PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

COVID-19 detection from lung ultrasound images

Melisa Mateu, Jimena Olveres Montiel, Boris Escalante-Ramírez

Melisa Mateu, Jimena Olveres Montiel, Boris Escalante-Ramírez, "COVID-19 detection from lung ultrasound images," Proc. SPIE 12138, Optics, Photonics and Digital Technologies for Imaging Applications VII, 121380C (17 May 2022); doi: 10.1117/12.2621962

SPIE.

Event: SPIE Photonics Europe, 2022, Strasbourg, France

COVID-19 detection from lung ultrasound images

Melisa Mateu[a], Jimena Olveres[b,c], and Boris Escalante-Ramírez[b,c]

[a]Posgrado en Ingeniería, Universidad Nacional Autónoma de México, Mexico City, Mexico
[b]Facultad de Ingenería, Universidad Nacional Autónoma de México, Mexico City, Mexico
[c]Centro de Estudios en Computación Avanzada, Universidad Nacional Autónoma de México, Mexico City, Mexico

ABSTRACT

Early-stage detection of Coronavirus Disease 2019 (COVID-19) is crucial for patient medical attention. Since lungs are the most affected organs, monitoring them constantly is an effective way to observe sickness evolution. The most common technique for lung-imaging and evaluation is Computed Tomography (CT). However, its costs and effects over human health has made Lung Ultrasound (LUS) a good alternative. LUS does not expose the patient to radiation and minimizes the risk of contamination. Also, there is evidence of a relation between different artifacts on LUS and lung's diseases coming from the pleura, whose abnormalities are related with most acute respiratory disorders. However, LUS often requires an expert clinical interpretation that may increase diagnosis time or decrease diagnosis performance. This paper describes and compares machine learning classification methods namely Naive Bayes (NB) Support Vector Machine (SVM), K-Nearest Neighbor (K-NN) and Random Forest (RF) over several LUS images. They obtain a classification between lung images with COVID-19, pneumonia, and healthy patients, using image's features previously extracted from Gray Level Co-Occurrence Matrix (GLCM) and histogram's statistics. Furthermore, this paper compares the above classic methods with different Convolutional Neural Networks (CNN) that classifies the images in order to identify these lung's diseases.

Keywords: Machine learning, deep learning, comparision, metrics.

1. INTRODUCTION

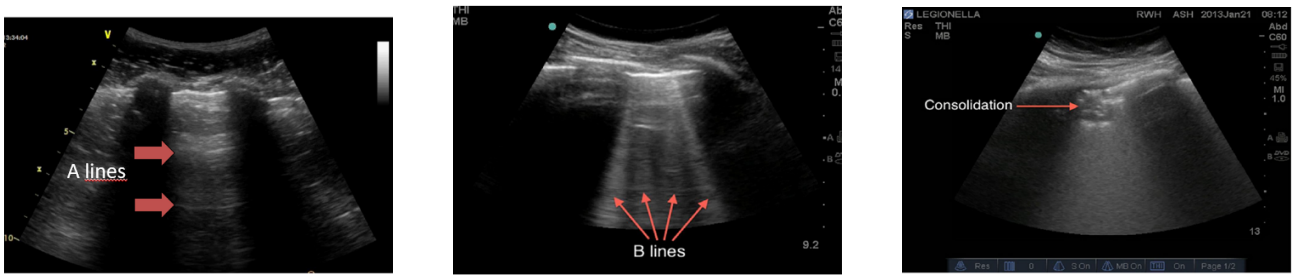
COVID-19 is an infectious disease caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-Cov-2). It is characterized by severe pneumonia or acute respiratory distress syndrome (ARDS) in about 20% of infected patients with high morbidity and mortality.[1,2] Radiological imaging techniques such as thoracic computed tomography (CT) is used for early screening, diagnosis and treatment of patients with suspected or confirmed COVID-19 infections.[3] Although CT has excellent ability to detect lung's changes, it has disadvantages of large radiation exposure, lack of portability for bedside imaging and risk of cross infections between patients.[1]

