

A Feasibility study for Deep Learning Based Automated Brain Tumor Segmentation using Magnetic Resonance Images

Shanaka Ramesh Gunasekara¹, HNTK Kaldera², Maheshi B. Dissanayake³

Dept. Of Electrical and Electronic Engineering, Faculty of Engineering,

University of Peradeniya, Sri Lanka.

¹ shanakag@eng.pdn.ac.lk, ² nikalal.k@eng.pdn.ac.lk, ³ maheshid@eng.pdn.ac.lk

Abstract: Deep learning algorithms have accounted for the rapid acceleration of artificial intelligence research in medical image analysis, interpretation, and segmentation with many potential applications across various sub-disciplines in medicine. However, only a limited number of research investigating these application scenarios is deployed into the clinical sector to evaluate the real requirement and the practical challenges of the model deployment. In this research, a deep convolutional neural network (CNN) based classification network and Faster RCNN based localization network were developed for brain tumour Magnetic resonance image (MRI) classification and tumour localization. A typical edge detection algorithm called Prewitt was used for tumour segmentation task, based on the tumour localization output. Overall performance of the proposed tumour segmentation architecture, was analyzed using objective quality parameters including Accuracy, Boundary Displacement Error (BDE), Dice score and confidence interval. A subjective quality assessment of the model was conducted based on the Double Stimulus Impairment Scale (DSIS) protocol using medical expertise input. It was observed that our segmented output's confidence level was in a similar range to that of experts.

Index Terms: CNN, Deep learning, Faster RCNN, Feasibility study, Magnetic resonance images, subjective analysis.

1. INTRODUCTION

Magnetic resonance imaging (MRI) is one of the most common types of imaging technique in radiology used to diagnose blood vessel damages, brain injuries and tumors, strokes, eye problems, bone infections, heart disease, and many more would otherwise require invasive procedures. MRIs are widely used for Brain tumour diagnosis. The Brain tumor is an abnormal growth in brain cells that leads to considerable morbidity or mortality. There are more than 120 tumor types, and all of those belong to two main tumour groups; primary brain tumors and secondary brain tumors. Typically, at the initial stage of diagnosing tumors, non-invasive medical imaging techniques such as MRI and Computer Tomography (CT) are widely used by the physicians [1].

During the past ten years, with the rapid improvement of artificial intelligence (AI) and deep learning algorithm, image classification and segmentation architectures and their applications have exhibited incredible performance growth. The improvements in these technologies directly affect the medical imaging sector [2]. Many researchers focused on developing more advanced, accurate algorithms useful and powerful in medical image analysis [3, 4]. As a result, computerized systems were able to interpret the medical imaging outputs such as MRI or CT scans with high efficiency and accuracy while getting closer to a medical practitioner's diagnosis. Therefore, AI-radiologists can provide substantial benefits in many clinical settings, such as prioritizing workflow, supporting clinical decision-making, large-scale screening initiatives and many more [5, 6].

In the recent past, there was a rapid acceleration in academic research in the health care sector investigating the AI's capabilities, specifically Deep learning, in medical image diagnosis. All of these algorithms were developed to match the performance of medical experts in a wide range of image analysis tasks including, skin cancer diagnosis and classification, retinopathy detection for diabetics [7], automated diagnosis from chest imaging for lung module detection [8], pulmonary tuberculosis classification [9] and chest radiography

diagnosis [10,11], breast cancer detection and diagnosis with mammography [12,13]. But only a few of these researches, present feedback of the design from the medical experts' point of view and feasibility study of the implementation of these developed algorithms in real-world applications[14][15]. For instance, [14], explore the main limitations and challenges of the AI models developed for the medical sector when translating them into clinical practice.

