Multimodal-based Pneumonia Etiology Classification

Maxime Fontana[1], Jooho Lee[1], Daniel Gray[1] Andrei Tihoan[1], Jose Alves[1], and Tom Jarman[1]

University of Sheffield, United Kingdom

Abstract. Pneumonia is a lung infection that results in breathing difficulties, fever and coughing. It is a common disease that can underlie more serious chronic diseases. While the existence of pneumonia can be confirmed fairly quickly, the majority of pneumonia cases are classified as idiopathic i.e. unknown cause, mainly due to insufficient information early on in the care process. This consequently leads to the sub-optimal treatment of patients. Thus, it is necessary to ease the early identification of the etiology of pneumonia cases in order to treat patients in the most optimal way. In order to emulate the cross-media aspect of information doctors consider at the time of diagnosis, we introduce in this paper a deep-learning multimodal binary classification algorithm utilising both static and chest X-ray image features. We show that when performing binary classification of viral and bacterial pneumonia etiologies, this multimodal approach improves over results attained by models trained on a single modality. Thus we show that a multimodal approach is suited to this problem and is able to improve on contemporary single-modality solutions.

Keywords: Pneumonia · Multimodal Deep Learning · Computer vision · Classification · MIMIC

1 Introduction

Pneumonia encapsulates diverse etiologies, of which the most commonly categorised are bacterial and viral pneumonia[1]. Pneumonia patients may be weakened by the disease leaving them susceptible to contracting additional illnesses, and if left untreated can be deadly [2]. Patients who are already immunocompromised, such as those with HIV, are particularly at risk and it makes pneumonia more complicated to treat [3]. For these reasons, it is crucial to identify pneumonia as early as possible.

Pneumonia can be classified based on its underlying cause. This disease is generally Community-acquired (CAP)[4], i.e. usually acquired by a complication due to an aspiration of infectious organisms, which is often a type of Influenza. Pneumonia can additionally be Hospital Acquired Pneumonia (HAP), such as from ventilators during treatment, also known as Ventilator Associated Pneumonia (VAP). VAP and HAP are less far less common than CAP, but have a

higher tendency of becoming terminal.[4] Although these origins can highlight different conditions, notably in the duration of incubation [5], during this study we have disregarded acquisition type and focused primarily on the symptoms. Furthermore, pneumonia is most common in patients aged over 70 or under 5[6]. We only consider patients aged over 15 in order to maintain stability, as young children exhibit different physiological statistics, making them challenging to consider alongside adults.

When it comes to diagnosing pneumonia in a patient, multiple viewpoints are considered. Healthcare professionals primarily rely on exhibitory symptoms such as a fever, shortness of breath, coughing or other respiratory afflictions. In addition to their initial considerations, a comprehensive diagnosis will also take into consideration each patients medical background. Based on this data the decision to perform different laboratory tests is made in order to either confirm a diagnosis or gain additional information about a patient's condition. In most pneumonia-related cases, a chest X-ray will be taken for a clearer understanding of the patient's condition [7]. In this study, we thereby aim to replicate such a multimodal process by providing a joint analysis of chest X-rays and a list of static data features that, when combined, helps gain insight into a patients condition at certain points in time.

Another key component of a patient's care process consists of analysing the cause of a disease. Unfortunately, tracing back the etiology of a pneumonia patient is a task that largely goes unsolved, with less than 10%[8] of cases resolving an etiology.

The most widely used and effective method for identifying pneumonia is using X-ray imaging. During diagnosis, radiologists identify white spots of infiltrates in the lungs as being indicative of viral pneumonia, whilst regions of white are more indicative of bacterial pneumonia. However, the monochromatic nature of X-rays makes it difficult to diagnose viral cases, as less dense regions of lung fluid are harder to detect [7]. Moreover, pneumonia does not always show up on X-rays when the disease is still in its preliminary stages, despite the patient exhibiting symptoms. It can be concluded that radiology in terms of exhaustive pneumonia diagnosis is quite error prone, and commonly results in the misclassification of pneumonia[9] due to the potential subjectivity inherent to the classification of X-ray images. To address the sparse research addressing etiology classification, we propose a multi-modal based binary classification algorithm to predict pneumonia etiology. Disregarding the etiology of infections when deciding on a treatment can have dire consequences, such as the aggravation of viral infections or exacerbation of health issues such as antibiotic resistance, due to the adoption of generalised treatment plans, including potential broad-spectrum antibiotics.

