ORIGINAL PAPER

A Comparative Analysis of Data Augmentation Approaches for Magnetic Resonance Imaging (MRI) Scan Images of Brain Tumor

Muhammad Farhan Safdar¹,
Shayma Saad Alkobaisi²,
Fatima Tuz Zahra³

¹Bahria University, Islamabad, Pakistan

²College of IT, UAEU, Al-Ain, UAE

³KRL Hospital, Islamabad, Pakistan

Corresponding author: Muhammad Farhan Safdar, MS. Department of Software Engineering, Bahria University, Islamabad, Pakistan Address: Sector E-8 Shangrila Road Islamabad, Pakistan. Phone: +92-334-5212458. E-mail: mfarhan166@gmail.com. ORCID ID: <http://www.orcid.org/0000-0001-7818-4805>.

doi: 10.5455/aim.2020.28.29-36

ACTA INFORM MED. 2020 MAR 28(1): 29-36

Received: Jan 05, 2020

Accepted: Feb 08, 2020

ABSTRACT

Introduction: Machine Learning (ML) is a rapidly growing subfield of Artificial Intelligence (AI). It is used for different purposes in our daily life such as face recognition, speech recognition, text translation in different languages, weather prediction, and business prediction. In parallel, ML also plays an important role in the medical domain such as in medical imaging. ML has various algorithms that need to be trained with large volumes of data to produce a well-trained model for prediction. **Aim:** The aim of this study is to highlight the most suitable Data Augmentation (DA) technique(s) for medical imaging based on their results. **Methods:** DA refers to different approaches that are used to increase the size of datasets. In this study, eight DA approaches were used on publicly available low-grade glioma tumor datasets obtained from the Tumor Cancer Imaging Archive (TCIA) repository. The dataset included 1961 MRI brain scan images of low-grade glioma patients. You Only Look Once (YOLO) version 3 model was trained on the original dataset and the augmented datasets separately. A neural network training/testing ecosystem named as supervisely with Tesla K80 GPU was used for YOLO v3 model training on all datasets.

Results: The results showed that the DA techniques rotate at 180o and rotate at 90o performed the best as data enhancement techniques for medical imaging. **Conclusion:** Rotation techniques are found significant to enhance the low volume of medical imaging datasets.

Keywords: Machine Learning, Data Augmentation, Medical Imaging.

1. INTRODUCTION

In recent years, Machine Learning (ML) has made a meaningful impact on our daily lives. ML is the sub-field of Artificial Intelligence (AI) that gives the computer an ability to learn and improve with experience to solve real-world problems by using algorithms and statistical models. It is used for text translation (1), face recognition (2), speech recognition (3) and even has a variety of work in medical imaging as segmentation and automatic detection and diagnosis of different diseases (4).

Machine Learning refers to supervised and unsupervised learning techniques. Its accuracies are a significant part of automatic detection. However, ML accuracies depend on how efficiently algorithms are tuned on data. ML algorithms

require a large volume of data for training. Due to the scarcity of reliable training data in a few domains such as medicals imaging, accuracies of ML algorithms are affected in most of the cases (5, 6).

Data Augmentation (DA) is used to enhance the size of low volume datasets (7). It refers to some common approaches such as scaling, rotation, translation, and flipping (8). However, these approaches have their accuracy levels depending on the type of images. As ML algorithms require a large volume of data for producing a well-generalized model (9), this research work is aimed to increase the limited volume of MRI scans of brain tumor images by applying different DA approaches and then compare the results of these approaches.

© 2020 Muhammad Farhan Safdar, Shayma Saad Al Kobaisi, Fatima Tuz Zahra

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Figure 1. Tumor on different regions of brain; (A) Tumor on Parietal lobe of the brain; (B) Tumor on Temporal lobe of the brain; (C) Tumor on Frontal lobe of the brain; (D) Tumor on Occipital lobe of the brain (13)

Medical imaging is an essential part of diverse fields of clinical practice and medical research. Radiologists classify diseases such as tumors from Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) scan images. Neuroscientists identify metabolic brain activity from MRI, Magnetic Resonance Spectrum Imaging (MRSI) or Positron Emission Tomography (PET) (10). A tumor is a mass of tissue that is formed by some abnormal cells (11). In normal cycle i.e. a normal healthy person, these cells die with the passage of time and new cells generate again in body but in tumor condition or abnormal condition, this cycle is disturbed. Cells are made in the body; even when the body does not require these cells and old cells do not die. As a result, the tumor starts to grow by more and more cells that are added to the mass (12).