The majority of the machine learning studies have been retrospective; they used historically labeled data in both training and testing. Therefore when it comes to the real-world scenario, the AI systems' performance might be worse as the real-world data differ from the training data set [14]. Hence, it is necessary to carry out prospective studies with the real field data before the implementation. Although many objective quality metrics imply the performance of the developed AI algorithms, none of these reflects the clinical applicability of the algorithm. Researchers need to emphasize how the proposed AI and deep learning models may improve patient care with reliable facts [16].

One potential approach is to compare the results subjectively with medical experts' opinions and obtain the algorithm's subjective quality analysis concerning the clinical needs.

This research article presents a subjective quality analysis of a proposed deep learning algorithm for brain tumour classification and segmentation. Section 2 summarizes the theoretical background and overview of the existing literature related to the design problem. Section 3 presents the methodology embraced with data preparation, tumour classification and localization, and finally segmentation. The results and findings of the subjective performance evaluation are presented and discussed in section 4. Finally, section 5 concludes the paper.

2. BACKGROUND WORK

2.1 Initiation towards CNN

A network architecture which consists of convolutional, activation, max pooling and fully connected layers is defined as a Convolutional neural network(CNN). CNNs are powerful tools used in a wide range of applications, such as classification of images, segmentation of images, etc. In CNN, a filter or kernel slides over the input image and generates a feature map. Through different kernel filters, the network can be customized to generate numerous feature maps and increase the ability to learn more correlation between neighboring pixels of the image. But increasing the number of filters may result in huge computational cost and take a long time to train the model. Furthermore, it can be more effective to look at smaller portions of the image for the interesting objects, rather than looking at an entire image at once to find specific features [17, 18, 19].

A CNN architecture blended with a region proposal network (RPN) is called Faster R-CNN. R-CNN is a classifier that classifies regions in an image as belonging to the object classes or not, while RPN creates regions of interest, called as anchors, for the input image and then predicts the probability of an anchor being in the background or foreground [20, 21]. Generally, there are nine standard anchors for the RPN network, with different dimensions' as illustrated in Fig. 1. In the RPN, it examines every anchor box to predict whether it contains an object or not. Furthermore, if there is more than one object in the input image, RPN will output several anchor boxes with each object's relevant locations. Later, RPN uses Region of Interest (ROI) pooling to identify the correct features belongs to a particular object. Afterwards, R-CNN uses the proposed anchors from

the RPN and classifies the object inside the bounding box to the existing classes. Also, in this stage, the fine network tune the bounding box coordinates to fit in the ROI.

2.2 Subjective quality analysis for medical images

Medical image analysis is only a tool for efficient diagnosis. The image analysis and the output should meet the physician's requirement and decision criteria, i.e the experts. Although most of these objective quality analysis metrics are developed based on the human vision system (HVS), it is impossible to model all the possible image distortions and levels that can identify with HVS along with human perception[22]. Hence, objective quality measures, would not justify the accuracy of the model, despite the promising numbers, specifically in medical imaging. To overcome the diagnostic deficiencies of objective quality assessment in medical imaging, subjective quality assessment, where experts in the medical field give their best judgment on the perceptual quality and the diagnostic information of the output, is widely recommended.

