Studying the Behavior of a Modified Deep Learning Model for Disease Detection Through X-ray Chest Images

Elma Zanaj, Lorena Balliu, Gledis Basha, Elunada Gjata, Elinda Kajo Meçe
Faculty of Technology of Information, Polytechnic University of Tirana, Tirana, Albania

***Abstract*—In modern medical diagnostics, Deep Learning models are commonly used for illness diagnosis, especially over X-ray chest images. Deep Learning approaches provide unmatched promise for early identification, prognosis, and treatment evaluation across a range of illnesses, by combining sophisticated algorithms with large datasets. It is crucial to research these models to lead to improved ones to progress toward disease identification's precision, effectiveness, and scalability. This paper presents the study of a CNN+VGG19 Deep Learning architecture (subsets of machine learning), both before and after its modification. The same dataset is used over the existing and modified models to compare metrics under the same conditions. They are compared using metrics like loss, accuracy, precision, sensitivity, and AUC. These metrics display lower values in the updated model than in the original one. The numbers demonstrate the occurrence of the overfitting phenomenon, which is most likely the result of the model's increased complexity for a small dataset. The noise in the images included in the dataset may also be the cause. As a result, it can be stated that regularization techniques should be applied; otherwise, layers of extraction and classification should not be added to the model to prevent overfitting.**

Keywords—Machine learning; big data; X-ray chest image; CNN; VGG19

I. INTRODUCTION

In recent years, Deep Learning techniques related to pulmonary disorders have become more and more useful and more and more studied. These research presented us with the reality of new or modern medicine. Numerous studies conducted in this area, used it, aiming to identify the ideal architecture for lung diagnostics. Chest X-ray (CXR) and Computed Tomography (CT) scans are often the most widely used datasets. Many researchers have examined these datasets separately or combined.

According to these studies, datasets that combine CXR images with CT scans are currently the most utilized. One of the base works in this area is the multi-class classification categorization of lung illnesses using Deep Learning architectures like CNN and VGG19, which both, separately are known for their simplicity and effectiveness.

A different Deep Learning model is applied for the first time in a study that is considered as a basis for the work that will be implied in this paper. In that study, a multiclass classification for lung conditions applies datasets of CXR images. CNN and VGG19 are utilized to implement the DL model. The dataset is split into training and validation sets following the standard procedure for using this kind of method. Therefore, the dataset was split into 80% training and 20% validation for a total of 5000 epochs of model training. This model generated very good performance metrics. The core model, in which this paper is based on, reached accuracy (96.48%), recall (93.75%), precision (97.56%), F1 score (95.62%), and area under the curve (AUC) (99.82%).

While the model's performance and results are outstanding, there is still much space for further research, particularly in Deep Learning architectures. Customization is used to maximize benefits from them. This would result in models that are more adaptable and efficient. To enhance the interpretation of pulmonary pathology, it is possible to combine both CT and CXR images. This is what this study includes within the framework of Deep Learning algorithms.

This paper studies CNN+VGG19, a Deep Learning architecture. Its core, or basic form got from [1], is studied first, followed by its modified form. Emphasizing the behavioral characteristics of the model in both versions is the primary goal. To see the comparison of the measures of loss, accuracy, precision, sensitivity, and AUC, this study will use a standardized dataset. As a result, conclusions concerning the model's efficacy and adaptability are stated via conclusions reached.

The Region of Interest (ROI) is another fascinating feature that is highly desirable to investigate. It can be a future flow to follow since it is becoming more and more common in many applications connected to object detection, image processing, and medical diagnostics. Based on recent research, it has been suggested that enhanced performance may result from incorporating feature extraction and classification layers with the preceding Deep Learning techniques. This would significantly simplify the duties and the application itself.

This DL model's results provide a significant contribution to the model's complexity. The two indicated layers that were added to the model together with the noise are not improving the model. The results will show that it will lead overfitting, which can be result on by noise in the data, a shortage of trainable samples, or a more complicated model. This study demonstrates that adding noise to the images did not improve their variance, but rather made it more difficult for the model to detect the proper logic patterns during training. Noise inclusion should be removed, and other tests done on a larger dataset with the modified model.

The structure of the paper is as follows: Section II provides an overview of the literature of previous research on lung image detection using CNN and deep learning. The study approach, the initial model, and the changes that resulted in the new model examined in this work are all included in Section III. The results are displayed at Section IV. Discussions continue in Section V. Finally, the Conclusion in Section VI is going to concentrate on resume of the work done and will give suggestions for further works.

II. Literature Review

Numerous studies have worked over the use of CNN and Deep Learning to detect or classify lung illnesses. Their goal is to identify the top Deep Learning model that allows the highly accurate diagnosis of different illnesses. We can look at cases one by one or with other methods or data.

