Ann Ig 2021 Nov-Dic; 33(6): 644-655. doi: 10.7416/ai.2021.2467.

Hierarchical convolutional models for automatic pneumonia diagnosis based on X-ray images: new strategies in public health

G. Maselli[1], E. Bertamino[2], C. Capalbo[2,3], R. Mancini[2,4], G.B. Orsi[2,5], C. Napoli[2,6], C. Napoli[1]

Key words: Accuracy, convolutional neural network, diagnostic process
Parole chiave: Accuratezza, modelli convoluzionali, processi diagnostici

Abstract

Background. *In order to help physicians and radiologists in diagnosing pneumonia, deep learning and other artificial intelligence methods have been described in several researches to solve this task. The main objective of the present study is to build a stacked hierarchical model by combining several models in order to increase the procedure accuracy.*

Methods. *Firstly, the best* convolutional network *in terms of accuracy were evaluated and described. Later, a stacked hierarchical model was built by using the most relevant features extracted by the selected two models. Finally, over the stacked model with the best accuracy, a hierarchically dependent second stage model for inner-classification was built in order to detect both inflammation of the pulmonary alveolar space (lobar pneumonia) and interstitial tissue involvement (interstitial pneumonia).*

Results. *The study shows how the adopted staked model lead to a higher accuracy. Having a high accuracy on pneumonia detection and classification can be a paramount asset to treat patients in real health-care environments.*

Conclusions. *Despite some limits, our findings support the notion that deep learning methods can be used to simplify the diagnostic process and improve disease management.*

[1] *Department of Computer, Control, and Management Engineering, Sapienza University of Rome, Rome, Italy*
[2] *Sant'Andrea Hospital, Rome, Italy*
[3] *Department of Molecular Medicine, Sapienza University of Rome, Rome, Italy*
[4] *Department of Clinical and Molecular Medicine, Sapienza University of Rome, Rome, Italy*
[5] *Department of Public Health and Infectious Diseases, Sapienza University of Rome, Rome, Italy*
[6] *Department of Medical Surgical Sciences and Translational Medicine, Sapienza University of Rome, Rome, Italy*

Annali di Igiene : Medicina Preventiva e di Comunità (Ann Ig)

ISSN 1120-9135 https://www.annali-igiene.it

Introduction

Transmission of infectious diseases is, worldwide, one of the most important problems of public health; therefore, screening, assessment, and early diagnosis are considered of primary importance as control measures (1-4). After the spread of the novel coronavirus, also known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the problem became a global emergency. It is demonstrated that the spread of the disease may be efficiently controlled thanks to some health management choices, in addition to developing structured care (5, 6). Besides that, due to the magnitude and the complexity of the problem, it is of paramount importance to develop early diagnosis systems not only based on laboratory tests, but also on alternative testing methods, such as diagnostic imaging (7, 8).

Pneumonia is a lung inflammation caused by different pathogens. Viral pneumonia is generally of milder severity, and symptoms occur gradually; it can become complicated to diagnose if a bacterial infection develops at same time with viral pneumonia. On the other side, bacterial pneumonia can be more severe, and can eventually affect many lobes of the lung. Fungal pneumonia generally occurs in patients with weak immune systems. Such a pneumonia can be dangerous and requires time for regress (9). Chest X-ray (CXR) and computed tomography (CT) images are central in pneumonia diagnosis; however, it is fundamental to promptly analyse such images in order to obtain an early diagnosis. Motivated by this, researchers are globally taking initiatives to assist health practitioners with cutting-edge technology that also aims to detect and possibly prevent the further spread of the etiological agent (8).

Recently, a number of researchers have proposed different artificial intelligence (AI) based solutions for different medical problems. Convolutional neural networks (CNNs), as an example, have allowed researchers to obtain successful results in medical issues, such as breast or brain cancer detection, staging and classification based on X-ray images (9). In order to help field experts, such as physicians and radiologists, in diagnosing pneumonia, deep learning and other AI methods have been adopted to solve this task in several researches (9-13).

Toğaçar et al. adopted a pre-trained CNN model with a similar layer structure for the pneumonia detection. Each model is separately applied to the dataset to extract the local discriminative features. By using the minimum redundancy maximum relevance (mRMR) algorithm, the dimension of the obtained deep features was reduced. These features were then combined to create a feature set given as input to several classifiers for the final classification (14).

Rahman et al. utilized four pre-trained models for transfer learning management and for analysing their performances. The authors show that the CNN Densenet201 model reaches the higher accuracy for the pneumonia detection (15).