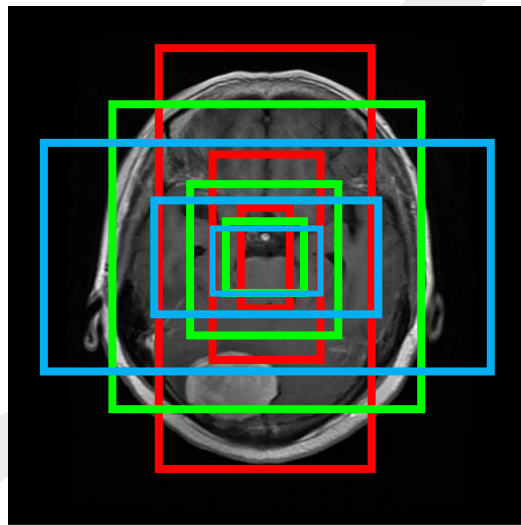


Fig. 1. Anchors used by RPN on input image

International Telecommunications Union Recommendations (ITU-R) define subjective evaluation methods into three categories, Single Stimulus Impairment Scale(SSIS), Double Stimulus Impairment Scale (DSIS) and Double Stimulus Continuous Quality Scale (DSCQS)[22]. As the DSIS method is more suitable for assessing visible impairments, we select the DSIS procedure for our quality assessment. In the DSIS method, five levels for impairment grading are defined, and the reference image is shown with the processed image simultaneously. A five Degree scale adopted in our study is defined as in Table 1.

Table 1. The DSIS Five Degree Scale

| Quality | Impairment |
|-------------|--------------------------------------|
| 5 Excellent | 5 Accurately Localized and segmented |
| 4 Good | 4 Perceptible, but not erroneous |
| 3 Fair | 3 Slightly erroneous |
| 2 Poor | 2 Erroneous |
| 1 Bad | 1 Very fallacious |

3. METHODOLOGY

3.1 Data preparation and preprocessing

The three most commonly used MRI sequences are T1-weighted, T2-Weighted and Fluid Attenuated Inversion Recovery (Flair) images. T1-weighted MR images with 2D slices which were recorded at Guangzhou, China, and General Hospital, Tianjing Medical University, China, from 2005 to 2010 were used to develop and test the algorithm presented in this research [23]. The dataset consists of 3064 T1-weighted contrast-enhanced grayscale MR images, belonging to 233 different patients. Those images belonged to three tumor categories, namely Meningioma (708 slices), Gliomas (1426 slices), and Pituitary (930 slices). In the research presented only axial MRIs of Meningioma and Glioma tumors were considered. The implemented end-to-end system architecture was shown in Fig. 2. The initial MRIs were of 512x512 resolution, and at the preprocessing stage, those images were down-sampled to 128x128 to reduce the training time and the complexity of the network.

3.2 Network architecture for tumor classification and localization

A classical shallow convolutional neural network was developed to classify Meningioma and Glioma brain tumors by authors [24, 25]. The CNN network consists of two convolutional layers, and each convolution layer is followed by a ReLU activation layer and a 2x2 maxpool layer. The first convolution layer has 3x3x20 kernel structure and it generates 20 different feature maps for every MR image input. The output of this convolution layer goes through a ReLU activation function and then a 2x2 maxpooling layer. At the second convolution layer, there are 10 filters with 3x3 and the outputs of the previous convolution layer are directly fed to the second set of Convolution, ReLU and maxpooling layers. The maxpooling layers reduce the dimensions of the output tensor by considering the max and min values of the processed window, which assists to avoid the model overfitting. After getting the 2D output feature map from the two convolution layers, it was converted into 1D feature map using a flattening operation and fed into a fully connected network for the classification task. The last layer of this fully connected network outputs the classification task's decision either as Meningioma tumor or as Glioma tumor. In the proposed model a faster R-CNN network follows the CNN network. The objective of the faster R-CNN is the tumor localization process. The RPN-based faster R-CNN network is modeled following a transfer learning approach with the pre-trained weights generated using the COCO dataset, to detect the tumor boundaries. The classified output from the CNN is fed into the Faster R-CNN network and it generates estimated coordinates of the tumor bounding box as the output. The bounding box's output coordinates belong to low resolution 128x128 MR image, it was mapped to the original 512x512 resolution at the post-processing stage. Finally, Prewitt contour algorithm is applied to the bounding box area for a finer tumor segmentation [26].

