

A machine learning approach on chest X-rays for pediatric pneumonia detection

Natali Barakat¹, Mahmoud Awad²  and Bassam A Abu-Nabah³

Abstract

Background: According to the World Health Organization (WHO), pneumonia is the leading infectious cause of death in children below 5 years old. Hence, the early detection of pediatric pneumonia is crucial to reduce its morbidity and mortality rates. Even though chest radiography is the most commonly employed modality for pneumonia detection, recent studies highlight the existence of poor interobserver agreement in the chest X-ray interpretation of healthcare practitioners when it comes to diagnosing pediatric pneumonia. Thus, there is a significant need for automating the detection process to minimize the potential human error. Since Artificial Intelligence tools such as Deep Learning (DL) and Machine Learning (ML) have the potential to automate disease detection, many researchers explored how such tools can be implemented to detect pneumonia in chest X-rays. Notably, the majority of efforts tackled this problem from a DL point of view. However, ML has shown a higher potential for medical interpretability while being less computationally demanding than DL.

Objective: The aim of this paper is to automate the early detection process of pediatric pneumonia using ML as it is less computationally demanding than DL.

Methods: The proposed approach entails performing data augmentation to balance the classes of the utilized dataset, optimizing the feature extraction scheme, and evaluating the performance of several ML models. Moreover, the performance of this approach is compared to a TL benchmark to evaluate its candidacy.

Results: Using the proposed approach, the Quadratic SVM model yielded an accuracy of 97.58%, surpassing the accuracies reported in the current ML literature. In addition, this model classification time was significantly smaller than that of the TL benchmark.

Conclusion: The results strongly support the candidacy of the proposed approach in reliably detecting pediatric pneumonia.

Keywords

Machine learning, chest X-rays, pediatric pneumonia detection, statistical feature extraction, healthcare

Submission date: 4 January 2023; Acceptance date: 11 May 2023

Introduction

Recent technological advancements in the healthcare industry have contributed to the digitization and storage of medical health records in large databases.^{1–3} Consequently, numerous efforts have been utilized to investigate the potential of using AI tools in the era of big data to assist medical practitioners in making more informed diagnostic decisions using available medical databases to ensure high patient care quality.^{4,5} The use of AI in the healthcare industry is promising because it

¹Engineering Systems Management Department, American University of Sharjah, College of Engineering, Sharjah, United Arab Emirates

²Industrial Engineering Department, American University of Sharjah, College of Engineering, Sharjah, United Arab Emirates

³Mechanical Engineering Department, American University of Sharjah, College of Engineering, Sharjah, United Arab Emirates

Corresponding author:

Mahmoud Awad, Industrial Engineering Department, American University of Sharjah, College of Engineering, Sharjah 26666, United Arab Emirates.
Email: miawad@aus.edu



can aid in preventing, treating, and diagnosing illnesses with expert-level accuracy while decreasing human error.⁶ To date, AI tools have been shown to be effective in detecting several diseases such as skin cancer,⁷ lung cancer,⁸ breast cancer,⁹ heart diseases,^{10,11} eye diseases,¹² tuberculosis,¹³ covid-19,¹⁴ and pneumonia.^{15–19} Within the scope of pneumonia detection, the implementation of AI has been investigated using multiple medical modalities including clinical data, computed tomography, ultrasounds, and chest X-rays.¹⁶

Notably, pneumonia is a critical respiratory disease that limits a person's necessary oxygen intake by filling the lung's air-sacs, also known as the alveoli, with fluid. This disease is caused by bacteria, viruses, or fungi and may be fatal if left untreated. According to the World Health Organization (WHO),²⁰ despite being curable, pneumonia is the “single largest infectious cause of death in children worldwide” where it accounted for 15% of the fatalities of children under the age of five in 2017, making its early detection essential to reduce the mortality rate. Currently, chest radiography is the most commonly utilized imaging modality for pediatric pneumonia detection.^{21,22} However, the work of Voigt et al.²³ proves that there is a poor interobserver agreement between radiologists when it comes to diagnosing pediatric pneumonia using the same chest X-ray images. As such, the authors recommend standardizing the pneumonia detection process and setting up compulsory training programs to reduce the high interobserver variability. Furthermore, in several low- to middle-income countries, pediatric pneumonia is diagnosed in chest X-rays by non-radiologist clinicians,²¹ and Fawole et al.²⁴ confirm the existence of variability in such clinicians' diagnosis. In addition, the authors suggest that even though training interventions have the potential to reduce the diagnosis variability of pediatric pneumonia in chest X-rays in the short run, further studies must be conducted to monitor whether such progress can be retained in the long run.

Thus, it can be deduced that there is a significant need for an automated approach that can enable medical practitioners to increase the reliability and accuracy of their pediatric pneumonia diagnosis in chest X-ray images, especially in under-developed countries such as those present in sub-Saharan Africa and South Asia where pneumonia is most widespread.²⁰ In response to this, numerous efforts have been utilized by researchers to evaluate the candidacy of Artificial Intelligence (AI) tools, such as Deep Learning (DL) and Machine Learning (ML), when it comes to automating the pneumonia diagnosis process. By analyzing the existing literature, it is observed that the majority of the efforts explored the potential of DL to detect pneumonia in chest X-ray images while little work has been conducted to explore that of ML.

Despite the high resulting accuracies of DL models and their popularity amongst researchers, several challenges have been raised to evaluate their clinical applicability. Firstly, DL models employ architectures that extract

features automatically from the data,¹⁵ and they behave like a black-box problem, which lowers the medical interpretability of the model output and undermines the clinical effectiveness of utilizing DL in healthcare. Secondly, DL models require a large volume of data to produce acceptable results.¹⁵ As a result, such models require computationally expensive systems over large training periods, which can be impractical in healthcare. Consequently, evaluating ML has the potential to overcome such challenges, and this is the main subject of interest for this research.

In contrast to DL models, ML models possess a higher potential for clinical interpretability as they permit choosing the feature extraction method, enabling the model to focus on features directly related to the symptoms of the disease in question. In addition, ML models possess the ability to provide comparable accuracies to DL models with significantly lower computational time and effort when fine-tuned. Furthermore, Bhardwaj et al.¹ suggest that ML has the potential to enhance the patient-doctor relationship while reducing the growing cost of healthcare. As such, the purpose of this paper is to propose an ML approach that can accurately and reliably detect pediatric pneumonia in chest X-rays with significantly reduced training time. It is demonstrated that using a Quadratic SVM model delivers a 97.58% accuracy surpassing the current ML accuracies reported in the literature with significantly smaller classification time than that of the used Transfer Learning (TL) benchmark. This strongly promotes ML for further development in pediatric pneumonia detection in the future.

