# Classification of Chest X-ray images using radiomic features and machine learning

**CH Yugandhar[1], Manjunatha Hiremath[2]**
[1]*Department of Computer Science, Christ University, Bengaluru,*
[2]*Department of Computer Science, Christ University, Bengaluru*
**Email:** *yugandhar.ch@res.christuniversity.in*

**KEYWORDS**
Radiograph, X-ray, Pneumonia, Data Analysis, Machine Learning, Gradient Boosted Trees, XGBoost

**ABSTRACT:**
Accurate identification of the existence of disease in a radiograph by a radiologist is highly essential. Various systems are being developed to assist radiologists to diagnose the disease with the best accuracy possible. Machine Learning algorithms have been used in classification tasks, particularly Convolutional Neural Networks, a variant of Neural Networks has been proven to outperform most algorithms. The success is attributed to this variant because they resort to optimal feature construction on their own instead of depending on a finite set of feature candidates. But CNNs in their basic form or various pre-built models like VGG Net, MobileNet, etc. are able to classify the images based on certain conditions i.e. have sufficient training data and expensive computing power. The classification of radiographic images for medical diagnosis can also be achieved usinga predefined set of features which are calculated via the extraction of quantitative metrics. Such a process is known as radiomics or radiogenomics (kumar et al, 2012).

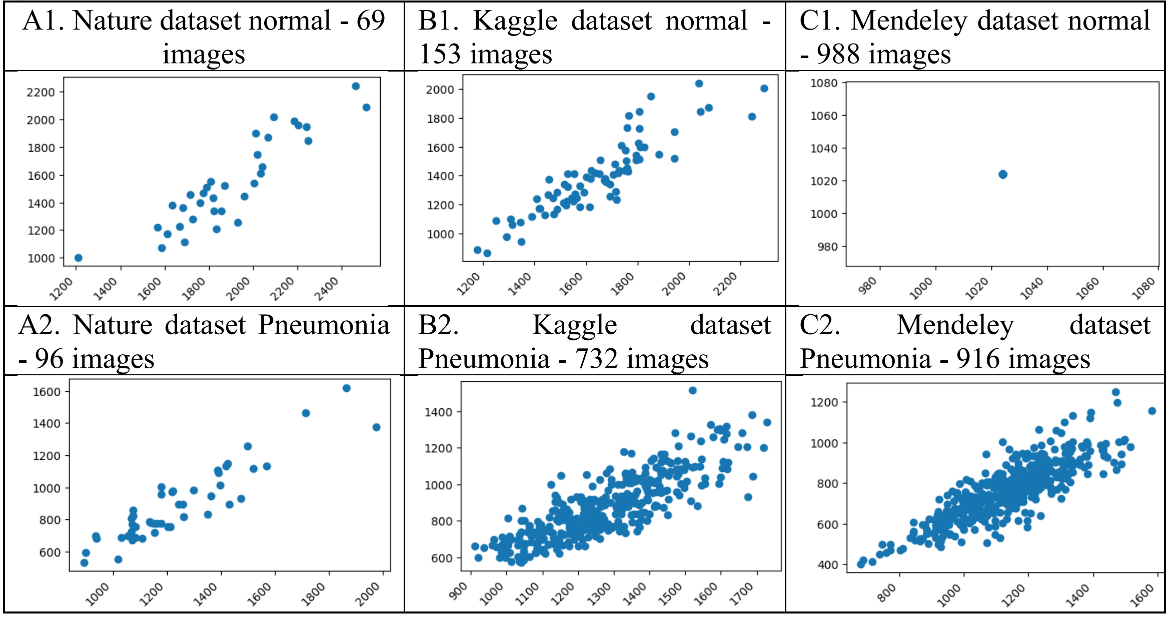
We have extracted 955 features (shape, texture, transform) from the X-ray images and have identified the features that are very effective in the classification of normal vs. Pneumonia. Among such features are wavelet transform features particularly low pass (LH) gray level uniformity, Correlation of GLCM, Inverse Difference Normalized, High Gray Level Zone Emphasis, Gray Level Non-Uniformity Normalized, high pass (HH) gray level uniformity, low pass (LH) run length uniformity, high pass (HH) run length gray level uniformity. We were able to achieve accuracy of 99% for our primary dataset and 90% for the secondary dataset which is unseen by the model in training phase. In this paper we particularly discussed and investigated the dataset open sourced by (Khuzani et al, 2021) in a paper submitted to Nature journal. We get the same accuracy levels whether we use all the features that were calculated after applying wavelet transform or the few above mentioned features.