In the coming section we introduce the datasets we utilise during our study and how we derived a study cohort from MIMIC-IV and MIMIC-CXR. We then present how we designed our Chest X-ray analysis algorithm by using a Convolutional Neural Network (CNN), and how different static features related to a patient's admission were captured as additional information. We conclude

the paper by presenting and discussing the classifications results obtained by our models.

2 Data

2.1 Datasets

In this study we use the publicly available **MIMIC-IV**[10] and **MIMIC-CXR**[11] datasets. The former is a large dataset composed of various medical records of patients from the *Beth Israel Deaconess Medical Center* Critical care units, spanning from the years 2008 to 2019. The latter is composed of 277,385 radiology studies including 377,100 radiographs from different view positions, linked to 64,588 patients from MIMIC-IV. Additionally, every radiograph study has a free-text radiology report embedded.

In our study we will use the preprocessed version of MIMIC-CXR, **MIMIC-CXR-JPG**. This dataset includes additional tables containing metadata from the original DICOM radiology formats. Additionally, the dataset contains structured labels derived from the free-text radiology reports with the Natural Language Processing tools NegBio[12] and **CheXpert**[13], used on radiology reports, they classify common pathologies, such as pleural effusion, into Positive, Negative, Uncertain and No Mention classes. We derive from these labels the information used for our cohort selection process.

2.2 Cohort Selection

As our model training algorithm focuses on supervised learning over the derived labels, a suitable cohort of samples had to be obtained from this large dataset. Figure 1 shows the process taken to select this cohort, and we then proceed to detail the decisions made during this process.

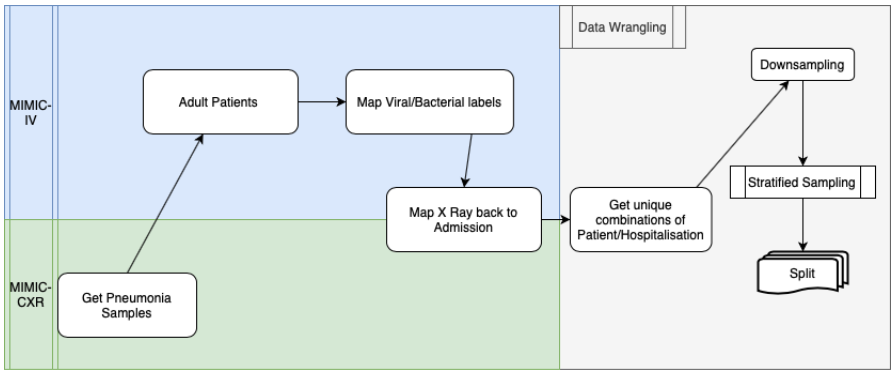

Fig. 1: Cohort Selection Pipeline

The first step consisted of acquiring pneumonia cases from **MIMIC-CXR**[11]. This consisted of retrieving radiology studies for which the report, encoded via the previously introduced NegBio[12] tool, presented a presence or uncertainty for pneumonia. We initially kept samples that both either indicated the certain presence of pneumonia or the uncertainty of its presence. As all samples are later filtered based on the appropriate ICD-9/10 diagnosis codes, the approximation over the uncertain samples is justified. We then remove patients aged under 15 as children present a different morphology and would therefore introduce too much bias in the data distribution.

The final labels for each sample were determined from the bank of ICD-9 and ICD-10 codes as shown below in Table 1. As ICD codes are ranked at diagnosis time on most to least prevalent, the highest ranking ICD code that relates to our binary classification label was taken as the ground truth. Subsequently, we generate ground-truth labels by identifying samples for which diagnoses, ranked as most important with respect to the provided patient, follow under the different codes banks presented in Table 1.