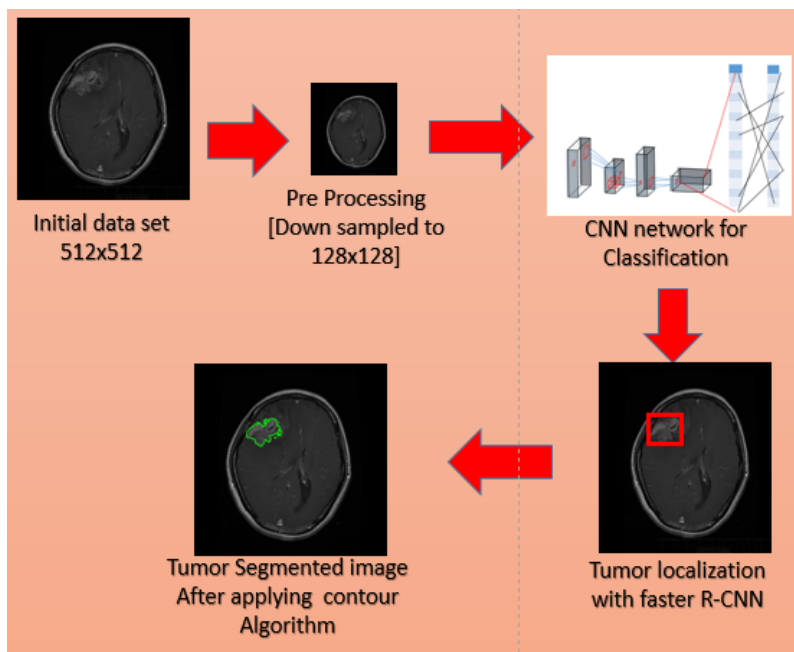


Fig. 2. Architecture of the network

3.3 Subjective quality evolution

The main objective of the research presented is to conduct a feasibility study, where the output of the AI model shown is compared against the decision of the medical practitioners. We have carried out a subjective quality evaluation of the proposed tumor segmentation architecture's output with this objective.

As expressed in the ITU recommendations, 5 to 35 number of subjects were required to perform a subjective equality assessment test, and all of these subjects should have normal vision and be experts in the medical sector. In general, 15 number of subjects are sufficient to perform a good subjective quality evaluation. Our study, we have selected 15 subjects, covering Neurologists, Medical Officers (MO) and Intern Home Officers (HO) according to the ITU recommendations. None of these experts was given access to the patient information or knowledge of the disease prevalence in the dataset used for this research. Before carrying out the assessment, a set of instructions consisting of the type of evaluation approach, the ranking scale, and the test duration were given to each subject. Environmental specifications were set to meet the room conditions, ambient lighting conditions, and viewing distance. Our subjective test was carried out by physically meeting selected evaluators with at least a good bachelor's degree in medicine, individually at their medical offices. About 10-15 minutes were given to doctors to read the instructions, adapt to the test environment and also lighting conditions. After that, a graphical user interface designed with Google Form was used for the evaluation, and this user interface was designed in a simple manner to assist the evaluators with low computer skills. The sample of the graphical user interface used for the evaluation process is shown in Fig. 3, and the link to the online questioner is at [27]. As in Fig. 3, the original MRI image and Segmented image are shown together according to ITU recommendations. The participants are asked to rate the prediction accuracy into five levels and rate of the prediction out of 100% at their discretion. Twelve images each were used for Glioma and Meningioma for the evaluation process, and all the data were recorded and analyzed. To make an unbiased evaluation process, MR images with incorrect segmentations were added to the subjective test in both Glioma and Meningioma data sets.

