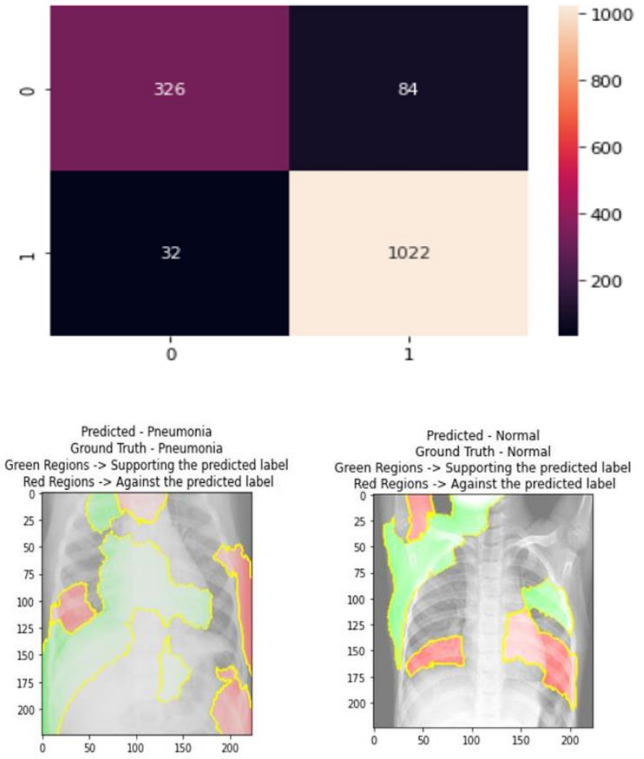
The first model applies LIME for explainability after using a CNN to distinguish pneumonia and normal cases in chest X-rays. Preprocessing of the dataset, which comprises chest X-ray pictures divided into the "NORMAL" and "PNEUMONIA" classes, is first step in the methodology. To guarantee uniformity throughout the dataset, the photos are scaled to 224x224 pixels after being loaded using the cv2 package. The matching labels—0 for Normal and 1 for Pneumonia—are recorded in arrays when the photos are retrieved from the appropriate directories. The dataset is then split into training and testing sets using train, test , split, with 75% going toward training and 25% going to ward testing. To properly scale the data for model input, the picture pixel values are normalized to a range of [0,1] by dividing by 255. Furthermore, one-hot encoding is used to encode the class labels (y train and y test), transforming them into a binary format appropriate for multiclass classification. A multi-layered, fundamental CNN model built with Ker-as makes up the model architecture. ReLU activation and 32 and 64 filters, respectively, come after the convolutional layers to extract features from the pictures. Following each convolutional layer, max-pooling layers down sample the feature maps' spatial dimensions. The retrieved attributes are then processed using a dense layer of 128 units that is fully linked. The probability distribution over the two classes (Normal and Pneumonia) is then predicted using an output layer that has two units and a soft max activation function. The model is built using the Adam optimizer, and cate gory cross entropy loss and accuracy are the assessment measures. The test set (X test scaled, y test) is used to assess the model's performance after it has been trained for five epochs using the training data (X train scaled, y train). The model's accuracy on the test data is approximately 93.44%. A confusion matrix is calculated using y test scaled and the predicted values y pred scaled to assess the classification performance (true positives, false positives, true negatives, and false negatives). Sea born is used to plot the confusion matrix as a heatmap for display. The model's predictions are interpreted using the LIME package to make them explain able. By locally approximating the model's choice using an interpretable model (such as a linear classifier), LIME provides an explanation. The LIME explainer provides feature importance in terms of which pixels most affected the model's choice to produce an explanation for each image in the test set. The model's forecast for every test image is explained using the explain instance function. Mark boundaries is used to highlight the areas of the image that either positively or adversely contributed to the anticipated label. Red parts show pixels that contradict the projected label, while green regions show supportive pixels in the final explanations. This enables us to determine which areas of the picture—such as particular lung regions—were used by the model to determine whether the X-ray was normal or indicative of pneumonia.

