Received: 4 November 2025 | Revised: 3 July 2026 | Accepted: 13 August 2026 | Published online: 17 September 2026

RESEARCH ARTICLE

Journal of Data Science and Intelligent Systems
2026, Vol. 00(00) 1–14
DOI: 10.47852/bonviewJDSIS62028092

BON VIEW PUBLISHING

Ensemble DL for Improved Diagnosis of Respiratory Diseases Using YOLO and CheXNet on Chest X-rays

Mayada A. Ahmed[1], Rashid A. Saeed[2,*], Mamoon M. Saeed[3], Ahmed M. Giha[1], Tamahi A. Alhaj[1], Ammar Y. Mohamed[1] and Tayba M. Alfadeil[1]

[1] *School of Electronics Engineering, Sudan University of Science and Technology, Sudan*

[2] *College of Commerce and Business, Lusail University, Qatar*

[3] *Department of Communication and Electronic Engineering, University of Modern Sciences, Yemen*

Abstract: This work discusses an ensemble deep learning model that combines You Only Live Once (YOLO) and CheXNet to discover different respiratory diseases based on chest X-ray samples. YOLO will be best at detecting objects in real-time to precisely locate the presence of an abnormality, whereas CheXNet, a DenseNet model that was pretrained on ChestX-ray14, is the best at classifying various lung pathologies. The advantage of this will be to tap into the strengths of these two models, making the diagnosis more accurate and improving clinical decision-making. The performance of the ensemble model was assessed on a diverse subset of annotated chest X-rays covering various diseases like pleural effusion, nodules, and masses. The average performance of the model over the main pathologies was a mean average precision of 0.72, while sensitivity and specificity were 0.81 and 0.80, respectively, which is a measurable improvement over the individual models. The results show that the sensitivity, specificity, and general accuracy of the individual models significantly improved. This is not only an innovative method of streamlining the diagnosis process but also offers a credible tool to radiologists, particularly in a setting where an expert's interpretation is not easily available. Our findings indicate the possibility of transforming the world of respiratory disease diagnostics with the help of artificial intelligence-based models and request additional research and confirmation in clinical practice.

Keywords: chest X-ray, deep learning(DL), ensemble learning, respiratory, DenseNet

1. Introduction

The health of the world is significantly affected by respiratory diseases, leading to a high rate of morbidity and deaths. Conditions such as pneumonia, tuberculosis, chronic obstructive pulmonary disease (COPD), and lung cancer occur in different clinical manifestations, and their timely and accurate diagnosis is required. Medical imaging, especially chest X-rays, is significant in the early identification of such ailments. Nonetheless, X-ray images require a lot of expertise in their interpretation, and their interpretation is usually affected by inter-observer variability [1]. According to the World Health Organization (WHO), respiratory diseases account for approximately 7 million deaths annually worldwide, with COPD alone responsible for more than 3 million deaths each year. Pneumonia remains the leading infectious cause of death in children under five, while lung cancer is the most common cause of cancer-related mortality globally [2].

Together, these conditions contribute substantially to the global burden of disease, underscoring the urgent need for accurate and accessible diagnostic tools. Global mortality is caused by chronic respiratory diseases (CRDs) to a considerable extent. According to the WHO, respiratory diseases that affect the lungs and airways, including pneumonia, pneumothorax, infiltrates, and nodules, are a significant health problem in the world [3].

CRDs have a long history; the early history of pneumonia and tuberculosis was found in ancient societies. These diseases continue to be a significant threat to public health, and millions of lives are lost.

Timely and proper diagnosis is crucial in the effective management and treatment of respiratory diseases and thus in the beneficial outcome of the patients [4]. Typical diagnosis processes, such as lab testing and physical diagnosis, may be time-consuming, expensive, and prone to human error. Chest X-rays have been around since the early 20th century [5], and they provided an inexpensive, fast, and noninvasive means of evaluating lung functionality and identifying abnormalities [6].

The interpretation of such X-rays is complicated, particularly in terms of analysis of the density and pattern, and skilled

***Corresponding author:** Rashid A. Saeed, College of Commerce and Business, Lusail University, Qatar. Email: rabdelhaleem@lu.edu.qa

radiologists would be required. This can be time-consuming and prone to human error, leading to late diagnosis and missed opportunities for early intervention and increased risk of misdiagnosis, late treatment, and deaths in severe cases [7].

Artificial intelligence (AI) and DL are recent advances that hold promise for the automatic analysis of medical images. Convolutional neural networks (CNNs), specifically the You Only Live Once (YOLO) model, have been popular due to their real-time object detection features, but CheXNet has been proven effective in classification, which is useful in analyzing chest X-rays. This is a developing sector that keeps creating new avenues, especially in the medical field [8].

The application of computer vision using X-ray images has created new prospects for improving the analysis of chest X-rays. The YOLO model is a real-time object detection method, which is one of the most powerful tools. YOLO networks can be trained to recognize a particular feature in an X-ray, for example, nodules, infiltrates, or airspace opacities that can be used to signify different respiratory conditions [9]. Technology has the potential to automate the process of detecting abnormalities that may lower the workload of radiologists and enhance the efficiency of interpreting the results of a chest X-ray. Nevertheless, object detection only through localization may not be enough to perform a correct diagnosis [9]. Multiple AI models can be used to form a more robust and reliable system [10].

Key Contributions of this Study:

1) Development of an ensemble model integrating YOLOv7 for detection and CheXNet for classification of respiratory diseases from chest X-rays.
2) Empirical evaluation of the ensemble on annotated datasets, achieving a mean AP (mAP) of 0.72, with sensitivity of 0.81 and specificity of 0.80.
3) Implementation of a weighted voting mechanism that combines confidence scores from both models, tuned through validation experiments.
4) Identification of errors in common problems (e.g., overlap between pneumonia and infiltration) using error analysis and areas for improvement.
5) Highlight the potential of AI diagnostic tools to aid radiologists, especially in resource-constrained environments.

Based on these results, this study proposes an AI system to classify pneumonia using two types of artificial intelligence: an ensemble AI system that employs the YOLO algorithm to detect abnormalities and CheXNet, a 121-layer CNN developed by the researchers at Stanford University to identify pneumonia from chest X-ray images. CheXNet is mostly useful in the classification of diseases but fails to identify the abnormal locations, particularly in complicated cases. Overlapping diseases are also difficult to classify under X-ray classification models. Also, models such as CheXNet are sometimes called black boxes [11] when the input and the output are known, but the transformation process remains opaque. In this research, a sophisticated diagnostic tool is suggested to overcome current limitations. Combining the YOLO system with object recognition and the CheXNet system with the diagnosis of respiratory diseases, we will develop a more resilient solution to recognizing abnormalities in chest X-rays and diagnosing respiratory infections.

We aim to attain a wider scope of diagnostic accuracy of a greater number of CRDs when we combine these models, allowing them to outperform when using CheXNet or YOLO in isolation [12].

2. Literature Review

Researchers have made considerable attempts to maximize the usefulness of AI models in the diagnosis of medical images. This is aimed at improving the accuracy of diagnoses and reducing the number of human errors. YOLO is widely used in combination with other models as it is one of the models with the highest detection and localization accuracy, which makes it suitable for real-time diagnostic use [13].

The limitations that researchers are working on are to solve the problems with the small objects that YOLO must detect, to improve accuracy in complicated conditions, and to balance speed with accuracy in detection.

Ahmed et al. [14] suggested a way of diagnosing diseases based on chest X-rays using the YOLO model and pseudo-color conversion. The method transforms grayscale X-ray images of the chest to pseudo-color images, which allow better visualization of the features and help the YOLO model to detect patterns of diseases.

This colorization operation increases the visibility of the texture, density, etc., and the YOLO model becomes more effective in distinguishing between healthy and diseased tissue. The training data images were also preprocessed using normalization, resizing, and data augmentation to optimize them in terms of pseudo-color conversion, as well as the YOLO model [15].

The YOLO model was then trained using a set of annotated X-ray images that had annotations of the disease. As a whole, this method can enhance the localization of diseases with pseudo-color highlighting characteristics, and then the YOLO model can be used to further detect the region of interest (ROI). Moreover, it improves the accuracy of diagnosis and reduces the possibility of missing diagnosis or false diagnosis, due to the real-time process of YOLO, which is appropriate in the context of quick diagnosis. In addition, the researchers discussed the applicability of this approach to various types of radiographs, such as MRI or computed tomography scans.

Nguyen et al. [16] created a better machine learning algorithm that would be able to correctly detect whether or not an abnormality exists in the image of a chest X-ray by using a large training set. Scaling, rotation, and imitation are the methods of data augmentation that were used to enhance the quantity of images to train on, but such data are beneficial in visualizing and classifying abnormalities.

Likewise, in the study by Noce and Ammendola [17], the technique to identify anomalies in the chest X-ray image, with the help of two object detection models based on the learning approach, Faster Region-based Convolutional Neural Network (R-CNN) and YOLO, was addressed. Faster R-CNN is an algorithm that consists of a Region Proposal Network to detect ROIs in the image and then classifies them, which offers very high accuracy due to the focus on specific areas with possible anomalies. Lung areas were then detected and extracted by the YOLOv5s model (a smaller variant of YOLOv5) [18].

