

Auto-outlier Fusion Technique for Chest X-ray Classification with Multi-head Attention Mechanism

Yuru Jing¹, Zixuan Li²

¹University College London, Gower Street, London, WC1E 6BT, UK

²University of Sheffield, Western Bank, Sheffield, S10 2TN, UK

Keywords: Auto-outlier Fusion Technique, Chest X-ray classification, Generalized max-pooling, Self-attention, multi-head attention, Deep learning

Abstract: A chest X-ray is one of the most widely available radiological examinations for diagnosing and detecting various lung illnesses. The National Institutes of Health (NIH) provides an extensive database, ChestX-ray8 and ChestXray14, to help establish a deep learning community for analysing and predicting lung diseases. ChestX-ray14 consists of 112,120 frontal-view X-ray images of 30,805 distinct patients with text-mined fourteen disease image labels, where each image has multiple labels and has been utilised in numerous research in the past. To our current knowledge, no previous study has investigated outliers and multi-label impact for a single X-ray image during the preprocessing stage. The effect of outliers is mitigated in this paper by our proposed auto-outlier fusion technique. The image label is regenerated by concentrating on a particular factor in one image. The final cleaned dataset will be used to compare the mechanisms of multi-head self-attention and multi-head attention with generalised max-pooling.

1. Introduction

The adoption of medical imaging as a diagnostic tool has risen dramatically in the last few decades, with several modalities increasing by over 50%.^[1] In comparison to other medical modalities such as computerised tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET), radiography is effective and inexpensive which regularly utilised in medical facilities such as hospitals. Chest radiography (Chest X-ray) is a potent, noninvasive investigative tool for revealing information about lung diseases without causing discomfort to the patient.^[2] In recent years, Artificial Intelligence techniques such as machine learning and deep learning have been widely utilised in the healthcare industry.^[3] It has become a sought-after option for image analysis tasks and has significantly contributed to the medical imaging field.^[4] In light of the fact that deep learning is a data-hungry paradigm, the National Institutes of Health (NIH) extended the "ChestX-ray8" database to a larger chest X-ray database in order to aid in the advancement of chest X-ray research, which is called "ChestX-ray14", including 112,120 frontal-view X-ray images of 30,805 distinctive patients with the text-mined 14 disease image labels shown as figure 1, where each image has multiple labels available.^[5]

Numerous earlier studies on lung diseases based on Chest X-rays have been conducted, including image-level predictions, segmentation, localization, image production, and domain adaption.^[6]

Predicting the labels of the ChestX-ray 14 dataset is the image-level prediction problem that has been the subject of most research.^[6] According to Ccalli et al.,^[6] ResNet is one of the most commonly exploited architectures in the classification of lung illnesses. Furthermore, the number of layers in specific neural architecture is also taken into account.^[6] ResNet-50 is implemented in the work of Owais et al.^[7] ResNet-18, ResNet-34, and ResNet-101 have also been utilised in the research of a variety of researchers^{[8][9][10]}. In addition to the consideration of the architecture change, the attention mechanism is a frequently used technique in deep learning that focuses on vital information and less on unimportant information to resolve information overload issues with restricted computing resources.^[11] Baltruschat et al. evaluate the performance of multiple ResNet blocks with a novel non-image fusion implementation to categorise the 14 illness labels provided by the ChestX-ray14 dataset and apply gradient-weighted class activation mapping, in the end to make the prediction results more interpretable.^[12] In order to improve the task of classifying 14 lung diseases from chest X-rays, Wang et al.^[13] first propose a single attention branch with gradient-weighted class activation and then design a unique triple attention learning method that combines three different attention modules with a multi-scale ensemble module.^[14] Outliers, which are findings that diverge so significantly from other observations as to raise suspicion that a distinct mechanism generated them, are another common concern in datasets. Typically, removing outliers improves and stabilises forecasts.^[15] There are many outlier removal strategies in both the classic statistical area and the sophisticated algorithm field. Conventional statistical techniques, such as the confident interval approach^[16] and the Interquartile Range Technique,^[17] is used to find outliers based on statistical principles. In terms of algorithm approach, the primary three categories for auto-outlier identification are linear, proximity, and ensemble.^[15] Breunig et al.'s^[18] Local Outlier Factor (LOF) algorithm is a proximity method that effectively uses low-dimensional feature spaces to determine how remote a place is from its immediate surroundings. A more sophisticated and successful method is proposed by Liu et al.,^[19] called isolation forest, which selects a feature from the feature space and a random split value based on a binary decision tree. One-Class Support Vector Machine,^[20] developed by Bounisar et al., is a linear method for capturing extreme density cases as outliers (OCSVM). In order to describe the data as a high-dimensional Gaussian distribution with potential covariances between feature dimensions, the Elliptical Envelope procedure was devised.^[21] Data outside of the ellipse are recognized as outliers.

