PneumoniaCXR: AI-Enabled Pneumonia Detection

Gary Kong, Drew Piispanen, Diqing Wu

Contents

Introduction

- **The COVID-19 Crisis**: Underscored the urgent need for accurate diagnostics.
- **Hard to Differentiate**: COVID-19 vs. non COVID-19 vs. normal lung issues can be hard to diagnose.
- **Understaffing of Medical Professionals**: Only doctors can diagnose these issues, but many are understaffed with the need for second opinions.
- **Our Approach**: Using feature sets (HOG, ResNet, Pyradiomics) we'll classify CXR images, aiming for improved diagnostics.

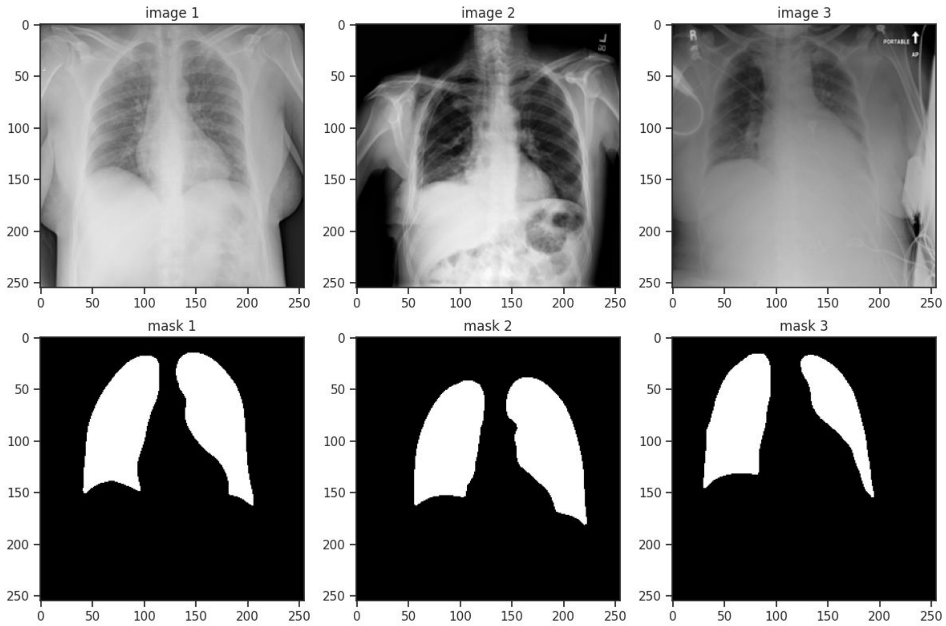
Data and Preprocessing

Dataset:

- **From Kaggle** - 32,103 CXR images (255, 255) of chest x-rays from all over the world, labeled to three classes:
    - COVID-19 pneumonia
    - Non COVID-19 pneumonia
    - Normal non-pneumonia
- **After processing**:
    - 10,701 COVID-19, 10,701 non-COVID, and 10,701 Normal

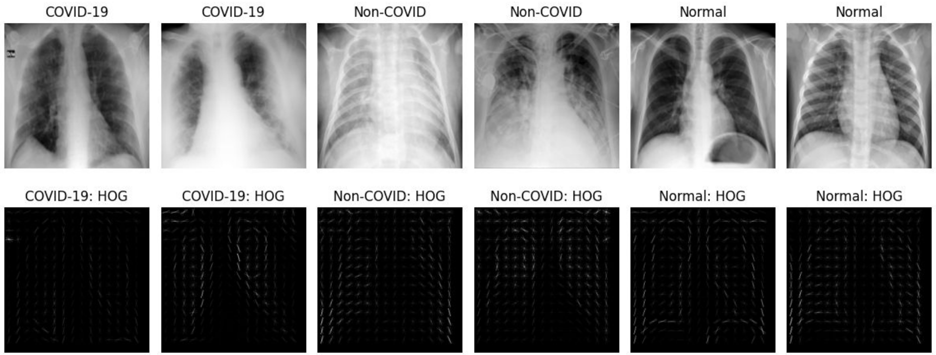
Data Preprocessing:

