



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

F M J Ebiele, T Ansah-Narh, S R T Djiokap, E. Proven-Adzri, M Atemkeng

► To cite this version:

F M J Ebiele, T Ansah-Narh, S R T Djiokap, E. Proven-Adzri, M Atemkeng. Conventional Machine Learning based on Feature Engineering for Detecting Pneumonia from Chest X-rays. ACM Annual Conference of the South African Institute of Computer Scientists and Information Technologists, Sep 2020, Cape Town, South Africa. <10.1145/3410886.3410898>. <hal-02909906>

HAL Id: hal-02909906

<https://hal.science/hal-02909906v1>

Submitted on 3 Aug 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

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

F. M. J. Ebiele
jaures.ebiele@aims.ac.rw
African Institute for Mathematical
Sciences (AIMS), Rwanda

T. Ansah-Narh
philusnarh@gmail.com
Ghana Space Science & Technology
Institute (GSSTI), Ghana

S. R. T. Djiokap
stive.djiokap@aims.ac.rw
African Institute for Mathematical
Sciences (AIMS), Rwanda
Institute of Fine Arts Foumban,
University of Dschang, Cameroon

E. Proven-Adzri
proven2012@gmail.com
Ghana Space Science & Technology
Institute (GSSTI), Ghana

M. Atemkeng*
m.atemkeng@ru.ac.za
Department of Mathematics, Rhodes
University, South Africa

ABSTRACT

Chest X-ray is the standard approach used to diagnose pneumonia and other chest diseases. Early diagnosis of the disease is very relevant in the life of people, but analyzing X-ray images can be complicated and needs the competence of a radiographer. In this paper, we demonstrate the potential of detecting the disease in chest X-rays using conventional machine learning classifiers. The principal component analysis is used for the data dimensionality reduction and features extraction then the extracted features are used to train several model classifiers. We obtained an accuracy of 90 %, using 95 % of the principal explained variance.

KEYWORDS

Pneumonia, chest X-rays, feature extraction, principal component analysis, supervised learning.

ACM Reference Format:

F. M. J. Ebiele, T. Ansah-Narh, S. R. T. Djiokap, E. Proven-Adzri, and M. Atemkeng. Conventional Machine Learning based on Feature Engineering for Detecting Pneumonia from Chest X-rays . In SAICSIT' 20: Conference of the South African Institute of Computer Scientists and Information Technologists, 7 pages (2020). <https://doi.org/10.1145/3410886.3410898>

1 INTRODUCTION

Typically, as we breathe in, air moves into the trachea through large tubes (bronchi) then via smaller tubes (bronchioles), and eventually into tiny sacs known as the alveoli. The airwaves and the alveoli are springy such that when air goes into the body, each air sac inflates like a balloon and when we exhale, the air sacs deflate. As shown in Fig. 1, inhaled oxygen diffuses from the alveoli to the blood in the capillaries, and carbon dioxide diffuses from the blood in the capillaries to the air in the alveoli. Meanwhile, the lungs can be infected either in the bigger air passage or in the alveoli. Pneumonia is a disease that affects the alveoli of the lungs and it is mostly caused by infection by viruses, bacteria, fungi, or other causal organisms. The infection causes the body's immune

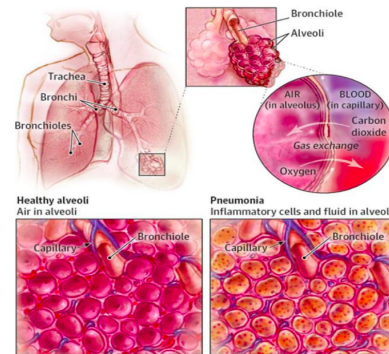


Figure 1: Lung anatomy and gas exchange (Reproduced from Kaggle).

system to become feeble by allowing the alveoli sacs to get filled with mucus and white blood cells. Hence, making it difficult for the incoming air to exchange oxygen with the red blood cell. The disease is one of the dominant causes of death in the world and very prevalent in developing countries [9, 21].

Chest X-ray is the best test for pneumonia diagnosis since this test enables the medical practitioner to investigate for any inflammation in the lungs. However, reading X-ray images can be very tricky as shown in Fig. 2 and requires domain expertise and experience. Therefore, it would be convenient to train a computer to read these X-ray images to automatically detect pneumonia. Note that the quick diagnosis by an artificial intelligence system would allow physicians to accurately confirm and start a treatment plan for pneumonia more quickly than with current clinical practice.