The majority of previous attention-based neural networks only implemented the attention mechanism at the very end of the architecture rather than implementing it at the beginning. Furthermore, to the best of our knowledge, none of the prior studies considered the outlier impact for the non-image features connected to the images and image labels or the multi-factor for a single image. Instead, they almost exclusively used all 112,120 X-ray images from the ChestX-ray14 dataset. Therefore, we first propose an auto-outlier fusion strategy in this study to exclude outliers from chest X-ray image data following the required preprocessing. Then, we set up Multi-class Focal loss mitigate the imbalance effect during the training process and substitute the global max-pooling layer with a deep generalized max-pooling layer, and at last improve the final performance by implementing the multi-head attention mechanism to either the end or the beginning, namely, self-attention mechanism. The primary contribution of this research is as follows:

- Explore the raw Chest X-ray images while excluding the multi-factor images to recognize outliers, develop an auto-outlier fusion technique consisting of a voting system incorporating a Local Outlier Factor (LOF), isolation forest, One-Class Support Vector Machine (OCSVM), and Elliptical Envelope algorithms.
- Set up multi-class focal loss to tackle the imbalance classes of the Chest X-ray, replace the global max-pooling with the deep generalized max-pooling, and compare the multi-head self-attention mechanism with the multi-head mechanism to further boost the performance.

2. Methods

2.1 Data Preprocessing

The Chest X-ray14 database contains both image and non-image data. Aside from directly combining nonimage data (gender and age), non-image data can also be utilized as a benchmark to eliminate outliers through statistical and algorithm methods to generate a more reliable and accurate forecast. The preprocessing procedure is separated into non-image data preprocessing and image data preprocessing.

2.1.1 Preprocessing Non-image Chest X-ray Data:

After checking for missing values, identifying outliers is a crucial step in excluding unreliable data points that may have been produced by other mechanisms and would ultimately affect our final prediction. Using a voting system, we propose an auto-outlier fusion technique that combines the traditional statistical method (interquartile range) with the advanced algorithms from proximity, linear, and ensemble principles, namely Local Outlier Factor (LOF), One-Class Support Vector Machine (OCSVM), Isolation Forest, and Elliptic Envelope, to make the final determination for outliers.

