A Dual-Branch Network for Diagnosis of Thorax Diseases From Chest X-Rays

G. Jignesh Chowdary and Vivek Kanhangad, Senior Member, IEEE

Abstract—Automated chest X-ray analysis has a great potential for diagnosing thorax diseases since errors in diagnosis have always been a concern among radiologists. Being a multi-label classification problem, achieving accurate classification remains challenging. Several studies have focused on accurately segmenting the lung regions from the chest X-rays to deal with the challenges involved. The features extracted from the lung regions typically provide precise clues for diseases like nodules. However, such methods ignore the features outside the lung regions, which have been shown to be crucial for diagnosing conditions like cardiomegaly. Therefore, in this work, we explore a dual-branch network-based framework that relies on features extracted from the lung regions as well as the entire chest X-rays. The proposed framework uses a novel network named R-I UNet for segmenting the lung regions. The dual-branch network in the proposed framework employs two pre-trained AlexNet models to extract discriminative features, forming two feature vectors. Each feature vector is fed into a recurrent neural network consisting of a stack of gated recurrent units with skip connections. Finally, the resulting feature vectors are concatenated for classification. The proposed models achieve state-of-the-art performance for both segmentation and classification tasks on the benchmark datasets. Specifically, our lung segmentation model achieves a 5-fold cross-validation accuracy of 98.18% and 99.14% on Montgomery (MC) and JSRT datasets. For classification, the proposed approach achieves state-of-the-art AUC for 9 out of 14 diseases with a mean AUC of 0.842 on the NIH ChestXray14 dataset.

Index Terms—Chest X-ray Classification, Chest X-ray Segmentation, Multi-label Classification, Thorax Disease Diagnosis.

I. INTRODUCTION

CHEST X-ray (CXR) is the most widely used noninvasive imaging technique for screening and diagnosing several thorax diseases, including pneumonia, cardiomegaly, and atelectasis. A CXR may contain more than one abnormality [1] and accurate diagnosis of these abnormalities relies heavily on the expertise of the trained medical professionals. The diagnosis of thorax diseases from CXR with naked eyes is time-consuming, and these diagnostic findings are prone to errors [2], especially in the presence of noise. Therefore, it is desirable to have an automated diagnostic system that aids in diagnosis of these diseases. Although a few methods have been proposed [1], [3], [4], more effort should be focused on performance enhancement to meet the standards required for their deployment in clinical settings.

In the recent literature, the methods employed for computer-aided diagnosis (CAD) can be broadly categorized into two categories. The CNN based methods [3] involve extraction of deep features that carry discriminatory information about multiple abnormalities using classical CNNs [5], [6], [7]. Such methods have also been developed for related tasks like the diagnosis of COVID-19 from CXR [8]. Generally, these methods do not perform well for the detection of small abnormalities like nodules. This may be because the extracted features do not carry sufficient information due to the limited spatial extent of such abnormalities in CXR. To improve the performance, researchers have explored attention-guided approaches [9], [10] to focus the model’s eye on the suspicious regions in CXRs. The second category includes methods [11], [12] that utilize medical knowledge (such as pathology interdependence and co-occurrence) by explicitly capturing dependencies among multiple labels/pathologies. These methods generally work well for multi-label CXR classification.

The primary sites for thorax abnormalities are the lungs. Therefore, the lung region in CXRs can be analysed for the detection of severe thorax conditions like pneumothorax, emphysema, effusion, and nodules. Recently, a few studies [13], [14], [15], [16], [17] have demonstrated the usefulness of lung segmentation methods for assisting medical professionals in identifying the suspicious regions. Focusing only on lung regions for CXR classification using the lung segmentation masks has been shown to improve the detection performance for abnormalities like nodules, pneumothorax, and emphysema [18], [19], [20].

A review of the literature indicates that there has not been any work that explored the combination of features extracted from the lung region as well as the entire CXR image, at the same time leveraging deep learning for multi-label CXR classification. This motivated us to explore a dual-branch network based approach that uses both local branch and global branch features for automated diagnosis of thorax diseases. The idea behind focusing on both global and local (the segmented lung) regions in...
our approach is that while most thorax diseases are limited to the lung region, conditions like cardiomegaly are better characterized by contextual features around the lungs. Cardiomegaly refers to an enlarged heart condition, which compresses the lungs resulting in breathlessness to the individual. In this case, smaller size of the lung regions and increased gap between the lungs are the crucial features for the diagnosis. The key contribution of this work can be summarized as follows:

1) A novel U-Net based segmentation model that extracts lung regions from CXRs.
2) A dual-branch network consisting of two pre-trained AlexNet models and a recurrent neural network block that learn effective representations for multi-label classification of CXRs.
3) This work advances the state-of-the-art in the multi-label classification of thorax diseases from CXRs, a significant step towards developing a CAD system for deployment in clinical settings.

The rest of this paper is organized as follows: Section II presents the related works, Section III describes the proposed framework, Section IV presents the details of the dataset, our experiments and discussion. Finally, Section V concludes the paper. Table I presents a complete list of abbreviations used in this paper.

