Vol.67: e24231088, 2024
https://doi.org/10.1590/1678-4324-2024231088
ISSN 1678-4324 Online Edition

BABT
Brazilian Archives of Biology and Technology

Article - Engineering, Technology and Techniques

Multimodal Deep Dilated Convolutional Learning for Lung Disease Diagnosis

Kanchi Anantharaman Varunkumar[1]
https://orcid.org/0000-0001-9281-2273

Mikhail Zymbler[2*]
https://orcid.org/0000-0001-7491-8656

Sachin Kumar[3]
https://orcid.org/0000-0003-3949-0302

[1]SRM Institute of Science and Technology, School of Computing, Department of Networking and Communication, Kattankulathur, Tamilnadu, India; [2]South Ural State University, Department of Computer Science, Chelyabinsk, Russia; [3]South Ural State University, Big Data and Machine Learning, Chelyabinsk, Russia.

Editor-in-Chief: Alexandre Rasi Aoki
Associate Editor: Fabio Alessandro Guerra

Received: 31-Oct-2023; Accepted: 08-Mar-2024

*Correspondence: mzym@susu.ru; (M.Z.).

HIGHLIGHTS

- Multimodal deep learning approach for lung disease diagnosis is proposed.
- Intermediate fusion is used to fuse different lung modalities.
- Proposed model performed better on multimodal data than single modality.

Abstract: Accurate and timely identification of pulmonary disease is critical for effective therapeutic intervention. Computed tomography (CT), chest radiography (x-ray) and positron emission tomography (PET) scans are some examples of traditional diagnostic methods that rely on single-modality imaging. However, these methods are not always accurate or useful. This paper presents a novel strategy to overcome this obstacle by developing a multimodal deep learning framework. Current diagnostic techniques mostly prioritize the analysis of a single modality, which limits the holistic understanding of lung diseases. This limitation hinders the accuracy of diagnoses and the ability to tailor therapies to individual patients. To address this disparity, the proposed research presents a novel multimodal deep learning framework that effectively incorporates data from CT, X-ray, and PET scans. This approach allows for the extraction of features that are unique to each modality. Fusion methods, such as late or early fusion, are used to effectively capture synergistic information from multiple modalities. Adding more convolutional neural network (CNN) layers and pooling operations to the model improves the ability to obtain abstract representations. This is followed by the use of fully connected layers for classification purposes. The model is trained using appropriate loss functions and optimized using gradient-based techniques. The proposed methodology shows a significant improvement in the accuracy of lung disease diagnosis compared to conventional methods using a single modality.

Keywords: Multimodal Deep Learning; Lung Disease; Precise Diagnosis.

INTRODUCTION

Lung diseases represent a significant global health dilemma, affecting a large number of patients and requiring accurate and timely diagnosis to ensure effective treatment [1]. Conventional diagnostic techniques, however useful, may be limited in effectively distinguishing the intricacies of different lung diseases [2]. This has spurred the search for novel methods that have the potential to transform the diagnostic field. The application of deep learning to the field of lung disease diagnosis has recently emerged as a viable approach to overcome the prevailing obstacles [3]. Lung diseases cover a wide range of problems including, but not limited to, infections, inflammatory disorders, and neoplastic growths [4]. Sophisticated diagnostic techniques are required to effectively understand the unique characteristics and complex nature of these diseases. These tools should provide detailed insights into the pathology and enable the development of targeted and tailored treatment approaches [5].

Accurate identification of disease is a fundamental aspect of successful medical management of respiratory disease [6]. Traditional diagnostic techniques, which typically rely on single-modality imaging such as CT scans or chest X-rays, may be limited in providing the comprehensive range of information required for accurate diagnosis [7]. This limitation underscores the urgent need for sophisticated diagnostic methods that can provide a more comprehensive understanding of lung disease [4]. Deep learning, which falls under the umbrella of artificial intelligence, has attracted considerable interest due to its potential to revolutionize medical diagnostics, particularly in the field of lung disease diagnosis [8, 9]. It is possible for deep learning models to make diagnostic processes more accurate and faster by using multi-layered neural networks to autonomously learn complex patterns and features from different data sets [10]. Deep learning shows potential in the field of pulmonary diseases due to its ability to extract small anomalies from imaging data and improve the decision-making process with more comprehensive information [11].