Hammam et al compared six pre-trained models and then selected the best three of them in terms of accuracy to build a stacking ensemble deep learning model for an early prediction of COVID-19 diagnosis (16).

Based on this background, aims of this study are: i) to evaluate, select and describe the best convolutional network in terms of accuracy; ii) to build a stacked model by using the most relevant features extracted by the selected models; iii) to build, over the stacked model with the best accuracy, a hierarchically dependent second stage model for inner-classification in order to distinguish inflammation of the pulmonary alveolar space (lobar pneumonia) and interstitial tissue involvement (interstitial pneumonia).

Methods

The study was performed using a publicly available dataset of validated CXR; the images (anterior-posterior projections) were selected from a retrospective cohort of patients from Guangzhou Women and Children's Medical Center, Guangzhou (17). All CXR images were part of patients' routine clinical care. All chest radiographs were initially screened for quality control by removing the low quality or unreadable scans. The diagnoses for the images were performed by two expert physicians before being used for training the artificial intelligence system. In order to account for any grading errors, the evaluation set was also checked by a third expert.

In this database, over a total of 5216 images, 16 images for the validation set and 624 for the test were contained. A validation set of just 16 images was not enough to perform a proper estimation of the model properties (16). To solve this problem, a further 80:20 split has been performed from the union of both the training and validation dataset. In this way, we have decreased the number of training sample but at the same time the amount of validation images has been increased.

In Figure 1, it is possible to see that the sample is highly imbalanced. Pneumonia samples (positive) have a much higher number than normal images. This means that much more samples of a class are present compared to the other.

Therefore, a correction on the weights of the two classes was applied, reducing the imbalance between class 0 (normal) and class 1 (pneumonia) weight. This correction is of a critical importance since in general, CNNs model works better when the training data are balanced.

CXR images were resized to a dimension of 224x224x3, also to avoid overfitting. Transformations as shear, zoom, rotation, width shift, height shift, brightness and horizontal flip were applied as standard data augmentation technique. Furthermore, since we deal with a binary classification task, images mode was set to binary.

1. Theoretical outlines of the pre-trained model chosen

First of all, the pre-trained models were selected on a previous experience by Hammam et al. (16). The best two models were chosen after an accuracy analysis with respect to other models, by using the above mentioned free available database

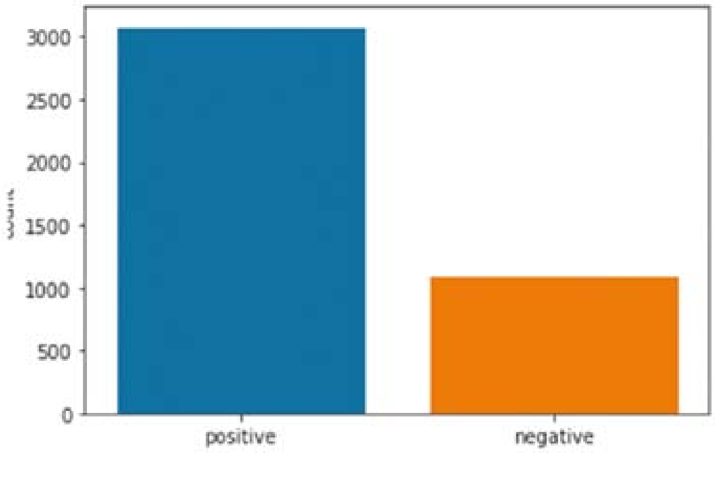


Figure 1 - Distribution of positive and negative samples.

of radiological images. *MobileNet* and *DenseNet121* models were the best ones in terms of testing accuracy, as shown in Table 1. Their features are summarized below, showing an overview of their architectures as well.

Table 1 - Testing accuracy of selected pre-trained models.

Model	Testing accuracy
MobileNet	0.916
DenseNet121	0.897
DenseNet169	0.893
InceptionResnetV2	0.879
Xception	0.825
MobileNetV2	0.806
InceptionV3	0.799
Resnet50	0.744

1a. MobileNet. Howard et al. proposed the *MobileNet* neural network in 2017 (18). The architecture of this network is based on separable convolutions. The latter are a form of factorized convolution which factorize a standard convolution in:

• depthwise convolution: applies a single filter to each input channel;

• pointwise convolution: applies a 1x1 convolution to combine the outputs of the depthwise convolution.

