Pneumonia diagnosis using chest x-ray images

Brooke Stevens
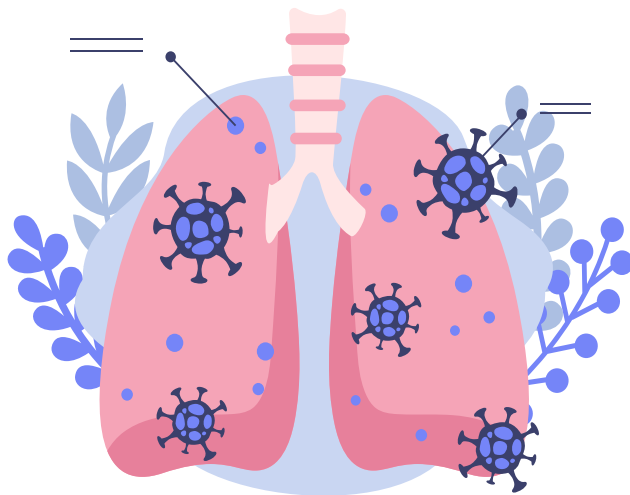
Agenda

01
Background

02
Problem

03
**Literature Review**

04
Approach

05
Results

06
Discussions and Conclusions

Background

- **2.5 million people** died from pneumonia in 2019, and **24%** were under 5 years of age

- Pneumonia is the **second most common cause of hospital admissions**, only behind women giving birth

- Pneumonia is the **leading cause of hospitalization of infants**

- Common risk factors include **air pollution**, a **lack of access to water** for hand hygiene, and **low temperatures**

- It disproportionately affects the **young**, the **elderly**, and the **immunocompromised**
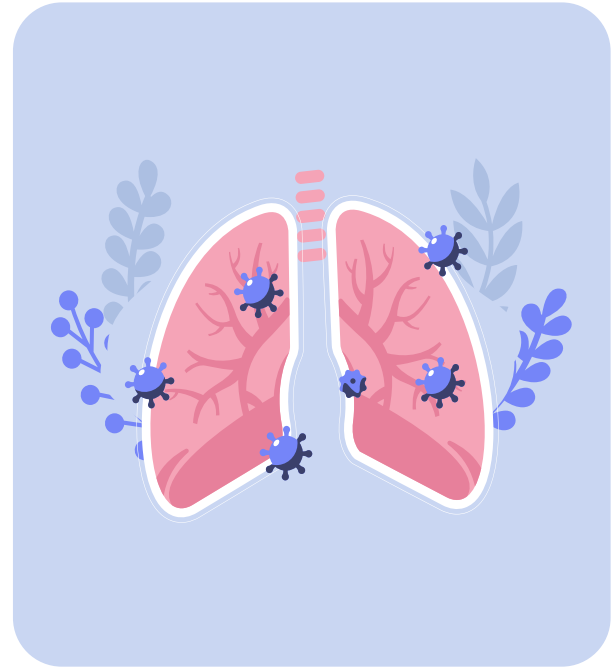
Problem

- Many respiratory diseases have **similar symptoms**, so it is often difficult for physicians to deliver confident diagnoses

- These symptoms include, but are not limited to: **cough, sneezing, sore throat, chest pain, fever, fatigue**

- Physicians would benefit from **real-time assistance** to diagnose pneumonia

  - Proper diagnoses allow for proper treatment methods for patients
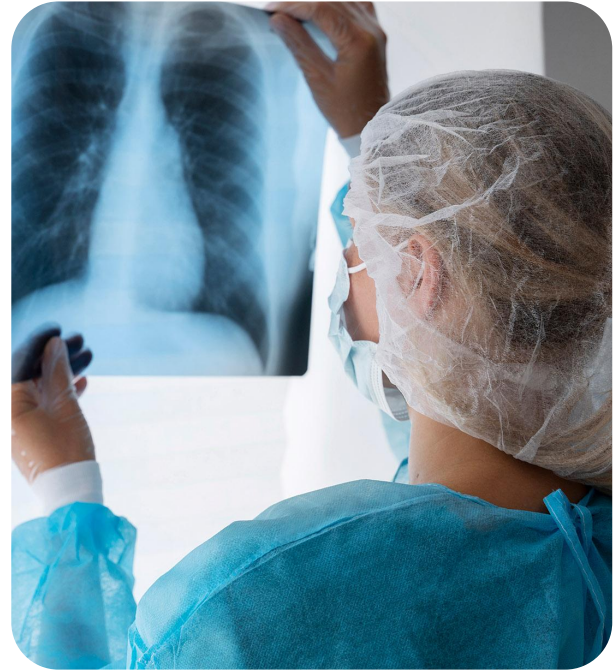
Literature Review

Liquid biopsy system tool, Species-Specific Bacterial Detector (SSBD) (Wang et al.):

- http://proxy.library.vanderbilt.edu/login?url=https://www.proquest.com/scholarly-journals/novel-fast-pathogen-diagnosis-method-severe/docview/2730538532/se-2?accountid=14816

Serological and PCR-based diagnosis method (Herrera et al.):