Literature review

Automated pneumonia detection

In the field of automated pneumonia detection in chest X-ray images, the majority of the utilized efforts explored the potential of DL and its subset, TL, as opposed to ML. To illustrate, Kundu et al.¹⁷ developed an automated Computer-Aided Diagnosis (CAD) framework using TL that can classify normal and pneumonic chest X-rays with an accuracy of 98.81%. This framework utilizes a weighted ensemble that accounts for decision scores obtained from the GoogLeNet, ResNet-18, and DenseNet-21 pre-trained DL networks. Similarly, Manickam et al.²⁵ compared the performance of three pre-trained architectures, namely ResNet50, InceptionV3, and InceptionResNetV2 to distinguish between normal, bacterial, and viral pneumonia classes in approximately 2300 chest X-rays obtained from a publicly available dataset. The authors applied data augmentation techniques such as rotation, horizontal and vertical flipping and shifting, and Gaussian blurring to overcome the class imbalance present in their data, yielding an accuracy of 93.06%, 92.67%, and 92.40% for the ResNet50, InceptionV3, and InceptionResNetV2 networks, respectively.

Moreover, Vrbanić et al.²⁶ employed a deep ensemble method based on Stochastic Gradient Descent with warm

restarts to classify pneumonia in chest X-ray images using 10-fold cross-validation which produced an accuracy of 96.26%. Despite their acceptable accuracies, the DL models presented above utilize deep features that are automatically extracted from the input data which decreases the interpretability of the acquired results. In addition, such models have several limitations since they require high computational power systems and a large training dataset, both of which are often absent in existing medical systems.^{13,27}

Even though TL attempts to reduce the volume of data required to produce acceptable accuracies, it is still more time consuming than ML models. Thus, various efforts have explored the effectiveness of utilizing a hybrid AI approach in which the feature extraction is performed by DL and the pneumonia classification is performed by ML in attempt to reduce the required classification time. This is illustrated in²² where Masad et al. proposed a hybrid model in which a pre-trained Convolutional Neural Network (CNN) was used for deep feature extraction while the binary pneumonia classification was performed by various classifiers including Softmax, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Random Forest (RF). The overall reported accuracies of these classifiers are 99%, 99.3%, 99%, and 98.6%, respectively where the RF classifier consumed the largest classification time. Similarly, Zein et al.¹⁵ paired the EfficientNetB0 pre-trained network with an SVM classifier to classify normal and pneumonic chest X-rays with an accuracy of 97%. It is observed that the accuracies resulting from TL and the hybrid approach are relatively similar despite the reduced classification time in the hybrid approach. Additionally, it should be noted that the hybrid approach does not contribute to increasing the output's interpretability.

Furthermore, identifying a region of interest (ROI) for the purpose of feature extraction through lung segmentation or image cropping has the potential to increase the interpretability of AI model outputs while decreasing the required time due to reducing the number of utilized features. In addition, Chandra and Verma²⁸ suggest that utilizing a feature-extraction ROI can contribute to enhancing the performance of the model by disregarding irrelevant anatomies in chest X-rays such as the heart and diaphragm, as this may ultimately reduce the probability of yielding false positive results. In fact, the current literature contains various efforts that utilized lung segmentation to predict pneumonia and other diseases in chest X-rays.²⁹

To illustrate, pertaining to TL, Hasan et al.³⁰ applied vertical cropping and image processing techniques such as Contrast Limited Adaptive Histogram Equalization (CLAHE) to 5856 chest X-rays to encourage the pre-trained networks to extract features from the lung nodule area exclusively. Using an 80:20 training to testing ratio, the achieved accuracies are 96.2% and 95.9% for the VGG-16 and VGG-19 networks, respectively. Moreover, pertaining to ML, Chandra and Verma²⁸ utilized lung segmentation to extract first-order statistical features from a

dataset comprising of 412 chest X-rays. Subsequently, the authors evaluated the performance of several ML classifiers such as Multi-Layer Perception (MLP), RF, Sequential Minimal Optimization (SMO), Logistic Regression (LR), and classification via regression, where LR yielded the highest accuracy of 95.39%. Since this accuracy is comparable to the accuracies achieved via DL in³⁰ despite the utilization of a significantly smaller dataset, combined with the fact that smaller datasets often result in a poorer model performance, it can be inferred that ML is quite promising when it comes to the fast, accurate, and interpretable detection of pneumonia in chest X-rays.

In addition to specifying feature-extraction ROIs, using different feature extraction and selection techniques can also minimize the computational effort of a ML model and enhance its performance. For instance, Akgundogdu³¹ proposed the utilization of feature extraction based on two-dimensional Discrete Wavelet Transforms (DWT) to detect pneumonia in chest X-ray images. This method was evaluated on the Artificial Neural Network (ANN), KNN, SVM, and RF classifiers yielding an accuracy of 95.85%, 94.5%, 93.41%, and 97.11%, respectively. Moreover, Ebiele et al.³² compared the performance of several ML classifiers before and after utilizing Principal Component Analysis (PCA) for feature extraction and selection, and the results support that utilizing PCA improved the accuracy, precision, recall, and F1 scores. Nevertheless, explicitly exploring the impact of feature selection on the ML model performance is not sufficiently addressed in the literature that addresses pneumonia detection in chest X-ray images. Therefore, exploring how ML models requirements, such as feature extraction, can be altered to provide an enhanced performance is a main focus of this paper.

Methodology

The proposed ML approach for pneumonia detection in pediatric chest X-rays is summarized in Figure 1 and guided by the following steps:

Step 1: Image preprocessing and data augmentation: In this step, all the chest X-ray images are rescaled to a uniform size and converted to grayscale. Then, the data augmentation techniques shown in Table 1 are applied to balance the "Normal" and "Pneumonia" classes. This is done to prevent the ML and DL models from becoming biased towards the dominant class.

In Figure 2, the leftmost image represents a "Normal" pre-processed chest X-ray image before data augmentation, while the rightmost image represents the same image after data augmentation.

Step 2: TL benchmark selection: To assess the candidacy of the proposed ML approach, its performance will be compared to that of a pre-trained DL network. AlexNet was selected for this purpose since it demonstrated its ability to provide a high pneumonia classification

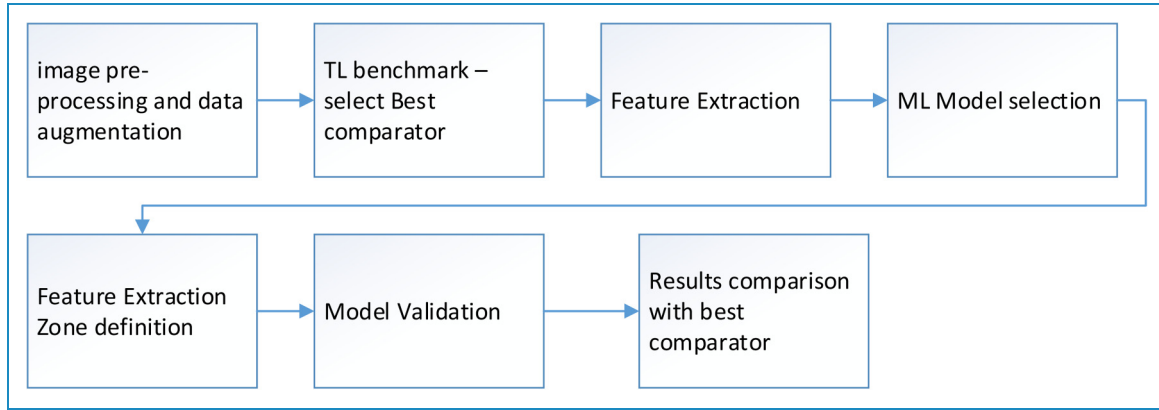


Figure 1. Methodology block diagram.