Basically, in a standard convolution we both filter and combine inputs into a new set of outputs in one step, while the depthwise separable convolution splits this into two layers:

• a separate layer for filtering;

• a separate layer for combining.

This is done to drastically reduce computation and model size. Finally, *MobileNet* uses both batch normalization and ReLU nonlinearities after all layers; moreover, a final average pooling reduces the spatial resolution to 1 before the fully connected layer. The latter does not present nonlinearities and feeds into a softmax.

1 b. DenseNet. Huang, et al. proposed the Densely CCN (*DenseNet*) as the next step to keep increasing the depth of deep convolutional networks (19). This solution was adopted firstly to solve the problems arising when the CNN go deeper, due to the fact that the path of information from the input layer until the output layer becomes so big that it vanishes before reaching the other side.

By connecting every layer directly with each other, the authors managed to solve the problem ensuring maximum information and gradient flow. One of the major advantages of using such a network is the fact that the *DenseNet* network, through the feature reuse, exploits its potential by avoiding relying on an extremely deep or wide architecture.

Despite of the classic CNNs, *DenseNet* does not need to learn redundant features and by adopting the type of connection aforementioned it require fewer parameters. Furthermore, by having very narrow layers, the network adds just a small set of new feature-maps.

In training phase, the *DenseNet* network can solve the aforementioned problem regarding the flow of information and gradients by relying on the fact that each layer has direct access to the gradients from the loss function and the original input image.

Another important aspect that should be mentioned is that the *DenseNet* concatenate the output feature maps of the layer with the incoming feature maps; therefore, there is no sum between them. In any case, to perform this concatenation the feature maps must have the same size. To address this problem, the *DenseNet* introduce the concept of *DenseBlocks*. Basically, *DenseBlocks* are utilized to guarantee that the dimension of the feature maps remains constant within a block, but the number of filter changes between them. Between the *DenseBlocks,* particular type of layers (called transition layers) are inserted. These

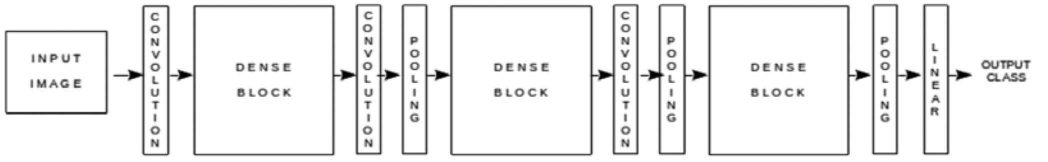


Figure 2 - The *DenseNet* architecture built by using *DenseBlocks* and transition Layers between them. The latter change feature-map sizes via convolution and pooling.

layers perform the down sampling applying batch normalization, a 1x1 convolution and a 2x2 layers. The basic architecture is shown in Figure 2.

After giving the basic theoretical outlines behind *DenseNet* networks, it is important to underline that the architecture implemented for our research is the *DenseNet121* one.

2. Building the Stacked Model

Mobilenet and *DenseNet* theoretical models were combined to create the stacked model in order to further improve the overall accuracy on predictions of the models as previously performed (16). The procedure foreseen to remove from both the pre-trained networks (*MobileNet* and *DenseNet121*) the last fully connected layer was implemented relying on the assumption that in the last convolutional layer the best weights for the feature detection are present. As a matter of fact, the last feature map should contain the most relevant extracted features for the final classification of the image. Based on this reasoning, we have improved the overall accuracy of both the best two models by merging the two last feature maps (one from each model) in a single flatten layer followed by two blocks composed by batch normalization, dense and dropout layers. As last procedure, the final sigmoid activation function was added to perform a binary classification.

The above assumptions ended up with a unique staked model that adopts the two pre-trained models as features extractors. This last third model receives the most relevant extracted feature to perform the final classification task. The overall architecture of the stacked model is shown in Figure 3.

3. Build, over the stacked model, a hierarchically dependent second stage for inner-classification

While the stacked model is able to classify pneumonia from the control group of healthy patients, this is not able to distinguish between lobar pneumonia and interstitial pneumonia. For this reason, a hierarchically dependent second stage model was added for inner-classification in order to distinguish the two kinds of pneumonia. This second stage started by the classification of pneumonia resulting by the use of the stacked model in order to obtain a more accurate classification (excluding, therefore, any classification interferences from the healthy control group). For this latter stage, a pre-trained *DenseNet201* model was used.