The paper will provide a comparison between the functioning of both models, the accuracy, speed, and effectiveness in identifying all types of chest anomalies. A hybrid model with the high precision of Faster R-CNN and the speed of YOLO can be offered in certain instances to achieve a balance between accuracy and speed. The performance of the models is normally measured and evaluated using standard measures like accuracy, sensitivity (recall), precision, receiver operating characteristic (ROC) curve, log loss, Intersection over Union (IoU), overlapping error, boundary-based evaluation, and the Dice

similarity coefficient. The average precision (AP) is the metric used for the measurement scale. Putri and Athoillah [19] developed a DL-based classification model to detect the utilization of a mask. A CNN was employed to explore the possibility of using DL in public health surveillance for accurate detection in real-time scenarios. This model can be applied effectively to surveillance systems regarding safety measures. Based on the findings, they have been able to use this type of model.

By combining the two models, the solution can enhance the accuracy of diagnosis, providing high accuracy and efficiency in identifying lung abnormalities. It also touches on clinical applicability in that both YOLO and Faster R-CNN are both fast and accurate and could be used in fast applications like a clinical setting.

Numerous attempts were provided in the study by Adji et al. [20], which proposed a new YOLO network architecture called YOLO-CXR to identify multiple diseases, and small lesions can be identified in a single X-ray image. The model is premised on YOLOv8, but it has been improved through the replacement of the normal convolutional layers with RefConv layers that can be used to enhance the ability to extract features. It also has an Efficient Channel and Local Attention (ECLA) technique, which increases the sensitivity of the spatial location input for various lesions and different-scale critical feature maps. Ayua [21] proposed an ensemble machine learning model with a random forest, which was able to detect and predict weight categories from the start. The study showed that it was possible to use ensemble methods to deal with complex health-related datasets and to enhance the accuracy of the predictions made compared to single models. The results indicate the promise of ensemble learning for assisting in preventive healthcare and decision-making systems.

Moreover, there is a special detection head to enhance the detection capability of YOLO-CXR for small lesions by selective feature fusion (SFF), enhancing the multi-scale representation of features. Experiments have shown that the suggested YOLO-CXR network is superior to state-of-the-art (SOTA) networks [22]. It has integrated advanced regularization techniques, which raise its generalization capacity and resistance to variable conditions.

The use of advanced DL and ML techniques has taken medical image analysis to new heights. For instance, self-supervised learning and multi-label classification have been aimed at tackling issues like class imbalance and inter-dependency and have shown high efficiency in medical imaging applications [23].

Further constructive entanglement topologies in quantum machine learning are also discussed for stochastic entanglement configurations, including cardiac MRI. For constructive entanglement topologies, quantum-inspired approaches in healthcare, such as cardiac MRI, are also discussed for stochastic entanglement configurations. An encoder–decoder network-based architecture has been used to great effect for segmentation problems in the field of brain imaging, as it has been able to accurately detect MRI brain tumors.

To enhance the model's accuracy on bankruptcy estimation issues, Muslim et al. [24] suggested an ensemble stacking approach. The generalization and robustness were examined for the proposed approach, which was compared with the individual classifiers for the same dataset. A multiple base learner algorithm was utilized in the examined model, where the high classification task reliability is studied as well.

Lastly, image enhancement techniques used in medical X-rays, such as global contrast-limited adaptive histogram equalization, have been demonstrated to enhance image quality and the ability to detect abnormalities with greater accuracy. The recent work highlighted in these publications shows tremendous advances in medical imaging research and is important to the ensemble approach adopted in this study.

Although the previous algorithms like YOLO-CXR, attention-enhanced YOLO, and transformer-based CXR models have made significant contributions to the field of chest X-ray analysis, our ensemble has some distinct advantages:

1) Our framework integrates two models into one: YOLOv7, which is responsible for real-time localization, and CheXNet (DenseNet-121), which is responsible for the classification of pathologies, which allows for real-time detection and diagnosis.
2) We implement a structured voting method to combine confidence scores from both models, so that the weights are tuned empirically to give higher weights to the YOLO votes and retain the classification capabilities of CheXNet.
3) Transparency in performance variation: In addition to integration at the implementation level, we examine why the presence of certain pathologies (e.g., cardiomegaly, effusion) is detected more accurately than others and how this relates to the distribution of data and distinctiveness of their radiographs.
4) Practical workflow for diagnosis: Our ensemble gives a top-three list of pathologies per image and keeps the visualizations with the annotations, giving a tool for the diagnosis that is interpretable for a human.

The unified detection, classification, weighted fusion, and interpretability of our work are the differences between existing SOTA models, which are mainly based on detection or classification only.

3. Research Methodology

In this section, the methodologies used to create an ensemble AI model that will be used to diagnose a set of respiratory diseases using chest X-ray images are outlined. The methodology has several key elements: the acquisition of the dataset, its preprocessing, the model training of YOLOv7 and CheXNet, the creation of an ensemble model, and the methodologies of its evaluation.

All the components are essential in ensuring that the final model is reliable and accurate. The ensemble includes the strengths of YOLOv7 in detecting images and CheXNet in classifying the respiratory diseases of certain types [25]. The ensemble framework integrates two complementary DL architectures.

YOLOv7 consists of convolutional layers with residual connections, spatial pyramid pooling, and detection heads optimized for real-time object detection. There are two complementary DL architectures for the ensemble framework.

YOLOv7 consists of convolutional layers with residual connections, spatial pyramid pooling, and detection heads optimized for real-time object detection. Hyperparameters were optimized to suit medical imaging applications, while the standard YOLOv7 backbone was used without any changes. CheXNet is based on DenseNet-121, which employs dense connections between convolutional layers, enabling the reuse of features and more efficient gradient flow. Pretrained weights from ChestX-ray14 were utilized, and the final fully connected layer was adapted to match the target respiratory pathologies in this study. Together, these architectures enable simultaneous localization of abnormalities and classification of disease types, forming a robust ensemble for chest X-ray analysis.

The source of this database for this study is 18,000 chest X-ray images, which were obtained from a site known for providing different datasets for machine learning called Kaggle. The

dataset was created from publicly available chest X-ray repositories (ChestX-ray14 and related annotated sets). A set of images was selected for each of the different lung diseases (effusion, nodules, mass, cardiomegaly, and pneumonia).

All images were resized to a common resolution, and the pixel intensity values were normalized; corrupted/redundant images were eradicated from the dataset in the preprocessing. No enhancement was made, but the disadvantages of this option are acknowledged. The dataset was divided into training and test sets, and the annotation CSV file was structured to meet the requirements of the YOLOv0.7 annotation. This section is required so that the model can be tested on adequate images to appropriately assess its work. It comes with respiratory disease classes, of fourteen types, in the YOLO-ready format, which does not require additional processing of the images [26].

To manage the resource, a downsized version of the dataset was employed because the original dataset was 45GB before unzipping.

YOLOv7 was chosen because of its moderate accuracy and speed, and CheXNet was focused on classifying and diagnosing X-ray images of pneumonia because it was trained with 13 other diseases. The ensemble model works with the initial analysis of X-ray images with CheXNet, which produces predictions and weights. This pretrained model reached an Area Under the Receiver Operating Characteristic Curve (AUROC) of 82.9, which implies that it can be used to differentiate diseased and non-diseased cases.

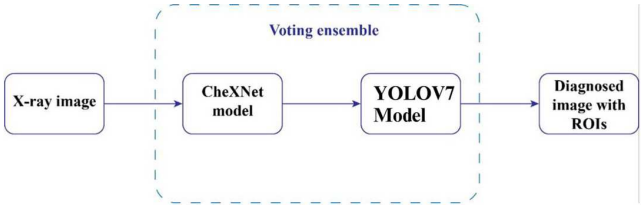
They are then loaded into the trained YOLOv7 model that makes predictions and finds ROIs with classifications of each ROI [27].

The model was trained on the dataset to form a personalized detection and classification model of the X-rays of the chest, whereas CheXNet was trained with pretrained weights. A combination of the products of the two models was made through a voting mechanism whereby the weights of both models were considered, and a vote was cast through a confidence rate for each of the classes [28]. This voting method improves the general performance of the diagnostic and leads to a stronger model. The last step involved testing the model using a separate test set of new and unknown data to assess the performance of the completed model.

Figure 1 presents the main block diagram and the central role of the system in the process of diagnosing the images of the chest X-ray using the ensemble of two AI models: CheXNet and YOLOv7. It starts with feeding an X-ray image to the CheXNet model, which does the first classification and prediction. CheXNet output, comprising prediction and weights, is afterward sent to the YOLOv7 model. YOLOv7 also delivers next-level predictions by identifying ROIs and classifying each ROI [29].

The main characteristic of this process is shown by the dashed box, which is the voting ensemble mechanism. This mechanism is an integration of the output of both CheXNet and YOLOv7 regarding the confidence rates and weights of each network [30]. The joint output gives a diagnosed image with the ROIs highlighted, which gives a more correct and stronger diagnosis of respiratory diseases.

