

Lung diseases classification based on auto-machine learning

Changhao Zhou

Department of Information System, City University of Hong Kong, HKSAR, China

chazhou7-c@my.cityu.edu.hk

Abstract. The escalating incidence of lung diseases and the substantial costs associated with their diagnosis underscore the significance of developing approaches that can enhance the effectiveness and precision of detection. In this regard, this study employs an auto-machine learning platform called Edge Impulse to train a greyscale dataset of lung diseases. Specifically, the study compares two classification models and two transfer learning models generated by edge impulse and BrainChip, respectively, and employs accuracy as a metric to assess the efficacy of these models. The research analyzes the potential causes behind the differing performances of the two authors and the two types of models and selects the optimal model for implementation in real-world devices. Additionally, the study takes into account the accuracy in detecting individual types of diseases to evaluate the model's performance in detailed classification. The findings reveal that the most effective model, the transfer learning model developed by Edge Impulse, performs exceptionally well in detecting abnormal lungs, particularly in diagnosing COVID-19. However, its classification capabilities with respect to identifying the specific type of disease remain limited.

Keywords: auto-machine learning, lung disease classification, edge impulse.

1. Introduction

Lung diseases, including COVID-19, pneumonia, and tuberculosis, are among the most prevalent and serious health problems globally. The early and accurate detection of these diseases is crucial for timely and effective treatment, as well as for mitigating their impact on public health. However, traditional diagnostic methods such as chest radiography and computed tomography (CT) scans can be time-consuming, expansive, and require specialized expertise. Moreover, the diagnose based on the images is subject to inter-observer variability, depending on the experience and subjective judgment of the medical professionals. These limitations can lead to the increased misdiagnosis rate, resulting in higher medical cost in both human resources and wealth, as well as impeding patients' treatment process.

Recently, artificial intelligence (AI) and machine learning (ML) have demonstrated great potential for improving the detection and diagnosis of lung diseases. ML algorithms can analyse vast amounts of data and identify patterns and correlations that are challenging to detect by human experts, enabling more accurate and efficient diagnosis. Furthermore, with the advancement of deep learning (DL), artificial neural networks with more complex and sophisticated structures can be trained to analyse medical images and signals with remarkable accuracy and precision. For instance, Cai and Gao introduced the application of intelligent imaging in the preliminary diagnosis process [1]. O.S. Albahri

and A.A. Zaidan researched on many AI techniques application in detecting COVID-19, such as multi-criteria decision analysis [2]. Zhang and Xie used synergic deep learning to classify medical images [3]. However, their models may not have the best performance, and are difficult to find the optimized parameters.

To solve the limitation mentioned above, Auto Machine Learning (AutoML) is an emerging field of AI that aims to streamline the entire machine learning pipeline, from data pre-processing to model selection and hyperparameter tuning [4-6]. The development of AutoML has been driven by the increasing demand for ML solutions and the shortage of experienced data scientists and ML experts. AutoML algorithms can automatically analyse the given dataset, select the most appropriate ML algorithms, and optimize their hyperparameters to achieve the best performance. This process can significantly reduce the time and effort required to develop ML models and can also increase the accuracy and robustness of the models by eliminating the potential for human biases and errors.

This paper will use AutoML to analyse a lung disease classification dataset. The dataset includes 8104 lung X-ray images, including three diseases of COVID-19, pneumonia, tuberculosis, and a normal type. In this paper, transfer learning and classification models will be used to train the dataset. After training the model, accuracy will be calculated to evaluate the model. If the accuracy is high enough, the model can be applied to devices to do auto detection. Else, this paper will find the potential reasons and make improvements to the model. Overall, this paper aims to find a solution to improve accuracy and efficiency of medical diagnosis by using AutoML.

2. Method

2.1. Dataset preparation

The dataset used in this study sourced from the famous competition platform called Kaggle [7]. It contains 6,305 grayscale X-ray images of lung diseases, including tuberculosis, COVID-19, pneumonia, and normal health lung. The dataset is divided into a training dataset of 4,847 (77%) images and test dataset of 1,458 (23%) images. The input dimension is set to size of 96px×96px. Figure 1 and Figure 2 show the data distribution, and Figure 3 provides some sample images of the collected dataset.

