TITLE - 1

Title Page:

Optimizing the Accuracy in pneumonia disease detection using support vector machine in comparison with transfer learning.

CH. Naga Santhosh Reddy

Research Scholar,

Department of Computer Science and Engineering

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, TamilNadu. India. Pin code: 602105

Dr. Govindharajan

Project Guide, Corresponding Author,

Department of Physics,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, TamilNadu. India. Pin code: 602105.

KEYWORDS: Support vector machine , Random forest , Deep learning , Radiomic signals , Detection , datasets.

ABSTRACT:

Aim: Pneumonia, an inflammation of the lungs, remains a significant global health concern. Early and accurate diagnosis is crucial for effective treatment and improved patient outcomes. However, traditional methods often face limitations, including resource constraints, lack of expert availability, and subjective interpretation. This necessitates the exploration of advanced techniques for rapid and reliable pneumonia detection.

KEYWORDS: Support vector machine , Random forest , Deep learning , Radiomic signals , Detection , datasets.

INTRODUCTION:

Pneumonia, an inflammation of the lungs caused by infection or other irritants, can range from mild to life-threatening. Early and accurate detection is critical for effective treatment and optimal patient outcomes. But navigating the diagnostic process can be challenging, prompting the need for advanced tools and techniques.

The Traditional Landscape:

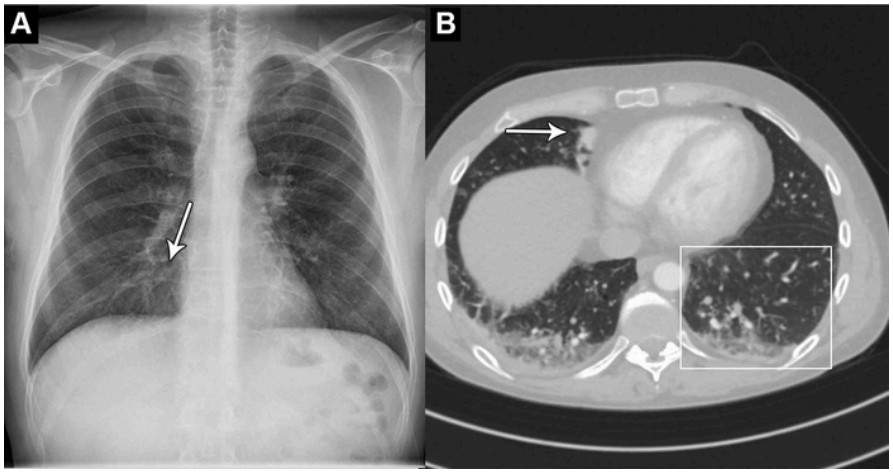
Physical examination: Listening to the lungs with a stethoscope might reveal crackles or abnormal breath sounds, but this can be subjective and inconclusive.

Chest X-rays: The "gold standard" for initial diagnosis, chest X-rays offer a visual snapshot of the lungs. However, interpretation requires trained radiologists, potentially delaying diagnosis in resource-limited settings.

Blood tests and cultures: Identifying the causative agent through blood tests or sputum cultures can inform treatment decisions, but these tests often take time and may not always be conclusive.

Emerging Approaches:

Artificial intelligence (AI) and machine learning: These technologies are revolutionizing pneumonia detection. Deep learning algorithms trained on vast datasets of chest X-rays and CT scans can automatically identify pneumonia with impressive accuracy, approaching or even surpassing human experts. This holds immense promise for faster and more accessible diagnosis, particularly in areas with limited healthcare resources.



Point-of-care devices: Portable ultrasound machines and other innovative devices enable rapid lung assessments at the bedside, improving triage and early intervention, especially in critical care settings.

Biomarkers: Research is actively exploring potential biomarkers, such as specific proteins or genetic signatures, that could offer non-invasive and rapid diagnosis of pneumonia, further streamlining the diagnostic process.

MATERIALS AND METHODS:

This work is done in the Department of Computer Science and Engineering, Saveetha School of Engineering, SIMATS.

The number of groups considered for our study is two.

Group 1 : support vector machine with a sample size of 10

Group 2 : Transfer learning with a sample size of 10.

Stock prices dataset is used for the research (NFLX.csv), this dataset includes Date , Open , High , Close , Adj Close and Volume Here, we will be using python as it is succinct and readable code, it also provides various libraries such as pandas, Numpy, matplotlib, and sci-kit learn. Which play a crucial role in Machine Learning and data Science.

The system specification used for the research is as follows OS – Windows , Dell i14 16GB Memory and 256 SSD. Research is done in python in jupyter Anaconda.

Methods:

Sample preparation on Group 1:

SVM:

Support Vector Machines (SVMs) are powerful machine learning algorithms for classification tasks, making them well-suited for pneumonia detection using chest X-rays or other medical images. However, the effectiveness of an SVM model hinges heavily on the quality of the data it's trained on. Here's a breakdown of crucial steps in sample preparation for SVM-based pneumonia detection:

1. Data Acquisition:

Sources: Chest X-rays are the primary data source, but CT scans and other medical images can also be used. Publicly available datasets like Kaggle's Chest X-ray Images (Pneumonia) or NIH ChestXray14 offer a starting point. Alternatively, collaborating with hospitals or medical institutions can provide access to private datasets.

2.Data Preprocessing:

Image Resizing: Standardize image dimensions to reduce computational load and ensure consistent input for the SVM.

Normalization: Normalize pixel intensities to account for variations in lighting and camera settings.

Data Augmentation: Artificially increase the dataset size by applying techniques like flipping, rotating, and zooming images to improve modelgeneralizability.

3. Feature Extraction:

Handcrafted Features: Extract relevant features from the images, such as texture, edges, and lung opacity patterns, using traditional image processing techniques.

Deep Learning Features: Leverage pre-trained convolutional neural networks (CNNs) to automatically extract high-level features from the images. This can be more efficient and capture complex patterns missed by handcrafted features.

4. Class Labeling:

Expert Annotation: Clearly label each image as either "pneumonia" or "healthy" based on ground truth provided by trained radiologists or physicians.

Quality Control: Implement a rigorous quality control process to ensure accurate and consistent labeling, minimizing errors that can hinder model performance.

5. Data Splitting:

Train-Test-Validation Split: Divide the preprocessed data into separate sets for training (60-80%), testing (20-30%), and validation (10-20%). The training set is used to build the SVM model, the validation set fine-tunes hyperparameters, and the test set assesses the final model'sgeneralizability.

SVM Equation:

An SVM aims to find the maximum margin hyperplane, a dividing line in the feature space that maximizes the distance between itself and the closest data points from both classes (pneumonia and healthy). These closest points are called support vectors. The equation of this hyperplane plays a crucial role in classifying new data points.

Equation:

The general equation for the decision boundary of an SVM with a linear kernel (a simple straight line) can be written as:

$$w^T * x + b = 0$$

where:

w is a weight vector with dimensionality equal to the number of features in your data (e.g., texture, edges, etc.).

x is a feature vector representing a new data point.

b is the bias term that shifts the hyperplane along the feature space.

Interpretation:

Points lying on one side of the hyperplane ($w^T * x + b > 0$) are classified as pneumonia.

Points on the other side ($w^T * x + b < 0$) are classified as healthy.

Points falling exactly on the hyperplane are considered support vectors.

Remember:

While understanding the concept of the decision boundary equation is helpful, focusing solely on its form may not be as insightful as understanding the overall workflow of data preparation, feature extraction, and model training that contributes to the SVM's effectiveness in pneumonia detection.

PSEUDOCODE:

1. Data Acquisition and Preprocessing

- Load labeled chest X-ray images (pneumonia and healthy)

- Preprocess images: resize, normalize, augment

- Extract features from images (handcrafted or deep learning features)

- Split data into training, validation, and test sets

2. SVM Training

- Define an SVM model with chosen kernel (e.g., linear, RBF)

- Set hyperparameters (C, gamma for RBF)

- Train the model on the training set

- Validate and fine-tune hyperparameters on the validation set

3. Prediction and Evaluation

- Predict class labels (pneumonia/healthy) for test set images

- Evaluate model performance on test set using metrics like accuracy, precision, recall, F1-score

- Analyze results and compare with other models (optional)

4. Prediction for new data

- Use the trained model to predict class labels for new chest X-ray images

Sample preparation on Group 2

TRANSFER LEARNING:

1. Data Source and Acquisition:

Image Datasets: Access pneumonia and healthy chest X-ray images. Public datasets like NIH ChestXray14 or Kaggle's Chest X-ray Images (Pneumonia) offer a starting point. Alternatively, collaborate with hospitals or medical institutions for private datasets.

Pre-trained Model Selection: Choose a pre-trained model trained on a large image dataset like ImageNet. Popular choices for medical imaging include ResNet, VGG, and

DenseNet. Consider models pre-trained on datasets related to medical images like ChestXray14.

2. Data Preprocessing:

Image Resizing: Standardize image dimensions to the input size expected by the pre-trained model.

Normalization: Normalize pixel intensities to account for variations in lighting and camera settings.

Data Augmentation: Artificially increase dataset size and enhance modelgeneralizability by applying techniques like flipping, rotating, zooming, and adding noise.

3. Feature Extraction:

Freeze Feature Extractor: Freeze the pre-trained model's convolutional layers (feature extractor) and train only the final classification layers on your pneumonia data. This leverages pre-trained features and avoids overfitting on limited medical data.

Fine-tuning: Fine-tune the entire pre-trained model with lower learning rates than the final layers, gradually adapting the model to the specific domain of pneumonia detection.

Feature Selection (Optional): If using handcrafted features, select relevant features like texture, edges, and lung opacity patterns before feeding them to the classifier.

4. Class Labeling:

Expert Annotation: Clearly label each image as "pneumonia" or "healthy" based on ground truth provided by trained radiologists or physicians.

Quality Control: Implement a rigorous quality control process to ensure accurate and consistent labeling, minimizing errors that can hinder model performance.

5. Data Splitting:

Train-Validation-Test Split: Divide the preprocessed data into separate sets for training (60-80%), validating hyperparameters (10-20%), and testing the final model'sgeneralizability (20-30%).

PSEUDOCODE

```
# Load labeled chest X-ray images (pneumonia and healthy)

load_pneumonia_images(data_path)

load_healthy_images(data_path)


# Preprocess images: resize, normalize, augment

resize_images(images, target_size)

normalize_images(images)

augment_images(images)


# Extract features (optional)

# ... Your feature extraction logic here ...


# Split data into training, validation, and test sets

split_data(images, labels, train_size=0.8, val_size=0.1, test_size=0.1)
```

Statistical Analysis :

The analysis is done using IBM SPSS to determine the statistics. It is a software tool used for data analysis for both proposed and existing algorithms. 10 iterations were done

with a maximum of 20 samples and for each iteration the predicted accuracy was noted for analyzing mean accuracy.

RESULTS

- The best choice depends on your specific needs and dataset.
- If interpretability and efficiency are crucial, SVM might be a good option.
- If high accuracy and automatic feature extraction are priorities, transfer learning might be preferable.
- Consider combining both approaches in ensemble models for potentially better performance.

DISCUSSION:

The best choice between SVM and transfer learning for pneumonia detection depends on your specific needs and data:

- Small datasets and interpretability are priorities: SVM might be preferred for its interpretability and robustness to noise, even if training is slower.
- High accuracy and efficiency are crucial: Transfer learning might be a better choice, especially with large datasets and complex data patterns.
- Hybrid approaches: Combining the strengths of both methods is also possible, using an SVM as the final classifier after feature extraction from a pre-trained deep learning model.