### II. RELATED WORKS

#### A. Lung Segmentation

Most previous works on lung segmentation in CXR images relied on traditional features like texture, shape, contour to design rule-based methods [21], [22]. With recent advancements in deep learning, CNN-based methods have also been developed for this task. Souza et al. [17] and Maity et al. [14] proposed two different segmentation models for lung segmentation. Their models were trained and tested on the MC [23], and JSRT datasets [24]. Tang et al. [13] proposed a segmentation network named XLSor, which was designed using criss-cross attention modules for extraction of contextual information in various directions around each pixel. The authors have also introduced the NIH dataset for segmentation task. Eslami et al. [25] proposed a multi-task generative adversarial network named MTdG that segments anatomical structures in CXRs and produces rib suppressed images. Singh et al. [26] designed a segmentation model named Deep LF-Net for segmenting lungs in CXRs. Their model integrates DeepLab architecture with custom designed atrous convolution module.

### B. Classification

For multi-label CXR classification, some of the recent studies have investigated the use of classical CNNs. These methods generally outperform handcrafted feature-based methods. Wang et al. [3] investigated the performance of several pre-trained models for classifying CXR images from the ChestX-ray14 dataset [3]. Their study indicates that ResNet outperforms the other models considered in their study. Ma et al. [27] presented a model named ChestXNet that enhances the classification performance. This model was developed by fine-tuning the pre-trained model – DenseNet121 [6].

A few studies have also explored attention mechanisms and have achieved enhanced performance for CXR classification. Guendel et al. [23] employed DenseNet121 for CXR classification through transfer learning. To further enhance the performance of individual disease diagnosis, Tang et al. [28] proposed a multi-task framework for simultaneous classification and localization of thorax diseases. They employed attention-guided curriculum learning scheme to achieve performance enhancement in disease localization and classification. Guan et al. [4] designed a CNN named CRAL with a class-specific attention learning scheme for multi-label classification of CXRs. Xi et al. [29] proposed a weakly supervised algorithm with hierarchical attention mining for localization of CXR abnormalities and their classification.

Chen et al. [12] proposed a graph convolutional neural network with label co-occurrence learning framework for thorax disease classification. Chen et al. [10] designed a model named Lesion Location Attention Guided Network (LLAGnet) for CXR classification. In addition to the features from the full CXRs, it uses features from lesion locations to enhance the classification performance.

### III. PROPOSED APPROACH

An overview of the proposed dual-branch model for automated diagnosis of thorax diseases from CXR is shown in Fig. 1. The model takes a CXR image as input, which is processed by two modules, namely, the lung region segmentation (LRS) module and the classification module. While the LRS module segments the lung region from the input CXR, and the classification module has two branches to extract deep features from the segmented lungs and the entire CXR image. The extracted
A. Lung Region Segmentation

In the proposed approach, the lung regions in the input CXR are segmented using a novel network named R-I UNet. The segmented lung regions undergo post-processing to reduce false-negatives in the R-I UNet predictions.

1) R-I UNet: The proposed R-I UNet has a four-level segmentation architecture, which is shown in Fig. 2. This semantic segmentation network is composed of three paths, namely Encoder, Decoder and Bridge. The encoder generates a compact representation for the input CXR. The decoder recovers pixel-wise classification from the encoded input. The bridge acts as a connection between the encoder and the decoder. The basic building block of this segmentation network is a Residual-Inception (R-I) block and all the paths are created using R-I blocks.
The structure of the proposed R-I block. + and × represent the predicted and the ground truth. C represents the input.

\[ L_s = 1 - \frac{1 + 2 \times N_s \times G_s}{1 + N_s + G_s} \] (2)

Here, \( N_s \) and \( G_s \) represent the predicted and the ground truth segmentation mask, respectively.

2) Post-Processing: The predicted segmentation mask typically has a considerable number of false positive and false negative predictions. To reduce these false predictions, we have used a set of morphological operations. Specifically, we have performed opening and area filtering to reduce false positives while retaining the two largest objects (lungs), followed by closing to reduce false negative predictions. The structuring elements used for opening and closing operations are cross and ellipse of size 5 × 5 and 7 × 7, respectively.

B. Classification

As can be seen in the structure of the proposed model presented in Fig. 1, the classification module consists of two parallel branches namely, global and local branch that extract features useful for discrimination. While the global branch extracts deep features from the entire CXR image, the local branch focuses on the two segmented lung regions for feature extraction. The input image to the local branch is of the same size as full CXR, with its non-lung pixels set to zero. Each of the two branches consists of a pre-trained AlexNet followed by a gated recurrent unit (GRU). The features extracted by the two branches are concatenated and passed to the final dense layer activated with the sigmoid function for prediction.

