

Viral and Bacterial Pneumonia Diagnosis via Deep Learning Techniques and Model Explainability

Hai Thanh Nguyen¹

College of Information and
Communication Technology
Can Tho University
Can Tho, Vietnam

Toan Bao Tran²

Center of Software Engineering, Duy Tan University,
Da Nang, 550000 Vietnam
Institute of Research and Development, Duy Tan University,
Da Nang, 550000 Vietnam

Huong Hoang Luong³

Department of Information Technology
FPT University
Can Tho, Vietnam

Trung Phuoc Le⁴

Department of Information Technology
FPT University
Can Tho, Vietnam

Nghi Cong Tran⁵

National Central University
Taoyuan, Taiwan

Abstract—Pneumonia is one of the most serious diseases for infants and young children, people older than age 65, and people with health problems or weakened immune systems. From numerous studies, scientists have found that a variety of organisms, including bacteria, viruses, and fungi, can be the cause of the disease. Coronavirus pandemic (COVID-2019) which comes from a type of pneumonia has been causing hundreds of thousands of deaths and is still progressing. Machine learning approaches are applied to develop models for medicine but they still work as a black-box are difficult to interpret output generated by machine learning models. In this study, we propose a method for image-based diagnosis for Pneumonia leveraging deep learning techniques and interpretability of explanation models such as Local Interpretable Model-agnostic Explanations and Saliency maps. We experiment on a variety of sizes and Convolutional neural network architecture to evaluate the efficiency of the proposed method on the set of Chest x-ray images. The work is expected to provide an approach to distinguish between healthy individuals and patients who are affected by Pneumonia as well as differentiate between viral Pneumonia and bacteria Pneumonia by providing signals supporting image-based disease diagnosis approaches.

Keywords—Interpretability; pneumonia; x-rays images; bacterial and viral pneumonia; image-based disease diagnosis

I. INTRODUCTION

According to the World Health Organization (WHO), pneumonia is one of the most infectious causes of death worldwide, it affects children and families everywhere and causes 50 thousand deaths each year. Recently, the Situation Report - 150 from WHO [1] about COVID-19 presented the number of infected active cases is up to 8, 2 million, the number of deaths is 445, 535. The patients can get pneumonia as a complication of viral infections such as COVID-19 or the common flu. Besides, the bacteria, fungi, and other microorganisms can also be the primary infectious agents of pneumonia, causing cough with phlegm or pus, fever, chills, and difficulty breathing. Pneumonia is the infection of one or both lungs and filled up with fluid and pus. Based on [2], up to 60% of the cases are related to respiratory virus infections. The study in [3] indicated the difference between viral and bacterial pneumonia

in children is based on the serum C reactive protein (CRP) but the sensitivity is not enough for use in clinical practice.

In the field of medicine, diagnostic radiology is significant and used for disease assessment. To detect the pneumoconiosis early and accurately after the clinical analysis, performing a chest X-ray is important and necessary. It is critical to preventing complications including death. The expert radiologists assessed the X-ray images for the pneumonia fluid in the lungs during the diagnosis. Specifically, Fig. 1 contains images of viral (left image), bacterial (middle image) and normal chest (right image) pneumonia. Viral pneumonia presents the diffuse interstitial pattern in both lungs, whereas, bacterial pneumonia typically exhibits a focal lobar consolidation.

Furthermore, chest X-ray (CXR or chest radiography) can reveal the abnormalities areas and not only produce images of the chest but also the nearby structures. Nevertheless, the X-ray images consist of black and white colors, it is quite difficult for detecting the infected areas in the images. Additionally, the technical level of radiologists is also important to make the diagnosis correctly. A study in [4] conducted an education in-person training for improving chest radiograph interpretation accuracy among non-radiologists clinicians.

In recent years, the complexity of medical data makes it more difficult for analyzing and diagnosing the disease. In parallel, the improvements in Machine Learning and Deep Learning have a certain influence on image processing in general and medicine in particular. The diagnosis process performs with Machine Learning or Deep Learning can help physicians investigate the medical images conveniently and reduce the analysis time. Several studies have resolved the challenging tasks such as medical image classification [5], [6], skin cancer detection using images [7], or 3D image biomedical segmentation [8].

Moreover, Deep Learning-based technologies have successfully demonstrated in clinical practice including clinical decision support systems (CDSS), diagnosis prediction, and predicting the invasiveness of lung adenocarcinoma manifesting based on radionics and clinical features [9]. Though, there

are still several challenges with Machine Learning. Selecting a dataset, creating a predictive model, and evaluating and refining the model, the most important thing is data [10]. The implementations of Machine Learning or Deep Learning in health care are influenced by the accuracy of medical data. Specifically, the annotation progress in the medical image is based on medical professional knowledge, medical industry standard, and medical system [6].

II. RELATED WORK

The exceptional improvement of Deep Learning and large datasets have facilitated for the replacement of artificial intelligence for human gradually. As we mentioned above, several studies have outperformed the performance of medical experts.

A study from Stanford University Machine Learning Laboratory [11] proposed a Deep Learning structure, namely CheXNet which contains 121 convolutional layers for detecting pneumonia. They evaluated the model on a large dataset - ChestX-ray14 [12], including over 100,000 chest X-ray images with 14 diseases. The authors resized the original images to 224×224 and applying the random horizontal flipping before training. Their proposed method can detect all 14 diseases in ChestX-ray14. However, it also contains several limitations. The X-ray images are in the frontal view for training and testing, but a study in [13] indicated that up to 15% of accurate diagnoses need the lateral view. Furthermore, the patient records are not allowed to use, which has affected negatively by radiologist diagnostic performance [14].

The authors in [15] studied the powerful performance of Residual Neural Network (ResNet) [16] on several diseases using the ChestX-ray14 dataset for classification tasks. The high spatial resolution of X-ray images is investigated by the extended ResNet-50 and the non-image features including patient age, gender, and view position are transformed into a non-image feature vector and concatenated with the image feature vector. In general, the integration with the non-image feature reached the best overall performance, the detailed analysis of the non-image feature has been provided.