Table 1: A table showing how each hospitalisation was classified into the binary system based on ICD-9/10 code.

ICD-9/10 Code	Description	Class
480	Viral Pneumonia	Viral
487.0	Influenza with pneumonia	Viral
488.01	Influenza due to identified avian influenza virus with pneumonia	Viral
488.11	Influenza due to identified 2009 H1N1 influenza virus with pneumonia	Viral
488.81	Influenza due to identified novel influenza A virus with pneumonia	Viral
484.1	Pneumonia in cytomegalic inclusion disease	Viral
481	Streptococcus pneumoniae pneumonia	Bacterial
482	Other bacterial pneumonia	Bacterial
484.5	Pneumonia in anthrax	Bacterial
J12.X	Viral pneumonia	Viral

Note: Any subcategories of codes were also considered to be the same class as the mentioned parent code.

As there is no direct mapping between the two datasets, linking a radiology report to a hospitalisation record, we align the X-ray images with the hospitalisation record according to the date and time of the study. Specifically, we consider whether the time of the study is between the admission and discharge times relating to the hospitalisation record for a patient.

Finally, we perform some operations to ensure the uniqueness of patient-hospitalisation pairs in order to avoid conflicting diagnosis for the same hospitalisation. We then perform down-sampling on the bacterial samples to mitigate the unbalanced nature of the labels, taking the same number of samples from the Bacterial class with comparatively more samples as the less populated Viral class. We lastly perform stratified sampling to acquire the necessary Training, Validation and Testing sets.

Given all of these methods We produce a cohort containing presenting the attributes summarised in Table 2.

Table 2: A table describing the size of each dataset and its distribution used for machine learning purposes.

Dataset	Viral Samples	Bacterial Samples
Training set	57	57
Validation set	11	11
Test set	23	23

3 Methods

3.1 Image Classification

Problem Elaboration The identification of pneumonia from Chest X-Rays is relatively straightforward and leaves little room for debate. However, identifying the etiology of a given case by only utilising the radiographic image is a much more complex task. In general, viral pneumonia is expressed by a 'diffuse' inflammatory reaction usually involving both lungs, whereas a typical bacterial case will show a more 'condensed' white area in sparse areas [14]. Moreover, bacterial pneumonia can typically be diagnosed as non-viral as it shows less distinctive characteristics. The figure below shows a clear difference between samples of different types:

A challenge to this classification task is that the stage of the disease has a large impact on how visible some of its identifiable features can be. Utilising earlier stage cases can harm the precision of the classifier and leads to less pronounced discriminating features being present.

Model We implemented a DenseNet-121 model, which is a type of Convolutional Neural Network that has proven to be highly accurate for complicated image classification tasks, as it allows the network to learn complex patterns at a good computational cost, by going deeper in the number of convolutional layers [16]. Figure 3 below depicts the architecture of this CNN.

Fig. 2: Pneumonia X-ray Differences [15]

In this study we used the open-source TorchXRayVision library and its X-Ray image handling functionality implemented for the Pytorch framework[17]. Utilising TorchXRayVision's standard transform process, each image was scaled to a size of 224x224 and normalised into a range of [-1024, 1024] per pixel value. To mitigate over-fitting, the training data was augmented prior to the training process by random rotation of up to 15 degrees, a translation of up to 10% of the total image size and random scaling by a factor between 0.5 and 1.5.

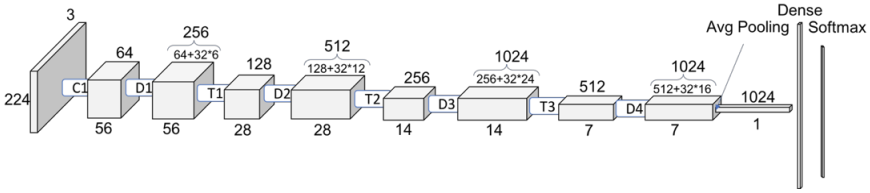

Fig. 3: DenseNet Architecture [18]

3.2 Static Data Classification

Static Features We use a variety of features extracted from the MIMIC-IV dataset. Globally, we can group our features in two categories. The first category is related to a **static feature** i.e. information unlikely to change throughout an admission for a given patient. The second category is classified as an **admission feature**, which is a measurement obtained during hospitalisation at a given time. These admission features are therefore either an approximation or related to a **patient's history** i.e. some diagnoses that are deemed to be correlated with a certain type of pneumonia.