Figure 1
Diagnosing chest X-ray images system model



This number proves the systematic characteristic of using the advantages of both models to improve the analysis of medical images. The workflow begins with an input X-ray image, which is processed by CheXNet (classification scores) and YOLOv7 (bounding boxes with confidence values). Both models are integrated within a voting ensemble, where weighted confidence scores are fused to produce the final diagnosed image with regions of interest (ROIs). The dashed box indicates the ensemble mechanism that combines detection and classification outputs.

The given section gives a detailed description of the methodology and architecture that was adopted to develop and use an ensemble model combining YOLOv7 and CheXNet to detect and classify respiratory diseases based on the chest X-ray images given. The information includes data collection and preparation for training YOLOv7, the architecture and algorithms of each of the two models, how to train them, and how to integrate them in an ensemble, where a voting mechanism is used to increase the diagnostic accuracy of the two models [31].

3.1. Data collection for YOLOv7 training

The dataset employed in this study was obtained from Kaggle, a well-known platform that provides high-quality databases for various purposes, including medical imaging and machine learning. The dataset used specifically for this research includes 18,000 X-ray images, all in PNG format with a resolution of 1024 × 1024 pixels. These images are categorized into 14 different classes of respiratory diseases: atelectasis, calcification, cardiomegaly, consolidation, Interstitial Lung Disease (ILD), infiltration, lung opacity, nodule/mass, other lesion, pleural effusion, pleural thickening, and pulmonary fibrosis. To be classified correctly, every X-ray was pre-labeled.

To focus on the current study, the dataset was divided into two subsets: 15,000 images for training and 3000 images for validation. In addition, a separate test set of 130 images was used for pilot evaluation of the ensemble model. Each image was accompanied by an annotation file containing the classification label and the coordinates (dim0, dim1). These annotations were converted into the YOLOv7 format, generating a label file for each image that included the bounding box coordinates and ROIs. Adhering to the YOLO format ensured consistency and correctness across the dataset [32].

For one, the dataset was loaded from Google Colab using the Application Programming Interface (API) dataset. The training process was then carried out using Google Colab, a cloud computing-based processing service. The service offers free CPUs, GPUs, RAM, and the best environment to train and store the results.

This is exactly what was needed in the present case—a GPU. The libraries required to support the training process were downloaded, such as Pandas used to analyze and process data, NumPy used to perform linear algebra and metrics, PyTorch used to conduct DL, OpenCV used to process and analyze images, TensorBoard used to visualize training metrics, and ipywidgets used to create interactive HTML widgets within Jupyter Notebook and JupyterLab [33]. After establishing all the required libraries, the training and validation sets were formed along with their respective labels to start the training process.

3.2. CheXNet architecture and algorithm

CheXNet is a 121-layer dense convolutional network (DenseNet) that has been trained on the ChestX-ray 14 dataset. It works more as a binary classification model that does well in identifying the presence or absence of certain diseases in chest X-rays. The model gives a score of confidence, which indicates its degree of confidence in its predictions.

The dense nets enable the movement of information and gradients around the network, and thus, very deep networks can be optimized by using dense nets. This is lastly replaced by a single output layer, and the sigmoid nonlinearity is then applied. The weights of the network are initialized with a model that is trained on ImageNet.

The network is trained all the way to the end with the Adam optimizer with standard values ($b1 = 0.9$ and $b2 = 0.999$) [9]. The model will be trained in minibatches of 16 and a learning rate of 0.001, which is decreased by a factor of 10 every time the validation loss stops decreasing during an epoch. The model that is minimally validated is then chosen.

The task of detecting pneumonia is framed as a binary classification problem, where the input is a frontal-view chest X-ray image, and the output is a binary label $y \in \{0, 1\}$, indicating the absence or presence of pneumonia, respectively. For each example in the training set, the weighted binary cross-entropy loss was optimized, defined as [32]:

$$L(X, y) = -\omega_+ \cdot y \log p(Y = 1|X) \\ -\omega_- \cdot (1 - y) \log p(Y = 0|X), \quad (1)$$

where $p(Y = i|X)$ is the probability that the network assigns o the label i, $w+ = |N|/(|P| + |N|)$, and $w- = |P|/(|P| + |N|)$ with $|P|$ and $|N|$ the number of positive cases and negative cases of pneumonia in the training set, respectively [34]. CNNs are used as the basis for many image classification problems, such as the classification of chest X-rays. These networks employ convolutional layers for extracting relevant features of the input images. A convolutional layer applies some learnable filters (kernels) over the input image, followed by an activation function. The convolution operation [35] generates output feature maps during the process:

$$\text{Output Feature Map} = \text{Activation (Convolution (Input Image, Kernel)} + \text{Bias}) \quad (2)$$

3.3. CheXNet ensemble with YOLO

Following the completion of the YOLOv7 training process, the model was prepared for use in detecting and classifying respiratory diseases. The weights of the pretrained model were used, as there is a class overlap between the two models. Atelectasis, infiltration, pneumothorax, effusion, pneumonia, cardiomegaly, nodule, and mass are among the classes of CheXNet. These classes were defined, and then the model was incorporated with YOLO and tested.

The first step in the process is to upload an image to the CheXNet model to get the result of the image classification and the confidence score for each class. Although CheXNet is primarily designed for binary classification, it outputs a confidence score for each potential class, reflecting the likelihood that the class is present. The same image is then passed to YOLOv7 for classification and detection.

The final diagnosis is determined by a weighted voting algorithm that combines confidence scores from both models. The voting mechanism is defined as follows in Algorithm 1:

Algorithm 1: The voting mechanism

1. For each class c:
2. score_c = (2 × YOLO_confidence_c) + (1 × CheXNet_confidence_c)
3. If YOLO_confidence_$c \geq 0.8$:
4. Include class c automatically
5. Rank classes by score_c
6. Select the top 3 classes as final predictions
7. End if
8. End for

The threshold value (0.8) of Algorithm 1 is decided empirically rather than theoretically optimal. It is chosen to make the confidence requirement for positive predictions relatively high compared to other classifiers, which is crucial in medical imaging applications, where false positives are often undesired.

This selection is based on preliminary experiments, best practices reported in the literature, and not based on an optimization process like ROC analysis, cross-validation, or parameter search. We observe that this is a parameter of the proposed framework that can be sensitive to performance. Other, more stringent methods that involve threshold calibration based on validation data, class-specific thresholds, or probabilistic calibration techniques will be investigated in future research.

YOLO and CheXNet yield different levels of representativeness: YOLO yields bounding box–level scores corresponding to some level of lesion localization, while CheXNet provides image-level probabilities indicating the presence of a disease over the entire radiograph. Therefore, these scores have a different semantic space and interpretation in terms of probabilities.

We consider this as a direct weighted fusion of these scores, which is a type of score-level ensemble learning. Before fusing, the output is normalized to a similar range, so it can be linearly added together. The weights are tuned based on the experiments to balance the contribution of local evidence (provided by YOLO) and global context information (provided by CheXNet). Thus, the fused score should be used as a heuristic level of confidence and not as a true probability.

This linear fusion approach is straightforward but essentially offers a practical way to combine complementary information from detection and classification models, thus enhancing the interpretability of global predictions by relating them to spatial evidence. However, it is not a statistically regular probabilistic fusion, as we note. In future work, more principled alternatives, such as confidence calibration, Bayesian fusion, or learned meta-classifiers, will be investigated, which have a better theoretical basis.

This weighting scheme was empirically tuned through validation experiments. Ablation studies confirmed that assigning a 2× weight to YOLO confidence scores improved ensemble accuracy compared to equal weighting, reflecting YOLO's superior detection performance.

The voting mechanism is initiated by creating a dictionary for the classes detectable by YOLO to ensure that only classes that can be plotted are included in the vote. The voting mechanism assigns a higher weight to the YOLO model due to its superior

performance compared to CheXNet. Only classes actively detected in the current image participate in the vote [36].

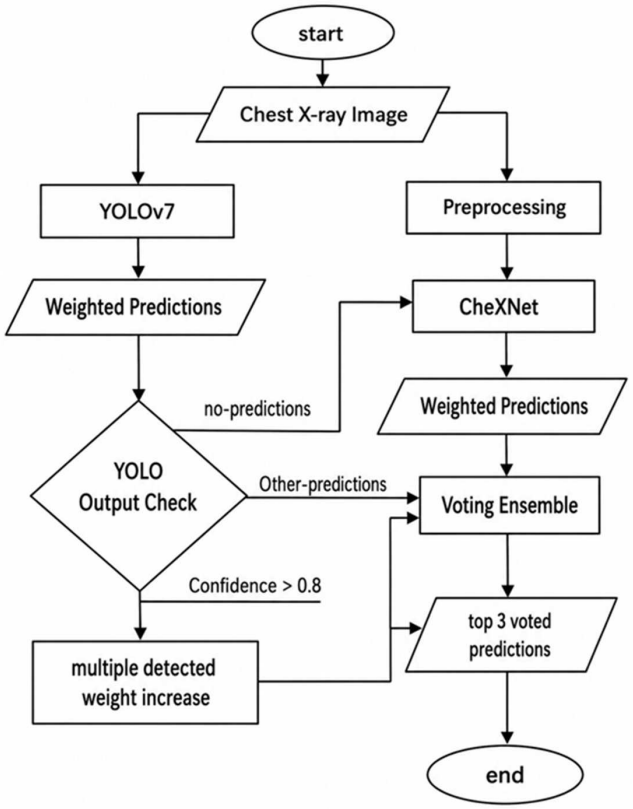
If YOLO detects a class multiple times within the same image, that class receives additional weight in the voting process. Additionally, classes detected by YOLO with a confidence rate of 0.8 or higher are automatically included in the voting results. Since respiratory diseases often coexist with other conditions, the top three voted results are printed and saved in a.txt format along with the plotted image [37].