The following approach [17] takes advantage of the Laplacian of Gaussian (LoG) filtering to improve the performance of the Convolutional Neural Networks. The considered dataset contains 247 radiograms from the publicly available Japanese Society of Radiological Technology dataset (JSRT). The original images are downsized to 96×96 and applied the LoG filter before training or testing. The performance of the Deep Learning model is evaluated based on the detection of the nodule in X-ray images, the results reached better performance in comparison with AlexNet [18] and GoogleNet [19].

The authors in [20] proposed an approach that supports the disease diagnosis with trained models and Gradient Class Activation Map (Grad-CAM) method [21]. The results from Grad-CAM have been generated based on the features that the models pay the most attention to. But the performance of Grad-CAM depends on how good the models are. This study also investigated the performance on a specific size of the image, which is 64×64 with a shallow Convolutional Neural Networks.

In an attempt to describe the chest radiographs of patients with bacterial and viral pneumonia, the study in [22] let the

radiologists reviewed the chest X-ray from pneumonia patients. The results stated that the comparison of bacterial and viral pneumonia is insignificant differences and chest radiographs are hard to recognize between bacterial and viral pneumonia.

In this study, we propose a CNN-based method to distinguish between bacterial and viral pneumonia, stratify healthy samples and patients. We also explain output from trained models with model-agnostic and saliency maps to extract signals in images for the diagnosis. Our contributions include:

- We introduce Convolutional Neural Networks to discriminate healthy individuals and patients who were affected by viral pneumonia or bacterial pneumonia. We also distinguish viral pneumonia and bacterial pneumonia from chest radiography. The performances on several categories are evaluated (For instance, normal-viral-bacterial, normal-viral, normal-bacterial, and viral-bacterial).
- Various sizes of images, in specific, 64×64 , 96×96 , and 256×256 are also carried out to compare the performance.
- We leverage the advantages of Local Interpretable Model-agnostic Explanations (LIME) [23] and Saliency maps [24] for visualizing the discriminate features in the X-ray images. The disease diagnosis process can be easier with the proposed method.
- Oversampling technique is also implemented in the case where we face imbalanced image datasets for prediction tasks. As a result, the performance in image classification is improved.

In the remainder of this study, we introduce the considered dataset and model structures in Section III. The works with LIME and Saliency maps are explained in Section IV. Our experimental results are presented in Accuracy, Area Under the Curve (AUC) of the Convolutional Neural Networks for several classification tasks in Section V. We conduct and summarize some remarks in Section VI.

III. CONVOLUTIONAL NEURAL NETWORK FOR PNEUMONIA IMAGES

A. Dataset

The publicly available dataset from Guangzhou Women and Children's Medical Center, Guangzhou [25] is considered to evaluate our method. The quality control and quality assurance have been done by two expert physicians. In specific, the low quality or hard-to-read images were removed, then, the images were classified by the experts. The dataset includes 5856 X-ray images in which, 1583 images are labelled *Normal* and 4273 for *Pneumonia*. To categorize the viral and bacterial pneumonia, we split the *Pneumonia* class into two sub-classes, namely *Viral* and *Bacterial* with 1493 and 2780 samples respectively, more details are in Table I. Fig. 1 visualizes the samples of 3 classes as we mentioned above. For the classification tasks, we split the dataset into a training set and validation set randomly with a ratio of 9 : 1.

We have three binary classification tasks and one multi-class (three classes) classification task. The X-ray images are

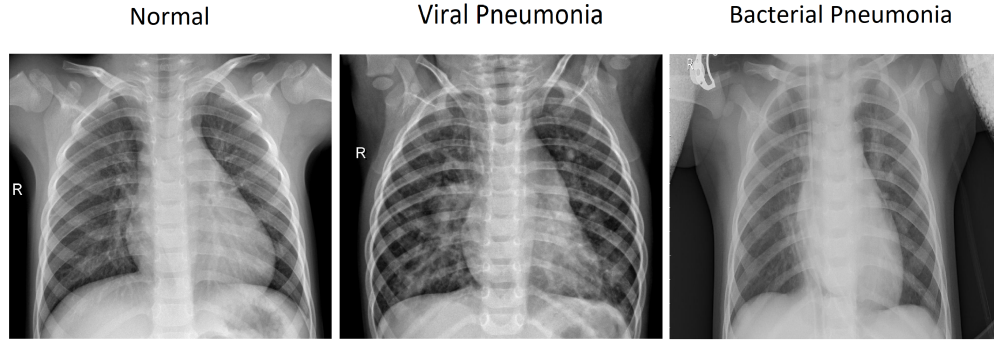


Fig. 1. Examples of Viral Pneumonia, Bacterial Pneumonia and Normal Chest Samples via X-ray Images.

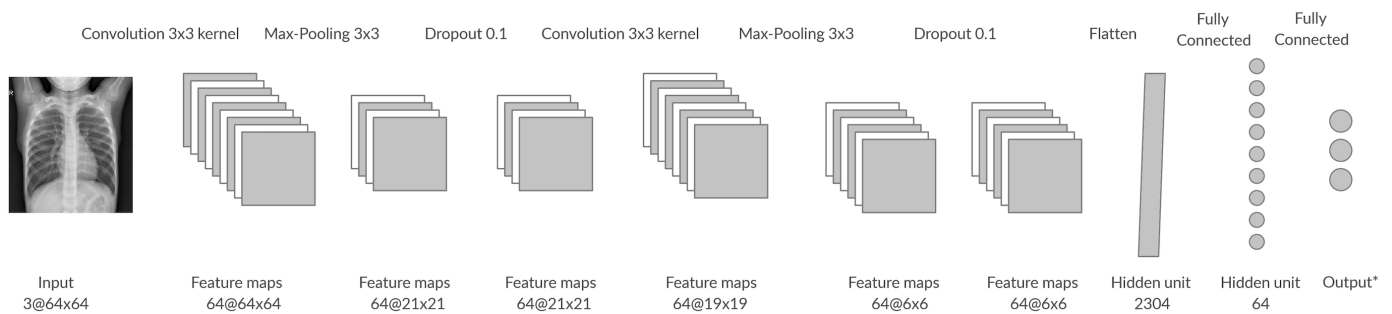


Fig. 2. A Convolutional Neural Networks Architecture for the Classification Tasks. “**” denotes the Number of Output Neurons depends on the Classification Tasks (if the Binary Classification Task needs One Output whereas, for the Classification of Three Classes (the Normal Sample-Viral Pneumonia-Bacterial Pneumonia, the Network Produces Three Output).