4. Training Phase

The training phase is modelled by inserting *EarlyStopping*, R*educeLROnPlateau* and *ModelCheckpoint* functions.

With *EarlyStopping* it is possible to monitor the validation loss to understand if the model is overfitting. On this view, *EarlyStopping* function is a callback allowing to specify the performance measure to monitor the trigger and stop the training process. To quantify a minimum improvement, a minimum delta was also stated. We always restored the model weights from the epoch with the best

Figure 3 - The stacked model architecture. MobileNet and DenseNet121 are utilized as feature extractors, while other layers along with the sigmoid activation function are added by hand.

value of the monitored quantity (validation loss in our case).

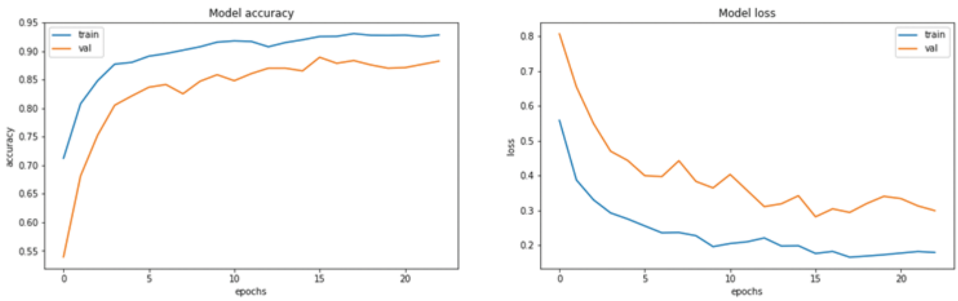
ReduceLROnPlateau is a function useful to reduce slightly the learning rate as soon as the validation loss has stopped improving. Also in this case, a minimum delta is needed for ensuring the new optimum and to only focus on significant changes.

Finally, *ModelCheckpoint* is used to save the model which is considered the best according to the validation loss minimum value.

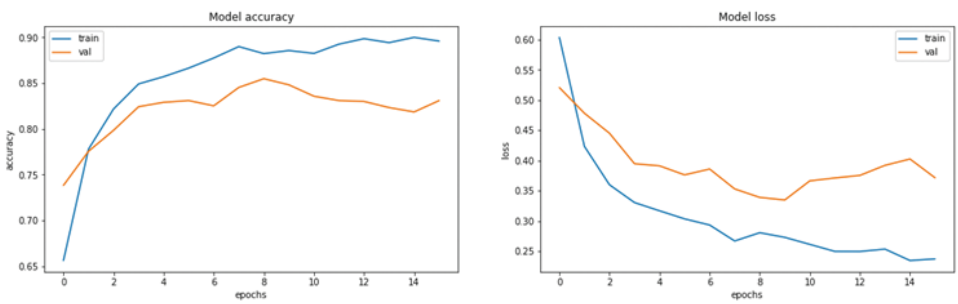
Results

Figure 4 shows the performance comparison on training and validation losses and accuracy between *MobileNet*, *DenseNet121* and the stacked Model. The stacked model is able to further reduce validation loss, leading to better results. More precisely, *MobileNet* reached a training loss of 0.1666, a training accuracy of 0.9322, a validation loss of 0.2814 and validation accuracy of 0.8892. *DenseNet121* reached a training loss of 0.2843, a training accuracy of 0.8807, a validation loss of 0.3346 and validation accuracy of 0.8481. Finally, the stacked model reached a training loss of 0.1754, a training accuracy of 0.9289, a validation loss of 0.2250 and validation accuracy of 0.9154.

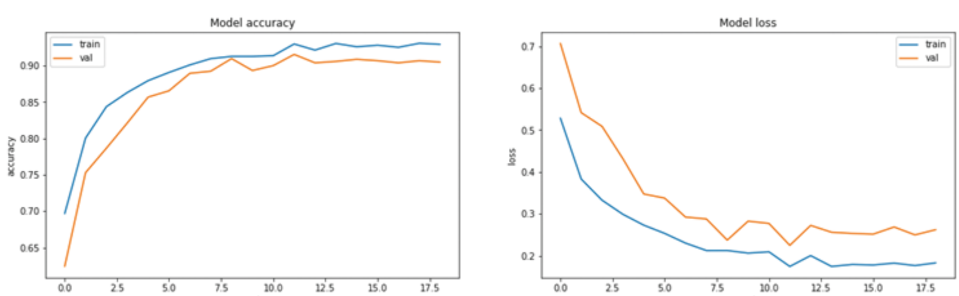
The expectations are satisfied by the test set as well. As a matter of fact, with the stacked model we reach a higher overall



a) MobileNet Accuracy and Loss on Training and Validation sets



b) DenseNet121 Accuracy and Loss on Training and Validation sets



c) Stacked Model Accuracy and Loss on Training and Validation sets

Figure 4 - The plots referring to a) MobileNet, b) DenseNet121 and c) Stacked model of the losses and accuracy on the training and validation sets.