On the other hand, lung ultrasound (LUS) is an emerging non-invasive bedside technique in the diagnosis of interstitial lung syndrome. LUS has provided the physical bases and patterns in COVID-19 patients, suggesting that it can be a useful tool to diagnose and monitor this sickness.[2,4] The main LUS findings in COVID-19 are B lines, which are represented by vertical hyperechoic artifacts that depart from the pleura to the bottom of the screen,[5,6] as shown in Fig. 1b. The presence of B lines suggests ARDS and can be related to COVID-19. Besides B lines, LUS has more artifacts that are related to the majority of acute respiratory disorders because they involve the pleura. For these reasons some artifacts can be used to detect potential lung's conditions. For instance A Lines can be related to healthy patients. These type of lines are motionless and regularly spaced lines horizontal to the pleura,[2,6] as shown in Fig. 1a. In addition, consolidations are another artifact that can be related to pneumonia.[6] The echo structure of the lung itself becomes visible with characteristic air inside the alveolus or surrounded by inflammation or pus,[1] as shown in Fig. 1c.

These different LUS artifacts allow to study medical images in order to identify possible patterns that may lead to the automatic diagnosis of the disease. Machine learning (ML) and deep learning (DL) have become

Further author information: (Send correspondence to M.M.)
M.M.: E-mail: mateu.melisa@gmail.com

Optics, Photonics and Digital Technologies for Imaging Applications VII, edited by
Peter Schelkens, Tomasz Kozacki, Proc. of SPIE Vol. 12138, 121380C



(a) (b) (c)

Figure 1: Artifacts of LUS: (a) A lines, (b) B lines and (c) consolidations[1]

established disciplines in applying artificial intelligence to mine, analyze, and recognize patterns from data. The advances of those techniques can benefit clinical decision making. As new data emerges, the use of computer-aided systems is becoming increasingly applied in clinical settings.[7]

The purpose of this paper is to compare the supervised machine learning classification methods trained by several LUS images. By using the images features extracted from Gray Level Co-Occurrence Matrix (GLCM) and histogram's statistics these algorithms can classify the lung images of patients with COVID-19, pneumonia, and healthy ones. These ML methods were compared with different Convolutional Neural Networks (CNN) that also classified the same images in order to identify lung diseases. Lastly, this paper evaluates ML and DL methods using metrics like accuracy, precision and recall to identify the best method for the classification task.

2. METHODS

2.1 Dataset

The LUS image database was constructed from two different datasets. Firstly from a Kaggle's repository[8] with 988 images of COVID-19, 731 images of pneumonia and 1,276 images of patients with no affections. The second dataset was collected from the Butterfly's COVID-19 gallery,[9] it consisted of LUS videos; eight of COVID-19, nine of pneumonia and two of normal lungs. After we extracted the frames from each video, the dataset consisted of 1,562 COVID-19 images, 1,705 pneumonia's images and 832 images of normal lungs. A random selection of 700 images per class; COVID-19, pneumonia and regular patients, was generated from the combination of both datasets. These 2,100 images were used for the training and evaluation of machine and deep learning methods.

2.2 Features

The machine learning algorithms consisted of several features extracted from the images on the database. The images were first resized to a resolution of 254×254 pixels, then the features were extracted and finally different feature selection methods were tested when training the ML algorithms.

2.2.1 Extraction

The features extracted for each image consisted of GLCM and histogram's statistics. For GLCM 25 patches were extracted per image and to each patch six statistics were calculated; dissimilarity, correlation, energy, contrast, homogeneity and angular second moment (ASM). In total one image had 150 textural features. In addition, there were five statistics calculated per histogram: mean, standard deviation, entropy, skewness and kurtosis. Therefore, each image had a total of 155 features used to classify them.

2.2.2 Selection

Due to the large amount of features per image, feature selection methods were used to determine the most important features for the ML classification task. Univariate feature selection method picked ten features, the same number of features was selected by the principal component analysis (PCA) and six features were retained with the feature importance method.

2.3 Machine Learning

We trained four classic supervised methods: Naive Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbor (K-NN) and Random Forest (RF). All 155 features and also the sets of reduced ones, mentioned in the Sec. 2.2.2, were used separately to train these ML algorithms for a three-classes classification task. To evaluate these method's performance, the metrics from Sec. 2.5 were calculate for each one of them. The results are shown in Sec. 3.

