Unsupervised Chest X-ray Opacity Classification using Minimal Deep Features

Mohd Zulfaezal Che Azemin[1], Mohd Izzuddin Mohd Tamrin[2], Mohd Adli Md Ali[3], Iqbal Jamaludin[4]
Kulliyyah of Allied Health Sciences, International Islamic University Malaysia, Kuantan, Pahang, Malaysia[1, 4]
Kulliyyah of ICT, International Islamic University Malaysia, Gombak, Kuala Lumpur, Malaysia[2]
Kulliyyah of Science, International Islamic University Malaysia, Kuantan, Pahang, Malaysia[3]

***Abstract*—Data privacy has been a concern in medical imaging research. One important step to minimize the sharing of patient's information is by limiting the use of original images in the workflow. This research aimed to use minimal deep learning features in detecting anomaly in chest X-ray (CXR) images. A total of 3,504 CXRs were processed using a pre-trained deep learning convolutional neural network to output ten discriminatory features which were then used in the k-mean algorithm to find underlying similarities between the features for further clustering. Two clusters were set to distinguish between "Opacity" and "Normal" CXRs with the accuracy, sensitivity, specificity, and positive predictive value of 80.9%, 86.6%, 71.5% and 83.1%, respectively. With only ten features required to build the unsupervised model, this would pave the way for future federated learning research where actual CXRs can remain distributed over multiple centers without sacrificing the anonymity of the patients.**

***Keywords*—Unsupervised classification; minimal deep features; convolution neural network; chest x-ray; airspace opacity**

I. Introduction

A key ingredient of anonymity in the development of machine learning models does not have to use original data in the learning process. Chest X-ray (CXR) is one of the most common modalities used in deep learning research where the images need to be labelled for further classification steps [1]. While the identity of the subjects is normally stripped from the data, it was shown that the CXRs may still contain details that could be used to identify the patients [2].

Supervised learning in radiology has grown rapidly in recent years. The applications, among others, include identifying abnormalities only from radiology reports [3], pneumonia detection using deep neural network ensemble [4], transfer learning for COVID19 classification [5], and multimodal strategy to increase the model performance [6]. Models constructed using supervised methodology requires labels annotated by experts as ground truth during training and testing phases. The problem is aggravated when the concordance rate among the radiologists is not optimal [7].

To address the problems with lack of ground truth, unsupervised approach has been proposed in recent research. The works range from the use of feature engineering based on pixel values and texture features [8] to the use of generative adversarial network (GAN) to detect anomaly in the CXRs [9]. GAN-based methods have shown promising results in identifying abnormal images but disease-free CXRs are still required in the training stage where the confirmation of experts is still needed.

This research work aims to address the problems of hyperparameters fine-tuning and lack of labels inherent in a supervised learning architecture by using minimal deep features from a pre-trained deep learning convolutional neural network applied on the k-means algorithm, which is used to find underlying similarities in the features for classification of CXRs.

II. Materials and Methods

The images were downloaded from ChestX-ray14 [1], a representative collection of CXRs for thoracic disorders for a general population, curated by the National Institutes of Health (NIH) Clinical Centre, USA, which primarily consist of all frontal CXRs in the centre. A total of 2,166 "airspace opacity" labels were obtained from previously published Google Health study [7]. This abnormality constitutes the most common finding in the dataset. "No findings" were assigned to the 1,388 CXRs when the original NIH and Google Health labels did not reveal any abnormality. Fig. 1 summarizes the classification used in this study.

As shown in Fig. 2, a pre-trained GoogLeNet deep learning convolutional neural network (CNN) was adopted to extract features from the CXRs [10]. The CNN was pretrained on more than a million images publicly available data set (http://www.image-net.org) which can classify image classes of everyday objects including pencil, coffee mug, keyboard, and animals. Each CXR resulted in 1,024 deep features which uniquely characterized the image. The features were extracted using MATLAB R2021a (MathWorks Inc., MA).