The first features we added were gender and age. MIMIC-IV is HIPAA compliant [19], therefore all patients over the age of 89 were joined together into an age group of 91 to prevent identification, which has a low impact on the accuracy of our model as pneumonia most acutely affects adults over the age of 70.

We also collect the white blood cell count and hematocrit score for each patient, as these demonstrate a higher proportion of white blood cells, which are indicative of the body fighting off an infection.

Finally we collect a selection of chronic illnesses which are correlated with pneumonia. Specifically, we gathered diagnoses of AIDS, metastatic cancer, Mycoplasma Pneumoniae, and Staphylococcus.

We have excluded acute illnesses, such as influenza, that would otherwise be counterproductive to our study due to their high correlation with specific etiologies, as these features would be derived from icd9 codes submitted alongside the etiology. Therefore, when using our model to predict an etiology in real life we would not have an accurate representation of these illnesses.

Brief descriptions of these static features, as well as their value ranges are shown in Table 5.

Table 3: A table describing each static feature, and the pre-standardised range of possible values.

Feature	Description	Range of values (pre-standardised)
Age	The age of the patient during the hospital admission	$15 \leq x \leq 91$
Gender	Whether the patient is male or female	$x \in 0, 1$
AIDS	Whether the patient has been previously diagnosed with Acquired Immunodeficiency Syndrome	$x \in 0, 1$
Hematocrit	The patient's red blood cell percentage min/max/mean	$15 \leq x \leq 60$
MSCancer	Whether the patient has been previously diagnosed with metastatic cancer	$x \in 0, 1$
Mycoplasma	Whether the patient has been previously diagnosed with Mycoplasma Pneumoniae	$x \in 0, 1$
Staphylococcus	Whether the patient has been previously diagnosed with Staphylococcus Aureus	$x \in 0, 1$
Whitebloodcells	The patient's white blood cell count min/max/mean	$0 \leq x \leq 1000$

When considering features related to a chronic illness, we do not treat them as positive during the admission in which they are first diagnosed, as that data is unlikely to have been available at the time the etiology was determined.

When the patient does not have prior diagnoses that relate to a chronic illness, we consider them to not have that illness. We can only take the absence as data to be negative, as it has not been diagnosed.

In the event that a patient has no prior diagnoses recorded, we leave these features missing rather than assuming them to be negative. The presence of each feature is shown in Table 4

Table 4: Cohort patient attributes w.r.t the features

Attribute	Viral	Bacterial	Overall
Number of Samples	86	721	807
Age (μ/σ)	(64.85/18.14)	(64.38/16.91)	(64.43/17.04)
Gender (Male/Female)	(40/46)	(397/324)	(437/370)
Prior AIDS Diagnoses	3	14	17
Prior Metastatic Cancer Diagnoses	4	31	35
Prior Mycoplasma Pneumoniae Diagnoses	1	2	3
Prior Staphylococcus Diagnoses	0	171	171

When data was missing, we replaced the missing values with the mean value from that feature. We then standardised each feature by subtracting the mean and dividing by the standard deviation to calculate the standardised scores.

