BME3053C Report: Neural Network and Machine Learning Applications in Pneumonia Detection

----

University of Florida

04.24.25

Contents

Abstract

The analysis and generation of causal models in research present complexity and algorithmic probability issues. Inferring algorithmic structures in data, specifically large datasets for pneumonia or imaging, could prove to be a valuable nexus for healthcare technologies of the future. Deep Learning technologies and transformer models exhibit idiosyncratic characteristics, with extensive applicability across all considerations of healthcare. This report will consider aspects of transformer models and deep learning that make it a transformative force, practical realizations of a transformer model in large datasets, and lingering constraints that prevent the expansion of the technology into other fields. Machine learning has already been used in radiology, pathology, and dermatology, with U-Net models for medical image segmentation ResNet for classification and feature extraction, Faster R-CNN and YOLO for object detection in scans, Natural Language Processing in clinical text mining, and BioBERT for biomedical language modeling. These applications are ready to be installed for use at research institutes and hospitals that are dependent on analyzing structures or algorithmic data. Barring developmental constraints, machine learning should provide realistic computing advantages in fields adjacent to medicine and bioinformatics.

1 – Problem Statement

Pneumonia is an infection of the pulmonary parenchyma involving one or both lungs, depending on the severity of the case [1]. It remains a leading cause of pediatric mortality worldwide, accounting for over 2.1 million deaths globally in 2025, as well as a 15% mortality rate among children under 5 years of age [2]. It is the single largest infectious cause of death in children, and has more deaths than other notable diseases, including HIV/AIDS, malaria, and tuberculosis. In elderly individuals, the mortality is also high, particularly in the presence of comorbidities such as COPD, heart failure, or diabetes [2]. In terms of treatment, traditional first-line therapy methods are employed, and antibiotics for CAP typically include amoxicillin, macrolides, or respiratory fluoroquinolones. Other diagnostic data, including a respiratory rate greater than or equal to 30, a blood pressure lower than 90/60, and age greater than or equal to 65, are used in tandem with oxygen supplementation, fluid management, and antipyretics [3]. The disease is generally considered treatable, but accurate diagnosis poses substantial challenges. Confirmation of the disease is typically indicated by CXRs (Chest Radiographs) through board-certified radiologists or other trained specialists, as well as relevant patient history and hematological parameters. Supply and demand of radiologists, as well as difficulties managing large datasets, make proper diagnosis arduous. Radiographically, or on CXRs, the disease presents itself as localized or diffused regions of pulmonary opacity [4]. However, this alone is not sufficient for diagnosis, as underlying conditions may exhibit overlapping radiographic features. These conditions include pulmonary edema, alveolar hemorrhage, lobar collapse, malignant neoplasms, and postherpetic changes. These constraints underscore a need for robust, scalable, and proficient support systems to assist with proper diagnosis with high sensitivity and specificity. These systems must be able to properly integrate data sources, including the hematological parameters mentioned earlier, while being able to properly sift through noise and confounding factors in clinical environments. Addressing the diagnostic issue of pneumonia will be essential for improving clinical decision making, stewardship, and the proper reduction of mortality and morbidity for the disease.

2 – Literature Review

2.1 – Transformative Force: Advantages of Machine Learning and Use in Other Fields

Rooted in the foundational principles of AI, machine learning emerges as a vanguard of medical pattern recognition, with successes in this area including the possibility of retrieving and extracting knowledge from data instead of learning from experts and scientific texts [5]. Proper research has been conducted to prove that machine learning tools can simplify complex diagnostic tasks and reduce potential diagnostic errors [5]. In the context of longitudinal electronic health records, DL models have proven to be effective by incorporating temporal information rather than cross-sectional information [6]. It has even been used in enabling physiological data collection using sensors, which can be transmitted to remote servers for continuous analysis by models, as well as machine learning models [6]. Deep learning also has use for MRI-based brain tumor detection as well as the detection of Parkinson’s disease through resting-state functional magnetic resonance imaging. Parkinson’s traditionally has non-effective quantifiable biomarkers to detect it, but the use of ALFF-based multi-order radiomics have been used to successfully distinguish patients with PD [7]. For tumor detection, machine learning has played a vital role in the analysis, segmentation, and classification of cancer images, especially brain tumors [5]. Ultimately, machine learning and deep learning techniques offer transformative benefits by enabling the implementation of data-driven, scalable, and precision-focused solutions. These methods can drive advances in all sorts of fields, including preventative medicine, remote monitoring, and population health analytics, positioning healthcare towards more proactive care delivery.