A brain tumor is a collection of unwanted cells in the brain. It can be cancerous or non-cancerous. Based on the brightness of the cerebrospinal fluid (CSF), MRI brain images can be divided into three types of images including T1-weighted, T2-weighted and Fluid Attenuated Inversion Recovery (FLAIR). The CSF tissue shows as dark in T1-weighted and FLAIR while bright in T2-weighted images (MRI Basics). Figure 1 shows low-grade glioma tumors in different regions of the brain which were taken from the Tumor Cancer Imaging Archive repository (TCIA) (13).

An ML algorithm named as You Only Look Once (YOLO) (14) is a state-of-the-art ML model for object detection, classification, and localization (15,16). Previously YOLO model achieved good accuracy results on Common Objects in Context (COCO) dataset (17). For classification and localization of objects in the input image, YOLO processed the entire image at once, unlike the sliding window approach that is followed by the Convolutional Neural Network (CNN) architecture that is why YOLO proved to be much faster (16).

The primary purpose of this study is to apply DA approaches on a small dataset of MRI scans of brain tumor images and compare their classification results to distinguish which technique(s) are more suitable for medical images. The major contribution of this research work is to highlight the most suitable DA technique(s) for medical imaging based on their results.

Literature Review

Training of the ML model over low volume dataset becomes a challenging task. Low volume of data increases the chances of model training either as over fitted or underfitting. In both cases, the accuracy of such a model is affected. In medical imaging, many other authors adopted different ways to avoid over fitted or under fitted ML models training.

Zhe Zhu et al. (18) evaluated deep learning models' performance to discriminate breast cancer molecular subtypes from MRI images. Dataset included 270 cases with 6480 images that were reviewed by the institutional board. Three different deep learning models: one that was learning from scratch that was only trained on tumor patches, a second that was transfer learning where the model was trained on natural images but it was fine-tuned on tumor patches and a third was off the shelf deep learning model where the model was trained on only natural images and used for classification with Support Vector Machine (SVM). Area Under ROC Curve (AURC) was calculated as a performance metric. Study results showed that the best Area Under Curve (AUC) 0.65 was achieved by off the shelf deep learning model for distinguishing molecular subtypes. On training, the transfer learning approach achieved AURC of 0.60 and deep learning from scratch approach achieved AURC of 0.58.

Shanzen Pang et al. (19) identified the cholelithiasis and classified the gallstone from CT scan images. They created their dataset by collecting 223846 CT scan images of 1369 patients. The ML model namely YOLO v3-arch was trained on the CT scan images. The accuracy of classification and identification were produced by the 10-fold cross-validation techniques. Study results showed that their proposed model achieved 80.3% accuracy in the identification of muddy gallstone and 92.7% in the identification of granular gallstone while 86.50% accuracy was achieved in identifying the cholelithiasis.

Natalia Antropova et al. (20) classified breast tumor cells as Benign or Malignant. Dataset included three different imaging modalities from which 690 were MRI cases, 245 were mammographic cases and 1125 were ultrasound cases. A non-linear support vector machine

was exploited for classification with conventional computer-aided diagnosis (CADx) and Convolutional Neural Network (CNN) features. AURC was calculated for performance evaluation. The study results showed that AURC for MRI cases was 0.89, AURC for Mammography cases was 0.86 and AURC for ultrasound cases was 0.90. Sukh H. et al. (21) evaluated the performance of deep ensemble learning on MRI brain images. Dataset included MRI brain images of 805 patients. They proposed a framework that was a combination of deep learning and sparse regression. The CNN model used in their study contained three layers named convolutional, pooling and fully connected layers. Deep ensemble Sparse Regression Network (DeepESRNet) was compared with two different methods named as Multi-Output Linear Regression (MOLR) and Joint Linear and Logistic Regression (JLLR). Study results showed that the accuracy of DeepESRNet with MOLR was 90.28% and DeepESRNet with JLLR was 91%.

