



Systems and Biomedical Engineering Department
Third Year 2023



Artificial Intelligence in Medicine

SBME 3021 – Final Project

Supervised by:

Dr. Inas A. Yassine

Presented by:

Name	Sec.	B.N.
Ahmed Hesham AbdElEllah Mohamed	1	8
Hussein Mohamed Saad AbdElHafez	1	26
AbdElRahman Muhsen AbdElLatif AbdElGhany	1	51
Mariam Magued Habib Bebawy	2	27

Abstract

While the field of artificial intelligence keeps growing, in this paper we focus on the basics of models and how they work. Our research has led us to the topic of Pneumonia classification due to the recent events of COVID-19. In this paper we walk through the phases and steps taken to reach decent results of 80% taking into consideration the issue of missing and/or lacking data.

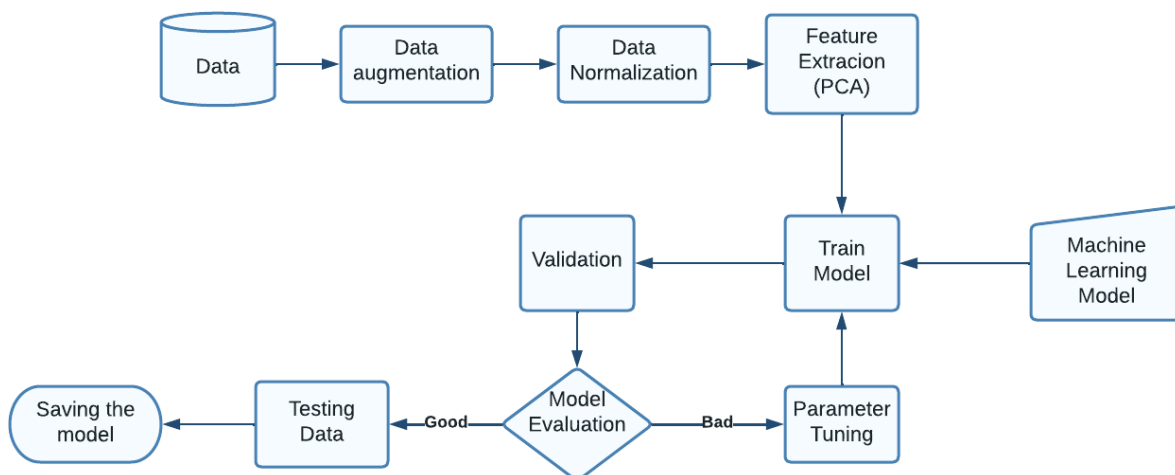
Introduction

Unhygienic living conditions, where high levels of pollution may prevail, are home to a respiratory infection called Pneumonia. It is caused by bacteria or viruses, and may cause pleural effusion. The air sacs may be filled with fluids, causing coughing, fever, chills, and difficulty breathing. Pneumonia can range in seriousness from mild to life-threatening. It is most serious for infants and young children, people older than age 65, and people with health problems or weakened immune systems.

In this paper, we tackle this problem. After thorough research, it had been decided that we would train a machine learning model to classify Pneumonia cases based on the collected data.

Disclaimer: bacterial and viral classification is not considered and is out of the scope of this paper.

Before going into detail of the project implementation, the pipeline used throughout the code is presented in the following diagram.



Literature Review

Explainable Machine Learning for COVID-19 Pneumonia Classification With Texture-Based Features Extraction in Chest Radiography

This paper aims to classify COVID-19 pneumonia based on X-ray images by investigating radiomic features and classification models to differentiate between COVID-19 pneumonia and other lung patterns. Lung segmentation for three zones of the chest and histogram normalization are used before feature extraction phase. Using the interpretable Shapley Additive Explanations (SHAP) recursively feature elimination with cross-validation is used to select features. The models used for hyperparameters optimization are XGBoost and random forest ensemble trees using random search.

The performance metrics of both the XGBoost and Random Forest ensemble learning were tested and compared against each other to prove that the best classification model is the XGBoost, with an accuracy of 0.82 and a sensitivity of 0.82.