- **Contrast Normalization**: Applied z-score scaling to normalize intensity across images.
- **Image Cropping:** Removed blank space and zoomed out images. Applied to both image and mask.
- **Orientation Standardization**: OpenCV image alignment, adjusting spines to align perpendicularly. Applied to both image and mask.
- **Balancing Classes**: Implemented random undersampling to balance the 3 classes.



Feature Engineering

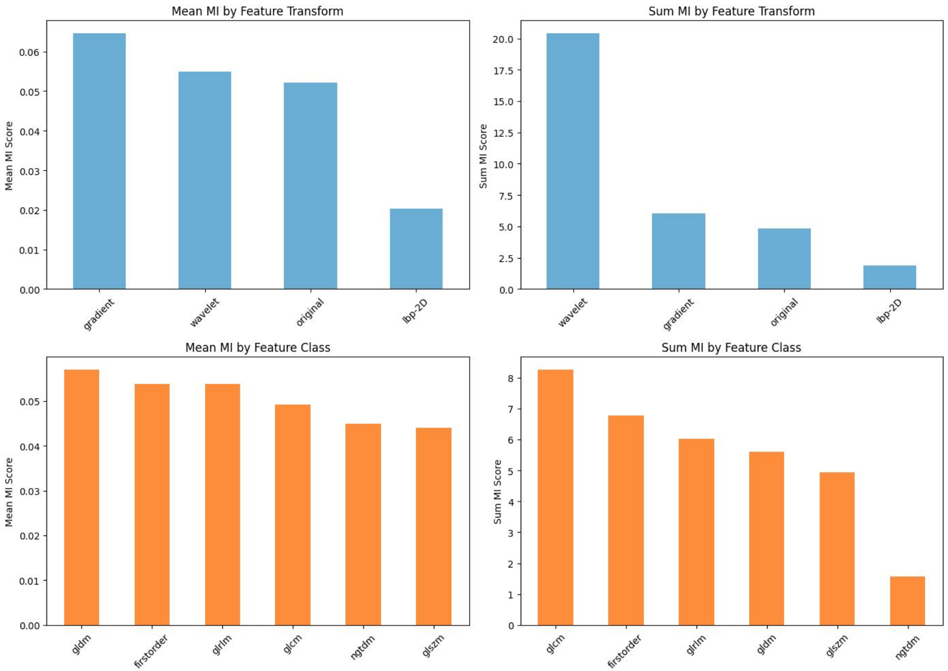
Histogram of Oriented Gradients (HOG)



Parameters	Initial Configuration	Search Values	Parameter Importance	Best Configuration
Image Size	(128, 255)	(128,255), (255,128), (255,255)	30.9%	(255, 255)
Orientations	9	7, 8, 9	4.3%	9
Pixels per Cell	(16, 6)	(8,8), (12,12), (16,16)	58.7%	(16, 16)
Cells per Block	(2, 2)	(2,2), (3,3)	6.2%	(2, 2)