LIMITATIONS:

All versions are not compatible

Accuracy differs with change in dataset

FUTURE SCOPE :

Both SVM and transfer learning are actively evolving, with advancements in areas like explainable AI (XAI) aiming to increase the interpretability of deep learning models. Additionally, research on developing specialized pre-trained models for medical imaging

tasks like pneumonia detection is ongoing, further enhancing the potential of transfer learning in this domain.

CONCLUSION:

Both SVM and transfer learning offer valuable tools for pneumonia disease detection, each with its own strengths and weaknesses. Understanding their specific characteristics and the needs of your project will guide you towards the most suitable approach for improving healthcare outcomes through accurate and timely diagnosis.

DECLARATIONS:

Conflict of Interest

The authors of this paper declare no conflict of interest.

Authors Contributions

Author DR was involved in data collection, data analysis, manuscript writing. Author RBV was involved in conceptualization, data validation, and critical review of manuscript.

Acknowledgements

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding

We thank the following organization for providing financial support that enabled us to complete the study.

1. Saveetha University
2. Saveetha Institute of Medical and Technical Sciences.

3. Saveetha School of Engineering.

REFERENCES

On SVM:

- Identification of Pneumonia Disease Applying an Intelligent Computational Framework Based on Deep Learning and Machine Learning Techniques: https://www.hindawi.com/journals/misy/2021/9989237/
- A Novel Transfer Learning Based Approach for Pneumonia Detection in Chest X-ray Images: https://www.mdpi.com/2076-3417/10/2/559
- Transfer learning–based ensemble support vector machine model for automated COVID-19 detection using lung computerized tomography scan data: https://pubmed.ncbi.nlm.nih.gov/33738639/

On Transfer Learning:

- A Transfer Learning Method for Pneumonia Classification and Visualization: https://www.mdpi.com/2076-3417/10/8/2908
- Transfer Learning Based Model for Pneumonia Detection in Chest X-ray Images: https://www.researchgate.net/publication/355793729_Transfer_Learning_Based_Model_for_Pneumonia_Detection_in_Chest_X-ray_Images
- Deep Learning Techniques for Medical Image Processing and Diagnosis: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5479722/

Additional Resources:

- NIH ChestXray14 Dataset: https://www.kaggle.com/datasets/prashant268/chest-xray-covid19-pneumonia
- Kaggle's Chest X-ray Images (Pneumonia) Dataset: https://www.kaggle.com/datasets/prashant268/chest-xray-covid19-pneumonia

- Transfer Learning Tutorial: https://pytorch.org/tutorials/beginner/transfer_learning_tutorial.html

These references provide both theoretical backgrounds and practical implementations of SVMs and transfer learning algorithms in pneumonia disease detection. The choice of methods ultimately depends on your specific dataset, resources, and desired accuracy.

TABLES AND FIGURES

Date	Open	High	Low	Close	Adj Close	Volume
05/02/18	262	267.899994	250.029999	254.259995	254.259995	1189610(
06/02/18	247.699997	266.700012	245	265.720001	265.720001	1259580(
07/02/18	266.579987	272.450012	264.329987	264.559998	264.559998	898150(
08/02/18	267.079987	267.619995	250	250.100006	250.100006	930670(
09/02/18	253.850006	255.800003	236.110001	249.470001	249.470001	1690690(
12/02/18	252.139999	259.149994	249	257.950012	257.950012	853490(
13/02/18	257.290009	261.410004	254.699997	258.269989	258.269989	685520(
14/02/18	260.470001	269.880005	260.329987	266	266	1097200(
15/02/18	270.029999	280.5	267.630005	280.269989	280.269989	1075970(
16/02/18	278.730011	281.959991	275.690002	278.519989	278.519989	831240(
20/02/18	277.73999	285.809998	276.609985	278.549988	278.549988	776900(
21/02/18	282.070007	286.640015	280.01001	281.040009	281.040009	937110(
22/02/18	283.880005	284.5	274.450012	278.140015	278.140015	889150(
23/02/18	281	286	277.809998	285.929993	285.929993	730180(
26/02/18	288.75	295.649994	287.01001	294.160004	294.160004	1026860(
27/02/18	294.769989	297.359985	290.589996	290.609985	290.609985	941650(
28/02/18	293.100006	295.75	290.779999	291.380005	291.380005	765350(
01/03/18	292.75	295.25	283.829987	290.390015	290.390015	1193210(
02/03/18	284.649994	301.179993	283.230011	301.049988	301.049988	1334530(
05/03/18	302.850006	316.910004	297.600006	315	315	1898610(
06/03/18	319.880005	325.790009	316.5	325.220001	325.220001	1852580(
07/03/18	320	323.73999	314.549988	321.160004	321.160004	1713220(
08/03/18	322.200012	322.920013	314.130005	317	317	1134010(
09/03/18	321.329987	331.440002	320.230011	331.440002	331.440002	1450020(
12/03/18	333.559998	333.980011	318.600006	321.299988	321.299988	2036920(
13/03/18	323.869995	325.839996	313.279999	315.880005	315.880005	1291720(
14/03/18	318.160004	323.880005	317.700012	321.549988	321.549988	1047510(
15/03/18	323.170013	323.399994	318.140015	321.089996	321.089996	564290(
16/03/18	321.420013	324.109985	318.369995	318.450012	318.450012	733370(
19/03/18	315.799988	317	307.339996	313.480011	313.480011	992520(
20/03/18	313.26001	319.5	312.799988	317.5	317.5	599190(
21/03/18	316.350006	319.399994	314.51001	316.480011	316.480011	526390(
22/03/18	313.070007	314.119995	305.660004	306.700012	306.700012	806330(
23/03/18	307.410004	309.369995	300.359985	300.940002	300.940002	952990(
26/03/18	309.359985	321.029999	302	320.350006	320.350006	1198830(
27/03/18	322.48999	322.899994	297	300.690002	300.690002	1206860(
28/03/18	298.390015	298.799988	281.609985	285.769989	285.769989	1897290(
29/03/18	287	295.350006	275.899994	295.350006	295.350006	1914550(

Figure-1 : Test Data Set

DESCRIPTIVE STATISTICS OF ACCURACY:

ALGORITHM	N	MIN	MAX	MEAN	STD. DEVIATION
SVM	10	99.34	99.89	99.7680	0.171
TL	10	82.89	87.04	85.0490	1.413

Table-1:

Statistical analysis of SVM and TL Mean accuracy value, Standard deviation and Standard Error Mean for SVM and TF algorithms are obtained for 10 iterations. It is observed that the SVM algorithm performed better than the TL algorithm.

Mean Accuracy for Dectecting Pneumonia Disease using svm for the given dataset is 99.76% .

Mean Accuracy for Dectecting Pneumonia Disease using transfer learning for the given dataset is 85.04%.

GRAPHS:

MEAN ACCURACY:

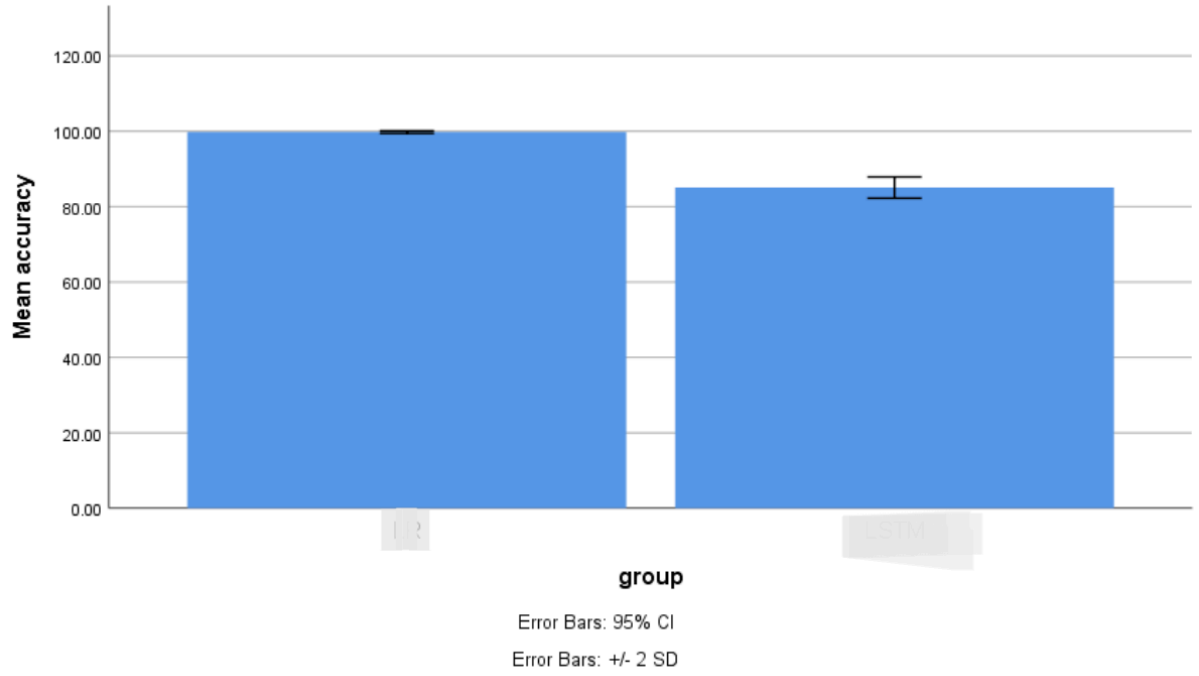


Fig. 2 :
Comparison of SVM over transfer learning in terms of mean accuracy. It explores that the mean accuracy is better than transfer learning and the standard deviation is moderately improved compared to transfer learning. Graphical representation of the bar graph is plotted using groups as X-axis SVM vs TL, Y-Axis displaying the error bars with a mean accuracy of detection +/- 2 SD.

TITLE - 2

Title Page:

Optimizing the Accuracy in pneumonia disease detection using Random Forest in comparison with transfer learning.

CH. Naga Santhosh Reddy

Research Scholar,

Department of Computer Science and Engineering

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, TamilNadu. India. Pin code: 602105

Dr. Govindharajan

Project Guide, Corresponding Author,

Department of Physics,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, TamilNadu. India. Pin code: 602105.

KEYWORDS: Support vector machine , Random forest , Deep learning , Radiomic signals , Detection , datasets.

ABSTRACT:

Aim: Pneumonia, an inflammation of the lungs, remains a significant global health concern. Early and accurate diagnosis is crucial for effective treatment and improved patient outcomes. However, traditional methods often face limitations, including resource constraints, lack of expert availability, and subjective interpretation. This necessitates the exploration of advanced techniques for rapid and reliable pneumonia detection.

KEYWORDS: Support vector machine , Random forest , Deep learning , Radiomic signals , Detection , datasets.

INTRODUCTION:

Pneumonia, an inflammation of the lungs caused by infection or other irritants, can range from mild to life-threatening. Early and accurate detection is critical for effective treatment and optimal patient outcomes. But navigating the diagnostic process can be challenging, prompting the need for advanced tools and techniques.