2.4 Deep Learning

Using the same resolution as in Sec. 2.2 three CNN's were trained. Transfer learning techniques were used to learn the classification task of COVID-19, pneumonia and regular images. Adam optimizer was used with a learning rate of $1E-3$ for each CNN. Cross Entropy was used as the loss function.

The trained CNN were ResNet18, MobilNet V2 and GoogleNet. They were evaluated with the metrics from Sec. 2.5 and the results are shown in Sec. 3.

2.5 Metrics

To evaluate categorization predictions in deep and machine learning methods, three types of metrics were computed. The accuracy determines how correct the diseases classifications were, precision shows how many of the predictions were correct and recall indicates how correctly the disease was recognized.

The values obtained and their comparison are shown in the tables 1 - 4 on the Sec. 3.

3. RESULTS

3.1 Machine Learning

For accuracy training results, different values can be observed depending on the machine learning method used, as shown in Fig. 2 and in table 1.

Also accuracy values are the same from almost all the methods independently of the feature selection technique that was used. The only difference between values comes with the Random Forest method but the difference is not significant. The RF method achieved the highest accuracy value compared to the other ML methods.

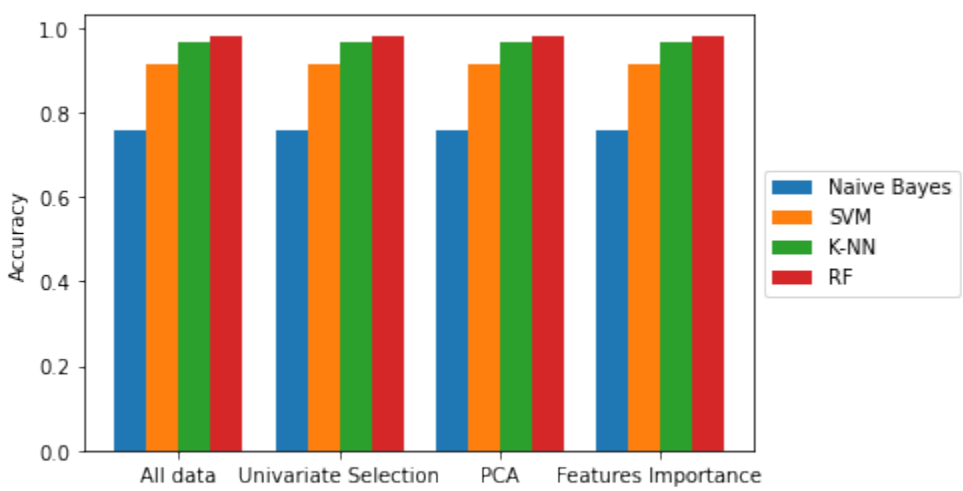


Figure 2: Accuracy comparison between the different methods of machine learning and feature selection

Figures 3 and 4 show a comparison between precision and recall values obtained for every class. These values are shown in table 2

Table 1: Accuracy comparison between ML and feature selection methods

Machine Learning Methods	**Features - Accuracy**			
	All data	Univariate Selection	PCA	Feature Importance
Naive Bayes	75.67	75.67	75.67	75.67
SVM	91.38	91.38	91.38	91.38
K-NN	96.62	96.62	96.62	96.62
Random Forest	98.10	**98.24**	**98.24**	98.19

Table 2: Precision (P) and Recall (R) comparison between ML and features selection methods