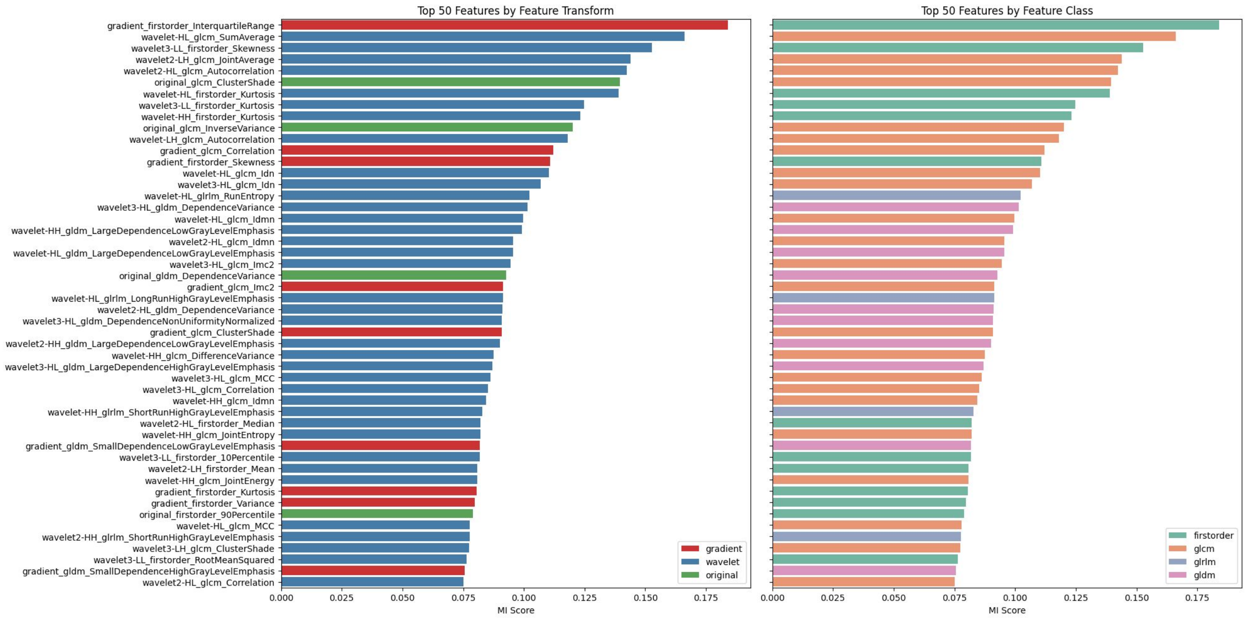
- **Relevance to Pneumonia:** Orientations, locations and strength of edges are relevant to consolidations and ground-glass opacities.
- **Hyperparameter Tuning:** Optimized image size, orientation pixels per cell, and cells per block to maximize mutual information
- **Initial Configuration:** Based our initial settings on a similar study of pneumonia detection.
- **Key Parameters:** Pixels per cell and image size were most crucial for maximizing mutual information.
- **Optimized Image Size:** Modified the image size from the initial configuration to enhance performance.

Radiomics: Capturing Textural Cues (1 of 2)



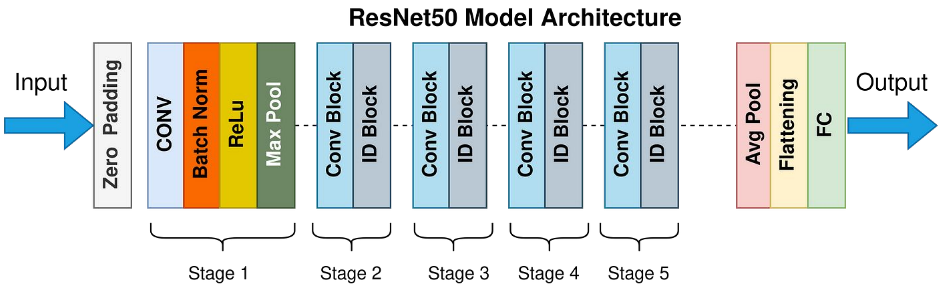
- **Texture and Disease:** Radiomics extracts quantitative features from medical images, revealing textural patterns correlated with underlying pathologies.
- **Transformations Matter:** Gradient and wavelet transformations enhanced the discriminative power of our radiomics features.
- **Feature Class Insights:** GLCM, GLDM, first-order statistics, and GLRLM emerged as particularly informative for pneumonia detection.