TABLE I. INFORMATION ON X-RAY IMAGE DATASET [25]

Class	Bacterial	Normal	Viral	Total
Samples	2780	1583	1493	5856

the input for all problems and the output is to predict the label of the input which indicates the normal, absence, or presence of viral and bacterial pneumonia.

To compare the difference of the CNN performance on the size of images, we test both images of 64×64 , 96×96 , and 256×256 . The purpose of using two various sizes of images is to investigate the performance of Convolutional Neural Networks for classification tasks. We need a good model for explaining the predictions with LIME and Saliency maps.

B. Convolutional Neural Networks and Settings

The architecture of Convolutional Neural Networks includes two Convolutional layers containing 64 filters of 3×3 for each layer, followed by a Max-Pooling of 2×2 (stride 2), a dropout rate of 0.1 and a Fully Connected layer with 64 neurons. The architecture is illustrated in Fig. 2. For binary classification tasks, we set the number of neurons in the output layer to 1 with a *sigmoid* activation function. Otherwise, the output for classifying patients, viral pneumonia, and bacterial

pneumonia is shown in Fig. 2 with three neurons. For multi-class classification, we use the softmax activation function.

The considered CNN is carried out with Adam optimizer [26] as the optimized function with standard parameters. The cross-entropy loss is also implemented for optimization purposes. We used the default learning rate of 0.001. During training, if the loss is not improved after every 5 consecutive epochs, the training section will be stopped by the Early Stopping method and the model with the lowest validation loss will be saved.

IV. LOCAL INTERPRETABLE MODEL-AGNOSTIC EXPLANATIONS

The LIME method can explain the predictions of any machine learning models, the main purpose of this technique is to understand the model by perturbing the input and understanding the change of the predictions. More specifically, LIME visualizes the contribution of each feature to the prediction from the input, it also allows determining which feature changes will affect the prediction mostly. The explanation of an input x can be obtained by the formula 1 [23].

$$\xi(x) = \operatorname{argmin}_{g \in G} \mathcal{L}(f, g, \pi_x) + \Omega(g) \quad (1)$$

The authors in [23] denote $g \in G$ as an explanation, $\Omega(g)$ is a measure of complexity of $g \in G$. $f(x)$ is the score of the

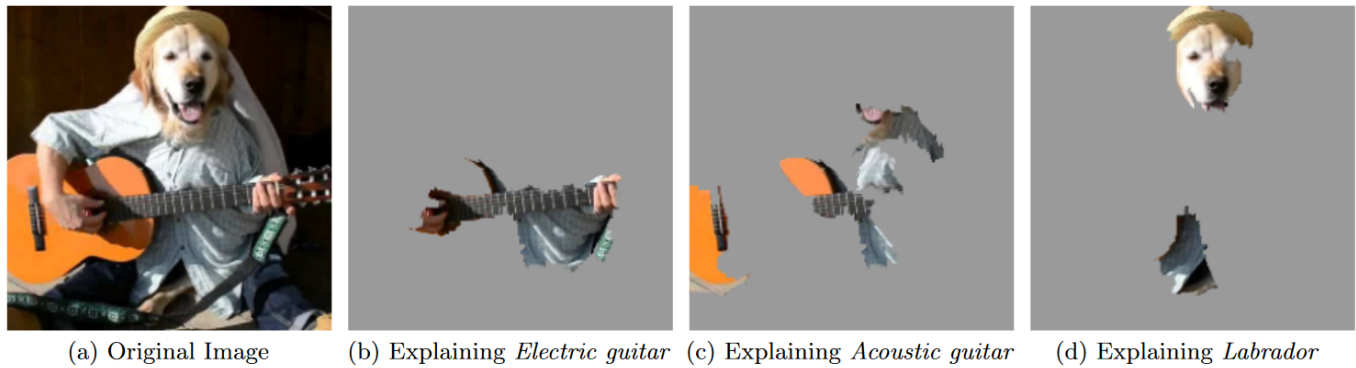


Fig. 3. An Example of the Explanations of an Image which Predicted by Inception Neural Network [23], [27].

relevant class. $\pi_x(z)$ is only found with respect to a proximity measure in the neighborhood of x . The $\mathcal{L}(f, g, \pi_x)$ is responsible of g to f while a low loss is desirable indicating high local fidelity. An example of LIME explanation is illustrated in Fig. 3 [23]. Fig. 3a contains the original image whereas Fig. 3b, Fig. 3c, and Fig. 3d display the explanation for *ElectricGuitar* class, *AcousticGuitar* class, and *Labrador* class respectively.

We also applied Saliency maps to our CNN for discriminating the features in images. Saliency maps offer a visualization of somewhere in the image that the model pays attention to and contributes the most to predictions. Saliency maps are usually visualized as a heatmap where the highlighted pixels are the important points that affect the decision of the model. Fig. 4 illustrates the important regions within an image that contribute the most to the output, by calculating the gradient of a class output concerning the input image via back-propagation.

V. EXPERIMENTAL RESULTS

A. Metrics for Comparison

The performance of the model for classification tasks is evaluated by computing the average accuracy on 10-folds stratified-cross validation. The classification accuracy is defined by the number of correct predictions divided by the total number of predictions and multiplied by 100. We also computed the Receiver Operating Characteristic Curve (ROC-AUC) for assessing the performance. ROC is a probability curve and AUC represents the degree or measure of separability. It presents the capability of distinguishing between classes of the model.

In this section, we present the results of different tasks including performance comparison on the image sizes, the discriminant performance comparison of healthy individuals and bacterial pneumonia, between normal samples and viral pneumonia samples, and exhibitions of differences of viral and bacterial pneumonia.