Mingxia Liu et al. (22) evaluated a deep learning model for the diagnosis of brain diseases. A data-driven learning technique named a landmark-based deep learning approach was used in their study. Three datasets ADNI-1, ADNI-2, and MIRIAD were used. Landmark based deep learning can learn images from one end to the other end, which extracted both local and global features. Study results showed that on the ADNI-1 dataset their proposed approach achieved 92.75% accuracy with AUC of 97.16%. On the ADNI-2 dataset, it achieved 91.09% accuracy with AUC of 95.86%. On the MIRIAD dataset, their approach achieved 92.75% accuracy with AUC of 97.16%. Rahul Paul et al. (23) predicted the survival rate in patients having lung adenocarcinoma by classification of disease. Dataset included 960 CT Lung images. Several architectures of pre-trained CNN with a combination of traditional features was used as a classifier. AURC was used for performance evaluation. The study results showed, one of the pre-trained models namely VGG-f with 5 traditional features achieved the best classification accuracy of 90% with 0.935 AURC.

Jiamin Liu et al. (24) evaluated the performance of Regional Convolutional Neural Network (R-CNN) for colitis detection on CT scan images. Dataset included 56 patients with 448 images. In the first step, their applied technique created 3000 independent region proposals on each slice of CT scan film by using selective search. R-CNN with 7 hidden layers was used for feature extraction. Once region proposals were created, the fixed-length feature vector was extracted from region proposals. Finally, colitis was classified and confidence score was assigned to it by using the Support Vector Machine classifier. Their study results showed 85% sensitivity with 1 false-positive/image.

Yanping Xue et al. (25) classified osteoarthritis in the hip joint from X-ray images. The dataset contained 420 images and was divided into osteoarthritis and normal images by two radiologists; both had greater than 20 years of experience. 219 images were in the normal group & 201 images were in the osteoarthritis group.

Deep CNN with 16 layers called VGG-16 was used for classification. Neural Network performance was compared with one expert radiologist, who had 10 years of experience in radiology. Neural Network showed 92.8% accuracy, 95% sensitivity, and 90.7% specificity. Mohammad Tariq et al. (26) detected an abnormality in Chest X-Ray images through deep Convolutional Neural Network (d-CNN). In their study, three datasets named Indiana, JSRT, and Shenzhen were used. The heat map was applied to the abnormalities area. Different pre-trained deep convolutional neural network performances were evaluated on the given datasets. Applied neural network models were fine-tuned on X-Ray images as well. Twenty abnormalities on chest X-Rays were experimented in the datasets using deep learning models. One of the abnormalities, named cardiomegaly showed the accuracy of 86% on AlexNet, 92% on VGG-19 and 87% on ResNet.

Joseph Redmon et al. (27) evaluated YOLO model for object detection from real-time and artwork images. The authors proposed a unified architecture of the YOLO model with 24 convolutional layers followed by 2 fully connected layers. They used 1x1 reduction layers followed by convolutional layers. The network was trained on ImageNet 1000 class dataset. YOLO performance was compared with other object detectors such as R-CNN, Faster R-CNN and Deformable Parts Model (DPM). Their proposed architecture performed 45 frames per second. A simple network with a single convolutional network was run on each image. The study results showed that in YOLO performance, background and localization error was 4.7% and 19% respectively while correct object detection was 65.5%. In Faster R-CNN performance, background and localization error was 14.6%, 8.6% respectively and correct object detection was 71.6%.

2. METHODS

In this study, two publicly available datasets were obtained from The Cancer Imaging Archive (TCIA) repository (13). The datasets downloaded from TCIA repository consisted of T1-weighted, T2-weighted and FLAIR images. However, T2-weighted images were excluded from the datasets due to the dissimilarity of T2-weighted images with T1-weighted and FLAIR images as defined earlier in the introduction section. The first dataset consisted of 900 MRI images of 159 patients that suffered from low-grade glioma and the second dataset consisted of 1061 MRI images of 199 patients that suffered from low-grade glioma. These two datasets were merged as one because individually in both datasets, images were not enough to consider for applying DA techniques and training. The Digital Image Communication in Medicine (DICOM) images were converted into jpg format using MicroDicom software. As the same dimensions of input images are required for better training of ML algorithms (28), so the dimensions of all input images were modified to 256x256 using MATLAB function. After modifying, very few images about 18-20 were showing loss of important infor-

mation about tumor object so we concluded these images from the dataset.