Recent works [2, 12, 16, 18, 19] in image classification on pneumonia detection use convolutional neural network (CNN) architecture to extract features and build the classification model. In this work, we demonstrate the feasibility of detecting pneumonia in chest X-rays using the traditional supervised learning approach. Here, we define a set of target groups and develop a model to observe the images with labels. Nonetheless, physical inspection and identifying images could be an exhausting task particularly, when the

images are massive in number. It is, therefore, very necessary to automate this process using ML algorithms.

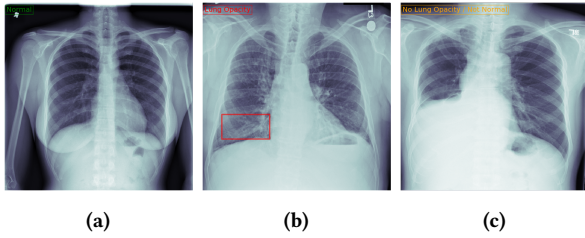


Figure 2: (a) No pathology detected, (b) pathology detected, (c) no lung opacity/not normal.

The paper is organized as follows: Section 2 investigates statistically the pneumonia dataset used for this work. Section 3 presents how we handle the curse of dimensionality and extracts the significant features using the PCA method. A discussion of the classification models is presented in Section 4. Section 5 presents the performance metrics of the model classifiers. Next, we discuss the results in Section 6 and conclude in Section 7.

2 EXPLORATORY DATA ANALYSIS

2.1 Dataset

The dataset used in this paper contains 28989 bounding boxes, including target labels, and 25684 unique patient IDs. Note that this dataset has different rows for the same patient ID, as each row describes one observation per patient. The auxiliary dataset recorded consists of 11500 No Lung Opacity/Not Normal, 8525 Normal, and 8964 Lung Opacity patient cases as presented in Fig. 3a. The data source is obtained from the National Institutes of Health (NIH) Chest X-ray¹. The X-ray images used are in Digital Imaging and Communications in Medicine (DICOM²) format. The raw pixel images are saved in 128×128 8-bit encoded (that is, 2^8 gray-scales) NumPy arrays. The stacked bar charts in Fig. 3b depict the ratio of male to female for individual cases. Here, we can observe that there are more male patients than females for the reported cases. Even though the dataset recorded an increase in the number of male patients, the age distribution of pneumonia cases for both male and female patients is the same as shown in Fig. 4. This plot shows that patients between 50 – 60 years are more prone to pneumonia disease.

2.2 Data pre-processing

For this work, we considered all lung opacity cases with bounding boxes for pneumonia and all normal cases. Here, the normal cases have a ‘Target’ value of 0 whilst pneumonia cases are labeled 1. To handle the imbalance in the dataset, we randomly sample the number of pneumonia images in order to have the same number as that of the normal images as displayed in Fig. 5. The balanced chest X-ray images are then used to train and test the model classifiers by adopting the 80/20 training/test split scheme.

¹<https://www.rsna.org/en/education/ai-resources-and-training/ai-image-challenge/RSNA-Pneumonia-Detection-Challenge-2018>

²The images can be opened using the Python package pydicom.

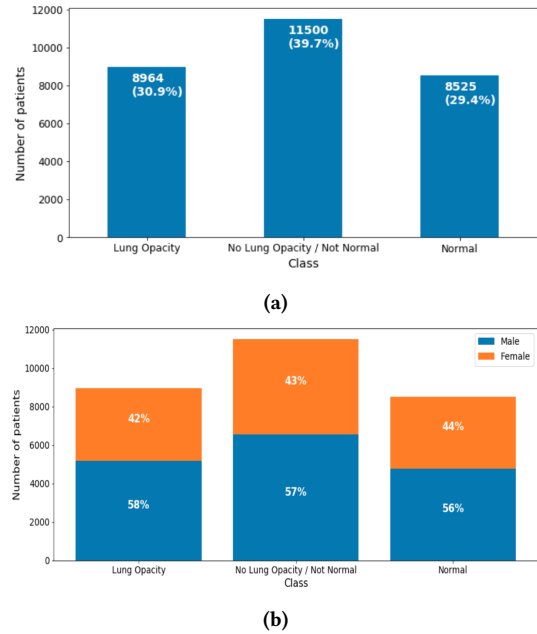


Figure 3: (a) Bar charts showing the number of observations per class. (b) Stacked bar charts showing the distribution of gender for each patient case.