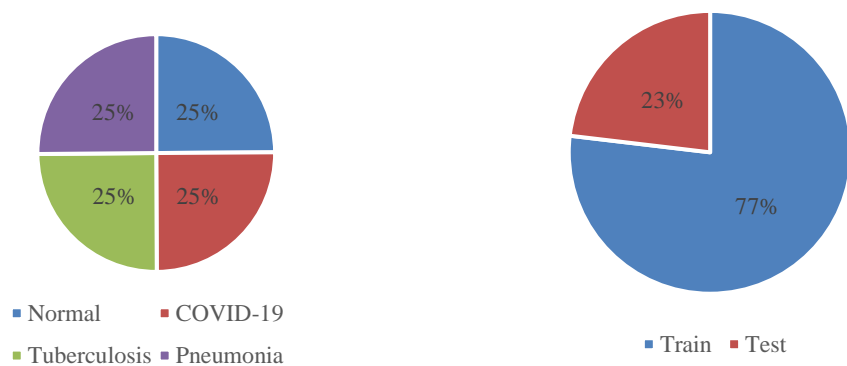


Figure 1. Data distribution.

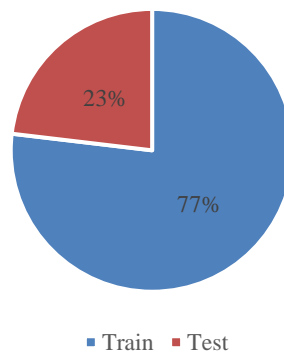


Figure 2. Train and test split.

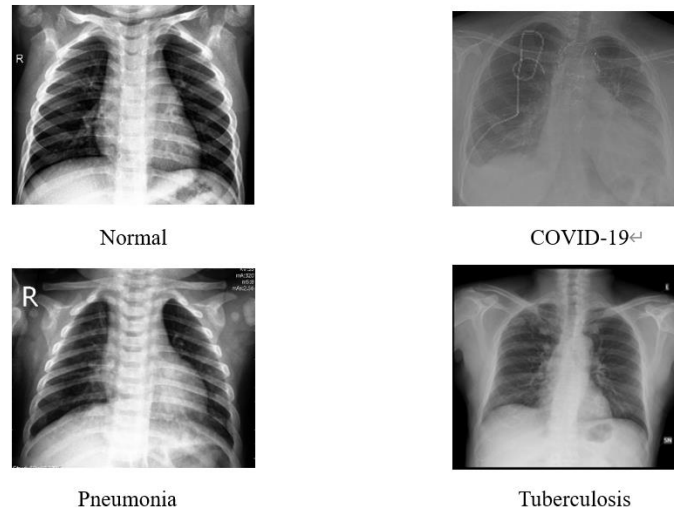


Figure 3. The sample images of the collected dataset.

2.2. Edge impulse-based lung diseases recognition model

2.2.1. Introduction of edge impulse. Edge Impulse which is considered in this study is a platform that enables developers to build and deploy machine learning models on embedded devices. It provides a user-friendly interface for collecting data, training models, and deploying them on microcontrollers. Edge Impulse also supports transfer learning, which allows users to take advantage of pre-trained models and adapt them to their specific use cases.

2.2.2. Introduction of CNN. Convolutional neural networks (CNNs) are a type of neural network that are commonly used in computer vision applications, such as traffic detection and controlling, face recognition, object detection and medical diagnosis [8-10]. They are particularly effective at processing images due to their ability to identify features at different scales. CNNs consist of a series of layers that perform convolutions and pooling operations to extract features from the input data. These features are then fed into a fully connected layer for classification

2.2.3. Introduction of transfer learning. Transfer learning is a technique used in machine learning to leverage pre-trained models for new tasks. Rather than starting from scratch, transfer learning enables researchers to use a pre-trained model that has already learned a set of features and apply it to a new dataset.

2.2.4. Modeling process. This study used Edge Impulse to train a CNN for image classification based on the transfer learning technology, and the process is shown in Figure 4.

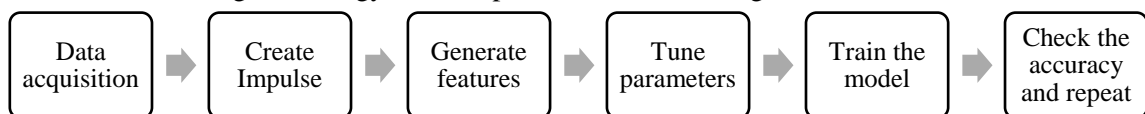


Figure 4. The procedure of the model building based on the Edge Impulse platform.