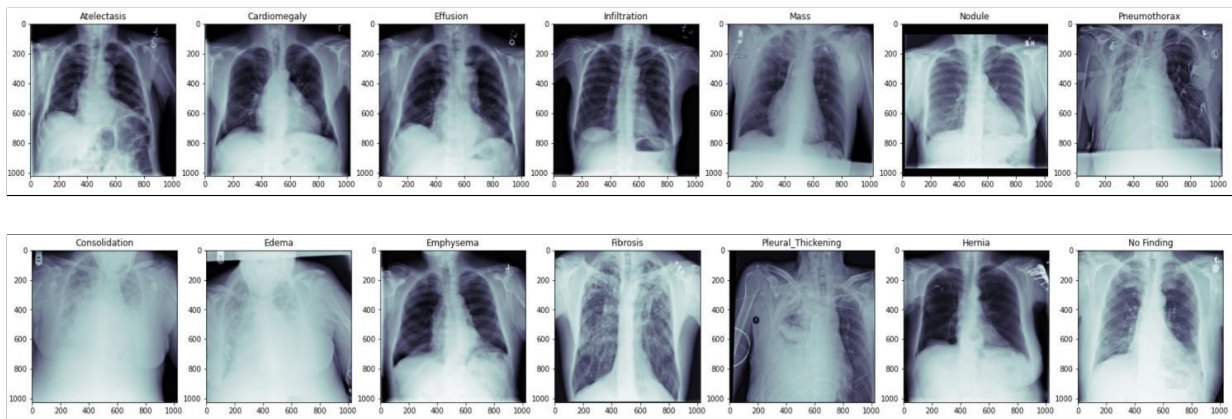


Figure 1: The example of thirteen prevalent thoracic disorders and one healthy case in chest X-rays.

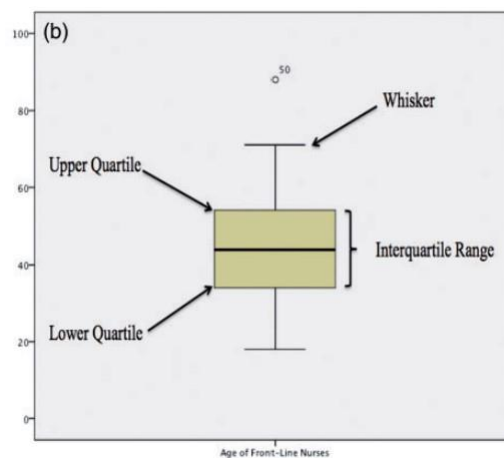


Figure 2: Example of outlier detection in a box whisker plot [23]

Interquartile Range Technique (IQR): The IQR method, which employs the difference between the third (Q3) and first quartiles (Q1) to gauge the spread of the middle 50% of data, successfully

identifies and deals with outliers in statistics. The entire procedure can be depicted as the figure 2 in a box whisker plot. Any point that lies outside the whisker, the bounds of $Q1 - 1.5IQR$ or $Q3 + 1.5IQR$, is regarded as a potential outlier.^[22]

Auto-Outliers Detection Techniques: In this work, three distinctive categorical methods in auto-outliers detection are considered: proximity, linear, and ensemble.^[15] The local Outlier Factor (LOF) algorithm assigns each data point a score depending on the size of its local neighbourhood that describes how isolated it is.^[18] This approach is practical for low-dimensional feature space, and it is more likely that those with a higher score will be outliers.^[15] One-Class Support Vector Machines (OCSVM) can find outliers by locating the extreme points of the density function,^[20] and this technique is applicable to both classification and regression datasets.¹⁵ Isolation forest is based on a binary decision tree constructed using the selected features and arbitrary splits.¹⁹ It is appropriate for values without scaling and when value distributions cannot be assumed. Those outliers are identified with the fewest splits compared to the remaining data.^[15] Elliptical Envelope identifies outliers outside the ellipse based on potential covariances between feature dimensions.^[21]

Proposed Auto-Outlier Fusion Technique: The auto-outlier fusion technique is a simple, unique design that integrates the findings of the five outlier detection strategies mentioned above: Interquartile Range (IQR), Local Outlier Factor (LOF), One-Class Support Vector Machines (OCSVM), Isolation Forest, and Elliptical Envelope.