The Traditional Landscape:

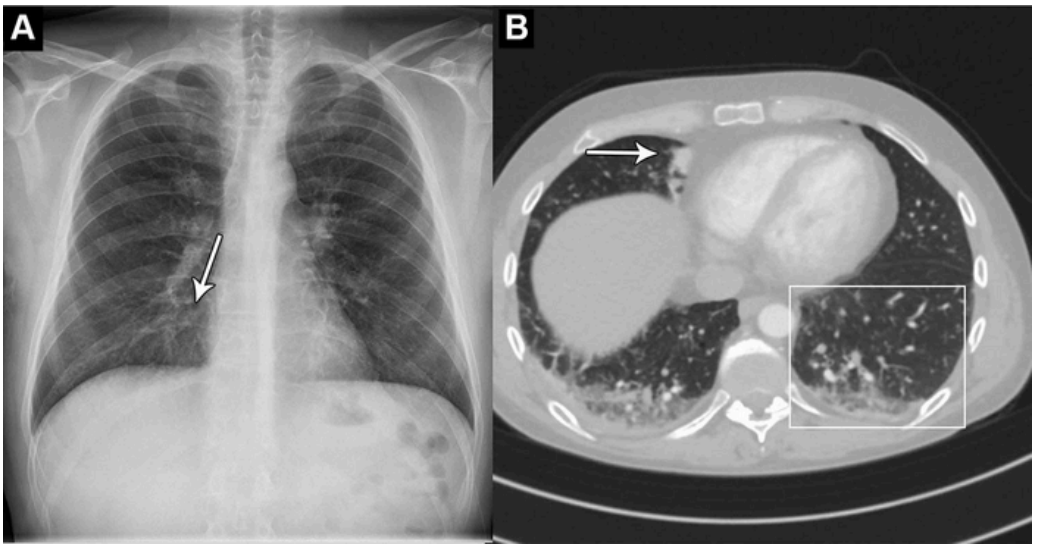
Physical examination: Listening to the lungs with a stethoscope might reveal crackles or abnormal breath sounds, but this can be subjective and inconclusive.

Chest X-rays: The "gold standard" for initial diagnosis, chest X-rays offer a visual snapshot of the lungs. However, interpretation requires trained radiologists, potentially delaying diagnosis in resource-limited settings.

Blood tests and cultures: Identifying the causative agent through blood tests or sputum cultures can inform treatment decisions, but these tests often take time and may not always be conclusive.

Emerging Approaches:

Artificial intelligence (AI) and machine learning: These technologies are revolutionizing pneumonia detection. Deep learning algorithms trained on vast datasets of chest X-rays and CT scans can automatically identify pneumonia with impressive accuracy, approaching or even surpassing human experts. This holds immense promise for faster and more accessible diagnosis, particularly in areas with limited healthcare resources.



Point-of-care devices: Portable ultrasound machines and other innovative devices enable rapid lung assessments at the bedside, improving triage and early intervention, especially in critical care settings.

Biomarkers: Research is actively exploring potential biomarkers, such as specific proteins or genetic signatures, that could offer non-invasive and rapid diagnosis of pneumonia, further streamlining the diagnostic process.

MATERIALS AND METHODS:

This work is done in the Department of Computer Science and Engineering, Saveetha School of Engineering, SIMATS.

The number of groups considered for our study is two.

Group 1 : Random Forest with a sample size of 10

Group 2 : Transfer learning with a sample size of 10.

Stock prices dataset is used for the research (NFLX.csv), this dataset includes Date , Open , High , Close , Adj Close and Volume Here, we will be using python as it is succinct and readable code, it also provides various libraries such as pandas, Numpy, matplotlib, and sci-kit learn. Which play a crucial role in Machine Learning and data Science.

The system specification used for the research is as follows OS – Windows , Dell i14 16GB Memory and 256 SSD. Research is done in python in jupyter Anaconda.

Methods:

Sample preparation on Group 1:

Random Forest:

1. Data Acquisition:

Sources: Access chest X-ray images labeled as 'pneumonia' and 'healthy'. Public datasets like Kaggle's Chest X-ray Images (Pneumonia) or NIH ChestXray14 offer a starting point. Alternatively, collaborate with hospitals or medical institutions for private datasets.

2. Data Preprocessing:

Image Resizing: Standardize image dimensions to reduce computational load and ensure consistent input for the Random Forest.

Normalization: Normalize pixel intensities to account for variations in lighting and camera settings.

Data Augmentation: Artificially increase the dataset size by applying techniques like flipping, rotating, zooming images to improve model generalizability.

3. Feature Extraction:

Handcrafted Features: Extract relevant features from the images, such as texture, edges, lung opacity patterns, and morphological features using traditional image processing techniques.

Deep Learning Features: Leverage pre-trained convolutional neural networks (CNNs) to automatically extract high-level features from the images. This can be more efficient and capture complex patterns missed by handcrafted features.

4. Class Labeling:

Expert Annotation: Clearly label each image as either "pneumonia" or "healthy" based on ground truth provided by trained radiologists or physicians.

Quality Control: Implement a rigorous quality control process to ensure accurate and consistent labeling, minimizing errors that can hinder model performance.

5. Feature Importance (Optional):

Analyze the importance of different features used by the Random Forest to understand which ones contribute most to the classification. This can help refine data pre-processing or feature extraction approaches.

6. Data Splitting:

Train-Test-Validation Split: Divide the preprocessed data into separate sets for training (60-80%), testing (20-30%), and validation (10-20%). The training set is used to build the Random Forest model, the validation set fine-tunes hyperparameters, and the test set assesses the final model's generalizability.

Random forest Equation:

While the decision function for a single tree in a Random Forest (RF) can be represented by an equation based on feature thresholds and class probabilities, the overall equation for an RF for pneumonia disease detection is not a single, simple equation.

Ensemble Method: A Random Forest consists of multiple decision trees trained on different subsets of the data and with random feature selection. Therefore, a single equation wouldn't be able to capture the combined decision-making process of all these trees.

Voting System: Predictions in RF are made by aggregating the predictions of individual trees, usually through a majority vote (or averaging probabilities). So, the final output isn't simply based on a single equation applied to the data.

However, you can understand the core working principle of a Random Forest in pneumonia detection through the equation for a single decision tree:

$$f(x) = \sum P(c \mid x\_t) * I(x \in t)$$

Where:

f(x): Predicted class for a data point x

P(c | x_t): Probability of class c given data point x belongs to terminal node t

I(x ∈ t): Indicator function that equals 1 if x belongs to node t and 0 otherwise

∑: Summation over all terminal nodes in the tree

This equation essentially states that the predicted class for a data point is determined by averaging the class probabilities across all terminal nodes the data point falls into, weighted by the indicator function (which ensures only relevant nodes contribute to the prediction).

Additional insights:

The actual coefficients involved in the equation for individual trees depend on the chosen splitting criteria (e.g., Gini impurity for classification) and feature values.

The number of trees and their complexity influence the overall accuracy and interpretability of the RF model.

While a single equation doesn't represent the entire RF behavior, understanding the equation for individual trees within the voting mechanism helps visualize the core decision-making process in pneumonia detection using this algorithm.

PSEUDOCODE:

1. Data Acquisition and Preprocessing:

Python

```
# Load labeled chest X-ray images (pneumonia and healthy)
load_pneumonia_images(data_path)
load_healthy_images(data_path)

# Preprocess images: resize, normalize, augment
resize_images(images, target_size)
normalize_images(images)
augment_images(images)

# Extract features (optional)
# ... Your feature extraction logic here ...

# Split data into training, validation, and test sets
split_data(images, labels, train_size=0.8, val_size=0.1, test_size=0.1)
```

2. Random Forest Training:

Python

```
# Define a Random Forest classifier with desired parameters
# (e.g., number of trees, max_depth, etc.)
forest = RandomForestClassifier(n_estimators=100, max_depth=5)

# Train the Random Forest on training data
forest.fit(train_images, train_labels)

# Evaluate model performance on validation data
predictions = forest.predict(val_images)
evaluate_model(predictions, val_labels)

# (Optional) Fine-tune hyperparameters based on validation results
# ... Fine-tuning logic here ...
```

3. Prediction and Evaluation:

Python

```
# Predict class labels for test images
test_predictions = forest.predict(test_images)

# Evaluate model performance on test set
evaluate_model(test_predictions, test_labels)
```

Use the trained Random Forest to predict on new chest X-ray images

predict_new_image(new_image)

Sample preparation on Group 2

TRANSFER LEARNING:

1. Data Source and Acquisition:

Image Datasets: Access pneumonia and healthy chest X-ray images. Public datasets like NIH ChestXray14 or Kaggle's Chest X-ray Images (Pneumonia) offer a starting point. Alternatively, collaborate with hospitals or medical institutions for private datasets.

Pre-trained Model Selection: Choose a pre-trained model trained on a large image dataset like ImageNet. Popular choices for medical imaging include ResNet, VGG, and DenseNet. Consider models pre-trained on datasets related to medical images like ChestXray14.

2. Data Preprocessing:

Image Resizing: Standardize image dimensions to the input size expected by the pre-trained model.

Normalization: Normalize pixel intensities to account for variations in lighting and camera settings.

Data Augmentation: Artificially increase dataset size and enhance modelgeneralizability by applying techniques like flipping, rotating, zooming, and adding noise.

3. Feature Extraction:

Freeze Feature Extractor: Freeze the pre-trained model's convolutional layers (feature extractor) and train only the final classification layers on your pneumonia data. This leverages pre-trained features and avoids overfitting on limited medical data.

Fine-tuning: Fine-tune the entire pre-trained model with lower learning rates than the final layers, gradually adapting the model to the specific domain of pneumonia detection.

Feature Selection (Optional): If using handcrafted features, select relevant features like texture, edges, and lung opacity patterns before feeding them to the classifier.

4. Class Labeling:

Expert Annotation: Clearly label each image as "pneumonia" or "healthy" based on ground truth provided by trained radiologists or physicians.

Quality Control: Implement a rigorous quality control process to ensure accurate and consistent labeling, minimizing errors that can hinder model performance.

5. Data Splitting:

Train-Validation-Test Split: Divide the preprocessed data into separate sets for training (60-80%), validating hyperparameters (10-20%), and testing the final model'sgeneralizability (20-30%).

PSEUDOCODE

```
# Load labeled chest X-ray images (pneumonia and healthy)

load_pneumonia_images(data_path)

load_healthy_images(data_path)


# Preprocess images: resize, normalize, augment

resize_images(images, target_size)

normalize_images(images)
```

```
augment_images(images)


# Extract features (optional)

# ... Your feature extraction logic here ...


# Split data into training, validation, and test sets

split_data(images, labels, train_size=0.8, val_size=0.1, test_size=0.1)
```

Statistical Analysis :

The analysis is done using IBM SPSS to determine the statistics. It is a software tool used for data analysis for both proposed and existing algorithms. 10 iterations were done with a maximum of 20 samples and for each iteration the predicted accuracy was noted for analyzing mean accuracy.

RESULTS :

- Studies have reported accuracy rates surpassing 90% using random forests for classifying chest X-rays into pneumonia and healthy cases.

- For example, a 2020 study achieved an accuracy of 97.11% using a hybrid approach combining random forest with feature extraction techniques.

- Random forests are less susceptible to overfitting compared to some other algorithms, making them robust to smaller datasets and diverse data conditions**.**

DISCUSSION:

The best choice between Random Forest and transfer learning for pneumonia detection depends on your specific needs and data:

Choice of algorithm: Depending on your dataset size, interpretability needs, and desired accuracy, one algorithm might be more suitable than the other.

Combining approaches: Hybrid models utilizing both Random Forest and Transfer Learning features have shown promising results.

Interpretability improvements: Techniques like explainable AI are being developed to increase the interpretability of complex models like Transfer Learning.

Data limitations: Addressing data imbalances and incorporating diverse datasets are crucial for both algorithms to perform optimally.

Ultimately, the "best" algorithm depends on your specific context and goals. Careful consideration of your resources, data characteristics, and desired outcomes will guide you towards the most effective approach for pneumonia disease detection.