Model We use a Multi-Layer Perceptron with the configuration shown in figure 4

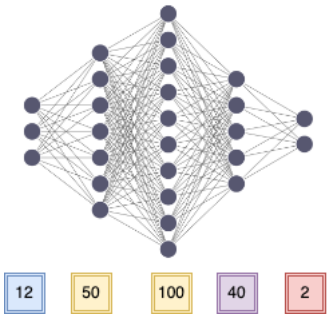

Fig. 4: MLP Architecture

3.3 Multi-Modal Architecture

The solution we propose consists of 3 different blocks summarised in the figure below :

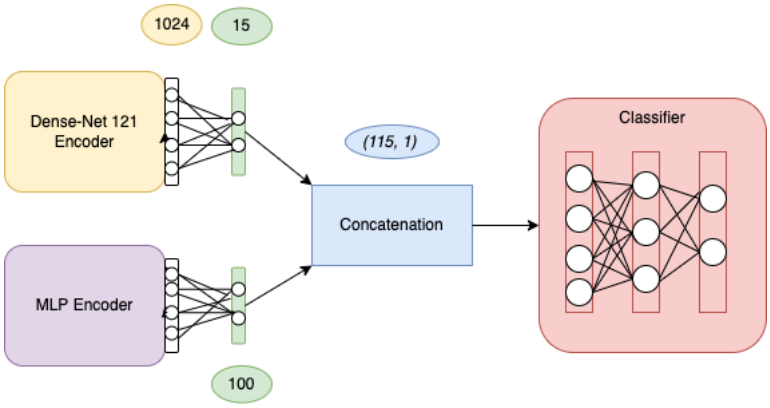

Fig. 5: Multi-Modal Neural Network Architecture

Each individual modality has a corresponding encoder. Subsequently, simple late fusion with concatenation is used in order to merge the 2 modalities. This concatenated result is passed to a Multi-layer-perceptron in order to make the final classification.

4 Results and Discussion

4.1 Results

The results were obtained through the utilisation of the open-source framework MultiBench [20]. This provided several functions that aided in both unimodal and multimodal training and evaluation. We have implemented our previously described multimodal architecture and other training functionality for model assessment.

The following configuration describes the final model training and evaluation process:

- Initialisation : We initialise the CNN encoder with pre-trained weights provided by TorchXRayVision : *'densenet121-res224-all'*, which was trained over datasets including MIMIC-CXR, NIH and CheXPert[21]. The classifier and the MLP encoder are initialised with random values between 0 and 1.
- Loss Function : Cross Entropy, as defined in the PyTorch library class `CrossEntropyLoss`.
- Optimiser : The Adam optimiser[22] was used in order to perform stochastic gradient descent.
- Multimodal Fusion : Simple late fusion with concatenation was utilised. In spite of its simplicity, it performs relatively well on medical data, even when compared with more sophisticated fusion methods.[20].

Overfitting was a severe problem due to the smaller sample size used for this study. Therefore, measures such as early stopping were used when training to

mitigate overfitting on the training set. Performance metrics averaged over 20 runs are reported below in Table 5, and Figures 6, 7 and 8.

Table 5: Performance Metrics

Model	Accuracy	Precision ($\boldsymbol{V}$)	Precision ($\boldsymbol{B}$)	F1-Score ($\boldsymbol{V}$)	F1-Score ($\boldsymbol{B}$)
DenseNet-121	0.523	0.508	0.457	0.439	0.428
Static	0.593	0.566	0.636	0.600	0.535
Multi-Modal	0.610	0.633	0.639	0.599	0.583

$\boldsymbol{V}$ refers to the Viral Pneumonia class, and $\boldsymbol{B}$ refers to the Bacterial Pneumonia class.

