Artificial Intelligence in Healthcare Project (DS-552) Topic: Pneumonia Detection with Model Uncertainty Estimation

Amandeep Viswas*, **Indumouli Nandy**†, **Rahul Singha**‡, **Sourit Khamaru**§
Batch: MSDSM-01, Group: 2
IIT Ropar - IIM Amritsar Joint MSc Program in Data Science and Management
*2023DSS1004, †2023DSS1016, ‡2023DSS1027, §2023DSS1034

***Abstract*—This project explores a deep learning-based approach to detecting pneumonia from chest X-ray images and introduces a mechanism to estimate the model's uncertainty in its predictions. A Convolutional Neural Network (CNN) model was developed and trained on a dataset of labeled X-rays, distinguishing between pneumonia and normal cases. Data augmentation and normalization techniques were applied to enhance model performance, and an uncertainty estimation mechanism was integrated to quantify prediction confidence. Results indicate satisfactory recall and F1 scores, with the uncertainty estimation helping flag cases where the model exhibited low confidence. This approach offers potential for a more reliable diagnosis in clinical settings, where identifying uncertain predictions can guide further medical assessment.**

***Index Terms*—Pneumonia Detection, Deep Learning, Convolutional Neural Network, Uncertainty Estimation, Biomedical Imaging**

I. Introduction

Pneumonia remains a leading cause of morbidity and mortality globally, often requiring timely diagnosis for effective treatment. Chest X-rays (CXRs) are widely used for diagnosis; however, accurately interpreting them is challenging, especially in early or mild cases. Artificial Intelligence (AI) in radiology has shown promise for automating and improving diagnostic accuracy. This project develops a CNN model to classify CXR images as either pneumonia or normal and introduces an uncertainty estimation feature, enabling the identification of predictions with high uncertainty. This work demonstrates a robust model that not only identifies pneumonia but also flags low-confidence cases, potentially aiding radiologists in prioritizing cases that may need further review.

A. Context of the Project

Pneumonia is a major global health issue, leading to significant morbidity and mortality each year. Effective treatment relies on rapid and accurate diagnosis, often requiring CXRs that need skilled radiologists for interpretation. However, subtle pneumonia cases can be overlooked, resulting in delayed treatment. With AI's advancements, automated diagnostic tools are emerging as a way to assist clinicians and improve healthcare outcomes. Recent statistics indicate that pneumonia affects millions, with particularly high incidences in vulnerable populations, including children and the elderly. The integration of machine learning techniques in this domain could significantly mitigate human error and enhance diagnostic precision.

B. Motivation for the Work

This project aims to enhance diagnostic speed and accuracy in detecting pneumonia using CXRs, particularly in areas with limited radiological expertise. Convolutional Neural Networks (CNNs) are effective in image classification tasks, making them suitable for this application. By developing a CNN model with an uncertainty estimation feature, this work helps radiologists identify high-priority cases, potentially leading to faster, more accurate diagnoses and improved patient care. The motivation also stems from the necessity for scalable solutions that can operate in low-resource settings, where access to trained medical personnel may be restricted. Furthermore, addressing the challenge of model interpretability in AI is critical, as clinicians must trust and understand the outputs generated by AI systems to incorporate them into their workflow effectively.

II. Related Work / Background

Past studies have shown the efficacy of deep learning in detecting lung conditions, particularly pneumonia, from X-ray images. Research by Paul Mooney et al. [1] and others has validated CNN models in distinguishing between pneumonia and healthy cases. While these models achieve high accuracy, they lack mechanisms to communicate uncertainty, which is crucial in clinical decision-making. Our approach differs by not only classifying images but also quantifying prediction confidence, potentially adding an extra layer of reliability in practical applications.

Previous works have focused on achieving high classification accuracy, but few have adequately addressed the interpretability of model decisions. The introduction of uncertainty estimation methodologies is essential as it allows clinicians to understand when to rely on model predictions and when to seek additional diagnostic information. This is particularly relevant in clinical environments where the stakes of misdiagnosis are high.

III. Methodology

A. Data Collection and Preprocessing

The dataset used, sourced from Kaggle, comprises labeled CXR images divided into ”pneumonia” and ”normal” categories. Images were preprocessed, including resizing to 256x256 pixels and converting to grayscale to ensure consistency and reduce computational complexity. Additionally, data augmentation was implemented using rotation, shifting, zooming, and horizontal flipping to enhance generalization. Data normalization techniques, such as scaling pixel values to a range of [0, 1], were also applied to improve model convergence during training. The dataset was split into training, test, and validation sets, with a standard 89-10-1 percentage distribution, to ensure a robust evaluation of model performance.