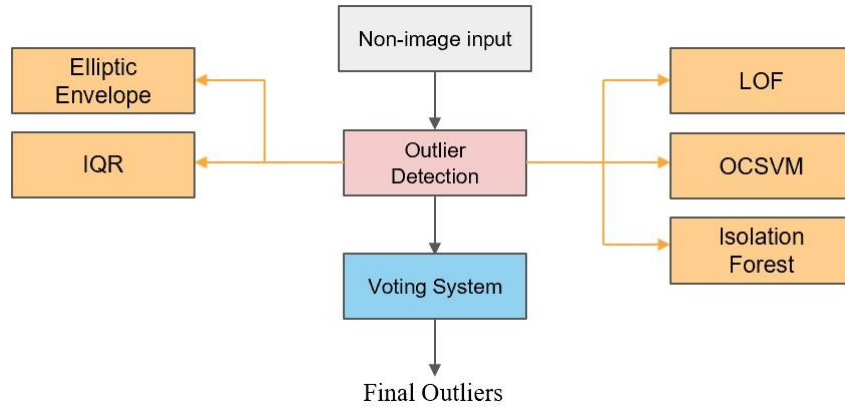


Figure 3: Proposed auto-outlier fusion technique

The internal voting system determines the final outliers, identified as outliers when three or more of the above three procedures perceive it as an outlier, which will be employed in our subsequent experiment. The figure 3 depicts our developed auto-outlier fusion process.

2.1.2 Preprocessing Image Chest X-ray Data:

Data preprocessing: In terms of image data preparation, we use a variety of ways to convert and supplement the image Chest X-ray data that we have. Because the original image size from the Chest X-ray is too large to process, we first resize 1024×1024 into 224×224 , then normalise all images by calculating their mean and standard deviation, randomly rotate images 5 degrees, and randomly sample image width and height after scaling from 0.9 to 1.1.

2.2 Practical Approaches and Evaluation Metrics

Deep Generalized Max-Pooling (DGMP): DGMP is a new technique that modifies the standard Max Pooling layer by re-weighting all descriptors so that the effect of common and rare ones is equalised rather than computed geographically independently.^[24] According to Vincent et al.,^[24] the final activation volume of a convolutional neural network has the following dimensions: height h ,

width w , and the number of activation maps, depth d . Activation maps are situated at i , where i spans the entire depth d , and Φ_i is the local embedding. A weighted average of each point in the activation volume produces the global descriptor ζ , which optimises and takes into account all activation maps rather than just one activation map when performing the maximum global pooling.^[24] The deep generalized max-pooling process are shown in the figure 4 and its dual function can be extracted from Murray et al.'s^[25] work as follows:

$$\alpha_{gmp,\lambda} = \underset{\alpha}{\operatorname{argmin}} \|\phi^T \phi \alpha - 1_n\|^2 + \lambda \|\phi \alpha\|^2 \quad (1)$$

$$= \underset{\alpha}{\operatorname{argmin}} \|K\alpha - 1_n\| + \lambda \alpha^T K \alpha \quad (2)$$

Where K is Gram matrix, yielding the closed-form solution:

$$\alpha = (K + \lambda I_n)^{-1} 1_n \quad (3)$$

Where λ is the hyperparameter we can tune through cross validation process.

Focal Loss (FL): Lin et al.^[26] devised Focal Loss (FL), which reweights distinct classes using the weighted crossentropy (CE) loss to overcome class imbalance issues, as seen below.

$$CE(p, y) = CE(p_t) = -\log(p_t) \quad (4)$$

$$p_t = \begin{cases} p, & y = 1 \\ 1 - p, & \text{otherwise} \end{cases} \quad (5)$$

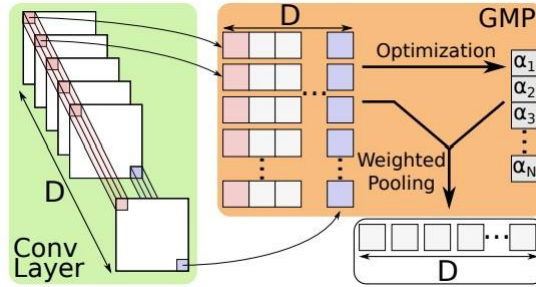


Figure 4: Overview of Deep Generalized Max Pooling. A D -unknown linear optimization issue is solved by expressing each local activation vector along the depth axis of the activation. The output consists of a weighted sum of local activation vectors.^[24]

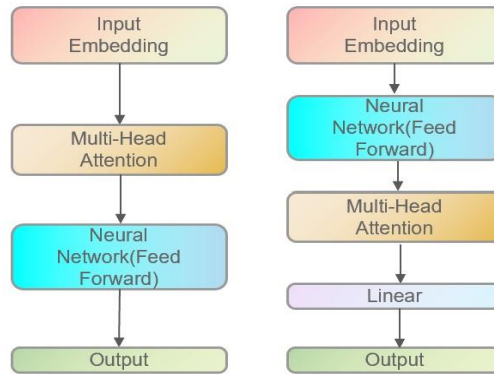


Figure 5: The left figure is self multi-head attention and the right figure is traditional multi-head attention.

$$FL(p_i) = -(1 - p_i)^\gamma \log(p_i), \gamma \geq 0 \quad (6)$$

Where p is the probability for the class label y , $(1 - p)^y$ is a modulating factor of the binary CE loss with tunable focusing parameter.

