

A Study on Heart Segmentation Using Deep Learning Algorithm for MRI Scans

Shakeel Muhammad Ibrahim^{*}, Muhammad Sohail Ibrahim[†], Muhammad Usman[‡],
Imran Naseem[§] and Muhammad Moinuddin[¶]

^{*}Department of Computer Science, University of Karachi, Karachi-75270, Pakistan.

Email: shak.ibrahm.97@gmail.com

[†]College of Electrical Engineering, Zhejiang University, Hangzhou, Zhejiang, China 310027.

Email: msosail@zju.edu.cn

[‡]Department of Computer Engineering, Chosun University, Gwangju, Republic of Korea.

Email: usman@chosun.kr

[§]College of Engineering, Karachi Institute of Economics and Technology, Karachi 75190, Pakistan.

Email: imrannaseem@pafkiet.edu.pk

[¶]CEIES, King Abdulaziz University, Saudi Arabia.

Email: mmsansari@kau.edu.sa

Abstract—Among all body organs heart is a one of the most vital of organs of human body. Dysfunction of heart function even for a couple of moments can be fatal, therefore, efficient monitoring of its physiological state is essential for the patients with cardiovascular diseases. In the recent past, various computer assisted medical imaging systems have been proposed for the segmentation of the organ of interest. However, for the segmentation of heart using MRI, only few methods have been proposed each with its own merits and demerits. For further advancement in this area of research, we analyze automated heart segmentation methods for magnetic resonance images. The analysis are based on deep learning methods that processes a full MR scan in a slice by slice fashion to predict desired mask for heart region. We design two encoder-decoder type fully convolutional neural network models (1) Multi-Channel input scheme (also known as 2.5D method), (2) a single channel input scheme with relatively large size network. Both models are evaluated on real MRI dataset and their performances are analysed for different test samples on standard measures such as Jaccard score, Youden's index and Dice score etc. Python implementation of our code is made publicly available at <https://github.com/Shak97/iceest2019> for performance evaluation.

Index Terms—Medical Imaging, Segmentation, Cardiovascular Imaging, Machine learning, Neural Networks, Deep Learning, Fully Convolution Neural Network (FCNN), Magnetic Resonance Imaging (MRI).

I. INTRODUCTION

Heart is one of the vital organs of human body. It is the most hardworking organ that keeps circulating the blood in the body, thus playing indispensable role in maintaining the energy levels of the body. A minor dysfunction of heart even for a short time interval can be fatal, therefore efficient monitoring of its physiological state is essential for patients with cardiovascular diseases (CVD). To accomplish this important task, variety of medical imaging modalities have been introduced in the last few decades such as ultrasound, X-Ray computed tomography (CT), and magnetic resonance (MR) imaging etc.

The Ultrasound (US) is a real-time, radiation free and non-invasive and the safest imaging modality. However, it has some limitations such as, low contrast, perspective visualization (2D projection of 3D object), low penetration, and small field of view etc [1]. On the other hand, CT provides 3D tomography with high quality speckle free images that have desired contrast information which can help in better diagnosis. But similar to X-Ray, CT imaging is also a radiation based imaging therefore it is used as an offline imaging system that suits best in emergency situations. Due to its ionization characteristics, it is unsuitable for the patients who are at high risk e.g., pregnant women or young children etc [2].

MR imaging (MRI) is the second most safest imaging modality. Unlike CT which works on the principle of radiation, in MRI, the body is magnetized by powerful magnets and the spin of the protons in hydrogen or other elements of choice (depending on the application of MRI) are synchronized with the resonant frequency. The change in the resonance of the spin is measured in Fourier domain and with the help of inverse Fourier transform high quality MR images are produced. To scan the region of interest (ROI), variety of scanning schemes are used which affect the quality, interpretation of pixel intensity, and acquisition time. By properly manipulating the control parameters, the desired quality of MRI can be obtained. A drawback of obtaining MRI is its scan time, which is several times higher than the CT or US imaging. However, for offline applications where highest contrast and superior resolution is required without the risk of radiation exposure, MRI is considered as a gold standard option. One such application is the detailed diagnosis of heart for its physical structural and metabolic functional analysis [3].

