Deep Learning Methods for Diagnostic Imaging

An overview

William Bach
CSSE
UWB
Bothell, WA
bach2690@uw.edu

ABSTRACT

Diagnostic imaging plays a pivotal role in modern healthcare, enabling the non-invasive assessment of various medical conditions and diseases. The integration of machine learning algorithms into diagnostic imaging has revolutionized medical practice, offering enhanced accuracy, efficiency, and patient care. This paper provides an overview of the machine learning algorithms applied in diagnostic imaging, emphasizing their significance in improving diagnostic accuracy, reducing human error, and improving clinical decision-making. Further, we will utilize common and emerging machine learning techniques to analyze medical image datasets to better understand the applicable machine learning methodologies as they pertain to medical imaging.

CCS CONCEPTS

• Machine Learning • Deep Learning • Convolutional Neural Networks • Spatial Transformers

1 Introduction

The field of diagnostic imaging has undergone a remarkable transformation with the advent of machine learning algorithms. Historically, techniques such as X-rays, magnetic resonance imaging (MRI), computed tomography (CT), and ultrasound have served as incredibly useful tools in the medical field, evolving since their inception in the early 20th century [1]. These techniques have been integral in diagnosing and monitoring a variety of medical conditions, revolutionizing patient care with their ability to non-invasively reveal internal structures and abnormalities.

In recent years, a major shift has occurred with the inclusion of machine learning algorithms into the domain of diagnostic imaging. This change has sparked a revolution in medical practices, significantly enhancing diagnostic accuracy and overall patient care [2]. Machine learning, particularly through its ability to process and analyze large datasets, has brought about a massive change in how medical images are interpreted. By identifying patterns and anomalies that might elude human detection, these algorithms reduce the likelihood of diagnostic errors and accelerate the formulation of treatment [3].

Central to this revolution is the advent of artificial intelligence (AI), with deep learning algorithms, mimicking the human brain's structure and functionality [4]. These advanced algorithms excel at deciphering complex patterns within large datasets, making them well-suited to the realm of analyzing medical images. Their implementation spans across various aspects of diagnostic imaging, from classification and segmentation to the detection of subtle abnormalities [5].

The historical progression from analog to digital imaging laid the groundwork for this transformative era in medical diagnostics, enabling intricate image analysis and more efficient storage methods [1]. At the same time, the rise of machine learning, especially in the form of supervised learning, has allowed for more sophisticated and accurate interpretation of these digital images [6]. Particularly, deep learning and convolutional neural networks (CNNs) have significantly advanced the field, demonstrating high accuracy in image recognition and classification tasks within medical imaging [3]. The impact of these AI applications is evident across various imaging techniques, enhancing diagnostic abilities in procedures ranging from X-rays to ultrasounds [7].

Despite these improvements, these advancements are not without challenges. The issues of data privacy, the need for extensive, well-annotated datasets, and the potential for biases in AI models represent substantial challenges that the field must address [8]. Looking ahead, emerging trends such as the integration of AI with telemedicine and predictive analytics herald a new era in healthcare, promising even more personalized and efficient patient care [8].

This paper delves into these developments, exploring some of the various machine learning algorithms in diagnostic imaging with a particular focus on CNNs. By analyzing medical image datasets through common and emerging machine learning techniques, we aim to provide a deeper understanding of the methodologies that are reshaping medical imaging. In doing so, we highlight how these advancements can not only improve diagnostic accuracy but also reduce human error and enhance the speed and efficacy of clinical decision-making.

1.1 Convolutional Neural Networks

CNNs have revolutionized the field of medical imaging, thanks to their ability to accurately process and analyze visual data. CNNs are noted by their distinctive layered architecture that mimics the human visual system, allowing for the automatic detection of features within images [9].

A typical CNN architecture is comprised of several types of layers: convolutional layers, pooling layers, fully connected layers, and the output layer [10]. The convolutional layers act as feature extractors; they apply various filters to the input images to create feature maps that highlight specific attributes like edges, textures, or shapes. Pooling layers follow, which reduce the spatial size of the feature maps, retaining only the most important features helping to reduce computational complexity. After several convolutional and pooling layers, the high-level reasoning in the neural network is done via fully connected layers. These layers flatten the feature maps and connect every neuron in one layer to every neuron in the next layer, giving rise to the name 'fully connected'. The final layer uses an activation function, such as the sigmoid function for binary classification, or the softmax function for multi-class classification, to output probabilities corresponding to each class [11].