Although deep learning shows great promise in the field of lung disease diagnosis, there are still persistent obstacles that need to be addressed. The limited availability of annotated and diverse datasets poses a challenge in training models that can achieve high levels of robustness [12]. The interpretability of deep learning models in medicine remains a significant challenge. It is essential to understand the reasoning behind the judgments made by these models in order to gain confidence in clinical settings. In addition, researchers are constantly striving to address the ongoing difficulty of ensuring the generalizability of models across different patient populations and disease subtypes.

The primary goal of this study is to improve the accuracy of lung disease diagnosis through the development of a multimodal deep learning system [13]. The proposed system aims to effectively incorporate data from CT, chest x-ray, and PET scans to provide a comprehensive and nuanced understanding of lung disease, overcoming the limitations of traditional single-modality approaches. The novelty of the proposed method lies in the use of a multimodal deep convolutional learning model that goes beyond the limits of single-modality analysis. The proposed method extracts modality-specific features from the data by combining different Convolutional Neural Network (CNN) branches with dilated convolutions for each imaging mode. The use of advanced fusion techniques, whether used as late or early fusion methods, makes it easier to obtain information from multiple sources that work well together. The addition of CNN layers and pooling operations improves the model's ability to learn abstract representations, which ultimately leads to better diagnostic accuracy. In general, this work presents a novel approach that utilizes many modes of information to improve the proposed understanding of lung diseases, overcoming the limitations of traditional single-modality methods.

MATERIAL AND METHODS

This section will provide a detailed overview of the proposed methods and the data set used.

Overview of the Proposed Architecture

This system architecture design aims to exploit the unique advantages of multiple imaging modalities, such as CT, chest x-ray, and PET scans, in a multimodal deep learning context. The architectural design utilizes CNN with a unique emphasis on dilated convolutions. Each image modality is processed separately using specialized CNN branches. The integration of dilated convolutions within these branches enables the extraction of features that are unique to each modality, thereby improving the model's ability to capture subtle information.

The use of dilated convolutions in advanced fusion techniques involves the fusion of feature maps from many modalities through the process of intermediate fusion. The process of final fusion involves the fusion of feature maps to produce a complete representation. In the context of additional CNN layers and pooling, intermediate processing involves the use of conventional CNN layers to facilitate hierarchical feature extraction. Pooling is incorporated into the network architecture to reduce spatial dimensions. The output feature map is an enhanced and abstracted feature map. In the context of classification using fully connected layers, the flatten operation is used to convert the feature map into a one-dimensional vector. Then, fully connected layers are used along with Rectified Linear Unit (ReLU) activation to effectively simulate complicated features. In the output layer, the final fully connected layer is utilized with softmax activation for the purpose of classification, as depicted in Figure 1.