Radiomics: Capturing Textural Cues (2 of 2)



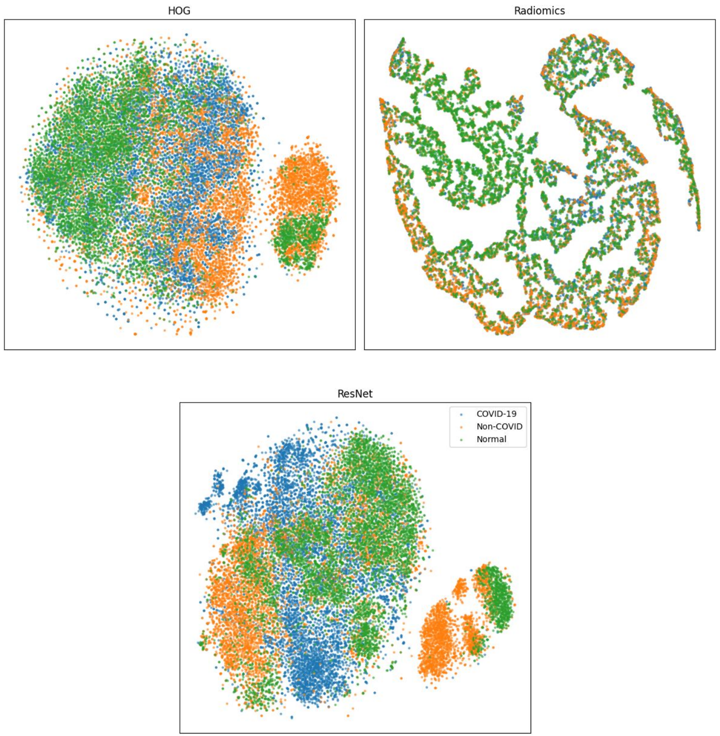
- **Reducing Redundancy:** Reduced correlated features to enhance efficiency and avoid bias towards correlated features in PCA
- **Wavelet Dominance:** Wavelet-transformed features proved highly informative.
- **Diverse Feature Classes:** Top features include First-order, GLCM, GLDM, and GLRLM

Deep Learning with ResNet50



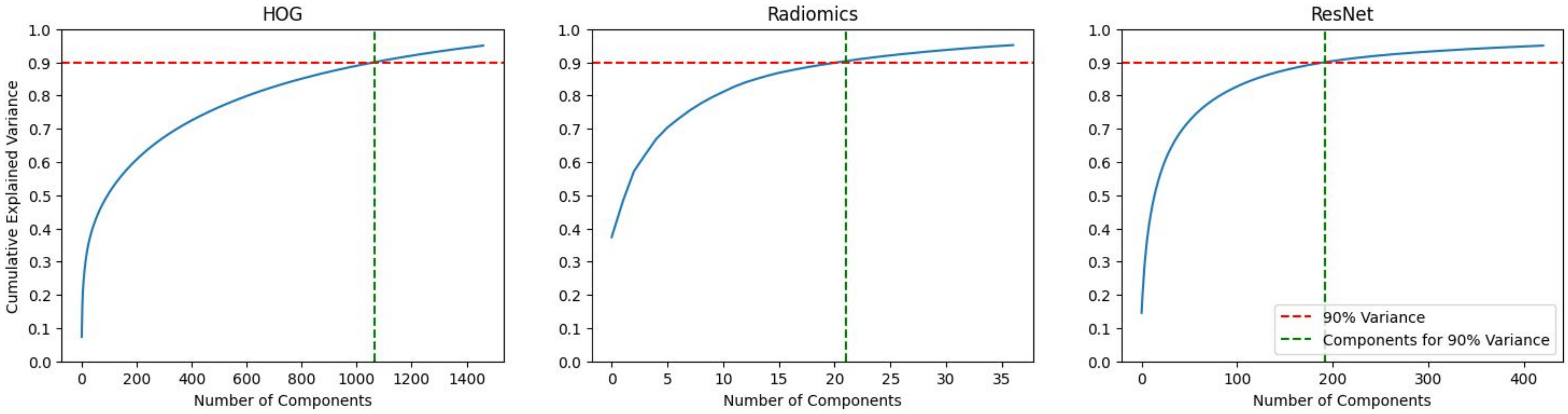
- **Proven in Medical Imaging**: ResNet demonstrates strong performance in medical image analysis, including pneumonia classification.
- **ResNet50 for Efficiency:** We selected ResNet50 for its balance of depth and computational efficiency.
- **Pre-trained Power**: Employed a pre-trained ResNet50 model on ImageNet for robust feature extraction.
- **Adaptation**: Images were resized and normalized for compatibility with the pre-trained model.