With the advent of modern medical facilities it has now become an standard practice to store and analyse biomedical data to further improve the diagnostic processes. The type of data ranges from molecular level to the 3D MRI scan of fully

human body. Substantially, large medical imaging datasets are being introduced and the demand for automated medical assistance system is increasing subsequently. A thorough research in the domain of biomedical engineering has been observed in the recent years. The medical data is mostly heterogeneous as it is obtained after a variety of different clinical analysis procedures and the wide range of medical imaging modalities. Manual processing of such data is extremely difficult task which costs both the time and money to analyze the underlying information. In order to effectively utilize the acquired data, machine learning techniques have been effectively used for a wide number of biomedical applications such as protein prediction [4], [5], [6] and cancer classification [7], [8].

Various computer assisted medical imaging systems have been proposed for variety of biomedical applications such as classification of tumor, segmentation of organ of interest, and registration of multi-modality images etc. For computer assisted surgeries and diagnosis systems, the detection and identification of ROIs is one of many challenging tasks. The age, size and gender of the humans have direct impact on the shapes and sizes of the internal organs. Therefore a great degree of variation is taken into account. The complications in identifying the upper and lower part of the heart can put even an experienced radiologist in a great inconvenience.

Supplementary hitches are added while seeking computer based assistance. To design a computer based medical imaging system, there is a need of high quality images in which ROIs can be easily distinguished e.g., in MRI. To the best of our knowledge only a few methods exist for the segmentation of heart [9], [10], [11]. Previously the segmentation task was performed using the conventional methods such as edge detection [9]. 2D and 3D CNN were investigated for the tasks of heart segmentation in [10]. The use of few parameters were explored and it was revealed that the use of an optimized 2D approach yields better results than that of the 3D counterpart. More recently, in the authors proposed a shape reconstruction neural network (SCNN) and spatial constraint network (SCN) to achieve two different tasks. Firstly, SCNN was developed to maintain the shape of the heart in the segmentation results, secondly SCN was utilized to solve the problem of large variation in the 2D slices. In [12] cardiopath classification of heart disease was performed. The ROIs were generated using the you only look once (YOLO) based object detection method. The proposed approach produced adequate accuracy results compared to the previous methods. A cluster based approach named vantage point is proposed in [13]. Complete feature vector was used and representation pattern in higher dimensions feature space were found.

For further advancement in this area of research, in this paper, we implemented an automated heart segmentation method for MR images. The proposed approach is a deep learning based method that processes a full MR scan in a slice by slice fashion to predict desired mask for heart region.

The contributions of this research are as follows:

- 1) A fully automated computer assisted method is proposed for the segmentation of heart region from MRI.

- 2) Two U-Net [14] like fully convolutional network models are designed for the said task (i) multi-channel input scheme (also known as 2.5D method), (ii) a single channel input scheme with relatively large size network.
- 3) A comparative analysis for both schemes are presented exhibiting their performances.
- 4) The proposed models are evaluated on real MRI dataset and their performance is analyzed for different standard measures such as Jaccard score and Dice score etc.
- 5) Python implementation of our method is made public for the performance evaluation.

We organized this paper in following sections: in Section II, we discuss the details of our proposed method followed by the experimental results and discussion in Section III and the conclusion of this study is presented in section IV.

II. METHOD

The method is based on fully convolutional neural network, which is trained using the concept of data driven supervised learning, where representative examples are used to fine tune the learning parameter of the model. Typically medical imaging datasets are huge in size and have highly imbalanced distributions which make it difficult to design machine learning based model using conventional approaches. In this regard, the deep learning has proven to show great success [15]. MR scan produce 3D tomographic view of the body which is computationally very expensive to process with deep learning algorithms [10]. Therefore, to perform the heart segmentation from MRI scans, we divide 3D cubes into multiple 2D images. For these 2D image planes, we trained a convolutional neural network that generates desired masks for each input image. Details of dataset, network configuration and training strategies are discussed in subsequent sections.

A. Dataset

The dataset is downloaded from the medical segmentation decathlon (MSD) challenge dataset from (<https://decathlon-10.grand-challenge.org/home/>). The MSD challenge dataset is specially designed for generalized segmentation task and consists of 3D MR scans of 10 different body parts including liver, pancreas and heart. For this study, we used 'Task2' dataset which is comprised of 20 heart MRI scans. Out of these 20 scans, 14 and 2 scans are randomly selected for the training and validation of the model respectively, while remaining 4 scans were used as test scans. Each 3D scan is converted into a set of images, having dimensions of $320 \times 320 \times 1$. In particular there are 1578, 218, and 455 pairs of input/label images in training, validation, and test datasets respectively.