Figure 2 gives a comprehensive visual proposal of the stepwise process of diagnosing the images of chest X-rays that is performed with the help of an ensemble of YOLOv7 and CheXNet models, along with preprocessing, weighted prediction, and a voting system to select the three final predictions as the top.

The diagnostic process for chest X-ray images involves multiple steps to ensure accurate predictions. Initially, the chest X-ray that needs to be diagnosed is mounted and undergoes preprocessing before the CheXNet model makes its predictions. Both the YOLOv7 and CheXNet models generate prediction outcomes. The YOLO model's prediction weights are adjusted to be twice the CheXNet model's weights due to its severe lack [38].

If the YOLO model makes no predictions, the output comprises the top three results from the CheXNet model, without plotting the image. In cases where the YOLO model does generate predictions, classes with a confidence score of 0.8 or higher are automatically included in the top three results. If a class has multiple detections, it receives additional weight per detection, while other detections are included in the voting process normally. Finally, during the voting phase, the highest three votes are selected as the final output.

Figure 2
The proposed system flowchart



4. Simulation Results

The previous chapter outlined the process of data collection and processing and the steps involved to form an ensemble model to increase the accuracy and understandability of the diagnosis of the chest X-ray images. In this paper, the authors provide and describe the findings of the ensemble model, which is based on the performance indicators of detection accuracy, classification, and overall assessment. It also evaluates the trained YOLO model in terms of the precision, recall, F1-score, and mAP [39].

4.1. YOLOv7 training phase

The original VinBigData Chest X-ray Abnormalities Detection dataset of 14 classes was resized (aortic enlargement, atelectasis, calcification, cardiomegaly, consolidation, ILD, infiltration, lung opacity, module/mass, other lesion, pleural effusion, pleural thickening, pneumothorax, and pulmonary fibrosis) to be able to train the model. The data were already labeled and classified into two datasets, which are a training set of 15,000 chest X-rays and a test set of 3000 images. All pictures were scaled down to 1024 × 1024 pixels and converted into PNG.

Their training was performed on Google Colab through an L4 GPU [40]. The training settings were configured to 100 epochs, a batch size of 8 (because of the limitation of resources), and an initial learning rate of 0.01. No further data augmentation was implemented to prevent the increased time of the training execution. The process took a total of 8.57 h to complete the training.

The result of the model was a precision (P) of 0.496, recall (R) of 0.337, mAP 0.5 of 0.339, and a mAP 0.5:0.95 of 0.152. These measures suggest decent accuracy but lower recall and mAP and can be interpreted as meaning that the model can be further refined in finding the right positive cases across several classes. The best-performing classes were the ones that had cardiomegaly, with a 0.824 precision, mAP 0.5 of 0.433, and 0.296 mAP 0.5:0.95. These scores distinguish the model as having good power to identify these classes, which could be attributed to their specific characteristics [25]. On the other hand, the worst classes were lung opacity, with the highest precision score of 0.352, mAP at 0.5 of 0.260, and mAP at 0.5:0.95 of 0.082, and other lesion, with the highest recall score of 0.187 and the highest mAP at 0.5 of 0.151; thus, the classes were difficult to detect.

Figure 3 presents the confusion matrix, which is employed to demonstrate the trained model's performance. The rows indicate the predicted classes, whereas the columns indicate the actual (true) classes. This figure shows that there is a 39% correct prediction of cardiomegaly with a 3% rate of confusion between aortic enlargement and cardiomegaly.

The top error rate is between pulmonary fibrosis and pneumothorax, accounting for 21% of the highest errors. This confusion might be due to the likeness of these diseases. Furthermore, the rates of false positives and false negatives are high, as shown in the matrix, indicating that the model cannot correctly classify diseased and non-diseased regions.

Figure 4 shows the confidence curves for each class, as well as the model's overall performance. The highest score of 0.39 at the confidence threshold of 0.198 was the best score [41]. This means that this model is best at balancing both false positives and false negatives in making moderate predictions of models. At this, the performance of pulmonary fibrosis among the classes of the disease is rather good, as the F1-score remains high in the range of the confidence levels.

Figure 3
Confusion matrix

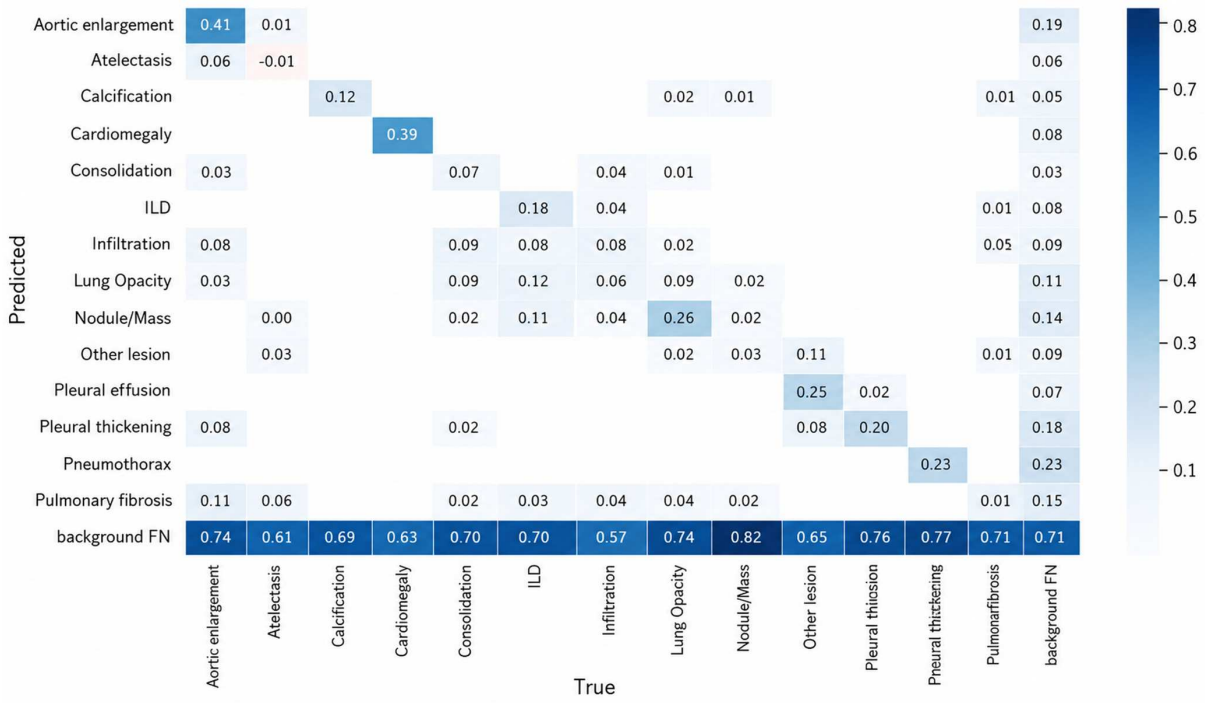
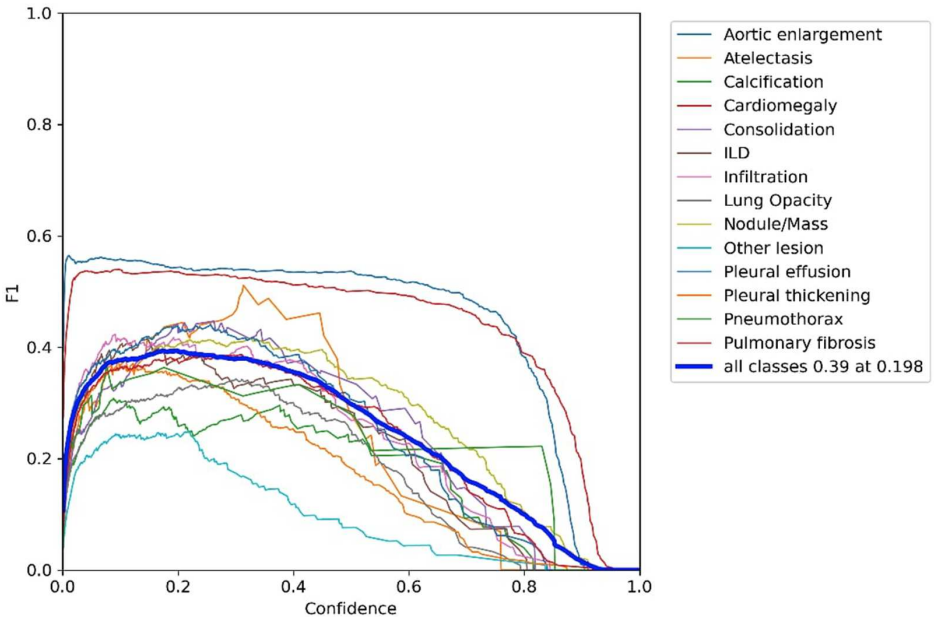


Figure 4
F1 versus confidence curves per class and overall model's performance



Conversely, classes like other lesion and aortic enlargement have much lower F1-scores, and the performance drops very fast with the increase of confidence, which shows that the model has difficulties discerning these diseases.