Table 1. Proposed data augmentation techniques and their corresponding parameter values.

Technique	Parameter value
Random horizontal reflection	On
Random vertical shear (degrees)	[0, 15]
Random horizontal shear (degrees)	[0, 15]
Random rotation (degrees)	[−3, 3]

accuracy that is comparable to the top performing TL models in the existing literature.¹⁷ In this step, AlexNet is modified and trained using the preprocessed dataset to serve as a benchmark for accuracy and training time comparison with the proposed ML approach illustrated in the subsequent steps. This benchmark is selected since the DL literature does not explicitly report the model training time.

Step 3: Statistical feature extraction: In this step, the following sixteen statistical features are extracted from each chest X-ray image: maximum, minimum, mean, mode, standard deviation, skewness, kurtosis, median, 2.5% quantile, 5% quantile, 10% quantile, 90% quantile, 95% quantile, 97.5% quantile, absolute energy, and entropy. These features have an interpretable dimension as the fluid that accumulates in the lungs due to pneumonia is visualized as white areas in chest X-rays. Successively, the extracted features are scaled such that their values range from 0 to 1.

Step 4: ML model selection: The data is divided into training and testing sets using a ratio of 70:30. Then, a predictive model is built using the training set and evaluated using the testing set. In this study, several ML models are utilized to predict pneumonia using the 10-fold cross-validation option in MATLAB.

Step 5: Defining the feature-extraction regions: In step 3, a set of sixteen statistical features is extracted from a single ROI, namely the whole image. Notably, this step aims to investigate the effect of increasing the number of feature-extraction ROIs per image on the resulting accuracy and training time. This is done by dividing each image into smaller feature extraction regions and extracting the same set of features from each region, which increases the length of the feature array. The investigated schemes are 4, 16, 64, and 256 equally sized regions per image as illustrated in Figure 3.

For every scheme, the fault detection explained in step 4 will be conducted to determine the scheme, which yields the highest pneumonia classification accuracy.

Step 6: Model validation: ANOVA is used to assess the reliability of the utilized ML models.

Step 7: Comparison with the TL benchmark: In this step, the ML model performance is compared to that of the TL model mentioned in step 2 to evaluate its candidacy in terms of training time and pneumonia classification accuracy.

Moreover, the proposed methodology will be evaluated using a desktop computer with an Intel® Xeon® Processor E5–1650 at 3.2 GHz and 16 GB of RAM.

Dataset

The publicly available dataset used to validate the proposed methodology is published by Kermayn et al.³³ It is comprised of 5856 pediatric chest X-ray images belonging to pediatric patients from Guangzhou Women and Children’s Medical Center, where the patients’ age ranges between one and five years old. Notably, the dataset is highly imbalanced where 4273 images are labeled as pneumonic while 1583 images are labeled as normal. Moreover, Figure 4 provides a comparison between a normal and a pneumonic chest X-ray from the utilized dataset.

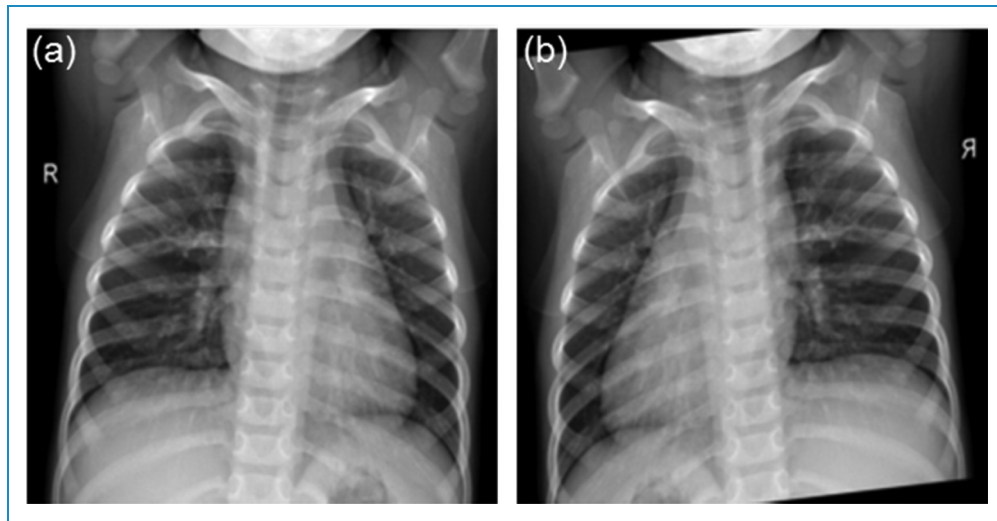


Figure 2. An example of a normal chest X-ray image before (a) and after (b) data augmentation.