CNNs are a part of Artificial Neural Networks (ANN) and help with medical image analysis, through feature extraction and learning [2]. CNN is well renowned for its superior performance in several 2D and 3D medical image processing applications, including classification, segmentation, and detection. A feature filter that is placed on top and slides along the input layer of the neural network executes the convolution process. As a result, is generated a feature map. The layer that performs the convolution process is called the convolution layer. 2D CNN convolution is done through extracting features only from a 2-dimensional space [3]. The value of a unit at (x,y) in the i layer in the j feature, defined as: $\mathbf{v}_{\mathbf{ij}}^{\mathbf{xy}}$ is given by Eq. (1).

$$v_{ij}^{xy}=f\left(b_{ij}+\sum_m \sum_{p=0}^{P_i-4} \sum_{q=0}^{Q_i-1} w_{ijm}^{pq}\, v_{(i-q)m}^{(x+p)(y+q)}\right) \quad (1)$$

Where: f is the activation function, b_{ij} is the addition of the feature map, m is the number of filters in the (i-1) layer, w_{ijm}^{pq} is the value for the position (p, q) of the connected particle in the map of the k characteristics and Pi and Qi represent respectively the length and width of the particle.

3D CNN convolution is performed through the same concept but applied over a 3-dimensional space. Eq. (1) is modified and expanded to lead to Eq. (2) as follows:

$$v_{ij}^{xy}=f\left(b_{ij}+\sum_m \sum_{p=0}^{P_i-4} \sum_{q=0}^{Q_i-1} \sum_{r=0}^{R_i-1} w_{ijm}^{pqr} v_{(i-q)m}^{(x+p)(y+q)(z+r)}\right) \quad (2)$$

Where: Ri is the size of the 3D particle together with the third spatial dimension and w_{ijm}^{pqr} is the value of the (p, q, r) particle related to the previous m-layer map.

VGG19 has 19 layers and is a complicated convolutional neural network. According to [4], increasing recognition or classification accuracy requires a certain level of network depth. At [5], chest X-ray images are divided into three primary groups using a hybrid DCNN technique. DCNN hybrid network is created by the Inception module and VGG blocks together. A precision rate of 99.25%, a Kappa-score of 99.10%, an AUC of 99.43%, an F1-score of 99.24%, and a recall of 99.25% were reached when applied to the dataset by the suggested strategy. The results of the experiments demonstrate the efficiency and robustness of the hybrid DCNN mechanism utilized in this work.

The LungNet22 is developed using new layers and hyper parameters, building upon the VGG16 [6]. After the fifth block of VGG16, two blocks are connected to produce LungNet22. In the sixth block are three structures and a GlobalAveragePooling2D structure, while in the seventh block is a scatter structure linked to a dense structure. AUC values and the ROC curve were among the performance indicators that were computed to confirm the model's effectiveness. The proposed model obtained an estimated accuracy of 98.89% with the use of the Adam Optimizer.

Certain transfer learning models, including InceptionV3, AlexNet, DenseNet121, VGG19, and MobileNetV2 in study [7], incorporate pre-processed images. MobileNetV2 has outperformed the other models with an overall classification accuracy of 91.6%. In the next step, this model is built up to maximize the performance of MobileLungNetV2. The enhanced model, MobileLungNetV2, obtains an exceptional classification accuracy of 96.97% on the pre-processed data. It is found that the model has the following values: 96.71%, 96.83%, and 99.78% for precision, recall, and sensitivity, respectively, using a confusion matrix for each class.

Another study in [8] uses CNN to learn lung illness images to deeper into the multi-category classification method. Training data came from the Cheonan Soonchunhyang University Hospital dataset, which included tuberculosis, and the National Institutes of Health (NIH) datasets, which were split into: Normal, Pneumonia, and Pneumothorax categories. Preprocessing of the center crop was carried out while maintaining a 1:1 aspect ratio to increase performance. Weights from ImageNet improved learning, and Multi GAP was used to optimize each layer's features. With an accuracy of 85.32%, it thus achieved the best performance out of all the examined models. The average score for the predictions was 96.1%, with a sensitivity of 92.2%, specificity of 97.4%, and assessment time of 0.2 s.

An intriguing work, [9], uses an enhanced augmentation strategy to modify the CNN model for identifying lung cancer biopsy images by employing the pre-trained Visual Geometry Group19 (VGG19) model. These two methods greatly improve lung cancer diagnosis in histopathology images. The recorded accuracy reached 97.73%. The suggested approach outperforms the current methods in the experiment findings. The amount of the training data has a direct impact on the neural network's performance. A large set of data can be used to train efficiently a neural network. For moreover, this method also reduces network pattern overlap. This work combines color leveling and transformation to improve image techniques. This technique works well, especially at reducing tone and intensity changes in input images. This increases the classifiers' ability to forecast the future.

There are three phases that contribute to [10]'s experimental investigation. To prepare the images for use as input in the CNN model, each image is first processed by scaling them to 224 by 224 pixels and converting them to RGB format. Next, the data (the image intensity) is transformed to the range 0 to 1 to normalize it. In the next phase, the image is randomly rotated 15 degrees clockwise to expand the data because there aren't enough trained images in the work and to ensure sure the model generalizes.