<i>Model</i>	<i>Accuracy</i>	<i>F1-score</i>	<i>Sensitivity</i>	<i>Precision</i>
<i>XGBoost</i>	0.82	0.82	0.82	0.82
<i>Random Forest</i>	0.77	0.78	0.81	0.75

The weaknesses in this paper are prominent in the data acquisition phase. According to said paper, even though the data is widely available, the consistency is not; most databases suffer from having missing and/or mislabeled data. Also, the used dataset offered data from pediatric and adult patients leading to an increase of age-related biases. Another issue is the lack of COVID and non-COVID data from the same medical centers; it is worth looking into due to the fact that medical centers use different scanners and protocols which causes mismatch in between the data. These limitations caused the performance metrics of this paper's model to be significantly lower than other papers with the same objective.

The methods of the model proved important in the classification of COVID-19 pneumonia from other lung patterns; it shows the significance of certain lung parts (middle left and superior right lung zones) in the classification process. This model may be used as a first tool in the classification process of individuals with a COVID -19 diagnosis as the accuracy score isn't high enough to be used on its own.

Conventional Machine Learning based on Feature Engineering for Detecting Pneumonia from Chest X-rays

Data Preprocessing

In this paper, they were trying to solve the pneumonia detection problem using classical machine learning algorithms. The auxiliary dataset recorded consists of 11500 No Lung Opacity/Not Normal, 8525 Normal, and 8964 Lung Opacity patient cases. They randomly sampled the number of pneumonia images in order to have the same number as that of the normal images. To extract the feature parameters, they adopted the Principal Component Analysis (PCA) approach to extract the relevant principal components from the captured X-ray images to reduce the complexity of this model and prevent overfitting.

Evaluation Metrics and Results

They used different evaluation metrics such as Accuracy, Specificity, Precision, Recall, True Positive Rate, False Positive Rate, F1 score, and AUC score. They used different classifiers such as Logistic regression, K-Nearest Neighbors, Decision Tree, Random Forest, Multi-Layer perceptron, Gaussian Naïve Bayes, Bernoulli Naïve Bayes, Gradient Boosting Classifier, Support vector machine, linear support vector machine. Here are some of the results they have reached. They tried training with and without PCA, we shall only show the results with PCA.

<i>Classifier</i>	<i>AUC – ROC score</i>	<i>Accuracy score</i>	<i>Time</i>
<i>Logistic Regression</i>	0.918	0.85	57.15
<i>K-Nearest Neighbors</i>	0.923	0.86	200.63
<i>Decision Tree</i>	0.813	0.81	92.72
<i>Random Forest</i>	0.914	0.81	16.34
<i>Multi-layer Perceptron</i>	0.916	0.89	143.84
<i>Gaussian Naive Bayes</i>	0.723	0.64	5.8
<i>Gradient Boosting Classifier</i>	0.93	0.86	384.94
<i>Linear Support-Vector machine Classifier</i>	0.81	0.81	96.9
<i>Bernoulli Naive Bayes</i>	0.866	0.78	6.27
<i>Support-Vector machine Classifier</i>	0.93	0.9	1346.73

Conclusion

The performance measure of the model classifiers recorded higher scores when they used the principal component of the chest X-ray images. Support-Vector machine Classifier recorded the highest score in accuracy and AUC score but it took too much time relatively even after applying PCA.

Dimensionality reduction is a very crucial step as it can lead to variance loss, but they showed good selection of principal components as they contain 95% of the explained variance.

PCA computes the coordinate rotation based on the statistics of the training dataset, without considering the class labels. Therefore, even though the class labels may be perfectly linearly separable (both before

and after PCA), it does not imply that they will be separable independently by each of the coordinates transformed. Thus, the linear decision boundary is not parallel to any of the transformed coordinate axes. Due to this, we expect the performance of the Naive Bayes classifier like the Gaussian Naive Bayes GNB to produce a relatively lower score because the model classifier (GNB) assumes the features are conditionally independent.

Support Vector Machine (SVM) for Medical Image Classification of Tumorous

This paper worked on a set of MRI images for brain tumors, where a model based on Support Vector machines will be developed. In the first stage some preprocessing processes, feature extraction, and feature reduction are applied. In the second step, SVM will be trained.

