

Classification of Cardiomegaly Using X-ray Images and Segmentation of MRI-Derived Hippocampus

Chinmay Mokashi
School of Biomedical Informatics
University of Texas Health Science Center
Houston, TX, USA
Chinmay.Mokashi@uth.tmc.edu

Samia Islam
School of Biomedical Informatics
University of Texas Health Science Center
Santa Clara, CA, USA
Samia.R.Islam@uth.tmc.edu

Sarita Soebianto
School of Biomedical Informatics
University of Texas Health Science Center
Houston, TX, USA
Sarita.Soebianto@uth.tmc.edu

Abstract— The purpose of this study is to address the challenge of creating the most optimal pattern recognition algorithms for classification of cardiomegaly based on X-ray images of the chest, and segmentation of the hippocampus based on MRI scans of the brain. Pattern recognition in deep neural network has the potential to aid clinicians in interpreting medical scans. Accurate diagnosis is important in healthcare. Thus, our aim is to obtain high accuracy of image classification and segmentation using deep learning methods. For image classification, the labeled chest X-ray images were provided. Our model included using different transfer learning methods to extract the features and then applying logistic regression and Support Vector Machine (SVM) for classification of the presence or absence of cardiomegaly. The results indicate the best performing models were ResNet-50 and InceptionV3 when used with logistic regression (AUC score of 0.84, and 0.823). For the segmentation task, we were provided with MRI scans of the brain and we performed image segmentation on the anterior and posterior areas of the hippocampus. We used the nnUNet method because of its plug and play benefit. The results show the anterior and posterior dice scores were 0.893 and 0.879, respectively. Thus, deep learning methods are capable of performing classification and segmentation on medical images, thereby, assisting in clinical decision making.

Keywords—deep learning, pattern recognition, image classification, segmentation, cardiomegaly, hippocampus

I. INTRODUCTION

Deep learning is a subset of artificial intelligence (AI) that applies computer algorithms to find meaningful representations of raw data through multiple layers of abstraction [1]. Deep neural networks (DNNs) are able to recognize patterns, and can thus help interpret medical scans, pathology slides, skin lesions, retinal images, electrocardiograms, endoscopy, faces, and vital signs [2]. Recently, medical fields have started to look into the potential of using AI technology such as deep learning as an aid for clinicians to interpret medical scans more accurately. Accuracy of detection and diagnosis of cancer and/or many other diseases depends on the expertise of individual clinicians [3]. That being said, it is prone to bias as human interpretation may differ from one to another.

Deep learning can be used to classify images. For classification tasks, higher layers of representation amplify aspects of the input that are important for discrimination and suppress irrelevant variations [1]. In this project, we will use pattern recognition algorithms in deep learning to classify cardiomegaly images and healthy images. Cardiomegaly is a condition of an enlarged heart that can be seen on any imaging test. Cardiomegaly is a symptom of cardiac insufficiency that occurs in various diseases and conditions, including high blood pressure, coronary artery disease, heart valve disease, and pulmonary hypertension [4]. X-rays of the

chest, being non-invasive with minimum radiation, as well as relatively low-cost, are widely used as a medical imaging test for early detection of cardiomegaly [4].

For cardiomegaly classification, we propose using pre-trained models, specifically ResNet-50, InceptionV3 and VGG-16, for creating our feature matrix from the images provided in the training data set. We will then apply logistic regression and Support Vector Machine (SVM) for classification of the images that are provided as part of the test data set. Cross validation will also be applied to some of our models. Finally, AUC scores will be used to evaluate the performance of the different models.

In recent years, deep learning has become a robust tool in image segmentation [5]. Applying image segmentation to the follow-ups is necessary in addition to proper disease detection and diagnosis [6]. Medical image segmentation separates the pixels belonging to organs or lesions from the background in images such as CT or MRI scans, with the purpose of delivering critical information about the shapes and volumes of these organs. [5]. However, it is known to be challenging due to the need for high accuracy and high precision. The absence of an accepted reference standard for clinical imaging data also creates a stumbling block for quantifying the accuracy of segmentations of medical images [7]. Precision is determined by the reproducibility of the segmentations obtained repeatedly from the same image [8]. Different image-derived parameters/metrics have been proposed for segmentation and quantification.

