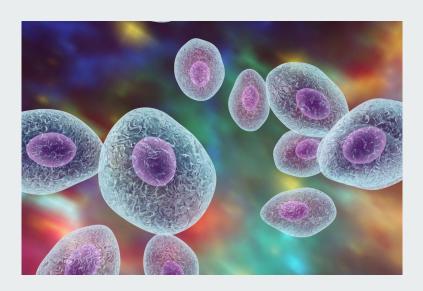
# Chest X-Ray Image Pneumonia Classification

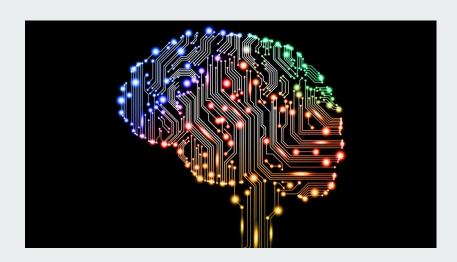
By: Joe Marx and Sandra Welbeck

## **Project Overview**



- We trained a Convolutional Neural Network to help detect pneumonia in chest x-ray images.
- Pneumonia is a contagious lung infection that disproportionately affects the elderly and young and is the leading cause of death among children under 5 years old.<sup>1</sup>
- Pneumonia is also one of the most expensive conditions seen during hospitalizations in the U.S.<sup>2</sup>

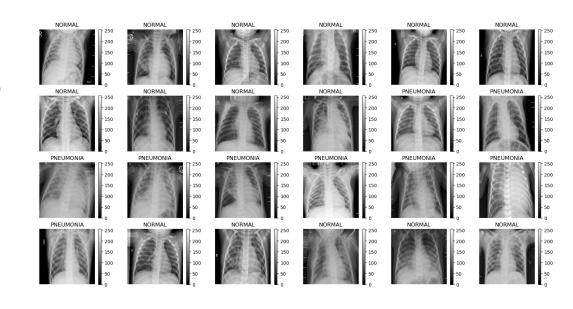
#### **Business Problem**



 Detecting pneumonia solely from x-rays is a difficult task for the human eye, so we sought to train a Convolutional Neural Network to recognize pneumonia in x-ray images to help speed the process of pneumonia detection and help prevent cases from going undetected.

#### **Data Source**

- Kaggle's Chest X-Ray Images (Pneumonia)<sup>3</sup>
- Over 5.5 thousand jpeg images split into train, test, and validation folders
- Each folder consisted of a "Pneumonia" and "Normal" folder
- 16 images in original validation set
  - Training: 4616 images
  - o Testing: 640 images
  - Validation: 600 images



conv2d_10_in	put: InputLayer		conv2d_1	10: Conv2D	[	max_pooling2d_	10: MaxPooling2D	]	conv2d_1	1: Conv2D	]	max_pooling2d_1	1: MaxPooling2D	7	conv2d_1	2: Conv2D	max_pooling2d	_12: MaxPooling2D	] [	conv2d_1	3: Conv2D		max_pooling2d_1	3: MaxPooling2D
input:	output:	-	input:	output:	-	input:	output:	-	input:	output:	⊢∙	input:	output:	-	input:	output:	input:	output:	-	input:	output:	-	input:	output:
[(None, 256, 256, 1)]	[(None, 256, 256, 1)]		(None, 256, 256, 1)	(None, 256, 256, 32)		(None, 256, 256, 32)	(None, 128, 128, 32)		(None, 128, 128, 32)	(None, 128, 128, 64)		(None, 128, 128, 64)	(None, 64, 64, 64)		(None, 64, 64, 64)	(None, 64, 64, 128)	(None, 64, 64, 128	(None, 32, 32, 128)	] [	(None, 32, 32, 128)	(None, 32, 32, 128)	] [	(None, 32, 32, 128)	(None, 16, 16, 128)

### **Approach**

- Keras package, prediction model was built iteratively and updated using metric optimization.
- Training set had a class imbalance— class weights were implemented
- Data augmentation of the training set to account for class imbalance
- Final model consists of over 3 million parameters



	conv2d_14: Conv2D							
	input:	output:						
(N	None, 16, 16, 12	8) (None, 16, 16, 256)						

input:		output:	-
(None, 16, 1	6, 256)	(None, 8, 8, 256)	

1	input:	output:
	None, 8, 8, 256)	(None, 8, 8, 256



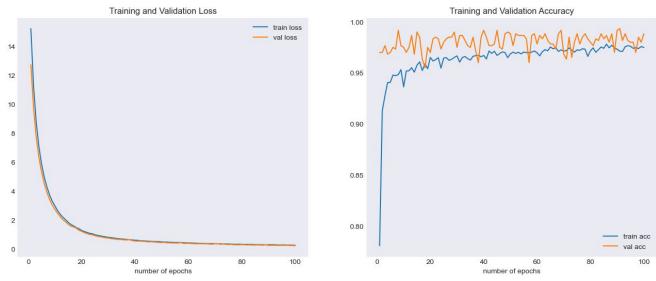
	flatten_2:			
-	input:	output:	-	
	(None, 4, 4, 256)	(None, 4096)		

ı	dropout_2			
	input:	output:	-	8
ı	(None, 4096)	(None, 4096)		(No

dense_4	: Dense		dense_5:	Dense
input:	output:	-	input:	outpu
(None 4096)	(None 512)	1	(None, 512)	None

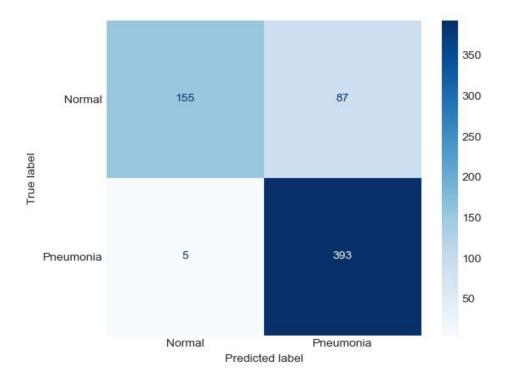
#### Results

- On our validation set, our model achieved 99% accuracy, 99% recall, and 98% percent precision.
- When tested on unseen data, our model received 86% accuracy, 99% recall, and 82% precision.



# Conclusions and Recommendations

- The majority of the positive pneumonia cases were accurately detected from our model
- We recommend that the doctor should do additional testing to detect false positive cases



## **Next Steps**



- Pre-trained Models/ other modeling architectures
- Principal Component Analysis (PCA)
- Different Image Augmentation
- Additional methods to address class imbalance
- Train using a variety of other parameters
- Separating pneumonia cases into viral and bacterial
- Test classifier on other medical images

# Questions?

#### **Sources**

- 1. <a href="http://54.209.11.195/patients/patient-resources/resources/top-pneumonia-facts.pdf">http://54.209.11.195/patients/patient-resources/resources/top-pneumonia-facts.pdf</a>
- 2. Ibid., <a href="http://54.209.11.195/patients/patient-resources/resources/top-pneumonia-facts.pdf">http://54.209.11.195/patients/patient-resources/resources/top-pneumonia-facts.pdf</a>
- 3. <a href="https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia">https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia</a>