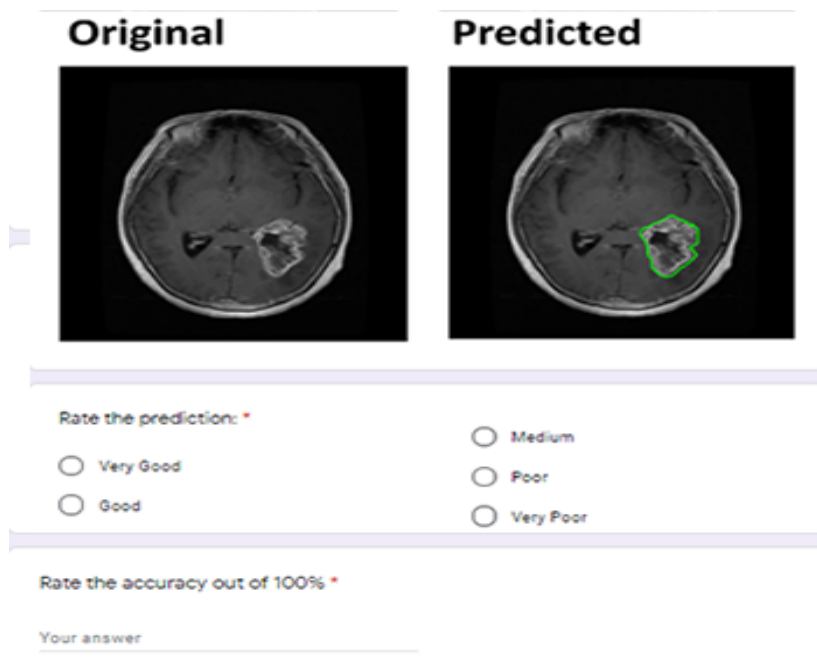


Fig. 3. Graphical User Interface for DISS test evaluation

4. RESULTS AND DISCUSSION

The performance analysis was carried out on a laptop computer with Intel(R) 8 Series Chip set processor, 4.00GB RAM and 64-bit operating system, x64-based processor. The objective quality evaluation of the segmentation is carried out using confidence interval, accuracy, Dice Score, and Boundary Displacement Error (BDE) metrics by comparing the tumor segmentation boundary of the proposed architecture with the ground truth. The results are tabulated in Table 2.

The objective parameters were calculated using standard definitions of them. Dice score is defined as the harmonic mean of precision and recall as in the eq. (1).

$$Dice\ Score = \frac{2TP}{(2TP + FP + FN)} \quad (1)$$

The accuracy of each segmentation was calculated according to the eq. (2).

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \quad (2)$$

The boundary displacement error, measures the average displacement error of one boundary pixel and the closest boundary pixels by means of the eq. (3)

$$\mu_{LA}(u, v) = \left\{ \frac{(u - v)}{(L - 1)} ; 0 < u - v \right\} \quad (3)$$

Here, TP is true positive, FP is False positive, FN is False negative and TN is True negatives.

Table 2. Overall performance of the system

| Boundary estimation method | Performance value |
|----------------------------|-------------------|
| Confidence interval | 94.78 % |
| Dice score | 0.90 |
| Accuracy | 0.9125 |
| BDE | 3.581 |

As of our subject sample, neurologists are the most experienced and qualified personals on tumor detection using MRI. Hence, their prediction accuracy can be treated as the ground truth to compare the proposed system model's output. The predicted confidence interval for tumor annotation using neural network and neurologists was shown in Fig. 4 and Fig. 5 for Meningioma and Glioma. According to the figures, most of the time average confidence level for tumor annotation was obtained using the proposed model tally closely with the neurologists' prediction results. Because of this similarity, we can deduce that the proposed model would work well in the field, to aid the medical professionals with a satisfying level of accuracy.

Further, we noticed a significant difference in the rating level between neurologists and non-specialists (i.e. Medical Officers (MO) and House officers (HO)) as in Fig. 6 and Fig. 7 for both Glioma and Meningioma identification. This is because the medical image diagnosis depends on the doctors' experience level and the level of training they had on subspecialties of medicine, such as neurology, radiology, and neurosurgery. According to Fig. 6 and Fig. 7, the MO/HO prediction could carry more errors than the neurologists' prediction. Since our proposed algorithm was able to perform accurately in the same capacity as of neurologists, this proposed model can be used to assist the diagnostic process of brain tumors in the absence of expertise, so that the patient can be directed for further medical examination by a qualified specialist.