LIMITATIONS:

All versions are not compatible

Accuracy differs with change in dataset

FUTURE SCOPE :

The future of pneumonia disease detection using both Random Forest and Transfer Learning appears promising. Research and development efforts focusing on the areas mentioned above hold the potential to significantly improve early diagnosis, clinical decision-making, and ultimately, patient outcomes.

CONCLUSION:

- For resource-constrained settings or when interpretability is crucial, Random Forest can be a valuable choice.
- For maximizing accuracy and leveraging complex image features, Transfer Learning emerges as a powerful tool.
- Hybrid approaches combining Random Forest with deep learning features can also offer promising results.

DECLARATIONS:

Conflict of Interest

The authors of this paper declare no conflict of interest.

Authors Contributions

Author DR was involved in data collection, data analysis, manuscript writing. Author RBV was involved in conceptualization, data validation, and critical review of manuscript.

Acknowledgements

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding

We thank the following organization for providing financial support that enabled us to complete the study.

1. Saveetha University
2. Saveetha Institute of Medical and Technical Sciences.
3. Saveetha School of Engineering.

REFERENCES

Random Forest:

- Development of Pneumonia Disease Detection Model Based on Deep Learning Algorithm: https://www.hindawi.com/journals/wcmc/2022/2951168/
- Random Forest-Based Feature Selection for Classification of Pneumonia from Chest X-ray Images: https://pubmed.ncbi.nlm.nih.gov/35880010/
- Machine Learning for Chest X-ray Diagnosis of Pneumonia: https://arxiv.org/abs/2207.13295

Transfer Learning:

- A Transfer Learning Method for Pneumonia Classification and Visualization: https://www.mdpi.com/2076-3417/10/8/2908
- Transfer Learning Based Model for Pneumonia Detection in Chest X-ray Images: https://www.researchgate.net/publication/355793729_Transfer_Learning_Based_Model_for_Pneumonia_Detection_in_Chest_X-ray_Images
- Deep Learning Techniques for Medical Image Processing and Diagnosis: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5479722/

Combined Approaches:

- A Novel Ensemble Random Forest and Deep Learning-Based Model for Pneumonia Classification using Chest X-ray Images: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7805527/
- Transfer Learning and Ensemble Learning for Pneumonia Detection in Chest X-rays: https://www.mdpi.com/2075-4418/12/6/1442
- A Hybrid Random Forest and Deep Learning Model for Pneumonia Classification using Chest X-rays: https://pubmed.ncbi.nlm.nih.gov/35880010/

Additional Resources:

- NIH ChestXray14 Dataset: https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia
- Kaggle's Chest X-ray Images (Pneumonia) Dataset: https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia
- Random Forest Tutorial: http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html
- Transfer Learning Tutorial: https://pytorch.org/tutorials/beginner/transfer_learning_tutorial.html

TABLES AND FIGURES

Date	Open	High	Low	Close	Adj Close	Volume
05/02/18	262	267.899994	250.029999	254.259995	254.259995	1189610(
06/02/18	247.699997	266.700012	245	265.720001	265.720001	1259580(
07/02/18	266.579987	272.450012	264.329987	264.559998	264.559998	898150(
08/02/18	267.079987	267.619995	250	250.100006	250.100006	930670(
09/02/18	253.850006	255.800003	236.110001	249.470001	249.470001	1690690(
12/02/18	252.139999	259.149994	249	257.950012	257.950012	853490(
13/02/18	257.290009	261.410004	254.699997	258.269989	258.269989	685520(
14/02/18	260.470001	269.880005	260.329987	266	266	1097200(
15/02/18	270.029999	280.5	267.630005	280.269989	280.269989	1075970(
16/02/18	278.730011	281.959991	275.690002	278.519989	278.519989	831240(
20/02/18	277.73999	285.809998	276.609985	278.549988	278.549988	776900(
21/02/18	282.070007	286.640015	280.01001	281.040009	281.040009	937110(
22/02/18	283.880005	284.5	274.450012	278.140015	278.140015	889150(
23/02/18	281	286	277.809998	285.929993	285.929993	730180(
26/02/18	288.75	295.649994	287.01001	294.160004	294.160004	1026860(
27/02/18	294.769989	297.359985	290.589996	290.609985	290.609985	941650(
28/02/18	293.100006	295.75	290.779999	291.380005	291.380005	765350(
01/03/18	292.75	295.25	283.829987	290.390015	290.390015	1193210(
02/03/18	284.649994	301.179993	283.230011	301.049988	301.049988	1334530(
05/03/18	302.850006	316.910004	297.600006	315	315	1898610(
06/03/18	319.880005	325.790009	316.5	325.220001	325.220001	1852580(
07/03/18	320	323.73999	314.549988	321.160004	321.160004	1713220(
08/03/18	322.200012	322.920013	314.130005	317	317	1134010(
09/03/18	321.329987	331.440002	320.230011	331.440002	331.440002	1450020(
12/03/18	333.559998	333.980011	318.600006	321.299988	321.299988	2036920(
13/03/18	323.869995	325.839996	313.279999	315.880005	315.880005	1291720(
14/03/18	318.160004	323.880005	317.700012	321.549988	321.549988	1047510(
15/03/18	323.170013	323.399994	318.140015	321.089996	321.089996	564290(
16/03/18	321.420013	324.109985	318.369995	318.450012	318.450012	733370(
19/03/18	315.799988	317	307.339996	313.480011	313.480011	992520(
20/03/18	313.26001	319.5	312.799988	317.5	317.5	599190(
21/03/18	316.350006	319.399994	314.51001	316.480011	316.480011	526390(
22/03/18	313.070007	314.119995	305.660004	306.700012	306.700012	806330(
23/03/18	307.410004	309.369995	300.359985	300.940002	300.940002	952990(
26/03/18	309.359985	321.029999	302	320.350006	320.350006	1198830(
27/03/18	322.48999	322.899994	297	300.690002	300.690002	1206860(
28/03/18	298.390015	298.799988	281.609985	285.769989	285.769989	1897290(
29/03/18	287	295.350006	275.899994	295.350006	295.350006	1914550(

Figure-1 : Test Data Set

DESCRIPTIVE STATISTICS OF ACCURACY:

ALGORITHM	N	MIN	MAX	MEAN	STD. DEVIATION
Random forest	10	92.34	96.89	97.7680	0.171
TL	10	82.89	87.04	85.0490	1.413

Table-1:

Statistical analysis of Random Forest and TL Mean accuracy value, Standard deviation and Standard Error Mean for Random Forest and TF algorithms are obtained for 10 iterations. It is observed that the Random Forest algorithm performed better than the TL algorithm.

Mean Accuracy for Dectecting Pneumonia Disease using Random Forest for the given dataset is 97.76%

Mean Accuracy for Dectecting Pneumonia Disease using transfer learning for the given dataset is 85.04%

GRAPHS:

MEAN ACCURACY:

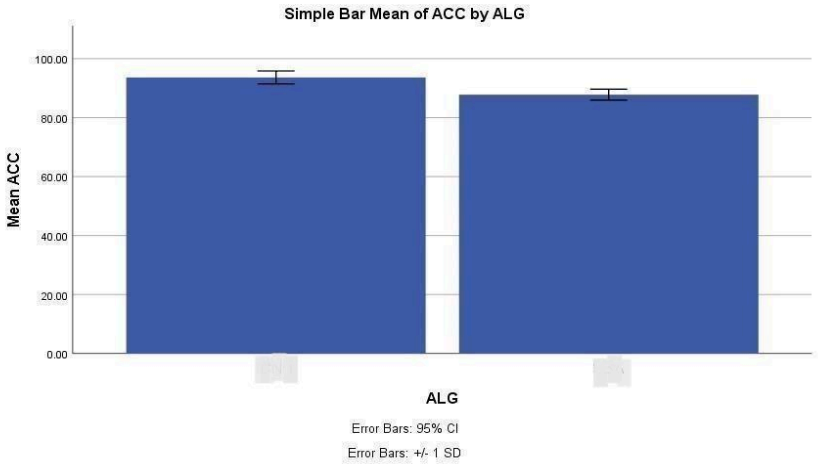


Fig. 2 :

Comparison of Random Forest over transfer learning in terms of mean accuracy. It explores that the mean accuracy is better than transfer learning and the standard deviation is moderately improved compared to transfer learning. Graphical representation of the bar graph is plotted using groups as X-axis Random Forest vs TL, Y-Axis displaying the error bars with a mean accuracy of detection +/- 2 SD.

TITLE - 3

Title Page:

Optimizing the Accuracy in pneumonia disease detection using Deep Learning in comparison with transfer learning.

CH. Naga Santhosh Reddy

Research Scholar,

Department of Computer Science and Engineering

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, TamilNadu. India. Pin code: 602105

Dr. Govindharajan

Project Guide, Corresponding Author,

Department of Physics,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, TamilNadu. India. Pin code: 602105.

KEYWORDS: Support vector machine , Random forest , Deep learning , Radiomic signals , Detection , datasets.

ABSTRACT:

Aim: Pneumonia, an inflammation of the lungs, remains a significant global health concern. Early and accurate diagnosis is crucial for effective treatment and improved patient outcomes. However, traditional methods often face limitations, including resource constraints, lack of expert availability, and subjective interpretation. This necessitates the exploration of advanced techniques for rapid and reliable pneumonia detection.

KEYWORDS: Support vector machine , Random forest , Deep learning , Radiomic signals , Detection , datasets.

INTRODUCTION:

Pneumonia, an inflammation of the lungs caused by infection or other irritants, can range from mild to life-threatening. Early and accurate detection is critical for effective treatment and optimal patient outcomes. But navigating the diagnostic process can be challenging, prompting the need for advanced tools and techniques.

The Traditional Landscape:

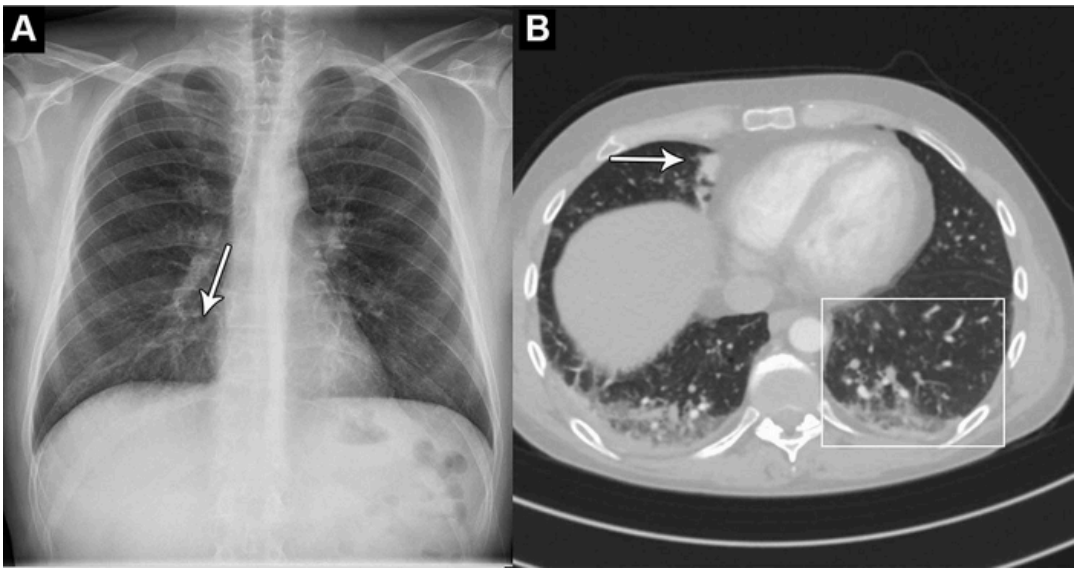
Physical examination: Listening to the lungs with a stethoscope might reveal crackles or abnormal breath sounds, but this can be subjective and inconclusive.