To analyze the effect of DA on medical images enhancement, different DA approaches including flip vertical, flip horizontal, rotate at 1800, rotate at 900, blur, noise, shear, and crop & scale were applied on the original dataset to increase the low volume data. An ecosystem Supervisely (29) was used for applying DA approaches to original images dataset and training purposes. Supervisely provides a powerful tool named Data Transformation Language (DTL) which was used to apply DA approaches on the original images' dataset.

The purpose of the DTL tool is to apply different DA techniques on the original dataset and tags the images as a train or validate by using a simple program that consists of JSON tags. A DTL program consists of three basic layers that are Data layer, the Transformation layer, and the Save layer. To produce the dataset against each DA approach, the DTL program with the specific function was run multiple times on the original images' dataset (as source dataset). Implementation details of DTL function for each DA approach are given below:

I. Flip Vertical

To obtain the vertically flip dataset, all images of the original dataset flipped vertically to change the location of the glioma patch in images. In the transformation layer of the DTL program, the value of parameter "action" was set to "flip" and "axis" to "vertical". To increase the size of the original dataset, the resulting flip vertical dataset was merged with the original images' dataset.

II. Flip Horizontal

To obtain the horizontally flip dataset, all images of the original dataset flipped horizontally to change the location of the glioma patch in images. In the transformation layer of the DTL program, the value of parameter "action" was set to "flip" and "axis" to "horizontal". To increase the size of the original dataset, the resulting flip horizontal dataset was merged with the original images' dataset.

III. Noise

To obtain the noise dataset, the noise filter was added in all images of the original dataset. In the transformation layer of the DTL program, the value of parameter "action" was set to "noise", the mean and standard deviation was set to 0 and 35 respectively. Mean and standard deviation indicates the density of noise in the image. The noise was added at the different levels by changing mean and standard deviation values but the suitable level of noise was found when "mean" was set to "0" and "standard deviation" was set to "35". To increase the size of the original dataset, the resulting noise dataset was merged with the original images' dataset.

IV. Rotate at 900

To obtain the 900 rotation dataset, all images of the original dataset were rotated at 900. In the transformation layer of the DTL program, the value of parameter "action" was set to "rotate" and in settings tags

min/max degrees was set to 900 because the resulting images were required at the exact angle of 900. To increase the size of the original dataset, the resulting dataset was merged with the original images' dataset.

V. Rotate at 1800

To obtain the 1800 rotation dataset, all images of the Original dataset were rotated at 1800. The value of parameter "action" was set to "rotate" and in settings tags min/max degrees was set to 1800 because the resulting images were required at the exact angle of 1800. To increase the size of the original dataset, the resulting dataset was merged with the original images' dataset.

VI. Shear

To obtain the shear dataset, the shear effect was applied to all images of the original dataset. To apply the shear effect on images, MATLAB version 2013a was used instead of the DTL program of Supervisely because it does not provide a function for shear effect. MATLAB function `affine2d` was used for this purpose which takes three parameters.

`affine2d ([100; 510; 001]);`

The above function parameter shows the values of the 3x3 affine matrix for a given object or image. To increase the size of the original dataset, the resulting shear dataset was merged with the original images' dataset.

VII. Crop & Scale

To obtain the crop & scale dataset, all images of the original dataset were cropped at a fixed rectangle and then scaled up. In the DTL program, two transformation layers were defined, one for crop and second for scale. In the first transformation layer of DTL program, in above syntax, "action" was set to "crop" and dimensions for crop rectangle were set for all images such as top: 48px, left: 64px, right: 71px and bottom: 37px. The standard size of the rectangle was applied to all images to perform the crop & scale DA approach. However, very few images about 2-5 lost the important data from images i.e. tumor object and these images were concluded from the dataset.

In the first transformation layer of DTL program, parameter "action" was set to "resize", "width & height" of images was set to "256x256" and "aspect ratio" was set to "true" to keep the quality of the image. To increase the size of the original dataset, the resulting crop & scale dataset was merged with the original images' dataset.

VIII. Blur

To obtain the Gaussian blur dataset, a blur effect was added to all images of the original dataset. In the transformation layer of the DTL program, the value of parameter "action" was set to "blur". Parameter "sigma" that indicates the density of blur effect was set to "2,3" as 2 for minimum and 3 for maximum. To increase the size of the original dataset, the resulting Gaussian blur dataset was merged with the original images' dataset. Table 1 shows the details of the datasets created from the original dataset by applying the different DA approaches.