Class	**All data**				**Class**	**Univariate Selection**			
	NB	*SVM*	*K-NN*	*RF*		*NB*	*SVM*	*K-NN*	*RF*
COVID-19	P: 0.72 R: 0.76	P: 0.89 R: 0.89	P: 0.95 R: 0.97	P: 0.98 R: 0.98	*COVID-19*	P: 0.72 R: 0.76	P: 0.89 R: 0.89	P: 0.95 R: 0.97	P: 0.97 R: 0.99
Pneumonia	P: 0.83 R: 0.66	P: 0.96 R: 0.90	P: 0.99 R: 0.94	P: 0.98 R: 0.97	*Pneumonia*	P: 0.83 R: 0.66	P: 0.96 R: 0.90	P: 0.99 R: 0.94	P: 0.98 R: 0.97
Regular	P: 0.73 R: 0.84	P: 0.88 R: 0.94	P: 0.95 R: 0.98	P: 0.98 R: 0.97	*Regular*	P: 0.73 R: 0.84	P: 0.88 R: 0.94	P: 0.95 R: 0.98	P: 0.97 R: 0.98
Class	**PCA**				**Class**	**Feature Importance**			
	NB	*SVM*	*K-NN*	**RF**		*NB*	*SVM*	*K-NN*	*RF*
COVID-19	P: 0.79 R: 0.76	P: 0.89 R: 0.89	P: 0.95 R: 0.97	**P: 0.98** **R: 0.98**	*COVID-19*	P: 0.72 R: 0.76	P: 0.89 R: 0.89	P: 0.95 R: 0.97	P: 0.98 R: 0.99
Pneumonia	P: 0.83 R: 0.66	P: 0.96 R: 0.90	P: 0.99 R: 0.94	**P: 0.98** **R: 0.97**	*Pneumonia*	P: 0.83 R: 0.66	P: 0.96 R: 0.90	P: 0.99 R: 0.94	P: 0.98 R: 0.97
Regular	P: 0.73 R: 0.84	P: 0.88 R: 0.94	P: 0.95 R: 0.98	**P: 0.98** **R: 0.98**	*Regular*	P: 0.73 R: 0.84	P: 0.88 R: 0.94	P: 0.95 R: 0.98	P: 0.97 R: 0.97

According to the comparisons the best ML method in all the feature selection methods was RF, it shows higher values than NB, SVM and K-NN in all the cases. Taking RF as the best ML method we can observe that the highest precision and recall was achieved with the PCA method. Thus, considering all metrics (accuracy, recall and precision) the best performance is PCA - Random Forest combination.

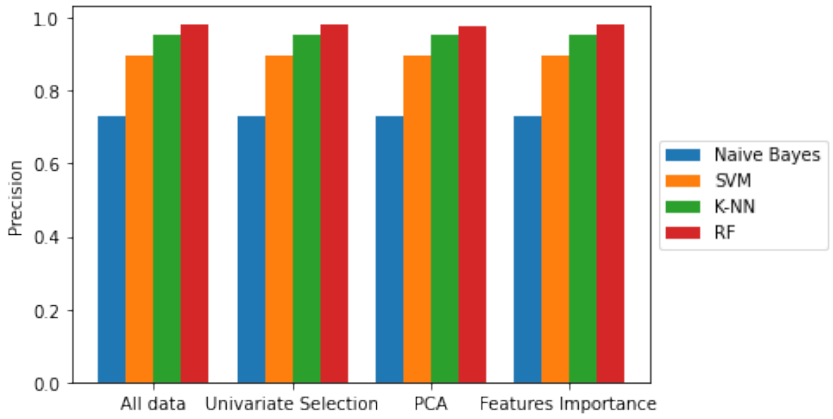
3.2 Deep Learning

Table 3 shows a comparison between the DL methods trained. All three displayed satisfactory accuracy above 0.90. In precision and recall the ResNet 18 architecture showed the worst performance compared to the other two. Although MobilNet V2 and GoogleNet have the same accuracy value, their precision and recall values vary. The best performance of the three metrics was achieved by Google Net architecture.