For the segmentation challenge, we are provided with MRI images of the brain. The hippocampus, a region of the brain, is considered as the center of a brain network supporting encoding, consolidation, and retrieval of memory [9]. It is central to the study of memory in humans and in recent years, the regional specialization and organization of hippocampal functions have been elucidated in experimental models and in human neurological and psychiatric diseases [10]. In diagnosis and disease management, morphological analysis of the hippocampus is important [10] and MRI scans can quantify the volume and assess the shape of the hippocampus in a noninvasive manner [11]. The gray levels of the hippocampus in MRI are very similar to other neighboring structures, such as the amygdala, the caudate nucleus and the thalamus [12]. Thus, automatic segmentation of the hippocampus in MRI is challenging.

In order to perform segmentation of the anterior and posterior areas of the hippocampus, we propose using nnUNet because of its ease of use. It was developed fairly recently and can segment features automatically from 2D and 3D images [13]. We will train the data, in 3D full resolution and after running the inference, we will evaluate the performance using DICE scores. We will also apply cross validation to our model.

II. METHODS

A. Image Classification

The aim for this task is to perform image classification to identify cardiomegaly based on chest X-ray images. A subset of the chest X-ray dataset from NIH were provided.

The dataset includes a total of 21,966 labeled images of healthy patients (767) and patients with cardiomegaly (21,199). Additionally, the dataset has been pre-split into a training set and a testing set. The testing set was set with a balanced number of 300 healthy samples and 300 cardiomegaly samples. Fig 1. shows an example of healthy and cardiomegaly chest X-ray images.

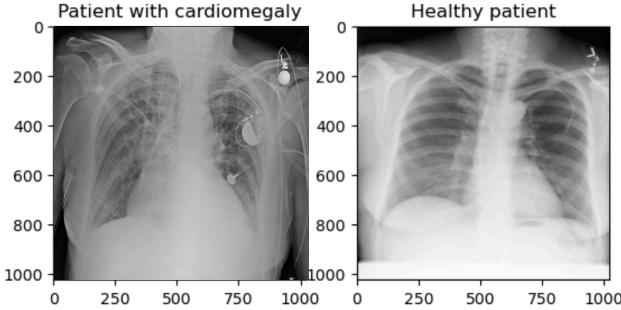


Fig. 1. Chest X-ray example

We used a pipeline to classify the X-ray images as seen in fig. 2.

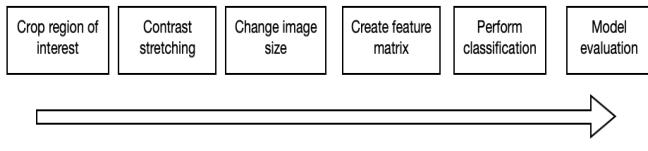


Fig. 2. Pipeline flowchart

Crop region of interest. Initially, we applied an 8 by 8 grid to the images focused on the left side of the sternum where the heart is located and cropped out the other irrelevant parts of the X-ray. However, in [14], we noticed in the X-ray images that they included for cardiomegaly patients, the enlargement of the heart was visible on the right side of the sternum. Therefore, we modified our cropping region to include both sides of the sternum and re-ran our best performing models.

Contrast stretching. Next, we performed contrast stretching. As mentioned in [15], contrast stretching, or normalization, stretches the range of intensity values to a full range of pixel values that the image allows, thus improving the contrast of the image. Example of before and after contrast stretching is in fig. 3.

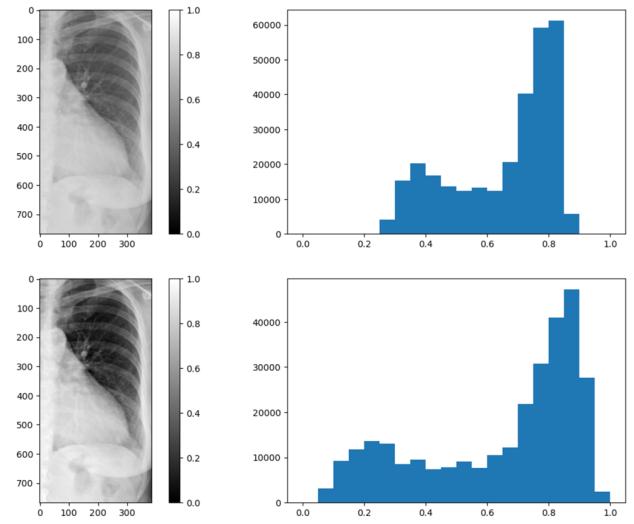


Fig. 3. Before (top) and after (bottom) contrast stretching with histogram on left chest region