(1) Data acquisition

The first step is to upload training dataset and test dataset separately. Each data item is labeled according to its type of lung disease.

(2) Create impulse

The next step is the impulse design. A processing block and a learning block have to be added to build the network.

(3) Generate features

Next, change the colour depth into grayscale. Then the image block will normalize image data and convert each pixel into a single value according to ITU-R BT.601 conversion and will be shown in a feature explorer graph. Figure 5 illustrates the distribution of the four categories in the feature space.

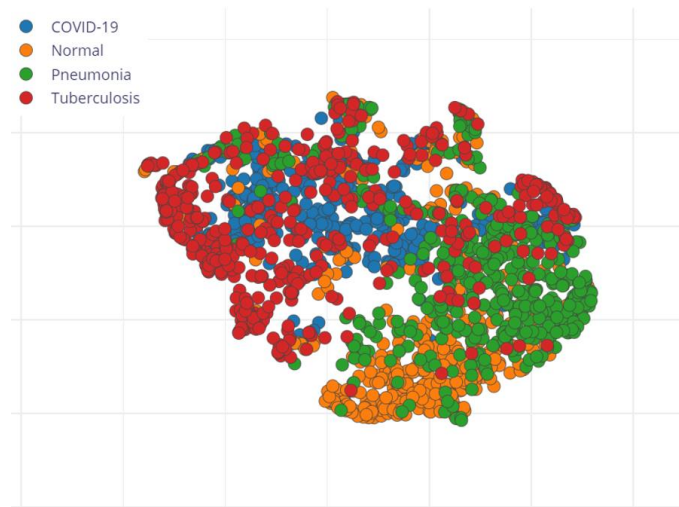


Figure 5. The procedure of the model building based on the Edge Impulse platform.

(4) Tune parameters

Input parameters or use default parameters. In this case, the parameters are set as default ones.

(5) Train the model

Run the model, and the output will be shown in a scatter plot and a confusion matrix.

(6) Check the performance and repeat the processes

Some performance measurements are shown in the result page. After checking the results such as accuracy, inferencing time, RAM storage, another model will be used to train the dataset.

2.3. Implementation details

The two main hyperparameters utilized in this study are the number of cycles and the learning rate. Changing different couple of parameters will have different accuracy.

2.3.1. Number of cycles. Number of cycles refers to the number of times that the training data is shown to the model during the training process. Increasing the number of training cycles can help the model learn more patterns in the data, but it can also increase the risk of overfitting if the model is too complex relative to the amount of data available. Decreasing the number of training cycles can make the model less prone to overfitting, but it can also result in underfitting if the model is not complex enough to capture all the patterns in the data.

2.3.2. Learning rate. Learning rate is a hyperparameter that controls the step size at which the model adjusts its weights during training. Specifically, the learning rate determines the size of the update made to the model's weights at each iteration of the training process. A high learning rate means that the weights will be updated by a large amount at each iteration, which can result in faster convergence but may also cause the model to overshoot the optimal weights and become unstable. A low learning rate means that the weights will be updated by a smaller amount at each iteration, which can result in slower convergence but may also help the model avoid overshooting and converge to a more stable solution.

In this case, the parameters are set as default ones. The numbers of cycles for two transfer learning are both 20, while the numbers for classification are 10. The learning rate is all set to 0.0005.

3. Results and discussion

3.1. The performance of the model

In this study, four models were employed to train a given dataset: namely Transfer Learning (Images) by Edge Impulse, Classification by Edge Impulse, Classification (Keras) by BrainChip and Transfer Learning (Keras) by BrainChip. The performance outcomes of these models were then recorded and presented in Table 1.

Table 1. The performance of various models.

Model	Author	Accuracy	Los s	Inferencing time
Transfer Learning (Images)	Edge Impulse	88.1%	0.4	102ms
Classification	Edge Impulse	83.9%	0.5	
Transfer Learning (Keras)	BrainChip	80.6%	1.1	
Classification (Keras)	BrainChip	80.2%	1.4	566ms

According to the Table 1, the transfer learning model designed by edge impulse achieves the superior performance, which the highest accuracy of 88.1% and the least loss of 0.41. Besides, this model also exhibits a comparatively shorter inferencing time.