t-SNE Visualizations



- **Dimensionality Reduction for Visualization:** t-SNE projects high-dimensional features into 2D or 3D space for easier exploration.
- **Insights into Class Separability:** t-SNE helps us assess if features naturally cluster by pneumonia type.
- **Partial Clustering:** HOG and ResNet features show some clustering, suggesting potential discriminative power.
- **Radiomics Diversity:** Lack of clustering in radiomics features could be due to their diverse nature.
- **t-SNE Limitations:** Challenges in preserving relationships between heterogeneous features in lower dimensions.

Principal Component Analysis



	HOG	Resnet	Radiomic	Total
all features	13689	2048	241	15978
90% explained	1065	193	22	1280
80% explained	601	84	11	696

- Concatenated for classification model training

Results

Feature Importance

- Feature importance from Random Forest Model using 90% explained variance PCA
  - No one set of features shows significantly more importance than the others across different metrics
    - Sum of importance
    - Average importance per feature
    - Max single feature importance
    - etc.

	HOG	Resnet	Radiomic
sum importance	50.14%	26.59%	23.27%
num of features	1065	193	22
average importance per feature	0.05%	0.14%	1.06%
max feature importance	6.01%	4.34%	2.91%

Top 10	importance
hog_0	6.0%
resnet_1	4.3%
hog_1	4.2%
resnet_0	3.4%
radio_2	2.9%
radio_10	2.9%
radio_3	2.2%
resnet_3	2.2%
radio_1	2.1%
radio_5	2.1%

Classification Performance

Hyperparameter Tuning
- SVM
- Logistic Regression
- Gradient Boosting
- Random Forest

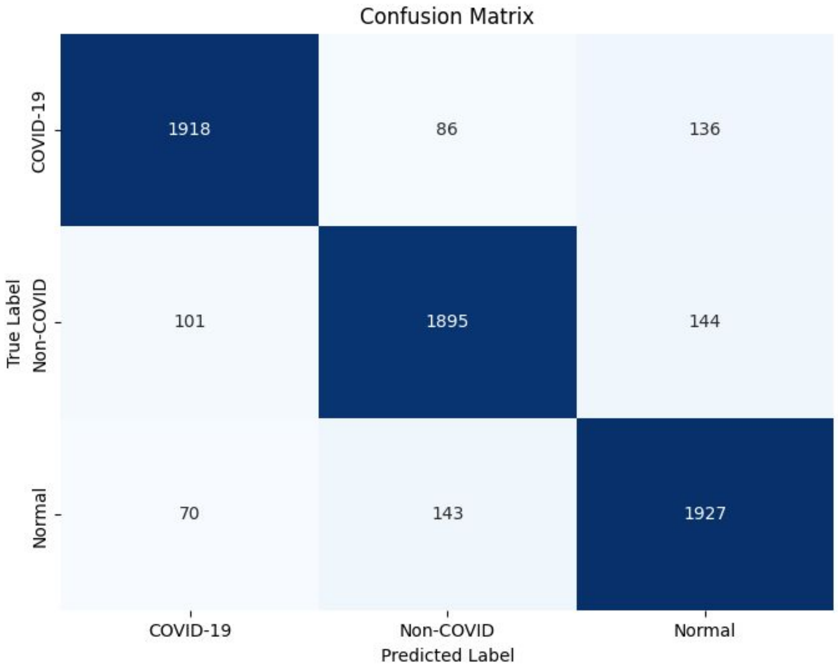
Best set of parameters is also used for efficiency comparison