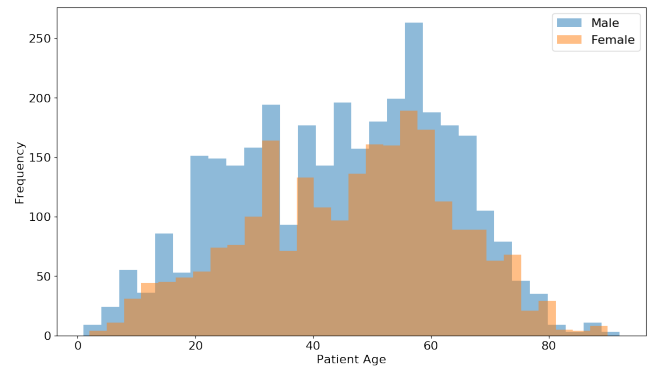


Figure 4: The distribution of patient age and the number of pneumonia cases.

In the conventional ML process, we need to extract and select the feature parameters necessary to give the topological information of the training dataset. This is a very crucial stage that distinguishes it from deep learning techniques. In this work, we adopt the principal component analysis (PCA) approach to extract the relevant principal components from the captured X-ray images.

3 PRINCIPAL COMPONENT ANALYSIS

The *curse of dimensionality* can affect the accuracy of the model classifier since there are more pixel values in high dimensional space. In order to reduce the complexity of this model and prevent

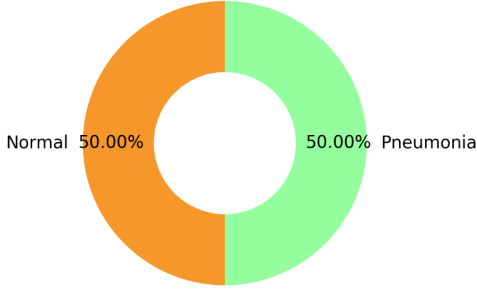


Figure 5: Pie chart showing the data balance.

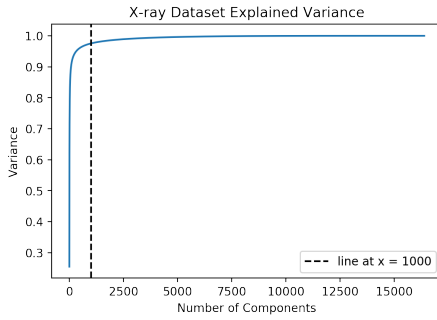


Figure 6: Representation of the cumulative explained variance ratio as a function of the number of components.

over-fitting, we use the PCA method to perform the feature extraction. The following steps present the transformation to select the relevant features:

- Compute the mean vector.
- Construct the training data in a mean-adjusted matrix.
- Generate the covariance matrix.
- Compute the eigenvalues with the corresponding eigenvectors.
- Calculate the basis vectors.
- Denote each sample as a linear combination of basis vectors.

A detailed discussion to derive the algebraic solution and application to PCA is presented in [7, 17]. To perform PCA on the chest X-ray images, we first assume that each image is $m \times m$ pixels and then handle each image as a vector of size m^2 without data loss. Thus, we consider an image as a point in m^2 dimensional space. The curve in Fig. 6 estimates the number of principal components required to depict the data. This is done by computing at the cumulative explained variance ratio against the number of components. We can observe from the graph that the first 1000 components contain $\approx 95\%$ of the variance. For this work, we train the model classifiers discussed in Section 4 with both the 1000 principal components as well as the original (128×128) -dimensional variance to measure their performances.

4 CLASSIFICATION METHODS

The pneumonia detection performance of the proposed method is measured using ten benchmark classifiers. The following bullet points are short descriptions of each model classifier:

- **Logistic Regression (LR):** LR is a classification learning algorithm used to measure the probability of an event. Note that the term event depicts a categorical dependent variable, and the values of the dependent variable can be binary, ordinal, or multinomial. In this work, we focus on binary events where the dataset contains target labels catalogued as 1 (pneumonia state) or 0 (normal state). Recall that to map any real value into a range of $[0, 1]$, we use the Logistic sigmoid function defined in Equation 1:

$$\Lambda_{\omega, \eta}(\theta_k) \stackrel{\text{def}}{=} \frac{1}{\exp[-(\omega\theta_k + \eta)] + 1} \quad (1)$$

where θ_k is a M-dimensional selected features of the training data. The representation $\Lambda_{\omega, \eta}$ denotes that the model Λ has parameters ω and η , where $\omega \in \mathbb{R}^n$ is the M-dimensional weight vector and $\eta \in \mathbb{R}$. If we consider λ_k to be an event, then this can be modeled as a linear function of θ_k . Therefore, using the sigmoid model in Equation 1, we now compute the maximum likelihood (ML) of the training data to get

$$L_{\omega, \eta} \stackrel{\text{def}}{=} \prod_{k=1}^N \Lambda_{\omega, \eta}(\theta_k)^{\lambda_k} (1 - \Lambda_{\omega, \eta}(\theta_k))^{(1-\lambda_k)}. \quad (2)$$