YOLO version 3 was obtained from the darknet library (30). The architecture of YOLO v3 model consists of 106 convolutional layers and 1024 filters. To extract the features and to reduce the output channels from multiple images during the training phase, the model used 3x3 and 1x1 filters respectively. Along with the convolutional filters, the model also consists of some skip connections like residual to skip the layers of non-linear processing. In the configuration parameters, we have set and tested the different range of parameters including Learning Rate (LR) starts from 0.1, 0.01, 0.001 & 0.0001, epochs from 8, 10, 12, 15 & 20, batch size for training and validation from 5, 8, 10, 15 & 25 to get the better training parameters. Finally, the LR was set to 0.0001, the number of epochs was set to 15 and the batch size for training and validation was set to 12 because accuracy was showing higher by setting these parameters. So, the same configurations were set for all other datasets. The same YOLO v3 architecture and configuration was implemented for all augmented and original images datasets. YOLO v3 model was trained on the GPU NVIDIA Tesla K80. As we used the high computational machine so each dataset took 45–55 minutes to get trained. Fig. 2 shows the architecture of YOLO v3 model.

3. RESULTS

To analyze the results of the DA approaches, eight different experiments were performed. The images were divided into 80% of training and 20% of the evaluation ratio. However, we have extracted the separate 50 images for testing purposes from the TCIA data source. These images were not part of the training dataset. The annotation shape was set as a rectangle that belongs to a defined class named “Low-Grade Glioma Tumor”. For each experiment training loss, validation loss, Intersection Over Union (IOU) and test accuracy were calculated. The purpose of calculating training loss, validation and IOU was to overcome the training/validation error. Finally, each model was evaluated on the test accuracy results.

The descriptions of the results are given below:

Experiment 1: Original Dataset

The experiment was performed on the original dataset. Results showed that during training, the highest training loss was 4.58 and the lowest training loss was 0.14. At the epoch 15th, training loss was 0.25 and evaluation loss was 1.23. Training loss was high at epoch zero but it showed some improvement at each next epoch. The training loss showed some stability in loss after the 13th epoch. Similarly, evaluation loss was stable at epoch 1 to 5 but slightly increased between 5 to 9 epochs. At 12th and 14th epoch evaluation loss was high. The model achieved the best training loss and evaluation loss at epoch 13th. On completion of training, IOU was calculated as 0.72 at epoch 15th. The trained model on original dataset images achieved 68% accuracy.

Experiment 2: Vertical Flip Dataset

The experiment was performed on the vertical flip

dataset. Results showed that during training, the highest training loss was 2.75 and the lowest training loss was 0.08. On completion of training, training loss was 0.08 and evaluation loss was 1.33. Training loss was high at epoch zero but showed some improvement at each next epoch. Similarly, evaluation loss showed fluctuations throughout the model training. The average training loss was 0.43 and evaluation loss was 0.96 from epochs 1 to 15th. On completion of training, IOU was calculated as 0.72 at epoch 15th. The trained model on the vertical flip dataset achieved 70% accuracy.

Experiment 3: Horizontal Flip Dataset

The experiment was performed on the horizontal flip dataset. Results showed that during training, the highest training loss was 3.45 and the lowest training loss was 0.07. On completion of training, training loss was 0.07 and evaluation loss was 0.84. Training loss was started as high at epoch zero but showed some improvement at each next epoch. Training loss showed some stability in loss after epoch 13th. Similarly, evaluation loss was not much stable. The average training loss was 0.46 and evaluation loss was 0.96 from epoch 1 to epoch 15th. On completion of training, IOU was calculated as 0.80 at epoch 15th. The trained model on the horizontal dataset achieved 72% accuracy.

Experiment 4: Noise Dataset

The experiment was performed on the noise dataset. Results showed that during training, the highest training loss was 3.45 and the lowest training loss was 0.12. On completion of training, training loss was 0.12 and evaluation loss was 1.80. Training loss was high at epoch zero but showed improvement at each next epoch. However, training loss was not stable at all. Similarly, evaluation loss was not stable throughout the model training due to noise added in each image of the noise dataset. The average training loss was 0.48 and evaluation loss was 0.94 from epoch 1 to 15th. On completion of training, IOU was calculated as 0.77 at epoch 15th. The trained model on the noise dataset showed 60% accuracy.