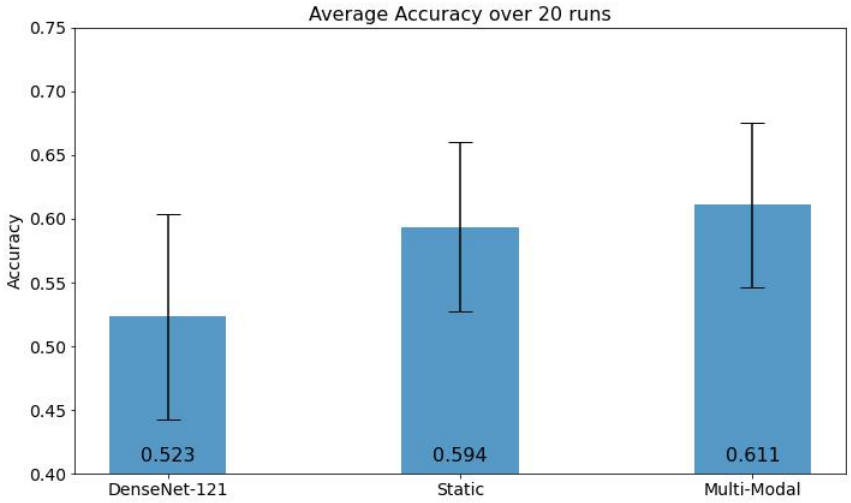

Fig. 6: Average Accuracies over 20 runs

4.2 Discussion

The aforementioned solution for determining the cause of pneumonia relies on strong assumptions and therefore we will discuss their different impacts to this study. We will also bring in different topics worth exploring to mitigate these weaknesses, with the aim of providing a starting point to extend this study.

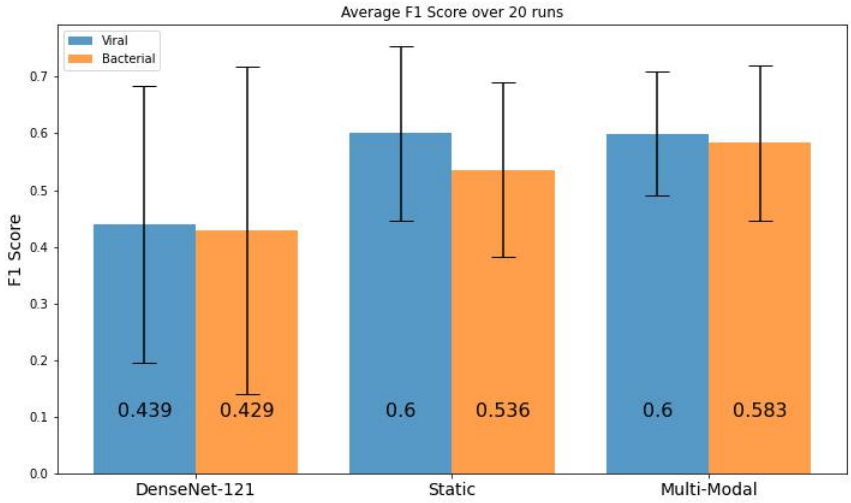

Fig. 7: Average F1 Scores per class over 20 runs

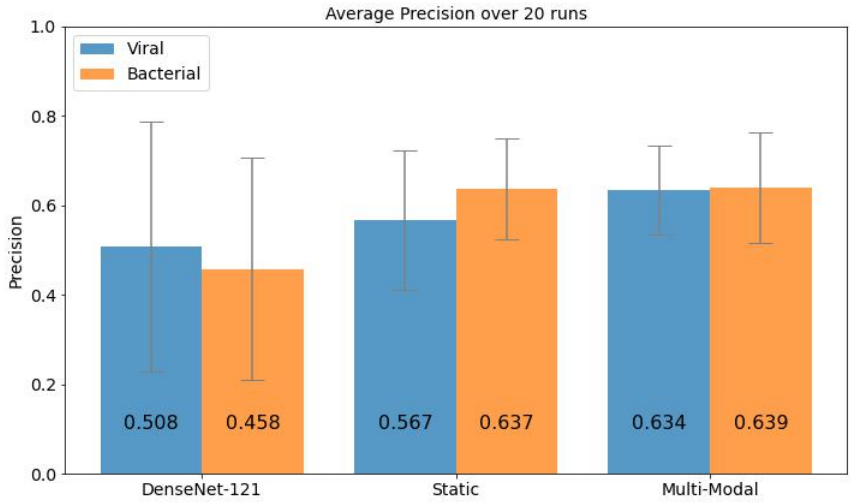

Fig. 8: Average Precision per class over 20 runs

Curse of Bias The pipeline we introduced in the previous sections is heavily reliant on premises that are bias-prone. The course of action under which our ground-truth is established is further prone to errors.