## Introduction

In this paper we have used techniques and a finite set of features to classify the images as normal vs. Pneumonia. The goal of this research is to find the few key features that would classify the images with optimal accuracy. Deep learning models such as CNNs have produced promising results in image segmentation as well as classification (Scapicchio et al, 2021) and many products have been in use in the market to diagnose patients (Ridhi et al, 2024). But such successes were achieved at the cost of large input datasets which are not so easily available (Hatt et al, 2019) and using very powerful or custom-builtalgorithms (Hussein et al, 2024) and were implemented on expensive computing infrastructure. So, our approach is to extract all possible features from the images and identify the relevant features and use an algorithm well accepted or prominent in the industry and implement using reasonable computing power. Images acquired for diagnostics purposes are not just pictures, they are really data for analysts and computer science engineers in the form of 2D or 3D arrays. Source data for Radiomics is acquired through various modalities like X-ray, ultrasound, MRI, PET etc. There is intrinsic variability of the data that gets extracted from images of the above-mentioned modalities. The study of radiomics began a few years ago (kumar et al, 2012) but now it is very mature, and the results of the studies are being used in the industry settings, and various software and hardware products are being built. There are few societies in North America as well as Europe that have standardized quantitative imaging in order to ensure reliability (Scapicchio et al, 2021). One such initiative is the Image Biomarker Standardization Initiative (IBSI), which we used as the standard for this paper. Additionally, we used Pyradiomics, a Python-based package, which was developed following the IBSI standard (Griethuysen et al, 2017).

## I. Materials and Methods

This data contains a total of 235 images out of which 165 (69 normal images and 96 Pneumonia images) images are used for training and 70 images (30 normal images and 40 Pneumonia images) are used for testing the model that was developed (Khuzani et al, 2021). Pixel is a short form for a picture element that shows the intensity or the gray value at specified indices in the image array. A radiographic image that is of size 1024x750 will have gray level intensities in the form of a 2D matrix. A very detailed analysis of anatomy of these images is conducted as the purpose of this study was to use data analysis and build machine learning models to classify images into normal and pneumonia classes. The data collected in this paper was observed and compared with 2 other datasets as shown below. At the most basic level 1) the image sizes of threedifferent datasets are compared to each other using scatter plots 2) distribution of data of the pixel intensities using box plots 3) Frequency distribution of one feature through histograms are depicted in Fig1, Fig2, and Fig3 respectively.


A1. Nature dataset normal - 69 images

B1. Kaggle dataset normal - 153 images

C1. Mendeley dataset normal - 988 images

A2. Nature dataset Pneumonia - 96 images

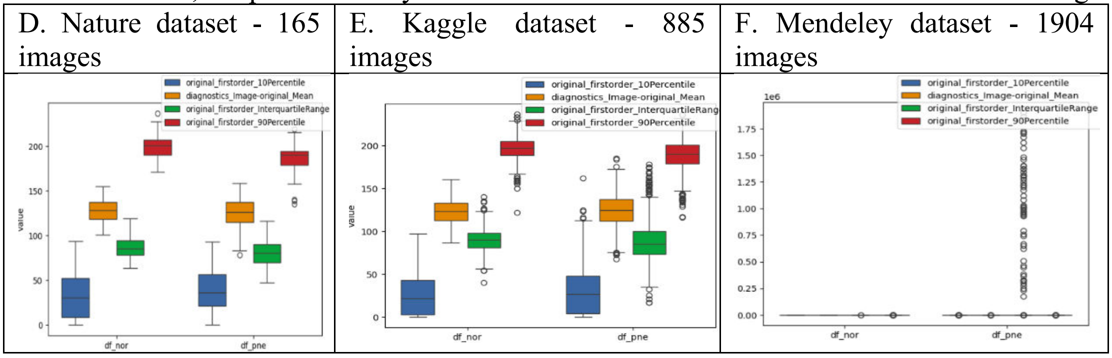
B2. Kaggle dataset Pneumonia - 732 images

C2. Mendeley dataset Pneumonia - 916 images

Fig 1: A1 shows the normal images of Nature paper dataset, B1 shows the normal images of Kaggle dataset, C1 shows the normal images of Mendeley dataset. A2, B2, C2 are the Pneumonia images of the same datasets respectively. An average intensity of two nearby pixels was computed to draw these charts.