B. Network Specification and Training

We design 5 different models which can be classified into two groups (1) thin networks with 1, 3, 5, and 7 input channels. (2) thick network with only 1 input channel. All other hyper-parameters in both groups are same. Fig.1 shows the generalized block diagram of models. The variable parameters are defined by number of input channels (CH) and number of

filters (NF). As shown in Fig.1, the proposed design consists of two types of modules **Mod_1** and **Mod_2**, the configuration of which are depicted in Fig.2. In particular, **Mod_1** consists of two convolution layers, two batch normalization layers and the **K** number of filters, which for instance in *C3* equals to $2 \times \text{NF}$, whereas, **Mod_2** contains an additional concatenation layer along with the two convolutional and batch normalization layers. As stated earlier the configuration of thin networks is $\text{NF} = 16$ and $\text{CH} = \{1, 3, 5, 7\}$ and for thick network $\text{NF} = 16$ and $\text{CH} = 1$ is chosen. In all models, ReLU activation function is used in every layer except the output layers where sigmoid function is used as an activation function. The Dropout and Max-pooling layers are also used as specified in Fig.1.

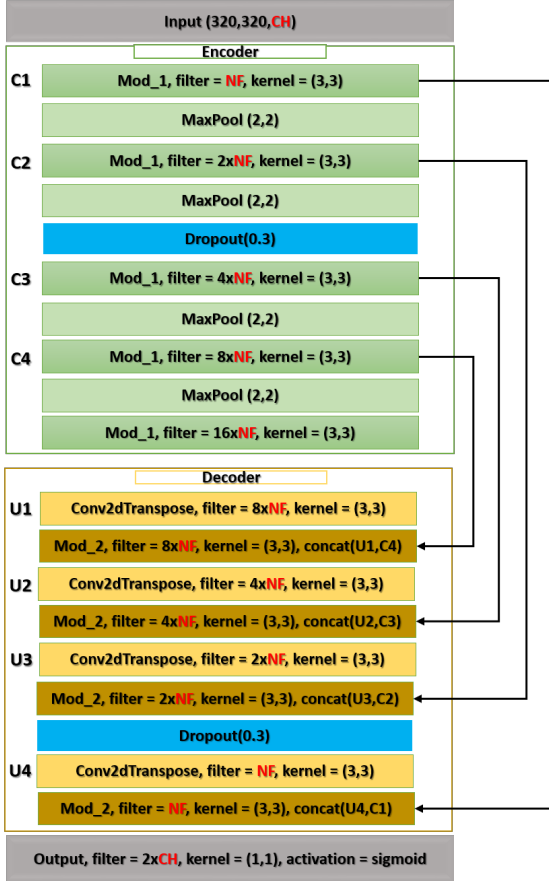


Fig. 1: Architecture of our convolutional neural network model.

The model is implemented in Python TensorFlow using Keras library. To train a neural network, variety of algorithms have been proposed [16], [17], [18]. Each has its own merits and demerits. However in general, the Adam optimizer is considered as the best choice for variety of problems. Therefore, to train our model we opt to use Adam optimizer, keeping the default value of all parameters. In the proposed method, we minimized the categorical cross entropy loss function \mathcal{L} for 10 epochs. The categorical cross entropy loss function is defined

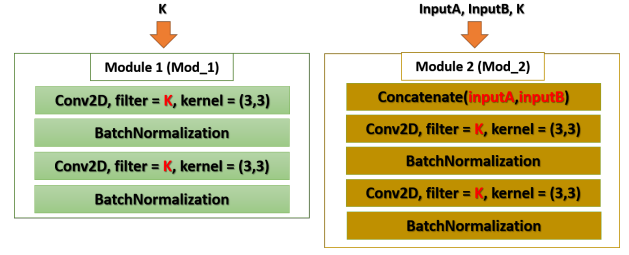


Fig. 2: Schematic diagram of individual modules of Fig. 1

as

$$\mathcal{L}(y, \hat{y}) = - \sum_{r=0}^R \sum_{q=0}^Q (y_{q,r} \log(\hat{y}_{q,r})), \quad (1)$$

where y is the original mask and \hat{y} is the predicted mask in vector form, while Q is the number of training samples and R is the number of categories which is 2 in this study i.e., (0: background pixel and 1: segment pixel). For each epoch, we monitor the training and validation losses along with the Dice scores (defined in (5)).