In the confusion matrix shown in Table 2, the prediction of COVID-19 performs the best, while the prediction of pneumonia is relatively low. And most misclassification occurred in classifying other items into COVID-19.

Table 2. Confusion matrix of transfer learning model by edge impulse.

Actual\Predicted	COVID-19	Normal	Pneumonia	Tuberculosis
COVID-19	97.6%	0.4%	0.8%	1.2%
Normal	3.8%	86.0%	9.8%	0.4%
Pneumonia	11.7%	6.1%	82.3%	0%
Tuberculosis	13.0%	0%	0.8%	86.2%
F1 Score	0.87	0.89	0.85	0.92

3.2. Discussion

From the results, the transfer learning model by edge impulse performs the best. And the transfer learning models are better than the classification models with the same author. This is mainly because the transfer learning model is designed to tune a pre-trained image classification on existing dataset. However, the classification model is used to learn patterns from data, and can apply these patterns to new data. The greyscale images are hard to extract patterns from, if the dataset is not pre-trained. Therefore, the transfer learning is better than classification in this study. Besides, this result shows that edge impulse is a better model designer, for having a better performance in a same type of model designed by BrainChip.

As for an individual type, the performance in detecting COVID-19 is better, but many other items are misclassified to this type of disease. This result has two probabilities: one is that other diseases do not have unique patterns to be identified; the other is that COVID-19 has certain special features, which is common in other types of diseases. So that the machine may misclassify some other diseases as COVID-19. And if a case is actually a lung disease, it has a low rate to be diagnosed as normal. This means that it is not easy to miss detection of lung diseases.

Therefore, this model is good at detecting lung diseases, because it performs well in recognizing abnormal cases. However, since the model may misclassify other types of lung diseases into COVID-19, the detailed classification performance of this model needs to be improved.

4. Conclusion

This study employs edge lung disease data to identify an optimal machine learning model on the edge impulse platform for facilitating the diagnosis of lung diseases. The findings suggest that transfer learning models outperform classification models, owing to their distinct approaches to training the data. Specifically, the transfer learning model designed by edge impulse exhibits superior performance, particularly in detecting COVID-19. Nonetheless, the model is not immune to misclassifications, as it may erroneously classify other diseases as COVID-19.

The study implies that in the future, the model could be implemented in real-world devices to enhance the diagnosis of lung diseases, given its high overall accuracy in detecting abnormal cases. However, the model requires refinement to enable more precise classification.

References

- [1] Cai L et al 2020 A review of the application of deep learning in medical image classification and segmentation *Ann Transl Med* Jun 8(11) 713
- [2] Albahri O S et al 2020 Systematic review of artificial intelligence techniques in the detection and classification of COVID-19 medical images in terms of evaluation and benchmarking: Taxonomy analysis, challenges, future solutions and methodological aspects, *Journal of Infection and Public Health* 13 (10) pp 1381-1396
- [3] Zhang J et al 2019 Medical image classification using synergic deep learning, *Medical Image Analysis* 54 pp10-19
- [4] He X Zhao K Chu X 2021 AutoML: A survey of the state-of-the-art *Knowledge-Based Systems* 212: 106622
- [5] Karmaker S K Hassan M M Smith M J et al 2021 Automl to date and beyond: Challenges and opportunities *ACM Computing Surveys (CSUR)* 54(8): 1-36
- [6] Gijssbers P LeDell E Thomas J et al 2019 An open source AutoML benchmark *arXiv preprint arXiv:1907.00909*
- [7] Kaggle 2023 Lungxraysgrayscale <https://www.kaggle.com/datasets/samuel156/lungxrays-grayscale>
- [8] Ren S He K Girshick R et al 2015 Faster r-cnn: Towards real-time object detection with region proposal networks *Advances in neural information processing systems* 28
- [9] Yu Q Chang C S Yan J L et al 2019 Semantic segmentation of intracranial hemorrhages in head CT scans 2019 *IEEE 10th International Conference on Software Engineering and Service Science (ICSESS)* IEEE 112-115
- [10] Coşkun M Uçar A Yildirim Ö et al 2017 Face recognition based on convolutional neural network 2017 *International Conference on Modern Electrical and Energy Systems (MEES)* IEEE 376-379