From Fig 1, we can clearly observe that there is a lot of variation in the sizes of images. This intrinsic variability is expected among different modalities and even in same modality as X-ray because there are many OEMs that make these machines and in addition to that year and model of make also will have strong impact on the spatial resolution (Mayerhoefer et al, Pg.491, 2020). These variations and similarities of 3 different datasets are clearly depicted in Fig 2. Intensity values and the data distribution of image F are quite different from D and E datasets. Such variations will have to be accounted for before being fed into any algorithm of investigator's choice.

Further to the above, the pixel intensity distribution of the same 3 datasets above is shown in Fig2.


D. Nature dataset - 165 images

E. Kaggle dataset - 885 images

F. Mendeley dataset - 1904 images

Fig 2: D, E, F show the distribution of intensity levels of gray level values of normal images and pneumonia images of Nature paper dataset, Kaggle dataset, Mendeley dataset respectively. 1st

boxplot is 10th percentile, 2nd boxplot is Mean, 3rd boxplot is the interquartile range, 4th boxplot is the 90th percentile of gray level intensity values.

Fig 3. shows the histogram of a key feature known as wavelet low pass GLRLM Gray Level Non-Uniformity. Again, it can be clearly observed that there are variations and similarities even in the transform features that are calculated from the raw pixel values.

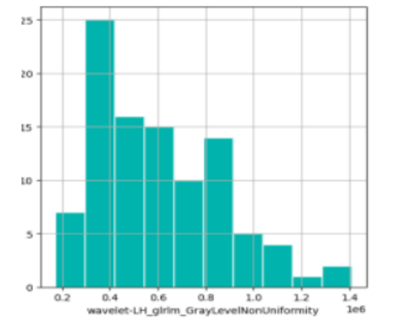
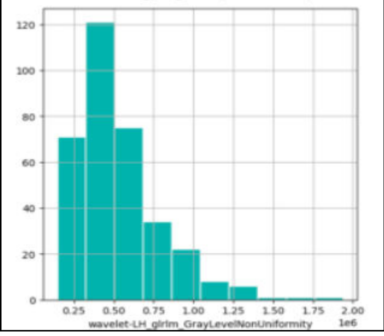
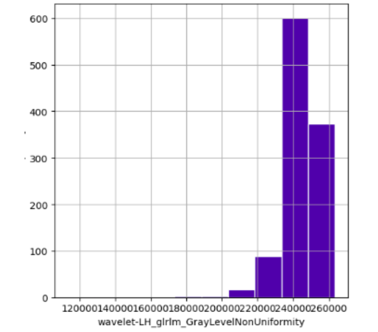
| G. Nature dataset normal - 69 images | H. Kaggle dataset normal - 153 images | I. Mendeley dataset normal - 988 images |
|---|---|---|
|  |  |  |

Fig 3: G, H, I show the distribution of a feature wavelet-LH_glrlm_GrayLevelNonUniformity of normal images in Nature paper dataset, Kaggle dataset, Mendeley dataset respectively.

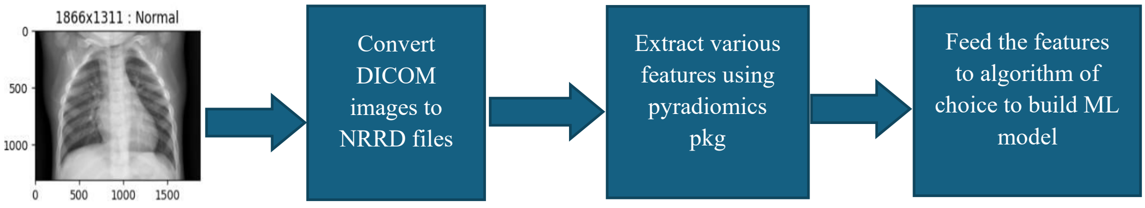
The features that are extracted are of three classes namely, shape, intensity, and texture based. Few of the very important features and their mathematical expressions are listed in Tab1.