The third step uses transfer learning to identify objects or images in new categories. It has not been trained for the new ones. In this case was implemented a multi-step Keras "fine-tuning" approach. The network is first "frozen" to prevent the backpropagation passport from reaching any of the layers below the top. Second, at the conclusion of the network, the connected completion nodes are removed and replaced with newly set up ones. After that, training is limited to the connected completeness layer's upper layer.

Due to the recent developments in the medical diseases, most of the articles of the recent years focused in the Covid-19 detection. The research regarding the classification of lung diseases is not limited to a restricted number of models anyway. Research has a wide scope, and different studies bring different results. We can mention the following articles as examples:

In study [11] it is emphasized the how important are the predictions related to the X-ray images. The paper focus in chest diseases like tuberculosis (TB), COVID-19, and pneumonia. The study is done based on the analysis of three CNN models: VGG19, Resnet50V2, and Densenet201.

Evaluation of the predictions are done bases on indicators like Accuracy and Loss. All three models demonstrate great accuracy and consistency. But there are taken in consideration other aspects including training efficiency and complexity of the architecture. After comparison the best option out of the three is considered Resnet50V2.

The aim of project [12] is the creation of a Medical Diagnosis Support System (MDSS) created for the x-ray images related to Covid-19. The diseases taken in consideration are COVID-19, Normal, Pneumonia, and Tuberculosis. MDSS uses a combination of pre-trained convolutional neural networks (CNNs) based on Transfer Learning (TL) classifiers. The accuracy of Covid-19 detection increased using of a parallel deep feature extraction method based on Deep Learning (DL). The concatenation classifier was noticed to give good accuracy rates.

The study presented in study [13] shows an innovative two-dimensional CNN (2D-CNN) architecture developed especially for COVID-19 classification. The aim of the model is to make a clear division between viral pneumonia, which is typical of COVID-19 and other forms of pneumonia or a fully healthy lung image.

The design (especially the depth distinct layer layout) and of the suggested 2D-CNN architecture is done to maximize the accuracy of the model disease prediction. The model performed well in preliminary testing, attaining high levels of sensitivity, specificity, and accuracy. It is also important to mention that the design of the model enables a smooth and easy integration into existing medical imaging workflows.

The [14] study utilizes an approach that preprocesses chest photos using techniques such as histogram equalization and sharpening. Feature maps are used in the model to include a self-attained mechanism that further improves the performance of CNNs. Based on the stimulations, it is seen that the Inception-Resnet CNN is a more flexible and efficient way to classify and CT images than classic segmentation techniques. With better results of accuracy, sensitivity, the Inception-Resnet model demonstrates its efficacy in COVID-19 classification.

In study [15], it is employed a deep learning approach, specifically using CNNs, to enhance pneumonia detection. It is utilized the VGG-19 model, part of the CNN framework, on both original and augmented CXR image datasets. Augmented images were generated from the existing dataset to improve the model's performance. The techniques implemented in this approach include image scaling, data augmentation, deep learning with Keras, batch normalization, and utilizing weights from the pre-trained VGG-19 model. The proposed model achieved a 95% accuracy rate on the augmented CXR image dataset.

In study [1], presented a DL classification model with the aim to identify the most prevalent chest conditions. The aims are to develop a DL framework and categorize various forms of pneumonia, lung cancer, emphysema, TB, and, most recently, COVID-19. After reviewing the literature, this study appears to be the first to categorize all six classes simultaneously using a single DL framework. According to experimental findings obtained from [1], VGG19 + CNN model reached an 96.48% accuracy, 93.75% recall, 97.56% precision, 95.62% F1 score, and 99.82% AUC.

In conclusion, many models have been trained by studies utilizing CXR and CT image datasets. These models include CNN hybrid network, LungNet22, MobileNetV2, CNN GD, Inception V3, Resnet-50, VGG-16, and VGG-19. CXR images make disease diagnosis easier compared to CT images with these patterns.

III. Research Methodology

The work featured in this paper is based on study [1], which uses for its testing a combination of CNN and VGG19 models.

The models used in the existing and in the new model created are still the CNN+VGG19 Deep Learning architecture. The same dataset is used over the existing and modified models to compare metrics under the same conditions.

The first part of the section describes the core model, what does it combine and how the classification and extraction process is done through it. The second part of the section gives a detailed information of the dataset used, data image partitioning into 20% random validation and 80% training. The last part of the section is described the use of Deep Learning techniques through VGG19 and CNN algorithms and the changes done over the base model.

All the base and new model codes are written and run in Python language by using the libraries that this language offers.

A. Base Model and Changes

Base models used on the original work are transformed in the new one through modifications as described below. The main aim and the input effort is to improve performance results.

1) Combination of CXR and CT images: To modify Deep Learning architectures usually are used several frameworks. It is a common practice in the field of computational medicine and Machine Learning in medicine, combining X-ray (CXR) and Computed Tomography (CT) images to create a dataset used then in Deep Learning models.

2) Increasing the levels of classification and extraction: Increasing the classification and extraction levels in Deep Learning models is an important approach to advance the capabilities and performance of Deep Learning models in various tasks.