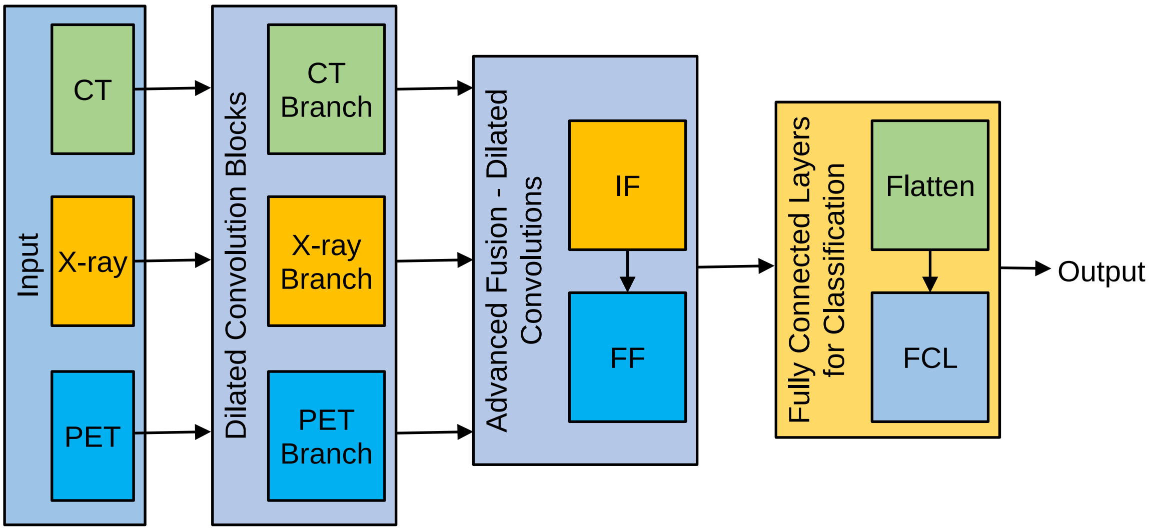
Input
CT
X-ray
PET
Dilated Convolution Blocks
CT Branch
X-ray Branch
PET Branch
Advanced Fusion - Dilated Convolutions
IF
FF
Fully Connected Layers for Classification
Flatten
FCL
Output

Figure 1. Proposed Framework

Dilated Convolutions in CNN Branches

Dilated convolutions are included as a special feature within these branches. Dilated convolutions, also known as extended convolutions, refer to a convolution operation in which the filter kernel includes intervals or gaps between its values. Unlike conventional convolutions, which move a filter at a constant step, dilated convolutions incorporate gaps or dilation rates within the filter. This allows the filter to capture a wider range of information from the input data.

The inclusion of dilated convolutions within the CNN branches serves a distinct and deliberate purpose. This facilitates the network's ability to extract features from the input data in a more comprehensive manner. The ability to capture subtle and scattered patterns across many imaging modalities is particularly beneficial in the field of lung disease diagnosis. By integrating dilated convolutions into the proposed CNN branches, the proposed method can significantly increase the receptive fields of these branches. This enhancement allows them to capture and incorporate information from a wider spatial context, as shown in Figure 2. In a regular 2D convolution, the output Y at a given spatial location (i,j) is computed according to Equation 1.

$$Y(i,j) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} X(i+m, j+n).K(m,n) \tag{1}$$

Where X is the input sensor, K is the convolutional kernel, and M and N are the dimensions of the kernel. For dilated convolutions, we introduce a dilation factor d, which introduces spacing between the values of the convolutional kernel, allowing it to capture a wide range of spatial information. The output Y_d for dilated convolutions is computed as given in Equation 2.

$$Y_d(i,j) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} X(i+d.m, j+d.n).K(m,n) \tag{2}$$

The dilated convolution operation is often expressed in terms of the original convolution operation with modified indices. The dilated convolution operator $*_d$ is defined as in Equation 3.

$$Y_d = X_{*d}K \tag{3}$$

This encapsulates the dilated convolution operation, where Y_d is the output, *X* is the input, and *K* is the kernel with the dilation factor incorporated. The procedure is explained as follows.

Procedure 1: Dilated Convolution

```
Input: Input tensor $X$ for a specific imaging modality; Convolution kernel $K$ with dilation factor $d$
Output: Output tensor $Y_d$ after dilated convolutions
Initialization:
    Initialize the output tensor $Y_d$ with zeros.
Traversal:
    Traverse the input tensor $X$ with a nested loop over spatial locations $(i,j)$.
Dilated Convolution Operation:
    For each spatial location $(i,j)$, compute the dilated convolution operation
    Update the corresponding entry in the output tensor $Yd$.
End //Continue traversal until all spatial locations have been processed.
output tensor $Yd$
End
```