Table 2 - Comparison of Testing accuracy and Losses between Stacked model, MobileNet and DenseNet121

Model	Testing accuracy	Testing loss
Stacked model	0.935	0.196
MobileNet	0.916	0.225
DenseNet121	0.897	0.268

accuracy and a lower loss on the test set, demonstrating a more robust and stable model. In Table 2 the final accuracy on the test set is shown.

To better understand how the three models performed in these binary classifications a comparison between true and predicted labels is shown in a heat map (Figure 5). This is done essentially to have an idea of how a good classification model should be, but also how it could be further improved when dealing with disease diagnosis, since this is sometimes crucial for patients' life.

In order to distinguish between lobar and interstitial pneumonia, since our stacked model outperforms the two based models used, we use the latter as pneumonia classifier. The input that was previously classified as pneumonia (373), has been classified again with a *DenseNet201* model, separately trained for this task. In figure 6 the confusion matrix related to this second stage is shown. The second-stage classifier has reached an accuracy of 97.1% when distinguishing between lobar and interstitial pneumonia (table 3). Considering that the first stage classification (obtained with our stacked model) obtained a 93.7% of accuracy, and taking into account also the misclassifications with respect to the healthy control group, then the final accuracy of the overall hierarchical system classifier resulted 91%, as reported in both table 3 and figure 7.

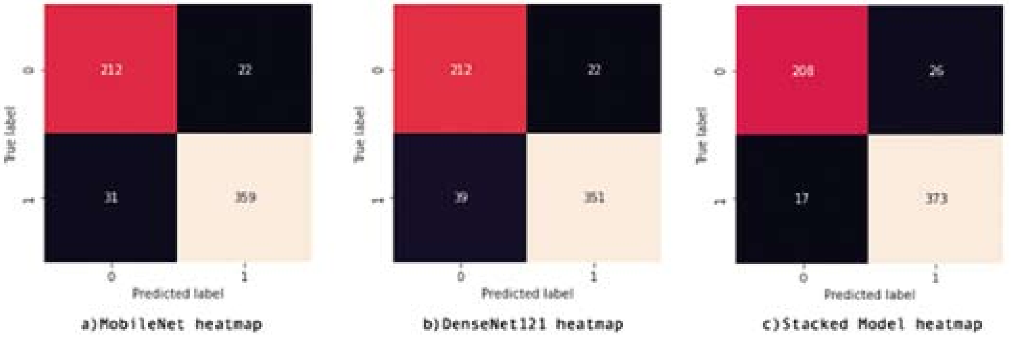


Figure 5 - The confusion matrix for each model containing true and predicted labels for both normal images (234) with class labelled with 0 and pneumonia images (390) labelled with 1. The confusion matrices compare the a) MobileNet, b) DenseNet121 and c) Stacked model.

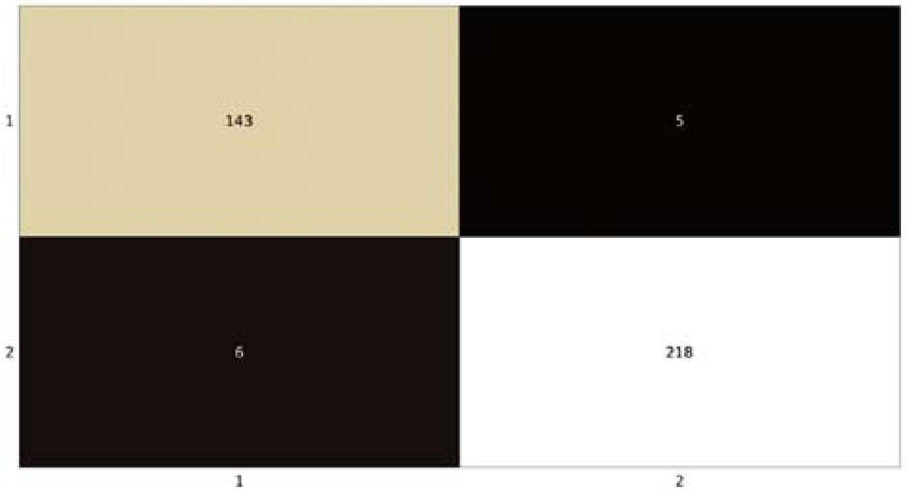


Figure 6 - The confusion matrix of the last stage for internal classification between Lobar and Interstitial pneumonia, respectively identified as class 1 and class 2. These latter stage have been applied for testing to the 373 pneumonia images successfully classified with the proposed stacked model containing true and predicted labels for both normal images (234) with class label 0 and pneumonia images (390) labelled with 1.