The main reason for this ambiguity is attributed to diagnoses being provided by different doctors. Our initial filtering techniques are based on radiology re-

ports obtained via the use of the NLP tool: NegBio[12], which can make wrong classifications.

Our second source of ground-truth comes from the billing established at a patients discharge time, as a result, we do not know if the procedure from start to finish is reliable.

Additionally, the way the dataset is constructed introduces ambiguity as diagnosis are ranked based on their severity and a ranking cannot fully reflect the importance of these certain diagnoses. However it is evident that our proposed method, in the way it is constructed, cannot remove the human bias introduced in the structure of the data collected. Humans generally make mistakes and any solution relying on such information will not be capable of breaching the human decision-making barrier.

Finally, a predominant source of uncertainty emerges from the non-representativity of our samples that causes high variance in performance calculations reported in figure 6.

Curse of Approximation Artificial Intelligence (AI) systems are in their very nature, approximation systems. The trade-off is always to consider what the necessary amount of data is in order to generalise well to new data. Our solution is no exception. A prominent strength of our study is the small amount of features the prediction requires, and therefore, the realistic application it brings. However, our analysed data is often approximated to be over the same time period; for instance it is very unlikely that the different blood measurements and X-rays were obtained at the same instant.

Additionally, trying to estimate the amount of features necessary to carry out the classification task would be difficult. Moreover, one would be buried in complexity in an attempt to render all the different modalities a doctor has access to during the diagnosis procedure.

Further Work From an engineering perspective, research in Multimodal fusion has introduced sophisticated methods that may be more suitable to our purposes, especially considering the different dimensions of our latent representations. MultiBench[20] is designed to handle this testing and such experimentation could significantly improve over the current results.

As possible extensions to the presented model and to mitigate the two different problems introduced above, one could firstly consider turning this fully supervised method into a hybrid method such as semi-supervised. The justification for such an idea lies in the fact that there are few obtainable data samples from the datasets, and also to mitigate the strong assumption with regard to the ground-truth bias. Similarly, a fully multimodal unsupervised method could potentially perform better and help differentiate between all the different etiologies pneumonia can have.

Moreover to incorporate another dimension in our model, another potentially rich source of data is accessible through MIMIC-IV; specifically, time-series data acquired throughout an admission depicting different measurements obtained

from the patient over a small timeline. However, one would have to account for the considerable amount of missing data for the concerned patients.

5 Conclusion

We have proposed a novel method to the rarely tackled and important problem of estimating the two major causes of pneumonia in order to improve the treatment of patients. We have shown that classifying in a multimodal context, using both static information related to a patient over different timescales and chest X-rays, outperforms the respective single-modality with an average improvement in accuracy of approximately 2%. We have also used a range of open-source tools to derive a dataset tailored to our purposes. With respect to this dataset and our framework, we have critically assessed the different advantages and drawbacks of our study and accordingly, we have discussed some possible extensions that could potentially present better performance and eventually become useful in a pragmatic setting.