In this work, we have used transfer learning technique for AlexNet. Specifically, we have replaced its dense layers with a single dense layer consisting of 14 neurons with ReLU activation function. The output of this layer is fed into a GRU block. A GRU is an improved version of the standard RNN, which is a kind of feed-forward neural network primarily used for sequence learning tasks. In the proposed approach, a GRU block formed using 5 GRUs with skip connections help exploit feature dependencies [33], thereby enhancing the classification performance of the model. The structure of our GRU block is shown in Fig. 4.

The skip connections not only help to overcome the problem of vanishing gradients but also make the loss function less chaotic, making its minimization easier [34], [35]. The processing in a single GRU can be mathematically represented as follows:

\[ Z_t = \sigma_g(W_z x_t + U_z H_{t-1} + b_z) \] (3)

As can be seen in Fig. 2, the proposed R-I UNet has four R-I blocks in its encoder path. In each of these blocks, a stride of 2 is used in the first convolutional layer to obtain a down-sampled feature map. Correspondingly, the decoder path is composed of four R-I blocks. The upsampled feature map from the lower level and the feature map at the same level from the corresponding encoder path are concatenated and fed into each of the R-I blocks in the decoder. At the end of the decoder path, a 1 × 1 convolutional layer with a sigmoid activation function is employed to obtain the desired segmentation output.

Our lung segmentation network is trained using a dice coefficient-based loss function, which measures the shape similarity between the ground truth and the predicted lung masks. Specifically, the segmentation loss is defined as follows:

\[ L_s = 1 - \frac{1 + 2 \times N_s \times G_s}{1 + N_s + G_s} \] (2)

Fig. 3. The structure of the proposed R-I block.

In this work, we have used transfer learning technique for AlexNet. Specifically, we have replaced its dense layers with a single dense layer consisting of 14 neurons with ReLU activation function. The output of this layer is fed into a GRU block. A GRU is an improved version of the standard RNN, which is a kind of feed-forward neural network primarily used for sequence learning tasks. In the proposed approach, a GRU block formed using 5 GRUs with skip connections help exploit feature dependencies [33], thereby enhancing the classification performance of the model. The structure of our GRU block is shown in Fig. 4. The skip connections not only help to overcome the problem of vanishing gradients but also make the loss function less chaotic, making its minimization easier [34], [35]. The processing in a single GRU can be mathematically represented as follows:

\[ Z_t = \sigma_g(W_z x_t + U_z H_{t-1} + b_z) \] (3)

As can be seen in Fig. 2, the proposed R-I UNet has four R-I blocks in its encoder path. In each of these blocks, a stride of 2 is used in the first convolutional layer to obtain a down-sampled feature map. Correspondingly, the decoder path is composed of four R-I blocks. The upsampled feature map from the lower level and the feature map at the same level from the corresponding encoder path are concatenated and fed into each of the R-I blocks in the decoder. At the end of the decoder path, a 1 × 1 convolutional layer with a sigmoid activation function is employed to obtain the desired segmentation output.

Our lung segmentation network is trained using a dice coefficient-based loss function, which measures the shape similarity between the ground truth and the predicted lung masks. Specifically, the segmentation loss is defined as follows:

\[ L_s = 1 - \frac{1 + 2 \times N_s \times G_s}{1 + N_s + G_s} \] (2)

Here, \( N_s \) and \( G_s \) represent the predicted and the ground truth segmentation mask, respectively.

2) Post-Processing: The predicted segmentation mask typically has a considerable number of false positive and false negative predictions. To reduce these false predictions, we have used a set of morphological operations. Specifically, we have performed opening and area filtering to reduce false positives while retaining the two largest objects (lungs), followed by closing to reduce false negative predictions. The structuring elements used for opening and closing operations are cross and ellipse of size 5 × 5 and 7 × 7, respectively.

B. Classification

As can be seen in the structure of the proposed model presented in Fig. 1, the classification module consists of two parallel branches namely, global and local branch that extract features useful for discrimination. While the global branch extracts deep features from the entire CXR image, the local branch focuses on the two segmented lung regions for feature extraction. The input image to the local branch is of the same size as full CXR, with its non-lung pixels set to zero. Each of the two branches consists of a pre-trained AlexNet followed by a gated recurrent unit (GRU). The features extracted by the two branches are concatenated and passed to the final dense layer activated with the sigmoid function for prediction.

In this work, we have used transfer learning technique for AlexNet. Specifically, we have replaced its dense layers with a single dense layer consisting of 14 neurons with ReLU activation function. The output of this layer is fed into a GRU block. A GRU is an improved version of the standard RNN, which is a kind of feed-forward neural network primarily used for sequence learning tasks. In the proposed approach, a GRU block formed using 5 GRUs with skip connections help exploit feature dependencies [33], thereby enhancing the classification performance of the model. The structure of our GRU block is shown in Fig. 4. The skip connections not only help to overcome the problem of vanishing gradients but also make the loss function less chaotic, making its minimization easier [34], [35]. The processing in a single GRU can be mathematically represented as follows:

\[ Z_t = \sigma_g(W_z x_t + U_z H_{t-1} + b_z) \] (3)
The structure of the residual GRU block.

\[ R_t = \sigma_g(W_r x_t + U_z H_{t-1} + b_r) \]  
\[ H_t = \tanh(W_h x_t + U_r (r_t) \otimes H_{t-1}) \]  
\[ H_t = (1 - Z_t) \otimes H_{t-1} + Z_t \otimes H_t \]

In (3)–(6), \( x_t \) represents the input vector, \( H_t \) represents the output vector, \( R_t \) is the candidate activation vector, \( Z_t \) is the update gate, \( r_t \) is the reset gate, \( b \), and \( U, W \) represent the vector and parameter matrices, respectively.