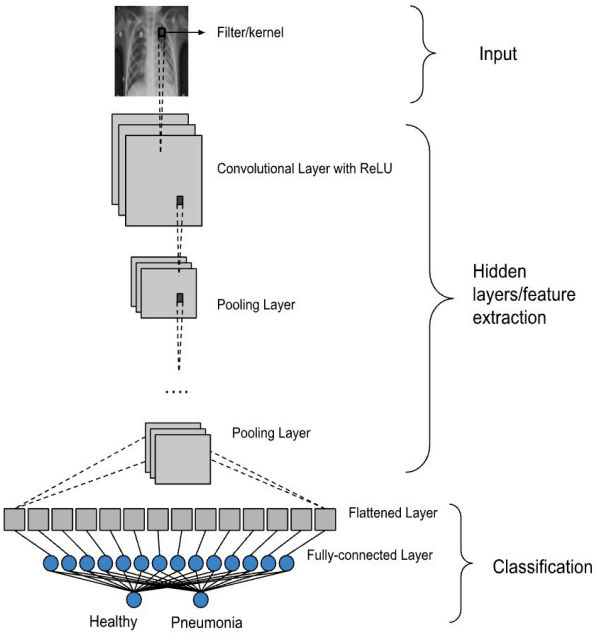

Figure 1: A representation of the various layers inside CNNs as they might be used in diagnostic imaging.

This hierarchical structure of CNNs allows for the transformation of raw pixel data into an abstract representation that can be used for classification tasks, making CNNs particularly effective for tasks such as identifying pathologies in X-ray or MRI images [7]

1.2 Dataset

In this study, I utilize a previously preprocessed dataset of chest X-rays, sourced from Kaggle [12], which is based on the dataset originally compiled and used by Kermany et al. in their article 'Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning' [13]. This dataset comprises 5856 X-ray images, each labeled to assist in distinguishing between two primary categories: 'normal', denoting images from healthy individuals, and 'opacity', indicating the presence of pneumonia in the subject.

The dataset is partitioned into three distinct subsets to facilitate training and evaluation of the model: the training set includes 4192 images (comprising 1082 'normal' and 3110 'opacity' images), the validation set consists of 1040 images (267 'normal' and 773 'opacity'), and the testing set contains 624 images (234 'normal' and 390 'opacity'). This distribution ensures a representation of both conditions, allowing the convolutional neural network to learn and generalize effectively.

The choice of this dataset is motivated by its balance and its relevance to current medical challenges. It presents a realistic scenario for deploying AI in medical diagnostics.