Multi-head Attention Mechanism: Information from several representation subspaces at various places can be combined with more than one head of attention via a multi-head attention method. The third type of attention map is often concatenated first with various attention map from various attention mechanisms.^[27] The classic multi-head attention mechanism and the more focused self-attention mechanism vary in that the former sets up the multi-head attention mechanism at the very end while the latter does so before the neural network. We have built our multi-head attention utilizing three conventional attention processes. The information on the inter-channel relationships of features can be extracted using the channel attention mechanism, the information on the inter-spatial relationships of features can be derived using the spatial attention mechanism,^[28] and the information of large vital regions, which is the advanced channel attention, can be extracted using the coordinate attention mechanism at a low computational cost.^[29] The graph 5 depicts the multi-head attention mechanism for setting up in the various settings.

Evaluation Metrics: Although accuracy is frequently employed as a machine learning measure, it has little uniqueness, discriminative power, and informativeness and is biased toward data from the majority class.^[30] F1 score has been proven superior to accuracy in classifier optimization for imbalance classification. In the medical industry, two metrics called specificity and sensitivity are employed to assess the proportions of both positive and negative patterns that are accurately identified. The most crucial classification metric in our work will be the F1 score.^[30]

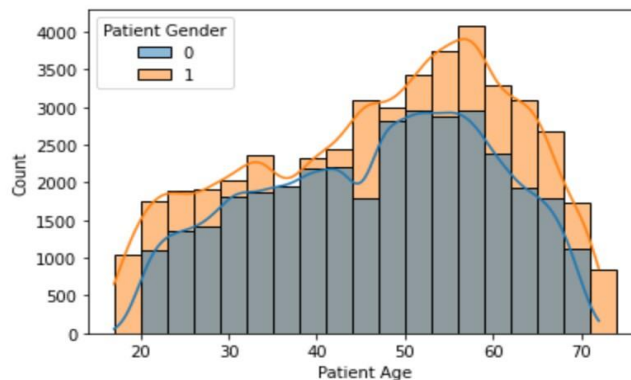


Figure 6: Patient age distribution grouped by the gender, which indicates male and female as 1 and 0, respectively.

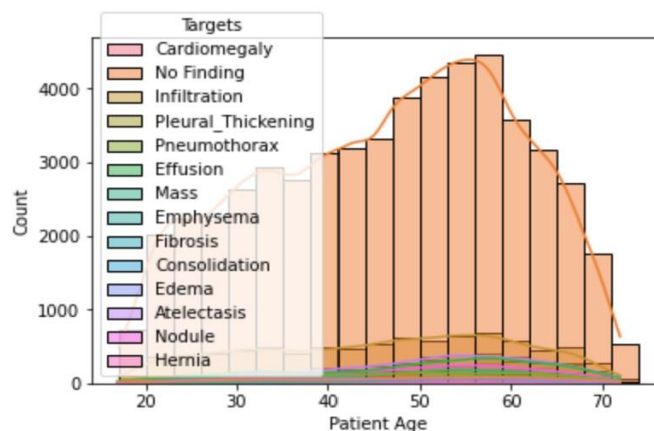


Figure 7: Patient age distribution grouped by the lung diseases.