The outputs of the two GRU blocks that are part of the global and local branches are concatenated to form a 28-dimensional feature vector. This feature vector is then fed into another dense layer consisting of 14 neurons with the sigmoid activation function to generate probability scores for classification.

We have frozen the convolutional base of AlexNet while training. Its newly added dense layer along with the residual GRU and the final dense layer are jointly trained using the focal loss [36], which is defined as follows:

\[ FL = \alpha (1 - p_y) \gamma \log(p_y) \]

In the above equation, \( p_y \) is the probability predicted for each class by the model, \( \alpha \) is the weighting factor, and \( \gamma \) is the focusing parameter. In this work, \( \alpha \) and \( \gamma \) are empirically set to 0.5, and 3, respectively using validation set. We have employed the focal loss function primarily because it focuses on hard samples, resulting in a reduced number of misclassified samples compared to the standard cross-entropy loss. The steps involved in processing a CXR at inference time are detailed in Algorithm 1.

### IV. DATA AND EXPERIMENTS

#### A. Datasets

In this work, we have used three publicly available datasets, namely, JSRT, Montgomery (MC), and NIH ChestXray14, for performance evaluation. While the first two datasets have been used to evaluate the segmentation model, ChestXray14 has been used to evaluate the proposed dual-branch network based framework. A detailed overview of these datasets is presented below.

The JSRT dataset [24] is created by the Japanese Radiological Society (JRT) in collaboration with the Japanese Society of Radiological Technology (JSRT). This dataset consists of 247 posterior-anterior (PA) view CXR images of 2048 \( \times \) 2048 pixels. Out of these 247 CXRs, 90 CXRs are of healthy lungs without any abnormalities, and the rest 154 CXRs contain lung nodules. The ground truth masks for lung segmentation in CXRs of the JSRT dataset are provided in [37].

The Montgomery County dataset (MC) [23] is created by the Department of Health and Human Services of Montgomery County, Maryland. This dataset contains 138 frontal CXR images of 4020 \( \times \) 4892 or 4892 \( \times \) 4020 pixels. There are 80 CXRs of healthy people and 58 CXRs of patients infected with Tuberculosis. This dataset also contains segmentation ground truth annotated by experienced radiologists.

ChestXray14 dataset [3] consists of 112,120 frontal CXR images of 30,805 patients collected from 1992 to 2015. Among them, 51,708 CXRs are categorized into 14 thorax abnormalities, and the rest are labeled as “no findings”. The images in this dataset are of 1024 \( \times \) 1024 pixels with 8-bit depth. The official dataset split is created by randomly splitting the data on the patient level into train (~70%), validation (~10%), and test partitions (~20%) while ensuring that all images belonging to a patient are included in only one of these sets [3].

#### B. Performance Metrics

We have used the following metrics for evaluating the proposed lung segmentation network:

\[ \text{Accuracy (ACC)} = \frac{TP + TN}{TP + TN + FP + FN} \]  
\[ \text{Recall (REC)} = \frac{TP}{TP + FN} \]  
\[ \text{Specificity (SPE)} = \frac{TN}{TN + FP} \]
TABLE II
HYPERPARAMETERS

| Hyperparameters       | Models         |
|-----------------------|----------------|
|                       | R-I UNet  | Dual-branch Network |
| Epochs                | 200      | 150                 |
| Batch size            | 16       | 8                   |
| Initial learning rate | 0.0005   | 0.0003              |
| Optimizer             | Adam     | SGD with momentum set to 0.7 |
| Dense layer neurons   | -        | 14                  |

\[
\text{Dice coefficient (DC)} = \frac{2TP}{2TP + FP + FN} \quad (11)
\]

\[
\text{Jaccard score (JS)} = \frac{TP}{FP + TP + FN} \quad (12)
\]

In the above equations, TP represents the number of pixels that are correctly classified as belonging to the lung region (true positives), TN represents the number of pixels that are correctly classified as belonging to the background (true negatives), FP and FN represent the number of pixels that are wrongly classified as belonging to the lung region (false positives) and background (false negatives), respectively.

As with the existing methods, we have computed area under the ROC curve (AUC) to evaluate the performance of the proposed dual-branch network based CXR classification framework. Specifically, we have computed AUC for each class and compared with the existing approaches.