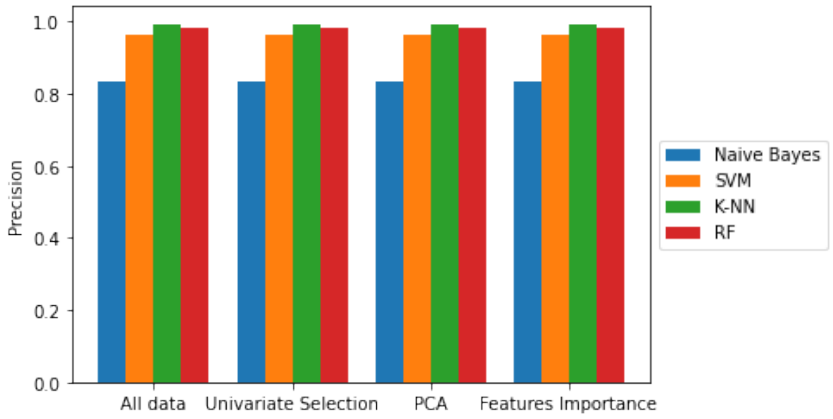
Table 3: Comparison between DL methods

Metric	**CNN architecture**		
	ResNet 18	*MobilNet V2*	*GoogleNet*
Accuracy	0.91	0.94	0.94
Precision - Recall			
COVID-19	0.89 - 1.00	1.00 - 0.92	1.00 - 1.00
Pneumonia	0.93 - 0.93	1.00 - 0.88	0.93 - 0.93
Regular	0.89 - 0.80	0.85 - 1.00	0.90 - 0.90

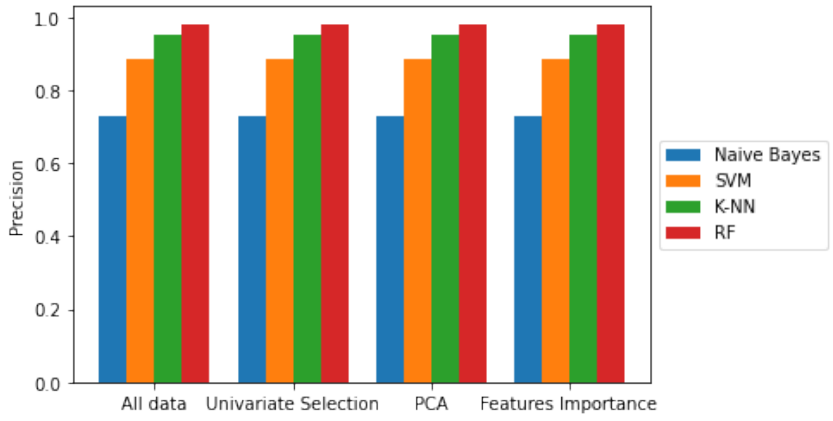
The loss and accuracy curves are shown in Figs. 5, 6 and 7. All loss curves coincide in a fast decay and