Preprocessing

Firstly, some preprocessing methods are applied to the image dataset before applying SVM. The Discrete Wavelet Transform (DWT) is used in the first step to remove noise from images. In the second step of preprocessing, the features have been reduced.

Here Authors need to reduce the features in the image while maintaining the underlying data. They used PCA (principal component analysis) to reduce dimensionality.

Conclusion

SVM yielded good results in the classification of brain tumors but the paper didn't discuss the result by any metrics.

An Improved SVM Classifier for Medical Image Classification

This paper discussed a better way of using SVM, by using the rough set theory RST which is able to deal with large sets of data and eliminate redundant information. This paper managed to improve the result of SVM in classifying digital mammography, by 3.42% from 92.94% without using the RST to 96.56%.

On Breast Cancer Detection: An Application of Machine Learning Algorithms on the Wisconsin Diagnostic Dataset

The diagnosis of breast cancer is targeted in this paper using the Wisconsin Diagnostic Breast Cancer dataset, which consists of digitized images of a fine needle aspirate (FNA) of breast masses. This paper offers a comparison of multiple machine learning algorithms, along with some neural network models which we shall not be discussing.

According to said paper, no preprocessing was done on the data to achieve higher results. However, the data was split into train and test splits with a ratio 70 to 30 %, and the hyperparameters were manually selected after multiple thorough trials.

In this paper, the machine learning models used are: Linear Regression, Softmax Regression, Nearest Neighbour (L1-NN & L2-NN), and SVM. Others, which are out of our scope, include GRU-SVM and MLP. All models held great results exceeding 90% test accuracy.

<i>Model</i>	<i>Accuracy</i>	<i>FPR</i>	<i>FNR</i>	<i>TPR</i>	<i>TNR</i>
<i>Linear Regression</i>	0.961	0.102	0	1	0.898
<i>SoftMax Regression</i>	0.977	0.058	0	1	0.942
<i>L1-NN (Manhattan)</i>	0.936	0.063	0.065	0.935	0.938
<i>L2-NN (Euclidean)</i>	0.947	0.094	0.028	0.972	0.906
<i>SVM</i>	0.961	0.064	0.025	0.975	0.936

From the given metrics we can conclude that SoftMax Regression, followed by SVM and Linear Regression hold the best test accuracy results of over 95%. The training time varied from one model to another, for instance Linear Regression finished training in 35 seconds, while SoftMax Regression finished in 25 seconds only, others aren't explicitly mentioned in the paper.

The implementations in this paper don't have obvious weaknesses due to the high performance of the binary classification of breast cancer and the satisfactory statistical results. Furthermore, to make these methods/models more elaborate, cross validation methods can be used to provide more accurate prediction performance.

Support vector machine classification of dimensionally reduced structural MRI images for dementia

Before this paper there were previous works discussing machine learning techniques to classify MRI scans. They utilized relatively small samples of subjects, testing and training support vector machines on relatively small samples containing fewer than 100 subjects. Additionally, these works used the raw image data to train and test the classifiers, that is, raw voxel intensity values were fed into the algorithms. An MRI scan with 2563 voxels has over 15 million voxels. When replicating the procedure outline in previous works by training a linear-kernel SVM classifier using a subset (10%) of the raw voxel intensities, they observe severe over fitting of the classifier (i.e., training accuracy of 100%), which may introduce significant bias if applied out of sample.

In this paper they were trying to improve the generality of the classifier by reducing dimensionality of the feature set. Rather than testing and training on the entire imaged volume, they reduce the dimensionality of the feature set to only 11 features per example by using quantities computed from the images and principal component analysis (PCA), then they assess the performance of the classifiers for various sets of features via 10-fold cross-validation. Dimensionality reduction of the data also increases the computational tractability of the problem, enabling one to execute the classifier on a standard laptop computer

They compute eigen brains for two separate slices, one axial slice and one coronal slice, selecting slices on which gray matter and white matter volume have been shown to correlate with impaired cognitive ability. We select PCA components by parametrically testing the classifiers; the combination of coronal and axial components (#4 and #7, respectively) that maximize test accuracy. Using these methods, feature set includes 11 values for each subject (for example Age, gender, total white matter, total grey matter) All features are mean-subtracted and normalized by the variance of each feature. Lastly, they classify subjects with an CDR score greater than as 0 as demented, and so 100 of the 416 subjects classified with dementia

Using the dimensionally reduced feature set, the SVM classifier performed with a training accuracy of 86.4% (90.1%) and testing accuracy of 85.0% and (85.7%) for a linear and (gaussian) kernel. The test precision and recall for each classifier was 68.7% (68.5%) and 68.0% (74.0%), respectively. The Matthews correlation coefficient for the linear and (gaussian) kernels was 0.594 (0.616).