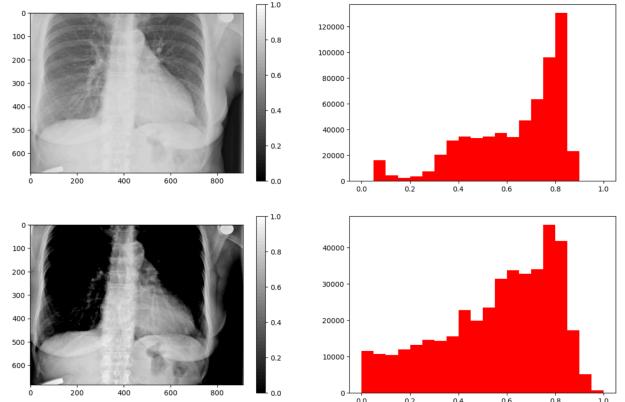


Fig. 4. Before (top) and after (bottom) contrast stretching with histogram on larger chest region

Change image size. After contrast stretching, we scaled the images to new dimensions, based on the requirements of the specific transfer learning methods we used. For example, a 299x299 dimension format is required by InceptionV3.

Create feature matrix. Zhou [14] performed classification of cardiomegaly using pre-trained models. Following their suit and from what we learnt from our course materials, we applied ResNet-50 and InceptionV3 for creating our feature matrix. In addition, we applied VGG-16, a model trained on over 14 million images which has generated a high accuracy in image classifications in the past [16].

Perform classification. Next, we performed logistic regression, which we learnt in our course. We also applied Support Vector Machines (SVM) on the dataset. SVM classifies data by drawing a hyperplane between two classes after training on labeled data [17]. It is a widely used classifier for image data [18]. Furthermore, we applied cross validation on some of our models. We split the healthy images from the training data into smaller chunks to take into account the disproportionately large number of healthy images compared to cardiomegaly images. Each chunk had the same number as the total cardiomegaly images in the training data, except for the last remaining chunk. This sample from the training data was used without replacement

in each fold. This way we ensured a more balanced distribution between the healthy and cardiomegaly images.

Model evaluation. Model evaluation was performed against the test set by plotting the ROC curves for the various algorithms. The AUC value and F1 score for each of these algorithms were also computed.

B. Hippocampus Segmentation

For this task, we were asked to perform image segmentation on the anterior and posterior areas of the hippocampus based on MRI data. We were provided with a dataset made available by Vanderbilt University Medical Center in Nashville, Tennessee.

Our dataset included:

- Training and testing sets
- Ground truth labels for anterior and posterior

We used the nnUNet method to perform the segmentation because of its plug-and-play benefit [19]. The nnUNet method is used for segmentation based on deep learning that “automatically configures itself” [19]. We performed the necessary pre-processing to ensure that the data format was compatible with nnUNet. We trained the data in 3D full resolution, ran the inference and performed DICE score evaluation. When training, we used 5-fold cross validation.