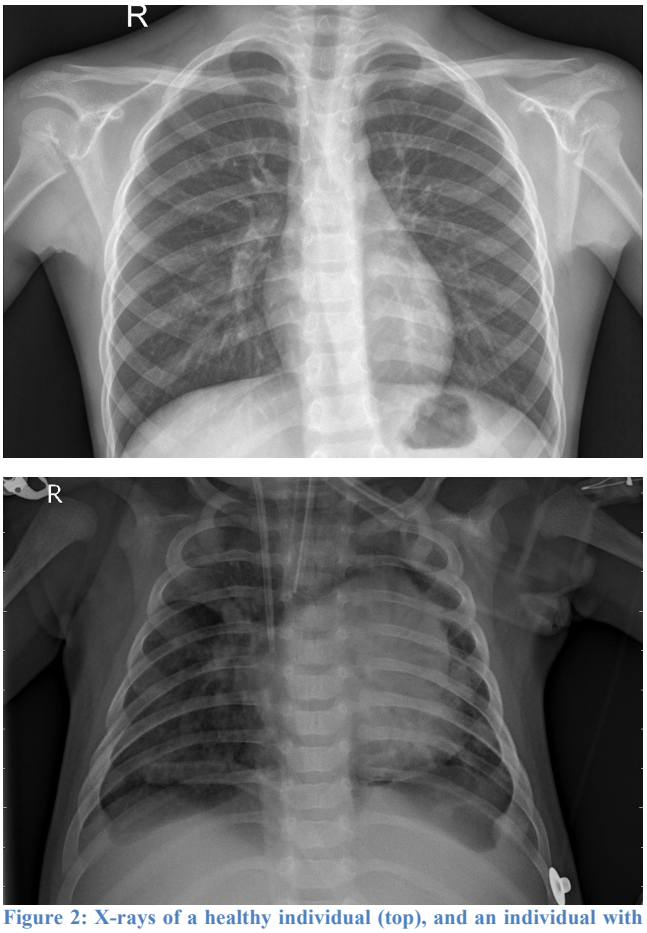

Figure 2: X-rays of a healthy individual (top), and an individual with pneumonia (bottom) derived from Kermany et al.'s dataset [13].

1.3 Data Preprocessing

My approach to preprocessing chest X-ray images involves several techniques tailored to optimize the performance of our CNNs [14]:

1. **Image Normalization:** Each X-ray image is normalized to have pixel values between 0 and 1. This process involves dividing the pixel values by 255 (the maximum pixel value), which aids in the convergence of the neural network during training by providing a consistent scale.
2. **Resizing Images:** Given the varying dimensions of raw X-ray images, I resized every image to a fixed size of 500x500 pixels. This uniformity is required for batch processing and ensures that the CNN receives inputs of a consistent shape.
3. **Data Augmentation**: To increase the diversity of the dataset and reduce the risk of overfitting, I applied data augmentation techniques such as random rotations, zooming, and horizontal flipping using python keras' built-in ImageDataGenerator class. These transformations simulate variations that could occur in clinical settings, enabling the CNN to learn more robust features [15].
4. **Label Encoding:** The images are categorized into two classes—'healthy' and 'pneumonia'. I employ one-hot encoding to convert categorical labels into a binary matrix representation, which is necessary for the classification output of the CNN.
5. **Greyscale Conversion:** Because color information is not critical for my analysis, all x-ray images are converted to greyscale to reduce computational load.

2 Methods

To better understand CNN architectures and their impact on diagnostic imaging classification, I used an incremental approach. I began with a simplified CNN model to explore the various model components and their roles in image classification. This step-by-step method also provides a framework to compare the accuracy and performance implications of different architectural choices.

My simplified CNN starts with a single two-dimensional convolutional layer, equipped with 32 kernels of size 3x3 and a stride of 2 to reduce the spatial dimensions of the output from the outset, helping to lower the model's complexity. The ReLU activation function was selected for its ability to introduce non-linearity while mitigating the vanishing gradient problem, a common issue in deep neural networks [16].

$$f(x) = \max(0, x)$$

Equation 1: The ReLU activation function

After the convolutional layer is a max pooling layer with a pool size of 2x2, which serves to further reduce the dimensionality of the feature maps. This operation not only compresses the image data but also helps in making the detection of features invariant to scale and orientation. The flattened layer that follows transforms the two-dimensional feature maps into a one-dimensional vector, preparing the data for the final classification step. The fully-connected layer at the end, with a sigmoid activation function, is optimized for binary classification—healthy versus pneumonia in X-ray images. This output layer maps the learned features to the probability of the presence of pneumonia.

$$S(x) = 1/(1 + e^{-x})$$

Equation 2: Sigmoid activation function

The model includes a binary cross-entropy loss function and Adaptive Moment Estimator (Adam) optimizer [17], with the goal being to utilize adaptive learning rate methods, which adjust the learning rate during training for faster convergence and improved performance.

2.0.1 Simple Model Results

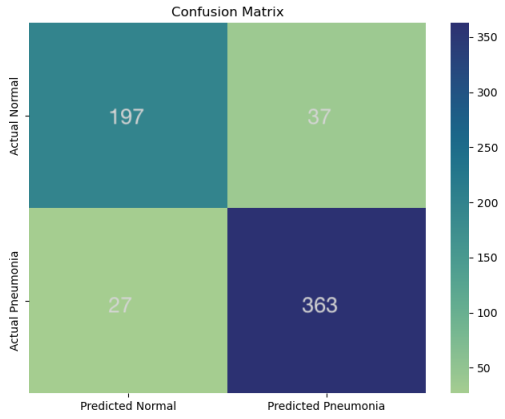

Figure 3: Confusion matrix with the simplified CNN

After training and validating the simplified CNN, I found a testing accuracy of 89.74%. Such a high degree of accuracy from a basic architecture is indicative of the powerful feature extraction capabilities found in even the simplest CNNs. The confusion matrix displayed in figure 3 provides further insights into the model's performance, revealing that out of the total pneumonia cases, the model correctly identified 363 cases (true positives), but also misclassified 37 cases as healthy (false positives). Conversely, the model accurately recognized 197 healthy individuals (true negatives), with 27 misclassifications (false negatives). The higher number of true positives relative to false