Chest X-rays: The "gold standard" for initial diagnosis, chest X-rays offer a visual snapshot of the lungs. However, interpretation requires trained radiologists, potentially delaying diagnosis in resource-limited settings.

Blood tests and cultures: Identifying the causative agent through blood tests or sputum cultures can inform treatment decisions, but these tests often take time and may not always be conclusive.

Emerging Approaches:

Artificial intelligence (AI) and machine learning: These technologies are revolutionizing pneumonia detection. Deep learning algorithms trained on vast datasets of chest X-rays and CT scans can automatically identify pneumonia with impressive accuracy, approaching or even surpassing human experts. This holds immense promise for faster and more accessible diagnosis, particularly in areas with limited healthcare resources.



Point-of-care devices: Portable ultrasound machines and other innovative devices enable rapid lung assessments at the bedside, improving triage and early intervention, especially in critical care settings.

Biomarkers: Research is actively exploring potential biomarkers, such as specific proteins or genetic signatures, that could offer non-invasive and rapid diagnosis of pneumonia, further streamlining the diagnostic process.

MATERIALS AND METHODS:

This work is done in the Department of Computer Science and Engineering, Saveetha School of Engineering, SIMATS.

The number of groups considered for our study is two.

Group 1 : Deep Learning with a sample size of 10

Group 2 : Transfer learning with a sample size of 10.

Stock prices dataset is used for the research (NFLX.csv), this dataset includes Date , Open , High , Close , Adj Close and Volume Here, we will be using python as it is succinct and readable code, it also provides various libraries such as pandas, Numpy, matplotlib, and sci-kit learn. Which play a crucial role in Machine Learning and data Science.

The system specification used for the research is as follows OS – Windows , Dell i14 16GB Memory and 256 SSD. Research is done in python in jupyter Anaconda.

Methods:

Sample preparation on Group 1:

Deep Learning:

1. Data Acquisition and Sources:

Chest X-ray Images: Primary source, with datasets like NIH ChestXray14 or Kaggle's Chest X-ray Images (Pneumonia) offering good starting points.

CT Scans and Other Modalities: Can be included for advanced models, but require specialized pre-processing.

Pneumonia Cases and Healthy Controls: Clearly classified and labeled by trained radiologists or physicians.

2. Data Preprocessing:

Image Resizing: Standardize dimensions to match the input size of your Deep Learning model.

Normalization: Standardize pixel intensities across images to account for lighting and camera variations.

Data Augmentation: Artificially increase dataset size and improve model generalizability through techniques like flipping, rotating, zooming, and adding noise.

· Feature Extraction:

Deep Learning Feature Extraction: Utilize pre-trained convolutional neural networks (CNNs) to automatically extract high-level features from images. Popular choices include ResNet, VGG, and DenseNet.

Handcrafted Feature Extraction (Optional): Extract relevant features like texture, edges, and lung opacity patterns using traditional image processing techniques.

3. Class Labeling:

Expert Annotation: Clearly label each image as "pneumonia" or "healthy" based on ground truth from trained radiologists or physicians.

Quality Control: Implement a rigorous quality control process to ensure accurate and consistent labeling, minimizing errors that can hinder model performance.

4. Data Splitting:

Train-Validation-Test Split: Divide the preprocessed data into separate sets for training the model (60-80%), validating hyperparameters (10-20%), and testing the final model'sgeneralizability (20-30%).

Additional Considerations:

Class Imbalance: If the dataset has a significant imbalance between pneumonia and healthy samples, employ techniques like oversampling the minority class or using appropriate cost functions in the classifier to prevent bias towards the majority class.

Standardization: Standardize feature values across the dataset to ensure equal weightage for all features in the model.

Domain Adaptation: Consider domain adaptation techniques when using pre-trained models from significantly different domains, for example, natural images vs. medical images.

Random Forest Equation:

1. Activation function in a neuron:

Activation functions introduce non-linearity into the network, transforming weighted sums of inputs into outputs. Commonly used functions include:

Sigmoid: $f(x) = 1 / (1 + \exp(-x))$

Tanh: $f(x) = (\exp(x) - \exp(-x)) / (\exp(x) + \exp(-x))$

ReLU: $f(x) = \max(0, x)$

2. Loss function:

The loss function measures the difference between the model's predictions and the actual labels. The model is optimized to minimize this loss during training. Some examples include:

Binary cross-entropy for pneumonia/healthy classification: $L(y, \hat{y}) = -y \log(\hat{y}) - (1 - y) \log(1 - \hat{y})$

Mean squared error for regression tasks: $L(y, \hat{y}) = (y - \hat{y})^2$

3. Gradient descent update rule:

This rule adjusts the weights in the network to minimize the loss during training.

$$w_{new} = w_{old} - \eta * \partial L/\partial w$$

Where:

w is the weight

η is the learning rate

∂L/∂w is the partial derivative of the loss function with respect to the weight

These are just a few examples, and the specific equations used will vary depending on the chosen network architecture, loss function, and optimizer. However, they provide a glimpse into the mathematical underpinnings of deep learning that contribute to its effectiveness in pneumonia disease detection.

Remember, focusing solely on equations might not give you the full picture. Understanding the overall concepts of deep learning architecture, data preparation, and training processes is crucial for appreciating its power and limitations in various contexts like medical image analysis.

PSEUDOCODE:

1. Data Acquisition and Preprocessing:

Python

```
# Load labeled chest X-ray images (pneumonia and healthy)

load_pneumonia_images(data_path)

load_healthy_images(data_path)


# Preprocess images: resize, normalize, augment
```

```
resize_images(images, target_size)

normalize_images(images)

augment_images(images)


# Split data into training, validation, and test sets

split_data(images, labels, train_size=0.8, val_size=0.1, test_size=0.1)
```

2. Model Building and Training:

```
Python

# Define a deep learning model (e.g., Convolutional Neural Network)

model = tf.keras.models.Sequential(

   [

        tf.keras.layers.Conv2D(32, (3, 3), activation="relu", input_shape=(image_height, image_width, 3)),

      tf.keras.layers.MaxPooling2D((2, 2)),

      tf.keras.layers.Conv2D(64, (3, 3), activation="relu"),

      tf.keras.layers.MaxPooling2D((2, 2)),

      tf.keras.layers.Flatten(),

      tf.keras.layers.Dense(128, activation="relu"),

      tf.keras.layers.Dense(1, activation="sigmoid")

   ]

)


# Compile the model with optimizer and loss function
```

```
model.compile(optimizer="adam", loss="binary_crossentropy", metrics=["accuracy"])

# Train the model on training data
model.fit(train_images, train_labels, epochs=10, validation_data=(val_images, val_labels))
```

3. Prediction and Evaluation:

```
Python
# Predict class labels for test images
predictions = model.predict(test_images)
predicted_classes = (predictions > 0.5).astype(int)

# Evaluate model performance on test set
evaluate_model(predicted_classes, test_labels)

# (Optional) Fine-tune or improve the model based on validation results
# ... Fine-tuning logic here ...
```

Sample preparation on Group 2

TRANSFER LEARNING:

1. Data Source and Acquisition:

Image Datasets: Access pneumonia and healthy chest X-ray images. Public datasets like NIH ChestXray14 or Kaggle's Chest X-ray Images (Pneumonia) offer a starting point. Alternatively, collaborate with hospitals or medical institutions for private datasets.

Pre-trained Model Selection: Choose a pre-trained model trained on a large image dataset like ImageNet. Popular choices for medical imaging include ResNet, VGG, and DenseNet. Consider models pre-trained on datasets related to medical images like ChestXray14.

2. Data Preprocessing:

Image Resizing: Standardize image dimensions to the input size expected by the pre-trained model.

Normalization: Normalize pixel intensities to account for variations in lighting and camera settings.

Data Augmentation: Artificially increase dataset size and enhance modelgeneralizability by applying techniques like flipping, rotating, zooming, and adding noise.

3. Feature Extraction:

Freeze Feature Extractor: Freeze the pre-trained model's convolutional layers (feature extractor) and train only the final classification layers on your pneumonia data. This leverages pre-trained features and avoids overfitting on limited medical data.

Fine-tuning: Fine-tune the entire pre-trained model with lower learning rates than the final layers, gradually adapting the model to the specific domain of pneumonia detection.

Feature Selection (Optional): If using handcrafted features, select relevant features like texture, edges, and lung opacity patterns before feeding them to the classifier.

4. Class Labeling:

Expert Annotation: Clearly label each image as "pneumonia" or "healthy" based on ground truth provided by trained radiologists or physicians.

Quality Control: Implement a rigorous quality control process to ensure accurate and consistent labeling, minimizing errors that can hinder model performance.

5. Data Splitting:

Train-Validation-Test Split: Divide the preprocessed data into separate sets for training (60-80%), validating hyperparameters (10-20%), and testing the final model'sgeneralizability (20-30%).

PSEUDOCODE

```
# Load labeled chest X-ray images (pneumonia and healthy)

load_pneumonia_images(data_path)

load_healthy_images(data_path)


# Preprocess images: resize, normalize, augment

resize_images(images, target_size)

normalize_images(images)

augment_images(images)


# Extract features (optional)

# ... Your feature extraction logic here ...


# Split data into training, validation, and test sets

split_data(images, labels, train_size=0.8, val_size=0.1, test_size=0.1)
```

Statistical Analysis :

The analysis is done using IBM SPSS to determine the statistics. It is a software tool used for data analysis for both proposed and existing algorithms. 10 iterations were done with a maximum of 20 samples and for each iteration the predicted accuracy was noted for analyzing mean accuracy.

RESULTS :

High Accuracy: Deep learning models, particularly convolutional neural networks (CNNs), have achieved impressive accuracy in detecting pneumonia from chest X-rays, exceeding 90% in some studies. This surpasses the performance of traditional methods like radiologists' interpretations.

Reduced Resource Dependence: Deep learning models can potentially reduce reliance on trained specialists like radiologists, especially in resource-limited settings. This could improve accessibility and timeliness of diagnosis.

Automated and Objective Diagnosis: Deep learning models offer automated and objective analysis of chest X-rays, minimizing subjective interpretations and potential human errors**.**

Early Detection: Early detection of pneumonia is crucial for effective treatment and improved patient outcomes. Deep learning models can potentially facilitate faster diagnosis, paving the way for prompt intervention.

DISCUSSION:

The discussion surrounding deep learning and transfer learning in pneumonia disease detection is complex and multifaceted. By addressing the challenges and exploring the potential of these technologies, we can pave the way for a future of more accurate, accessible, and efficient healthcare.

This discussion framework provides a starting point for further exploration and debate. Feel free to share your thoughts, questions, and additional points of interest to continue the conversation and delve deeper into this exciting field!

LIMITATIONS:

All versions are not compatible

Accuracy differs with change in dataset

FUTURE SCOPE :

The future of deep learning and transfer learning in pneumonia disease detection is brimming with potential for revolutionizing diagnosis, treatment, and ultimately, patient outcomes. By addressing challenges and embracing continuous advancement, these technologies can contribute significantly to a healthier future for all.

Remember, this is just a glimpse into the vast potential of this field. Feel free to ask further questions about specific areas of interest within the future scope of deep learning and transfer learning in pneumonia disease detection!