The two models used are: CNN and VGG19. The proposed framework in this study is divided into three phases: pre-processing, feature extraction and classification.

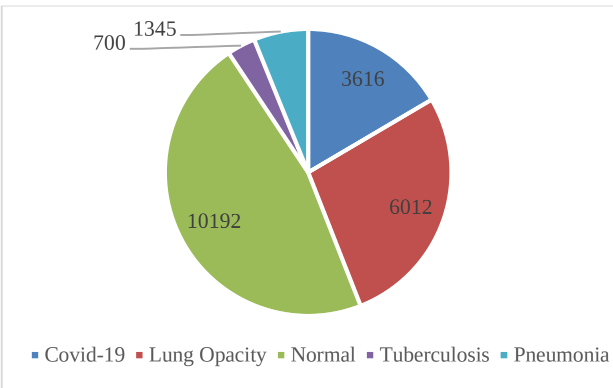

Fig. 1. Distribution of images dataset.

B. Dataset

Classification of the images is done under five main categories: Normal, Covid-19, Lung Opacity, Tuberculosis and Pneumonia.

As mentioned in this study, there are used two main data sources that feed the trained models: CXR images and CT images. Fig. 1 describes the dataset used. It is a total of 21’865 images out of which: 3’616 are images of lungs affected by Covid-19, 6’012 images of lung opacity, 10’192 images of normal lungs, 700 images of lungs of affected by tuberculosis and 1’345 images of lungs affected by viral pneumonia. The model starts with the procession of the images, which was trained using the images in our dataset. Rescaling the images is necessary to ensure that the pixels accept only values between 0 and 1. If any pixels are vacant, the value of the closest pixel is used to fill them.

This image set will be divided into two parts: twenty percent for training and the remaining portion for validation. This suggests that the model will become used to the visual pattern by using the training set. Following that, the capacity of the model to locate and classify the validation set images within any certainty degree of accuracy, is used for evaluating its performance. The Gaussian noise function has been added to this data set in the modified model.

C. The Redesigned Model

The fundamental model is modified through interfering with the feature extraction convolution layers and altering the filter types that eventually perform image classification. The first model, VGG19, is initialized without any fully connected layers. The input is specified as 224x224 pixels and three RGB channels and are utilized pre-trained weights from the ImageNet dataset. The first changed model has its trainable parameter set to false, meaning that the VGG19 layers are not moving. This implies that no changes will be made to the pre-trained weights when the next layers are being trained.

The initial model had trainable parameters for all cases. There were 24’622’341 parameters in all, and each epoch's average training time was seven minutes in the original model. The total number of parameters employed in the modified model is 31’832’837. An epoch's training takes three minutes on average.

A Reshape layer is applied to the redesigned model on top of the VGG19 layer. The output of the preceding layers is now transformed into a shape with the dimensions (7, 7, 512). It is then added a convolution layer with the following parameters:

- 512: The quantity of filters (or channels in the final image).
- 5: The 5x5 kernel size.
- padding="same": This sets a padding between the circles' matching zeros to maintain the same dimensions between the input and output images.
- kernel_initializer='random_normal': This initialization technique sets the filter weights using values generated by a random normalization.
- bias_initializer='zeros': Set the bias to zero at starting.
- regularizers = kernel_regularizer.l2(0.01): Apply L2 regularization with a 0.01 filter weight penalty factor to lessen overfitting.
- bias_regularizer=regularizers.l2(0.01): Apply L2 regularization with the same penalty factor for biases as well.

Other convolution layers that include up to 64 filters, a further halving of the original number, come after it. Classification is the one performed further using fully connected layers. In contrast to the original model, a dense layer containing 1024 neurons has been implemented.

Adam's optimization function will be applied to both models, with the default and most used learning rate of 0.000009.

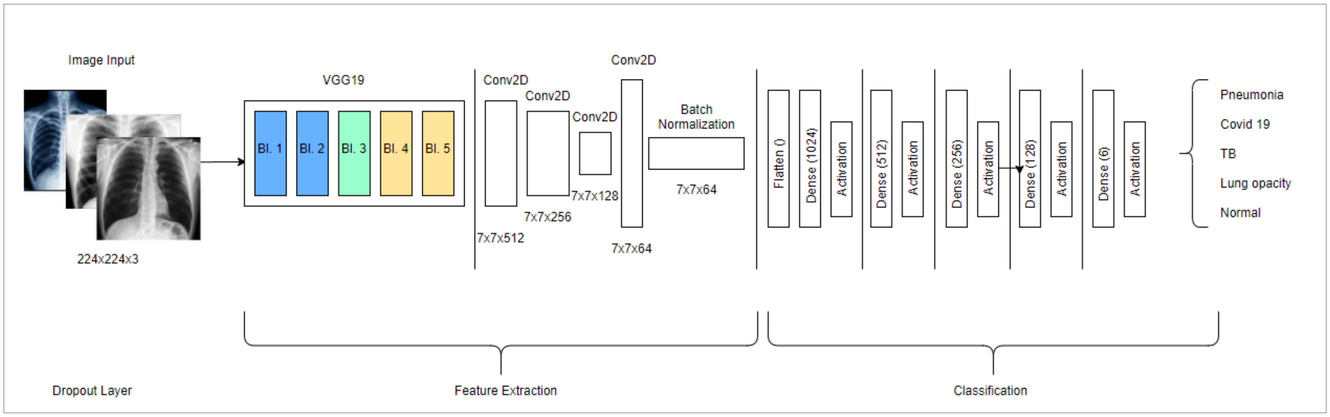

Fig. 2. Architecture of the modified model.