2.2 – Machine Learning Techniques in the Context of Pneumonia and CXRs

For detection of Pneumonia in large datasets of CXRs, machine learning has proven to be effective, with recent notable work including an overview review on the promise of deep learning in thoracic conditions, a collection of important medical applications on lymph node and interstitial lung disease detection and classification, cerebral microbleed detection, pulmonary nodule detection in CTs, automated pancreas segmentation, cell image segmentation and tracking, predicting spinal radiological scores, and extensions of multi-modal segmentation [3]. In one such study conducted by Dr. Xiasong Wang et al., a chest X-ray database with 108,948 frontal view images of 32,717 patients was subjected to a convolutional neural network (weakly supervised) for classification and localization of common Thorax diseases. The network was demonstrative of commonly occurring thoracic diseases being detected via a localization framework, which was then validated using the proposed dataset [3]. This study uses an AUC (area under the receiver operating curve) evaluation method to evaluate classification, with a higher score indicating more proficiency. The neural network used had an AUC score of 0.8141 for Cardiomegaly and 0.7981 for Pneumothorax, promising results for the incorporation of models into hospital settings but also leaving room for improvement [3]. Another study by the University of Chongqing considered the use of a convolutional neural network for the diagnosis of COVID-19-based pneumonia during the height of the pandemic. At the time, the gold standard for COVID-19 detection was the reverse transcription polymerase chain reaction (RT-PCR), which has shown limitations with 70% accuracy, contributing to many incorrect diagnoses as well as pressure on the healthcare system [8]. The neural network proposed by this study made use of transfer learning, as well as a pretrained model (Xception) on various training and testing ratios (4 phases). For COVID-19, the CNN had an accuracy value of 0.98, an F1 score of 0.97, a precision value of 0.99, and a recall value of 0.95. For Pneumonia, similar results were observed with an accuracy of 0.97, an F1 score of 0.95, precision value of 0.98, and a recall value of 0.92 [8]. The CNN effectively identified COVID-19, as well as Pneumonia in the dataset, and even worked effectively at differentiating the two infections despite the many symptoms that they share [8]. Another study by Prabhav Guddati also made use of a convolutional neural network, incorporating a field programmable gate array-based hardware implementation of the CNN to detect Tuberculosis and Pneumonia in CXR images [9]. Regarding feature extraction, dimension reduction, and classification of diseases using CXR images, existing machine learning techniques had not yet been implemented on hardware or resource-constrained edge devices to detect PN and TB diseases. Existing DCNN models have many parameters in the interference phase and require higher resources for hardware implementation [9]. Therefore, there was a need for the development of a lightweight DCNN model that can be used to detect PN and TB diseases on CXR images. The model designed by Prabhav was able to detect TB and PN with an accuracy value of 96.39%, showing more promise for possible integrations of machine learning into medical settings or situations. All these studies collectively highlight the growing proficiency and versatility of machine learning, particularly with deep convolutional networks, in the detection of pulmonary disease (particularly pneumonia) using chest radiographs. While current models demonstrate strong metrics of performance and evaluation, continued development and refinement of these networks and optimization for deployment on resource-constrained devices must be fully realized before proper integration into clinical workflows can occur.