C. Performance Metrics

We used following metrics to analyze the performance of the CNN models presented in this study.

- 1) True positive rate (TPR), recall, or sensitivity

$$\text{TPR} = \frac{TP}{TP + FN} \quad (2)$$

- 2) False positive rate (FPR), fallout, or (1- specificity)

$$\text{FPR} = \frac{FP}{FP + TN} \quad (3)$$

- 3) Positive predicted value (PPV), or precision

$$\text{PPV} = \frac{TP}{TP + FP} \quad (4)$$

- 4) F1 / Dice score

$$\text{Dice} = \frac{2TP}{2TP + FP + FN} \quad (5)$$

- 5) Jaccard index (JI) or Jaccard similarity coefficient

$$\text{JI} = \frac{TP}{TP + FP + FN} \quad (6)$$

- 6) Youden index (YI) or informedness

$$\text{YI} = \text{TPR} - \text{FPR} \quad (7)$$

where TP, FP, TN, and FN are the number of true positive, false positive, true negative, and false negative predictions of the number of pixels respectively.

III. EXPERIMENTAL RESULTS AND DISCUSSION

In this study we implemented five different models which are categorized into two groups (1) thin networks with $NF = 16$ and $CH = \{1, 3, 5, 7\}$ and (2) thick network with $NF = 64$ and $CH = 1$. For all models, same datasets and hyper-parameters are used. The thin models with $CH = \{1, 3, 5, 7\}$ achieved dice score of 0.65, 0.73, 0.75, and 0.77 respectively, which indicate that the model with same number of learning parameters (NF), performs better if adequate neighbourhood information is provided. Although, the computational cost of the prediction is defined by the NF and number of layers, however for each image we need to process CH number of images simultaneously. Which means the memory requirement for each training epochs will increase. Another method to exploit the memory is to increase the size of network, which eventually increase the computational complexity as well.

To analyse the effect of model complexity we design a relatively thick model with $NF = 64$ and $CH = 1$. This model presented significant performance improvement compared to thin models. For better understanding of the performance we calculated variance statistics of our predicted masks and also analyze the visual difference in ground truth and prediction of the thick network. Fig. 3 shows the dice coefficient and loss curves of 10 epochs for training and validation phases. Fig. 3(a) shows the dice coefficient verses numbers of training epochs. A significant increase in the dice similarity is observed with subsequent increase in the number of epochs. The behavior of loss function with respect to the increase in epochs for the training phase is shown in Fig. 3(b). The dice coefficient and loss function behavior of validation phase is depicted in Fig. 3(c) and (d) respectively. Since no further increment in the validation performance is observed, therefore we stop the training at 10 epochs.

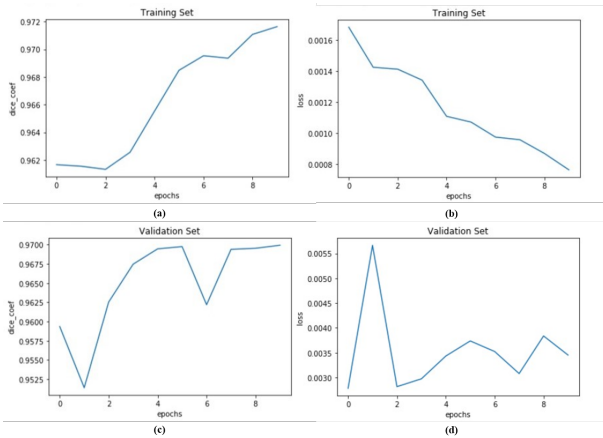


Fig. 3: (a) Dice coefficient curves for training dataset, (b) Cross entropy loss curve for training dataset, (c) Dice coefficient curves for validation dataset, (d) Cross entropy loss curve for validation dataset.

The visual examples of prediction are shown in Fig. 4. It is evident from the figure that the ground truth has been ac-

TABLE I: Performance statistics of model#5

Performance metrics	
Recall	0.7697 ± 0.3519
Fallout	0.0007 ± 0.0012
Precision	0.8471 ± 0.1602
Dice score	0.8216 ± 0.2067
Jaccard index	0.7361 ± 0.2260
Youden's index	0.7685 ± 0.3513

curately predicted achieving a significant dice score of 0.973, 0.953, and 0.924, for the examples 1, 2, and 3 respectively.