C. Training Strategy

As mentioned previously, the proposed segmentation model – R-I UNet has been evaluated on JSRT and Montgomery dataset. We have adopted the standard five-fold cross validation protocol for evaluation. The model is trained for 200 epochs with a batch-size of 16 and the initial learning rate set to 0.0005, which is reduced by a factor of 20 for every 40 epochs.

The proposed classification framework has been evaluated on ChestXray14 dataset. We have used the official dataset splits provided as part of the dataset. During training, the input CXRs are resized to $224 \times 224$ pixels. Our model is trained for 150 epochs with a batch-size of 8 and the initial learning rate set to 0.0003, which is reduced by a factor of 15 for every 20 epochs. We have used the SGD optimizer with a momentum of 0.7. All our experiments have been performed using the Keras framework on Google Colab with a single 16 GB P100 GPU. Table II summarises hyperparameters of our segmentation and classification models. All of these hyperparameters are tuned using validation sets. To avoid overfitting, we have employed the early stopping strategy during training. Specifically, if the loss is not improving for ten epochs, the training is stopped automatically, and the best model weights are restored.

D. Results and Discussion

1) Segmentation: In this section, we present the performance of the proposed R-I UNet on the two datasets. Table III presents performance of the individual models in each iteration of the five-fold cross validation. This table also presents mean performance estimates.

To benchmark the performance of the proposed segmentation model, we have considered a set of state-of-the-art segmentation models, namely, UNet++ [38], DeepLabV3+ [39], Mask R-CNN [40], and UNet [41]. We have followed the previously described strategy while training these models. The performance of these models on JSRT and MC datasets is presented in Table IV. The results presented in Table IV indicate that the proposed R-I UNet provides improved lung segmentation results on both the datasets. To ascertain the performance of our model, we have compared the lung segmentation results qualitatively. Fig. 5 shows the results achieved by different models on five CXRs. While lungs are clearly visible in the first and the second input CXRs, the third image is a challenging case with low contrast between lungs and neighboring region. The fourth and fifth input CXRs are even more challenging as parts of lungs are not clearly visible due to certain medical conditions.

Our qualitative analysis indicates that all of these models achieve good segmentation results on the first two CXRs with the predicted lung contours being close to the ground truth. On the third CXR, only our model and DeepLabV3+ achieve satisfactory results. On the fourth and the fifth CXR, all state-of-the-art models achieve poor results with high false negative predictions.
On the contrary, our model’s predictions are close to the ground truth.

In addition, we have compared our model with the existing lung segmentation approaches. For this purpose, we have evaluated our model using three different protocols to make fair comparisons with the existing approaches. Specifically, we have performed evaluations by splitting each dataset into training and testing set in the 70:30 and 80:20 ratios and using 5-fold cross-validation. As can be seen in Table V, our model provides a considerable improvement over the existing methods. Importantly, our model achieves higher recall on both the datasets, which is one of the most important metrics used to evaluate a computational model in medical informatics.

2) Classification: Table VI presents the CXR classification results. Specifically, this table presents the AUC for individual classes. For comparison, we have reported the performance of a set of existing approaches, specifically CRAL [4], CheXGCN [12], Li et al. [43], Wang et al. [3], Xi et al. [29], Li et al. [44], LLAGnet [10] which have also been evaluated on ChestXray14 using the same dataset splits.

Our approach achieves state-of-the-art classification performance for 9 out of 14 diseases. The LLAGnet [10] achieves the best classification results for 3 out of the remaining 5 diseases, while Li et al. [44] achieves the highest AUC for Effusion and CheXGCN [12] achieves approximately the same AUC for Emphysema. Importantly, our approach provides a considerable improvement in AUC for most of those 9 diseases. It also achieves the highest mean AUC, which indicates its better overall classification performance.

Further, we have compared our approach with LLAGnet [10] in terms of the classification accuracy (ACC_c), recall (REC_c) and specificity (SPE_c). Here the subscript is used to distinguish these classification metrics from the ones used for segmentation. Initially, we have calculated TP, TN, FP, and FN for each class...
TABLE VI
COMPARISON OF THE CLASSIFICATION PERFORMANCE WITH THE EXISTING APPROACHES