The deep features were fed into RapidMiner Studio Educational Version 9.10 (RapidMiner Inc., MA) as illustrated in Fig. 3. Forward selection was employed to increase the relevance and minimize features redundancy, which reduced the features to ten. k-Means algorithm was applied on the feature pool to determine a set of 2 clusters (i.e., "Opacity" vs "No Findings") and assigned each CXR to a cluster. Opacity is the term used in radiology to describe the presence of whiteness region in the lungs. It was chosen because this type of anomaly is considered as one the most prevalent in CXRs [11]. The clusters comprised CXRs with similar features with the similarity determined based on a distance measure between them. "Map Clustering on Labels" module estimates a

Supported by Ministry of Higher Education of Malaysia under the Fundamental Research Grant Scheme (ID: FRGS19-181-0790 / FRGS/1/2019/ICT02/UIAM/02/4.)

mapping between the given clustering and prediction by adjusting the given clusters with the given labels to estimate the best fitting pairs. Table I summarizes the algorithms used for the distance measure. The algorithms were chosen based on their availability in RapidMiner.

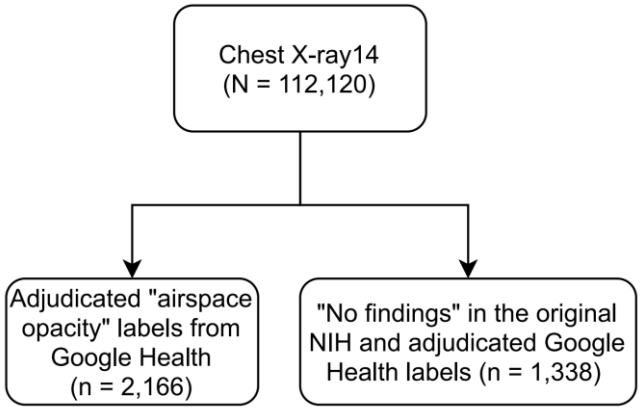

Fig. 1. CXR Images used in this Study, "Opacity" = 2,166 vs "No findings" = 1,388.

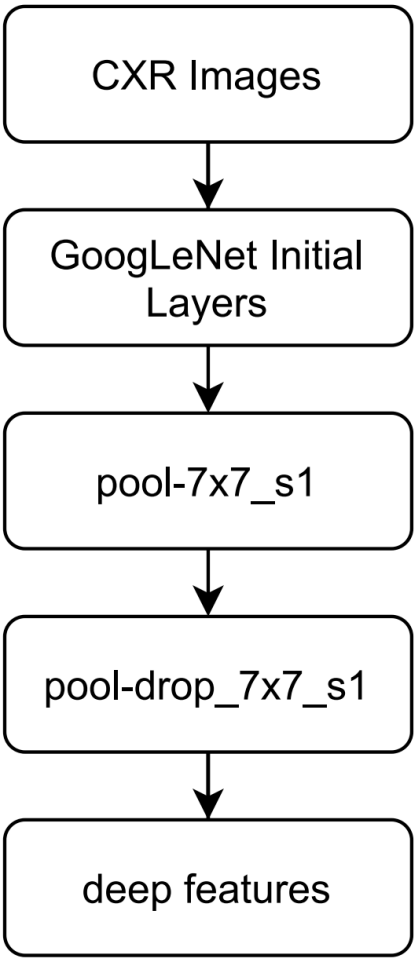

Fig. 2. Deep Features Extracted from GoogLeNet Convolution Neural Network Implemented in MATLAB.

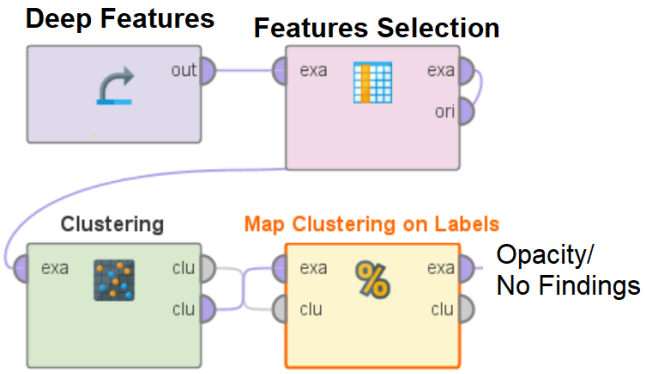

Fig. 3. Unsupervised Clustering using Deep Features on RapidMiner.