Advanced Fusion using Dilated Convolutions

The study uses sophisticated fusion methods to effectively integrate data from three imaging modalities, namely CT, chest x-ray and PET scans. Modality integration is further refined by the use of dilated convolutions, which add sophistication and complexity to the fusion process. The use of dilated convolutions is very important for the fusion process because it adds a new way to capture long-range relationships and contextual information. Unlike traditional convolutional methods, dilated convolutions allow the model to expand its receptive field while minimizing the impact on the number of parameters. Diluted convolutions are very important for obtaining synergistic information from different types of data, which helps us to better understand the complex patterns in lung disease. Using dilated convolutions to incorporate contextual information improves diagnostic accuracy by considering long-range dependencies.

Intermediate Fusion with Dilated Convolutions

In intermediate fusion using dilated convolutions, let $F_1, F_2, F_3, \ldots, F_n$ represent the feature maps obtained from distinct modalities following independent processing through dilated convolutions. The intermediate fusion *IF* procedure is denoted as in Eq. 4.

$$IF(F_1, F_2, F_3, \ldots, F_n) = \sum_{i=1}^{n} F_{i*d} W_i \tag{4}$$

Where, W_i denotes the learnable weights for each modality.

Final Fusion with Dilated Convolutions

In order to get to the final fusion, the feature maps that come from the intermediate fusion (IF) are processed further using dilated convolutions. Let F_{int} represents the intermediate fused feature map. The ultimate fusion FF procedure is denoted as in Equation 5.

$$FF(F_{int}) = F_{int*d} W_f \tag{5}$$

Where, W_f represents the final set of learnable weights for fusion.

The pseudocode outlining the fusion process employing dilated convolutions is shown.

Pseudocode 2: Advanced Fusion with Dilated Convolutions

```
Input: Feature maps $F1,F2,…,Fn$ from different imaging modalities after individual processing through dilated convolutions.
Output: Final fused feature map $F_f$.
Initialize an $IS$ to store intermediate feature maps. //Intermediate Fusion
    For each modality $i$:
        $ISi = Fi *_d Wi$ //Apply dilated convolutions
        Add IS$i$ to ISIS.
    Obtain $F_{int}$. // Concatenate feature maps in ISIS on channel axis to obtain $F_{int}$
    $F_f = F_{int} *_d W_f$ //Apply dilated convolutions for final fusion // Final Fusion
    Further process the fused feature map $F_f$ through convolutional layers.
Final fused feature map $F_f$
//integrated information from modalities using fusion with dilated convolutions.
```

Additional CNN Layers and Pooling

In the proposed framework, additional CNN layers and pooling procedures are added after the initial processing of different imaging modalities using dilated convolutions and the advanced fusion stage. This step is critical for refining the learned representations, improving abstraction, and preparing the features for the final classification.

Following the sophisticated fusion process, the feature maps exhibit a comprehensive and intricate depiction of the amalgamated information derived from several modalities. By adding more CNN layers, it is possible to improve the extraction of hierarchical features and patterns. The presence of these layers enhances the model capacity to discover intricate linkages within the amalgamated elements.

Pooling operations, such as max pooling or average pooling, are utilized to reduce the spatial dimensions of the feature maps through down-sampling. Pooling is a technique that aids in the reduction of computing burden by prioritizing the most prominent features and facilitating translation invariance. Additionally, it facilitates the capture of the most informative elements of the combined features. The following section outlines the incorporation of extra CNN layers and pooling processes. Let F_f represent the fused feature map after the advanced fusion stage. The additional CNN layers is mathematically expressed as in Equation 6.

$$F_{CNN} = ReLU(W_1 * F_f + b_1) \tag{5}$$

Where, W_1 denotes the learnable weights of the convolutional layer, $b_1$is the bias term, * is the convolutional operator, ReLU (Rectified Linear Unit) activation function. The max pooling operation used is given in Equation 6.

$$F_P = MAXP(F_{CNN}, PS, S) \tag{6}$$

MAXP represents the max pooling function, PS is the pooling window size, and s denotes the stride of the pooling window.