III. RESULTS

A. Image Classification

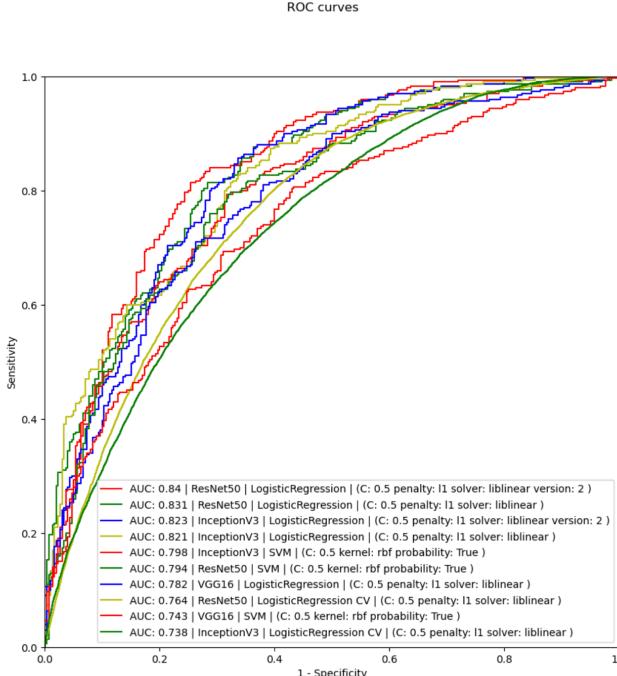


Fig. 5. ROC curves

Fig. 5. shows the ROC curves obtained from the various runs. The best performing models were ResNet-50 and InceptionV3 used with logistic regression. When run on the smaller region of the chest that was focused only on the left side of the sternum, the AUC scores for ResNet-50 and InceptionV3 were 0.831 and 0.821, respectively. ResNet-50 and InceptionV3 with SVM models performed fairly with an AUC score of 0.794 and 0.798. VGG-16 models had a lower AUC score of 0.782 with logistic regression and 0.743 with SVM. The models with logistic regression and cross

validation performed worst at 0.764 for ResNet-50 and 0.738 for InceptionV3.

The ResNet-50 and InceptionV3 logistic regression ROC curves denoted by version 2 were obtained when model was run using larger chest X-ray area. There was a very slight improvement in performance, with an AUC score of 0.84 and 0.831, respectively.

The F1 and AUC results for all the models is given in Table I.

TABLE I F1 SCORES

transfer_learning_type	model_type	train.threshold	train.f1	train.auc	test.threshold	test.f1	test.auc
InceptionV3	LogisticRegression	CV	0.864340	0.69427	0.672730	0.580378	0.636977
ResNet50	SVM		0.042580	0.934890	0.909088	0.022467	0.740620
InceptionV3	SVM		0.053291	0.997854	0.999546	0.017511	0.750000
VGG16	SVM		0.039617	0.954248	0.904048	0.024681	0.690000
ResNet50_v2	LogisticRegression		0.071546	0.501530	0.583723	0.013097	0.643289
VGG16	LogisticRegression		0.022514	0.154713	0.895247	0.018453	0.717428
InceptionV3	LogisticRegression		0.029279	0.183285	0.926540	0.013103	0.759375
ResNet50	LogisticRegression		0.036196	0.223339	0.936347	0.012171	0.773885
ResNet50	LogisticRegression	CV	0.963258	0.73019	0.668327	0.487795	0.671736
InceptionV3_v2	LogisticRegression		0.009492	0.064306	0.672800	0.009550	0.646400

The age and gender of the patients were also provided in the data sets, as well as whether the X-ray images were posterior-anterior (PA) or anterior-posterior (AP) view positions. Performance of ResNet-50 logistic regression model for training and testing sets across age, gender and X-ray view positions are given in Tables II, III and IV, respectively. These tables are for the case when model was run on the larger chest X-ray regions.

TABLE II RESNET-50 LOGISTIC REGRESSION MODEL PERFORMANCE ACROSS AGE GROUPS

age_binned	train	FN	FP	TN	TP	precision	recall	f1
0 (0, 10]	0	14	1	4	5	0.833333	0.263158	0.4
1 (10, 30]	0	100	5	52	43	0.895833	0.300699	0.450262
2 (30, 40]	0	88	5	49	38	0.883721	0.301587	0.449704
3 (40, 50]	0	115	6	69	40	0.869565	0.258065	0.39801
4 (50, 60]	0	118	14	69	61	0.813333	0.340782	0.480315
5 (60, 80]	0	118	12	75	91	0.883495	0.435407	0.583333
6 (80, 100]	0	4	1	1	2	0.666667	0.333333	0.444444

age_binned	train	FN	FP	TN	TP	precision	recall	f1
0 (0, 10]	1	303	1	187	2	0.666667	0.00655738	0.012987
1 (10, 30]	1	2476	25	1635	63	0.715909	0.0248129	0.0479635
2 (30, 40]	1	2211	13	1563	66	0.835443	0.0289855	0.0560272
3 (40, 50]	1	2372	30	2076	116	0.794521	0.04646238	0.088079
4 (50, 60]	1	2339	52	2312	129	0.712707	0.052269	0.073952
5 (60, 80]	1	1710	60	1880	186	0.756098	0.09801013	0.173669
6 (80, 100]	1	66	11	64	13	0.541667	0.164557	0.252427
7 (100, 450]	1	1	1	0	4	nan	0	nan