Therefore, if $\lambda_k = 1$, then $(1 - \Lambda_{\omega, \eta}(\theta_k))^{(1-\lambda_k)} = 1$. On the other hand, if $\lambda_k = 0$, then $\Lambda_{\omega, \eta}(\theta_k)^{\lambda_k} = 1$. For a binary case, we model λ_k as a Bernoulli Distribution [13]. Nevertheless, the average logistic loss is computed using the log-likelihood as;

$$\ln L_{\omega, \eta}(\theta_k) = \sum_{k=1}^N \lambda_k \ln \Lambda_{\omega, \eta}(\theta_k) + (1 - \lambda_k) \ln (1 - \Lambda_{\omega, \eta}(\theta_k)), \quad (3)$$

which is a smooth and convex function. The method has been widely used to address issues such as cancer survivability [3], customer churn [14], and football match results [15].

- **K-Nearest Neighbours (KNN):** KNN is one of the simplest classification techniques. The principle of KNN consists of assigning to each unlabeled example the majority class among its k -nearest neighbors in the training set. The Euclidean metric is used to measure the closeness between samples (for continuous variables) in a given dataset.
- **Decision Tree (DTree):** A DTree is a classifier that performs recursive partitions over the space of the instance [6]. A typical decision tree consists of internal nodes, edges and leaf nodes. The internal nodes represent a check on an attribute or a subset of attributes. The edge is a particular value or range of values of the input attributes. Therefore, the combination of an internal node and its associated edge(s) split the instance space into two or more partitions. Every leaf node is a tree terminal node with a class name. Each path from the root node to the leaf node forms a classification

rule. The technique is able to handle non-linearity [1, 5] and the interaction effect between two features [4].

- **Random Forest (RF):** RF is an ensemble learning method used for classification and regression. It creates multiple decision trees on different random subsets of the original data at the time of training. When generating the prediction, it returns the class model (classification) or the mean prediction (regression) of the individual trees. The RF algorithm is described in the following steps:
 - (a) Draw α bootstrap samples from the original data.
 - (b) Build a classification tree for each of the bootstrap experiments such that at each node, instead of choosing the best split between all predictors, randomly select β predictors and choose the best split between those variables.
 - (c) Predict new data by selecting the majority class of the α trees for classification or by computing the average for regression.
- **Multi-layer Perceptron (MLP):** A MLP is an artificial neural network (ANN) belonging to the feedforward class such that the connections are always guided from the lower layers to the upper layers. The design has one input layer, one output layer, and one or more hidden layers between the input and the output layers. The number of neurons in the input layers is the same as the number of variables in the dataset. Also, the number of neurons in the output layers is equal to the number of classes in the dataset. During the training, the weights of the connections are updated to achieve a minimal difference between the output of the network and the expected output.
- **Naive Bayes (NB):** This classification technique uses Bayes' theorem with the condition that the existence of a feature in one domain does not influence the existence of any other feature in the same domain. The main interest in applying this technique is to estimate the posterior probabilities such that the probability of a target (C) given some known features (f_e) is defined as $P(C|f_e) = \frac{P(C)P(f_e|C)}{P(f_e)}$. Here, the posterior probability of a domain is $P(C|f_e)$, the prior probability of a domain is given as $P(C)$, $P(f_e|C)$ is the likelihood function which is the probability of a feature given a domain, and $P(f_e)$ is the probability of a feature.
- **Gradient Boosting Classifier (GBC):** The approach used in GBC is to combine weak models, mostly decision trees in an ensemble in order to improve the final results. The main hyper-parameters to calibrate in GBC in order to improve the model accuracy are the number of trees, the step size at each iteration, and the depth of trees.
- **Support-Vector machine Classifier (SVC):** Its approach consists of creating a boundary between two classes from one or more features that are used for decision making. The boundary of judgment, known as the hyper-plane, is oriented in such a way that it is as far as possible from the nearest data points of each category. The closest points are named support vectors [10].

5 MODEL PERFORMANCE EVALUATION

Once the learning algorithm has been developed using the training dataset, we need to measure how good the model classifier is. We use either the test data or the k-fold cross-validation technique to assess the performance of the model classifier. Here, the classifier's performance is estimated with 10-fold cross-validation. The assessment is generally done using the performance metric classification approach [8, 11]. Note that the metrics for ML classification can be evaluated from the confusion matrix³. For instance, the confusion matrix in Table 1 shows that the row part of the table represents the predicted category whilst the other axis (the column of the table) depicts the original label. TP and TN depict the respective counts for positive and negative cases that are validly identified. FP and FN denote the counts for negative and positive misclassification instances, respectively. Table 2 presents the classification metrics derived from Table 1.