Pseudocode 3: Additional CNN Layers and Pooling

```
Input: Fused feature map $F_f$ obtained from the advanced fusion stage.
Output: Refined and abstracted feature map $F_o$ ready for the final classification.
Initialize the feature map $F_{CNN}$ as $F_f$.
    For each additional CNN layer //Apply Additional CNN Layers
        $F_{CNN}$ = ReLU($Wi * F_{CNN} + bi$) // Apply a convolutional with $Wi$ and $bi$:
    Repeat //for the desired number of additional CNN layers.
    $F_P = MAXP(F_{CNN}, PS, S)$//Apply max pooling to downsample spatial dimensions
    Output Feature Map $F_o$ // is obtained after the application of additional CNN layers and pooling.
End
```

Fully Connected Layers for Classification

Finally, the fully connected layers with a Softmax function is used to receive the predicted class probabilities. The pseudocode is given as follows

Pseudocode: Fully Connected Layers for Classification

```
Input: Flattened feature vector $F_f$ obtained after CNN layers and pooling operations
Output: Predicted class probabilities.
Initialize $F_{fc}$ as $F_f$.
    For each fully connected layer:
        $F_{fc}$=ReLU($W_{fc}\cdot F_{fc}+b_{fc}$)
        Repeat //for the desired number of fully connected layers.
    End
    Apply final FCL // Softmax activation at Output Layer
End // Prediction output represents the predicted class probabilities.
```

RESULTS AND DISCUSSION

The proposed method is experimented and validated on the combined dataset, which is obtained from RIDER Lung CT - The Cancer Imaging Archive (TCIA) Public Access - Cancer Imaging Archive Wiki [14], Chest X-Ray Images - Pneumonia - kaggle.com [15] and 4D MRI dataset [16]. The experiments are performed in python tool on an i7 processor with 16 GB of RAM and 16 GB of GPU. K-fold cross validation with different values are used to validate the performance of the model as in Table 1. Accuracy, precision, recall, and F-measure was used to evaluate the performance of the proposed model.

Table 1. Results on different K-Fold cross validation

Metric	3k Fold	5k Fold	10k Fold	15k Fold
Accuracy	0.82	0.88	0.94	0.92
Precision	0.85	0.92	0.96	0.94
Recall	0.82	0.84	0.91	0.89
F-measure	0.83	0.88	0.92	0.899