| **Class** | **Feature** | **Expression** |
|---|---|---|
| **Shape** | Mesh surface | $A = \sum_{i=1}^{N_f} A_i$ |
| | Pixel surface | $A_{pixel} = \sum_{k=1}^{N_v} A_k$ |
| | Perimeter | $P = \sum_{i=1}^{N_f} P_i$ |
| | Major axis length | $major\ axis = \sqrt[4]{\lambda_{major}}$ |
| | Minor axis length | $minor\ axis = \sqrt[4]{\lambda_{minor}}$ |
| | Mean | $mean = \frac{1}{N_p} \sum_{i=1}^{N_p} X(i)$ |
| | Entropy | $entropy = -\sum_{i=1}^{N_g} p(i) log_2(p(i) + \epsilon$ |
| | Energy | $energy = \sum_{i=1}^{N_p} (X(i) + c)^2$ |

| **Class** | **Feature** | **Expression** |
|---|---|---|
| **Intensity** | Skewness | $skewness = \frac{\mu_3}{\sigma^3} = \frac{\frac{1}{N_p}\sum_{i=1}^{N_p}(X(i)-\bar{X})^3}{\left(\sqrt{\frac{1}{N_p}\sum_{i=1}^{N_p}(X(i)-\bar{X})^2}\right)^3}$ |
| | Kurtosis | $kurtosis = \frac{\mu_4}{\sigma^4} = \frac{\frac{1}{N_p}\sum_{i=1}^{N_p}(X(i)-\bar{X})^4}{\left(\frac{1}{N_p}\sum_{i=1}^{N_p}(X(i)-\bar{X})^2\right)^2}$ |
| **Texture** | Gray level Non-uniformity | $GLN = \frac{\sum_{i=1}^{N_g}\left(\sum_{j=1}^{N_s}P(i,j)\right)^2}{N_z}$ |
| | Run Length Non-uniformity | $RLN = \frac{\sum_{j=1}^{N_r}\left(\sum_{i=1}^{N_g}P(i,j\|\theta)\right)^2}{N_r(\theta)}$ |
| | Short Run Emphasis | $SRE = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{P(i,j\|\theta)}{j^2}}{N_r(\theta)}$ |
| | Long Run Emph*asis | $LRE = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}P(i,j\|\theta)j^2}{N_r(\theta)}$ |
| | Short Run Low Gray Level Emphasis | $SRLGLE = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{P(i,j\|\theta)}{i^2 j^2}}{N_r(\theta)}$ |
| | Short Run High Gray Level Emphasis | $SRHGLE = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{P(i,j\|\theta)i^2}{j^2}}{N_r(\theta)}$ |
| | Long Run Low Gray Level Emphasis | $LRLGLE = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{P(i,j\|\theta)j^2}{i^2}}{N_r(\theta)}$ |
| | Long Run High Gray Level Emphasis | $LRHGLE = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{P(i,j\|\theta)i^2j^2}{i^2}}{N_r(\theta)}$ |

Tab1. Few important features are listed here with their expressions.

Exploration and understanding of data have been done using the data visualization techniques mentioned above. Data extracted has approximately 955 features that were extracted using pyradiomics package via the following process flow shown in Fig 4.



Convert DICOM images to NRRD files

Extract various features using pyradiomics pkg

Feed the features to algorithm of choice to build ML model

Fig 4: Process flow shows conversion of DICOM images to NRRD files using SimpleITK package and then features are extracted to a .csv file which in turn will be fed to the desired algorithm for classification tasks.

| **Original** | **Gradient** | **Lbp-2d** | **Logarithm** | **Exponential** | **Square** | **Squareroot** | **Wavelet-LH** | **Wavelet-LL** | **Wavelet-HH** | **Wavelet-HL** |
|---|---|---|---|---|---|---|---|---|---|---|
| **95** | **86** | **86** | **86** | **86** | **86** | **86** | **86** | **86** | **86** | **86** |

Out of the 955 features that were extracted the following features are the list of features at high level. Tab2. List of all features calculated from the base pixels of the image. This list contains original image features which include shape features and also many transform-based features.

Each of above transform features along with original (raw pixels) image has 1) Firstorder, 2) Gray Level Co-occurrence Matrix (GLCM), 3) Gray Level Run Length Matrix (GLRLM), 3) Gray Level Size Zone Matrix (GLSZM), 4) Gray Level Dependence Matrix (GLDM) features except for the original feature which has shape features in addition to the above.

These features were fed into 2 different algorithms namely, GBDT (Gradient Boosted Decision Trees) and DNNs (Deep Neural Networks). We chose XGBoost (GBDT) as it is the most suitable algorithm for tabular data and it outperforms DNNs (Shwartz-Ziv, Armon, 2021). The XGBoost model uses an ensemble of trees trained sequentially, where new trees are built to correct the residuals of earlier models, and their outputs are combined to produce the final prediction (Wang et al, 2024). Given there are n observations and m features

$$X \in R^{m \times n}, \qquad y \in R^{n}$$

X: Feature matrix with n observations and m features

y: vector of dependent values or target class corresponding the n observations

$$\hat{y}_i = \sum_{t=1}^{T} f_t(X_i)$$

Where $f_t$ is a decision tree trained on the residuals of previous predictions.