| Diseases         | CRAL [4] | CheXGCN [12] | Li et al. [43] | Wang et al. [3] | Xi et al. [29] | Li et al. [44] | LLAGnet [10] | ACPL [45] | Ours   |
|------------------|----------|--------------|----------------|----------------|---------------|----------------|---------------|----------|--------|
| Hernia           | 0.917    | 0.929        | 0.69           | 0.87           | 0.91          | 0.93           | 0.916         | 0.858    | 0.957  |
| Pleural thickening| 0.778    | 0.795        | 0.73           | 0.68           | 0.79          | 0.793          | 0.798         | 0.776    | 0.829  |
| Fibrosis         | 0.83     | 0.834        | 0.77           | 0.79           | 0.83          | 0.847          | 0.832         | 0.818    | 0.792  |
| Emphysema        | 0.908    | 0.944        | 0.88           | 0.83           | 0.93          | 0.918          | 0.939         | 0.933    | 0.942  |
| Edema            | 0.85     | 0.850        | 0.81           | 0.81           | 0.84          | 0.854          | 0.851         | 0.849    | 0.861  |
| Consolidation    | 0.754    | 0.751        | 0.71           | 0.70           | 0.74          | 0.758          | 0.754         | 0.755    | 0.793  |
| Pneumothorax     | 0.857    | 0.876        | 0.81           | 0.80           | 0.88          | 0.863          | 0.877         | 0.869    | 0.85   |
| Pneumonia        | 0.729    | 0.739        | 0.65           | 0.66           | 0.72          | 0.742          | 0.729         | 0.738    | 0.769  |
| Nodule           | 0.773    | 0.800        | 0.70           | 0.67           | 0.79          | 0.75           | 0.790         | 0.777    | 0.824  |
| Mass             | 0.834    | 0.840        | 0.78           | 0.69           | 0.83          | 0.836          | 0.841         | 0.824    | 0.827  |
| Infiltration     | 0.702    | 0.699        | 0.67           | 0.66           | 0.71          | 0.711          | 0.703         | 0.714    | 0.764  |
| Effusion         | 0.829    | 0.832        | 0.79           | 0.76           | 0.83          | 0.830          | 0.834         | 0.835    | 0.812  |
| Cardiomegaly     | 0.88     | 0.893        | 0.84           | 0.81           | 0.89          | 0.908          | 0.885         | 0.89     | 0.935  |
| Atelectasis      | 0.781    | 0.786        | 0.73           | 0.70           | 0.78          | 0.776          | 0.783         | 0.795    | 0.845  |
| Mean             | 0.816    | 0.826        | 0.755          | 0.745          | 0.817         | 0.824          | 0.824         | 0.817    | 0.842  |

TABLE VII
COMPARISON OF THE CLASSIFICATION PERFORMANCE WITH LLAGnet [10]

| Disease   | ACCc (%) | RECc (%) | SPBc (%) | PREc (%) | F1 (%) |
|-----------|----------|----------|----------|----------|--------|
| LLAGnet   |          |          |          |          |        |
| Ours      |          |          |          |          |        |
| Mean      | 76.7     | 84.13    | 74.3     | 83.7     | 83.68  |

by comparing the predicted scores with the thresholds set for individual classes, as has been done in [10]. The results presented in Table VII indicate that our approach provides consistently better classification performance. Additionally, we have computed precision (PREc) and F1-score (F1), which are defined as follows:

\[
\text{Precision (PRE}_c\text{)} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (13)
\]

\[
\text{F1 - score (F1)} = \frac{2 \times \text{PRE}_c \times \text{RE}_c}{\text{PRE}_c + \text{RE}_c} \quad (14)
\]

The superior performance of our dual-branch network can be attributed to its two keys aspects, i.e., its ability to segment lungs more accurately and learn a better representation for classification by combining features from the entire CXR as well as the segmented lung region. This approach effectively learns multiscale features for CXR classification. The proposed model has been trained on CXR images of size \(224 \times 224\) pixels. Some of the previous works [3], [10] have studied the effect of input image size on the classification performance. Using higher resolution input image (e.g. \(512 \times 512\) or \(1024 \times 1024\)) has only led to marginal improvement in the overall classification performance. We have not compared our results with [37] as this existing work has not used official dataset splits for performance evaluation, and therefore a fair comparison cannot be made.

a) Qualitative analysis: We present results of a qualitative evaluation of our approach and compare the results with LLAGnet [10]. To perform this analysis, we have selected the same set of CXRs that has been used in [10] for qualitative evaluation. The images and their predictions by our approach and LLAGnet are shown in Fig. 6. Specifically, we present top-8 predicted scores for each test sample and highlight the positive classes in red color. As can be seen, the...
prediction scores generated by our approach for positive classes are significantly higher than the ones generated for negative classes. While these differences can be observed in the case of LLAGnet as well, the margins are considerably lower. We have analysed the performance of our approach on failure cases presented in [10]. The results presented in Fig. 7 clearly indicate that our approach learns more discriminative representations for CXR classification. The predicted score for the positive class is the highest for each of the images except for the third CXR. In this case too, the predicted score for the positive class (consolidation) is quite high and likely to exceed any threshold set appropriately.

b) Grad-CAM visualization: We have generated Grad-CAM visualizations to gain a better understanding of our classification network, specifically its global branch. Fig. 8 provides a visual explanation indicating the image regions our network focuses on for its predictions. As can be seen, the localized region in each case corresponds to an abnormality in CXR. For example, consider the case of Cardiomegaly. The enlarged cardiac silhouette can be seen in the CXR, and Grad-CAM localizes this region well. The correspondence between a CXR abnormality and the localized region can also be observed for most of the other disease classes as well. These Grad-CAM visualizations indicate that our dual-branch classification network focuses on discriminative regions for its predictions.

c) Ablation study: We have carried out ablation studies to assess the effectiveness of the two branches in the proposed dual-branch network for CXR classification. In this set of experiments, we have trained and evaluated the global and local branch individually on ChestXray14 dataset. The AUC for each class along with the mean AUC are presented in Table VIII. As can be observed, the local branch achieves marginally higher AUC compared with the global branch. Lung nodules, for example, are better diagnosed when the model is trained on segmented lungs. The plausible reason is that the segmentation eliminates noisy regions and other structures such as heart and thoracic spine that are likely to impact the diagnosis of these small lesions. Importantly, the results of this study indicate that both the local and the global branches learn effective representations and that the proposed dual-branch network clearly benefits from the fusion of features extracted by its individual branches.