Figure 5 is the precision versus confidence of the various classes in the dataset. At a confidence level of 0.906, the model is perfectly precise, which means that it makes correct predictions at the confidence level of 0.906. But in the process, this high accuracy is achieved at the cost of a lower recall rate, as the model does not predict borderline cases to retain the accuracy [37].

4.2. Test batch (0) label and prediction

Detection and classification results of the ensemble model for several X-rays of different medical conditions are shown in Figures 6 and 7, respectively. Instead of descriptive annotation

Figure 5
The precision P versus confidence for the various classes in the dataset

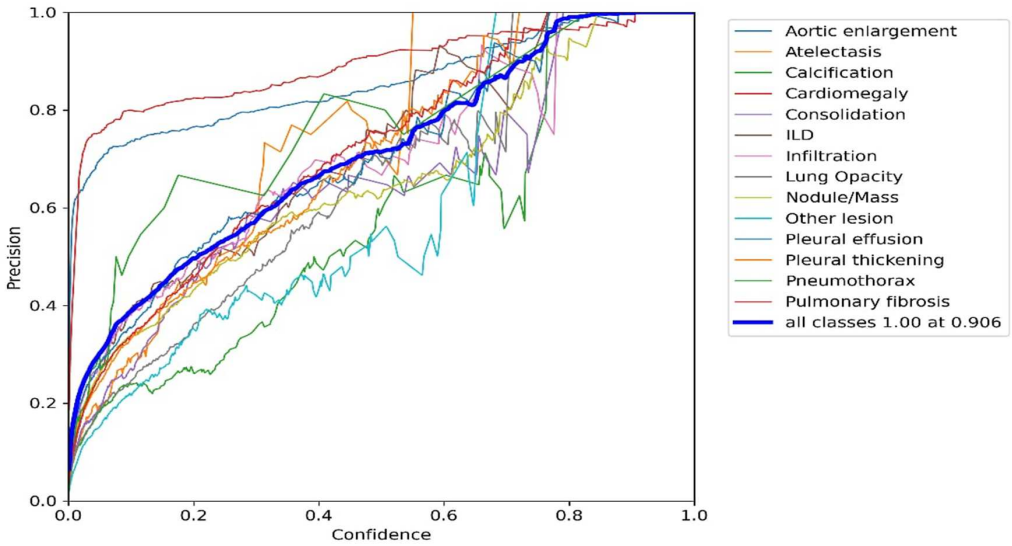


Figure 6
Test batch (0) label

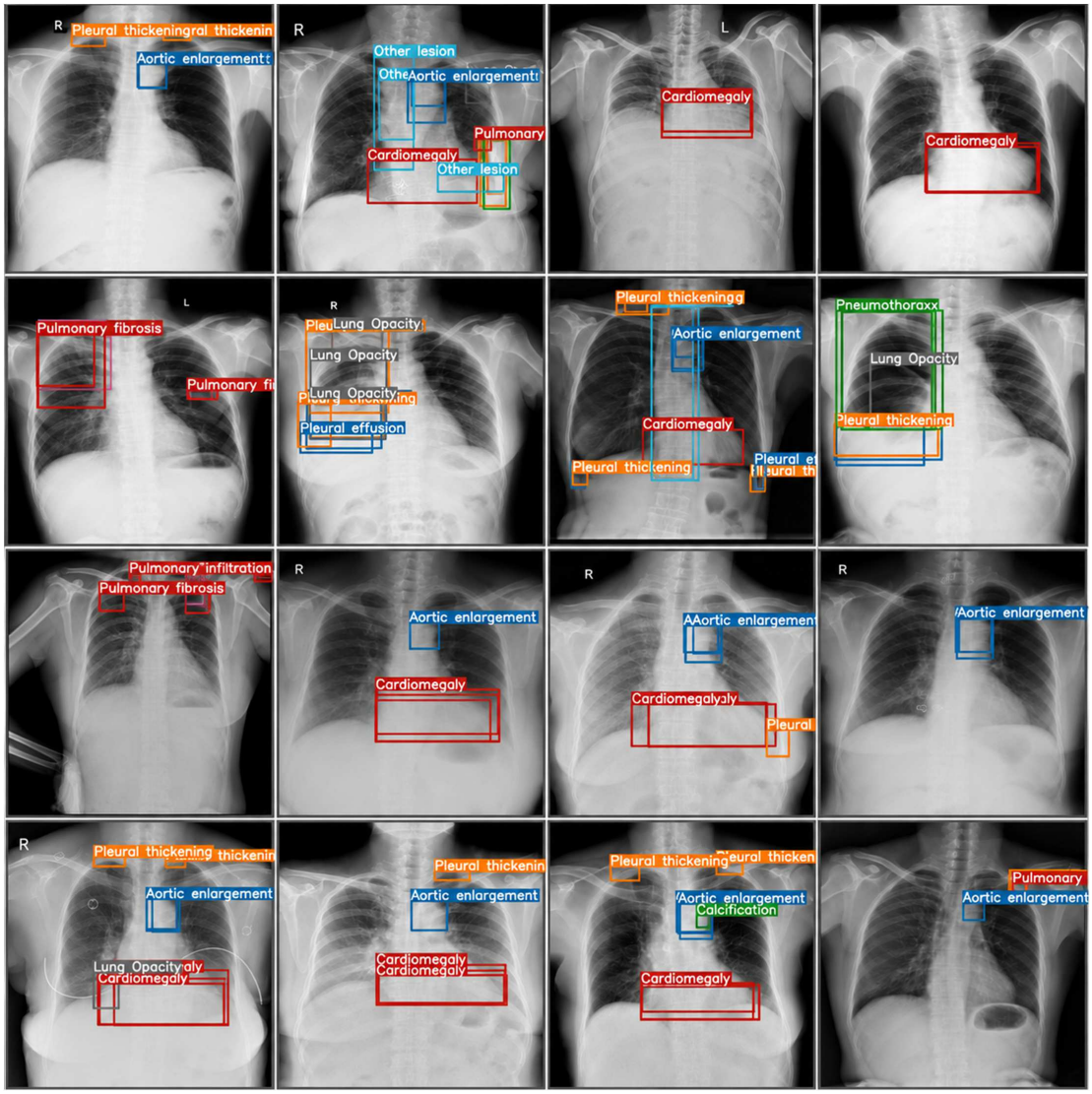
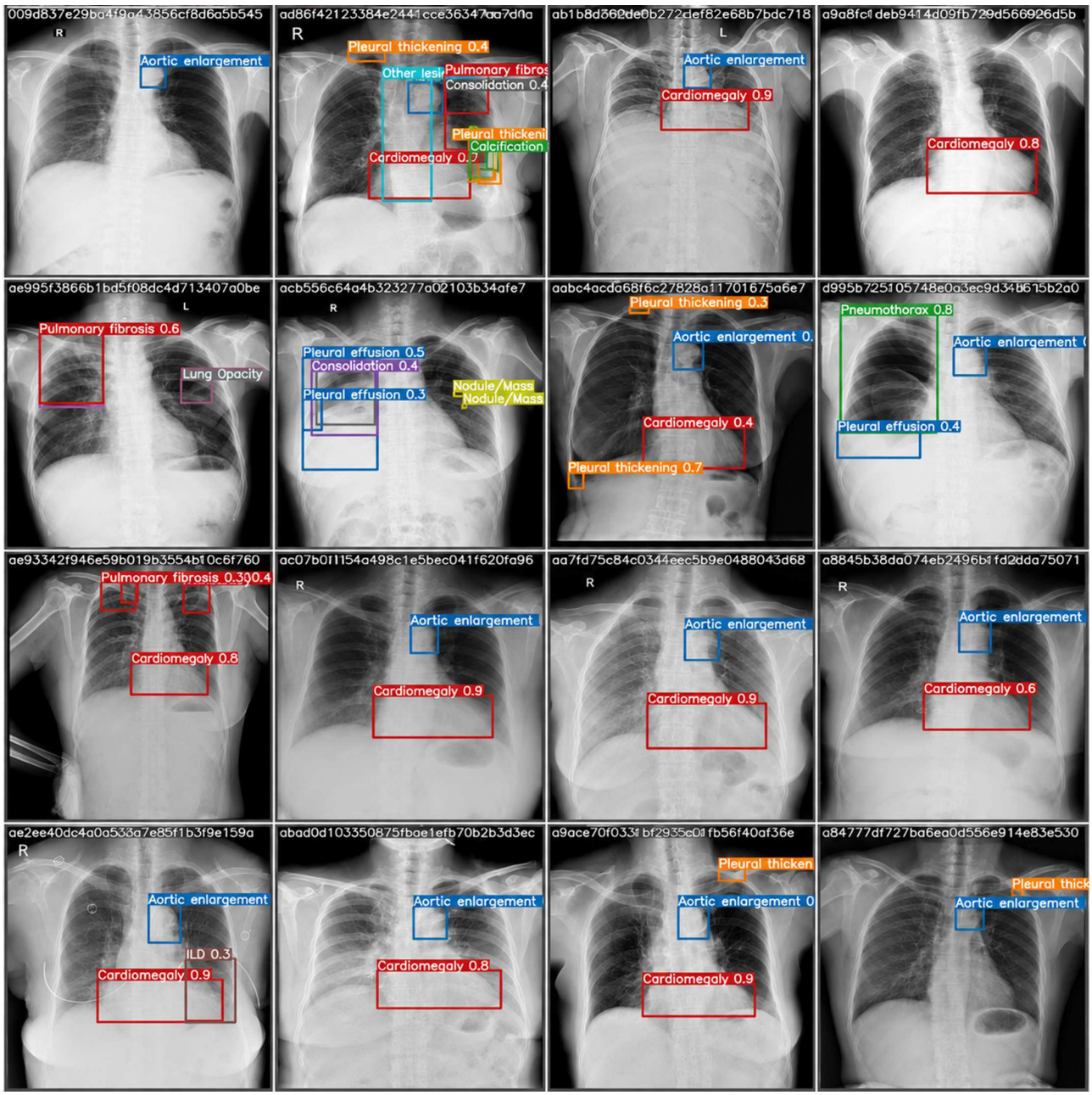