negatives suggests that the model is more inclined to err on the side of sensitivity rather than specificity.

While the overall accuracy is high, the presence of both false positives and false negatives indicates areas for improvement. False positives could lead to unnecessary treatment for healthy individuals, and false negatives could result in missed diagnoses for patients with pneumonia.

2.1 Adding Layers

Building on my simplified CNN model, I introduced additional layers to enhance its capacity for feature extraction and classification. The updates are as follows:

- **Additional Convolutional Layer**: A new 3x3 convolutional layer with 32 kernels and the ReLU activation.
- **Additional Max Pooling Layer:** Following the new convolutional layer, an added max pooling layer with a 2x2 window.
- **Expanded Fully-Connected Layer:** The addition of another fully-connected layer with 64 units expands the model's ability to combine the features into more abstract representations.

These architectural additions aim to provide the CNN with more layers of feature learning and non-linear transformation capabilities. By deepening the network, the expectation is to enable the extraction of more nuanced patterns that could lead to improved accuracy.

2.1.1 Intermediate Model Results

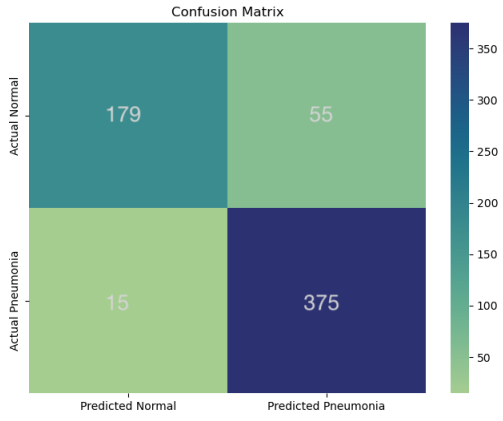

Figure 4: Confusion matrix for the intermediate model architecture.

Upon including additional layers into my CNN model, we observed a testing accuracy of 88.78%. Interestingly, this represents a slight decrease from the 89.74% accuracy of the simpler model. Such a reduction might suggest the onset of model complexity that does not yet translate to better performance on the given dataset, potentially due to factors like overfitting.

The confusion matrix reveals that the model predicted 375 true positives, correctly identifying most pneumonia cases. However, it also produced 55 false positives, which is an increase from the previous model. Conversely, the number of false negatives decreased to 15, indicating improved sensitivity.

2.2 Deep Model Architecture

The most comprehensive model utilized employs a sequential and layered approach with the addition of the following layers from the prior model iteration [14]:

- **Convolutional Layer:** 64 kernels with a window size of 3x3, and a ReLU activation. This addition brings the total convolutional layers up to three.
- **Max Pooling Layer:** Following this convolutional layer, an added max pooling layer with a 2x2 window.
- **Fully-Connected Layer:** The addition of another fully-connected layer with 128 units.