Workflow Phases

The implementation of this project was split into three consecutive phases: data acquisition and pre-processing, feature extraction, and model testing. Each phase is clearly explained throughout this paper.

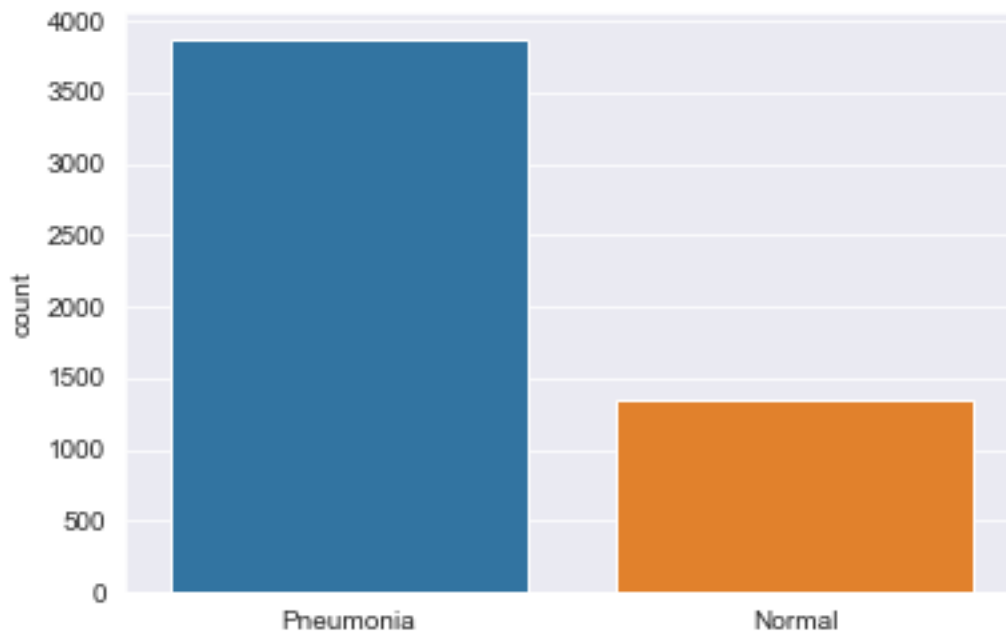
Phase I – Data Acquisition and Pre-processing

Data acquisition and preparation

In this project, we aim to use the obtained dataset of chest x-rays to detect cases of pneumonia. The dataset – obtained from Kaggle, link provided – is divided into train, test, and validation sets. Each set contains the two needed sets of classes: normal, and abnormal/pneumonia case. Furthermore, in the pneumonia case files, one can notice that there exist two types of pneumonia case: bacterial, and viral, both contributing to the training of the model, though considered as the same class; “Pneumonia”.

Data augmentation

In order to see how biased the data may be to each class, we can use the labels to make a simple plot for visualization.



One can notice right away that the data is heavily biased to the ‘Pneumonia’ label, which eventually leads to the occurrence of overfitting when training and testing the data. A simple solution to this issue may be performing data augmentation on the ‘Normal’ data.