3. Experiments and Results

3.1 Explore the Data

The dataset of chest X-ray images used in this study was obtained from the National Institutes of Health (NIH) which is also one of the Kaggle competitions.^[5]The original dataset includes 14 distinct lung diseases and one health condition known as "No Findings." Ccalli et al.^[6] developed the notion of focusing on significant relationships between distinct labels. Because consolidation is more prominent and has considerable overlap with pneumonia, and is commonly used to identify patients with pneumonia, we eliminate the "Pneumonia" label and keep the "Consolidation" label for the following experiments. After deleting the redundant features from the non-image data (Follow-up number, patient id, view position, original image space), we further investigated the data by grouping patient ages by gender and various types of lung diseases. According to the figure 6 and figure 7, the patient age distributions are comparable regardless of the gender or kind of the diseases. There is essentially no association between patient age and gender, according to the coefficient of the correlation between the gender and patient age from various lung diseases, which ranges from -0.13 to 0.13.

3.2 Experiment setup

In this research, we generate the final outliers using our proposed auto-outlier fusion technique to non-image data (gender and age). The version without outliers will eliminate multi-factor pictures in order to maintain a single factor image as the final clean dataset. Then, we use this clean data set to resize, normalise, rotate, and scale pictures for additional preprocessing and enhancement. Ultimately, we use the ImageNet-pretrained ResNet101 architecture with deep generalised max-pooling to incorporate traditional multi-head and self-multi-head attention in our neural network model and conduct a comparative study.

3.3 Performance analysis and results

3.3.1 Remove outliers through auto-outlier fusion technique:

Table 1: The number of different categories for diseases in the clean dataset.

Disease Label	Counts
Atelectasis	3795
Cardiomegaly	951
Effusion	3503
Infiltration	8472
Mass	1958
Nodule	2497
Pneumothorax	1941
Consolidation	1113
Edema	589
Emphysema	745
Fibrosis	637
Pleural Thickening	999
Hernia	80
No Finding	53875

After removing the cases with multiple factors (more than one disease in the image), the dataset is

reduced to 91,002 images from the initial dataset of 112,210. The application of our auto-outlier fusion algorithm will further eliminate over 9,000 outliers, yielding 81,155 samples in our final clean dataset. The table 1 shows the details of separate illness groups in the clean dataset.

3.3.2 Comparison Analysis

In our comparison experiment, we fixed the number of epochs, the learning rate of the Adam optimiser, and the batch size of our training procedure separately by 30, 10^{-4} , 64. The full dataset is divided into the training set (80%) and the testing set (20%). One benchmark is implementing ResNet-101 with deep generalised max-pooling (DGMP) in both the clean and outlier datasets. ResNet-101 with global max-pooling (MP) is another benchmark. The ResNet-101 architecture with different multi-head attention setups is compared to the aforementioned two benchmarks, and the results are displayed in the table 2 below.

In the above studies, GMP improves performance by about 0.03 based on the F1 score over the original configuration with MP. The clean dataset version outperforms the dataset with the outlier’s version, with values of 0.6804 and 0.6602, respectively. The implementation of attention methods improves performance even more than the version without attention mechanisms. Self-attention is exceeded by multi-head attention. As a result, the best model is ResNet-101 architecture with GMP and multi-head attention mechanism, which has an F1 score of 0.7217 after five-fold cross-validation.

Table 2: Comparison of benchmarks with different multi-head attention mechanisms. DGMP and MP represent deep generalized max-pooling and the global max-pooling, Self-attention is the multi-head attention mechanism applied before the neural network. CV is the cross validation.

Model	SPE	F1	Loss	Accuracy
ResNet-101 + MP (Clean Benchmark)	0.9711	0.6545	0.7174	65.52%
ResNet-101 + GMP (Outlier Benchmark)	0.9721	0.6602	0.7223	66.32%
ResNet-101 + GMP (Clean)	0.9733	0.6804	0.7925	68.03%
ResNet-101 + GMP + self-attention (Clean)	0.9785	0.6847	0.8025	68.34%
ResNet-101 + GMP + multi-head (Clean)	0.9822	0.6923	0.8839	69.23%
ResNet-101 + GMP + multi-head (Clean) + 5 fold CV	0.9915	0.7217	0.9712	72.62%