Model	Number of Trials	Search Time Per Trial (s)	Parameter	Search Values	Best Value
SVM	25	1577/25	C	[1e-4, 1e-3, 1e-2, 1e-1, 1, 1e1, 1e2]	1
			kernel	'linear', 'sigmoid', 'rbf'	rbf
Logistic Regression	25	1584/25	C	[0.01, 0.1, 1.0, 10]	0.1
			penalty	[‘L1’, L2’]	L1
Gradient Boosting	15	30411/15	learning_rate	[0.01, 0.05, 0.1]	0.1
			n_estimators	[100,200]	100
			subsample	[0.5, 1.0]	1.0
Random Forest	25	1336/25	n_estimators	[50, 100, 200]	200
			max_depth	[5, 10]	10
			criterion	["gini", "entropy"]	“entropy”
			min_samples_split	[2, 5]	5

Classification Performance

Model	Accuracy			F1-Score		
	Training Set	Validation Set	Test Set	Training Set	Validation Set	Test Set
SVM	0.97	0.87	**0.89** 🤗	0.97	0.87	**0.89** 🤗
Logistic Regression	0.91	0.88	**0.90** 🤗	0.91	0.88	**0.90** 🤗
Gradient Boosting	0.86	0.85	0.85	0.86	0.85	0.85
Random Forest	0.92	0.81	0.82	0.92	0.81	0.82

Classification Performance – Logistic Regression



	precision	recall	f1-score	support
COVID-19	0.93	0.89	0.91	2140
Non-COVID	0.88	0.90	0.89	2140
Normal	0.88	0.90	0.89	2140
accuracy			0.90	6420
macro avg	0.90	0.90	0.90	6420
weighted avg	0.90	0.90	0.90	6420

Misclassification Examples

True Label: Covid-19
Predicted Label (misclassify): Normal

True Label: Normal
Predicted Label (misclassify): Covid-19

True Label: Non-Covid
Predicted Label (misclassify): Normal

True Label: Normal
Predicted Label (misclassify): Non-Covid

Efficiency

- 90% PCA

Model	Per Trial Tuning Time (s)	Training Time (s)	Prediction Time (s)			Accuracy	F1-Score
			Training Set	Validation Set	Test Set	Test Set	Test Set
Logistic Regression	63.36	17	0.02	0.01	0.01	0.90	0.90
SVM	63.08	21	50	12	16	0.89	0.89
Gradient Boost	2094.07	2089	0.30	0.06	0.08	0.85	0.85
Random Forest	53.4	111	0.60	0.14	0.17	0.82	0.85

- 80% PCA: faster but not significantly
- All features: crash the runtime 🙁

Generalizability

- Logistic Regression & Gradient Boosting
  - Good generalizability
  - The minimal drop between validation and test set accuracy suggests the model generalizes well to unseen data

- SVM & Random Forest
  - Potential Overfitting?
  - Visible accuracy drop from train to validation/test
  - But not necessarily as it is not significant

Model	Accuracy			F1-Score		
	Training Set	Validation Set	Test Set	Training Set	Validation Set	Test Set
SVM	0.97	0.87	**0.89**	0.97	0.87	0.89
Logistic Regression	**0.91**	**0.88**	**0.90**	0.91	0.88	0.90
Gradient Boosting	0.86	0.85	0.85	0.86	0.85	0.85
Random Forest	0.92	0.81	0.82	0.92	0.81	0.82

Conclusions

Created an effective model to classify COVID-19, non-COVID pneumonia, and normal chest x-rays.

- 90% model accuracy
- Good generalizability
- No one feature set had greater importance over the others.
- Future studies could work with CXR machine-specific models, and future model experimentation.