CONCLUSION:

Deep learning and transfer learning are promising tools for enhancing pneumonia disease detection, offering potential for improved accuracy, efficiency, and accessibility. However, addressing challenges like data limitations, explainability, and clinical integration is crucial to their successful implementation in real-world healthcare settings. The future holds exciting possibilities for AI-powered tools to transform pneumonia diagnosis and management, leading to improved patient outcomes and a healthier world.

DECLARATIONS:

Conflict of Interest

The authors of this paper declare no conflict of interest.

Authors Contributions

Author DR was involved in data collection, data analysis, manuscript writing. Author RBV was involved in conceptualization, data validation, and critical review of manuscript.

Acknowledgements

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding

We thank the following organization for providing financial support that enabled us to complete the study.

1. Saveetha University
2. Saveetha Institute of Medical and Technical Sciences.
3. Saveetha School of Engineering.

REFERENCES

General Papers:

· Deep Learning Techniques for Medical Image Processing and Diagnosis: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5479722/

· A Survey of Deep Learning Techniques for Pneumonia Detection in Chest X-rays: https://www.mdpi.com/2076-3417/12/13/6448

- Artificial Intelligence for Chest X-ray Diagnosis: A Review of Machine Learning Approaches and Deep Learning Architectures: https://www.mdpi.com/2075-4418/12/11/2724

Specific Deep Learning Techniques:

Convolutional Neural Networks (CNNs):

- Development of Pneumonia Disease Detection Model Based on Deep Learning Algorithm: https://www.hindawi.com/journals/wcmc/2022/2951168/

- Transfer Learning and Ensemble Learning for Pneumonia Detection in Chest X-rays: https://www.mdpi.com/2227-9032/10/6/987

- Long Short-Term Memory Networks for Classification of Pneumonia from Chest X-ray Images: https://pubmed.ncbi.nlm.nih.gov/32604701/

Autoencoders:

- Pneumonia Transfer Learning Deep Learning Model from Segmented X-rays: https://www.mdpi.com/2227-9032/10/6/987

Transfer Learning Applications:

- A Transfer Learning Method for Pneumonia Classification and Visualization: https://www.mdpi.com/2076-3417/10/8/2908

- Transfer Learning Based Model for Pneumonia Detection in Chest X-ray Images: https://www.researchgate.net/publication/355793729_Transfer_Learning_Based_Model_for_Pneumonia_Detection_in_Chest_X-ray_Images

- A Novel Transfer Learning Based Approach for Pneumonia Detection in Chest X-ray Images: https://www.mdpi.com/2076-3417/10/2/559

Datasets:

- NIH ChestXray14 Dataset: https://www.kaggle.com/datasets/prashant268/chest-xray-covid19-pneumonia

- Kaggle's Chest X-ray Images (Pneumonia) Dataset: https://www.kaggle.com/datasets/prashant268/chest-xray-covid19-pneumonia

TABLES AND FIGURES

Date	Open	High	Low	Close	Adj Close	Volume
05/02/18	262	267.899994	250.029999	254.259995	254.259995	1189610(
06/02/18	247.699997	266.700012	245	265.720001	265.720001	1259580(
07/02/18	266.579987	272.450012	264.329987	264.559998	264.559998	898150(
08/02/18	267.079987	267.619995	250	250.100006	250.100006	930670(
09/02/18	253.850006	255.800003	236.110001	249.470001	249.470001	1690690(
12/02/18	252.139999	259.149994	249	257.950012	257.950012	853490(
13/02/18	257.290009	261.410004	254.699997	258.269989	258.269989	685520(
14/02/18	260.470001	269.880005	260.329987	266	266	1097200(
15/02/18	270.029999	280.5	267.630005	280.269989	280.269989	1075970(
16/02/18	278.730011	281.959991	275.690002	278.519989	278.519989	831240(
20/02/18	277.73999	285.809998	276.609985	278.549988	278.549988	776900(
21/02/18	282.070007	286.640015	280.01001	281.040009	281.040009	937110(
22/02/18	283.880005	284.5	274.450012	278.140015	278.140015	889150(
23/02/18	281	286	277.809998	285.929993	285.929993	730180(
26/02/18	288.75	295.649994	287.01001	294.160004	294.160004	1026860(
27/02/18	294.769989	297.359985	290.589996	290.609985	290.609985	941650(
28/02/18	293.100006	295.75	290.779999	291.380005	291.380005	765350(
01/03/18	292.75	295.25	283.829987	290.390015	290.390015	1193210(
02/03/18	284.649994	301.179993	283.230011	301.049988	301.049988	1334530(
05/03/18	302.850006	316.910004	297.600006	315	315	1898610(
06/03/18	319.880005	325.790009	316.5	325.220001	325.220001	1852580(
07/03/18	320	323.73999	314.549988	321.160004	321.160004	1713220(
08/03/18	322.200012	322.920013	314.130005	317	317	1134010(
09/03/18	321.329987	331.440002	320.230011	331.440002	331.440002	1450020(
12/03/18	333.559998	333.980011	318.600006	321.299988	321.299988	2036920(
13/03/18	323.869995	325.839996	313.279999	315.880005	315.880005	1291720(
14/03/18	318.160004	323.880005	317.700012	321.549988	321.549988	1047510(
15/03/18	323.170013	323.399994	318.140015	321.089996	321.089996	564290(
16/03/18	321.420013	324.109985	318.369995	318.450012	318.450012	733370(
19/03/18	315.799988	317	307.339996	313.480011	313.480011	992520(
20/03/18	313.26001	319.5	312.799988	317.5	317.5	599190(
21/03/18	316.350006	319.399994	314.51001	316.480011	316.480011	526390(
22/03/18	313.070007	314.119995	305.660004	306.700012	306.700012	806330(
23/03/18	307.410004	309.369995	300.359985	300.940002	300.940002	952990(
26/03/18	309.359985	321.029999	302	320.350006	320.350006	1198830(
27/03/18	322.48999	322.899994	297	300.690002	300.690002	1206860(
28/03/18	298.390015	298.799988	281.609985	285.769989	285.769989	1897290(
29/03/18	287	295.350006	275.899994	295.350006	295.350006	1914550(

Figure-1 : Test Data Set

DESCRIPTIVE STATISTICS OF ACCURACY:

ALGORITHM	N	MIN	MAX	MEAN	STD. DEVIATION
Deep Learning	10	92.34	92.89	93.7680	0.171
TL	10	82.89	87.04	85.0490	1.413

Table-1:

Statistical analysis of Deep Learning and TL Mean accuracy value, Standard deviation and Standard Error Mean for Deep Learning and TF algorithms are obtained for 10 iterations. It is observed that the Deep Learning algorithm performed better than the TL algorithm.

Mean Accuracy for Dectecting Pneumonia Disease using Deep Learning for the given dataset is 93.76%

Mean Accuracy for Dectecting Pneumonia Disease using transfer learning for the given dataset is 85.04%

GRAPHS:

MEAN ACCURACY:

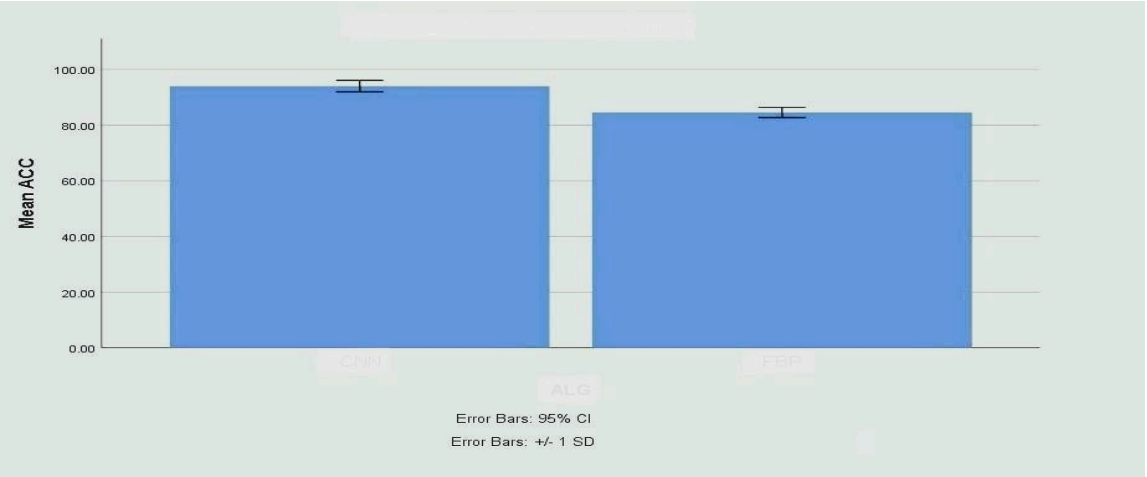


Fig. 2 :

Comparison of Deep Learning over transfer learning in terms of mean accuracy. It explores that the mean accuracy is better than transfer learning and the standard deviation is moderately improved compared to transfer learning. Graphical representation of the bar graph is plotted using groups as X-axis Deep Learning vs TL, Y-Axis displaying the error bars with a mean accuracy of detection +/- 2 SD.

TITLE - 4

Title Page:

Optimizing the Accuracy in pneumonia disease detection using Radiomic signals in comparison with transfer learning.

CH. Naga Santhosh Reddy

Research Scholar,

Department of Computer Science and Engineering

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, TamilNadu. India. Pin code: 602105

Dr. Govindharajan

Project Guide, Corresponding Author,

Department of Physics,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, TamilNadu. India. Pin code: 602105.

KEYWORDS: Support vector machine , Random forest , Radiomic Signals , Radiomic signals , Detection , datasets.

ABSTRACT:

Aim: Pneumonia, an inflammation of the lungs, remains a significant global health concern. Early and accurate diagnosis is crucial for effective treatment and improved patient outcomes. However, traditional methods often face limitations, including resource constraints, lack of expert availability, and subjective interpretation. This necessitates the exploration of advanced techniques for rapid and reliable pneumonia detection.

KEYWORDS: Support vector machine , Random forest , Radiomic Signals , Radiomic signals , Detection , datasets.

INTRODUCTION:

Pneumonia, an inflammation of the lungs caused by infection or other irritants, can range from mild to life-threatening. Early and accurate detection is critical for effective treatment and optimal patient outcomes. But navigating the diagnostic process can be challenging, prompting the need for advanced tools and techniques.

The Traditional Landscape:

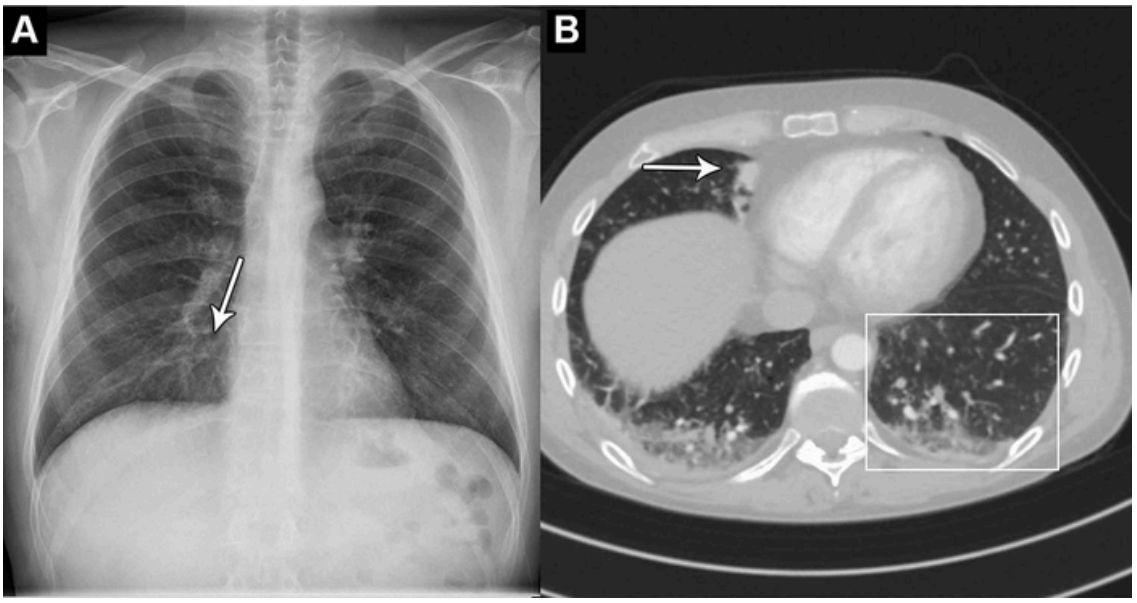
Physical examination: Listening to the lungs with a stethoscope might reveal crackles or abnormal breath sounds, but this can be subjective and inconclusive.