Below is shown the DL model in a pseudo-code summarized form. Fig. 2 shows the VGG19+CNN model architecture after modifications. It there can be visibly identified the changes done in the extraction and the classification part. We'll look at the metrics for AUC, precision, sensitivity, and accuracy to assess the model's performance in this task.

Algorithm 1: DL Model

Input: Dataset with chest images (CXR + CT)

Pre-processing: Image resizing to 224*224*3 and Square the image pixel values inside the interval [0, 1].

Splitting training data (80, 20): 20% is used for validation and 80% for training.

Using a new VGG19+CNN model for feature extraction.

Image classification from fully connected networks.

The cross-entropy loss will also be included. This will be done by comparing the model's predictions with the actual data labels. If the model receives higher validation accuracy values than the originally saved model, it will replace the current weight file. The initial model is trained in the tests done under this study, in its initial form, for 1000 epochs. The metric data, from the training and validation sets, were examined.

Following, the new model is trained for 5000 epochs. Same, the metric data from the training and validation sets were examined and compared with the ones reached from the original model training.

IV. RESULTS

The model's performance was evaluated considering loss, accuracy, precision, AUC, and recall or sensitivity. Each of the metrics is analyzed both for the base [1] and the new model separately in the below sections. The comparison of results obtained from the training, of both the original and new model, together with the comments regarding each of them will be analyzed in the following results.

A. Loss

Loss is a measure of how well the model's prediction matches the true labels or targets. It determines the error between the predicted output and the actual output during training. We need it to be as small as possible so that the model can improve and make better predictions.

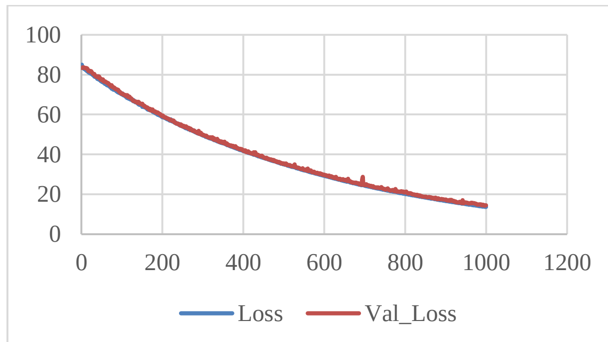

Fig. 3. Loss of the original model.

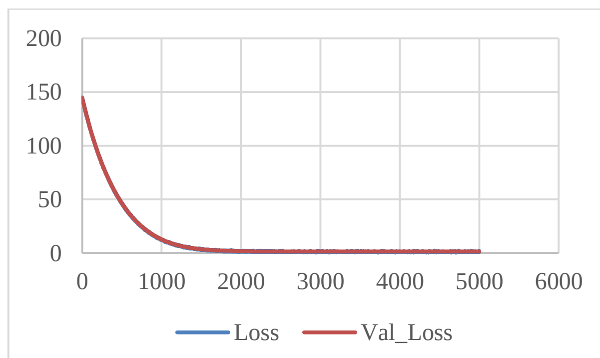

Fig. 4. Loss of the modified model.

Fig. 3 shows the case of original model, where the loss has experienced an exponential drop, reaching a value of 14.4 after 1000 epochs. The same results are observed on both the training and validation datasets, the first loss metric decreases, ruling out the possibility of overfitting. Fig. 4 shows the loss metric decays exponentially for both training and validation sets in the new model.

B. Accuracy

The accuracy is usually used to show the overall performance of the model. The goal of the metric is to be high in both the training and validation sets.

Usually, the accuracy of training is higher than that of validation, especially when the model is overfitting. Fig. 5 and Fig. 6 relate to the training accuracy, respectively over the original and new model. It has a deviation between 0 and 1, but on the other hand, the validation accuracy is zero in the case of the new model in contrast to the original one. This means we have overfitting on the new model.

C. Precision

Precision is the most useful metric, especially when the false positives are high. A high precision indicates that the model has fewer false positives, i.e. it is making fewer incorrect positive predictions. Precision of the original model is presented at Fig. 7.

At the end of the epochs, we see there is a high value around 1. Fig. 8 relates to the precision of the new model. It shows the training precision deviates from 0 to 1, but the validation one is zero, which means that we are facing the overfitting phenomenon.