4. Discussion

Previous studies have several drawbacks. To our knowledge, no existing studies have addressed overlapping labels such as "Consolidation" and "Pneumonia," the effect of multi-factors in one image, outliers impact in non-image data with the connected label, or the participation of self-attention. Our study compensates for all of the above flaws by additional data cleaning and our proposed novel design, auto-outliers fusion technique. Nevertheless, there are still some limitations that can be overcome in future research. To begin with, our work specified the number of epochs and learning rate; however, more epochs and learning rates can be tested in future trials. Following that, in our suggested auto-outlier fusion approach and multi-head attention mechanisms, more distinct types of single-outlier detection techniques and attention mechanisms can be included to strengthen the designs. Last but not least, in addition to the attention mechanism, fuzz logits can be used to improve model learning as the attention mechanism.

5. Conclusion

In this paper, we construct auto-outlier strategies to compensate for the shortcomings of earlier

research and multi-class focal loss to mitigate the impact of imbalanced classification. Utilise then the multi-head attention mechanism and deep generalised max-pooling to enhance the neural network further. The deep generalised max-pooling outperforms the conventional global max-pooling, regardless of whether the dataset contains outliers or is clean. Our proposed auto-outlier fusion method improves the final experimental prediction outcomes. Based on our tests' comparative analysis, the multi-head attention mechanism performs better than the self-multi-head attention.

Acknowledgments

We thank you for the data resources from The National Institutes of Health (NIH), ChestX-ray 14⁵ to help us investigate our research, and thank you for some attention mechanism referenced source codes from Github.

References

- [1] Smith-Bindman, R., Miglioretti, D. L., Johnson, E., Lee, C., Feigelson, H. S., Flynn, M., Greenlee, R. T., Kruger, R. L., Hornbrook, M. C., Roblin, D., et al.. "Use of diagnostic imaging studies and associated radiation exposure for patients enrolled in large integrated health care systems, 1996-2010," *Jama* 307(22), 2400–2409 (2012).
- [2] Gazda, M., Plavka, J., Gazda, J., and Drotar, P.. "Self-supervised deep convolutional neural network for chest x-ray classification," *IEEE Access* 9, 151972–151982 (2021).
- [3] Davenport, T. and Kalakota, R.. "The potential for artificial intelligence in healthcare," *Future healthcare journal* 6(2), 94 (2019).
- [4] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciampi, F., Ghahfarooian, M., Van Der Laak, J. A., Van Ginneken, B., and Sánchez, C. I.. "A survey on deep learning in medical image analysis," *Medical image analysis* 42, 60–88 (2017).
- [5] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., and Summers, R.. "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in [2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)], 3462–3471 (2017).
- [6] Calli, E., Sogancıoğlu, E., van Ginneken, B., van Leeuwen, K. G., and Murphy, K.. "Deep learning for chest x-ray analysis: A survey," *Medical Image Analysis* 72, 102125 (2021).
- [7] Owais, M., Arsalan, M., Mahmood, T., Kim, Y. H., Park, K. R., et al.. "Comprehensive computer-aided decision support framework to diagnose tuberculosis from chest x-ray images: data mining study," *JMIR medical informatics* 8(12), e21790 (2020).
- [8] Oh, Y., Park, S., and Ye, J. C. "Deep learning covid-19 features on cxr using limited training data sets," *IEEE transactions on medical imaging* 39(8), 2688–2700 (2020).
- [9] Yoo, H., Kim, K. H., Singh, R., Digumarthy, S. R., and Kalra, M. K. "Validation of a deep learning algorithm for the detection of malignant pulmonary nodules in chest radiographs," *JAMA network open* 3(9), e2017135–e2017135 (2020).
- [10] Kusakunniran, W., Karnjanapreechakorn, S., Siriapisith, T., Borwarnginn, P., Sutassananon, K., Tongdee, T., and Saiviroonporn, P.. "Covid-19 detection and heatmap generation in chest x-ray images," *Journal of Medical Imaging* 8(S1), 014001 (2021).
- [11] Niu, Z., Zhong, G., and Yu, H.. "A review on the attention mechanism of deep learning," *Neurocomputing* 452, 48–62 (2021).
- [12] Baltruschat, I. M., Nickisch, H., Grass, M., Knopp, T., and Saalbach, A.. "Comparison of deep learning approaches for multi-label chest x-ray classification," *Scientific reports* 9(1), 1–10 (2019).
- [13] Wang, H., Jia, H., Lu, L., and Xia, Y.. "Thorax-net: an attention regularized deep neural network for classification of thoracic diseases on chest radiography," *IEEE journal of biomedical and health informatics* 24(2), 475–485 (2019).
- [14] Wang, H., Wang, S., Qin, Z., Zhang, Y., Li, R., and Xia, Y.. "Triple attention learning for classification of 14 thoracic diseases using chest radiography," *Medical Image Analysis* 67, 101846 (2021).
- [15] Agarwal, A. and Gupta, N.. "Comparison of outlier detection techniques for structured data," *arXiv preprint arXiv:2106.08779* (2021).
- [16] Bakar, Z. A., Mohamad, R., Ahmad, A., and Deris, M. M.. "A comparative study for outlier detection techniques in data mining," in [2006 IEEE conference on cybernetics and intelligent systems], 1–6, IEEE (2006).
- [17] Vinutha, H., Poornima, B., and Sagar, B.. "Detection of outliers using interquartile range technique from intrusion dataset," in [Information and decision sciences], 511–518, Springer (2018).
- [18] Breunig, M. M., Kriegel, H.-P., Ng, R. T., and Sander, J.. "Lof: identifying density-based local outliers," in [Proceedings of the 2000 ACM SIGMOD international conference on Management of data], 93–104 (2000).
- [19] Liu, F. T., Ting, K. M., and Zhou, Z.-H.. "Isolation forest," in [2008 eighth IEEE international conference on data