We have also studied the effectiveness of GRU blocks in the proposed dual-branch network. To this end, we have performed two sets of experiments. Firstly, we have trained and evaluated the performance of each branch separately without its GRU block. Secondly, we have trained the network after removing the GRU blocks from both the branches and evaluated the classification performance of the dual-branch network. The results presented in Table VIII clearly indicate that capturing dependencies in feature sequences using GRU leads to significantly improved classification accuracy.

d) Failure cases: We have also analysed CXRs that have been misdiagnosed by our approach. Fig. 9 shows a few failure cases. In general, multi-label classification of these CXRs appears to be a non-trivial task due to different reasons, including severe lung conditions. For example, the first CXR is a very low contrast image, due to which the lungs are not clearly visible. In the second image, there appears to be a tube device that may have caused misdiagnosis. In the third image, one of the lungs is not visible, which makes the diagnosis of multiple conditions difficult. Failing to handle these cases appears to be a limitation of our approach.

Fig. 6. Qualitative comparison of the classification performance with LLAGnet [10].
Our extensive evaluation of the proposed models indicates that they advance the state-of-the-art in lung segmentation and multi-label classification of thorax diseases from CXRs. However, on the flip side, our segmentation model has significantly more parameters than the existing ones. Specifically, it has 140 million (M) trainable parameters, while the other segmentation models in Table IV, Mask R-CNN, UNet, DeepLabV3+, and UNet++, have only 64 M, 34 M, 11 M, and 4 M parameters, respectively. On the other hand, our dual-branch classification model is lighter with 0.7 M trainable parameters. The average inference time of our approach is 9.53 seconds. This can be a limiting factor in adopting our models for some real-world applications.

V. CONCLUSION AND FUTURE DIRECTIONS

In this paper, we have presented a dual-branch network for CXR-based diagnosis of thorax diseases. The proposed framework consists of a segmentation model named R-I UNet and a classification network. The proposed R-I UNet uses U-Net model as its backbone, wherein novel residual-inception blocks replace the convolutional layers. The classification network consists of two branches, namely, the local and the global branches. The local branch extracts features from the segmented lungs, while the global branch extracts features from the input CXR. These feature sequences are processed by two GRU blocks independently and their outputs are concatenated to obtain a single feature vector, which is passed through a dense layer activated by sigmoid function to generate prediction scores. We have employed transfer learning techniques to design our classification network. Our experimental results suggest that the proposed framework can be adopted for a more accurate diagnosis of thorax diseases from CXRs. Our study also indicates that while it is essential to focus on the lung region, contextual features also provide clues and their fusion improves the performance of CXR-based automated disease diagnosis.
Fig. 8. Grad-CAM visualizations from the global branch of the proposed classification model.

Fig. 9. Failure cases of our approach.

REFERENCES

[1] B. Chen, J. Li, X. Guo, and G. Lu, “Dualchexnet: Dual asymmetric feature learning for thoracic disease classification in chest X-rays,” Biomed. Signal Process. Control, vol. 53, 2019, Art. no. 101554.

[2] M. S. Bogner, Human Error in Medicine. Boca Raton, FL, USA: CRC Press, 2018.

[3] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, “ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 2097–2106.

[4] Q. Guan and Y. Huang, “Multi-label chest X-ray image classification via category-wise residual attention learning,” Pattern Recognit. Lett., vol. 130, pp. 259–266, 2020.

[5] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 770–778.

[6] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 4700–4708.

[7] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 7132–7141.

[8] D. A. D. Júnior et al., “Automatic method for classifying COVID-19 patients based on chest X-ray images, using deep features and PSO-optimized XGBoost,” Expert Syst. Appl., vol. 183, 2021, Art. no. 115452.

[9] J. Cai, L. Lu, A. P. Harrison, X. Shi, P. Chen, and L. Yang, “Iterative attention mining for weakly supervised thoracic disease pattern localization in chest X-rays,” in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention, 2018, pp. 589–598.

[10] B. Chen, J. Li, G. Lu, and D. Zhang, “Lesion location attention guided network for multi-label thoracic disease classification in chest X-rays,” IEEE J. Biomed. Health Inform., vol. 24, no. 7, pp. 2016–2027, Jul. 2020.

[11] L. Yao, E. Poblenz, D. Dagunts, B. Covington, D. Bernard, and K. Lyman, “Learning to diagnose from scratch by exploiting dependencies among labels,” 2017, arXiv:1711.10501.