TABLE III RESNET-50 LOGISTIC REGRESSION MODEL PERFORMANCE ACROSS GENDERS

gender	train	FN	FP	TN	TP	precision	recall	f1
0 F	0	273	24	155	152	0.863636	0.357647	0.505824
1 M	0	284	20	164	128	0.864865	0.31068	0.457143

gender	train	FN	FP	TN	TP	precision	recall	f1
0 F	1	5851	111	3949	318	0.741259	0.0515481	0.0963928
1 M	1	5627	81	5772	257	0.760355	0.0436778	0.0826101

TABLE IV RESNET-50 LOGISTIC REGRESSION MODEL PERFORMANCE ACROSS VIEW POSITIONS

view_pos	train	FN	FP	TN	TP	precision	recall	f1
0 AP	0	34	2	37	31	0.939394	0.476923	0.632653
1 PA	0	523	42	282	249	0.85567	0.322539	0.468485

view_pos	train	FN	FP	TN	TP	precision	recall	f1
0 AP	1	947	29	1285	58	0.666667	0.05772114	0.106227
1 PA	1	10531	163	8436	517	0.760294	0.0467958	0.0881651

B. Hippocampus Segmentation

For each MRI volume, we calculated the DICE scores for the anterior and posterior of the hippocampus. The results are given in Table V. As can be seen from the table, the anterior and posterior dice scores ranged from 0.797-0.950 and 0.770-0.939, respectively.

TABLE V DICE SCORES

pid	anterior_dice	anterior_precision	anterior_recall	posterior_dice	posterior_precision	posterior_recall
0 006	0.917	0.978	0.863	0.904	0.931	0.878
1 014	0.865	0.865	0.866	0.820	0.820	0.819
2 017	0.917	0.956	0.881	0.901	0.922	0.882
3 024	0.875	0.840	0.912	0.839	0.873	0.807
4 035	0.906	0.943	0.871	0.882	0.926	0.842
5 038	0.848	0.837	0.860	0.834	0.861	0.809
6 040	0.920	0.939	0.901	0.914	0.887	0.942
7 042	0.847	0.828	0.866	0.779	0.812	0.749
8 046	0.895	0.922	0.870	0.866	0.873	0.859
9 049	0.925	0.914	0.937	0.900	0.909	0.891
10 074	0.880	0.859	0.903	0.888	0.865	0.912
11 075	0.903	0.873	0.935	0.914	0.951	0.880
12 087	0.943	0.937	0.948	0.932	0.914	0.952
13 098	0.925	0.923	0.927	0.921	0.955	0.889
14 102	0.858	0.975	0.765	0.864	0.835	0.894
15 106	0.930	0.915	0.945	0.920	0.890	0.952
16 108	0.928	0.890	0.970	0.905	0.900	0.911
17 124	0.923	0.897	0.950	0.927	0.933	0.921
18 138	0.902	0.834	0.981	0.905	0.926	0.886
19 143	0.909	0.947	0.875	0.915	0.900	0.931
20 145	0.940	0.938	0.941	0.895	0.871	0.920
21 163	0.846	0.777	0.928	0.829	0.939	0.743
22 164	0.797	0.692	0.940	0.770	0.906	0.669
23 170	0.867	0.796	0.951	0.868	0.859	0.877
24 171	0.947	0.960	0.935	0.921	0.932	0.911
25 174	0.916	0.927	0.905	0.899	0.948	0.855
26 184	0.937	0.950	0.924	0.936	0.901	0.974
27 188	0.838	0.761	0.932	0.875	0.903	0.849
28 189	0.950	0.952	0.948	0.921	0.938	0.905
29 197	0.938	0.935	0.940	0.939	0.956	0.922
30 199	0.937	0.958	0.917	0.903	0.917	0.890
31 227	0.939	0.939	0.939	0.933	0.935	0.931
32 230	0.923	0.915	0.931	0.916	0.917	0.915
33 233	0.851	0.843	0.859	0.857	0.828	0.889
34 234	0.884	0.863	0.905	0.876	0.896	0.857
35 253	0.903	0.958	0.854	0.883	0.891	0.875
36 265	0.919	0.899	0.940	0.897	0.937	0.860
37 279	0.836	0.807	0.868	0.867	0.850	0.885
38 290	0.870	0.832	0.912	0.846	0.865	0.828
39 304	0.906	0.922	0.890	0.852	0.825	0.880
40 305	0.850	0.869	0.832	0.816	0.833	0.801
41 317	0.848	0.843	0.853	0.845	0.814	0.877
42 318	0.821	0.928	0.736	0.813	0.767	0.866
43 320	0.850	0.758	0.967	0.865	0.921	0.815
44 331	0.888	0.843	0.937	0.878	0.922	0.837
45 334	0.852	0.840	0.865	0.818	0.799	0.839
46 335	0.820	0.792	0.850	0.835	0.801	0.873
47 338	0.878	0.870	0.887	0.849	0.876	0.823
48 351	0.878	0.853	0.903	0.885	0.925	0.848
49 366	0.921	0.968	0.878	0.890	0.885	0.896
50 372	0.936	0.932	0.941	0.901	0.884	0.917
51 380	0.921	0.948	0.894	0.902	0.855	0.955