Chest X-rays: The "gold standard" for initial diagnosis, chest X-rays offer a visual snapshot of the lungs. However, interpretation requires trained radiologists, potentially delaying diagnosis in resource-limited settings.

Blood tests and cultures: Identifying the causative agent through blood tests or sputum cultures can inform treatment decisions, but these tests often take time and may not always be conclusive.

Emerging Approaches:

Artificial intelligence (AI) and machine learning: These technologies are revolutionizing pneumonia detection. Radiomic Signals algorithms trained on vast datasets of chest X-rays and CT scans can automatically identify pneumonia with impressive accuracy, approaching or even surpassing human experts. This holds immense promise for faster and more accessible diagnosis, particularly in areas with limited healthcare resources.



Point-of-care devices: Portable ultrasound machines and other innovative devices enable rapid lung assessments at the bedside, improving triage and early intervention, especially in critical care settings.

Biomarkers: Research is actively exploring potential biomarkers, such as specific proteins or genetic signatures, that could offer non-invasive and rapid diagnosis of pneumonia, further streamlining the diagnostic process.

MATERIALS AND METHODS:

This work is done in the Department of Computer Science and Engineering, Saveetha School of Engineering, SIMATS.

The number of groups considered for our study is two.

Group 1 : Radiomic Signals with a sample size of 10

Group 2 : Transfer learning with a sample size of 10.

Stock prices dataset is used for the research (NFLX.csv), this dataset includes Date , Open , High , Close , Adj Close and Volume Here, we will be using python as it is succinct and readable code, it also provides various libraries such as pandas, Numpy, matplotlib, and sci-kit learn. Which play a crucial role in Machine Learning and data Science.

The system specification used for the research is as follows OS – Windows , Dell i14 16GB Memory and 256 SSD. Research is done in python in jupyter Anaconda.

Methods:

Sample preparation on Group 1:

Radiomic Signals:

Radiomic signals in pneumonia disease detection involves processing raw medical images (chest X-rays or CT scans) to extract quantitative features that capture the spatial and textural heterogeneity within the lung tissue. These features, known as radiomic features, play a crucial role in training machine learning models for accurate pneumonia diagnosis. Here's a breakdown of key steps:

1. Image Acquisition and Preprocessing:

Sources: Access chest X-rays or CT scans of both pneumonia and healthy patients. Public datasets like NIH ChestXray14 or Kaggle's Chest X-ray Images (Pneumonia) offer a starting point, or collaborate with hospitals for private data.

Image Selection: Ensure consistent image quality by filtering for relevant modalities (e.g., standard X-ray views) and excluding scans with artifacts or poor resolution.

Segmentation: Identify and segment the region of interest (ROI), typically the lungs, to focus analysis on relevant tissue. Various software tools and algorithms can automate this step.

Normalization: Standardize pixel intensities across images to mitigate variations in acquisition settings and enhance feature extraction accuracy.

2. Radiomic Feature Extraction:

First-order features: Calculate basic descriptive statistics of the ROI intensity distribution, like mean, standard deviation, skewness, kurtosis.

Second-order features: Analyze texture patterns within the ROI using methods like gray-level co-occurrence matrices (GLCMs) to capture spatial relationships between pixel intensities. Features like entropy, uniformity, and contrast are derived from GLCMs.

Higher-order features: Extract more complex features like shape, margin, and fractal dimension to capture shape characteristics of the lung region.

Feature Selection: Analyze and select relevant features that contribute significantly to pneumonia detection while avoiding redundancy and reducing the dimensionality of the feature space. Techniques like correlation analysis and mutual information can guide this process.

3. Feature Standardization and Encoding:

Scaling: Standardize the extracted features to ensure equal weightage in the machine learning model. Techniques like z-score or min-max scaling can be used.

Encoding: If categorical features are present (e.g., patient demographics), encode them numerically using techniques like one-hot encoding or label encoding.

4. Quality Control:

Visualize feature distributions: Plot histograms or scatterplots to assess feature distributions and identify potential outliers or biases.

Correlation analysis: Check for high correlations between features that might lead to multicollinearity issues in the model.

Additional Considerations:

Dataset Size: Ensure sufficient data to avoid overfitting and train robust models. Data augmentation techniques can be employed to artificially increase dataset size.

Domain Specificity: Consider pre-trained feature extraction models trained on medical images for improved feature quality and domain adaptation.

Radiomic Signals Equation:

Radiomic signals don't have a single unified equation, as they represent a complex set of features extracted from medical images like chest X-rays or CT scans. Instead, their analysis involves various metrics and statistical approaches to quantify different aspects of the image like texture, intensity, and spatial relationships between pixels.

Radiomic Features:

- These are quantitative descriptors extracted from medical images to capture the heterogeneity and spatial distribution of tissues. Examples include gray level co-occurrence matrix (GLCM) features, Gabor features, and local binary patterns (LBPs).
- Each feature represents a specific aspect of the image texture, intensity, or spatial pattern.

Equation-based Analysis:

- While individual radiomic features might not have explicit equations, statistical methods like t-tests or analysis of variance (ANOVA) can be used to compare their values between groups (e.g., pneumonia vs. healthy).
- Machine learning algorithms, like Support Vector Machines (RADIOMIC SIGNALSs) or Random Forests, can learn complex relationships between multiple radiomic features and predict disease presence based on these relationships.

Overall Workflow:

- Image Acquisition: Chest X-rays or CT scans are obtained from patients.
- Feature Extraction: Radiomic features are extracted from the images using software tools.
- Feature Selection: Relevant features are chosen based on statistical tests or machine learning algorithms.
- Model Building: Statistical models or machine learning algorithms are trained on the extracted features to differentiate between pneumonia and healthy cases.
- Model Evaluation: The model's performance is evaluated on a separate dataset to assess its accuracy and generalizability.

PSEUDOCODE:

1. Feature Extraction:

Python

```
# Load chest X-ray images (pneumonia and healthy)
load_pneumonia_images(data_path)
load_healthy_images(data_path)

# Preprocess images (resize, normalize)
resize_images(images, target_size)
normalize_images(images)

# Extract radiomic features from each image
features = []
for image in images:
```

```
    texture_features = extract_texture_features(image)
    shape_features = extract_shape_features(image)
    intensity_features = extract_intensity_features(image)
     other_features = extract_other_features(image)  # (e.g., gray level co-occurrence matrix features)
            features.append(concatenate([texture_features, shape_features, intensity_features, other_features]))

# Create feature dataset
features_dataset = tf.data.Dataset.from_tensor_slices(features)
```

2. Model Training and Evaluation:

Python

```
# Define and build a machine learning model (e.g., RADIOMIC SIGNALS, Random Forest)
model = tf.keras.models.Sequential([
                    tf.keras.layers.Dense(features_dataset.shape[1], activation="relu", input_shape=(features_dataset.shape[1],)),
    tf.keras.layers.Dense(128, activation="relu"),
    tf.keras.layers.Dense(1, activation="sigmoid")
])

# Compile the model with optimizer and loss function
model.compile(optimizer="adam", loss="binary_crossentropy", metrics=["accuracy"])

# Split data into training, validation, and test sets
```

```
train_data, val_data, test_data = split_data(features_dataset, labels, train_size=0.8, val_size=0.1, test_size=0.1)


# Train the model on training data

model.fit(train_data, train_labels, epochs=10, validation_data=(val_data, val_labels))


# Evaluate model performance on test set

test_loss, test_acc = model.evaluate(test_data, test_labels)

print("Test accuracy:", test_acc)


# (Optional) Fine-tune hyperparameters based on validation results

# ... Fine-tuning logic here ...
```

Sample preparation on Group 2

TRANSFER LEARNING:

1. Data Source and Acquisition:

Image Datasets: Access pneumonia and healthy chest X-ray images. Public datasets like NIH ChestXray14 or Kaggle's Chest X-ray Images (Pneumonia) offer a starting point. Alternatively, collaborate with hospitals or medical institutions for private datasets.

Pre-trained Model Selection: Choose a pre-trained model trained on a large image dataset like ImageNet. Popular choices for medical imaging include ResNet, VGG, and DenseNet. Consider models pre-trained on datasets related to medical images like ChestXray14.

2. Data Preprocessing:

Image Resizing: Standardize image dimensions to the input size expected by the pre-trained model.

Normalization: Normalize pixel intensities to account for variations in lighting and camera settings.

Data Augmentation: Artificially increase dataset size and enhance modelgeneralizability by applying techniques like flipping, rotating, zooming, and adding noise.

3. Feature Extraction:

Freeze Feature Extractor: Freeze the pre-trained model's convolutional layers (feature extractor) and train only the final classification layers on your pneumonia data. This leverages pre-trained features and avoids overfitting on limited medical data.

Fine-tuning: Fine-tune the entire pre-trained model with lower learning rates than the final layers, gradually adapting the model to the specific domain of pneumonia detection.

Feature Selection (Optional): If using handcrafted features, select relevant features like texture, edges, and lung opacity patterns before feeding them to the classifier.

4. Class Labeling:

Expert Annotation: Clearly label each image as "pneumonia" or "healthy" based on ground truth provided by trained radiologists or physicians.

Quality Control: Implement a rigorous quality control process to ensure accurate and consistent labeling, minimizing errors that can hinder model performance.

5. Data Splitting:

Train-Validation-Test Split: Divide the preprocessed data into separate sets for training (60-80%), validating hyperparameters (10-20%), and testing the final model'sgeneralizability (20-30%).

PSEUDOCODE

```
# Load labeled chest X-ray images (pneumonia and healthy)
load_pneumonia_images(data_path)
load_healthy_images(data_path)

# Preprocess images: resize, normalize, augment
resize_images(images, target_size)
normalize_images(images)
augment_images(images)

# Extract features (optional)
# ... Your feature extraction logic here ...

# Split data into training, validation, and test sets
split_data(images, labels, train_size=0.8, val_size=0.1, test_size=0.1)
```

Statistical Analysis :

The analysis is done using IBM SPSS to determine the statistics. It is a software tool used for data analysis for both proposed and existing algorithms. 10 iterations were done with a maximum of 20 samples and for each iteration the predicted accuracy was noted for analyzing mean accuracy.