Table 3 - Final results of the overall classifier, in terms of discrimination accuracy between lobar and interstitial pneumonia and the obtained overall accuracy taking into account the overall 2-stage classification.

2^{nd} stage classifier accuracy	Overall hierarchical system accuracy
97.1%	91.0%

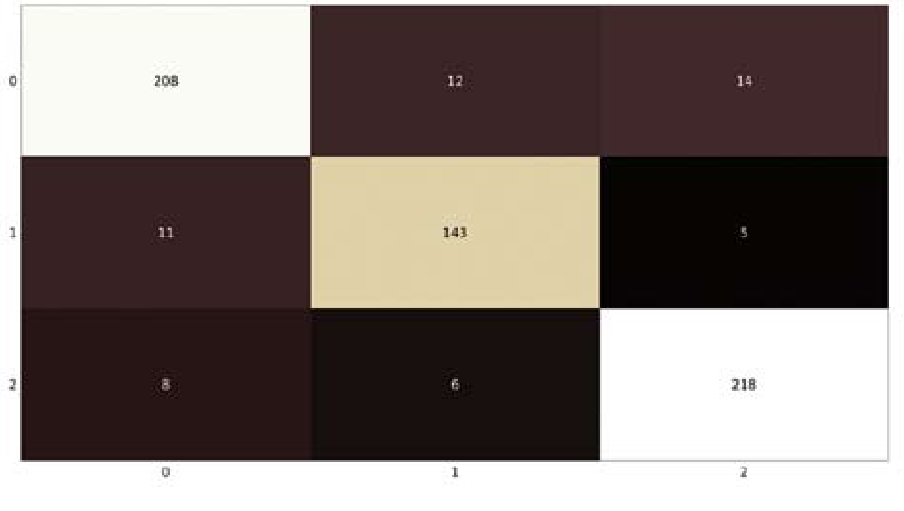


Figure 7 - The overall confusion matrix of the entire hierarchical system between the healthy control group (identified with the label 0), and the lobar or interstitial pneumonia, respectively identified as class 1 and class 2.

Discussion and Conclusions

Pneumonia is a common respiratory infection, affecting approximately 450 million people a year and occurring in all parts of the world. Laboratory methods are actually the gold standard to confirm a lung infection and to try to identify the type of organism causing the pneumonia.

The real-time PCR process is one of the fastest diagnostic methods, and it takes approximately 4–6 hours to obtain the test results (8). This can cause a diagnosis delay at early stages, worsening the prognosis of the pneumonia and, in some cases, allowing the contagion of other people inadvertently. Therefore, it is necessary to perform research and to develop new methods that help to provide computer-aided diagnosis to reduce pneumonia-related mortality or infection diffusion, especially in the developing countries (9). To shorten the diagnosis, radiology is fundamental in confirming the disease and monitoring its progression over time. Furthermore, computational models in the area of artificial intelligence and deep learning have been efficiently used in solving problems related to medical imaging (9, 11). Our study reports the results of the performance evaluation of the most used models, demonstrating that *MobileNet* and *DenseNet121* showed the highest accuracy. In particular, *DenseNet-121* is a model already pretrained, Ho et al (20) used a pretrained *DenseNet121* for the classification of 14 thoracic diseases, and Chouhan et al (9) suggested a novel deep learning framework for the detection of pneumonia using the concept of transfer learning. In this approach, five different models, including *DenseNet121*, were analysed and combined.

In our study, the accuracy further increases when these two models are combined in a staked one, able to further reduce validation loss, leading to better results. Our second stage accuracy level is similar to the one of Chouhan et al (9). Both the studies used the same free available database of images (17).

Nevertheless, it should be noted that image processing supported by the application of AI procedures may be a technology providing fast and accurate results, but, considering that it deals with patients' lives, the highest accuracy of the model must be ensured.