(a) COVID-19

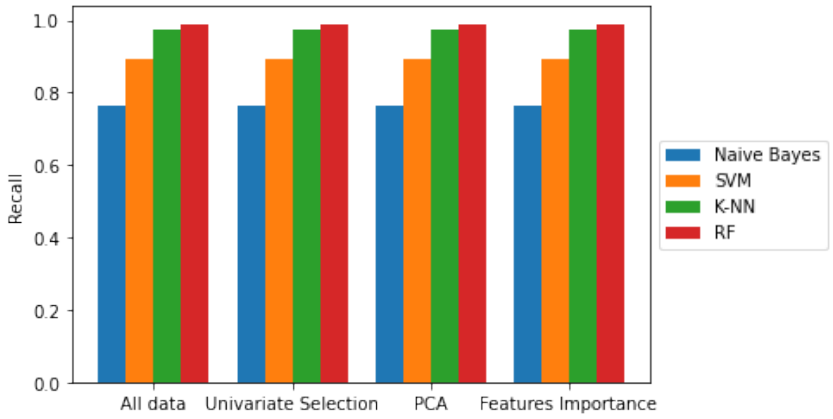


(b) Pneumonia

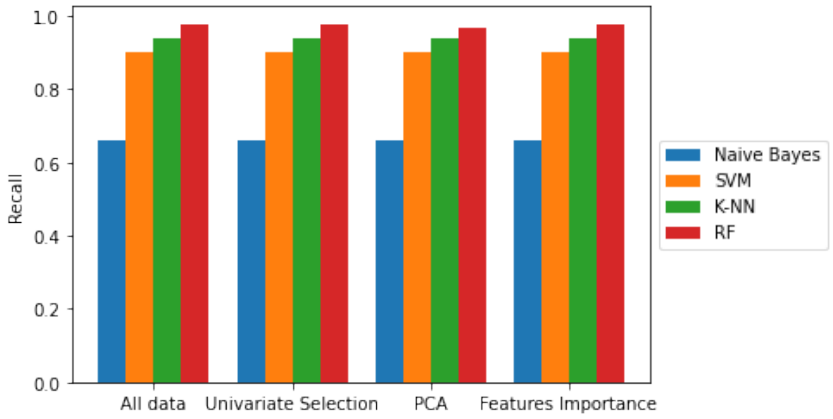


(c) Regular

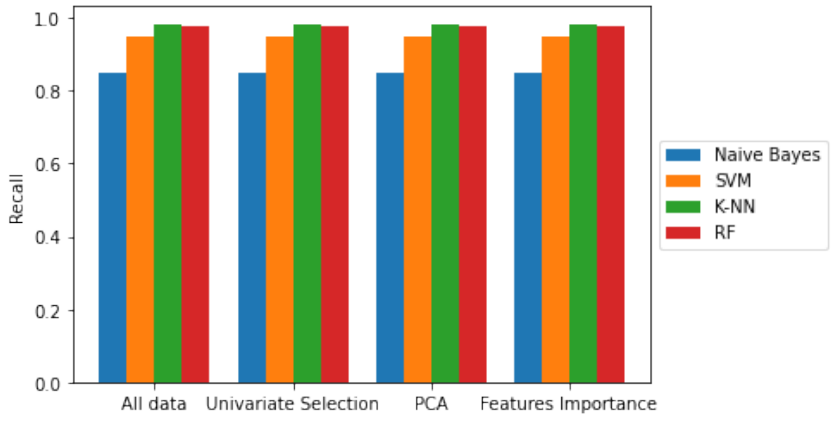
Figure 3: Precision comparison between the different methods of ML and feature selection