Table 1: Confusion Matrix

		Actual	
		True	False
Predicted	True	True Positive (TP)	False Positive (FP)
	False	False Negative (FN)	True Negative (TN)

Meanwhile, the Receiver Operating Characteristic (ROC) curve is a graphical representation of the TPR against the FPR. It is used to create an equilibrium between the TPs, and FPs. Fig. 7 is an illustration of a ROC curve. It entails four vital points. A(0,0) at the bottom left corner shows that $TPR = 0$ and $FPR = 0$. It implies that the model classifier has no positive classification rather all the valid classifications are negative. C(1,1) at the upper right corner shows that the model classifier validly classified all the positive samples whilst the negative samples are misclassified. D(1,0) at the bottom left corner shows that both the positive and negative samples are invalidly classified. B(0, 1) at the upper left corner is the vice-versa of D(1,0). Note, this denotes an *ideal* classification. Therefore, if a model classifier is located in the bottom right triangle, it is considered to have performed worse than the classifier in the top left triangle.

6 RESULTS AND DISCUSSION

The results reported in Fig. 8 present a comparison of the performance measure when the classification models are trained with 16384 dimensional feature space and 1000 principal components extracted from 16384 dimensional variance. The blue bar charts are the performance metrics produced with no PCA on the chest X-ray images whilst the orange bar charts are produced using PCA. We can observe that, in general, the performance measure of the model classifiers recorded higher scores when we use the principal component of the chest X-ray images. For instance, the SVC classifier (in Fig. 8a) recorded the highest accuracy score of $\approx 90\%$ (using PCA) and $\approx 85\%$ (using no PCA). Also, even though all the model classifiers in the ROC curves (in Fig. 8f) appear in the upper left

³This is a cross-tabulation that gives a rundown on how successful the model classifier is at predicting related examples of the target classes.

Table 2: Classification Metric Evaluation

Metrics	Formula	Measure
Accuracy (acc)	$\frac{TP+TN}{TP+FP+TN+FN}$	acc measures the ratio of the correct predictions to the total number of instances evaluated.
Sensitivity (sn)	$\frac{TP}{TP+FN}$	sn computes the fraction of positive patterns that are correctly classified.
Specificity (sp)	$\frac{TN}{TN+FP}$	sp measures the fraction of negative patterns that are correctly classified.
Precision (p)	$\frac{TP}{TP+FP}$	p measures the positive patterns that are correctly predicted from the total predicted patterns in a positive class.
Recall (r)	$\frac{TP}{TP+FN}$	r measures the fraction of positive patterns in a positive class.
F ₁ Score	$\frac{2pr}{p+r}$	F ₁ Score denotes the harmonic mean between recall and precision values.
False Positive Rate (FPR)	$\frac{FP}{FP+TN}$	FPR is the proportion of the negative samples that were misclassified. Hence, it complements the specificity (that is, 1 - specificity).
True Positive Rate (TPR)	$\frac{TP}{TP+FN}$	TPR = Recall = Sensitivity

triangle, the degree of separability (AUC) for using the PCA is relatively higher than when no dimension reduction is performed. For example, the SVC classifier recorded an AUC score of $\approx 96\%$ (using PCA) and $\approx 92\%$ (using no PCA). Note that the higher the AUC, the better the model classifier is able to distinguish between the target labels. Thus, the classification model is capable of estimating 0s as 0s as well as 1s as 1s. The log loss (in Fig. 8g) that measures the uncertainty of the prediction recorded a lesser score when we train the classification models with the relevant principal components. The SVC score produced in the log loss plots is ≈ 3.35 (using PCA) and ≈ 5.22 (using no PCA). Nonetheless, the computation time used during the classification model training process for using the PCA is less than that of the high dimensional space. For example, the

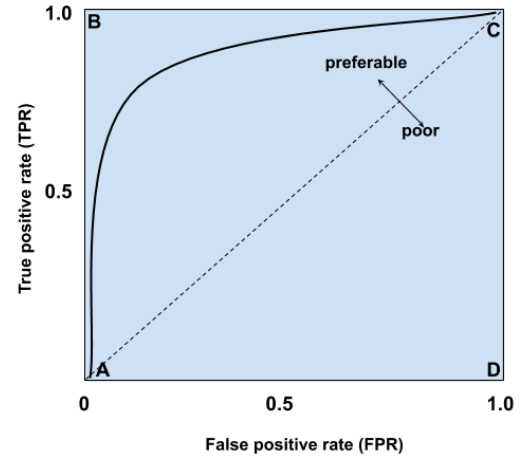


Figure 7: A ROC curve displaying the principal points for an equally recorded samples. This graph is reproduced from [20].

time taken to train the SVC classifier is ≈ 204 seconds (using PCA) and ≈ 6219 seconds (using no PCA)⁴.