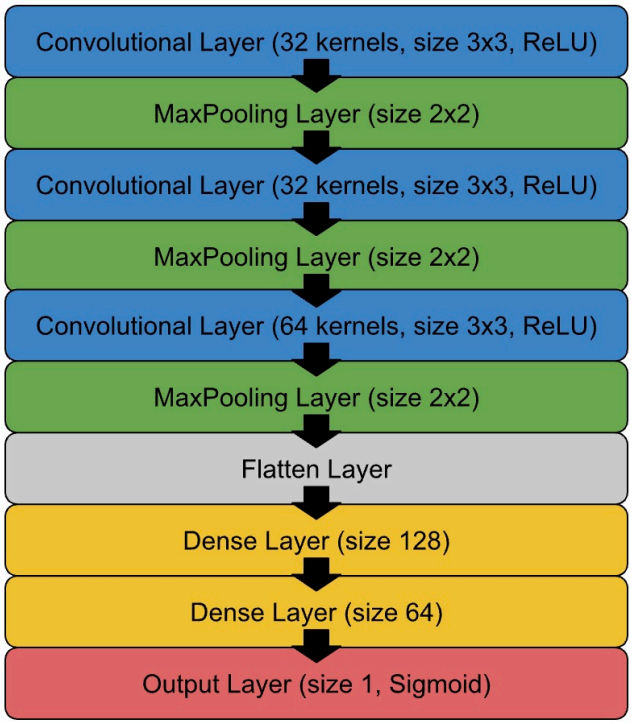

Figure 5: Architecture overview of the most complex CNN

This final model represents a more traditional and deeper CNN architecture. It is assumed that the added layers and increased number of neurons will provide a more nuanced analysis of the X-rays.

2.1.1 Deep Model Results

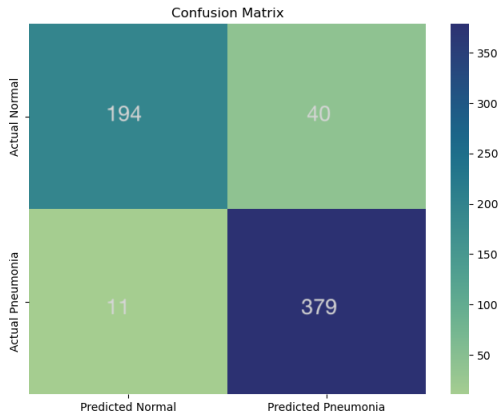

Figure 6: Confusion matrix for my deep CNN

The most sophisticated CNN model offers an overall improved testing accuracy of 91.83%. This improvement in accuracy, compared to the simpler models, indicates that the added complexity and depth of the architecture have successfully captured more distinctive features crucial for classification. Examining the confusion matrix, we see a substantial number of true positives (379), suggesting that the model is highly effective at identifying cases of pneumonia. Moreover, the reduction in false negatives to 11 indicates a heightened sensitivity, which is crucial in medical diagnostics to minimize the risk of overlooking potential illness. While there are 40 false positives, which reflects the model's tendency to over-predict pneumonia in some cases, this number is counterbalanced by the high true negative count of 194. This suggests that the model retains a strong specificity, correctly identifying healthy cases most of the time.

Overall, the performance of this more complex model demonstrates the benefits of a deeper network architecture in improving the reliability and accuracy of diagnostic imaging classification, making it a promising tool for aiding clinical decision-making in the detection of pneumonia.

3 Distinguishing Types of Pneumonia

To further refine diagnostic capabilities, I extended the dataset to distinguish not only between healthy individuals and those with pneumonia but also to identify the type of pneumonia—viral or bacterial. This first required hand-annotating the X-ray images to be either 'normal', 'viral', or 'bacterial' to ensure accurate classification. To maintain computational efficiency, I opted to use the initial simplest CNN model architecture. While more complex models can offer greater accuracy, they also require substantially more computational resources.

The most critical modification in the model itself was the replacement of the sigmoid output layer with a softmax activation function. This change is necessary for multi-class classification [10], as the softmax function provides a probability distribution across the three classes—healthy, viral pneumonia, and bacterial pneumonia—enabling the model to assign probabilities to each class for a given input image. In turn I switched from using a binary cross-entropy loss function to a categorical cross-entropy loss function, which is suitable for multi-class classification scenarios. The Adam optimizer was retained for its adaptive learning rate capabilities [17].

This multi-class classification approach represents a significant step forward in our project, offering the potential to provide more detailed diagnoses from chest X-ray images, which could be particularly beneficial in clinical settings for tailored patient care [18].