- https://link.gale.com/apps/doc/A444844867/AONE?u=tel_a_vanderbilt&sid=bookmark-AONE&xid=a47c262a

Sepsis prediction tool (HW1) (Sendak et al.):

- http://proxy.library.vanderbilt.edu/login?url=https://www.proquest.com/scholarly-journals/real-world-integration-sepsis-deep-learning/docview/2511970280/se-2?accountid=14816
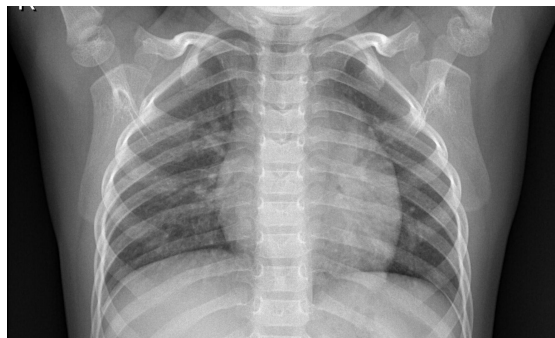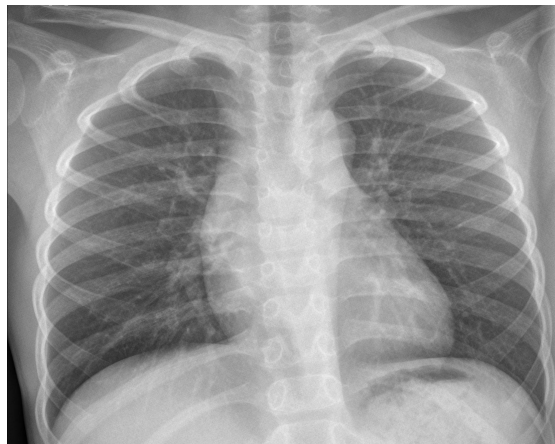
Approach

- Image data retrieved from Mendeley Data, courtesy of the University of California San Diego

  ○ Train normal: 1,349

  ○ Train pneumonia: 3,883

  ○ Test normal: 234

  ○ Test pneumonia: 390

- Test a wide variety of simple classification models

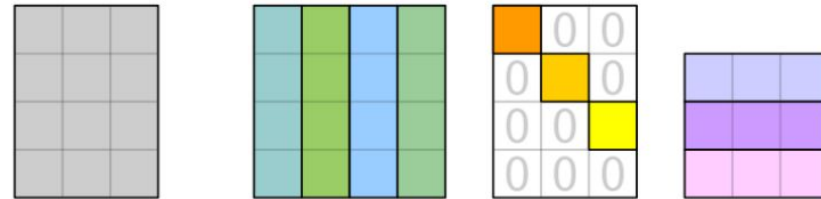- Fine-tune the model with the highest AUC
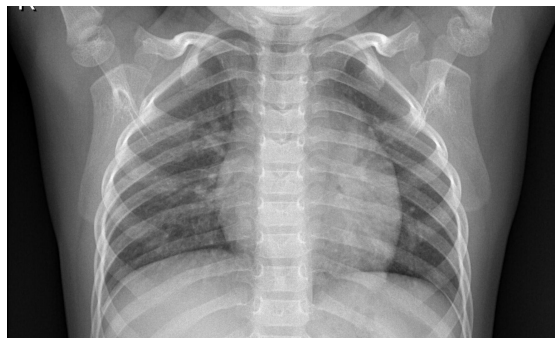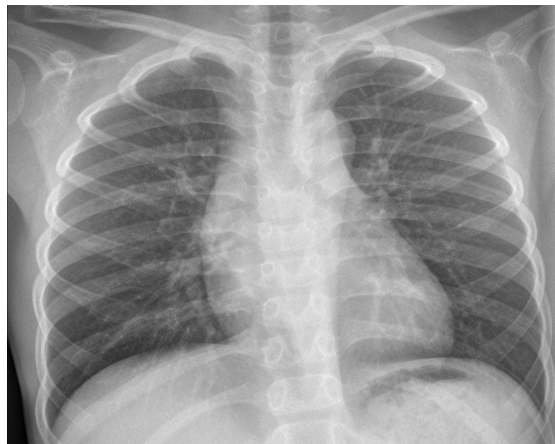
Approach

Method

- Scale all images to 64 x 64 and convert to numerical dataframe (12,286 features)