References

1. L. C. Jennings, T. P. Anderson, K. A. Beynon, A. Chua, R. T. R. Laing, A. M. Werno, S. A. Young, S. T. Chambers, and D. R. Murdoch, "Incidence and characteristics of viral community-acquired pneumonia in adults," *Thorax*, vol. 63, no. 1, pp. 42–48, 2008.
2. V. Jain, R. Vashisht, G. Yilmaz, and A. Bhardwaj, "Pneumonia pathology," 2022.
3. M. F. Di Pasquale, G. Sotgiu, A. Gramegna, D. Radovanovic, S. Terraneo, L. F. Reyes, J. Rupp, J. González del Castillo, F. Blasi, S. Aliberti, and M. I. Restrepo, "Prevalence and Etiology of Community-acquired Pneumonia in Immunocompromised Patients," *Clinical Infectious Diseases*, vol. 68, pp. 1482–1493, 08 2018.
4. J. Burnham and M. Kollef, "Cap, hcap, hap, vap: The diachronic linguistics of pneumonia," *Chest*, vol. 152, pp. 909–910, 11 2017.
5. S. Eccles, C. Pincus, B. Higgins, and M. Woodhead, "Diagnosis and management of community and hospital acquired pneumonia in adults: summary of nice guidance," *BMJ*, vol. 349, 2014.
6. R. Lozano, N. Fullman, J. E. Mumford, M. Knight, C. M. Barthelemy, C. Abbafati, H. Abbastabar, F. Abd-Allah, M. Abdollahi, A. Abedi, H. Abolhassani, A. E. Abosetugn, L. G. Abreu, M. R. M. Abrigo, A. K. Abu Haimed, A. I. Abushouk, M. Adabi, O. M. Adebayo, V. Adekanmbi, ..., and C. J. L. Murray, "Measuring universal health coverage based on an index of effective coverage of health services in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019," *The Lancet*, vol. 396, no. 10258, pp. 1250–1284, 2020.
7. C. J. Saul, D. Y. Urey, and C. D. Taktakoglu, "Early diagnosis of pneumonia with deep learning," 2019.
8. J. G. Bartlett, "Diagnostic Tests for Agents of Community-Acquired Pneumonia," *Clinical Infectious Diseases*, vol. 52, pp. S296–S304, 05 2011.
9. N. J. Brendish, A. K. Malachira, K. R. Beard, L. Armstrong, P. J. Lillie, and T. W. Clark, "Hospitalised adults with pneumonia are frequently misclassified as another diagnosis," *Respiratory Medicine*, vol. 150, pp. 81–84, 2019.

10. J. A., B. L., P. T., H. S., C. L. A., and R. Mark, *'MIMIC-IV' (version 1.0)*, 2021.
11. A. E. W. Johnson, T. J. Pollard, N. R. Greenbaum, M. P. Lungren, C.-y. Deng, Y. Peng, Z. Lu, R. G. Mark, S. J. Berkowitz, and S. Horng, "Mimic-cxr-jpg, a large publicly available database of labeled chest radiographs," 2019.
12. Y. Peng, X. Wang, L. Lu, M. Bagheri, R. Summers, and Z. Lu, "Negbio: a high-performance tool for negation and uncertainty detection in radiology reports," 2017.
13. I. Jeremy, R. Pranav, K. Michael, Y. Yifan, C.-I. Silviana, C. Chris, M. Henrik, H. Behzad, B. Robyn, S. Katie, S. Jayne, M. D. A., H. S. S., S. J. K., J. Ricky, L. D. B., L. C. P., P. B. N., L. M. P., and A. Y. Ng, "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison," 2019.
14. Pfizer, "Viral vs bacterial pneumonia understanding the difference," 2022.
15. D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, J. Dong, M. K. Prasadha, J. Pei, M. Y. Ting, J. Zhu, C. Li, S. Hewett, J. Dong, I. Ziyar, A. Shi, R. Zhang, L. Zheng, R. Hou, W. Shi, X. Fu, Y. Duan, V. A. Huu, C. Wen, E. D. Zhang, C. L. Zhang, O. Li, X. Wang, M. A. Singer, X. Sun, J. Xu, A. Tafreshi, M. A. Lewis, H. Xia, and K. Zhang, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131.e9, 2018.
16. G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," 2016.
17. J. P. Cohen, J. D. Viviano, P. Bertin, P. Morrison, P. Torabian, M. Guarrera, M. P. Lungren, A. Chaudhari, R. Brooks, M. Hashir, and H. Bertrand, "Torchxrayvision: A library of chest x-ray datasets and models," 2021.
18.
19. J. de Groot, "What is hipaa compliance?," 2022.
20. P. P. Liang, Y. Lyu, X. Fan, Z. Wu, Y. Cheng, J. Wu, L. Chen, P. Wu, M. A. Lee, Y. Zhu, R. Salakhutdinov, and L.-P. Morency, "Multibench: Multiscale benchmarks for multimodal representation learning," 2021.
21. J. P. Cohen, M. Hashir, R. Brooks, and H. Bertrand, "On the limits of cross-domain generalization in automated x-ray prediction," 2020.
22. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014.