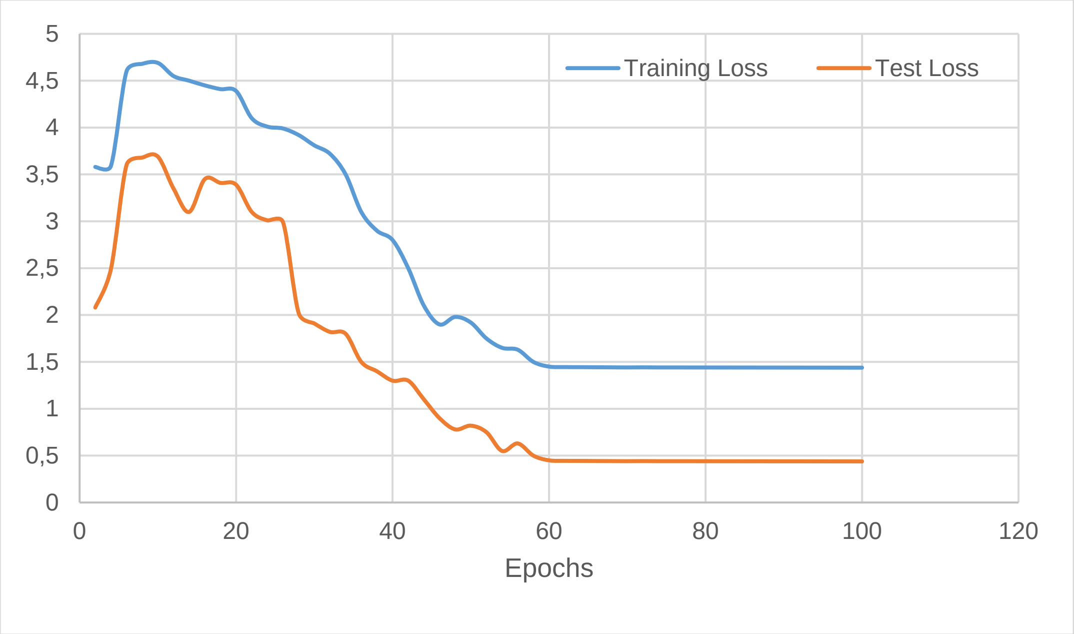

Figure 2. Training and test loss

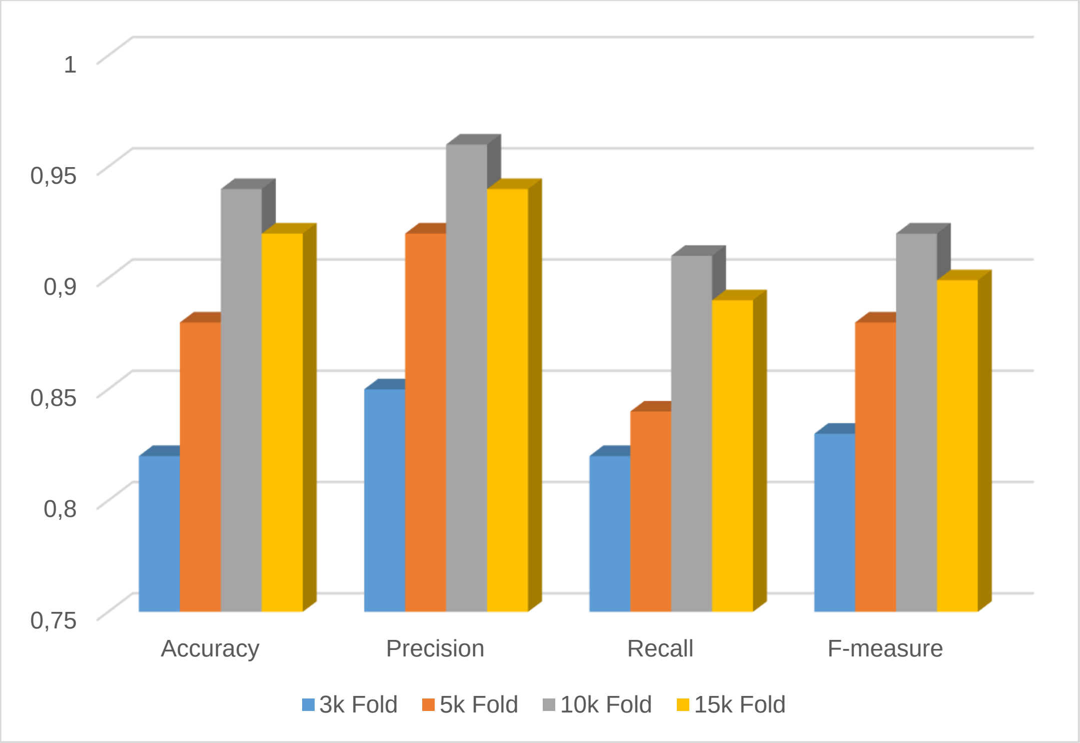

Figure 3. Performance evaluation of different folds

Multimodal deep learning in precise medical diagnosis is very important. However, the important criteria is the preparation of multimodal datasets and selection of best features that takes part in the diagnosis process. As, it is our first effort to combine the data from three different sources; there are no similar studies to be compared with our model. In order to check and validate the performance of our model, we tried different k fold cross validations. The value for k was considered as 3, 5, 10, and 15 for the evaluation. As shown in Table 1 and Figure 3, the performance of the model was better for k=10. Then, we trained our model for 10-fold cross validation and it achieved an accuracy of 94%. The training and test loss during different epochs are shown in Figure 2. The cross entropy is considered as loss function to validate the models performance during training and test phase. The formulas for loss function is given in Equation 7.

$$Loss = -\frac{1}{N}\sum y_i \log\big(p(P_i)\big) + (1 - P_i)\log\,(1 - p(P_i)) \tag{7}$$

Where, N represents total number of images corresponding to patients, P_i represents if the image contains lung disease (1) or not (0). $p(P_i)$ is the Softmax probability of each class in binary classification.