The second model classifies pneumonia and normal cases in chest X-rays using the pre trained InceptionV3 model for transfer learning. Grad-CAM is then used for visual explanations. Loading the chest X-ray pictures from directories with "NORMAL" and "PNEUMONIA" categories is the first step in the data preprocessing procedure. The pictures are preprocessed by scaling pixel values to the interval [0,1] and scaled to the necessary dimensions (299x299 for InceptionV3). To ensure that the normal and pneumonia photos are distributed appropriately, the dataset is divided into training and validation sets (80% for training and 20% for validation). The training data is subjected to data augmentation using Ker-as Image Data Generator, which applies random transformations like rotations, shifts, and flips to provide more varied training samples in order to enhance generalization. In terms of model architecture, the InceptionV3 model leverages learnt representations that are subsequently optimized for the pneumonia classification problem by acting as a feature extractor with pre-trained weights from ImageNet. For categorization, new fully connected layers are put on top of InceptionV3's basic layers, which are frozen (i.e., not trained). A flattening layer is used to convert the 2D feature maps into a 1D vector, two units for binary classification (normal or pneumonia), and two dense layers with 128 and 64 units with ReLU activation make up the final output layer. The model is put together using the Adam optimizer, binary cross-entropy loss, and accuracy as the evaluation metric. In order to train the model, the InceptionV3 base layers are left frozen while the weights of the additional layers are adjusted over a predetermined number of epochs using the training data. To fine tune the overall model and enhance performance on the particular job, it is optional to unfreeze parts of InceptionV3's higher layers. The accuracy and loss measures are used to assess the model's performance on the validation set. Grad-CAM is used to produce visual explanations for the model's predictions in order to increase explainability. By emphasizing the regions of the image that most influenced the model's selection, it aids in our understanding of the parts of the chest X-ray that it concentrated on. Grad-CAM creates a heatmap that is superimposed on the original image after calculating the gradients of the target class (normal or pneumonia) in relation to the output feature maps. This heatmap highlights the most relevant areas of the image that affected the model's decision to label anything, such as those showing pneumonia symptoms. To see the regions the model used to generate its predictions, the Grad-CAM heatmaps are super imposed over the original X-ray pictures. Healthcare practitioners can use these heatmaps to help them comprehend the model's decision-making process.

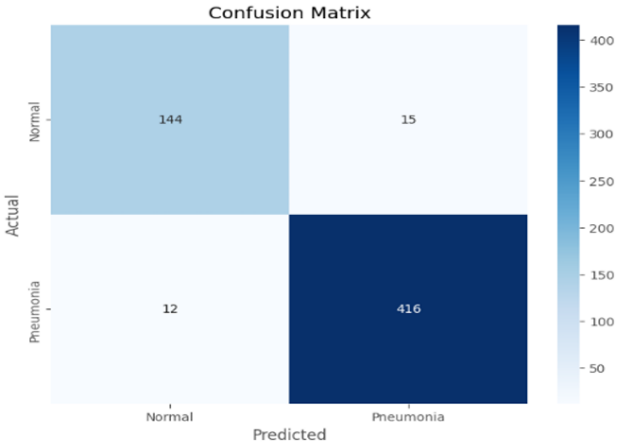
V. RESULTS

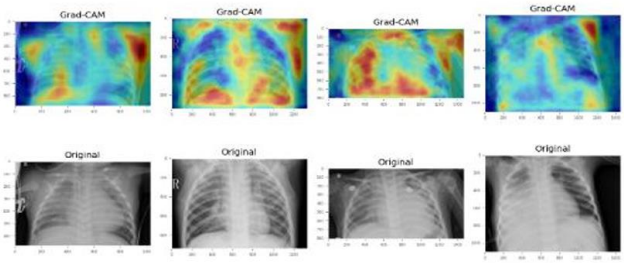
The initial deep learning model, constructed using a CNN, showed remarkable performance in differentiating between normal chest X-ray images and pneumonia. On the test dataset, the model's accuracy after training was roughly 93.44%. This shows how well the CNN model can

differentiate between chest X-rays that are normal and those that are affected by pneumonia. Its great accuracy can be attributed to the use of convolutional layers, which enabled the model to acquire key characteristics such as lung textures and aberrant patterns linked to pneumonia. The confusion matrix offered insightful information about how well the model worked. The relatively low frequency of false positives and false negatives indicates that the model can differentiate between normal and pneumonia X-rays. The model's ability to correctly categorize both groups was graphically supported by the confusion matrix heatmap. This model's integration with LIME, which makes the decision-making process transparent, is one of its main advantages. By emphasizing the areas of the test image that affected the model's choice, LIME was utilized to produce local explanations for each one. Red spots in the explaination indicated places that contradicted the expected label, such as pneumonia, whereas green portions supported the label. We were able to comprehend the model's focus thanks to these visual explanations, which showed that the model focused on particular lung regions that are frequently impacted by pneumonia. Because healthcare experts may validate the decision-making process based on the highlighted regions, this is essential for maintaining trust in the model.