TABLE I. BRIEF DESCRIPTION OF THE DISTANCE MEASURES USED WITH THE K-MEANS ALGORITHMS

Distance Measure	**Brief Description**
Euclidean Distance	The Euclidean distance, d between the CXR image and centroid, x and y are represented by this formula: $$d(x,y) = \sqrt{\sum_{k=1}^{10}(x_k - y_k)^2}$$ Where 10 is the total number of features, whereby x_k and y_k, are the k^{th} feature of the CXR image and the centroid respectively.
Camberra Distance	The Camberra distance, d between the CXR image and centroid, x and y are represented by this formula: $$d(x,y) = \sum_{k=1}^{10}\frac{\lvert x_k - y_k\rvert}{\lvert x_k + y_k\rvert}$$ Where 10 is the total number of features, whereby x_k and y_k, are the k^{th} feature of the CXR image and the centroid respectively.
ChebychevDistance	The Chebychev distance, d between the CXR image and centroid, x and y are represented by this formula: $$d(x,y) = max_k^{10}\lvert x_k - y_k\rvert$$ Where 10 is the total number of features, whereby x_k and y_k, are the k^{th} feature of the CXR image and the centroid respectively.
Correlation Similarity	The correlation similarity between CRX image and centroid, x and y are represented by this formula: $$\text{corr}(x,y) = \frac{\sum_{k=1}^{10}(x_k - \bar{x})\,(y_k - \bar{y})}{\sqrt{\sum_{k=1}^{10}(x_k - \bar{x})^2}\sqrt{\sum_{k=1}^{10}(y_k - y)^2}}$$ Where 10 is the total number of features, whereby x_k and y_k, are the k^{th} feature of the CXR image and the centroid respectively.
Cosine Similarity	The cosine similarity between CRX image and centroid, x and y are represented by this formula: $$\cos(x,y) = \frac{\sum_{k=1}^{10} x_k y_k}{\sum_{k=1}^{10} x_k^2 \sum_{k=1}^{10} y_k^2}$$ Where 10 is the total number of features, whereby x_k and y_k, are the k^{th} feature of the CXR image and the centroid respectively.
Dice Similarity	The dice similarity between CRX image and centroid, x and y are represented by this formula: $$\text{sim}(x,y) = \frac{2\sum_{k=1}^{10} x_k\, y_k}{\sum_{k=1}^{10} x_k^2 + \sum_{k=1}^{10} y_k^2}$$ Where 10 is the total number of features, whereby x_k and y_k, are the k^{th} feature of the CXR image and the centroid respectively.

Inner Product Similarity	The inner product similarity between CRX image and centroid, x and y are represented by this formula: $$sim(x,y) = \sum_{k=1}^{10} x_k y_k$$ Where 10 is the total number of features, whereby x_k and y_k, are the k^{th} feature of the CXR image and the centroid respectively.
Jaccard Similarity	The Jaccard similarity between CRX image and centroid, x and y are represented by this formula: $$\text{sim}(x,y) = \frac{\sum_{k=1}^{10} x_k\, y_k}{\sum_{k=1}^{10} x_k^2 + \sum_{k=1}^{10} y_k^2 - \sum_{k=1}^{10} x_k^2 y_k^2}$$ Where 10 is the total number of features, whereby x_k and y_k, are the k^{th} feature of the CXR image and the centroid respectively.
Kernel Euclidean Distance	The kernel Euclidean distance, k between the CXR image and centroid, x and y are represented by this formula: $$k(x,y) = e^{-\gamma D(x,y)^2}$$ Where gamma is a parameter that can be configured and $D(x,y)^2$ is the squared Euclidean distance between CXR image and the centroid.
Manhattan Distance	The Manhattan distance, d between the CXR image and centroid, x and y are represented by this formula: $$d(x,y) = \sum_{k=1}^{10} \lvert x_k - y_k \rvert$$ Where 10 is the total number of features, whereby x_k and y_k, are the k^{th} feature of the CXR image and the centroid respectively.
Max Product Similarity	The maximum product similarity between CRX image and centroid, x and y are represented by this formula: $$\text{sim}(x,y) = max_k^{10} x_k y_k$$ Where 10 is the total number of features, whereby x_k and y_k, are the k^{th} feature of the CXR image and the centroid respectively.
Overlap Similarity	The overlap similarity between CRX image and centroid, x and y are represented by this formula: $$\text{sim}(x,y) = \frac{\sum_{k=1}^{10} x_k\, y_k}{\min(\sum_{k=1}^{10} x_k^2, \sum_{k=1}^{10} y_k^2)}$$ Where 10 is the total number of features, whereby x_k and y_k, are the k^{th} feature of the CXR image and the centroid respectively.
KL Divergence	The Kullback-Leibler divergence, D_{KL} between the CXR image and centroid, x and y are represented by this formula: $$D_{KL}(x,y) = \sum_{k=1}^{10} x_k log \frac{x_k}{y_k}$$ Where 10 is the total number of features, whereby x_k and y_k, are the k^{th} feature of the CXR image and the centroid respectively.
Mahalanobis Distance	The Mahalanobis distance between the CXR image and centroid, x and y are represented by this formula: $$malalanobis(x,y) = (x-y)^T C^{-1}(x-y)$$ Where x is the vector of CXR image, y is the vector of the centroid, and C^{-1} is the inverse covariance matrix of the CXR image and the centroid.
Squared Euclidean Distance	The squared Euclidean distance, d^2 between the CXR image and centroid, x and y are represented by this formula: $$d(x,y)^2 = \sum_{k=1}^{10} (x_k - y_k)^2$$ Where 10 is the total number of features, whereby x_k and y_k, are the k^{th} feature of the CXR image and the centroid respectively.