mining], 413–422, IEEE (2008).

- [20] Bounsiar, A. and Madden, M. G.. “One-class support vector machines revisited,” in [2014 International Conference on Information Science & Applications (ICISA)], 1–4, IEEE (2014).
- [21] Hoyle, B., Rau, M. M., Paech, K., Bonnett, C., Seitz, S., and Weller, J.. “Anomaly detection for machine learning redshifts applied to sdss galaxies,” *Monthly Notices of the Royal Astronomical Society* 452(4), 4183–4194 (2015).
- [22] Chawsheen, A. H., Subhi Latif, I., et al.. “Detection and treatment of outliers in data sets,” *IRAQI JOURNAL OF STATISTICAL SCIENCES* 6(9), 58–74 (2006).
- [23] Mowbray, F. I., Fox-Wasylyshyn, S. M., and El-Masri, M. M.. “Univariate outliers: a conceptual overview for the nurse researcher,” *Canadian Journal of Nursing Research* 51(1), 31–37 (2019).
- [24] Christlein, V., Spranger, L., Seuret, M., Nicolaou, A., Král, P., and Maier, A.. “Deep generalized max pooling,” in [2019 International conference on document analysis and recognition (ICDAR)], 1090–1096, IEEE (2019).
- [25] Murray, N., Jégou, H., Perronnin, F., and Zisserman, A. “Interferences in match kernels,” *IEEE transactions on pattern analysis and machine intelligence* 39(9), 1797–1810 (2016).
- [26] Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P.. “Focal loss for dense object detection,” in [Proceedings of the IEEE international conference on computer vision], 2980–2988 (2017).
- [27] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I.. “Attention is all you need,” *Advances in neural information processing systems* 30 (2017).
- [28] Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S.. “Cbam: Convolutional block attention module,” in [Proceedings of the European conference on computer vision (ECCV)], 3–19 (2018).
- [29] Hou, Q., Zhou, D., and Feng, J.. “Coordinate attention for efficient mobile network design,” in [Proceedings of the IEEE/CVF conference on computer vision and pattern recognition], 13713–13722 (2021).
- [30] Hossin, M. and Sulaiman, M. N.. “A review on evaluation metrics for data classification evaluations,” *International journal of data mining & knowledge management process* 5(2), 1 (2015).