B. Comparison with the Different Size of Images

The chart in Fig. 5 visualizes the comparison between three sizes of images. We investigated the performance of model on both sizes 64×64 , 96×96 , and 256×256 for classifying three classes *Normal*, *Viral*, and *Bacterial*. The images

with a size of 64×64 reached the overall average accuracy of 0.862 and 0.834 of AUC on 10-folds stratified-cross validation whereas the images classification with the size of 96×96 is slightly better (it reached 0.873 and 0.843 of accuracy and AUC respectively). However, the increment of sizes is not helpful due to the shallow architecture and it is inconsonant for the large sizes images. As a result, the overall accuracy and AUC reached 0.855 and 0.826 for the size of 96×96 , respectively. The Accuracy and AUC column are horizontally stacked together, the left column represent for the Accuracy whereas the right of AUC. Fig. 6 visualizes the training and validation accuracy in 60 epochs where the learning is stopped by the overfitting issue. The initial performance is quite good, validation accuracy for the first epoch is 0.83 whereas the training validation reaches 0.74 and gets better after 10 epochs. The peak validation accuracy reaches 0.88 and is around 0.84 to 0.87.

C. Experimental Results on the Classification between Normal Samples and Bacterial Pneumonia Samples

This task, we examined the performance of classifying Bacterial and Normal classes. We trained the proposed architecture on 10-folds stratified-cross validation. The results are quite good, the average accuracy reached 0.958 and AUC of 0.988. We also plotted the Accuracy-AUC chart in Fig. 9. The explanations of LIME and saliency maps are visualized in Fig. 7 and Fig. 8, respectively. In both Fig. 7 and Fig. 8, the numbered images No. 1 and No. 2 belong to Bacterial class. Otherwise, No. 3 and No. 4 are Normal class. Furthermore, “label” in image caption means the true label of the image, we assume that 0, 1, and 2 represent for *Bacterial*, *Normal*, and *Viral* respectively. Also, “p” in the caption of the LIME’s output means the predicted label. They also describe the same thing for the upcoming Figures of LIME and saliency maps.

Specifically, in Fig. 7, the list of yellow dots in the images represents for the discriminate features that contribute to the final output of Convolutional Neural Networks. We only take the features that positively contribute to the prediction of the label for visualizing.

Furthermore, we also investigated the heatmap rendered by the saliency maps method which represents the conspicuity of the model which illustrated in Fig. 8. In the case of Bacterial pneumonia, the areas of inflammation in the lungs are pointed

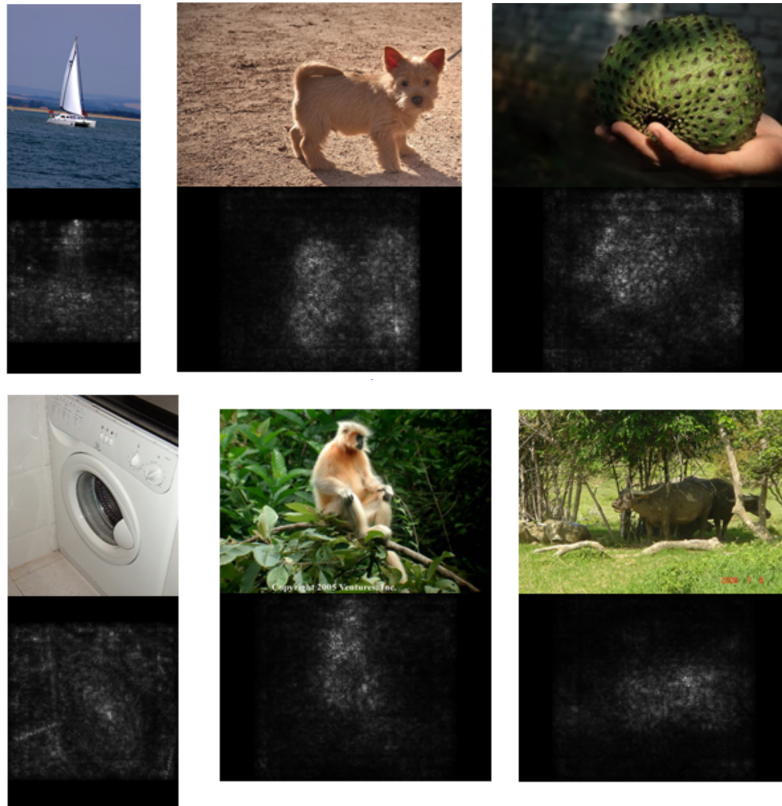


Fig. 4. Visualizations of Samples on the Image-Specific Classes by the Saliency Maps. [24].

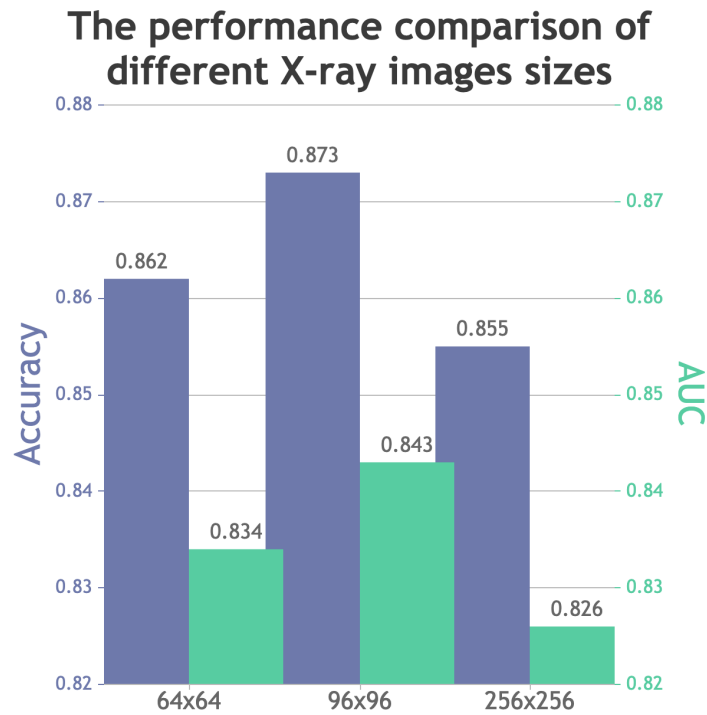


Fig. 5. Performance Comparison of Different Images Sizes on the Considered Image Dataset.