3.0.1 Multi-Class Classifier Model Results

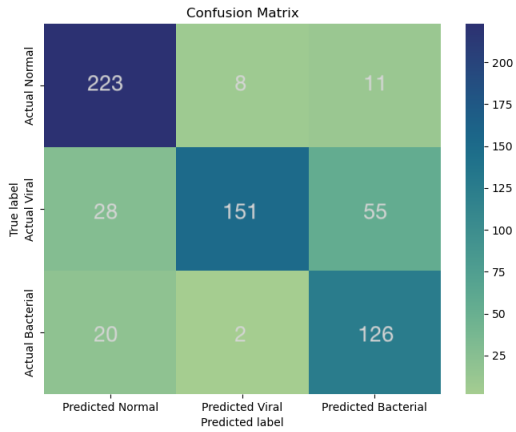

Figure 7: Confusion matrix for my multi-class classifier CNN

The testing accuracy of this multi-class CNN model was found to be 80.13%. While this result is notable, it is lower compared to the 91.83% accuracy obtained in the binary classification model (healthy vs. pneumonia). This discrepancy highlights the increased difficulty associated with multi-class classification tasks.

Looking at the confusion matrix for the multi-class model, we see the following:

- **Healthy Individuals:** The model shows strong performance in identifying healthy cases with 223 true positives, but also misclassifies 8 as viral pneumonia and 11 as bacterial pneumonia.
- **Viral Pneumonia:** The model correctly identifies 151 cases of viral pneumonia, but 28 cases are misclassified as healthy and 55 as bacterial pneumonia.
- **Bacterial Pneumonia:** The model identifies 126 cases correctly, with 20 misclassified as healthy and 2 as viral pneumonia.

These results indicate that while the model is reasonably effective at distinguishing between the three classes, its precision is somewhat lower than in the binary classification scenario. The binary model's higher accuracy can be attributed to its simpler task of differentiating between just two classes. However, the multi-class model faces the more complex challenge of discerning finer distinctions between three categories, leading to more instances of misclassification. This highlights the trade-offs between model complexity and accuracy, particularly in the context of diagnostic imaging, where distinguishing between different disease types can be significantly more challenging than a simple healthy vs. disease classification.

4 Spatial Transformers

Spatial Transformers, first introduced by Jaderberg et al. [19], are specialized network modules that apply a spatial transformation to feature maps. These transformations enable the network to spatially modify the input data, making it more robust to variations in orientation, position, and scale of the target objects in the images. This capability is useful in medical imaging, where variations in image capture angles, positions, and scales are common [20].

To reap these benefits, I attempted to integrate a Spatial Transformer Network (STN) with my simplified CNN, placed before the convolutional layers. This positioning allows the STN to preprocess the input data, aligning and normalizing the images before feature extraction occurs. This form of preprocessing is expected to enhance the model's ability to extract relevant features more effectively, thereby improving the accuracy of the classification.

The primary functionality of the STN in the model is to afford the CNN greater invariance to rotation, scale, and translation. By enabling the network to automatically correct for such variations, I anticipated a significant improvement in the model's performance, especially in scenarios where the orientation and scaling of the X-ray images are potentially diverse. This integration of STNs into the binary classification model aligns with the growing trend in deep learning to develop more adaptive and flexible architectures, especially in fields where data variability poses a significant challenge, such as medical imaging [21].

4.0.1 Hybrid Model Results

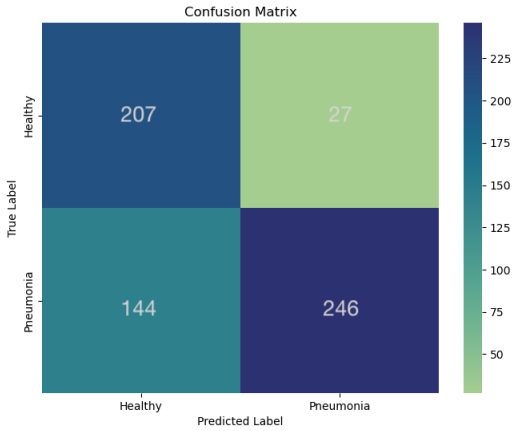

Figure 8: Confusion matrix of the CNN model integrated with a STN

The integration of a STN into the CNN model for binary classification of chest X-rays resulted in a testing accuracy of 72.60%. While this accuracy is significantly lower than previous models without an STN, part of this discrepancy could possibly be explained by model implementation and increased computational intensity of this hybrid approach. The training time was hard coded with an upper limit of 25 epochs, and while this limit was more than sufficient for prior model architectures, the hybrid model ran through all 25 epochs while continuing to show improvements in accuracy, suggesting that it was still in the process of learning and adjusting to the dataset. The STN's role in spatially transforming the input data adds a significant layer of complexity to the training process. This complexity, while beneficial for handling variations in the images, also requires more computational resources and potentially more training time to fully realize its advantages.