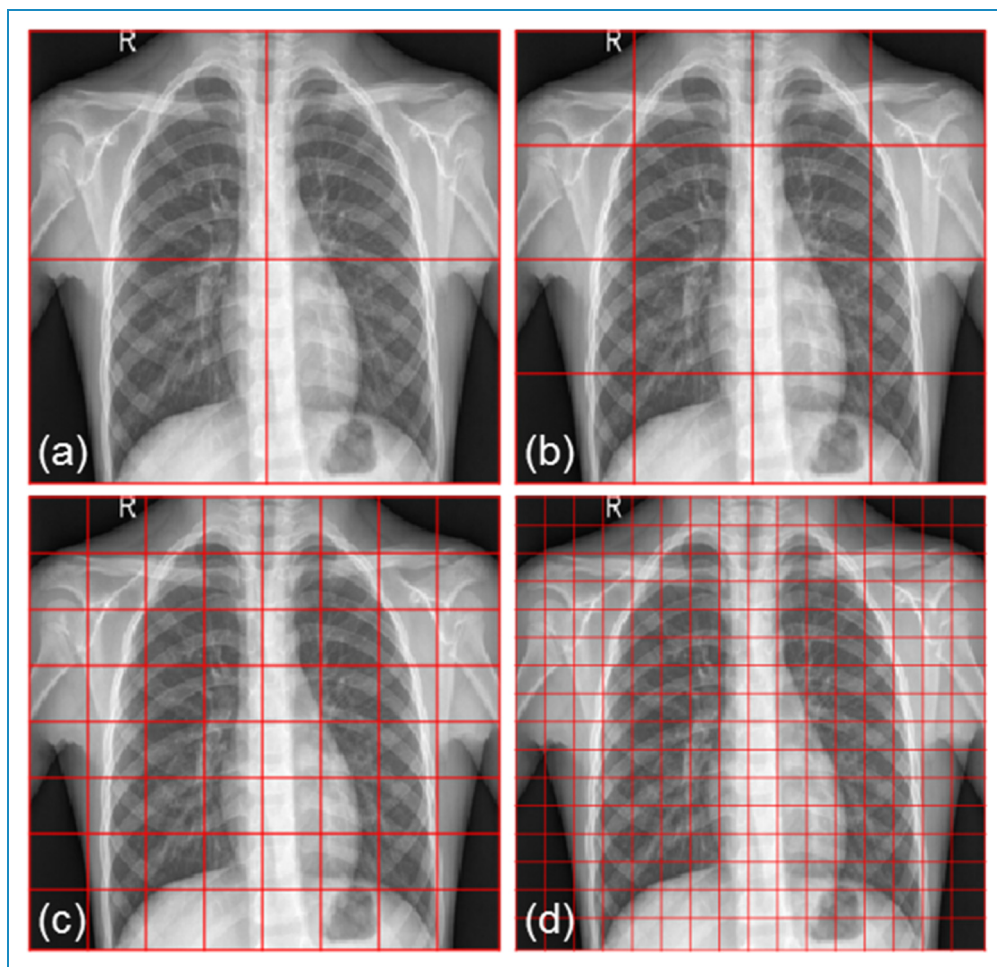


Figure 3. A schematic of the investigated ROIs; (a) 4 ROIs, (b) 16 ROIs, (c) 64 ROIs, and (d) 256 ROIs.

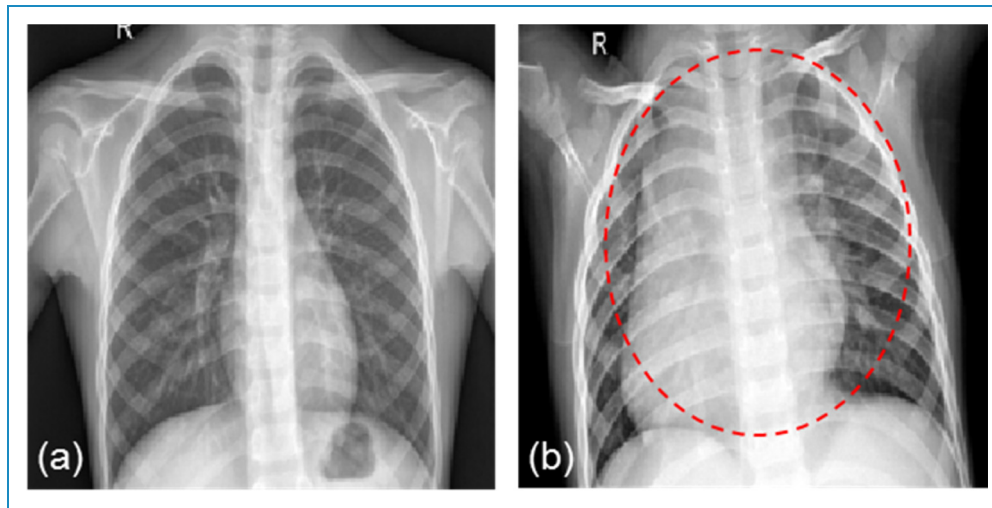


Figure 4. An example of a normal (a) and a pneumonic (b) chest X-ray from the utilized dataset.

Table 2. Description of the original and augmented dataset division utilized for Transfer Learning.

Dataset	Class	Training (70%)	Testing (15%)	Validation (15%)	Total
Original	Pneumonia	2991	641	641	4273
	Normal	1108	238	237	1583
	Total	4099	879	878	5856
Augmented	Pneumonia	2991	641	641	4273
	Normal	2991	641	641	4273
	Total	5982	1282	1282	8546

Table 3. A summary of the Transfer Learning benchmark results.

Dataset	Accuracy (%)	Precision (%)	Specificity (%)	Recall (%)	F1 (%)	Training Time
Original	97.61	98.44	98.29	98.29	98.36	107 min 14 s
Augmented	97.89	98.13	98.12	97.67	97.90	156 min 12 s

It can be observed that the lung nodule area in the pneumonic image appears brighter than that of the normal image due to the accumulation of fluid.

Results and discussion

This section presents and discusses the results of the TL benchmark and the proposed ML technique.

Transfer learning benchmark

The deep pre-trained AlexNet network is chosen to detect pediatric pneumonia in chest X-rays before and after image augmentation to serve as a time and accuracy comparison benchmark for the results obtained by utilizing the proposed ML methodology. The original and augmented datasets are divided into 70% training, 15% testing, and 15% validation sets as shown in Table 2.

Table 4. Comparing the performance of the Transfer Learning benchmark to the current Deep Learning literature.

Publication	Dataset size	Model	Accuracy (%)	Precision (%)	Specificity (%)	Recall (%)	F1 (%)	AUC (%)
Kundu et al. ¹⁷	5856 images ³³	GoogLeNet, ResNet-18, and DenseNet-121 Ensemble	98.81	98.82	–	98.80	98.79	98.35
	26,601 images		86.89	86.89	–	87.02	86.95	86.85
Manickam et al. ²⁵	5229 images	InceptionV3	92.67	88.70	–	92.70	90.65	–
		InceptionResNetV2	92.40	88.88	–	93.20	90.98	–
		ResNet50	93.06	88.97	–	96.78	92.71	–
Hasan et al. ³⁰	5856 images ³³	VGG-16	96.20	97.70	93.90	97.00	97.30	95.70
		VGG-19	95.90	97.10	91.90	97.40	97.20	95.50
Zein et al. ¹⁵	5856 images ³³	EfficientNetB0	96.70	99.90	99.60	95.60	97.70	97.60
		EfficientNetB0 and SVM	97.00	100.00	100.00	95.80	97.70	98.00
Masad et al. ²²	5856 images ³³	CNN and Softmax	98.97	97.90	99.22	98.31	–	–
		CNN and SVM	98.97	97.90	99.22	97.90	–	–
		CNN and KNN	98.51	97.89	99.22	96.67	–	–
		CNN and RF	97.15	94.96	98.12	94.56	–	–
Vrbančič et al. ²⁶	5858 images ³⁴	SGDRE	96.26	97.32	92.74	97.57	97.44	95.15
TL Benchmark	8546 images	AlexNet	97.89	98.13	98.12	97.67	97.90	–

Subsequently, AlexNet is trained then utilized to predict the classes of the testing set where Table 3 provides a summary of the obtained results.

By analyzing the results presented above, it can be inferred that utilizing data augmentation contributed to a slight improvement in the testing accuracy at the expense of an added 50 min of training time. In addition, the performance of AlexNet is comparable to that of the models presented in the DL literature (see Table 4), which justifies utilizing it as an accuracy benchmark.

Furthermore, since AlexNet is generally less computationally expensive than the other pre-trained models,³⁵ utilizing it as a benchmark for time comparison is justifiable.

Machine learning results

Pertaining to ML, the augmented dataset is divided using the scheme described in Table 5.

Table 5. Description of the augmented dataset division utilized for Machine Learning.