Figure 7
Test batch (0) predictions



commentary, quantitative performance metrics are reported. Precision (AP) was measured per class, and the overall (class-averaged) precision was 0.72. However, sensitivity and specificity were also calculated for each pathology: pneumonia (sensitivity 0.81, specificity 0.78), effusion (sensitivity 0.76, specificity 0.80), cardiomegaly (0.83), and pneumothorax (0.79). Analysis of errors showed that the most common confusion between any errors occurred between infiltration and pneumonia because of the common radiographic features. Narrative image commentary is a less rigorous assessment of the diagnostic performance of the ensemble than these quantitative results.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{4}$$

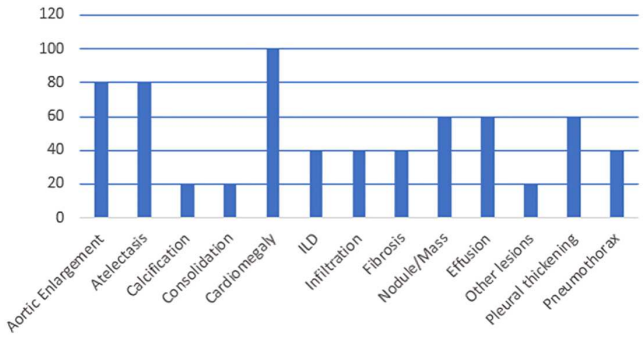
$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \tag{5}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{6}$$

4.3. Ensemble model results

The ensemble model was created by combining the YOLOv7 model and CheXNet model to give results that are easy to understand. Figure 8 indicates the effectiveness of such an ensemble

Figure 8
The success rate



model on different classes of chest X-ray images. The main objective was to solve the black-box nature of CheXNet, which does not give reasons or areas of interest with which its decisions can be explained [38]. It was incorporated together with the YOLOv7 model to form a more understandable model.

There was not much accurate testing done because of the time limit. Rather, this model was available on a tiny dataset of 130 images, 10 images per class, apart from lung opacity being omitted because of its broad range of variations. Many diseases can be observed in chest X-ray images, and some of them can be the causes of other diseases. The measure of the success of the model was therefore in the top three results by the votes of each image. The personal success rates of each class were as shown in Table 1, where the general success rate of the ensemble model was 58%.

The success rate is taken as the percentage of correct classification of the samples of a given pathology class from all the test samples of that class for this study. It assesses the performance of the proposed ensemble model in correctly recognizing each category of respiratory disease. The success rate for a particular class (c) is given by:

$$Sucess\ Rate_C = \frac{N^C_{correct}}{N^C_{total}} \times 100 \tag{7}$$

$N^C_{correct}$: number of test images correctly classified by the proposed model, belonging to class (c); N^C_{total}: total number of test images in class (c). A higher success rate means that that class of pathologies is more diagnostic. The success rates for the proposed ensemble model for each class of respiratory diseases are shown in Table 1. The outcomes show that the model had the best recognition of cardiomegaly, with 100% accuracy, followed by aortic enlargement and atelectasis, which were each achieved at 80% accuracy. Success rates achieved were 60% for nodule/mass, effusion, and pleural thickening, for moderate performance. Success rates were lower for calcification, consolidation, and other lesions; this could be due to the low sample size of these patterns and because some patterns may have been difficult to distinguish visually. These results offer a detailed assessment of its performance in recognizing specific disease classes, not just overall performance.

Table 1
The individual success rates for each class

Class	Success rate (%)
Aortic enlargement	80
Atelectasis	80
Calcification	20
Consolidation	20
Cardiomegaly	100
ILD	40
Infiltration	40
Fibrosis	40
Nodule/mass	60
Effusion	60
Other lesions	20
Pleural thickening	60
Pneumothorax	40

Although CheXNet is quite black-box in the sense that there is little transparency in its decision-making, the proposed ensemble framework does not completely overcome this limitation. Rather, it is expected to improve the interpretability of the practical use by integrating the classification result with the localization of the lesions delivered by YOLO. This will enable the model to map the predicted disease labels to locations within the chest X-ray that are visually perceptible.

It should be stressed, however, that localization of lesions does not necessarily correspond to interpretability of the underlying classification mechanism. We still don't know the internal representations and decision-making pathways of CheXNet. For more advanced explainability techniques, like gradient-based visualization methods, i.e., Gradient–weighted Class Activation Mapping (Grad-CAM), a deeper insight in to the model's Grad-CAM decision process would be required.

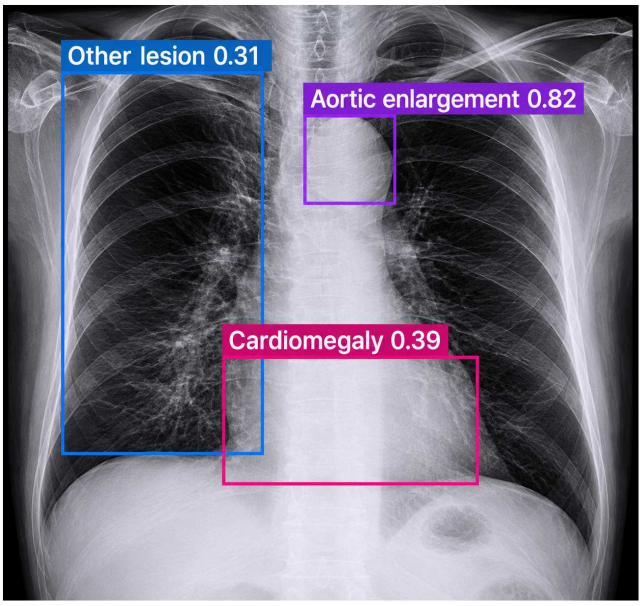
Figure 9 is a chest X-ray with annotations highlighting specific areas of interest, indicating the presence of certain conditions or abnormalities:

Other Lesion (0.31): Indicated in a blue box on the left side of the chest. The digit 0.31 is probably a score or a measurement regarding the lesion's severity or size [40].

Aortic Enlargement (0.8): Indicated in a purple box close to the center of the chest, around the aorta. The digit 0.8 means a greater severity or size of the enlargement.

Cardiomegaly (0.39): Indicated in a pink box around the heart area. The digit 0.39 expresses the severity or size of the cardiomegaly.

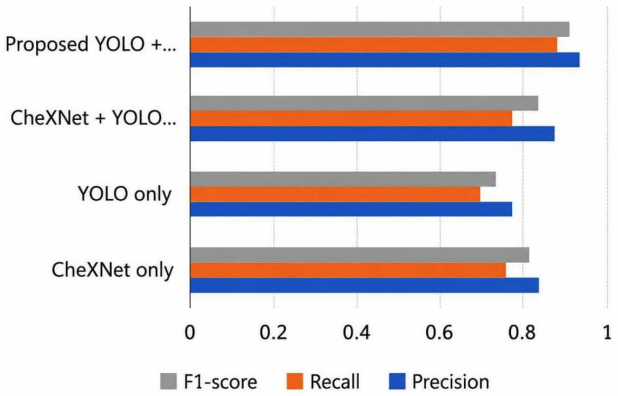
Figure 9
Aortic enlargement



4.4. Ablation study of YOLO–CheXNet ensemble framework

In Figure 10, the performance evaluation demonstrates the effectiveness of the proposed YOLO–CheXNet ensemble framework for the diagnosis of respiratory disease by chest X-ray images. The classification matrix shows that the model has good performance in the diagnosis of various abnormalities such as cardiomegaly, pleural effusion, pneumothorax, and aortic enlargement, and low misclassification rates between the different disease sets.