## II. Results and Discussion

Accuracy of the model with XGBoost algorithm and features 1) Correlation (GLCM_corr), 2) Inverse Difference Normalized (GLCM_Idn), 3) High Gray Level Zone Emphasis (GLSZM_HGLZE), 4) Gray Level Non-Uniformity Normalized (GLSZM_GLNN) used is 100% where as DNN and the same above-mentioned features is 98%. These features mentioned above when compared to the rest of the features were key and sufficient to classify the images into normal and pneumonia classes. One way ANOVA test was performed on these four variables between training data and test data and results are as follows. GLCM_corr (F-statistic: 75.17, p-value = 7.43e-16), GLCM_Idn (F-statistic: 184.31, p-value = 2.54e-31), GLSZM_HGLZE (F-statistic: 303.84, p-value = 4.07e-44), GLSZM_GLNN (F-statistic: 347.65, p-value = 4.25e-48) indicate that their means are not the same. The F-statistic value is very high, and the p-value is very small when compared to the threshold of 0.05 with which we reject the null hypothesis $H_o$. This fact supports the argument that there is a signal in the data which separates the normal images from pneumonia images.

The correlation among the four above-mentioned variables is depicted in Fig 5.

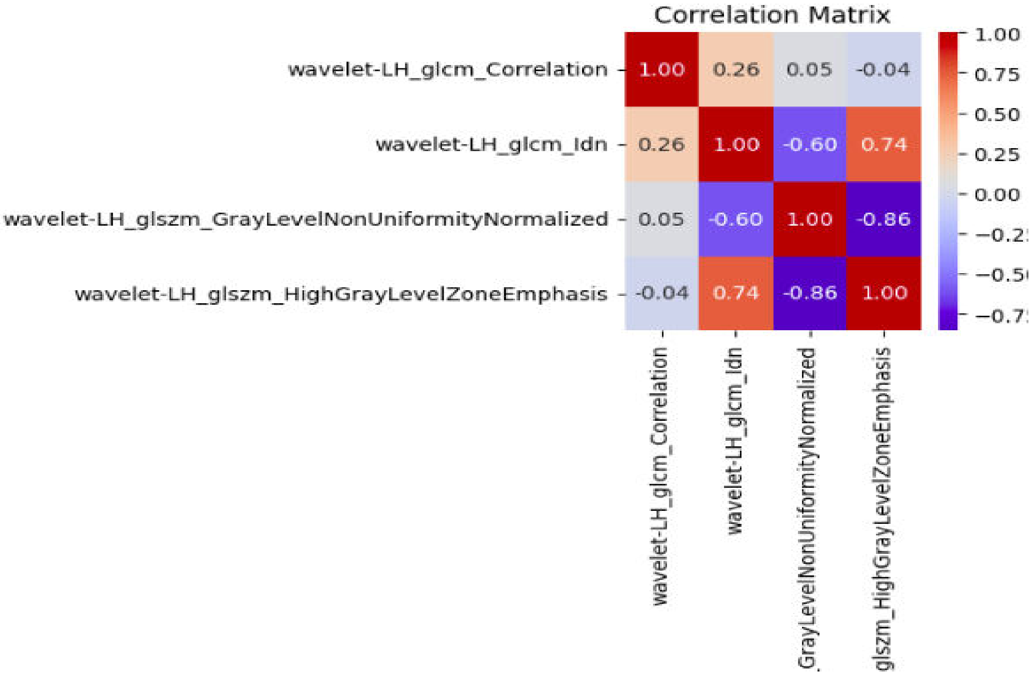


Fig5: Correlation among the variables Correlation(GLCM), Inverse Difference Moment Normalized (GLCM), High Gray Level Zone Emphasis (GLSZM),Gray Level Non-Uniformity Normalized (GLSZM)

The receiver operating characteristics on test data is as follows. While overfitting is a known trait of XGBoost algorithm, the fact is that it gave 100% accuracy (Fig 6) with the test data when all the wavelet transform features (344) listed in Tab2 were used. We had to iterate to find out which few features give the highest accuracy by removing those features that are having highest correlation among themselves. For example, Run Length Non-Uniformity Normalized of GLRLM contributes equally well, but it is a redundant feature. Also, manyof the features of the original raw pixels don't prove to have any positive effect on the accuracy of the model. In the similar fashion Short Run Emphasis and Long Run Emphasis of GLRM are also not having any positive effect on the model accuracy. As listed in the previous paragraph the four features that are giving the best results are from GLCM and GLSZM while GLRLM features are producing equally good performance and GLDM features are not performing as expected.