In general, the SVC recorded the overall best performance with an AUC score of $\approx 96\%$, followed by the Gradient Boosting Classifier GBC with an area of $\approx 94\%$ (both cases, using PCA).

7 CONCLUSION

In this work, we have discussed how we use classification learning algorithms to detect pneumonia. The main challenge we face in using this technique is the curse of dimensionality, which arises when working with high dimensional space data. In this paper, we discussed how PCA is used to reduce the high dimension of the X-ray images and also, to extract the critical features for the model classification. Some of the key results we obtained are:

- The overall model performance for using PCA out-performed that of the original space approach.
- The selected principal components chosen contain $\approx 95\%$ of the explained variance.
- The point to note regarding PCA is that it 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 as we can observe in Figs. 8a, 8b, 8c, 8d, 8f and, 8g. This is because the model classifier (GNB) assumes the features are conditionally independent.

⁴The models have been trained using a CPU server of 48 units with a RAM capacity of 200GB. Each model were trained using 10 parallel units.

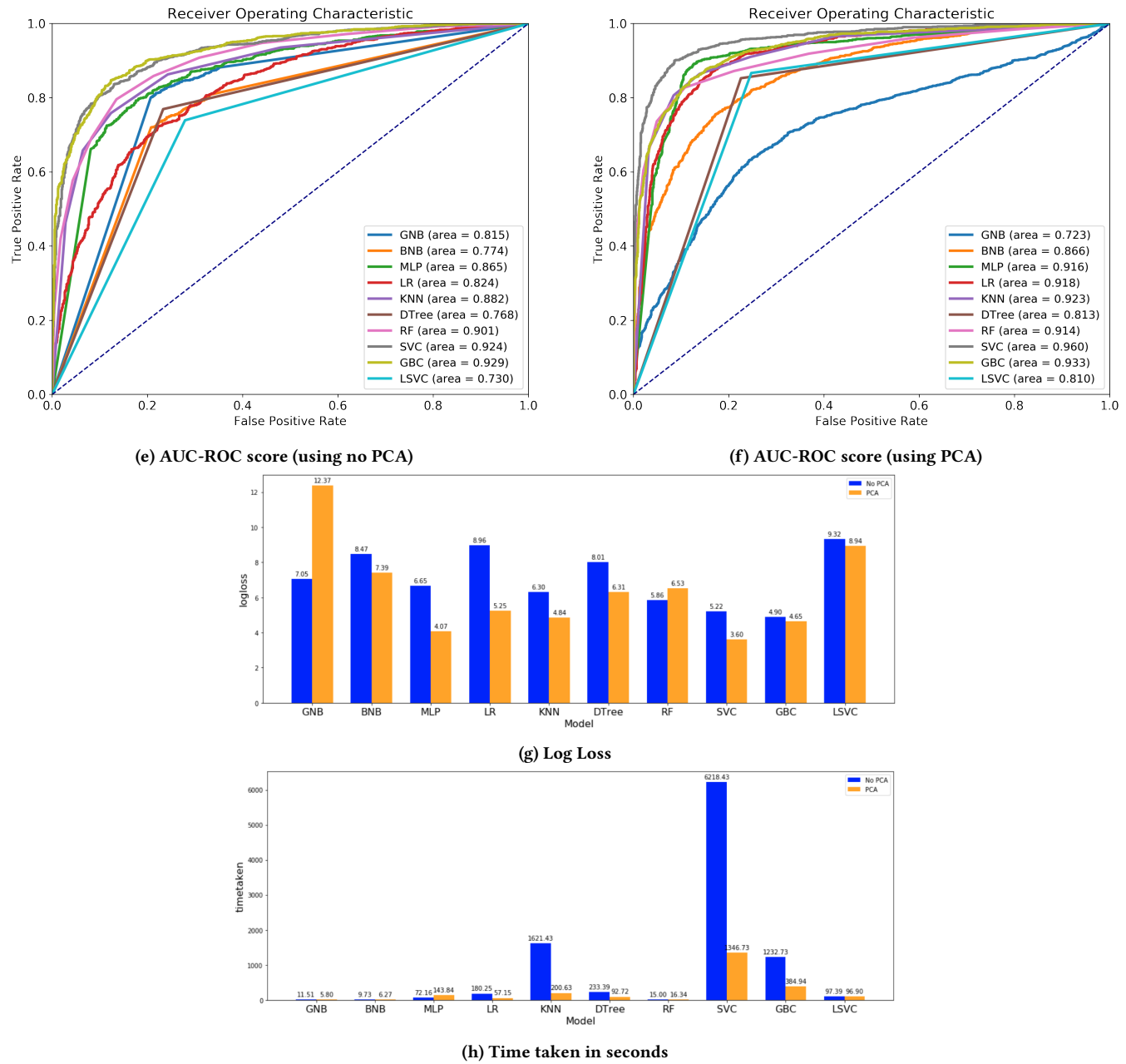
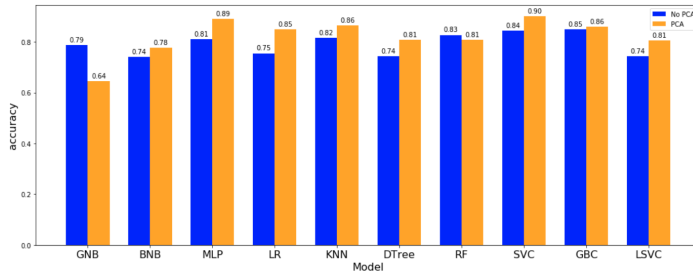