Table 2. Performance comparison on different modalities

Modality	Accuracy	Precision	Recall	F-measure
CT Imaging	0.91	0.92	0.89	0.90
X-ray Imaging	0.89	0.92	0.90	0.91
PET	0.88	0.90	0.89	0.89
Multimodal (CT+PET+X-ray)	0.94	0.96	0.91	0.92

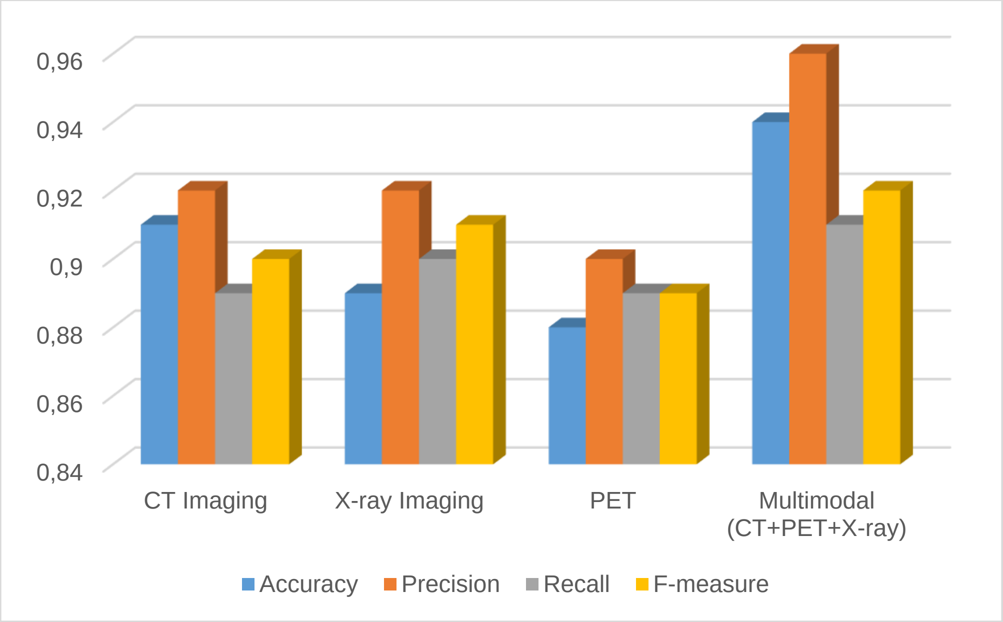

Figure 4. Performance comparison on different modalities

Further, we compared the performance of our model with 10-fold cross validation on multimodal dataset and on individual images from each modality as shown in Table 2 and Figure 4. It is clear that our model performed better with multimodal data and achieved better accuracy than on individual modalities.

CONCLUSION

The article presents a multimodal approach to handle data with three different modalities CT, PET, and X-ray. The data from different sources was first combined and then the modal was trained with intermediate fusion approach. Different K-fold cross validation was used and model performed well with 10 (K) fold cross validation. Further, the model was compared with multimodal and single modality data, and the results show that model is able to capture the features from different modalities to provide more accurate diagnosis than single modality data. The future experiments will consider testing the performance of the model on other multimodal data. In addition, other modalities such as clinical data can be incorporated and new modification to the model is planned in the future scope of the study.