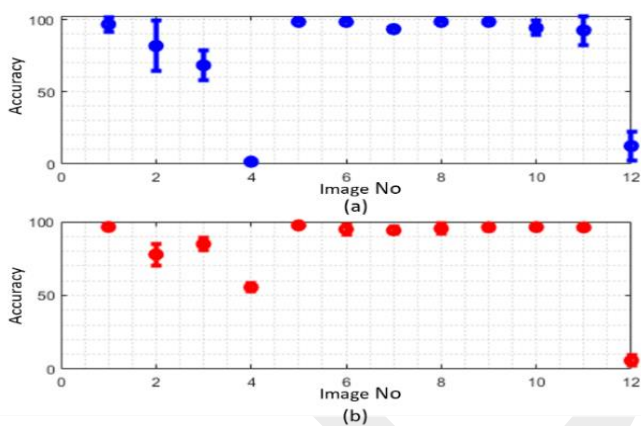


Fig. 4. Subjective evaluation for Meningioma between (a) Neurologist and (b) proposed network

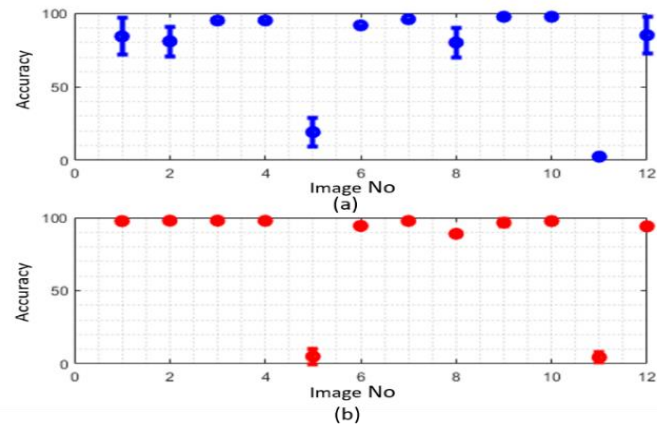


Fig. 5. Subjective evaluation for Glioma between (a) Neurologist and (b) proposed network

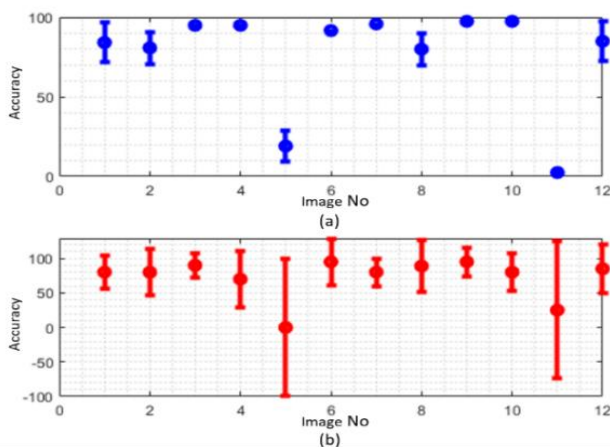


Fig. 6. Subjective evaluation for Glioma between (a) Neurologist and (b) MO/HO

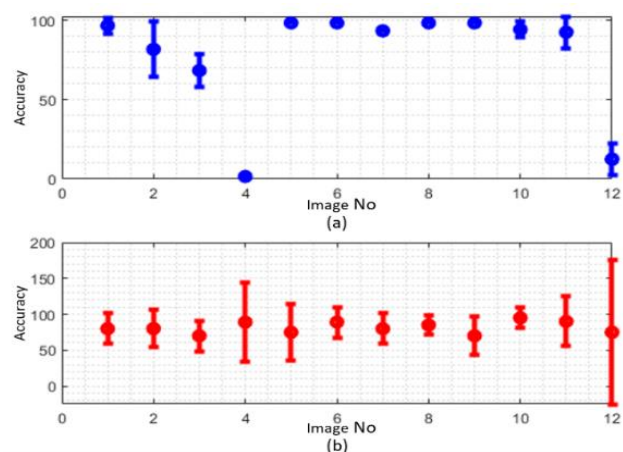


Fig. 7. Subjective evaluation for Meningioma between (a) Neurologist and (b) MO/HO