III. Result and Discussion

The performance of classification model was evaluated by its accuracy, sensitivity, specificity, and precision. The analysis of performance was expressed in true positive (TP), true negative (TN), false positive (FP), and false negative (FN). The measurements of each performance parameter were calculated as follows:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Sensitivity = TP / (TP + FN)

Specificity = TN / (TN + FP)

Positive Predictive Value (PPV) = TP / (TP + FP)

Table II provides the summary of the performance of the k-Means algorithm implemented using various distance measures. Generally, Mahalanobis Distance outperforms all other distance metrics with an accuracy, sensitivity, specificity, PPV of 80.9%, 86.6%, 71.5%, and 83.1%, respectively. This can be due to its ability to handle outliers and consideration of correlation between the features [12].

The unsupervised model with the highest accuracy shows higher sensitivity compared to specificity, which implies fewer opacity cases would be missed with the trade-off of higher false-positive results. Having a higher sensitivity is arguably more important than a higher specificity in the context of a disease screening. This is because the missed disease cases are more severe than falsely diagnosed with the disease, which can be further ruled out with a more accurate modality.

TABLE II. Performance of the Unsupervised CXR Opacity Classification using Deep Features with Various Distance Measures

Distance Measure	Accuracy	Sensitivity	Specificity	PPV
Euclidean Distance	75.3	81.6	65.2	79.1
Camberra Distance	71	63.7	82.9	85.8
Chebychev Distance	73.9	80.1	63.8	78.2
Correlation Similarity	61.8	45.3	88.4	86.4
Cosine Similarity	71.7	64.9	82.7	85.8
Dice Similarity	51.3	24.4	94.8	88.3
Inner Product Similarity	61.8	100	0	61.8
Jaccard Similarity	61.8	100	0	61.8
Kernel Euclidean Distance	75.2	81.3	65.4	79.2
Manhattan Distance	76.6	79.4	72	82.1
Max Product Similarity	52.4	77.1	12.2	58.7
Overlap Similarity	68.7	57.2	87.4	88
KL Divergence	60.8	44.6	87	84.7
Mahalanobis Distance	**80.9**	**86.6**	**71.5**	**83.1**
Squared Euclidean Distance	75.2	81.6	64.9	79

It is evident that without any training data for parameters fine-tuning, unsupervised approach results in a decent performance compared to the previously published state-of-the-art supervised classification method with the sensitivity, specificity and PPV of 86.1%, 89.7% and 91.6%, respectively [7]. It is important to highlight that while the supervised approach yields better performance, it requires dataset labelled by experts which is less cost effective.

Dice Similarity shows the lowest overall performance with an accuracy of 51.3%. Even though the specificity and PPV imply a good performance with more than 80%, the specificity is significantly lower than the average results of all distance measures.

For unsupervised opacity classification, previously published work based on GAN reported an area under the receiver operating characteristic of 0.838 [9]. However, this unsupervised algorithm still requires training dataset from normal CXRs to determine the anomaly score by calculating the pixel variation between the original and reconstructed images.