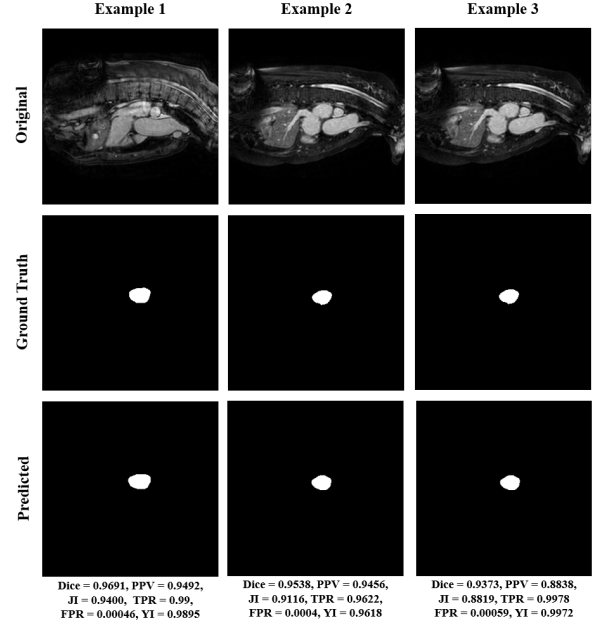


Fig. 4: Three good example results of heart segmentation using single channel CNN model

For the sake of comparison, some worst case scenarios in terms of prediction are also shown. These cases may occur due to various reasons such as over-fitting, inappropriate selection of input channels/filters, etc. Such schemes result in very low dice scores as shown in Fig. 5.

This model yielded a dice score of 0.8216 with a standard deviation of ± 0.2067 . The details statistical analysis is summarized in table I

This clearly corroborates the proposition that a simple encoder-decoder styled CNN model with skip-connections can effectively predict the segmentation of heart region from MRI scans. It is noteworthy to point out that the the performance in large region's segmentation is very good however, for small sized segment the performance is very low which is also evident in inferior dice scores.

IV. CONCLUSION

Herein, we demonstrated the application of deep learning for the segmentation of heart region from MRI scans. We design a fully convolution neural network based system which

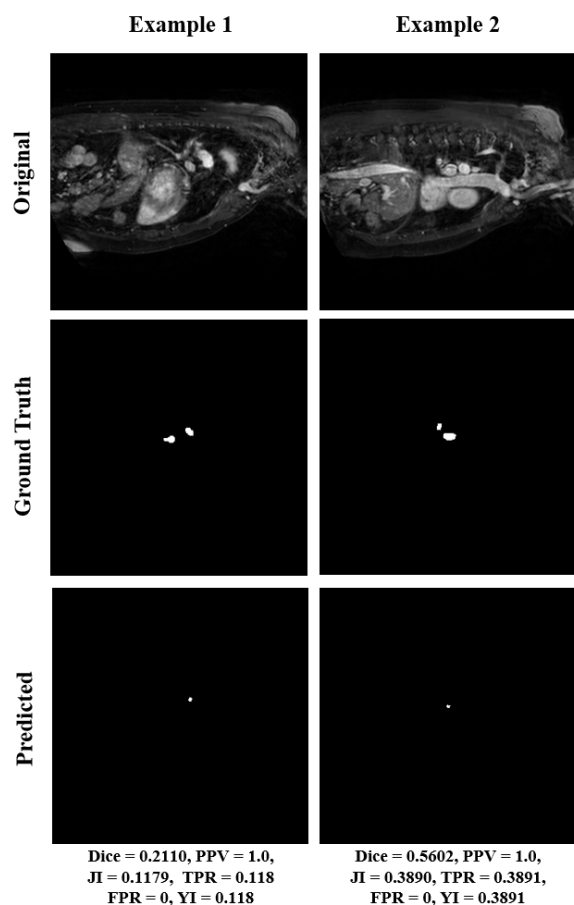


Fig. 5: Example results when single channel CNN model failed to predict correct mask.

is trained using the concept of data driven supervised learning, where representative examples are used to fine tune the learning parameter of the model. Two approaches presented in this study revealed that a significant improvement in the performance can be achieved by either processing multiple images simultaneously through multiple input channels or by increasing the number of filters in each convolution layer. Increasing the input dimensions or the network parameters both contribute to the constructive performances. It can therefore be concluded that a simple encoder-decoder styled CNN model with skip-connections can effectively predict the segmentation of heart region from MRI scans, specially in the images with large segmentation region. However, for small segments the performance is not satisfactory, and therefore, the efficacy of the advanced CNN architectures and loss functions that can improve the accuracy for small size segments can be explored in future studies.