Looking at the confusion matrix, we observe that the model correctly identified 207 healthy cases and 246 pneumonia cases. However, it also produced a significant number of false negatives (144), which indicates a tendency towards lower sensitivity.

Given the computational expense of training the model with an STN, further evaluation with extended training or using more powerful computational resources could provide a clearer picture of the STN's utility. The results indicate potential, but also highlight the need for a balance between model complexity, computation, and optimization.

5 Discussion and Further Work

This study's exploration into various Convolutional Neural Network (CNN) architectures, including the integration of a Spatial Transformer Network (STN), highlights the balance needed between model complexity and computational efficiency in medical imaging. The relatively accurate results obtained from

simplified CNN architectures are contrasted with the increased computational demands of the more complex architectures and the STN-hybrid model. From these experiments we saw varying degrees of sensitivity and specificity across different models. This highlights the challenge in fine-tuning models to accurately detect pneumonia while minimizing false positives. Future work could focus on extended training and optimization of the STN-hybrid model, potentially utilizing more powerful computational resources to better utilize the full benefits of spatial transformers.

Further work may include exploring the application of Graph Neural Networks (GNNs) [22] in medical image analysis. GNNs could offer an innovative approach to analyze the complex patterns and relationships in medical images. This exploration could extend to developing hybrid models that combine CNNs' feature extraction capabilities with GNNs' relational data processing strengths, with the goal of creating more accurate classifiers [23].

Finally, to ensure the practical applicability of these models, I would want to include a broader and more diverse dataset of medical images. This would provide a more comprehensive understanding of the models' efficacy across various patient demographics and a wider range of medical conditions.

ACKNOWLEDGMENTS

Inspiration for this project was drawn from Hardik Desmukh's medium article on building CNNs in python [14].

REFERENCES

[1] Eyal Bercovich and Marcia C. Javitt. 2018. Medical Imaging: From Roentgen to the Digital Revolution, and Beyond. *Rambam Maimonides Med J* 9, 4 (October 2018), e0034. https://doi.org/10.5041/RMMJ.10355

[2] Bradley J. Erickson, Panagiotis Korfiatis, Zeynettin Akkus, and Timothy L. Kline. 2017. Machine Learning for Medical Imaging. Radiographics 37, 2 (March 2017), 505–515. https://doi.org/10.1148/rg.2017160130

[3] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A. W. M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. 2017. A survey on deep learning in medical image analysis. Medical Image Analysis 42, (December 2017), 60–88. https://doi.org/10.1016/j.media.2017.07.005

[4] Mingyu Kim, Jihye Yun, Yongwon Cho, Keewon Shin, Ryoungwoo Jang, Hyun-jin Bae, and Namkug Kim. 2019. Deep Learning in Medical Imaging. Neurospine 16, 4 (December 2019), 657–668. https://doi.org/10.14245/ns.1938396.198

[5] Maryellen L. Giger. 2018. Machine Learning in Medical Imaging. Journal of the American College of Radiology 15, 3, Part B (March 2018), 512–520. https://doi.org/10.1016/j.jacr.2017.12.028

[6] Yoshua Bengio, Aaron Courville, and Pascal Vincent. 2013. Representation Learning: A Review and New Perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence 35, 8 (August 2013), 1798–1828. https://doi.org/10.1109/TPAMI.2013.50

[7] Ana Barragán-Montero, Umair Javaid, Gilmer Valdés, Dan Nguyen, Paul Desbordes, Benoit Macq, Siri Willems, Liesbeth Vandewinckele, Mats Holmström, Fredrik Löfman, Steven Michiels, Kevin Souris, Edmond Sterpin, and John A. Lee. 2021. Artificial intelligence and machine learning for medical imaging: a technology review. Phys Med 83, (March 2021), 242–256. https://doi.org/10.1016/j.ejmp.2021.04.016

[8] Eric J. Topol. 2019. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 25, 1 (January 2019), 44–56. https://doi.org/10.1038/s41591-018-0300-7