Funding: The study is supported by Russian science foundation regional grant number 23-21-10009.
Acknowledgments: We are thankful to our students in assisting in data preparation.
Conflicts of Interest: The authors declare no conflict of interest.
Availability of Materials and Data: Data were obtained from public sources and then combined to produce the multimodal data, which are available on request from the first author.

REFERENCES

1. Vasava RP, Joshiara HA. Different Respiratory Lung Sounds Prediction using Deep Learning. In: 4th International Conference on Electronics and Sustainable Communication Systems (ICESC). 2023 Jul 06-08; Coimbatore, India: IEEE. 2023. p. 1626-1630
2. Zhang L, Che Z, Li Y, Mu M, Gang J, Xiao Y, et al. Multilevel classification of knee cartilage lesion in multimodal MRI based on deep learning. Biomed Sig Proc Cont. 2023 May; 48 (5):104687.
3. Sreeja KA, Arshey M, Warrier GS, Pradeep A. Harmonizing the power of deep learning and traditional radiology for the advancement of lung disease diagnosis through chest x-ray image classification. Lat Am. J. Pharm. 2023; 42 (5):522-34.
4. Gaur L, Bhatia U, Jhanjhi NZ, Muhammad G, Masud M. Medical image-based detection of covid-19 using deep convolution neural networks. Multimedia syst. 2023; 29 (3):1729-38.
5. Gheibi Y, Shirini K, Razavi SN, Farhoudi M, Samad-Soltani T. CNN-Res: deep learning framework for segmentation of acute ischemic stroke lesions on multimodal MRI images. BMC Med Inform Dec Mak. 2023; 23 (1):1-14.
6. Mukhi SE, Varshini RT, Sherley SEF. Diagnosis of COVID-19 from multimodal imaging data using optimized deep learning techniques. SN Comp Sci. 2023;4(3):212.

7. Wang R, Chen LC, Moukheiber L, Seastedt K, Moukheiber M, Moukheiber D, et al. Enabling chronic obstructive pulmonary disease diagnosis through chest X-rays: A multi-site and multi-modality study. Intern J Med Infor. 2023 Oct; 178 (10):105211.
8. Priya SU, Tarun SG, Shamitha S, Rao AS, Prasad VB. Multimodal smart diagnosis of pulmonary diseases. In 2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT). 2023 Jun 05-06; Gharuan, India: IEEE. 2023. p. 33-40.
9. Dubey P, Tripathi P. Theoretical evaluation of transfer learning approaches for the identification of lung respiratory sounds. In 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC). 2023 Jul 06-08. Coimbatore, India: IEEE. 2023. p.1042-1047.
10. Li X, Qi B, Wan X, Zhang Z, Yang W, Xiao Y, *et al.* Electret-based flexible pressure sensor for respiratory diseases auxiliary diagnosis system using machine learning technique. Nano Energy. 2023 Sep; 114(9):108652.
11. Thanoon MA, Zulkifley MA, Mohd Zainuri MA, Abdani SR. A Review of Deep Learning Techniques for Lung Cancer Screening and Diagnosis Based on CT Images. Diagnostics. 2023;13(16):2617.
12. Sun H, Jiao J, Ren Y, Guo Y, Wang Y. Multimodal fusion model for classifying placenta ultrasound imaging in pregnancies with hypertension disorders. Pregnancy Hypertension. 2023;31,46-53.
13. Kumar S, Ivanova O, Melyokhin A, Tiwari P. Deep-learning-enabled multimodal data fusion for lung disease classification. Inform Medic Unlock. 2023 Oct; 42(1): 101367.
14. RIDER Lung CT - The Cancer Imaging Archive (TCIA) Public Access - Cancer Imaging Archive Wiki (https://wiki.cancerimagingarchive.net/display/public/rider+lung+ct). Accessed on 15 Jun 2023.
15. Chest X-Ray Images - Pneumonia - kaggle.com (https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia). Accessed on 15 Jun 2023.
16. 4D MRI dataset (Datasets – Biomedical Image Computing | ETH Zurich). Accessed on 17 Jun 2023.