Table VI shows the final summary of the mean scores. For the anterior, the mean DICE score was 0.893 and for the posterior, it was 0.879.

TABLE VI SUMMARY OF MEAN SCORES

anterior_dice	0.893
anterior_precision	0.887
anterior_recall	0.902
posterior_dice	0.879
posterior_precision	0.888
posterior_recall	0.873
dtype: object	

IV. DISCUSSION AND CONCLUSION

A. Image Classification

Eight models were used for the cardiomegaly image classification task. Their AUC values and F1 scores were compared. The ResNet-50 and InceptionV3 models used with logistic regression performed the best. Changing the region of interest to include the right side of the sternum did produce a very slight improvement. However, the impact was minimal, where AUC value went from 0.831 to 0.84. This improvement does not seem meaningful, given that more resources are needed to perform the algorithms on a much larger image area. Applying cross validation on the logistic regression models made their performance much worse. We believe that using the same cardiomegaly training images for all the folds resulted in overfitting of the model, leading to a poorer performance. In the future, we want to explore more ways of cross validation by splitting the training data set into smaller chunks that would include a subset of cardiomegaly images, and an equal number of healthy images. Given that the number of cardiomegaly images is only a fraction of the healthy images, we could use them with replacement, while the samples of the healthy images would be without replacement.

We were also interested in finding out if there was any inherent bias introduced in the images by age, gender and X-ray view positions. For instance, we wanted to know if there was any significant difference in the proportion of men and women between train and test sets. Preliminary analysis from the tables included in this report showed that there was no remarkable bias between the train and test sets. We plan on exploring this in the future where we will analyze whether age, gender and X-ray view positions played any statistically significant role in a diagnosis of cardiomegaly and whether the models would perform differently based on these features.

B. Hippocampus Segmentation

We performed image segmentation on the anterior and posterior areas of the hippocampus based on MRI data. As mentioned in the data challenge description, we computed the DICE score for each MRI volume followed by averaging them, rather than computing a single score on all the results, which would lead to giving more weight to larger hippocampi.

As mentioned in the introduction, automatic segmentation of the hippocampus from MRI scans is a challenging task. Given the simplicity of using the nnUNet method in our deep learning model, a relatively good performance was achieved in segmenting the anterior and posterior areas.

ACKNOWLEDGMENT

We thank Dr. Luca Giancardo for the opportunity he gave us in this challenge.

REFERENCES

- Y. LeCun, Y. Bengio, G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.
- E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," *Nat Med*, vol. 25, pp. 44–56, 2019. doi: 10.1038/s41591-018-0300-7.
- R. P. Kruger, J. R. Townes, D. L. Hall, S. J. Dwyer and G. S. Lodwick, "Automated radiographic diagnosis via feature extraction and classification of cardiac size and shape descriptors," *IEEE Transactions on Biomedical Engineering*, vol. BME-19, no. 3, pp. 174-186, May 1972, doi: 10.1109/TBME.1972.324115.