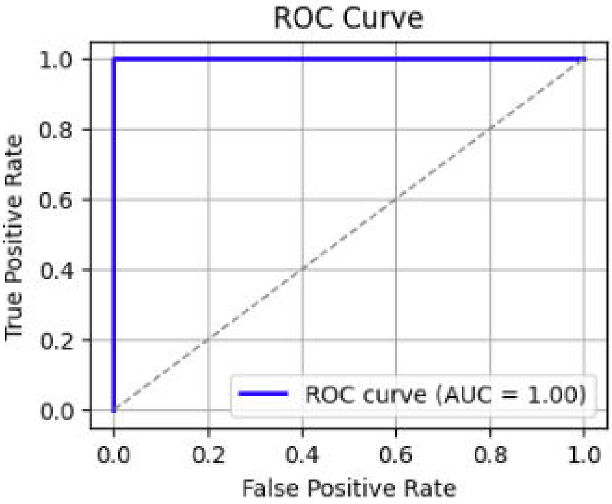


Fig6. RoC on the test data of 70 images.

The other transform features like Square, Squareroot, Exponential, Lbp-2D, Gradient, Logarithm were not contributing to optimal performance either individually or combined with any other features for this dataset.

Gray Level Co-occurrence Matrix (GLCM) describes the second-order joint probability function of an image region. Gray Level Size Zone Matrix (GLSZM) is quantification of gray level zones i.e. the number of pixels that share the same gray level intensity. Correlation and Inverse Difference Moment Normalized measures of GLCM transform that had the biggest impact on the accuracy along with GLSZM measures namely, High Gray Level Zone Emphasis and Gray Level Non-Uniformity Normalized. The feature importance scores given by the XGBoost'splot_importance function is depicted in Fig 7.

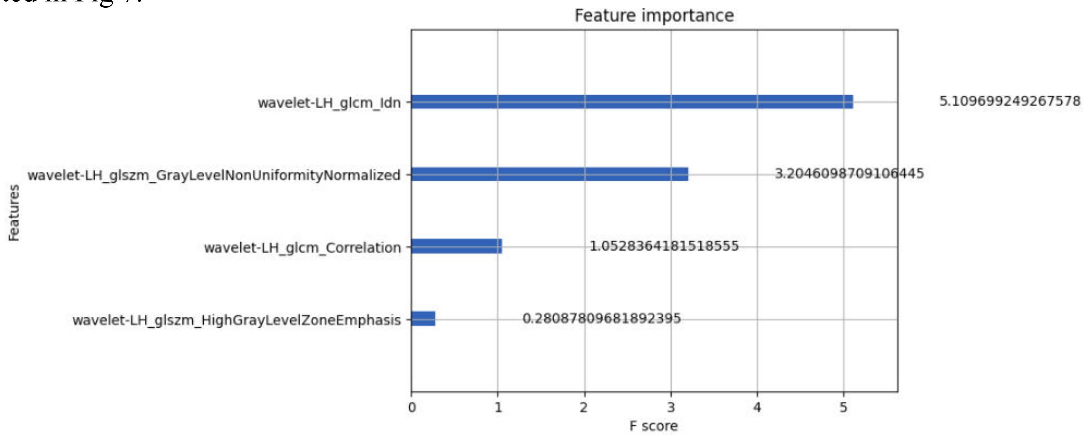


Fig7. The feature importance plot given by XGBoost'splot_importance function

**Correlation (GLCM)**: Descriptive statistics of the Correlation (Tab3) values show that normal images and pneumonia images have comparable correlation values.

$$Correlation = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)ij - \mu_x\mu_y}{\sigma_x(i)\sigma_y(j)}$$

| ***Metric*** | ***Train (normal)*** | ***Train (pneumonia)*** | ***Test(normal)*** | ***Test(pneumonia)*** |
|---|---|---|---|---|
| **count** | 69.00 | 96.00 | 30.00 | 40.00 |
| **mean** | 0.16 | 0.19 | 0.08 | 0.20 |
| **std** | 0.06 | 0.05 | 0.03 | 0.03 |
| **min** | 0.01 | 0.09 | 0.03 | 0.16 |
| **25%** | 0.12 | 0.16 | 0.05 | 0.17 |
| **50%** | 0.14 | 0.18 | 0.08 | 0.19 |
| **75%** | 0.21 | 0.21 | 0.11 | 0.21 |
| **max** | 0.28 | 0.32 | 0.15 | 0.30 |

Tab3. Shows the descriptive statistics of the Correlation measure of the GLCM feature.