	Training (70%)	Testing (30%)	Total	Validation
Pneumonia	2991	1282	4273	10 k-fold
Normal	2991	1282	4273	
Total	5982	2564	8546	

After that, the sixteen proposed statistical features are extracted from five feature-extraction schemes, namely the whole image and 4, 16, 64, and 256 equally sized regions per image. This is conducted to obtain insights regarding the impact of the feature-extraction ROI size on the model output. Consequently, the performance of twenty-six ML classifiers is evaluated using the MATLAB Classification Learner

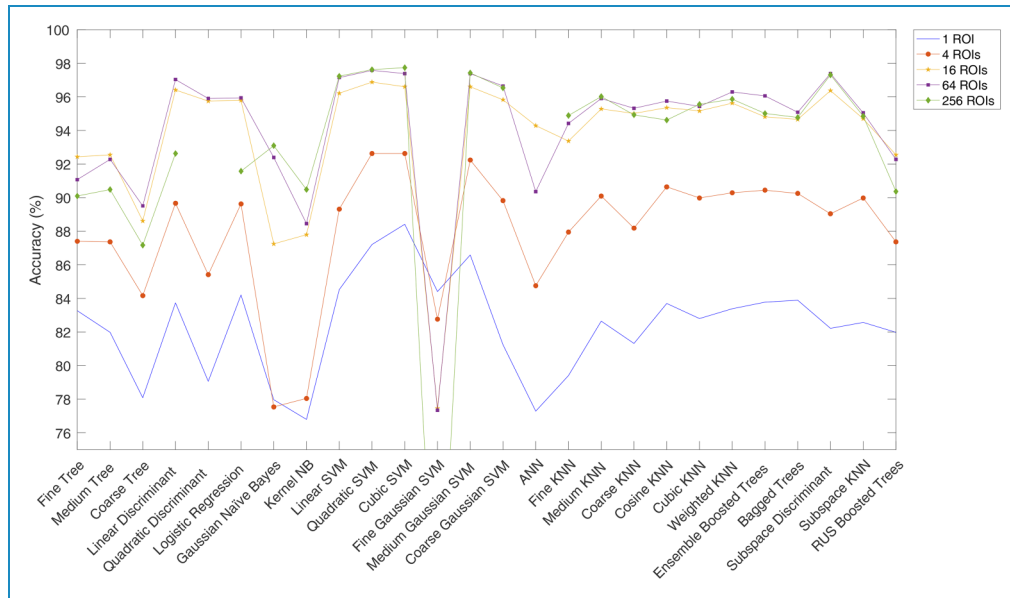


Figure 5. Plot of accuracy per ML classifier for each feature-extraction ROI scheme.

Table 6. Comparing the performance of the top Machine Learning models to the proposed Machine Learning model.

Publication	Dataset size	Model	Accuracy (%)	Precision (%)	Specificity (%)	Recall (%)	F1 (%)	AUC (%)
Chandra and Verma ²⁸	412 images ³⁷	MLP	95.39	97.46	97.57	93.20	95.29	95.40
		RF	94.42	95.07	95.15	93.69	94.38	94.40
		SMD	93.69	97.87	98.06	89.32	93.40	93.70
		Regression	94.66	97.92	98.06	91.26	94.47	94.70
		LR	95.63	97.48	97.57	93.69	95.55	95.60
Akgundogdu ³¹	5858 images ³⁴	RF	97.11	–	99.09	91.79	98.04	99.00
		ANN	95.85	–	98.01	90.02	92.14	98.20
		KNN	94.50	–	96.44	98.13	89.76	96.50
		SVM	93.41	–	94.62	95.22	87.48	90.80
Ebiele et al. ³²	NIH dataset ³⁸	SVM	90.00	92.00	–	88.0	90.00	96.00
		MLP	89.00	87.00	–	91.0	89.00	91.60
		KNN	86.00	87.00	–	86.0	86.00	92.30
		GBC	86.00	87.00	–	84.0	86.00	93.30
Proposed Approach	8546 images	Quadratic SVM (64)	97.58	97.97	97.96	97.21	97.59	97.00
		Cubic SVM (256)	97.74	98.36	98.34	97.15	97.75	98.00

and Neural Net Fitting apps on each feature-extraction scheme, and the resulting accuracy is displayed in Figure 5.

Thus, it can be inferred that a higher number of feature-extraction ROIs results in a higher testing accuracy where the top performing ML classifiers include several classifiers belonging to the SVM family as well as the subspace discriminant classifier. This is in line with³⁶ that demonstrates SVM's candidacy in medical applications as it tends to yield the highest classification accuracy. It should be noted that the Quadratic Discriminant and Artificial Neural Network (ANN) models did not converge for the 256 ROI scheme as they were computationally expensive. In addition, the accuracies achieved by the top performing classifiers for the 64 and 256 ROI schemes surpass those provided in the ML literature as summarized in Table 6.

When comparing these two schemes, it is observed they are quite similar in term of the testing accuracy, especially when it comes to the top performing classifiers. However, when factoring in the training time, the performance of the 64 ROI scheme is superior as illustrated in Figure 6. Hence, the 64 ROI scheme is optimal.

Statistical analysis

The SVM ML family provided the best results in terms of accuracy and time. However, the differences among the SVM methods were not that obvious. The analysis was repeated 10 times for all SVM tools results and the accuracy (training and testing), training time, TN, TP, FN, and FP

were obtained. Each time, the data was split to 30% test and 70% train randomly, then the normal class was augmented to match the pneumonia class (augmentation is also random). Next, ANOVA analysis was conducted for each outcome among the SVM family ML methods. Based on the p-value of 0.625, there is no statistically significant difference among the methods in terms of time, training accuracy, and true negatives (TN). However, there is strong evidence of a difference in terms of training and testing accuracy, false positives (FP), and true positives (TP) with p-values of 0.00. Figures 7 and 8 depict box plots for the results.

Notably, the time required for this scheme is significantly lower than the time needed by the TL benchmark as shown in Table 7. In addition, the quadratic SVM yielded an accuracy of 97.58% for the 64 ROI scheme, which is comparable to the 97.89% obtained via TL.

Hence, the results highlight the candidacy of ML as an accurate and computationally inexpensive tool for pediatric pneumonia detection in chest X-ray images. As mentioned earlier, the motivation behind utilizing ML in the proposed technique lies in its low computational expense and high potential for interpretability. While the former has been successfully confirmed by the obtained results, the latter possesses potential for further improvement. Thus, after realizing the optimal feature-extraction scheme, namely the 64 ROI scheme, the most important features are identified using the Classification and Regressions Tree (CART) method using Minitab and traced back to their