4. M. S. Lee, Y. S. Kim, M. Kim, M. Usman, S. S. Byon, S. H. Kim, B. I. Lee, and B. D. Lee, "Evaluation of the feasibility of explainable computer-aided detection of cardiomegaly on chest radiographs using deep learning," *Scientific Reports*, vol. 11, no. 1, pp. 16885, 2021, doi: 10.1038/s41598-021-96433-1.
5. M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: achievements and challenges," *Journal of Digital Imaging*, vol. 32, no. 4, pp. 582–596, 2019, doi: 10.1007/s10278-019-00227-x.
6. B. Foster, U. Bagci, A. Mansoor, Z. Xu, and D. J. Mollura, "A review on segmentation of positron emission tomography images," *Computers in Biology and Medicine*, vol. 50, pp. 76–96, 2014, doi: 10.1016/j.combiomed.2014.04.014.
7. S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation," *IEEE transactions on medical imaging*, vol. 23, no. 7, pp. 903–921, 2004, doi: 10.1109/TMI.2004.828354.
8. D. Nicoll and W. Detmer, Basic Principles of Diagnostic Test Use and Interpretation. New York: McGraw-Hill, 2001, ch. 1, pp. 1–16.
9. T. Bartsch, The Clinical Neurobiology of the Hippocampus. Oxford, UK: Oxford University Press, 2012.
10. T. Bartsch, and P. Wulff, "The hippocampus in aging and disease: From plasticity to vulnerability," *Neuroscience*, vol. 309, pp. 1–16, 2015, doi: 10.1016/j.neuroscience.2015.07.084.
11. G. B. Frisoni, "Structural imaging in the clinical diagnosis of Alzheimer's disease: problems and tools," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 70, no. 6, pp. 711–718, 2001. <http://jnp.bmjjournals.org/content/70/6/711.short>.
12. B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrave, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, and A. M. Dale, "Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain," *Neuron*, vol. 3, pp. 341–355, 2002, doi: 10.1016/s0896-6273(02)00569-x.
13. Rowland W. Pettit, Britton B. Marlatt, Stuart J. Corr, Jim Havelka, Abbas Rana, "nnU-Net deep learning method for segmenting parenchyma and determining liver volume from computed tomography images," *Annals of Surgery Open*, vol. 3, no. 2, pp. e155, Jun. 2022, doi: 10.1097/AS9.000000000000155.
14. S. Zhou, X. Zhang, R. Zhang, "Identifying cardiomegaly in ChestX-ray8 using transfer learning," *Stud Health Technol Inform*, vol. 264, pp. 482–486, Aug 2019, doi: 10.3233/SHTI190268.
15. R. Fisher, S. Perkins, A. Walker, E. Wolfart, "Contrast stretching." Image processing learning resources. <https://homepages.inf.ed.ac.uk/rbf/HIPR2/stretch.htm> (accessed Dec 4, 2022).
16. Q. Guan, Y. Wang, B. Ping, D. Li, J. Du, Y. Qin, H. Lu, X. Wan, J. Xiang, "Deep convolutional neural network VGG-16 model for differential diagnosis of papillary thyroid carcinomas in cytological images: a pilot study," *J. Cancer*, vol. 10, no. 20, pp. 4876–4882, Aug. 2019, doi: 10.7150/jca.28769.
17. S. Huang, N. Cai, P. P. Pacheco, S. Narandes, Y. Wang, W. Xu, "Applications of support vector machine (SVM) learning in cancer genomics," *Cancer Genomics Proteomics*, vol. 15, no. 1, pp. 41-51, Jan.-Feb. 2018, doi: 10.21873/cgp.20063.
18. L. H. Thai, T. S. Hai, N. T. Thuy, "Image classification using support vector machine and artificial neural network," *I. J. Information Technology and Computer Science*, vol. 5, pp. 32-38, May 2012, doi: 10.5815/ijitcs.2012.05.05.1.
19. F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, pp. 203-211, Dec. 2021, doi: 10.1038/s41592-020-01008-z.