5. CONCLUSION

The paper content illustrates an important real-world analysis for the medical field. The validation of the model for tumor segmentation can be used for further improvements to the system. The deep learning algorithm consists of a typical CNN for tumor classification into Meningioma and Glioma, and a faster R-CNN based model for tumor localization of both cases. Authors used fivefold cross-validation with figshare dataset to train the network. For the segmentation task, a Prewitt edge detection algorithm was used and objective quality matrices such as confidence interval, dice score, accuracy and BDE are calculated to evaluate the performance. The system was able to segment a tumor with an accuracy of 91.25%; dice score of 0.90 and a confidence interval of 94.78%.

Further, we tested our system's performance, both using objective and subjective quality metrics. A subjective quality evaluation was carried out by following the DSIS protocol. According to the subjective evaluations, a significant variation in the tumor detection task between Neurologists and non-specialists can be observed. Yet, the segmentation decision obtained from the proposed network is closer to the neurologists' rating level. Hence we can conclude that our model can be a reliable aid for brain tumor classification and segmentation in low human resource, expertise, and environments.

ACKNOWLEDGMENT

The authors wish to acknowledge the insightful and extremely helpful reviews and comments provide by the Neurologists, Radiologists, Medical Officers (MO) and Intern Home Officers (HO) in Teaching Hospitals and district hospitals in the country. A special thanks goes to Dr. S.C Weerasinghe, Neurologist, at Teaching hospital Anuradhapura for his guidance during the design and validation stages of the research. Also we thank all the medical professionals who took part in our subjective results analysis stage, despite their busy schedule.

REFERENCES

- [1] A. Kabir Anaraki, M. Ayati, and F. Kazemi, "Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms," *Biocybern. Biomed. Eng.*, vol. 39, no. 1, pp. 63–74, 2019.
- [2] C. G. Madamombe, "Deep Learning Techniques to Classify and Analyze Medical Imaging Data," *Int. J. Comput. Sci. Eng.*, vol. 7, no. 04, pp. 109–113, 2018.
- [3] S. S. Yadav and S. M. Jadhav, "Deep convolutional neural network based medical image classification for disease diagnosis," *J. Big Data*, vol. 6, no. 1, 2019.
- [4] Anwar, S. M., Majid, M., Qayyum, A., Awais, M., Alnowami, M., & Khan, M. K. (2018). Medical Image Analysis using Convolutional Neural Networks: A Review. *Journal of Medical Systems*, 42(11). doi:10.1007/s10916-018-1088-1
- [5] X. Liu *et al.*, "A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis," *Lancet Digit. Heal.*, vol. 1, no. 6, pp. e271–e297, 2019.
- [6] P. Rajpurkar *et al.*, "Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists," *PLoS Med.*, vol. 15, no. 11, pp. 1–17, 2018.
- [7] V. Gulshan *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA - J. Am. Med. Assoc.*, vol. 316, no. 22, pp. 2402–2410, 2016.
- [8] M. Cicero *et al.*, "Training and Validating a Deep Convolutional Neural Network for Computer-Aided Detection and Classification of Abnormalities on Frontal Chest Radiographs," *Invest. Radiol.*, vol. 52, no. 5, pp. 281–287, 2017.
- [9] C. Dasanayaka and M. B. Dissanayake, "Deep Learning Methods for Screening Pulmonary Tuberculosis Using Chest X-rays," *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.*, no. August, pp. 0–18, 2020.
- [10] Lakhani P, Sundaram B. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. *Radiology*. 2017 Aug;284(2):574-582. doi: 10.1148/radiol.2017162326. Epub 2017 Apr 24. PMID: 28436741
- [11] P. Maduskar, M. Muyoyeta, H. Ayles, L. Hogeweg, L. Peters-Bax, and B. Van Ginneken, "Detection of tuberculosis using