The second model, leveraging InceptionV3 for transfer learning, achieved significant success in pneumonia detection. After fine-tuning, the model demonstrated high accuracy on the validation set, achieving approximately 95% accuracy. For visual explainability, this model's use of Grad-CAM is among its most influential features. Grad-CAM made it possible for us to see which areas of the chest X-ray pictures were most important to the model's judgment. The heatmaps produced by Grad-CAM provided clear visualizations of areas where pneumonia-related abnormalities, such as lung infiltrations or consolidation, were present. These areas were marked with red in the heatmap, while the background remained uncolored, providing an intuitive and effective way to communicate the model's reasoning. Grad-CAM improves the model's interpretability, facilitating healthcare professionals comprehension of the rationale behind each pre diction. During training, data augmentation was used to increase the model's resilience. In this step, random transformations such as rotations, shifts, and flips were applied to the training dataset, increasing its diversity. The model benefited from this approach, as it was exposed to a wider variety of X-ray images, which helped it generalize better to unseen data. The augmented data ensured that the model did not overfit to specific pat terns in the original dataset, improving its performance on the validation set.





Metric	Normal	Pneumonia	Macro Avg	Weighted Avg
Precision	92%	97%	94%	95%
Recall	91%	97%	94%	95%
F1 – Score	91%	97%	94%	95%
Support	159	428	-	-

VI. DISCUSSIONS

Number The results demonstrate that both models are effective for pneumonia diagnosis, but the InceptionV3 model showed superior performance in terms of accuracy and robustness. The InceptionV3 model achieved a higher accuracy (95%) compared to the CNN model (93.44%), along with better precision, recall, and F1-scores for pneumonia detection. This highlights the advantage of transfer learning, as the pre-trained InceptionV3 model

leveraged knowledge from the ImageNet dataset and fine-tuned it for the pneumonia classification task. In terms of interpretability, the CNN model integrated with LIME provided pixel-level explanations for each prediction, which is valuable for understanding the regions influencing the model's decision, particularly in cases with ambiguous patterns. However, the explanations were relatively localized and less visually intuitive. On the other hand, Grad-CAM for InceptionV3 provided a more comprehensive visual explanation by overlaying heatmaps on chest X-rays, which highlighted pneumonia affected areas such as lung consolidations, enabling clinicians to better interpret the model's predictions. The InceptionV3 model is better suited for practical clinical applications because to this feature. The "black box" character of deep learning models is addressed by integrating explainable AI approaches like Grad-CAM and LIME. These techniques foster trust among healthcare professionals by ensuring that the models focus on medically relevant regions. This is especially critical in healthcare, where decision-making must be transparent and reliable. However, there are some limitations and future considerations. The dataset primarily included pediatric X-rays, which may limit generalizability to other age groups or demographics. Expanding the dataset to include diverse populations is essential for broader applicability. Additionally, while the models showed high accuracy, further testing on external datasets is needed to ensure robustness in different clinical settings. Future work should also focus on integrating these models into real-time diagnostic systems to assist radiologists in clinical environments.