The authors are aware of some limits of the study. It is demonstrated that several protocols used for automatic pneumonia diagnosis directly on X-Ray images are not reliable and that the neural networks are learning patterns in the dataset that are not correlated to the presence of pneumonia (21). These protocols might be biased and may learn to predict features by relying more on the source dataset than on relevant medical information (21). Therefore, creating a reliable testing protocol is a challenging task. In fact, even if our model performance was good, still 43 samples resulted incorrectly classified: 26 for false normal images and 17 for false pneumonia images. This could lead, in a real health-care environment, to mistakenly classify a patient as healthy or sick. In this research, we used two models, *MobileNet* and *DenseNet121*, as feature extractors, but the stacked model is more performant than the two chosen models, further increasing the accuracy.

Despite these limits, and the small and imbalanced dataset, the results obtained may be considered overall satisfactory, showing an efficient and robust model. Still, introducing a better and more balanced dataset could further improve the model's performances, resulting in more precise and accurate outcomes.

In conclusion, the hierarchical stacked model designed in the present study seems able to detect pneumonia with an accuracy higher than the single models; moreover, the accuracy is high, especially when distinguishing between lobar and interstitial pneumonia. Our findings support the notion that deep learning methods can be used to simplify the diagnostic process and improve disease management, highly contributing to public health purposes. However, it would be helpful to have a larger dataset, especially regarding the data validation process. In addition, it should be taken into account the possible misclassification (21).

Riassunto

Modelli convoluzionali sovrapposti per il rilevamento automatico di polmonite da rx torace, nuove strategie per la sanità pubblica

Introduzione. Al fine di supportare medici ed, in particolare, specialisti in radiologia nella diagnosi di polmonite, diversi sistemi di intelligenza artificiale sono stati descritti in letteratura.

Metodi. In primo luogo sono stati identificati i modelli convoluzionali più performanti in termini di accuratezza e successivamente descritti. Utilizzando le più rilevanti caratteristiche estratte dai due modelli più performanti è stato successivamente costruito un modello integrato. In fine, su questo modello integrato è stato impostato un successivo modello gerarchico per la sotto-classificazione delle polmoniti in lobari (infiammazione polmonare dello spazio alveolare) e interstiziali (infiammazione polmonare con interessamento del tessuto interstiziale).

Risultati. I risultati hanno mostrato che il modello integrato presenta una accuratezza maggiore dei due modelli di origine. Tale caratteristica applicata alla diagnostica automatica delle polmoniti può risultare di fondamentale importanza nella gestione dei pazienti in un'ottica di sanità pubblica.

Conclusioni. Nonostante alcuni limiti, il presente studio supporta le evidenze scientifiche che mostrano come sistemi di intelligenza artificiale possano essere utili per semplificare i processi diagnostici e migliorare il management di alcune patologie.

References

1. Napoli C, Dente MG, Kärki T, Riccardo F, Rossi P, Declich S; Network for the Control of Cross-Border Health Threats in the Mediterranean Basin and Black Sea. Screening for Infectious Diseases among Newly Arrived Migrants: Experiences and Practices in Non-EU Countries of the Mediterranean Basin and Black Sea. Int J Environ Res Public Health. 2015 Dec 8; **12**(12): 15550-8. doi: 10.3390/ijerph121215002.
2. Napoli C, Riccardo F, Declich S, et al. An early warning system based on syndromic surveillance to detect potential health emergencies among migrants: results of a two-year experience in Italy. Int J Environ Res Public Health. 2014 Aug 20; **11**(8): 8529-41. doi: 10.3390/ijerph110808529.
3. Illari S I, Russo S, Avanzato R, Napoli C. A cloud-oriented architecture for the remote assessment and follow-up of hospitalized patients. In: SYSTEM 2020: 5th Symposium for Young Scientists in Technology, Engineering and Mathematics, May 20 2020. CEUR Workshop Proceedings 2020; **2694**: 29-35.
4. Napoli C, Salcuni P, Pompa MG, Declich S, Rizzo C. Estimated imported infections of Chikungunya and Dengue in Italy, 2008 to 2011. J Travel Med. 2012 Sep-Oct; **19**(5): 294-7. doi: 10.1111/j.1708-8305.2012.00640.x. Epub 2012 Aug 8.
5. Capalbo C, Aceti A, Simmaco M, et al. The Exponential Phase of the Covid-19 Pandemic in Central Italy: An Integrated Care Pathway. Int J Environ Res Public Health. 2020 May 27; **17**(11): 3792. doi: 10.3390/ijerph17113792.
6. Roma P, Monaro M, Muzi L, et al. How to Improve Compliance with Protective Health Measures during the COVID-19 Outbreak: Testing a Moderated Mediation Model and Machine Learning Algorithms. Int J Environ Res Public Health. 2020 Oct 4; **17**(19): 7252. doi: 10.3390/ijerph17197252.
7. Capizzi G, Lo Sciuto G, Napoli C, Połap D, Woźniak M. Small lung nodules detection based on fuzzy-logic and probabilistic neural network with bioinspired reinforcement learning. IEEE Transactions on Fuzzy Systems 2020 June; **28**(6): 1178-89. doi: 10.1109/TFUZZ.2019.2952831.
8. Shorfuzzaman M, Masud M, Alhumyani H, Anand D, Singh A. Artificial Neural Network-Based Deep Learning Model for COVID-19 Patient Detection Using X-Ray Chest Images. J Healthc Eng 2021 Jun 5; **2021**: 5513679. doi: 10.1155/2021/5513679.
9. Chouhan V, Singh SK, Khamparia A, et al. A novel transfer learning based approach for pneumonia detection in chest x-ray images. Appl Sci. 2020; **10(2)**: 559. https://doi.org/10.3390/app10020559.
10. Woźniak M, Połap D, Napoli C, Tramontana