[9] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. Nature 521, 7553 (May 2015), 436–444. https://doi.org/10.1038/nature14539

[10] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2017. ImageNet classification with deep convolutional neural networks. Commun. ACM 60, 6 (May 2017), 84–90. https://doi.org/10.1145/3065386

[11] Karen Simonyan and Andrew Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. https://doi.org/10.48550/arXiv.1409.1556

[12] Paulo Breviglieri. 2019. Pneumonia X-Ray Images. Retrieved December 05, 2023 from https://www.kaggle.com/datasets/pcbreviglieri/pneumonia-xray-images

[13] Daniel S. Kermany, Michael Goldbaum, Wenjia Cai, Carolina C. S. Valentim, Huiying Liang, Sally L. Baxter, Alex McKeown, Ge Yang, Xiaokang Wu, Fangbing Yan, Justin Dong, Made K. Prasadha, Jacqueline Pei, Magdalene Y. L. Ting, Jie Zhu, Christina Li, Sierra Hewett, Jason Dong, Ian Ziyar, Alexander Shi, Runze Zhang, Lianghong Zheng, Rui Hou, William Shi, Xin Fu, Yaou Duan, Viet A. N. Huu, Cindy Wen, Edward D. Zhang, Charlotte L. Zhang, Oulan Li, Xiaobo Wang, Michael A. Singer, Xiaodong Sun, Jie Xu, Ali Tafreshi, M. Anthony Lewis, Huimin Xia, and Kang Zhang. 2018. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell 172, 5 (February 2018), 1122-1131.e9. https://doi.org/10.1016/j.cell.2018.02.010

[14] Hardik Deshmukh. 2020. Medical X-ray ♯ Image Classification using Convolutional Neural Network. Medium. Retrieved December 05, 2023 from https://towardsdatascience.com/medical-x-ray-%EF%B8%8F-image-classification-using-convolutional-neural-network-9a6d33b1c2a

[15] Connor Shorten and Taghi M. Khoshgoftaar. 2019. A survey on Image Data Augmentation for Deep Learning. J Big Data 6, 1 (July 2019), 60. https://doi.org/10.1186/s40537-019-0197-0

[16] Matías Roodschild, Jorge Gotay Sardiñas, and Adrián Will. 2020. A new approach for the vanishing gradient problem on sigmoid activation. Prog Artif Intell 9, 4 (December 2020), 351–360. https://doi.org/10.1007/s13748-020-00218-y

[17] Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. arXiv.org. Retrieved December 05, 2023 from https://arxiv.org/abs/1412.6980v9

[18] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, Matthew P. Lungren, and Andrew Y. Ng. 2017. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. arXiv.org. Retrieved December 05, 2023 from https://arxiv.org/abs/1711.05225v3

[19] Max Jaderberg, Karen Simonyan, Andrew Zisserman, and koray kavukcuoglu. 2015. Spatial Transformer Networks. In Advances in Neural Information Processing Systems, 2015. Curran Associates, Inc. . Retrieved December 05, 2023 from https://proceedings.neurips.cc/paper_files/paper/2015/hash/33ceb07bf4eeb3da587e268d663aba1a-Abstract.html

[20] Fahad Shamshad, Salman Khan, Syed Waqas Zamir, Muhammad Haris Khan, Munawar Hayat, Fahad Shahbaz Khan, and Huazhu Fu. 2023. Transformers in medical imaging: A survey. Medical Image Analysis 88, (August 2023), 102802. https://doi.org/10.1016/j.media.2023.102802

[21] Peng Xu, Xiatian Zhu, and David A. Clifton. 2023. Multimodal Learning With Transformers: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 45, 10 (October 2023), 12113–12132. https://doi.org/10.1109/TPAMI.2023.3275156

[22] Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. 2020. Graph neural networks: A review of methods and applications. AI Open 1, (January 2020), 57–81. https://doi.org/10.1016/j.aiopen.2021.01.001

[23] Kexin Ding, Mu Zhou, Zichen Wang, Qiao Liu, Corey W. Arnold, Shaoting Zhang, and Dimitri N. Metaxas. 2022. Graph Convolutional Networks for Multi-modality Medical Imaging: Methods, Architectures, and Clinical Applications. https://doi.org/10.48550/arXiv.2202.08916