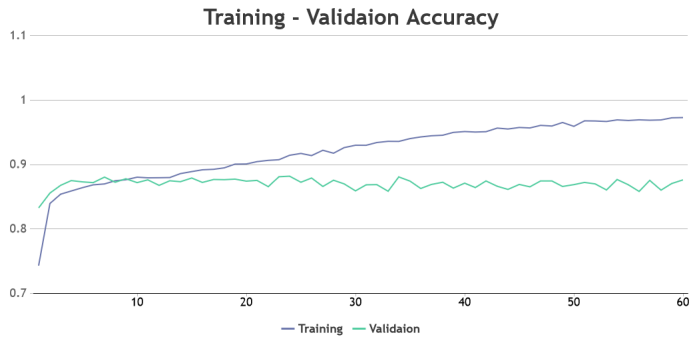


Fig. 6. Performance Visualization of Training and Validation Accuracy in 60 epochs.

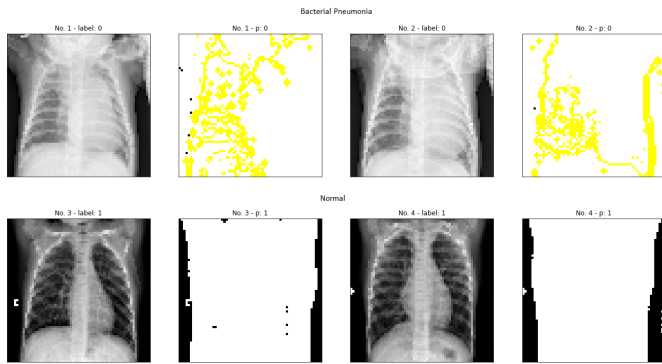


Fig. 7. The Exhibition of *Normal* and *Bacterial* Classes based on the Predicted Output from CNN using LIME.

out by the green bright pixels in the images. Otherwise, the heatmaps visualization presents the lungs with some noises.

D. Experimental Results on Normal Chest and Viral Pneumonia

The efficiency of classifying between Normal and Virus is quite good, the chart in Fig. 12 exhibits the average validation performance reaching 0.939 in Accuracy and an AUC of

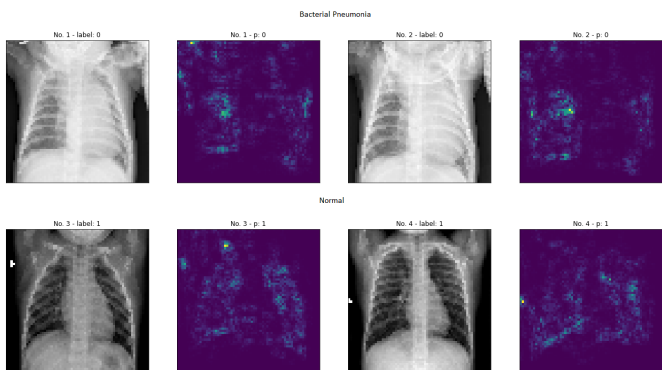


Fig. 8. Image Specific Class Saliency Maps for the Predicted *Normal* and *Bacterial* Classes.

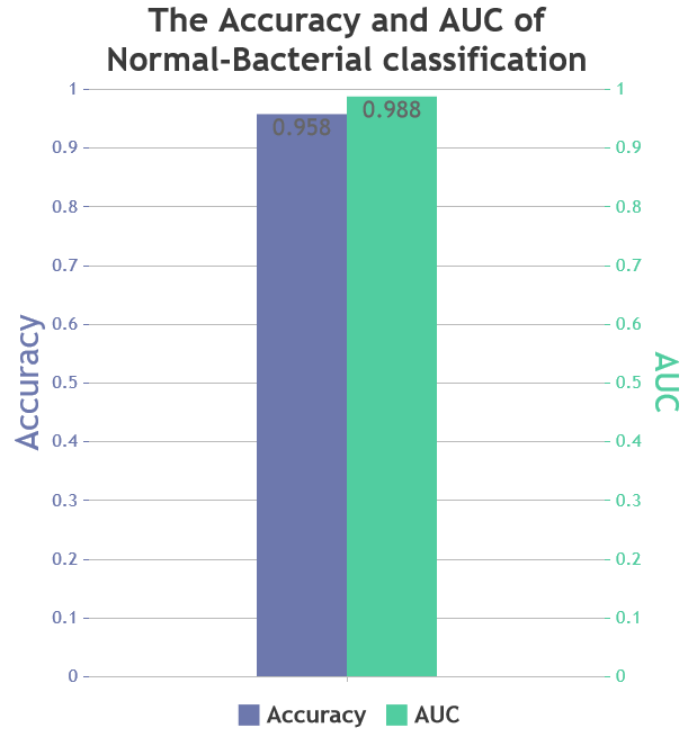


Fig. 9. The Validation Accuracy and AUC of the Normal-Bacterial Classification Task.

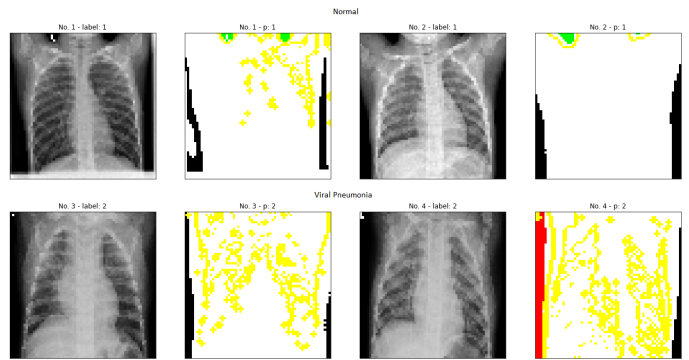


Fig. 10. The Explanation of *Normal* and *Viral* Classes based on the Predicted Output from CNN using LIME.

0.981. In comparison with the prior task, the performance slightly decrease. We also applied LIME and saliency maps for explaining the predicted results and visualized them in Fig. 10 and Fig. 11. As we mentioned above, the images come with “label” state their true label and “p” declare the predicted label.

The Normal images in Fig. 10 are described nearly the same with the normal in Fig. 7, the contribute features in images are different from the Bacterial. The explanations of Bacterial might provide insights into pneumonia, the abnormalities in lungs have been listed as yellow points.