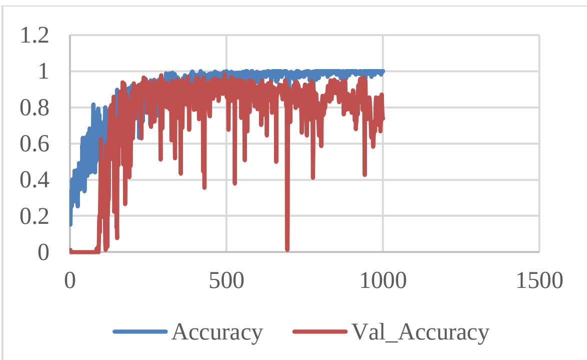

Fig. 5. Accuracy of the original model.

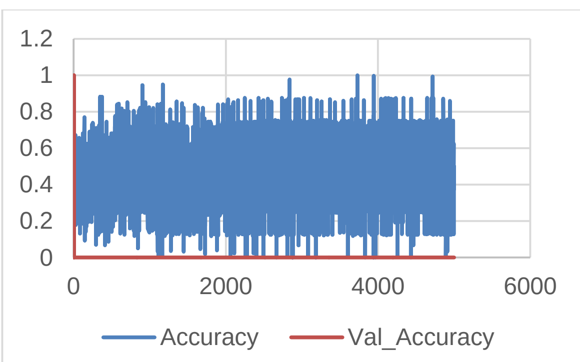

Fig. 6. Accuracy of the modified model.

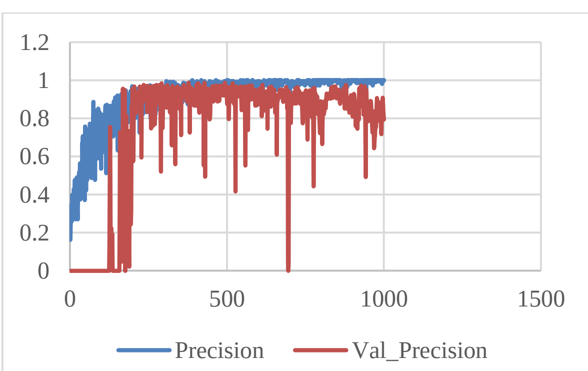

Fig. 7. Precision of the original model.

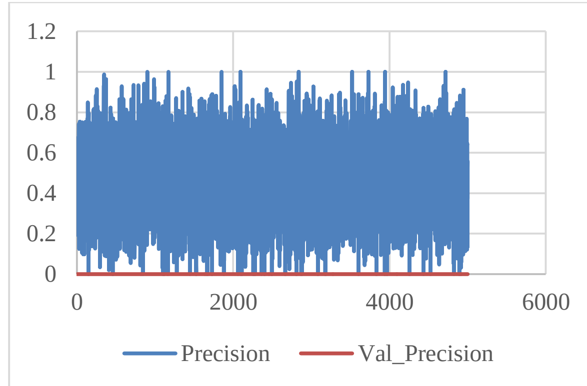

Fig. 8. Precision of the modified model.

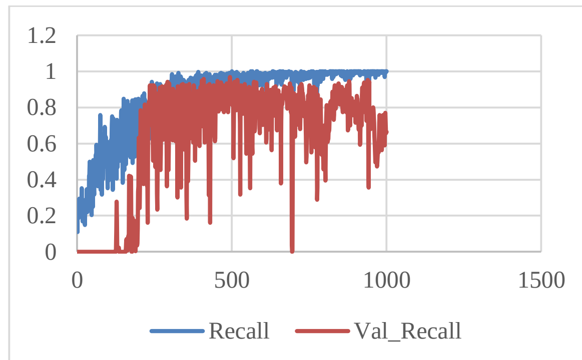

Fig. 9. Sensitivity of the original model.

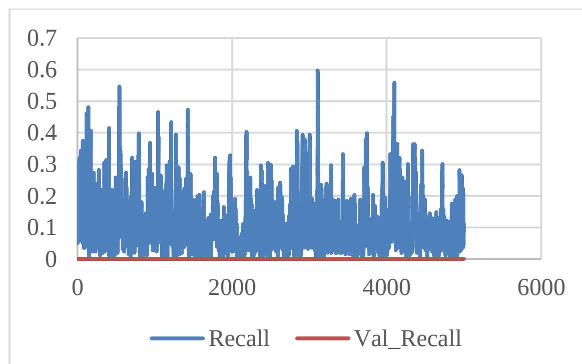

Fig. 10. Sensitivity of the new model.

D. Sensitivity

The sensitivity is particularly useful when the goal is to reach as many positive instances as possible, even at the cost of some false positives. Fig. 9 and Fig. 10 are displaying the sensitivity of the original and the new one. The values are a deviation between 0 and 1 and that of the validity is 0. Again, we are facing overfitting.

E. AUC

AUC measures the ability of the model to distinguish between positive and negative classes at different thresholds. A high AUC value indicates a better overall performance. Fig. 11 and 12 are the AUC of the original and new models. In the new one training varies between 0.5 and 1 and the AUC of validation is mostly zero, so we have the phenomenon of overfitting again.