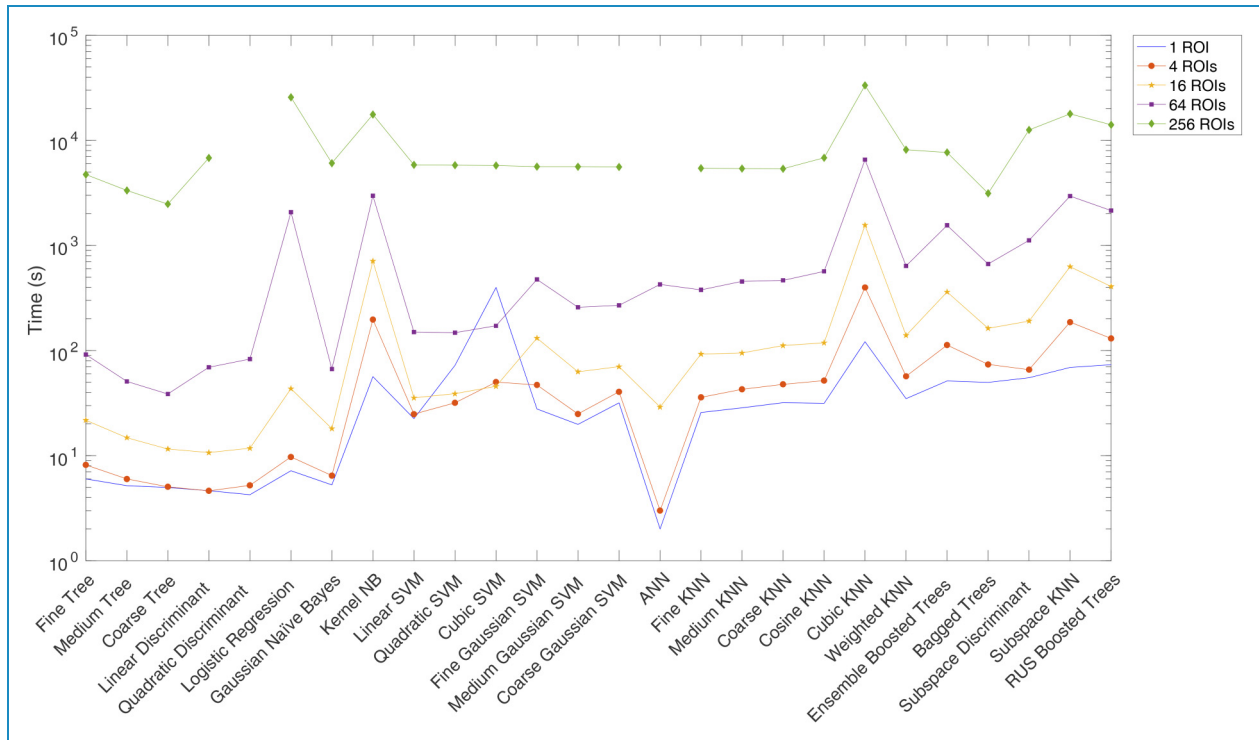


Figure 6. Plot of training time per ML classifier for each feature-extraction ROI scheme.

corresponding ROI locations in the chest X-ray images as shown in Figure 9.

This adds an interpretable dimension to the proposed approach as it brings critical ROIs from the X-rays to the radiologists’ attention while making the diagnosis decision

using our model. Moreover, to further enhance the interpretability, the future work will tackle reducing the X-ray images to a centered square enclosing the locations that contain the identified important features as illustrated in Figure 10.

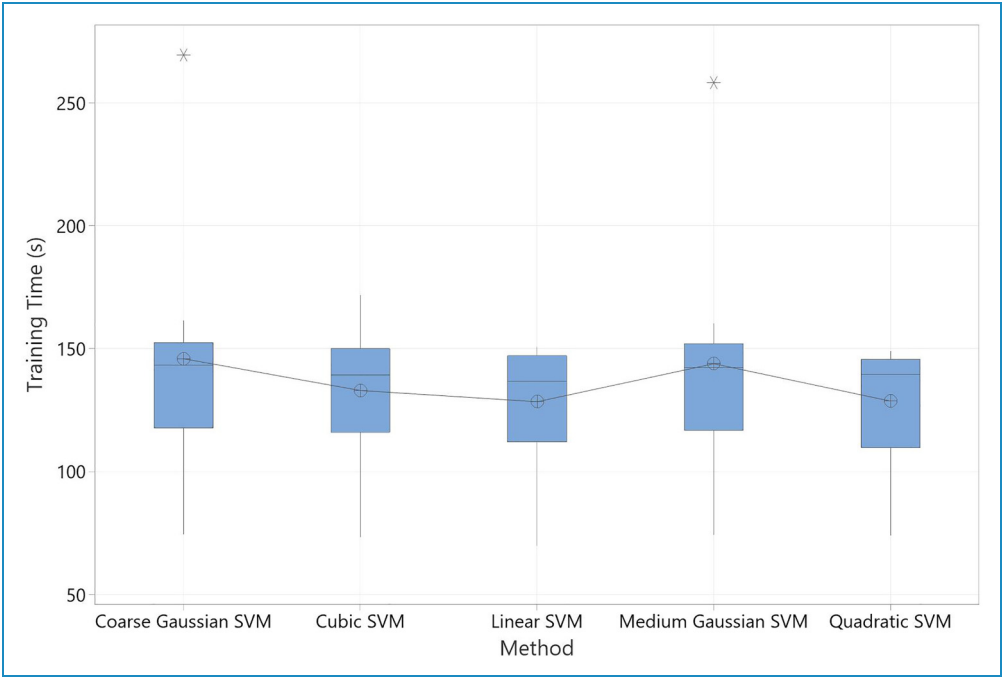


Figure 7. Box plot of training time.