[12] B. Chen, J. Li, G. Lu, H. Yu, and D. Zhang, “Label co-occurrence learning with graph convolutional networks for multi-label chest x-ray image classification,” IEEE J. Biomed. Health Inform., vol. 24, no. 8, pp. 2292–2302, Aug. 2020.

[13] Y.-B. Tang, Y.-X. Tang, J. Xiao, and R. M. Summers, “XLSOR: A robust and accurate lung segmentor on chest X-rays using criss-cross attention and customized radiorealistic abnormalities generation,” in Proc. Int. Conf. Med. Imag. Deep Learn., 2019, pp. 457–467.

[14] A. Maity, T. R. Nair, S. Mehta, and P. Prakash, “Automatic lung parenchyma segmentation using a deep convolutional neural network from chest X-rays,” Biomed. Signal Process. Control, vol. 73, 2022, Art. no. 103398.
[15] P. Chondro, C.-Y. Yao, S.-J. Ruan, and L.-C. Chien, “Low order adaptive region growing for lung segmentation on plain chest radiographs,” Neurocomputing, vol. 275, pp. 1002–1011, 2018.

[16] Y. Tian, J. Wang, W. Yang, J. Wang, and D. Qian, “Deep multi-instance transfer learning for pneumothorax classification in chest X-ray images,” Med. Phys., vol. 49, no. 1, pp. 231–243, Jan. 2022.

[17] J. C. Souza, J. O. B. Diniz, J. L. Ferreira, G. L. F. da Silva, A. C. Silva, and A. C. de Paiva, “An automatic method for lung segmentation and reconstruction in chest X-ray using deep neural networks,” Comput. Methods Programs Biomed., vol. 177, pp. 285–296, 2019.

[18] D. Gu, G. Liu, and Z. Xue, “On the performance of lung nodule detection, segmentation and classification,” Comput. Med. Imag. Graph, vol. 89, 2021, Art. no. 101886.

[19] Q. Wang et al., “Automated segmentation and diagnosis of pneumothorax on chest X-rays with fully convolutional multi-scale scse-densenet: A retrospective study,” BMC Med. Inf. Decis. Mak., vol. 20, no. 14, pp. 1–12, 2020.

[20] M. Negahdar and D. Beymer, “Lung tissue characterization for emphysema differential diagnosis using deep convolutional neural networks,” Med. Imag. Comput.-Aided Diagnosis Int. Soc. Opt. Photon., vol. 10950, 2019, Art. no. 109503R.

[21] T. Xu, M. Mandal, R. Long, I. Cheng, and A. Basu, “An edge-region force guided active shape approach for automatic lung field detection in chest radiographs,” Comput. Med. Imag. Graph, vol. 36, no. 6, pp. 452–463, 2012.

[22] L. Li, Y. Zheng, M. Kallergi, and R. A. Clark, “Improved method for automatic identification of lung regions on chest radiographs,” Academic Radiol., vol. 8, no. 7, pp. 629–638, 2001.

[23] S. Guendel, S. Grbic, B. Georgescu, S. Liu, A. Maier, and D. Comaniciu, “Learning to recognize abnormalities in chest X-rays with location-aware dense networks,” in Iberoamerican Congress on Pattern Recognition, Berlin, Germany: Springer, 2018, pp. 757–765.

[24] J. Shiraiishi, “Development of a digital image database for chest radiographs with and without a lung nodule: ROC analysis on radiologists’ performance in detection of pulmonary nodules,” Am J Roentgenol, vol. 174, no. 1, pp. 71–74, 2000.

[25] M. Eslami, S. Tabarestani, S. Albarqouni, E. Adeli, N. Navab, and M. Adjouadi, “Image-to-images translation for multi-task organ segmentation and bone suppression in chest X-ray radiography,” IEEE Trans. Med. Imag., vol. 39, no. 7, pp. 2553–2565, Jul. 2020.

[26] A. Singh et al., “Deep LF-Net: Semantic lung segmentation from Indian chest radiographs including severely unhealthy images,” Biomed. Signal Process. Control, vol. 68, 2021, Art. no. 102666.

[27] W.-D. K. Ma, J. Lewis, and W. B. Kleinj, “The HSIC bottleneck: Deep learning without back-propagation,” in Proc. AAAI Conf. Artif. Intell., 2020, pp. 34, pp. 5085–5092.

[28] Y. Tang, X. Wang, A. P. Harrison, L. Lu, J. Xiao, and R. M. Summers, “Attention-guided curriculum learning for weakly supervised classification and localization of thoracic diseases on chest radiographs,” in Proc. Int. Workshop Mach. Learn. Med. Imag., 2018, pp. 249–258.

[29] X. Ouyang et al., “Learning hierarchical attention for weakly-supervised chest X-ray abnormality localization and diagnosis,” IEEE Trans. Med. Imag., vol. 40, no. 10, pp. 2698–2710, Oct. 2021.