REFERENCES

- [1] Y. H. Yoon, S. Khan, J. Huh, J. C. Ye *et al.*, "Efficient b-mode ultrasound image reconstruction from sub-sampled rf data using deep learning," *IEEE transactions on medical imaging*, 2018.
- [2] E. Robbins, "Radiation risks from imaging studies in children with cancer," *Pediatric blood & cancer*, vol. 51, no. 4, pp. 453–457, 2008.
- [3] L. Wang, H. Nie, Q. Wang, G. Zhang, G. Li, L. Bai, T. Hua, and S. Wei, "Use of magnetic resonance imaging combined with gene analysis for the diagnosis of fetal congenital heart disease," *BMC medical imaging*, vol. 19, no. 1, p. 12, 2019.
- [4] I. Naseem, S. Khan, R. Togneri, and M. Bennamoun, "Ecmsrc: A sparse learning approach for the prediction of extracellular matrix proteins," *Current Bioinformatics*, vol. 12, no. 4, pp. 361–368, 2017.
- [5] S. Khan, I. Naseem, R. Togneri, and M. Bennamoun, "Rafp-pred: Robust prediction of antifreeze proteins using localized analysis of n-peptide compositions," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. PP, no. 99, pp. 1–1, 2017.
- [6] M. Usman and J. A. Lee, "Afp-cksaap: Prediction of antifreeze proteins using composition of k-spaced amino acid pairs with deep neural network," *arXiv preprint arXiv:1910.06392*, 2019.
- [7] S. Khan, I. Naseem, R. Togneri, and M. Bennamoun, "A novel adaptive kernel for the rbf neural networks," *Circuits, Systems, and Signal Processing*, pp. 1–15, 2016.
- [8] S. Khan, J. Ahmad, I. Naseem, and M. Moinuddin, "A novel fractional gradient-based learning algorithm for recurrent neural networks," *Circuits, Systems, and Signal Processing*, pp. 1–20, 2017.
- [9] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1395–1403.
- [10] C. F. Baumgartner, L. M. Koch, M. Pollefeys, and E. Konukoglu, "An exploration of 2d and 3d deep learning techniques for cardiac mr image segmentation," in *International Workshop on Statistical Atlases and Computational Models of the Heart*. Springer, 2017, pp. 111–119.
- [11] O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, J. Caballero, S. A. Cook, A. De Marvao, T. Dawes, D. P. O'Regan *et al.*, "Anatomically constrained neural networks (acnns): application to cardiac image enhancement and segmentation," *IEEE transactions on medical imaging*, vol. 37, no. 2, pp. 384–395, 2017.
- [12] Y. Chang, B. Song, C. Jung, and L. Huang, "Automatic segmentation and cardiopathy classification in cardiac mri images based on deep neural networks," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 1020–1024.
- [13] M. P. Heinrich and M. Blendowski, "Multi-organ segmentation using vantage point forests and binary context features," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 598–606.
- [14] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [15] A. Kumar, J. Kim, D. Lyndon, M. Fulham, and D. Feng, "An ensemble of fine-tuned convolutional neural networks for medical image classification," *IEEE journal of biomedical and health informatics*, vol. 21, no. 1, pp. 31–40, 2016.
- [16] S. Khan, I. Naseem, M. A. Malik, R. Togneri, and M. Bennamoun, "A fractional gradient descent-based rbf neural network," *Circuits, Systems, and Signal Processing*, pp. 1–22, 2018.
- [17] O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, J. Caballero, S. A. Cook, A. De Marvao, T. Dawes, D. P. O'Regan *et al.*, "Anatomically constrained neural networks (acnns): application to cardiac image enhancement and segmentation," *IEEE transactions on medical imaging*, vol. 37, no. 2, pp. 384–395, 2017.
- [18] A. Kumar, J. Kim, D. Lyndon, M. Fulham, and D. Feng, "An ensemble of fine-tuned convolutional neural networks for medical image classification," *IEEE journal of biomedical and health informatics*, vol. 21, no. 1, pp. 31–40, 2016.