E. Graphic object feature extraction system based on cuckoo search algorithm. Expert Systems with Applications 2016; **66**: 20-31. doi: 10.1016/j.eswa.2016.08.068.

11. Oh Y, Park S, Ye JC. Deep learning COVID-19 features on CXR using limited training data sets. IEEE Trans Med Imaging 2020 Aug; **39**(8): 2688-700. doi: 10.1109/TMI.2020.2993291. Epub 2020 May 8.
12. Jain R, Gupta M, Taneja S, Hemanth DJ. Deep learning-based detection and analysis of COVID-19 on chest x-ray images. Appl Intell. 2021; **51**: 1690-700. Epub 2020 Oct 9. https://doi.org/10.1007/s10489-020-01902-1.
13. Pereira R M, Bertolini D, Teixeira LO, Silla CN, Costa YMG. COVID-19 identification in chest x-ray images on flat and hierarchical classification scenarios. Comput Methods Programs Biomed. 2020 Oct; **194**: 105532. doi: 10.1016/j.cmpb.2020.105532. Epub 2020 May 8.
14. Toğaçar M, Ergen B, Cömert Z, Özyurt F. A Deep Feature Learning Model for Pneumonia Detection Applying a Combination of mRMR Feature Selection and Machine Learning Models. IRBM. 2020 Aug; **41**(4): 212-22. https://doi.org/10.1016/j.irbm.2019.10.006.
15. Rahman T, Chowdhury MEH, Khandakar A, et al. Transfer learning with deep Convolutional Neural Network (CNN) for pneumonia detection using chest X-ray. Appl Sci. 2020; **10**(9): 3233. https://doi.org/10.3390/app10093233.
16. Hammam AA, Elmousalami HH, Hassanien AE. Stacking Deep Learning for Early COVID-19 Vision Diagnosis. In: Hassanien AE, Dey N, Elghamrawy S, eds. Big Data Analytics and Artificial Intelligence Against COVID-19: Innovation Vision and Approach. Studies in Big Data. Cham: Springer, 2020: 297-307.
17. Kermany DS, Goldbaum M, Cai W, et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell. 2018 Feb; **172**(5): 1122-31. doi: 10.1016/j.cell.2018.02.010.
18. Howard AG, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv 2017; 1704.04861.
19. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21-26 July 2017. IEEE, 2017: 4700-8. doi: 10.1109/CVPR.2017.243.
20. Ho TKK, Gwak J. Multiple feature integration for classification of thoracic disease in chest radiography. Appl Sci. 2019; **9**(19): 4130. https://doi.org/10.3390/app9194130.
21. Maguolo G, Nanni L. A critic evaluation of methods for COVID-19 automatic detection from X-ray images. Inf Fusion. 2021 Dec; **76**: 1-7. doi: 10.1016/j.inffus.2021.04.008. Epub 2021 Apr 30.

Corresponding author: Enrico Bertamino, MD, Sant'Andrea Hospital, Via di Grottarossa 1035/1039, 00189 Rome, Italy
e-mail: ebertamino@ospedalesantandrea.it