Fig. 11 illustrates the highlighted features in images of Normal and Viral class by computing the gradient of output from the Convolutional Neural Networks. With Normal class,

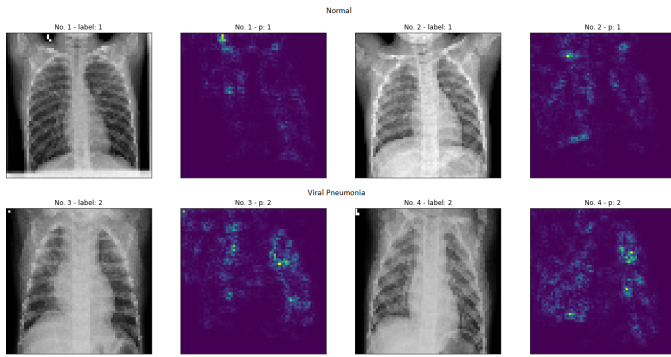


Fig. 11. Image Specific Class Saliency Maps for the Predicted *Normal* and *Viral* Classes.

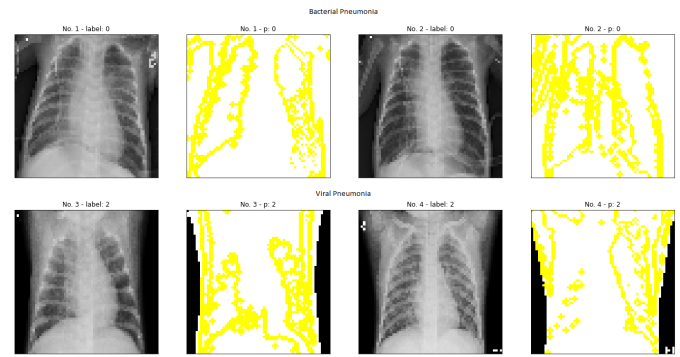


Fig. 13. The Explanation of *Bacterial* and *Viral* Classes based on the Predicted Output from CNN using LIME.

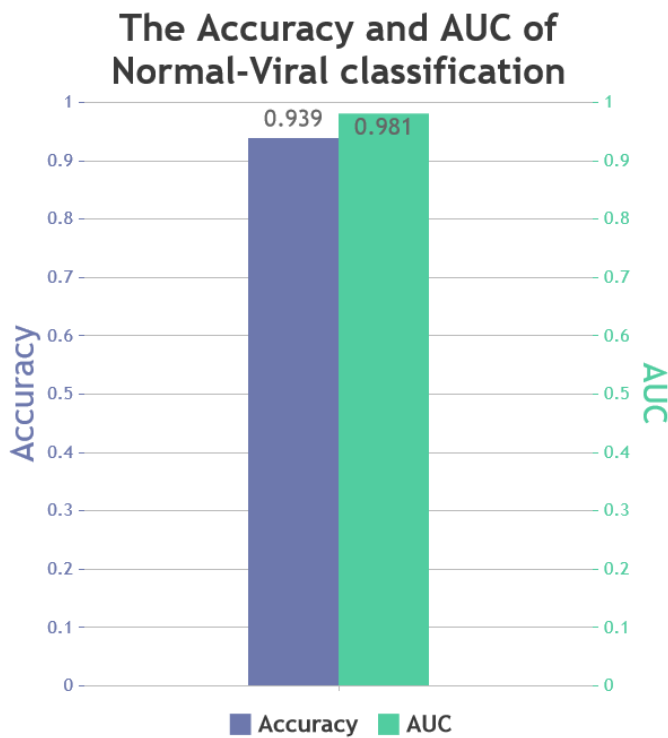


Fig. 12. The Validation Accuracy and AUC of the Normal-Viral Classification Task.

the most contributions to the final output have been around the lungs. On the other hand, the presence of pneumonia has been spotted by the superpixels.

E. Experimental Results on the Bacterial and Viral Pneumonia Classification

Bacterial and Viral pneumonia images are very similar and not easy to discriminate them. The classification performance is significantly reduced on this task due to the similarity of bacterial and viral pneumonia. Moreover, imbalanced data may influence performance. As we proposed in Section III, the Bacterial class contains 2780 images (65%) whereas the number of samples in Normal class is 1493 (35%). The validation

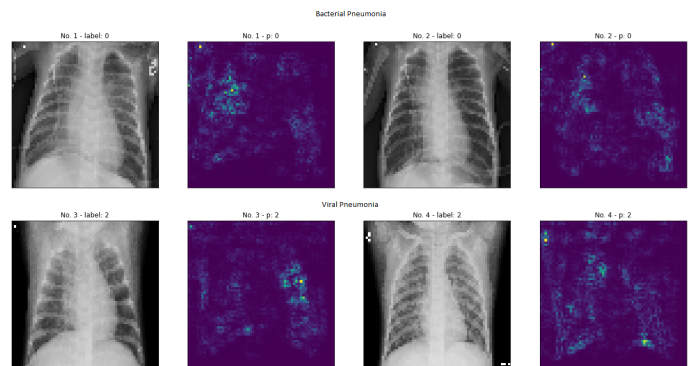


Fig. 14. Image Specific Class Saliency Maps for the Predicted *Bacterial* and *Viral* Class.

accuracy and AUC reached 0.756 and 0.799, respectively. To handle the imbalanced classification problem, we applied the oversampling method [28]. In specific, we randomly duplicated the samples in the training set of Viral class, the number of generated images is 1200. We retrained the model with the proposed parameters, the accuracy and AUC increased slightly, 0.808 for accuracy and 0.888 of AUC. We also visualized the performance in Fig. 15. The imbalanced column represented the accuracy and AUC of the imbalanced classification. Otherwise, the Balanced column proposed the increased performance after applying the oversampling method.

Fig. 13 and Fig. 14 present the discriminate features of Bacterial and Viral pneumonia which generated by LIME and saliency-maps-based method respectively. Observing the Bacterial and Viral X-ray images as non-radiologists is a challenge, it seems to be confused. The explanation of Bacterial images tends to be localized and consolidation whereas the Viral images are diffuse and interstitial markings. Furthermore, the classification performance has a certain influence on the LIME or saliency results.