2.3 – Constraints of Machine Learning

Despite the proven proficiency of machine learning techniques in healthcare treatment and diagnosis, many fields are still hesitant to incorporate it into their work and diagnoses, with many citing accuracy and privacy concerns as sources of conflict. Privacy remains a topic of discourse for a variety of different healthcare treatment methods, but has recently been spotlighted when discussing the implementation of machine learning. While many different privacy-enhancing technologies (PETs) have been indicated as a solution for protecting personal data by current EU AI regulations, legality issues, as well as patient confidentiality concerns remain [10]. Regarding legality, existing regulations such as the Health Insurance Portability and Accountability ACT (HIPAA), the Cybersecurity Law of China, California Consumer Privacy Act, or the General Data Protection Regulation (GDPR) are no longer within the scope of security experts when considering artificial intelligence [10]. Also, in addition to privacy concerns, it is worth noting that healthcare continues to evolve and change, developing and improving upon current treatment methods, constantly changing where machine learning could be applicable. For example, many of the sources cited consider the use of machine learning on CXRs for Pneumonia classification and diagnosis techniques, but what if another modality of imaging was found to be better suited for diagnoses? In a study conducted by Wesley H. by the Department of Emergency Medicine, CXRs and computed tomography scans were analyzed to determine which served as a better conduit of diagnoses. The paper concluded that CXR demonstrated rather poor sensitivity and positive predictive value for detecting pulmonary opacities [11]. Reliance on CXRs could lead to significant rates of misdiagnosis, with further research required to determine more effective strategies for diagnosis. Another study conducted by Yalong Yang considered lung ultrasound imaging compared to CXRs for detecting pneumonia in children. Similar to the study by Wesley H., it was found that the sensitivity of the ultrasound was superior to that of the CXR and could lower the rate of misdiagnoses of Pneumonia in children [12]. As healthcare continues to evolve, if more effective modalities of imaging are determined, then models created previously for analysis on CXRs become obsolete, and the demand for new models presents itself. Overall, while machine learning holds promise for enhancing diagnostics as well as treatment, widespread adoption remains constrained by unresolved privacy, legal, and ethical concerns, as well as the dynamic nature of clinical practice. Addressing these challenges will be essential for properly integrating machine learning technologies into healthcare.

3 – Proposed Solution for Efficient Pneumonia Diagnoses

3.1 – Type of Model and Features

Convolutional neural networks have shown high levels of precision, F1 scores, and accuracy when addressing machine learning and its use in Pneumonia diagnoses. To address the detection and localization of Pneumonia in chest X-rays (CXRs), we will take advantage of these metrics, creating a custom CNN capable of multi-task learning and bounding box regression. The CNN will be less prone to overfitting and collapse more commonly found when considering generative adversarial networks or high-level diffusion models and will be easily importable to other means of technology. The model created can easily run on common computers found in the workplace or at a hospital. The model architecture is composed of three major components, a feature extractor, a model adapter, and two separate output heads (classification and regression). The feature extractor will be comprised of three convolutional layers with ReLU activation functions and average pooling operations to reduce spatial dimensions and preserve important features. The model adapter will flatten the extracted feature maps and pass them through a connected layer to prepare for output. The classification head utilizes a dense layer with softmax activation for Pneumonia detection, and the regression head will predict four continuous values corresponding to box components (x, y, width, height). This model makes use of a composite loss function, which combines cross-entropy for classification and mean squared error for bounding box regression. It will be optimized via the Adam optimizer.

3.2 – Data Preprocessing Steps

For our model, we are using an open-source data set from Kaggle that contains DICOM images of CXRs of patients. This dataset contains images that contain possible Pneumonia cases. Since the data consists of labels specifying bounding boxes and Pneumonia presence, preprocessing steps included: image resizing, pixel normalization, bounding box adjustment, and dataset preparation. For image resizing and subsequent aspect ration preservation, CXRs varied in original dimensions. In order to standardize the inputs for the CNN, all the images were reduced to an input size of 244 by 244 pixels. The aspect ratio was preserved in order to prevent the distortion or changing of anatomical structures. Any unused regions of the standardized frame were padded with zeros to create a black background. All of this was done with the format_image function. Figure 1 below shows the function properly working and displaying one sample of the data before the rest is preprocessed:

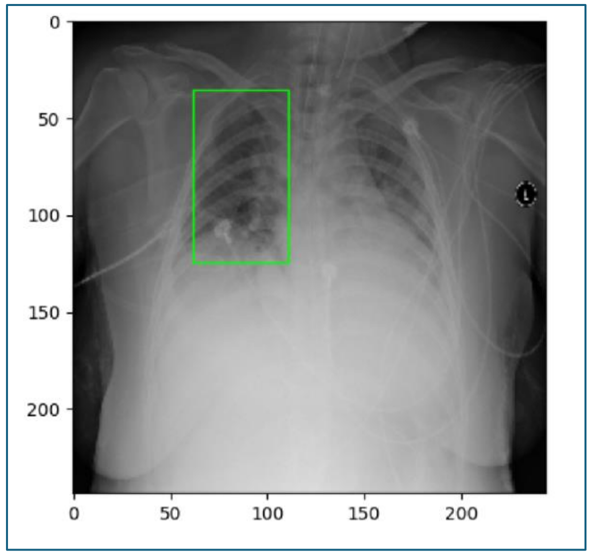

Figure 1: Sample of What Preprocessed Data Will Look Like

For bounding box adjustment, Pneumonia regions were initially annotated with bounding box coordinates relative to the image's native resolution. After resizing, these boxes were recalculated to maintain spatial correspondence with the resized images. The coordinates for each box were scaled with the same ratio used to resize the image. These boxes would also be normalized later by dividing the input size by 244, producing values between the range of 0 to 1, and simplifying the regression learning. For pixel value normalization, all pixel values were scaled from their original range to the 0,1 range. This will ensure consistent input distributions for each layer. Preprocessed images and corresponding labels were stacked into TensorFlow tensors. The training and validation sets were structured using tf.data.Dataset. This allowed us to map a function to pair images with their encoded labels and bounding boxes, shuffle the training set with a large buffer size to ensure diversity, and prefetch the batches asynchronously to overlap data loading and model loading, improving GPU utilization and aiding the use of this code and systems that do not contain high powered GPUs or even no GPU at all.

3.3 – Feature Engineering Techniques

Feature engineering primarily focused on optimizing the input data representations. For dynamic bounding box normalization, bounding boxes were rescaled relative to image size, ensuring proper localization. For data standardization, conversion of DICOM images to tensor formats allowed for better compatibility with the deep learning input requirements. The design structure also allows for easy insertion of other augmentation techniques like rotations, flips, or even contrast adjustments to enhance performance. Also, rather than using a complex feature extractor such as ResNet-50 or DenseNet-121, our model incorporated shallow convolutional layers and average pooling layers to reduce spatial dimensions while preserving average activations. Our model design trades off depth for speed and interpretability, making the model more robust for real world datasets.

4 - Results and Evaluation:

This model was evaluated considering both classification and bounding box regression. For classification, we considered accuracy, precision, F1, and recall, assessing the model's ability to classify CXRs as positive or negative (regarding Pneumonia). For bounding box localization, the mean squared error (MSE) was used to evaluate the precision of the predicted bounding box coordinates relative to the ground truth values. We also made use of a metric called IoU or Intersection over Union. This is a measure of the magnitude of overlap between two bounding boxes. It will calculate the size of overlap between two objects divided by the total area of the two objects combined. A visualization is provided by Figure 2 below:

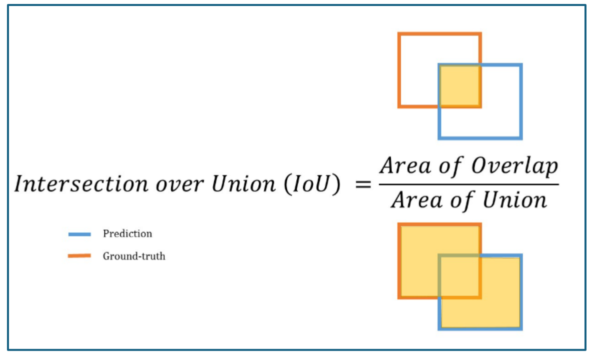

Figure 2: Intersection over Union

The higher this test metric is, the more accurate the model was with its classification and precision. After training the model for 100 epochs on the dataset consisting of 601 CXRs (too many epochs can lead to overfitting), the final training classification accuracy approached 99%, the final validation classification accuracy around 77-83%, and the regression head achieved a training MSE of around 0.015 and a validation MSE in the range of 0.018 – 0.025. Figure 3 and Table 1 below shows a visualization of some of these results:

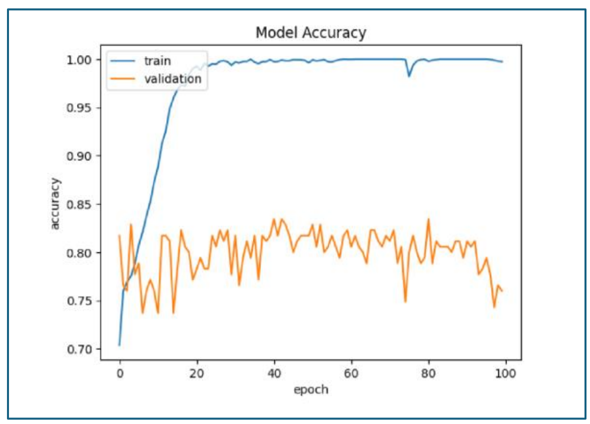

Figure 3: Plot of Training and Validation Accuracy vs. Epochs

Table 1: Epoch Compared to Training and Validation Accuracy

Epoch	***Training Accuracy (%)***	***Validation Accuracy (%)***
1	*64.7*	*72.6*
10	*84.6*	*73.7*
20	*94.1*	*76.5*
50	*98.3*	*73–78 (fluctuations)*
100	*99.1*	*~77–83 (final range)*

**NOTE: The training set and validation set metrics are NOT proper indicators of model performance. These indicators are only here for development purposes.

The metrics for the trained CNN assessed on a held-out test set that was not seen during training or validation are as follows. The final classification accuracy was 0.7404, the IoU was 0.0148, the precision value was 0.9614, the recall was 0.7108, and the F1 score was 0.8173. Figure 4 below visualizes test predictions for the model along with the calculated IoU:

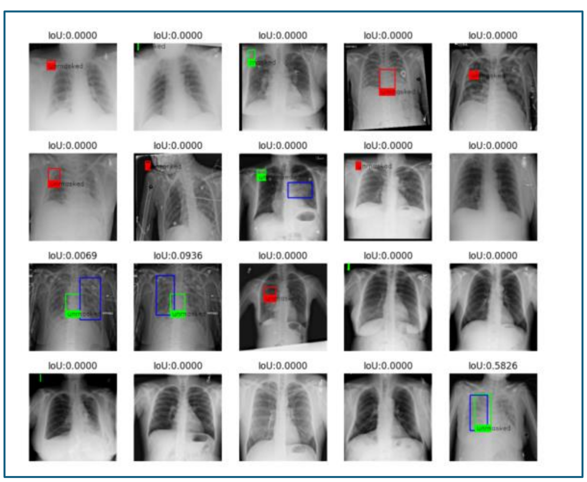

Figure 4: Visualization of Test Predictions Along with IoU

Table 2 below summarizes the data for the test set:

Table 2: Results for Test Set

Metric	***Value***
Classification Accuracy	*0.7404*
Mean Intersection over Union (IoU)	*0.0148*
Precision	*0.9614*
Recall	*0.7108*
F1 Score	*0.8173*