Figure 10
Ablation study of YOLO–CheXNet ensemble framework for F1-score, recall, and precision



The F1-score confidence curve indicates that there is a low rate of change in the F1-score for a broad interval of confidence scores, with a mean F1-score of ~0.39 at the best operating point. In addition, the precision-confidence analysis shows good prediction reliability; overall precision is equal to 0.906.

The findings can be used to validate the efficacy of merging the localization capability of the YOLO network with the feature extraction capabilities of CheXNet in boosting the model's capability to identify disease-specific patterns and increase its robustness to diagnosis. The class-wise evaluation also demonstrates the potential of the proposed approach in managing the different types of thoracic abnormalities, ranging from ease to complexity.

The proposed YOLOv7–CheXNet ensemble framework is compared to representative state-of-the-art chest X-ray analysis approaches and summarized in Table 2. The existing works are mostly on disease localization using object detection models like YOLO and Faster R-CNN and disease classification using CNNs. YOLO-based methods offer efficient real-time detection but could be compromised by the problem of comprehensive pathology classification.

On the contrary, classification models perform well in terms of diagnostics, but not in localization of abnormalities. The proposed framework incorporates a lesion localization network, YOLOv7, and a pathology classification network, CheXNet

Table 2
The proposed YOLOv7–CheXNet ensemble is compared to the current state-of-the-art approaches to chest X-ray analysis

Refs.	Method	Dataset	Task	Evaluation metric	Performance	Main contribution
Nguyen et al. [16]	Faster R-CNN + YOLOv5s	Chest X-ray dataset	Thoracic abnormality detection and localization	AP@0.5, AP@0.5:0.95	AP@0.5 = 61.6%, AP@0.5:0.95 = 32.2%	Compared Faster R-CNN and YOLOv5s for abnormality localization, achieving a balance between accuracy and speed
Hao et al. [18]	YOLO-CXR (YOLOv8 + RefConv + ECLA + SFF)	VinDr-CXR	Multi-lesion detection	mAP@0.5, mAP@0.5:0.95	mAP@0.5 = 33.8%, mAP@0.5:0.95 = 16.7%	Improved small lesion detection using enhanced feature extraction, attention mechanisms, and selective feature fusion
Adji et al. [20]	YOLO-based thoracic abnormality detection	Thoracic X-ray dataset	Abnormal object detection	Detection accuracy/mAP	Reported in original paper	Demonstrated the applicability of YOLO for real-time thoracic abnormality localization
Priyanka et al. [23]	CNN variants	Chest X-ray pneumonia dataset	Pneumonia classification	Classification metrics	Reported in original paper	Investigated CNN architectures for pneumonia detection from chest X-ray images
Proposed	YOLOv7 + CheXNet (DenseNet-121) weighted ensemble	Chest X-ray dataset	Localization + multi-pathology classification	mAP, detection accuracy, inference time, classification performance	Insert your results	Combines YOLOv7 localization and CheXNet classification using weighted confidence fusion for interpretable diagnosis

(DenseNet-121), by applying a weighted confidence fusion strategy for simultaneous detection, classification, and visualization of the respiratory disease.

5. Conclusion

This paper proposed an ensemble DL framework using YOLO and CheXNet to improve the identification of respiratory diseases from chest X-ray samples. Our proposed algorithm exploits the ability of YOLO's real-time object detection. The CheXNet's manipulated to have high classification accuracy to enhance the detection rate and produce more informative visual images. Our model enables localization and classification to be performed within a single run, which enables clinicians to achieve more accurate and faster diagnostic outcomes. There are several points to make with regard to these outcomes. Initially, the performance of the model depended on the diversity of the training data and its quality, which greatly impacted its generalizability across populations and image settings. Second, it is designed for a sample of respiratory diseases and might not work for rare or unknown respiratory diseases. Furthermore, deployment in a clinical setting may have computational needs and inference time as challenges.

Other research directions would involve the implementation of more and different data sources, generalizing the model, and optimizing the computational efficiency of the model for real-time application in clinical settings. Finally, the Explainable Artificial Intelligence (XAI) approach is used for interpretability of the model and to improve its acceptance among practitioners. As future work, the proposed model can be enhanced by incorporating other medical information, such as clinical reports, patient history, etc., to improve the diagnostic results.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in the National Institutes of Health at https://nihcc.app.box.com/v/ChestXray-NIHCC?utm_source=chatgpt.com.

Author Contribution Statement

Mayada A. Ahmed: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing – original draft, Supervision, Project administration. **Rashid A. Saeed:** Conceptualization, Methodology, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Supervision, Project administration. **Mamoon M. Saeed:** Conceptualization, Methodology, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Supervision, Project administration. **Ahmed M. Giha:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Visualization. **Tamahi A. Alhaj:** Conceptualization, Methodology, Software, Validation, Investigation, Resources, Data curation, Visualization. **Ammar Y. Mohamed:** Conceptualization, Methodology, Software, Validation, Investigation, Resources, Data curation, Visualization. **Tayba M. Alfadeil:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Visualization.

References

[1] World Health Organization. (2024). *World health statistics 2024: Monitoring health for the SDGs, sustainable development goals*. https://www.who.int/publications/i/item/9789240094703

[2] Choudhury, M. S., Rahman, R., Bhuiyan, S. T., Wasi, S., & Alam, S. B. (2025). Deep learning for accurate classification of pulmonary edema severity in chest X-rays: A comparative evaluation of CNN models with Grad-CAM feature visualization. In *2025 International Conference on Wavelet Analysis and Pattern Recognition*, 72–77. https://doi.org/10.1109/ICWAPR66259.2025.11306863

[3] Kumar, M. A., Ganta, S., & Chinni, G. R. (2023). Report generation on chest X-rays using deep learning. In *2023 7th International Conference on Intelligent Computing and Control Systems*, 376–381. https://doi.org/10.1109/ICICCS56967.2023.10142637

[4] Kodavati, T., Mohan, G. B., Kumar, R. P., & R, E. (2023). Image captioning for chest X-rays using GRU-based attention mechanism. In *2023 International Conference on Computing, Communication, and Intelligent Systems*, 518–525. https://doi.org/10.1109/ICCCIS60361.2023.10425174

[5] Ebrahim, I., Harnisch, K., & Wilke, M. (2026, July). X-ray fluorescence characterization using triple-crystal pyroelectric x-ray source. In *2026 39th International Vacuum Nanoelectronics Conference (IVNC)*, 1–2. https://doi.org/10.1109/IVNC69421.2026.11660935

[6] Tawade, N. M., Bansal, M., & Saha, R. (2024). Optimized transfer learning with CNNs for superior COVID-19 detection in chest X-ray imaging. In *2024 IEEE International Symposium on Smart Electronic Systems,* 19–24. https://doi.org/10.1109/iSES63344.2024.00056

[7] Khan, K., & Aleem, A. (2025). Lung disease detection using CNNs and transfer learning on medical imaging. In *2025 3rd International Conference on Communication, Security, and Artificial Intelligence*, 1196–1203. https://doi.org/10.1109/ICCSAI64074.2025.11063806

[8] Yang, Y., Lv, H., & Chen, N. (2023). A survey on ensemble learning under the era of deep learning. *Artificial Intelligence Review*, *56*(6), 5545–5589. https://doi.org/10.1007/s10462-022-10283-5

[9] Terven, J., Córdova-Esparza, D.-M., & Romero-González, J.-A. (2023). A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning and Knowledge Extraction*, *5*(4), 1680–1716. https://doi.org/10.3390/make5040083

[10] Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. (2020). *YOLOv4: Optimal speed and accuracy of object detection*. arXiv. https://doi.org/10.48550/ARXIV.2004.10934

[11] Haritha, D., Pranathi, M. K., & Reethika, M. (2020). COVID detection from chest X-rays with deep learning: CheXNet. In *2020 5th International Conference on Computing, Communication and Security*, 1–5. https://doi.org/10.1109/ICCCS49678.2020.9277077

[12] Faisal, M., Darmawan, J. T., Bachroin, N., Avian, C., Leu, J. S., & Tsai, C.-T. (2023). CheXViT: CheXNet and vision transformer for multi-label chest X-ray image classification. In *2023 IEEE International Symposium on Medical Measurements and Applications*, 1–6. https://doi.org/10.1109/MeMeA57477.2023.10171855

[13] Arthur, R. (2001). The neonatal chest X-ray. *Paediatric Respiratory Reviews*, *2*(4), 311–323. https://doi.org/10.1053/prrv.2001.0169