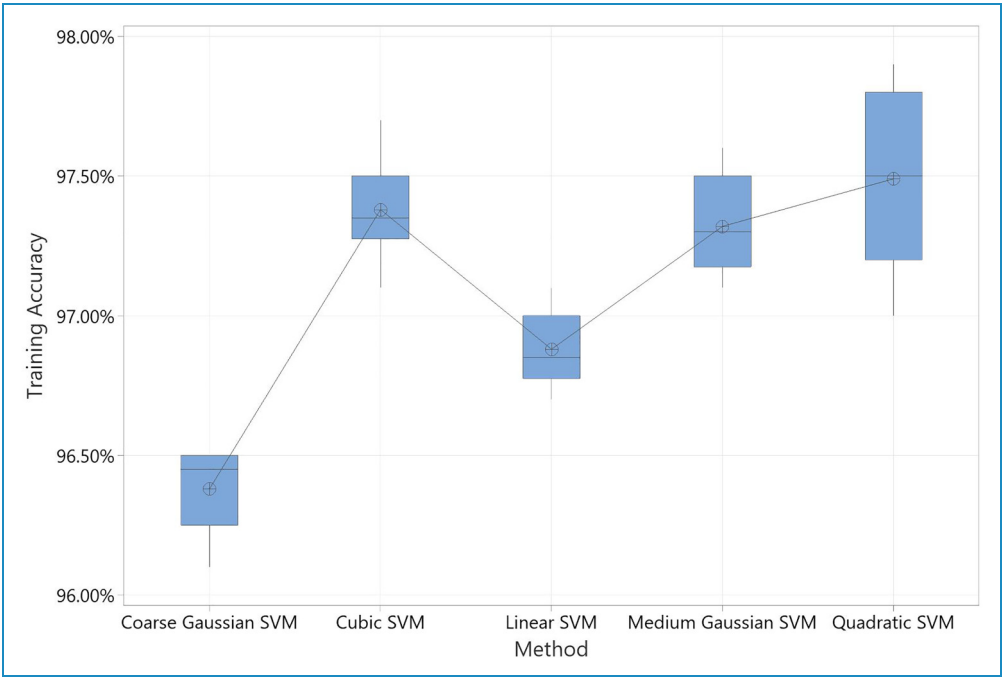
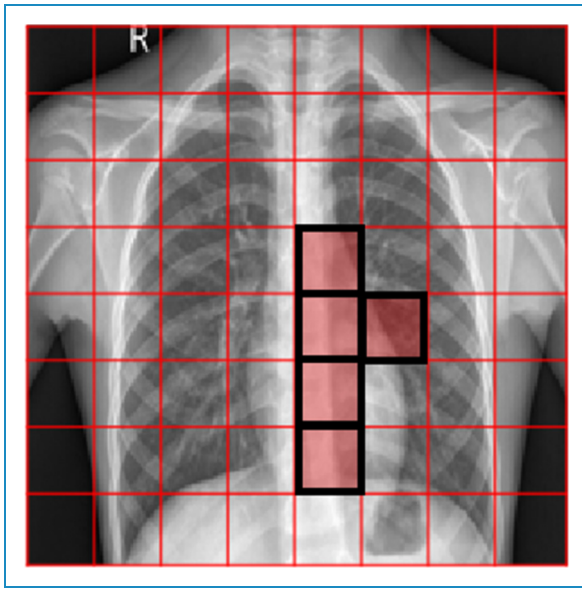
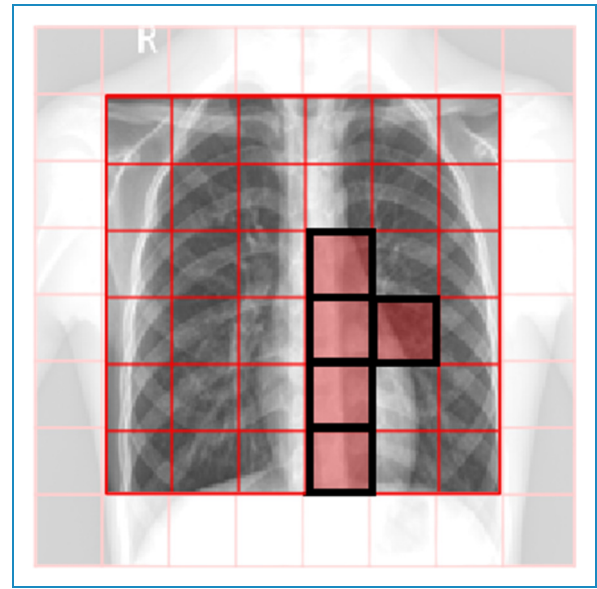


Figure 8. Box plot of training accuracy.

Table 7. Comparing the performance of the proposed Machine Learning approach with the Transfer Learning Benchmark.

Model	Dataset size	Accuracy (%)	Precision (%)	Specificity (%)	Recall (%)	F1 (%)	AUC (%)	Training time
TL Benchmark	8546 images	97.89	98.13	98.12	97.67	97.90	–	156 min 12 s
Quadratic SVM (64 ROI)	8546 images	97.58	97.97	97.96	97.21	97.59	97.00	2 min 28 s
Cubic SVM (256 ROI)	8546 images	97.74	98.36	98.34	97.15	97.75	98.00	95 min 59 s

**Figure 9.** A schematic illustrating the ROIs containing the most important features obtained from Minitab.**Figure 10.** An example of a reduced X-ray image.

Then, step 5 of the methodology will be repeated for the reduced X-ray images considering a higher number of feature-extraction regions than those in the optimal ROI scheme. This step will be reiterated until a satisfactory model performance is achieved, where this will answer whether focusing on a specific region in the X-rays is more efficient than considering the image as a whole.

Conclusion and future work

In this paper, a novel ML approach is proposed to reliably detect pediatric pneumonia in chest X-rays. This approach entails performing data augmentation to balance the classes of the utilized dataset, optimizing the feature extraction scheme, and evaluating the performance of several ML models. Moreover, the performance of this approach is compared to a TL benchmark to evaluate its candidacy.

After investigating the effect of varying the number of the feature extraction ROIs on the classification time and accuracy, it is inferred that the 64 ROI feature extraction scheme is optimal. Using this scheme, the Quadratic SVM model yielded an accuracy of 97.58%, which surpasses the accuracies reported in the current ML literature. In addition, this scheme's classification time is significantly smaller than that of the TL benchmark, making it practical and less computationally expensive. Hence, the results highlight the candidacy and potential of the proposed approach in detecting pediatric pneumonia in chest X-rays.

Moreover, the future work will investigate refining the optimal feature-extraction scheme to enhance the proposed approach's performance and interpretability. This will be conducted by reducing the X-ray images into a centered square enclosing the locations that contain the lungs. Then, the feature extraction scheme will be optimized for the reduced images by varying the number of ROIs. This

step will be reiterated until a satisfactory model performance is achieved, as this will provide insights on whether focusing on a specific region in the X-rays is more accurate and efficient than considering the image as a whole. Then, the most important features will be determined using several methods such as CART. Subsequently, these features will be traced back to their corresponding locations in the chest X-rays in order to aid medical practitioners in diagnosing pediatric pneumonia in an interpretable manner.

Lastly, it is important to note that this research is limited to the binary classification of pneumonia using pediatric X-rays with a certain resolution and a sufficient sample size. It is worth mentioning that the proposed method has a potential of being used for other applications such as detection of COVID-19 or shaft defects in a motor or other binary classification applications using X-ray images or any other different imaging modality. Similarly, the proposed method may be used on adult X-rays that vary anatomically from pediatric X-rays. Nevertheless, each one of these applications might have its own specific challenges and opportunities. As a result, generalizing this method to other applications needs to be evaluated.

Acknowledgements: The authors would like to thank Dr Hussam Alshraideh from the industrial engineering department at the American University of Sharjah for his support and technical advice. The work in this paper was supported, in part, by the Open Access Program from the American University of Sharjah. This paper represents the opinions of the author(s) and does not mean to represent the position or opinions of the American University of Sharjah.

Author contribution: All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Natalie Barakat, Mahmoud Awad and Bassam Abu-Nabah. The first draft of the manuscript was written by Natalie Barakat and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Consent to participate: Not applicable- A publicly available dataset is used in this study is. Data set is published by Kermany et al.³³

Consent to publish: Not-applicable- A publicly available dataset is used in this study is. Data set is published by Kermany et al.³³

Declaration of conflicting interests: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

Funding: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication

of this article: This work was supported by a scholarship fund provided by the Engineering Systems Management department at the American University of Sharjah.

Guarantor: Mahmoud Awad.