It was noticed that the phenomenon of overfitting appeared in the modified model.

V. Discussion

We observed from the performance metrics of the modified model that we're dealing with the phenomenon of overfitting. The below sections discuss the overfitting phenomena and some suggestions on how to fix it.

A. Overfitting

Overfitting occurs whilst a system gaining knowledge of the model learns the training information so nicely that it begins to pick up noise or fluctuations in the data as opposed to studying the underlying version that generalizes properly to new, in no way-seen facts.

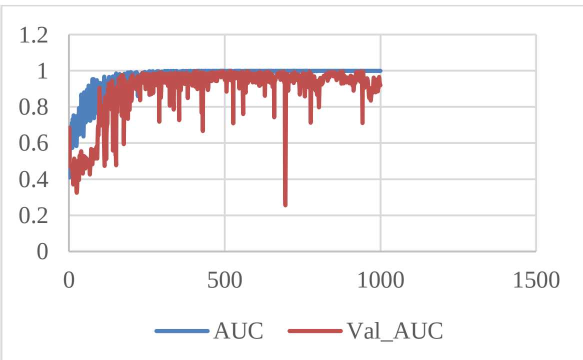

Fig. 11. AUC of original model.

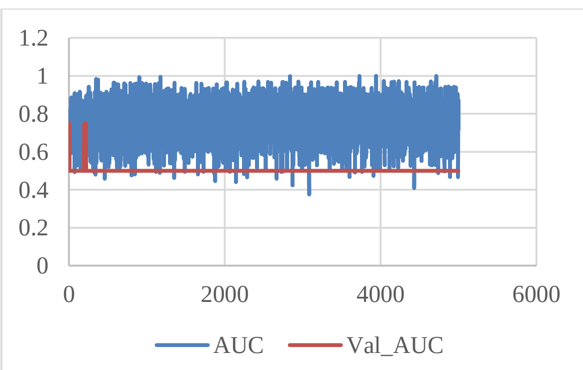

Fig. 12. AUC of the new model.

There are several elements that may have motivated the overfitting of our model:

1) Complexity of the version: Our changed model can be more complex than it should be compared to the size and complexity of the schooling facts given that it could memorize the trainable examples in preference to the underlying sample. This is the case while the deep getting-to-know version has many parameters, and our version went from 24’622’341 parameters to 31’832’837 parameters.

2) Lack of trainable records: While the training dataset is simply too small, the model cannot see many examples without the opportunity to master the variation underlying the model effectively. Compared to the complexity of the model, we can say that the dataset is exceedingly small.

3) Noise within the records: If the education statistics include noise or beside-the-point features, the version tends to mistakenly research those patterns as though they had been the patterns we need it to learn. In our version, we added Gaussian noise to growth the diversity of the records and make the version more robust.

B. Overfitting Adjustment Techniques

To keep away from overfitting in a Deep Learning model, numerous techniques may be used:

- L1/L2 regularization techniques optimize the loss function that penalizes massive weights, stopping the model from studying from very complicated underlying models.
- Dropout randomly turns on a fraction of neurons throughout education, forcing the version to study greater robust capabilities.
- Batch Normalization. Normalizes the activation of each layer to stabilize and accelerate training.
- Changes to the structure of the model making it less difficult with the aid of decreasing the wide variety of layers or gadgets. This reduces the capability of the version to memorize training records and encourages it to learn extra well-known patterns.

VI. Conclusion

Modifying architectures in Deep Learning is a powerful tool to improve the performance of models for various tasks. The changes made should be appropriate for the specific task and effectively define the needs of the model and the data available for training. In this paper, we aimed to create a new model bases on the architecture of the CNN+VGG19 model that was seen at [1]. Modifications are made to improve its performance metrics.

We studied the performance results of the modified model. It was observed that in the modified model the performance metrics deteriorated significantly compared to the original model. For each metric: accuracy, precision, sensitivity, F1-score, AUC, the model did not converge to a fixed value. It had all the time fluctuating values from 0.5 to 1. Moreover, the metric values of the validation set were lower than the values of the training set metrics. The difference was huge, so big that they went to zero most of the time.

Consequently, this means that we are dealing with the phenomenon of overfitting. It might have occurred due to the increase in the complexity of the model (adding of layer) for a not very large dataset. Another reason could be the noise of the images used in the dataset. To avoid overfitting, adjustment techniques such as L1/L2, batch normalization, dropout, or reducing the complexity of the model can be used.

The new model can only be used for small datasets because otherwise, it learns the pattern of images so well that it starts learning about image fluctuations, and noises, so making the model unusable. The introduction of noise elements should be eliminated as it was proved that it did not increase the variation of the images but pushed the model to not get the correct pattern of the logic detection during training.

The modified CNN+VGG19 model that we build did not achieve the desired improvements aimed. Our study provides valuable insights into the challenges of model modification and highlights areas for future research. In this work by increasing the complexity of the DL model for a not very large dataset or the increased noise of the images used in the dataset led to a poor performance of the model.