(a) COVID-19

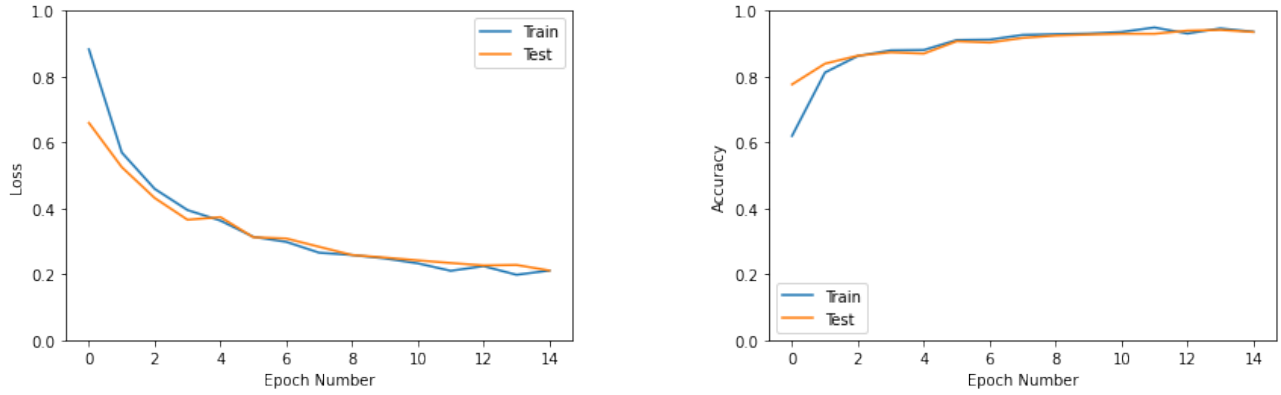


(b) Pneumonia



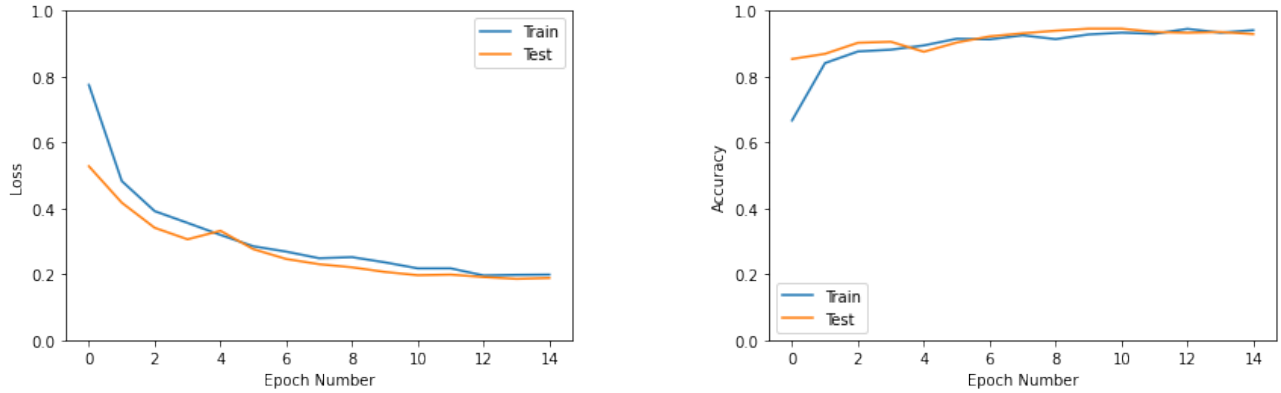
(c) Regular

Figure 4: Recall comparison between the different methods of ML and feature selection



(a) (b)

Figure 5: ResNet 18, (a) loss curve and (b) accuracy



(a) (b)

Figure 6: MobilNet V2, (a) loss curve and (b) accuracy

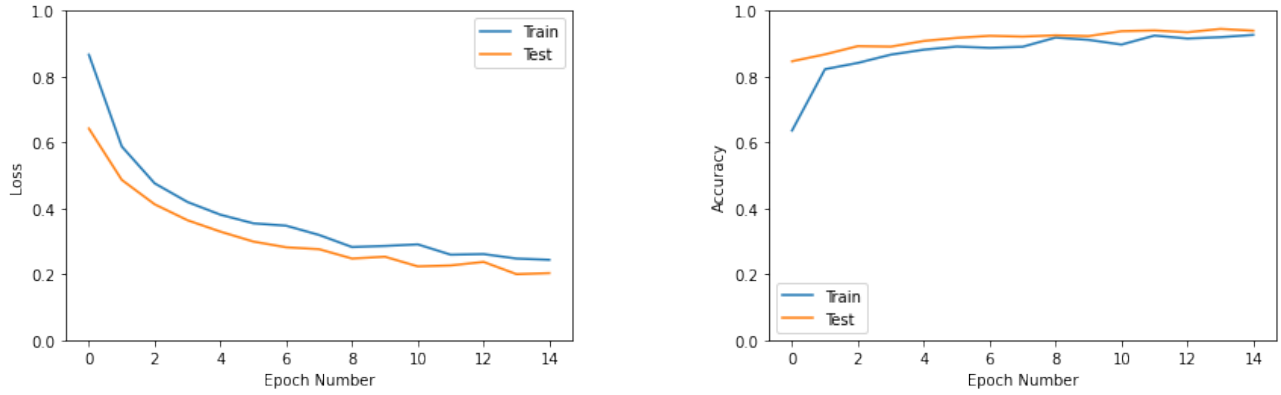
converge on the training and testing sets. Also, the accuracy curves show that high levels of accuracy are achieved in very few epochs. Their performances, in loss and accuracy, suggest that this classification task is easily achieved for the DL methods.

3.3 Comparison

In table 4 all metrics for ML and DL methods obtained are compared. The ML method selected for the comparison corresponds to the PCA feature selection since it showed the highest performance in the metrics evaluated as seen in tables 1 and 2.