We also presented the overview of three classes (Bacterial, Normal, and Viral) with LIME and saliency method in Fig. 16 and Fig. 17 respectively. The description of normal and each pneumonia type are explained as above.

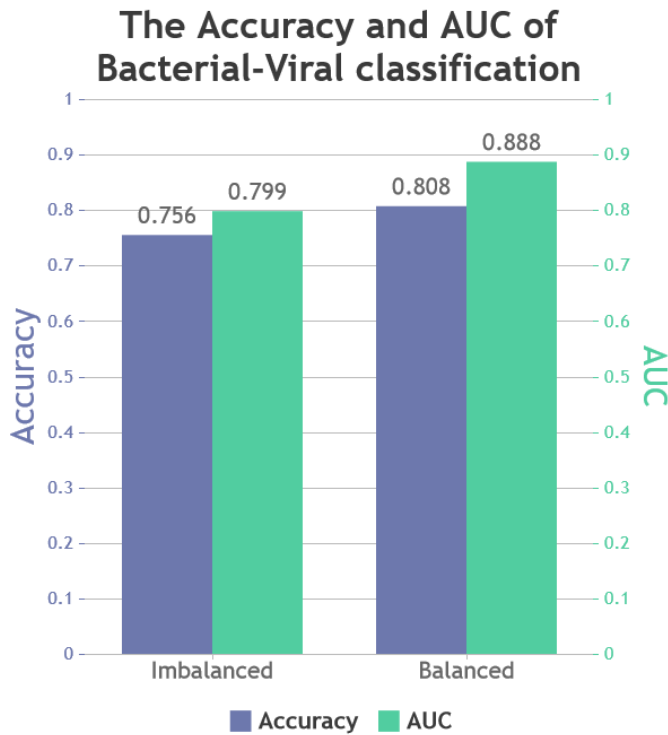


Fig. 15. The Validation Accuracy and AUC of Bacterial-Viral Classification Task for Imbalanced and Balanced Classification.

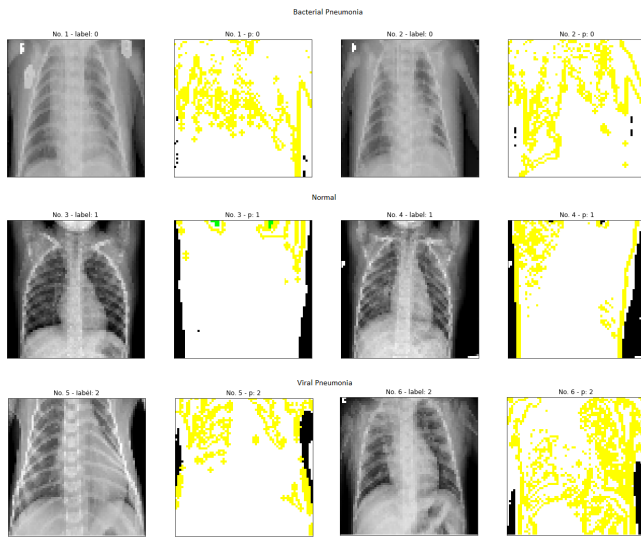


Fig. 16. The Explanation of *Bacterial*, *Normal*, and *Viral* Classes based on the Predicted Output from CNN using LIME.

VI. CONCLUSION

We presented an approach to combine the convolutional neural network and model explainability to support pneumonia diagnosis. The pneumonia classification X-ray images-based have been popular in recent years, applying several machine learning techniques for explaining the predictions

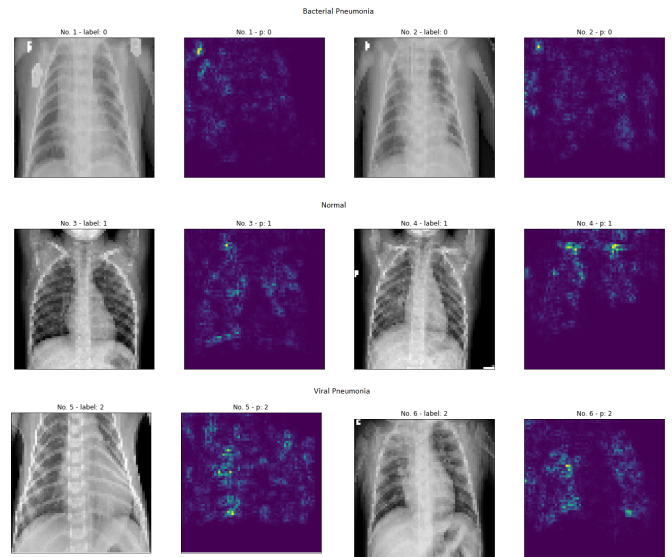


Fig. 17. Image Specific Class Saliency Maps for the Predicted *Bacterial*, *Normal*, and *Viral* Classes.

is a promising approach. As shown from the results, the proposed method is a promising approach so that the doctors who unfortunately own fewer experiences due to some limited conditions can leverage the proposed method to improve the diagnosis accuracy and speed up the pneumonia diagnosis. In our experimental results, the performance of explanations is almost based on how good the classifier is. Moreover, the complexity of the model also influences the results of high-resolution image classification. The proposed architecture learning model has acceptable results on several tasks but with our limitations, it still challenges classifying Viral and Bacterial pneumonia. The pneumonia diagnosis with X-ray images and explain the results by machine learning are significant and play a key role in medical diagnosis.

Various sizes of images are also investigated and evaluated. In our experiments, with higher images in the same CNN architecture, the performance can be improved but the results in classification with very large images, for instance, 256×256 , can be not high. The Very larger image sizes may need more sophisticated models to enhance the performance and they also require more computation resources. Due to the limitations of computation resources, we only use shallow CNN architectures with a few convolutional layers on the small images. Further research should take into account deeper architecture to enhance classification tasks.

The Oversampling technique is attempted and helps to improve the performance where we only face imbalanced images datasets. Further research can continue to investigate to find better approaches to imbalanced issues for image datasets.