The results of this DL model give an important contribution related to the model complexity. It is not needed to increase complexity by adding the two mentioned layers and the noise on the same model, to be used for multi-class lung disease detection. We faced the overfitting, which can be result of model complexity, lack of trainable data or the noise in data. Other tests in the future can be made by keeping the number of layers of extraction and classification of the model intact but removing the noise. The introduction of noise elements should be made as it was proven that it did not increase the variation of the images but lead the model to get an incorrect logic pattern detection during training.

Limitation related to the study is the dataset. The model performs well only on small datasets. In large datasets, the model memorizes the pattern of images fluctuations and noise making the model unusable.

Future studies can build upon these findings to develop more effective and robust models, which need to overlap the challenged faced and highlighted at the end of the project.

REFERENCES

[1] G. M. M. Alshmrani, Q. Ni, R. Jiang, H. Pervaiz and N. M. Elshennawy, "A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images," Alexandria Engineering Journal, vol. 64, pp. 923-935, 2022.

[2] Le L., Xiaosong W., Gustavo C., Lin Y. Deep Learning and Convolutional Neural Networks for Medical Imaging and Clinical Informatics (Advances in Computer Vision and Pattern Recognition), Springer; 1st ed. 2019 edition (October 1, 2019), ISBN : 978-3030139681.

[3] Stanford University. Cs231n: Convolutional neural Networks for visual recognition. http://cs231n.stanford.edu/, 2019.

[4] Bezdan, T.; Džakula, N.B. Convolutional Neural Network Layers and Architectures. In Proceedings of the SINTEZA 2019International Scientific Conference on Information Technology and Data Related Research, Novi Sad, Serbia, 20 April, 2019;pp. 445–451.

[5] T. Sanida, I.-M. Tabakis, M. V. Sanida, A. Sideris and M. Dasygenis, "A Robust Hybrid Deep Convolutional Neural Network for COVID-19 Disease Identification from Chest X-ray Images," in Information 14, no. 6: 310, 2023.

[6] F. M. J. M. Shamrat, S. Azam, A. Karim, R. Islam, Z. Tasnim, P. Ghosh and F. De Boer, "LungNet22: A Fine-Tuned Model for Multiclass Classification and Prediction of Lung Disease Using X-ray Images," in Journal of Personalized Medicine 12, no. 5: 680, 2022.

[7] F. J. M. Shamrat, S. Azam, A. Karim, K. Ahmed, F. M. Bui and F. De Boer, "High-precision multiclass classification of lung disease through customized MobileNetV2 from chest X-ray images," in Computers in Biology and Medicine,Volume 155,106646,ISSN 0010-4825, 2023.

[8] M. Hong, B. Rim, H. Lee, H. Jang, J. Oh and S. Choi, "Multi-Class Classification of Lung Diseases Using CNN Models," in Applied Sciences 11, no. 19: 9289, 2021.

[9] S. Wadekar and D. K. Singh, "A modified convolutional neural network framework for categorizing lung cell histopathological image based on residual network," in Healthcare Analytics, 2023.

[10] M. Y. Kamil, "A deep learning framework to detect Covid-19 disease via chest X-ray and CT scan images," in International Journal of Electrical and Computer Engineering (IJECE), 2021.

[11] Latheesh Mangeri, Gnana Prakasi O S, Neeraj Puppala and Kanmani P, "Chest Diseases Prediction from X-ray Images using CNN Models: A Study" International Journal of Advanced Computer Science and Applications(IJACSA), 12(10), 2021. http://dx.doi.org/10.14569/IJACSA.2021.0121026.

[12] Oussama El Gannour, Soufiane Hamida, Shawki Saleh, Yasser Lamalem, Bouchaib Cherradi and Abdelhadi Raihani, "COVID-19 Detection on X-Ray Images using a Combining Mechanism of Pre-trained CNNs" International Journal of Advanced Computer Science and Applications(IJACSA), 13(6), 2022.http://dx.doi.org/10.14569/IJACSA.2022.0130668.

[13] Nurlan Dzhaynakbaev, Nurgul Kurmanbekkyzy, Aigul Baimakhanova and Iyungul Mussatayeva, "2D-CNN Architecture for Accurate Classification of COVID-19 Related Pneumonia on X-Ray Images" International Journal of Advanced Computer Science and Applications(IJACSA), 15(1), 2024. http://dx.doi.org/10.14569/IJACSA.2024.0150191.

[14] Mohammed Sidheeque, P. Sumathy and Abdul Gafur. M, "Deep Learning and Classification Algorithms for COVID-19 Detection" International Journal of Advanced Computer Science and Applications(IJACSA), 13(9), 2022. http://dx.doi.org/10.14569/IJACSA.2022.0130940.

[15] R. Das, D. S. K. Nayak, C. P. Rout, L. Jena and T. Swarnkar, "Deep Learning Techniques for Identification of Pneumonia: A CNN Approach," 2024 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC), Bhubaneswar, India, 2024, pp. 1-5, doi: 10.1109/ASSIC60049.2024.10507933.