B. Model Architecture

A CNN architecture was implemented due to its proven performance in image classification tasks. The model consisted of several convolutional layers with ReLU activation, followed by max-pooling layers. A dropout layer was added to prevent overfitting. The final layer was a single neuron with a sigmoid activation function, suitable for binary classification (pneumonia vs. normal). The architecture was designed with the aim of balancing complexity and efficiency, ensuring that the model could learn relevant features while remaining computationally feasible. Additionally, hyperparameter tuning was conducted to optimize learning rates, batch sizes, and the number of epochs.

C. Uncertainty Estimation Mechanism

Monte Carlo Dropout was implemented to estimate uncertainty by making the model re-evaluate predictions multiple times with dropout layers active. The mean and variance of these predictions provided a measure of the model’s confidence. This mechanism helped identify cases where the model was less certain, flagging them for potential follow-up. By leveraging stochastic forward passes, the model can capture the inherent uncertainty associated with predictions, providing a quantifiable metric that can enhance clinical decision-making.

IV. Results

A. Model Performance

The CNN model was trained for 40 epochs. Despite some fluctuations in accuracy, recall was high, indicating the model’s sensitivity to pneumonia cases. Key performance metrics are as follows:

- **Accuracy**: 62.5%
- **Recall**: 100%
- **F1 Score**: 0.7692
- **AUC-ROC**: Moderate area under the curve, illustrating the model’s ability to distinguish between classes.

These results indicate that while the model is effective at detecting pneumonia, further tuning is required to improve overall accuracy and reduce false positives. The high recall score emphasizes the model’s capability to detect positive cases, which is critical in a healthcare setting where missing a pneumonia diagnosis can have severe consequences.

B. Confidence Scores and Uncertainty Detection

The uncertainty estimation effectively assigned confidence scores to predictions. Cases with low confidence were flagged for review, which could aid radiologists by highlighting cases needing further analysis. The ability to quantify uncertainty provides a means for radiologists to prioritize cases that warrant additional scrutiny, ultimately contributing to improved patient safety and care outcomes.

C. Visualization of Results

A confusion matrix and ROC curve illustrate the model’s performance. The confusion matrix shows the distribution of true positives, false positives, true negatives, and false negatives. The ROC curve further demonstrates the trade-off between true and false positive rates. Visualizing these metrics helps convey the model’s effectiveness and areas needing improvement to stakeholders and clinicians alike.

V. Discussion

The model’s high recall suggests it is well-suited for detecting pneumonia, as it successfully identifies the majority of positive cases. The F1 score of 0.7692, while satisfactory, indicates room for improvement in precision. Incorporating uncertainty estimation adds value by enhancing interpretability, as cases with high uncertainty can be prioritized for further review.

Comparison with previous studies reveals that while similar models achieve high accuracy, they lack uncertainty metrics. Our model’s approach of including uncertainty estimation contributes a unique aspect, improving decision support in clinical contexts where uncertainty can guide follow-up testing or expert review. Furthermore, the integration of AI in clinical workflows is vital for enhancing the efficiency of radiologists and reducing diagnostic bottlenecks.

VI. Conclusion and Future Work

This project presents a CNN-based model capable of detecting pneumonia from chest X-ray images with an added mechanism for uncertainty estimation. Key findings demonstrate that the model achieves high recall, effectively identifying pneumonia cases. The incorporation of uncertainty estimation offers a novel approach to highlight cases that may require further review, thus enhancing its potential clinical utility.

A. Future Work

Improvements could include exploring other architectures, such as transfer learning with pre-trained models, and fine-tuning the threshold for uncertainty. Additionally, validating the model on diverse datasets would help assess its robustness across different patient demographics. Future research could also focus on expanding the model to detect other respiratory conditions and integrating it with electronic health records for real-time decision support.

Compliance with Ethical Standards

This research utilized openly available datasets from Kaggle, where no additional ethical approval was required.

Acknowledgments

We extend our gratitude to **Prof. Sukrit Gupta (IIT Ropar) Sir, Pritam Sunil Mahajan Sir**, and the creators of the Kaggle dataset, and acknowledge the support of the Colab platform for com- putational resources.

References

[1] Mooney, P. T., et al., "Chest X-ray Images (Pneumonia)," Kaggle, 2016.
[2] He, K., Zhang, X., Ren, S., Sun, J., "Deep Residual Learning for Image Recognition," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.