- Use SVD to determine principal components and perform dimension reduction

- Employ simple sklearn models to determine the most promising classifier

- Fine-tune a hyperparameter of the chosen classifier to improve the prediction accuracy
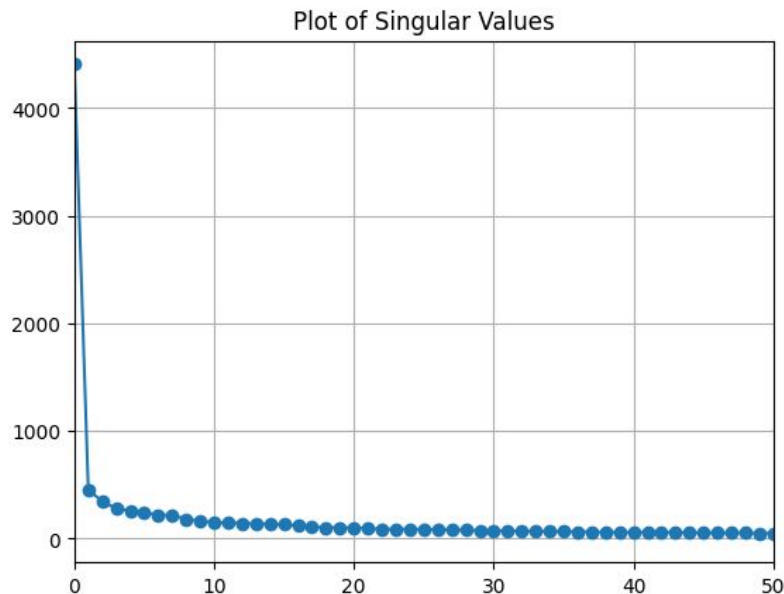
Approach

$$\mathbf{M}_{m \times n} = \mathbf{U}_{m \times m} \; \Sigma_{m \times n} \; \mathbf{V}^*_{n \times n}$$

$$X_{reduced} = X V_n$$

$n =$ number of principal components selected

Approach

Plot of Singular Values

$$X_{reduced} = XV_{20}$$

12,286 features → 20 features

Results

Classification Method	Accuracy	AUC
Logistic Regression	92.64%	0.9718
Random Forest	93.02%	0.9731
KNN	92.40%	0.9555
SVM	93.07%	0.9756

Approach

Model chosen for tuning: SVM

- Hyperparameter chosen: Kernel type

- Kernel used in preliminary model:

  - RBF

- Additional kernels to investigate:

  - Linear

  - Polynomial

  - Sigmoid

Input Space   Feature Space
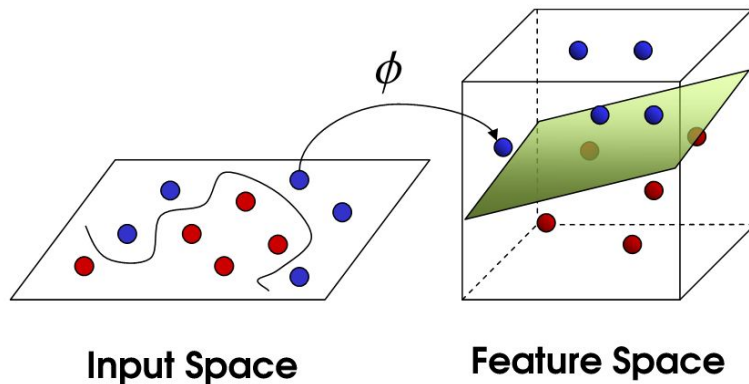
Results

Kernel Method	Accuracy	AUC
RBF	93.07%	0.9756
Linear	92.74%	0.9717
Polynomial	92.57%	0.9704
Sigmoid	78.69%	0.7981

Discussion

Principal Findings

- Simple classification results were very promising

  - All had accuracy within the 92-93% range

- The SVM classifier outperformed other models

- RBF kernel performance:

  - Accuracy: 93.07%

  - AUC: 0.9756

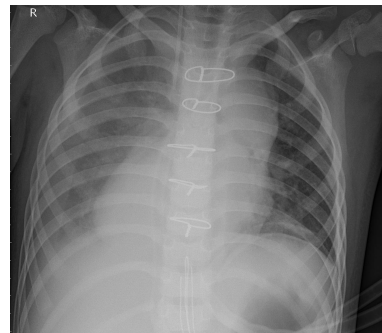# Discussion

Implications

- Performance is promising, but not good enough for confident diagnoses

    - Provide **recommendations** rather than concrete diagnoses

- This tool could be implemented into an x-ray image storage system

- Could also be a standalone tool

    - Extra steps, would likely see low usage in real-time care

Discussion

Limitations

- All images used are different sizes, lungs are positioned in different angles

    - Extra data manipulation would improve accuracy

- Choice of 64 x 64 image processing

- Only the most promising simple classification technique was fine-tuned

    - Future work could investigate whether other models are improved by fine-tuning

- Unexplored classification models (deep learning models)

References

https://www.thoracic.org/patients/patient-resources/resources/top-pneumonia-facts.pdf

https://www.clinicbarcelona.org/en/news/pneumonia-causes-2-5-million-deaths-around-the-world-each-year#:~:text=Globally%2C%20only%20half%20of%20children,under%205%20years%20of%20age

http://proxy.library.vanderbilt.edu/login?url=https://www.proquest.com/scholarly-journals/novel-fast-pathogen-diagnosis-method-severe/docview/2730538532/se-2?accountid=14816

https://link.gale.com/apps/doc/A444844867/AONE?u=tel_a_vanderbilt&sid=bookmark-AONE&xid=a47c262a

http://proxy.library.vanderbilt.edu/login?url=https://www.proquest.com/scholarly-journals/real-world-integration-sepsis-deep-learning/docview/2511970280/se-2?accountid=14816

https://data.mendeley.com/datasets/rscbjbr9sj/2