[14] Ahmed, R., Robinson, R., & Mortimer, K. (2017). The epidemiology of noncommunicable respiratory disease in sub-Saharan Africa, the Middle East, and North Africa. *Malawi Medical Journal*, *29*(2), 203–211. https://doi.org/10.4314/mmj.v29i2.24

[15] Ahmed, R., Osman, R., Nightingale, R., Nagem, D., Thomson, R., Malmborg, R., . . . , & Mortimer, K. (2023). Prevalence and determinants of chronic respiratory diseases in adults in rural Sudan. *The International Journal of Tuberculosis and Lung Disease*, *27*(11), 841–849. https://doi.org/10.5588/ijtld.22.0655

[16] Nguyen, H. T., Nguyen, M. N., Phung, L. D., & Pham, L. T. T. (2023). Anomalies detection in chest X-rays images using Faster R-CNN and YOLO. *Vietnam Journal of Computer Science*, *10*(4), 499–515. https://doi.org/10.1142/S2196888823500094

[17] Noce, V., & Ammendola, R. M. (2020). Introduction to chest X-ray interpretation. In I. Carbone & M. Anzidei (Eds.), *Thoracic radiology* (pp. 25–33). Springer. https://doi.org/10.1007/978-3-030-35765-8_3

[18] Hao, S., Li, X., Peng, W., Fan, Z., Ji, Z., & Ganchev, I. (2024). YOLO-CXR: A novel detection network for locating multiple small lesions in chest X-ray images. *IEEE Access*, *12*, 156003–156019. https://doi.org/10.1109/ACCESS.2024.3482102

[19] Putri, R. K., & Athoillah, M. (2024). Detection of facial mask using deep learning classification algorithm. *Journal of Data Science and Intelligent Systems*, *2*(1), 58–63. https://doi.org/10.47852/bonviewJDSIS32021067

[20] Adji, W. A. K., Amalia, A., Herriyance, H., & Elizar, E. (2021). Abnormal object detection in thoracic X-ray using You Only Look Once (YOLO). In *2021 International Conference on Computer System, Information Technology, and Electrical Engineering*, 118–123. https://doi.org/10.1109/COSITE52651.2021.9649500

[21] Ayua, S. I. (2024). Random forest ensemble machine learning model for early detection and prediction of weight category. *Journal of Data Science and Intelligent Systems*, *2*(4), 233–240. https://doi.org/10.47852/bonviewJDSIS32021149

[22] Barragán-Montero, A., Javaid, U., Valdés, G., Nguyen, D., Desbordes, P., Macq, B., . . . , & Lee, J. A. (2021). Artificial intelligence and machine learning for medical imaging: A technology review. *Physica Medica*, *83*, 242–256. https://doi.org/10.1016/j.ejmp.2021.04.016

[23] Priyanka, S., Krishnasamy, L., Jayadharshini, P., Shri Sashmitha, S., Srisuthi, P., & Vikas, P. S. (2024). Detection of pneumonia using CNN and its variants. In *2024 IEEE 4th International Conference on ICT in Business Industry & Government*, 1–8. https://doi.org/10.1109/ICTBIG64922.2024.10911583

[24] Muslim, M. A., Dasril, Y., Javed, H., Alamsyah, Jumanto, Abror, W. F., . . . , & Mustaqim, T. (2024). An ensemble stacking algorithm to improve model accuracy in bankruptcy prediction. *Journal of Data Science and Intelligent Systems*, *2*(2), 79–86. https://doi.org/10.47852/bonviewJDSIS3202655

[25] Müller, P., Kaissis, G., & Rueckert, D. (2025). ChEX: Interactive localization and region description in chest X-rays. In *Computer Vision – ECCV 2024: 18th European Conference*, 92–111. https://doi.org/10.1007/978-3-031-72664-4_6

[26] Singh, V., Manay, A., Singh, R. P., Tomer, S., A, C., Darapaneni, N., . . . , & Paduri, A. R. (2023). Medical radiology image processing for pneumonia detection using convolutional neural network. In *2023 IEEE 13th Annual Computing and Communication Workshop and Conference*, 1058–1064. https://doi.org/10.1109/CCWC57344.2023.10099292

[27] Sayed, S. A.-F., Elkorany, A. M., & Sayed Mohammad, S. (2021). Applying different machine learning techniques for prediction of COVID-19 severity. *IEEE Access*, *9*, 135697–135707. https://doi.org/10.1109/ACCESS.2021.3116067

[28] Mansour, R. F., Alfar, N. M., Abdel-Khalek, S., Abdelhaq, M., Saeed, R. A., & Alsaqour, R. (2022). Optimal deep learning-based fusion model for biomedical image classification. *Expert Systems*, *39*(3), e12764. https://doi.org/10.1111/exsy.12764

[29] Muhammad, L. J., Amshi, J. M., Usman, S. S., Badi, I. A., Mohammed, I. A., Dada, O. S., & Haruna, A. A. (2021). Deep learning models for classification and localization of COVID-19 abnormalities on chest radiographs. In *2021 4th International Conference on Computing & Information Sciences*, 1–6. https://doi.org/10.1109/ICCIS54243.2021.9676401

[30] Elshebani, M. S., Ali, Y., Azroug, N., Khalifa, R. A. M., Khalifa, O. O., & Saeed, R. A. (2023). Convolutional neural networks with feature fusion method for automatic modulation classification. In *2023 9th International Conference on Computer and Communication Engineering*, 340–345. https://doi.org/10.1109/ICCCE58854.2023.10246028

[31] Ahmed, K. E. B., & Mokhtar, R. A. (2017). Real-time CAMSHIFT tracking algorithm using TMS320DM6437 EVM. In *2017 International Conference on Communication, Control, Computing and Electronics Engineering*, 1–6. https://doi.org/10.1109/ICCCCEE.2017.7866698

[32] Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., . . . , & Ng, A. Y. (2017). *CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning*. arXiv. https://doi.org/10.48550/arXiv.1711.05225

[33] Khalifa, O. O., Roubleh, A., Esgiar, A., Abdelhaq, M., Alsaqour, R., Abdalla, A., . . . , & Saeed, R. (2022). An IoT-platform-based deep learning system for human behavior recognition in smart city monitoring using the Berkeley MHAD datasets. *Systems*, *10*(5), 177. https://doi.org/10.3390/systems10050177

[34] Kommidi, M., Chinta, A., Dachepally, T. K., & Chebrolu, S. (2022). Identification and localization of COVID-19 abnormalities on chest radiographs using ensemble deep neural networks. In *2022 1st International Conference on the Paradigm Shifts in Communication, Embedded Systems, Machine Learning and Signal Processing*, 21–26. https://doi.org/10.1109/PCEMS55161.2022.9807972

[35] Quach, L.-D., Quoc, K. N., Quynh, A. N., & Ngoc, H. T. (2023). Evaluating the effectiveness of YOLO models in different sized object detection and feature-based classification of small objects. *Journal of Advances in Information Technology*, *14*(5), 907–917. https://doi.org/10.12720/jait.14.5.907-917

[36] Zhan, W., Sun, C., Wang, M., She, J., Zhang, Y., Zhang, Z., & Sun, Y. (2022). An improved YOLOv5 real-time detection method for small objects captured by UAV. *Soft Computing*, *26*(1), 361–373. https://doi.org/10.1007/s00500-021-06407-8

[37] Raj, A., M, O. P., & N, M. (2024). Development of CheXNet-based web application to detect pneumonia using chest X-ray images. In *2024 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics*, 322–327. https://doi.org/10.1109/DISCOVER62353.2024.10750561

[38] Wijerathna, V., Raveen, H., Abeygunawardhana, S., & Ambegoda, T. D. (2022). Chest X-ray caption generation with CheXNet. In *2022 Moratuwa Engineering Research Conference*, 1–6. https://doi.org/10.1109/MERCon55799.2022.9906263

[39] Mishra, D., Sai, K. P., Sri HarshaVardhan, K., Saketh Ram, C. K., & Gabbi Reddy, K. (2025). Automated chest X-ray captioning and report generation using CheXNet and BioGPT. In *2025 IEEE International Conference on Advanced Visual and Signal-Based Systems*, 1–6. https://doi.org/10.1109/AVSS65446.2025.11149939

[40] Haque, A., Hasan, M., Suma, S. A., Vasker, N., Ashhab, M. M., Ali, M. S., . . . , & Hossain Khan, M. S. (2023). Detecting pneumonia from X-ray images of the chest using a deep convolutional neural network. In *2023 4th International Conference on Big Data Analytics and Practices*, 1–6. https://doi.org/10.1109/IBDAP58581.2023.10271996

[41] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *2017 IEEE Conference on Computer Vision and Pattern Recognition*, 3462–3471. https://doi.org/10.1109/CVPR.2017.369

How to Cite: Ahmed, M. A., Saeed, R. A., Saeed, M. M., Giha, A. M., Alhaj, T. A., Mohamed, A. Y., & Alfadeil, T. M. (2026). Ensemble DL for Improved Diagnosis of Respiratory Diseases Using YOLO and CheXNet on Chest X-rays. *Journal of Data Science and Intelligent Systems*. https://doi.org/10.47852/bonviewJDSIS62028092