REFERENCES

- [1] World Health Organization. Coronavirus disease Situation Report-150. Retrieved from <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>. June 2020.
- [2] Ruuskanen O, Mertsola J. Childhood community-acquired pneumonia. *Semin Respir Infect.* 1999;14(2):163-172. 1999.

- [3] Virkki R, Juven T, Rikalainen H, Svedström E, Mertsola J, Ruuskanen O. Differentiation of bacterial and viral pneumonia in children. *Thorax*;57(5):438-441. doi:10.1136/thorax.57.5.438. 2002.
- [4] Fawole, O.A., Kelly, M.S., Steenhoff, A.P. et al. Interpretation of pediatric chest radiographs by non-radiologist clinicians in Botswana using World Health Organization criteria for endpoint pneumonia. *Pediatr Radiol* 50, 913–922 (2020). <https://doi.org/10.1007/s00247-020-04625-0>. 2020.
- [5] Wang X, Yu J, Zhu Q, et al. Potential of deep learning in assessing pneumoconiosis depicted on digital chest radiography [published online ahead of print, 2020 May 29]. *Occup Environ Med*. 2020;oemed-2019-106386. doi:10.1136/oemed-2019-106386. 2020.
- [6] H. Tang and Z. Hu, "Research on Medical Image Classification Based on Machine Learning," in *IEEE Access*, vol. 8, pp. 93145-93154, 2020, doi: 10.1109/ACCESS.2020.2993887. 2020.
- [7] Esteva, A., Kuprel, B., Novoa, R. et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 115–118 (2017). <https://doi.org/10.1038/nature21056>. 2017
- [8] K. Tseng, Y. Lin, W. Hsu and C. Huang, "Joint Sequence Learning and Cross-Modality Convolution for 3D Biomedical Segmentation," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 3739-3746, doi: 10.1109/CVPR.2017.398. 2017.
- [9] Lan Song, Tongtong Xing, Zhenchen Zhu, Wei Han, Guangda Fan, Ji Li, Huayang Du, Wei Song, Zhengyu Jin, Guanglei Zhang, Hybrid Clinical-Radiomics Model for Precisely Predicting the Invasiveness of Lung Adenocarcinoma Manifesting as Pure Ground-Glass Nodule, *Academic Radiology*, 2020, ISSN 1076-6332. 2020. <https://doi.org/10.1016/j.acra.2020.05.004>.
- [10] Verghese A, Shah NH, Harrington RA. What This Computer Needs Is a Physician: Humanism and Artificial Intelligence. *JAMA*. 2018;319(1):19-20. doi:10.1001/jama.2017.19198. 2017.
- [11] Rajpurkar, Pranav, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Yi Ding, Aarti Bagul, Curtis P. Langlotz, Katie S. Shpanskaya, Matthew P. Lungren and Andrew Y. Ng. "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning." *ArXiv abs/1711.05225* (2017); n. pag. 2017.
- [12] Wang, Xiaosong, Peng, Yifan, Lu, Le, Lu, Zhiyong, Bagheri, Mohammadhadi, and Summers, Ronald M. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. *arXiv preprint arXiv:1705.02315*, 2017.
- [13] Raoof S, Feigin D, Sung A, Raoof S, Irugulpati L, Rosenow EC 3rd. Interpretation of plain chest roentgenogram. *Chest*. 2012;141(2):545-558. doi:10.1378/chest.10-1302. 2012.
- [14] Berbaum K, Franken EA Jr, Smith WL. The effect of comparison films upon resident interpretation of pediatric chest radiographs. *Invest Radiol*. 1985;20(2):124-128. doi:10.1097/00004424-198503000-00004. 1985.
- [15] Baltruschat, Ivo M., Hannes Nickisch, Michael Grass, Tobias Knopp and Axel Saalbach. "Comparison of Deep Learning Approaches for Multi-Label Chest X-Ray Classification." *Scientific Reports* 9 (2019). 2019.
- [16] He, Kaiming, Xiangyu Zhang, Shaoqing Ren and Jian Sun. "Deep Residual Learning for Image Recognition." 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016): 770-778. 2016.
- [17] Candemir, Sema, and Sameer Antani. "A review on lung boundary detection in chest X-rays." *International journal of computer assisted radiology and surgery* vol. 14,4 (2019): 563-576. doi:10.1007/s11548-019-01917-1. 2019.
- [18] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'12)*. Curran Associates Inc., Red Hook, NY, USA, 1097–1105. 2012..
- [19] C. Szegedy et al., Going deeper with convolutions, 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 1-9, doi: 10.1109/CVPR.2015.7298594. 2015.
- [20] Hai Thanh Nguyen, Hoang Thanh Huynh, Toan Bao Tran, Hiep Xuan Huynh. Explanation of the Convolutional Neural Network Classifying Chest X-Ray Images Supporting Pneumonia Diagnosis. *EAI Endorsed Transactions on Context-aware Systems and Applications*. ISSN: 2409-0026. Vol. 7. DOI: 10.4108/eai.13-7-2018.165349. 2020.
- [21] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization", 2017 IEEE International Conference on Computer Vision (ICCV), Venice, 2017, pp. 618-626, doi: 10.1109/ICCV.2017.74. 2017.
- [22] Robles A., San Gil A., Pascual V. Viral vs bacterial community-acquired pneumonia: radiologic features. *Eur. Respir. J*. 2011;38:2507.
- [23] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "Why should i trust you?: Explaining the predictions of any classifier". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM. 2016, pp. 1135–1144. 2016.
- [24] Simonyan, Karen et al. "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps." *CoRR abs/1312.6034* (2014)
- [25] Kermany, Daniel; Zhang, Kang; Goldbaum, Michael (2018), "Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification", *Mendeley Data*, v2 <http://dx.doi.org/10.17632/rscbjbr9sj.2>. 2018.
- [26] Diederik P. Kingma and Jimmy Lei Ba. Adam : A method for stochastic optimization. 2014. *arXiv:1412.6980v9*. 2014.
- [27] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *CVPR*, 2015
- [28] Nitesh V. Chawla, Debray, Kevin W. Bowyer, Lawrence O. Hall and W. Philip Kegelmeyer. SMOTE: Synthetic Minority Over-Sampling Technique. *Journal of Artificial Intelligence Research* 2002; 16: 321–357.