VII. RESEARCH SCOPE

By merging explainable AI techniques with deep learning models, this work seeks to improve pneumonia detection and address significant challenges in medical imaging and diagnosis. The study primarily aims to improve diagnostic accuracy by leveraging advanced models like InceptionV3 with transfer learning, achieving high diagnostic performance to ensure reliable results in clinical settings. Furthermore, the study aims to improve the interpretability of AI models by incorporating methods like Grad-CAM and LIME to offer localized and visual explanations for predictions, enabling medical practitioners to verify and have faith in the model's judgments. Another key objective is to address resource constraints by developing models capable of aiding pneumonia diagnosis in settings where access to skilled radiologists and advanced diagnostic tools is limited, ensuring scalability and accessibility. The research also seeks to support decision-making in healthcare by highlighting regions of interest in chest X-rays, enabling faster and more informed clinical decisions. The study lays the foundation for expanding the methodology to other demographics, such as adults and different clinical circumstances, even if the current dataset is concentrated on pediatric X-rays. Addition ally, this research advances the creation of real-time diagnostic tools that can be easily incorporated into medical procedures and give doctors quick, useful insights. Ultimately, by combining high accuracy with interpretability, the research promotes the adoption of AI in sensitive areas such as healthcare, where transparency and reliability are crucial. This research serves as a foundation for future advancements in AI-driven medical imaging, with the goal of improving health outcomes through timely, accurate, and transparent diagnoses across diverse clinical settings.

VIII. CONCLUSION

The InceptionV3 model, enhanced with Grad-CAM, offers a robust and interpretable solution for pneumonia diagnosis. Its combination of high diagnostic accuracy, visual explainability, and robustness makes it particularly well-suited for adoption in clinical environments, especially in resource-constrained settings.

REFERENCES

[1] Rajpurkar P, Irvin J, Ball RL, Zhu K, Yang B, Mehta H, et al. Deep learning for chest radio graph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radi ologists. PLoS Med. 2018;15(11):e1002686. doi: 10.1371/journal.pmed.1002686. PubMed PMID: 30457988. PubMed PMCID: PMC6245676.

[2] Elshennawy NM, Ibrahim DM. Deep-Pneumonia Framework Using Deep Learning Models Based on Chest X-Ray Images. Diagnostics (Basel). 2020;10(9):649. doi: 10.3390/diag nos tics10090649. PubMed PMID: 32872384. PubMed PMCID: PMC7554804.

[3] Chouhan V, Singh SK, Khamparia A, Gupta D, Tiwari P, Moreira C, Damaševičius R, De Albuquerque VH. A novel transfer learning based approach for pneumonia detection in chest X ray images. Applied Sciences. 2020;10(2):559. doi: 10.3390/app10020559.

[4] H. Wei, Y.Cheng, Pneumonia ct image scoring method, involves predict ing second score of ct image according to image feature and pre-trained scoring model, and fusing first score and the second score to determine f inal score of ct image.

[5] T. Kuth, R.Rupprecht, Computer aided diagnosis and therapy for patients suffering from pneumonia based on mri scans uses historical patient data to establish diagnosis and determine therapy.

[6] V. Chouhan, S. K. Singh, A. Khamparia, D. Gupta, P. Tiwari, C. Moreira, R. Damasevicius, V. H. C. de Albuquerque, A novel transfer learning based approach for pneumonia detection in chest x-ray images, Applied Sciences-Basel 10 (2).

[7] 1. I. D. Apostolopoulos, T. A. Mpesiana, Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks, Physical and Engineering Sciences in Medicine 43 (2) (2020) 635–640

[8] T. Mahmud, M. A. Rahman, S. A. Fattah, Covxnet: A multi-dilation convolutional neural network for automatic covid-19 and other pneumo nia detection from chest x-ray images with transferable multi-receptive feature optimization, Computers in Biology and Medicine 122.

[9] F. Piccialli, V. Somma, F. Giampaolo, S. Cuomo, G. Fortino, A survey on deep learning in medicine: Why, how and when?,

Information Fusion 66 (2021) 111–137. doi:10.1016/j.inffus.2020.09.006.

[10] K.-L. Hua, C.-H. Hsu, H. C. Hidayati, W.-H. Cheng, Y.-J. Chen, Computer-aided classification of lung nodules on computed tomography images via deep learning technique, Oncotargets and Therapy 8 (2015) 2015–2022.

[11] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R. M. Summers, Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning, Ieee Transactions on Medical Imaging 35 (5) (2016) 1285–1298.

[12] W. Zhu, C. Liu, W. Fan, X. Xie, Ieee, DeepLung: Deep 3D Dual Path Nets for Au tomated Pulmonary Nodule Detection and Classification, IEEE Winter Conference on Applications of Computer Vision, 2018.