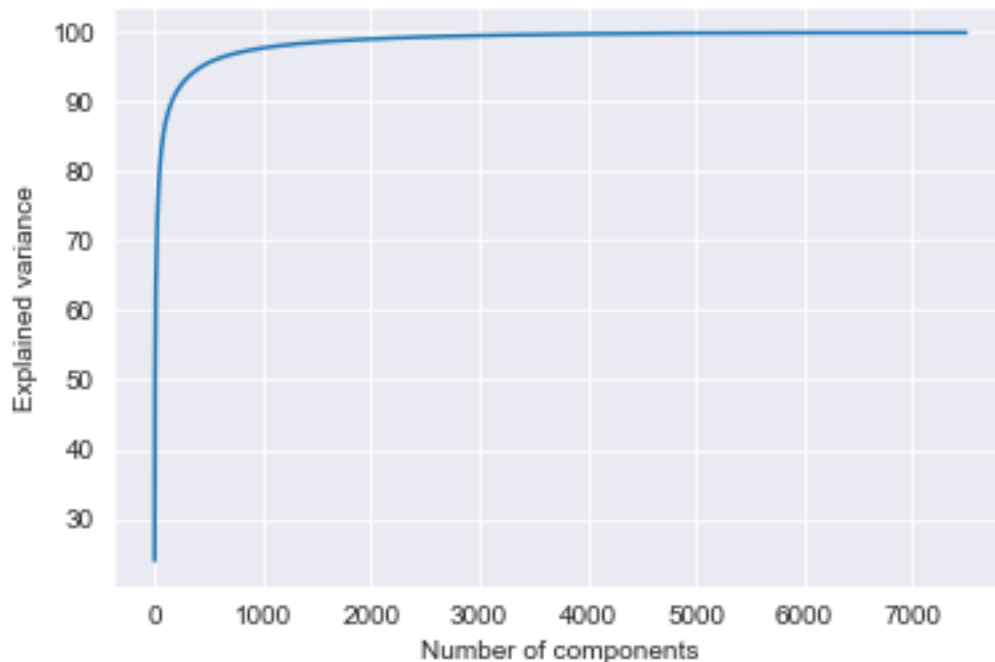
Data pre-processing

After inspecting the data, it was found that all images were of very high resolution of not less than 1500 x 1500 pixels. Naturally, it was decided that resizing the images to a smaller size would be the optimal option in this case to minimize processing time and model training speed. Through multiple tests, we settled on 100 x 100 pixels / image.

Upon further inspection, we realized that our chosen model would suffer from the problem of missing data due to the huge difference between the count of the images in the “Pneumonia” class versus the “Normal” one. The solution to this problem was to use some image manipulation libraries to perform data augmentation on the lacking class of images; i. e. The “Normal” class.

Phase II – Feature Extraction

During this phase, we only used the PCA pre-processing technique which helped in meaningful feature extraction. It was used to compress the images without the need to lose too much information, and to somewhat de-noise the data as it enhances the components with the greatest variance.



When performing the cumulative sum of the plot of the obtained PCA components, notice that the variance is almost a constant straight line after roughly 2500 components. Nevertheless, we picked 1000 components in order to reach the highest possible accuracy.

Phase III – Model Testing

Picking a Model

In order to pick a model, thorough research and investigation was required. We went through multiple papers that tackled similar issues to what we were trying to achieve; X-ray and image classification. Said papers usually used models like K-Nearest Neighbours, Support Vector Machine, XGBoost, Random Forests, Gradient Descent, and Regression models (i.e. Linear, Logistic, and SoftMax Regression).

The highest results were typically provided by SVM classification models, but other models -like KNN, logistic regression, and gradient boosting- yielded high results as well. Accordingly, It had decided to test said models with different variations and techniques.

The final models are as follows: SVM classification, Logistic regression, KNN, Ensemble Learning using { KNN, SVM, and Logistic regression }, and another Ensemble Learning model using { KNN, SVM, Logistic regression, and Gradient Boosting }.

Evaluation Metrics

The accuracy metric is used to compare the train, validation, and test scores, all to check for the way the data is fitted to the model, i.e. underfitted, perfectly fitted, overfitted. Seeing that after the data augmentation, the dataset is almost balanced, this metric is a viable option.

The AUC–ROC metric is also used to check the ability of the classifier of distinguishing between classes, where the higher the value, the better the performance. This metric is also viable to use considering the data was originally skewed towards the “Pneumonia” class.

The “time” metric wasn’t considered as it is heavily dependent on the machine being used and the kind of model one might be training.