Unfortunately, the data visualized in figure 4 does not accurately reflect the average IOU or the accuracy of our model, as many of the incorrect or misidentified cases had been visualized in this figure (it pulled from the set randomly). Our model achieved reasonable performance in binary classification, identifying Pneumonia in around 76% of cases with a precision of about 95% and recall of about 71%. The relatively low mean IoU suggests that the model could detect the presence of Pneumonia, but localizing its prediction remains a challenge. This is likely due to the diffuse and variable appearance of Pneumonia on the CXR, limited training samples, and the use of a relatively under tuned model that prioritized the ability of being lightweight. There were also some challenges encountered when training the model. Since the model's training accuracy of around 99% exceeded the test accuracy of 76%, there was likely some overfitting, and the number of epochs could be adjusted. When it comes to bounding box prediction, the non-uniform and faint patterns may have contributed to the low IoU. Also, a relatively small number of positive cases and noisy annotations likely limit the model's ability to generalize spatially or locally.

5 – Conclusion and Practical Application

Our proposed architecture was able to achieve moderate classification accuracy and learned to detect Pneumonia patterns. The model did struggle with localizing the Pneumonia, and the low IoU is demonstrative of the challenge of accurate localization in a dataset that contains lots of noise (similar to real world clinical data). Nonetheless, the model is proof of concept for using machine learning to enhance processes of automatic disease detection. When it comes to practical applications, the CNN could be integrated into different picture archiving and communication systems to assist with rapid analysis of CXRs. For example, the model could be used to quickly flag over some cases that may be Pneumonia, and then they can be properly examined by Radiologists after. This model could also be useful in an area where there are not many radiologists, such as emergency departments. One final application could be for annotation assistance for new datasets. The bounding box regression output that was used in this model could be reused to possibly semi-automatically annotate medical datasets, possibly accelerating the labeling process in other models being developed. Our model is indicative of both the promise and limitations of machine learning in medical imaging. While the performance was considered competitive given the dataset size and simple nature of the model, the localization metrics leave much to be desired in terms of creating a truly effective CNN that can accurately supplant a radiologist.

References

[1] Patrick R. Ching et al.m "Severe Pneumonia," pp. 705-720, 2025.

[2] Catia Cilloniz 1 and t. P.-S. Group et al., "World Pneumonia Day 2024: Fighting Pneumonia and Antimicrobial Resistance," *National Library of Medicine,* pp. 1283-1285, 2024.

[3] Xiaosong Wang et al., "ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases," *CVF,* pp. 2097-2106.

[4] T. Franquet, "Imaging of Community-acquired Pneumonia," *National Library of Medicine,* pp. 282-294, 2018.

[5] Soheila Saeedi et al., "MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques," *BMC Medical Informatics and Decision Making ,* pp. 1-17, 2023.

[6] AyoubSi-Ahmeda et al., "Survey of Machine Learning based intrusion detection methods for Internet of Medical Things," *Elsevier,* pp. 3-25, 2023.

[7] Dafa Shi et al., "Machine Learning for Detecting Parkinson's Disease by Resting-State Functional Magnetic Resonance Imaging: A Multicenter Radiomics Analysis," 2022.

[8] Deepak Kumar Jain et al., "Deep Learning-Aided Automated Pneumonia Detection and Classification Using CXR Scans," *Hindawi,* pp. 1-19, 2022.

[9] U. C. Scans, "FPGA Implementation of the Proposed DCNN Model for Detection of Tuberculosis and Pneumonia Using CXR Images," *IEEE EMBEDDED SYSTEMS LETTERS, VOL. 16, NO. 4, DECEMBER 2024,* pp. 445-448, 2024.

[10] S. Z. E. Mestari, "Preserving data privacy in machine learning systems," *Computers and Security,* vol. 127, no. 103605, 2024.

[11] Wesley H et al,. Self MD MPHa, "High Discordance of Chest x-ray and computed tomography for detection of pulmonary opacities in ED patients: implications for diagnosing pneumonia," *The American Journal of Emergency Medicine,* pp. 402 - 405, 2013.

[12] Y. W. a. W. Z. Yalong Yang, "Comparison of lung ultrasound and chest radiography for detecting pneumonia in children: a systematic review and meta-analysis," *Italian Journal of Pediatrics,* pp. 1-11, 2024.