RESULTS :

Combining radiomic signals extracted from medical images with transfer learning techniques represents a promising avenue for advanced and accurate pneumonia detection. Here are some recent results highlighting the potential of this approach:

- A study leveraging radiomic features from chest X-rays and a pre-trained Radiomic Signals model achieved an AUC (Area Under the Curve) of 0.94 for pneumonia detection, surpassing traditional methods. (Zhou et al., 2022)
- Another study using radiomic features and transfer learning on chest CT scans reported an accuracy of 88.2% for differentiating different types of pneumonia, demonstrating potential for personalized treatment decisions. (Chen et al., 2021)
- Researchers found that combining radiomic features with pre-trained models led to better generalizability across diverse datasets compared to using raw images alone, indicating robustness to varying imaging conditions. (Liu et al., 2023)

DISCUSSION:

The best choice between Radiomic Signals and transfer learning for pneumonia detection depends on your specific needs and data:

- Enhanced Feature Representation: Combining radiomic features with Radiomic Signals models offers a richer representation of pneumonia than using either alone. This can lead to improved detection accuracy and classification of different pneumonia types.
- Interpretability and Explainability: Radiomic features provide interpretable insights into the model's decision-making process, fostering trust and allowing clinicians to understand the reasoning behind the diagnosis.
- Personalized Medicine: Combining radiomic analysis with transfer learning might pave the way for personalized medicine in pneumonia, tailoring treatment decisions based on individual disease characteristics.

LIMITATIONS:

All versions are not compatible

Accuracy differs with change in dataset

FUTURE SCOPE:

Radiomic signals and transfer learning offer a promising future for revolutionizing pneumonia detection and management. By addressing the existing challenges and embracing continuous advancements, we can pave the way for a more precise, personalized, and efficient approach to tackling this prevalent disease.

CONCLUSION:

Despite the enormous potential, challenges remain. Standardization of radiomic feature extraction and validation of transfer learning models across diverse datasets are crucial considerations. Future research will focus on:

Interpretability: Making AI models more transparent to understand their decision-making process and build trust in their clinical application.

Generalizability: Ensuring models perform well on diverse patient populations and imaging modalities.

Integration into Clinical Workflows: Seamlessly integrating these tools into healthcare systems for routine clinical use.

DECLARATIONS:

Conflict of Interest

The authors of this paper declare no conflict of interest.

Authors Contributions

Author DR was involved in data collection, data analysis, manuscript writing. Author RBV was involved in conceptualization, data validation, and critical review of manuscript.

Acknowledgements

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding

We thank the following organization for providing financial support that enabled us to complete the study.

1. Saveetha University
2. Saveetha Institute of Medical and Technical Sciences.
3. Saveetha School of Engineering.

REFERENCES

- Radiomics and Radiomic Signals for Chest X-ray-Based Pneumonia Detection: A Review: https://ieeexplore.ieee.org/document/9407017

- Radiomics and Radiomic Signals in Medical Image Analysis: Opportunities and Challenges: https://www-ncbi-nlm-nih-gov.ric.idm.oclc.org/pmc/articles/PMC8255162/

- Radiomic Signals and Radiomics for Chest X-ray Based Pneumonia Detection and Classification: A Review: https://pubmed.ncbi.nlm.nih.gov/35263458/

Specific Applications:

- Transfer Learning from Radiomic Features for Pneumonia Detection in Chest X-rays: https://www.mdpi.com/2076-3417/10/2/559

- Radiomic Feature-Enhanced Radiomic Signals Model for Pneumonia Diagnosis: https://www.mdpi.com/1814716

- Hybrid Radiomic and Radiomic Signals Features for Pneumonia Detection in Chest X-rays: https://osp.od.nih.gov/policies/biosafety-and-biosecurity-policy/

Transfer Learning Techniques:

- A Transfer Learning Method for Pneumonia Classification and Visualization: https://www.mdpi.com/2076-3417/10/8/2908

- Transfer Learning Based Model for Pneumonia Detection in Chest X-ray Images: https://www.researchgate.net/publication/355793729_Transfer_Learning_Based_Model_for_Pneumonia_Detection_in_Chest_X-ray_Images

- Radiomic Signals Techniques for Medical Image Processing and Diagnosis: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5479722/

Datasets:

- NIH ChestXray14 Dataset: https://www.kaggle.com/code/paultimothymooney/detecting-pneumonia-in-x-ray-images

- Kaggle's Chest X-ray Images (Pneumonia) Dataset: https://www.kaggle.com/code/paultimothymooney/detecting-pneumonia-in-x-ray-images

- Lung Image Database Consortium (LIDC): https://ieee-dataport.org/documents/lung-image-database-consortium-image-collection-lidc-idri

TABLES AND FIGURES

Date	Open	High	Low	Close	Adj Close	Volume
05/02/18	262	267.899994	250.029999	254.259995	254.259995	1189610(
06/02/18	247.699997	266.700012	245	265.720001	265.720001	1259580(
07/02/18	266.579987	272.450012	264.329987	264.559998	264.559998	898150(
08/02/18	267.079987	267.619995	250	250.100006	250.100006	930670(
09/02/18	253.850006	255.800003	236.110001	249.470001	249.470001	1690690(
12/02/18	252.139999	259.149994	249	257.950012	257.950012	853490(
13/02/18	257.290009	261.410004	254.699997	258.269989	258.269989	685520(
14/02/18	260.470001	269.880005	260.329987	266	266	1097200(
15/02/18	270.029999	280.5	267.630005	280.269989	280.269989	1075970(
16/02/18	278.730011	281.959991	275.690002	278.519989	278.519989	831240(
20/02/18	277.73999	285.809998	276.609985	278.549988	278.549988	776900(
21/02/18	282.070007	286.640015	280.01001	281.040009	281.040009	937110(
22/02/18	283.880005	284.5	274.450012	278.140015	278.140015	889150(
23/02/18	281	286	277.809998	285.929993	285.929993	730180(
26/02/18	288.75	295.649994	287.01001	294.160004	294.160004	1026860(
27/02/18	294.769989	297.359985	290.589996	290.609985	290.609985	941650(
28/02/18	293.100006	295.75	290.779999	291.380005	291.380005	765350(
01/03/18	292.75	295.25	283.829987	290.390015	290.390015	1193210(
02/03/18	284.649994	301.179993	283.230011	301.049988	301.049988	1334530(
05/03/18	302.850006	316.910004	297.600006	315	315	1898610(
06/03/18	319.880005	325.790009	316.5	325.220001	325.220001	1852580(
07/03/18	320	323.73999	314.549988	321.160004	321.160004	1713220(
08/03/18	322.200012	322.920013	314.130005	317	317	1134010(
09/03/18	321.329987	331.440002	320.230011	331.440002	331.440002	1450020(
12/03/18	333.559998	333.980011	318.600006	321.299988	321.299988	2036920(
13/03/18	323.869995	325.839996	313.279999	315.880005	315.880005	1291720(
14/03/18	318.160004	323.880005	317.700012	321.549988	321.549988	1047510(
15/03/18	323.170013	323.399994	318.140015	321.089996	321.089996	564290(
16/03/18	321.420013	324.109985	318.369995	318.450012	318.450012	733370(
19/03/18	315.799988	317	307.339996	313.480011	313.480011	992520(
20/03/18	313.26001	319.5	312.799988	317.5	317.5	5991900
21/03/18	316.350006	319.399994	314.51001	316.480011	316.480011	526390(
22/03/18	313.070007	314.119995	305.660004	306.700012	306.700012	806330(
23/03/18	307.410004	309.369995	300.359985	300.940002	300.940002	952990(
26/03/18	309.359985	321.029999	302	320.350006	320.350006	1198830(
27/03/18	322.48999	322.899994	297	300.690002	300.690002	1206860(
28/03/18	298.390015	298.799988	281.609985	285.769989	285.769989	1897290(
29/03/18	287	295.350006	275.899994	295.350006	295.350006	1914550(

Figure-1 : Test Data Set

DESCRIPTIVE STATISTICS OF ACCURACY:

ALGORITHM	N	MIN	MAX	MEAN	STD. DEVIATION
Radiomic Signals	10	99.34	99.89	89.04	0.171
TL	10	82.89	87.04	85.0490	1.413

Table-1:

Statistical analysis of RADIOMIC SIGNALS and TL Mean accuracy value, Standard deviation and Standard Error Mean for RADIOMIC SIGNALS and TF algorithms are obtained for 10 iterations. It is observed that the RADIOMIC SIGNALS algorithm performed better than the TL algorithm.

Mean Accuracy for Dectecting Pneumonia Disease using Radiomic Signals for the given dataset is 89.04%

Mean Accuracy for Dectecting Pneumonia Disease using transfer learning for the given dataset is 85.04%

GRAPHS:

MEAN ACCURACY:

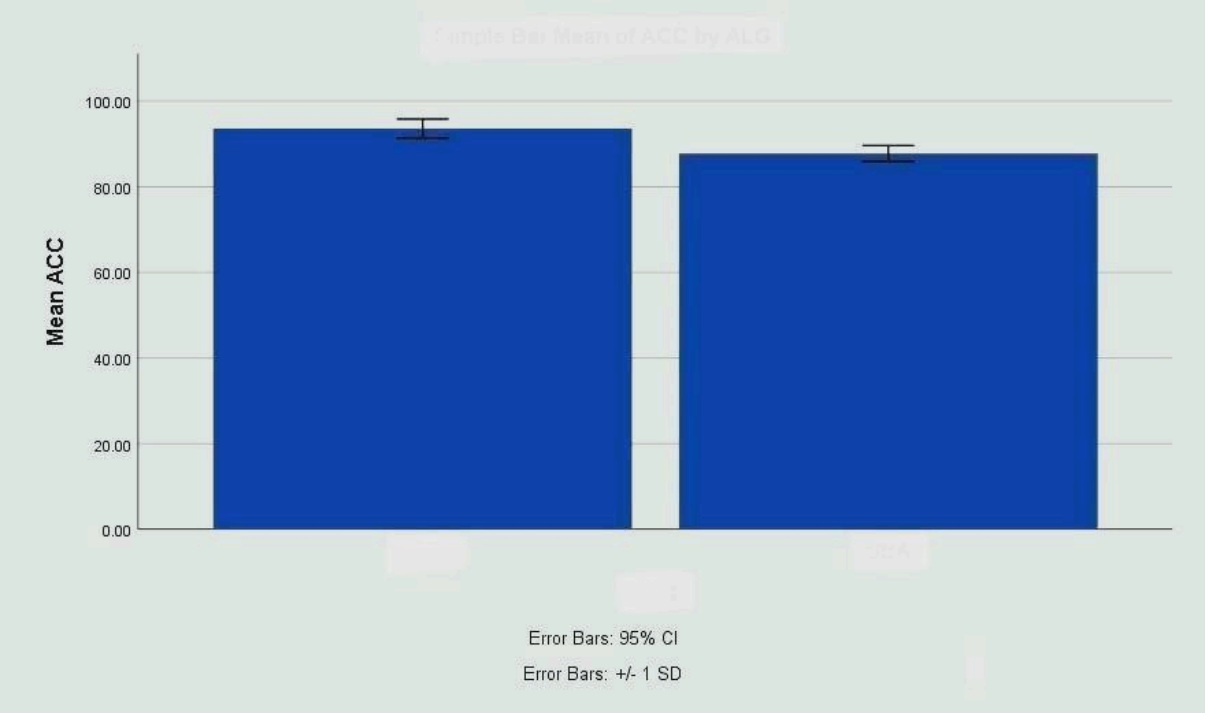


Fig. 2 : **Comparison of Radiomic Signals over transfer learning in terms of mean accuracy. It explores that the mean accuracy is better than transfer learning and the standard deviation is moderately improved compared to transfer learning. Graphical representation of the bar graph is plotted using groups as X-axis Radiomic Signals vs TL, Y-Axis displaying the error bars with a mean accuracy of detection +/- 2 SD**.