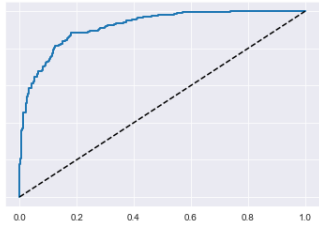
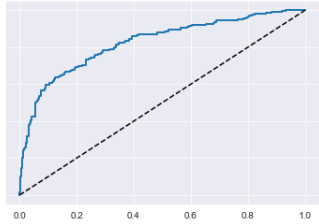
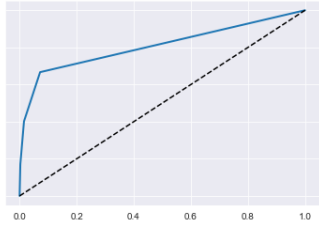
Model Testing

In order to have a fair comparison, each implemented model is compared against the rest, with the points of comparison being the scores obtained when training and testing. This is shown in the following table, (*figure: table*).

Saving the model

One simple approach to saving any of the models for future testing –without the need to re-train the model each runtime– is to use any pickling library to deploy said model and save it in a “.sav” file for future use.

figure: table

	<i>Train score %</i>	<i>Val score %</i>	<i>Test score %</i>	<i>AUC score %</i>	<i>ROC curve</i>
SVM	98.11	75.00	78.37	91.94	
Logistic regression	97.10	87.50	74.68	82.32	
KNN	96.01	62.50	76.60	80.75	
Ensemble learning – vl	99.25	81.25	77.56	—	—
Ensemble learning– vlI	98.66	75.00	75.80	—	—

Conclusion and future works

Through the use of PCA feature extraction, SVM proved to be the best model for the classification problem of this paper, i.e. Pneumonia classification, considering the other models' results.

In this project, a classical approach of machine learning algorithms was used for implementation. This should serve as a good start as it provides simpler workflow pipelines and development, which in turn is easier to process by hardware.

It is important to note that basic machine learning algorithms in this field are considered outdated given the huge advancements in the field of artificial intelligence. Deep learning models should prove more helpful in such cases.

References

1. Tutorial: image classification with scikit-learn – Kapernikov. (2018). Retrieved 12 June 2022, from <https://kapernikov.com/tutorial-image-classification-with-scikit-learn/>
2. Medical Image Pre-Processing with Python. (2021). Retrieved 12 June 2022, from <https://towardsdatascience.com/medical-image-pre-processing-with-python-d07694852606>
3. Basics of Machine Learning Image Classification Techniques. (2019). Retrieved 12 June 2022, from <https://iq.opengenus.org/basics-of-machine-learning-image-classification-techniques/>
4. Image Classification using Machine Learning - Analytics Vidhya. (2022). Retrieved 12 June 2022, from <https://www.analyticsvidhya.com/blog/2022/01/image-classification-using-machine-learning/>
5. Moura, L., Mattjie, C., Dartora, C., Barros, R., & Marques da Silva, A. (2022). Explainable Machine Learning for COVID-19 Pneumonia Classification With Texture-Based Features Extraction in Chest Radiography. *Frontiers In Digital Health*, 3. doi: 10.3389/fdgth.2021.662343
6. Grinsted, A., Moore, J., & Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes In Geophysics*, 11(5/6), 561-566. Retrieved from <https://hal.archives-ouvertes.fr/hal-00302394>
7. Journal, I. (2020). Support Vector Machine (SVM) for Medical Image Classification of Tumorous. *IJCSMC*. Retrieved from https://www.academia.edu/43373494/Support_Vector_Machine_SVM_for_Medical_Image_Classification_of_Tumorous_
8. Jiang, Y., Li, Z., Zhang, L., & Sun, P. An Improved SVM Classifier for Medical Image Classification. *Rough Sets And Intelligent Systems Paradigms*, 764-773. doi: 10.1007/978-3-540-73451-2_80
9. Papers with Code - On Breast Cancer Detection: An Application of Machine Learning Algorithms on the Wisconsin Diagnostic Dataset. (2022). Retrieved 15 May 2022, from <https://paperswithcode.com/paper/on-breast-cancer-detection-an-application-of>
10. Miller, V., Erlien, S., & Piersol, J. (2014). Support vector machine classification of dimensionally reduced structural MRI images for dementia. Retrieved 15 May 2022, from <https://arxiv.org/abs/1406.6568>