Table 4: Comparison between DL and ML methods

Metric	**Machine Learning Methods**				**Deep Learning Methods**		
	NB	*SVM*	*K-NN*	**RF**	*ResNet18*	*MobilNet V2*	*GoogleNet*
Accuracy	0.75	0.91	0.96	**0.98**	0.91	0.94	0.94
Precision - Recall							
COVID-19	0.72 - 0.76	0.89 - 0.89	0.95 - 0.97	**0.98 - 0.98**	0.89 - 1.00	1.00 - 0.92	1.00 - 1.00
Pneumonia	0.83 - 0.66	0.96 - 0.90	0.99 - 0.94	**0.98 - 0.97**	0.93 - 0.93	1.00 - 0.88	0.93 - 0.93
Regular	0.73 - 0.84	0.88 - 0.94	0.95 - 0.98	**0.98 - 0.98**	0.89 - 0.80	0.85 - 1.00	0.90 - 0.90



(a) (b)

Figure 7: GoogleNet, (a) loss curve and (b) accuracy

The all methods comparison table (table 4) evidenced that Random Forest presents the best accuracy performance and very high precision and recall values, showing little differences of ± 0.01 between them. In comparative, the remaining method's performance values vary with bigger differences, showing they can classify better some classes but are more susceptible to fail in others.

4. CONCLUSION

As the results show, ML and DL were able to successfully complete the three-classes classification task. However, ML methods have several advantages over DL ones. First, less computational power is required to train ML algorithms. For DL methods a GPU is needed whereas only a CPU is required on ML methods. Additionally, less data is required to train ML methods. ML algorithms were able to achieve good metrics with less LUS images than the DL algorithms.

In brief, artificial intelligence methods can be used to analyze and classify large amounts of medical information which can be useful as it can help doctors with diagnosis or treatments. This study has demonstrated that ML and DL methods are useful tools in a clinical setting. The approach for classifying LUS images of patients with COVID-19, pneumonia and without affections can be performed with a variety of machine learning and deep learning methods. Despite both methods achieved the task successfully, ML methods are easier to train as they require less computational power and can obtain good results with a smaller database.

ACKNOWLEDGMENTS

This work has been sponsored by UNAM PAPIIT grants TA101121, IV100420 and Consejo Nacional de Ciencia y Tecnología (CONACYT)

REFERENCES

[1] Sultan, L. R. and Sehgal, C. M., "A review of early experience in lung ultrasound in the diagnosis and management of covid-19," *Ultrasound in Medicine and Biology* **46**, 2530–2545 (9 2020).

[2] Allinovi, M., Parise, A., Giacalone, M., Amerio, A., Delsante, M., Odone, A., Franci, A., Gigliotti, F., Amadasi, S., Delmonte, D., Parri, N., and Mangia, A., "Lung ultrasound may support diagnosis and monitoring of covid-19 pneumonia," *Ultrasound in Medicine and Biology* **46**, 2908–2917 (11 2020).

[3] Bakheet, S. and Al-Hamadi, A., "Automatic detection of covid-19 using pruned glcm-based texture features and ldcrf classification," *Computers in Biology and Medicine* **137**, 104781 (10 2021).

[4] Musolino, A. M., Supino, M. C., Buonsenso, D., Ferro, V., Valentini, P., Magistrelli, A., Lombardi, M. H., Romani, L., D'Argenio, P., and Campana, A., "Lung ultrasound in children with covid-19: Preliminary findings," *Ultrasound in Medicine and Biology* **46**, 2094–2098 (8 2020).

[5] Sofia, S., Boccatonda, A., Montanari, M., Spampinato, M., Cocco, G., Accogli, E., Cipollone, F., and Schiavone, C., "Thoracic ultrasound and sars-covid-19: a pictorial essay," *Journal of Ultrasound* **23**, 217–221 (2020). ARDS - Sindrome de dificultad respiratoria aguda.

[6] Serafino, M. D., Notaro, M., Rea, G., Iacobellis, F., Paoli, V. D., Acampora, C., Ianniello, S., Brunese, L., Romano, L., and Vallone, G., "The lung ultrasound: facts or artifacts? in the era of covid-19 outbreak," *Radiologia Medica* **125**, 738–753 (8 2020).

[7] Apostolopoulos, I. D. and Mpesiana, T. A., "Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks," **43**, 635–640 (2020).

[8] "covid19 ultrasound original dataset." (Accessed: 8 june 2021).

[9] "Covid-19 lung ultrasound imaging: Fighting together." (Accessed: 16 february 2022).