Figure 8: Model Performance Evaluation. The bar plots in blue are the performance metrics for using 16384 dimensional feature space whilst the ones in orange are the performance metrics for using 1000 principal components as the feature space. *a* and *e* are graphical representations of the true positive rate (TPR) against the false positive rate (FPR) for each classification model using no PCA and PCA on the X-rays images, respectively.

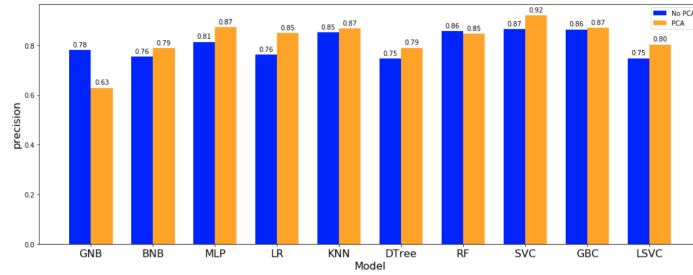
- The SVC classifier recorded the highest AUC score of $\approx 96\%$ (for using PCA) and $\approx 92\%$ (for not using PCA) for the pneumonia detection.

In summary, pneumonia detection can also be estimated with shallow ML classification algorithms. Meanwhile, the reduction of the high dimension of data like in this case is very crucial in

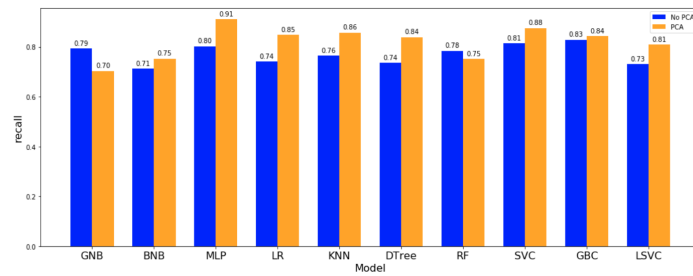
ML. PCA has shown to be a robust technique for dimensionality reduction. In addition, the PCA technique can be used to manipulate feature selection as well as improve the performance of the model classifier, as we have reported in this work. An extension of this work is to use the deep learning approach to extract the features



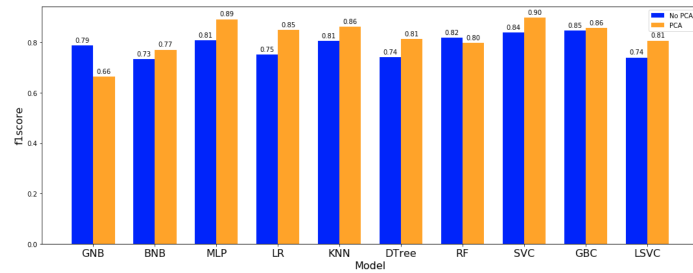
(a) Accuracy classification score



(b) Precision classification score



(c) Recall score



(d) f1 score

from the chest radiographic images and then use that to train and test the conventional ML classifiers.

8 ACKNOWLEDGEMENTS

T. Ansah-Narh is very grateful to the Data Science Intensive (DSI) Program, organised by the African Institute for Mathematical Sciences (AIMS) South Africa, in conjunction with SA-DISCNet. M. Atemkeng acknowledges support from Rhodes University. F.M.J. Ebiele and S.R.T. Djokap are grateful to the African Institute for Mathematical Sciences (AIMS) Rwanda for the use of their server.