- digital chest radiography: Automated reading vs. interpretation by clinical officers,” *Int. J. Tuberc. Lung Dis.*, vol. 17, no. 12, pp. 1613–1620, 2013.
- [12] L. Shen, “End-to-end training for whole image breast cancer diagnosis using an all convolutional design,” arXiv, no. August, pp. 1–12, 2017
- [13] S. Z. Ramadan, “Methods Used in Computer-Aided Diagnosis for Breast Cancer Detection Using Mammograms: A Review,” *J. Healthc. Eng.*, vol. 2020, 2020.
- [14] C. J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, and D. King, “Key challenges for delivering clinical impact with artificial intelligence,” *BMC Med.*, vol. 17, no. 1, pp. 1–9, 2019.
- [15] C. Cavaro-Menard, L. Zhang, and P. Le Callet, “Diagnostic quality assessment of medical images: Challenges and trends,” *2010 2nd Eur. Work. Vis. Inf. Process. EUVIP2010*, no. 1, pp. 277–284, 2010.
- [16] L. Faes *et al.*, “Automated deep learning design for medical image classification by health-care professionals with no coding experience: a feasibility study,” *Lancet Digit. Heal.*, vol. 1, no. 5, pp. e232–e242, 2019.
- [17] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, “A survey of the recent architectures of deep convolutional neural networks,” *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5455–5516, 2020.
- [18] Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). *Understanding of a convolutional neural network. 2017 International Conference on Engineering and Technology (ICET)*. doi:10.1109/icengtechnol.2017.8308186
- [19] K. O’Shea and R. Nash, "An Introduction to Convolutional Neural Networks", arXiv.org, 2021. [Online]. Available: <https://arxiv.org/abs/1511.08458>. [Accessed: 01- Jan- 2020].
- [20] S. Cui, L. Mao, J. Jiang, C. Liu, and S. Xiong, “Automatic semantic segmentation of brain gliomas from MRI images using a deep cascaded neural network,” *J. Healthc. Eng.*, vol. 2018, 2018.
- [21] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Region-Based Convolutional Networks for Accurate Object Detection and Segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 142–158, 2016.
- [22] ITU-T Recommendation (P.910), Subjective video quality assessment methods for multimedia applications, 2000
- [23] Cheng Jun, “brain tumor dataset”, figshare. Dataset, 2017, Available:<https://doi.org/10.6084/m9.figshare.1512427.v5>.
- [24] H. N. T. K. Kaldera, S. R. Gunasekara, and M. B. Dissanayake, “Brain tumor Classification and Segmentation using Faster R-CNN,” *2019 Adv. Sci. Eng. Technol. Int. Conf. ASET 2019*, 2019.
- [25] H. N. T. K. Kaldera, S. R. Gunasekara and M. B. Dissanayake, "MRI based Glioma segmentation using Deep Learning algorithms," 2019 International Research Conference on Smart Computing and Systems Engineering (SCSE), Colombo, Sri Lanka, 2019, pp. 51-6. doi: 10.23919/SCSE.2019.8842668
- [26] Rafael C. Gonzalez , Richard E. Woods, Digital Image Processing, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, 2001, pp 714-735,
- [27] “Aim of the project: Automate the tumor detection in MRI to aid the medical practitioners.” [Online]. Available: https://docs.google.com/forms/d/e/1FAIpQLSe379d9fteRilsusYgzb6Gr88VYbyOM1_Dxtg3Gbb4Lz6Ua6g/viewform?fbzx=-1925259441848764305. [Accessed: 29-Nov-2019].