Deep features enable the use of standard patterns and image features with a high degree of correlation with human perception in a different context [10]. In this paper, we have shown that using unsupervised classification, the features with similar characteristics could be grouped to classify x-ray images.

Irrelevant features were effectively discarded using a feature selection technique, with less than 1% of the total of 1,024 features were eventually used in the classification. The high compression rate demonstrates the possibility of a high throughput screening with lower hardware requirements.

Another advantage of using minimal features is the privacy preservation of the patients. It is almost impossible to reconstruct the original images using compact features. The importance of using minimum information is highlighted in the previous research [2]. The study demonstrated a potential attack might cross-reference the CXR images to obtain classified information even when all the personal identifiers were removed.

IV. Conclusion

To the best of our knowledge, this is the first attempt to report the use of unsupervised CXR opacity classification with minimal deep features. With only ten features required to build the unsupervised model without hyperparameters fine-tuning, this would pave the way for future federated learning research where actual CXRs can remain distributed over multiple centers without sacrificing the patients' anonymity.

Acknowledgment

This work was supported by the Ministry of Higher Education of Malaysia under the Fundamental Research Grant Scheme with identification number FRGS19-181-0790 / FRGS/1/2019/ICT02/UIAM/02/4.

References

[1] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," 2017. doi: 10.1109/CVPR.2017.369.

[2] K. Packhäuser, S. Gündel, N. Münster, C. Syben, V. Christlein, and A. Maier, "Is Medical Chest X-ray Data Anonymous?," CoRR, vol. abs/2103.08562, 2021, [Online]. Available: https://arxiv.org/abs/2103.08562.

[3] S. Towfighi, A. Agarwal, D. Y. F. Mak, and A. Verma, "Labelling chest x-ray reports using an open-source NLP and ML tool for text data binary classification," medRxiv, 2019, doi: 10.1101/19012518.

[4] I. Sirazitdinov, M. Kholiavchenko, T. Mustafaev, Y. Yixuan, R. Kuleev, and B. Ibragimov, "Deep neural network ensemble for pneumonia localization from a large-scale chest x-ray database," Computers & Electrical Engineering, vol. 78, pp. 388–399, 2019, doi: https://doi.org/10.1016/j.compeleceng.2019.08.004.

[5] S. G. Sundaram, S. A. Aloyuni, R. A. Alharbi, T. Alqahtani, M. Y. Sikkandar, and C. Subbiah, "Deep Transfer Learning Based Unified Framework for COVID19 Classification and Infection Detection from Chest X-Ray Images," Arabian Journal for Science and Engineering, 2021, doi: 10.1007/s13369-021-05958-0.

[6] G. Zamzmi, S. Rajaraman, and S. Antani, "UMS-Rep: Unified modality-specific representation for efficient medical image analysis," Informatics in Medicine Unlocked, vol. 24, p. 100571, 2021, doi: https://doi.org/10.1016/j.imu.2021.100571.

[7] A. Majkowska et al., "Chest radiograph interpretation with deep learning models: Assessment with radiologist-adjudicated reference standards and population-adjusted evaluation," Radiology, vol. 294, no. 2, 2020, doi: 10.1148/radiol.2019191293.

[8] P. Ghosh, S. K. Antani, L. R. Long, and G. R. Thoma, "Unsupervised segmentation of lungs from chest radiographs," in Medical Imaging 2012: Computer-Aided Diagnosis, 2012, vol. 8315, pp. 884 – 889. doi: 10.1117/12.911574.

[9] T. Nakao et al., "Unsupervised Deep Anomaly Detection in Chest Radiographs," Journal of Digital Imaging, vol. 34, no. 2, pp. 418–427, 2021, doi: 10.1007/s10278-020-00413-2.

[10] C. Szegedy et al., "Going deeper with convolutions," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2015, vol. 07-12-June-2015. doi: 10.1109/CVPR.2015.7298594.

[11] J. T. Wu et al., "AI Accelerated Human-in-the-loop Structuring of Radiology Reports," AMIA Symposium, vol. 2020, pp. 1305–1314, Jan. 2021, [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/33936507

[12] S. Ana-Maria Ramona, C. Marian Pompiliu, and M. Stoyanova, "Data Mining Algorithms for Knowledge Extraction," 2020. doi: 10.1007/978-3-030-43449-6_20.