ORCID iD: Mahmoud Awad  <https://orcid.org/0000-0003-2999-9084>

References

1. Bhardwaj R, Nambiar AR and Dutta D. A study of machine learning in healthcare. *Proceedings – International Computer Software and Applications Conference* 2017; 2: 236–241.
2. Au-Yong-Oliveira M, Pesqueira A, Sousa MJ, et al. The potential of big data research in HealthCare for medical doctors' learning. *J Med Syst* 2021; 45: 1–14.
3. Mendo IR, Marques G, de la Torre Díez I, et al. Machine learning in medical emergencies: a systematic review and analysis. *J Med Syst* 2021; 45. DOI: 10.1007/s10916-021-01762-3.
4. Zhang Z, Genc Y, Wang D, et al. Effect of AI explanations on human perceptions of patient-facing AI-powered healthcare systems. *J Med Syst* 2021; 45. DOI: 10.1007/s10916-021-01743-6.
5. Krittanawong C. The rise of artificial intelligence and the uncertain future for physicians. *Eur J Intern Med* 2018; 48: e13–e14.
6. Soellner M and Koenigstorfer J. Compliance with medical recommendations depending on the use of artificial intelligence as a diagnostic method. *BMC Med Inform Decis Mak* 2021; 21: 1–11.
7. Murugan A, Nair SAH and Kumar KPS. Detection of skin cancer using SVM, random forest and kNN classifiers. *J Med Syst* 2019; 43. DOI: 10.1007/s10916-019-1400-8.
8. Palani D and Venkatalakshmi K. An IoT based predictive modelling for predicting lung cancer using fuzzy cluster based segmentation and classification. *J Med Syst* 2019; 43. DOI: 10.1007/s10916-018-1139-7.
9. Zerouaoui H and Idri A. Reviewing machine learning and image processing based decision-making systems for breast cancer imaging. *J Med Syst* 2021; 45. DOI: 10.1007/s10916-020-01689-1.
10. Jafarian K, Vahdat V, Salehi S, et al. Automating detection and localization of myocardial infarction using shallow and end-to-end deep neural networks. *Applied Soft Computing Journal* 2020; 93: 106383.
11. Yıldırım Ö, Pławiak P, Tan RS, et al. Arrhythmia detection using deep convolutional neural network with long duration ECG signals. *Comput Biol Med* 2018; 102: 411–420.
12. Shu D, et al. Artificial intelligence and deep learning in ophthalmology. *The British Journal of Ophthalmology* 2019; 167–175. DOI: 10.1136/bjophthalmol-2018-313173.
13. Govindarajan S and Swaminathan R. Analysis of Tuberculosis in chest radiographs for computerized diagnosis using bag of keypoint features. *J Med Syst* 2019; 43: 11–13.
14. Khan M, et al. Applications of artificial intelligence in COVID-19 pandemic: a comprehensive review. *Expert Syst Appl* 2021; 185: 115695.
15. Zein O, Soliman M, Elkholy A, et al. Transfer learning based model for pneumonia detection in chest X-ray images. *International Journal of Intelligent Engineering and Systems* 2021; 14: 56–66.

16. Liz H, Sánchez-Montañés M, Tagarro A, et al. Ensembles of convolutional neural network models for pediatric pneumonia diagnosis. *Future Gener Comput Syst* 2021; 122: 220–233.
17. Kundu R, Das R, Geem ZW, et al. Pneumonia detection in chest X-ray images using an ensemble of deep learning models. *PLoS One* 2021; 16: e0256630.
18. El Asnaoui K. Design ensemble deep learning model for pneumonia disease classification. *Int J Multimed Inf Retr* 2021; 10: 55–68.
19. Stokes K, et al. Biomedical signal processing and control the use of artificial intelligence systems in diagnosis of pneumonia via signs and symptoms: a systematic review. *Biomed Signal Process Control* 2022; 72: 2–3.
20. Pneumonia. <https://www.who.int/news-room/fact-sheets/detail/pneumonia> (2019, accessed 10 October 2021).
21. Neuman MI, et al. Variability in the interpretation of chest radiographs for the diagnosis of pneumonia in children. *J Hosp Med* 2012; 7: 294–298.
22. Masad IS, Alqudah A, Alqudah AM, et al. A hybrid deep learning approach towards building an intelligent system for pneumonia detection in chest x-ray images. *International Journal of Electrical and Computer Engineering* 2021; 11: 5530–5540.
23. Voigt GM, et al. Interobserver agreement in interpretation of chest radiographs for pediatric community acquired pneumonia: findings of the pedCAPNETZ-cohort. *Pediatr Pulmonol* 2021; 56: 2676–2685.
24. Fawole OA, et al. Interpretation of pediatric chest radiographs by non-radiologist clinicians in Botswana using world health organization criteria for endpoint pneumonia. *Pediatr Radiol* 2020; 50: 913–922.
25. Manickam A, Jiang J, Zhou Y, et al. Automated pneumonia detection on chest X-ray images: a deep learning approach with different optimizers and transfer learning architectures. *Measurement (Lond)* 2021; 184: 109953.
26. Vrbanič G and Podgorelec V. Efficient ensemble for image-based identification of pneumonia utilizing deep CNN and SGD with warm restarts. *Expert Syst Appl* 2021; 187: 115834.
27. Hooda R, Mittal A and Sofat S. Segmentation of lung fields from chest radiographs-a radiomic feature-based approach. *Biomed Eng Lett* 2019; 9: 109–117.
28. Chandra TB and Verma K. Pneumonia detection on chest X-ray using machine learning paradigm. *Advances in Intelligent Systems and Computing* 2020; 1022.AISC: 21–33.
29. Tilve A, Nayak S, Vernekar S, et al. Pneumonia detection using deep learning approaches. In: [international conference on emerging trends in information technology and engineering, ic-ETITE 2020](#), 2020. DOI: 10.1109/IC-ETITE47903.2020.152.
30. Hasan MM, Md. Jahangir Kabir M, Haque MR, et al. A combined approach using image processing and deep learning to detect pneumonia from chest X-ray image. In: [3rd International conference on electrical, computer and telecommunication engineering, ICECTE 2019](#), 2019, pp.89–92. DOI: 10.1109/ICECTE48615.2019.9303543.
31. Akgundogdu A. Detection of pneumonia in chest X-ray images by using 2D discrete wavelet feature extraction with random forest. *Int J Imaging Syst Technol* 2021; 31: 82–93.
32. Ebiele J, Ansah-Narh T, Djioke S, et al. Conventional machine learning based on feature engineering for detecting pneumonia from chest X-rays. In: [Pervasivehealth: pervasive computing technologies for healthcare](#), 2020, pp.149–155. DOI: 10.1145/3410886.3410898.
33. Kermany DS, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* 2018; 172: 1122–1131. e9.
34. Kermany D, Zhang K and Goldbaum M. Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification, 2018. DOI: 10.17632/RSCBJBR9SJ.2.
35. Pretrained Deep Neural Networks – MATLAB & Simulink. <https://www.mathworks.com/help/deeplearning/ug/pretrained-convolutional-neural-networks.html> (accessed 4th December 2021).
36. Kalantari A, Kamsin A, Shamshirband S, et al. Computational intelligence approaches for classification of medical data: state-of-the-art, future challenges and research directions. *Neurocomputing* 2018; 276: 2–22.
37. Wang X, Peng Y, Lu L, et al. ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: [Proceedings – 30th IEEE conference on computer vision and pattern recognition, CVPR 2017](#), vol. 2017-Janua, 2017, pp.3462–3471. DOI: 10.1109/CVPR.2017.369.
38. RSNA Pneumonia Detection Challenge. <https://www.rsna.org/education/ai-resources-and-training/ai-image-challenge/RSNA-Pneumonia-Detection-Challenge-2018> (2018, accessed 5 December 2021).