**Inverse Difference Normalized(GLCM - IDN)**: Similar to the correlation feature we can observe that the Inverse Difference Normalized (Tab4) data also shows noticeable difference in the ranges of normal (Train and Test) together with pneumonia (Train and Test) features.

$$IDN = \sum_{k=0}^{N_g-1} \frac{p_{x-y}(k)}{1+\frac{k^2}{N_g^2}}$$

| *Metric* | *Train (normal)* | *Train (pneumonia)* | *Test(normal)* | *Test(pneumonia)* |
|---|---|---|---|---|
| count | 69.00 | 96.00 | 30.00 | 40.00 |
| mean | 0.98 | 0.98 | 0.98 | 0.98 |
| std | 0.01 | 0.00 | 0.00 | 0.00 |

| | | | | |
|---|---|---|---|---|
| min | 0.95 | 0.96 | 0.96 | 0.98 |
| 25% | 0.97 | 0.98 | 0.97 | 0.98 |
| 50% | 0.98 | 0.98 | 0.98 | 0.98 |
| 75% | 0.98 | 0.98 | 0.98 | 0.98 |
| max | 0.98 | 0.99 | 0.98 | 0.99 |

Tab4. Show the descriptive statistics of Inverse Difference Normalized which part of GLCM feature

**High Gray Level Zone Emphasis (GLSZM - HGLZE):** Clear distinction can be observed in the High Gray level emphasis (Tab5) measurements among the normal and pneumonia classes between the training data and test data. The quartiles of training data of normal images range from 72 to 90 whereas the testing dataset ranges from 56 to 90. Also, quartiles of training data of pneumonia images range from 162 to 186 whereas testing dataset range from 185 to 187.

$$HGLZE = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_s} p(i,j) i^2}{N_z}$$

| *Metric* | *Train (normal)* | *Train (pneumonia)* | *Test(normal)* | *Test(pneumonia)* |
|---|---|---|---|---|
| count | 69.00 | 96.00 | 30.00 | 40.00 |
| mean | 76.19 | 160.43 | 79.47 | 170.07 |
| std | 19.52 | 47.84 | 16.62 | 42.76 |
| min | 12.36 | 30.55 | 42.61 | 56.84 |
| 25% | 72.42 | 162.74 | 56.92 | 185.47 |
| 50% | 90.22 | 185.82 | 90.58 | 186.23 |
| 75% | 90.74 | 186.89 | 90.78 | 187.18 |
| max | 91.68 | 217.62 | 90.97 | 215.96 |

Tab5. Shows the descriptive statistics of the High Gray Level Zone Emphasis feature of the GLSZM feature.

**Gray Level Non-Uniformity Normalized (GLSZM - GLNN):** Similar to the High Gray Level Zone Emphasis Gray Level Non-Uniformity Normalized also has similar observations. There seems to be a signal between normal and pneumonia images when the mean or quartiles are observed.

$$GLNN = \frac{\sum_{i=1}^{N_g}(\sum_{j=1}^{N_s} P(i,j))^2}{N_z^2}$$

| *Metric* | *Train (normal)* | *Train (pneumonia)* | *Test(normal)* | *Test(pneumonia)* |
|---|---|---|---|---|
| count | 69.00 | 96.00 | 30.00 | 40.00 |
| mean | 0.44 | 0.35 | 0.44 | 0.33 |
| std | 0.02 | 0.05 | 0.01 | 0.05 |
| min | 0.39 | 0.20 | 0.41 | 0.18 |
| 25% | 0.43 | 0.32 | 0.44 | 0.31 |
| 50% | 0.44 | 0.35 | 0.44 | 0.33 |
| 75% | 0.46 | 0.38 | 0.45 | 0.35 |
| max | 0.48 | 0.48 | 0.47 | 0.43 |

Tab6. Shows the descriptive statistics of the Gray Level Non-Uniformity Normalized feature of the GLSZM feature.