REFERENCES

- [1] S. Aich, K. Younga, K. L. Hui, A. A. Al-Absi, and M. Sain. A nonlinear decision tree based classification approach to predict the parkinson's disease using different feature sets of voice data. In *2018 20th International Conference on Advanced Communication Technology (ICACT)*, pages 1–2, Feb 2018.
- [2] Ivo M Baltruschat, Hannes Nickisch, Michael Grass, Tobias Knopp, and Axel Saalbach. Comparison of deep learning approaches for multi-label chest x-ray classification. *Scientific reports*, 9(1):1–10, 2019.
- [3] M. Bozorgi, K. Taghva, and A. Singh. Cancer survivability with logistic regression. In *2017 Computing Conference*, pages 416–420, July 2017.
- [4] D. Braha and A. Shmilovici. On the use of decision tree induction for discovery of interactions in a photolithographic process. *IEEE Transactions on Semiconductor Manufacturing*, 16(4):644–652, Nov 2003.
- [5] K. A. Crockett, Z. Bandar, and A. Al-Attar. Soft decision trees: a new approach using non-linear fuzzification. In *Ninth IEEE International Conference on Fuzzy Systems. FUZZ- IEEE 2000 (Cat. No.00CH37063)*, volume 1, pages 209–215 vol.1, May 2000.
- [6] Wei Dai and Wei Ji. A mapreduce implementation of c4. 5 decision tree algorithm. *International journal of database theory and application*, 7(1):49–60, 2014.
- [7] Felipe L. Gewers, Gustavo R. Ferreira, Henrique F. de Arruda, Filippi N. Silva, Cesar H. Comin, Diego R. Amancio, and Luciano da F. Costa. Principal Component Analysis: A Natural Approach to Data Exploration. *arXiv e-prints*, page arXiv:1804.02502, Apr 2018.
- [8] Qiong Gu, Li Zhu, and Zhihua Cai. Evaluation measures of the classification performance of imbalanced data sets. In *International symposium on intelligence computation and applications*, pages 461–471. Springer, 2009.
- [9] Stephen R.C. Howie, Davidson H. Hamer, and Stephen M. Graham. Pneumonia. In Stella R. Quah, editor, *International Encyclopedia of Public Health (Second Edition)*, pages 500 – 508. Academic Press, Oxford, second edition edition, 2017.
- [10] Shujun Huang, Nianguang Cai, Pedro Penzuti Pacheco, Shavira Narrandes, Yang Wang, and Wayne Xu. Applications of support vector machine (svm) learning in cancer genomics. *Cancer Genomics-Proteomics*, 15(1):41–51, 2018.
- [11] Yasen Jiao and Pufeng Du. Performance measures in evaluating machine learning based bioinformatics predictors for classifications. *Quantitative Biology*, 4(4):320–330, 2016.
- [12] Can Jozef Saul, Deniz Yagmur Urey, and Can Doruk Taktakoglu. Early Diagnosis of Pneumonia with Deep Learning. *arXiv e-prints*, page arXiv:1904.00937, Apr 2019.
- [13] Masaaki Okabe, Jun Tsuchida, and Hiroshi Yadohisa. F-measure Maximizing Logistic Regression. *arXiv e-prints*, page arXiv:1905.02535, May 2019.
- [14] Peng Li, Siben Li, Tingting Bi, and Yang Liu. Telecom customer churn prediction method based on cluster stratified sampling logistic regression. In *International Conference on Software Intelligence Technologies and Applications International Conference on Frontiers of Internet of Things 2014*, pages 282–287, Dec 2014.
- [15] D. Prasetyo and D. Harlili. Predicting football match results with logistic regression. In *2016 International Conference On Advanced Informatics: Concepts, Theory And Application (ICAICTA)*, pages 1–5, Aug 2016.
- [16] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brand on Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, Matthew P. Lungren, and Andrew Y. Ng. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. *arXiv e-prints*, page arXiv:1711.05225, Nov 2017.
- [17] Jonathon Shlens. A tutorial on principal component analysis, 2014.
- [18] Raheel Siddiqi. Automated pneumonia diagnosis using a customized sequential convolutional neural network. In *Proceedings of the 2019 3rd International Conference on Deep Learning Technologies*, pages 64–70, 2019.
- [19] Okeke Stephen, Mangal Sain, Uchenna Joseph Maduh, and Do-Un Jeong. An efficient deep learning approach to pneumonia classification in healthcare. *Journal of healthcare engineering*, 2019, 2019.
- [20] Alaa Tharwat. Classification assessment methods. *Applied Computing and Informatics*, 2018.
- [21] Paul H. Wise. Infant mortality. In Stella R. Quah, editor, *International Encyclopedia of Public Health (Second Edition)*, pages 216 – 221. Academic Press, Oxford, second edition edition, 2017.