## III. Conclusion

Diagnosis of Pneumonia in radiographic images comes with many challenges namely, lack of trained radiologists and lack of equipment that can develop quality films. There have been many studies and Machine Learning technology-based solutions and products that would assist the radiologists with accurate diagnosis. These activities fall under classification of X-ray images into normal and pneumonia classes. But most studies have used Deep Neural Networks that demand larger amounts of input data than what is commonly available and expensive computing power. So, our approach was to use a finite set of constructed features using data analytics and commonly available Machine Learning algorithms and reasonably good computing infrastructure. We are able to achieve the above-mentioned goals in this study and we have identified a few features from the extracted feature set of 955 features and an ensemble algorithm XGBoost widely used in academia to achieve optimal classification accuracy. When we tested this model on unseen data, the model accuracy of that data was at 90.3% without any augmentation to data or the model.

### Research gap

Further research can be conducted using diverse datasets and various feature combinations to determine the common features that significantly influence classification accuracy across different datasets. Such exercise would reduce the necessity to depend on large amounts of data and complex algorithms to classify radiographic images.

### Conflict of Interest

There are no conflicts of interest associated with this study, which is conducted solely for academic and public research purposes.

**Funding**

There is no funding organization that supports this paper.

### References

1.Griethuysen et al, 2017: van Griethuysen, J. J. M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R. G. H., Fillon-Robin, J. C., Pieper, S., Aerts, H. J. W. L. (2017). Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Research, 77(21), e104–e107. `https://doi.org/10.1158/0008-5472.CAN-17-0339 <https://doi.org/10.1158/0008-5472.CAN-17-0339>`

2.Hatt et al, 2019: Mathieu Hatt, Catherine Cheze Le Rest, Florent Tixier, Bogdan Badic, Ulrike Schick, and Dimitris Visvikis; Radiomics : Data Are Also Images, J Nucl Med 2019; 60:38S–44S, DOI: 10.2967/jnumed.118.220582

3.Hussein, A.M., Sharifai, A.G., Alia, O.M. et al. Auto-detection of the coronavirus disease by using deep convolutional neural networks and X-ray photographs. Sci Rep 14, 534 (2024). https://doi.org/10.1038/s41598-023-47038-3

4.Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, Forster K, Aerts HJ, Dekker A, Fenstermacher D, Goldgof DB, Hall LO, Lambin P, Balagurunathan Y, Gatenby RA, Gillies RJ. Radiomics: the process and the challenges. Magn Reson Imaging. 2012 Nov;30(9):1234-48. doi: 10.1016/j.mri.2012.06.010. Epub 2012 Aug 13. PMID: 22898692; PMCID: PMC3563280.

5.Mayerhoefer ME, Materka A, Langs G, Häggström I, Szczypiński P, Gibbs P, Cook G. Introduction to Radiomics. J Nucl Med. 2020 Apr;61(4):488-495. doi: 10.2967/jnumed.118.222893. Epub 2020 Feb 14. PMID: 32060219; PMCID: PMC9374044.

6.Ravid Shwartz-Ziv, Amitai Armon, Tabular data: Deep learning is not all you need, Information Fusion, Volume 81, 2022, Pages 84-90, ISSN 1566-2535, https://doi.org/10.1016/j.inffus.2021.11.011.

7.Ridhi S, Robert D, Soren P, Kumar M, Pawar S, Reddy B Comparing the Output of an Artificial Intelligence Algorithm in Detecting Radiological Signs of Pulmonary Tuberculosis in Digital Chest X-Rays and Their Smartphone-Captured Photos of X-Ray Films: Retrospective Study JMIR Form Res 2024;8:e55641, doi: 10.2196/55641

8.Ruihan Wang, Mingyang Zhang, Fuzhong Gong, Shaohan Wang, Ran Yan, Improving port state control through a transfer learning-enhanced XGBoost model, Reliability Engineering & System Safety, Volume 253, 2025, 110558, ISSN 0951-8320, https://doi.org/10.1016/j.ress.2024.110558.

9.Scapicchio C, Gabelloni M, Barucci A, Cioni D, Saba L, Neri E. A deep look into radiomics. Radiol Med. 2021 Oct;126(10):1296-1311. doi: 10.1007/s11547-021-01389-x. Epub 2021 Jul 2. PMID: 34213702; PMCID: PMC8520512.

10.Zargari Khuzani, A., Heidari, M. & Shariati, S.A. COVID-Classifier: an automated machine learning model to assist in the diagnosis of COVID-19 infection in chest X-ray images. Sci Rep 11, 9887 (2021). https://doi.org/10.1038/s41598-021-88807-2