Experiment 5: Rotate at 90o Dataset

The experiment was performed on rotating at 90o dataset. Results showed that during training, the highest training loss was 4.00 and the lowest training loss was 0.06. On completion of training, training loss was 0.06 and evaluation loss was 0.87. Training loss was high at epoch zero but showed improvement at each next epoch. Similarly, evaluation loss showed fluctuated loss curves throughout the model training. The average training loss was 0.43 and evaluation loss was 0.94 from epoch 1 to epoch 15. On completion of training, IOU was calculated as 0.76 at epoch 15th. The trained model on rotating at 90o dataset achieved 92% accuracy.

Experiment 6: Rotate at 180o Dataset

The experiment was performed on rotating at the 180o dataset. Results showed that during training, the highest training loss was 3.49 at the start of the training and the lowest training loss was 0.06. On com-

Figure 2. Architecture of YOLO v3 Model

pletion of training, training loss was 0.09 and evaluation loss was 0.86. Training loss started as high at epoch zero but showed improvement at each next epoch. Training loss showed some stability in loss after epoch 13th. The evaluation loss was not much stable at all. The average training loss was 0.46 and evaluation loss was 0.9 from epoch 1 to 15th. On completion of training, IOU was 0.80 at epoch 15th. The trained model on rotating at the 1800 dataset achieved 96% accuracy.

Experiment 7: Crop & Scale Dataset

The experiment was performed on crop & scale datasets. Results showed that during training the highest training loss was 4.88 at the start of the training. On completion of training, training loss was 0.06 and evaluation loss was 0.60. Training loss was high at epoch zero but it showed some improvement at each next epoch. Training loss showed some stability after epoch 11th. Similarly, evaluation loss was not showing much stability throughout the training. The average training loss was 0.56 and evaluation loss was 0.77 from epoch 1 to 15th. On completion of training, IOU was calculated as 0.84 at the epoch 15th. The trained model on crop & scale dataset achieved 83% accuracy.

Experiment 8: Gaussian Blur Dataset

The experiment was performed on the Gaussian blur dataset. Results showed that during training the

Sr. #	Dataset	Avg. training loss from epochs 1-15	Avg. evaluation loss from epochs 1-15	Training loss at epoch 15th	Evaluation loss at epoch 15th	IOU	Test Accuracy
1	Rotate 180o	0.46	0.90	0.09	0.86	0.80	96%
2	Rotate 90o	0.43	0.94	0.06	0.87	0.76	92%
3	Crop & Scale	0.56	0.77	0.06	0.60	0.84	83%
4	Horizontal Flip	0.46	0.96	0.07	0.84	0.80	72%
5	Vertical Flip	0.43	0.96	0.08	1.33	0.80	70%
6	Original	0.59	1.00	0.25	1.23	0.72	68%
7	Shear	0.54	0.80	0.06	0.68	0.78	68%
8	Gaussian blur	0.44	0.72	0.11	0.46	0.72	66%
9	Noise	0.48	0.94	0.11	1.80	0.77	60%

Table 2. Summary of Experiments

Authors	Classification Problem	Methods	Data Augmentation	Accuracy
Shanchen Pang et al. (19)	Cholilithiasis and gallstone from CT scan images	You Only Look Once (YOLO)	No	92.7%
Joseph Redmon et al. (27)	Natural objects	You Only Look Once (YOLO)	No	65.5%
Data Augmentation & YOLO v3	Low-grade glioma Brain tumor from MRI scan images	You Only Look Once (YOLO)	Yes	96%

Table 3. Comparison of results with prior studies

highest training loss was 2.98 and the lowest training loss was 0.05. On completion of training, training loss was 0.11 and evaluation loss was 0.46. Training loss was higher at epoch zero and showed some improvement at each next epoch but it was not showing much stability in loss throughout the model training. Similarly, evaluation loss was not stable throughout the training. The reasons for these fluctuations were due to the blur effect added in all images of the Gaussian blur

dataset. The average training loss was 0.44 and evaluation loss was 0.72 from epoch 1 to 15th. On completion of training, IOU was calculated as 0.72 at epoch 15th. The trained model on the blur dataset achieved 66% accuracy.

Experiment 9: Shear Dataset

The experiment was performed on the shear dataset. Results showed that during training the highest training loss was 4.40 and the lowest training loss was 0.06. On completion of training, training loss was 0.06 and evaluation loss was 0.68. Training loss was higher at epoch zero but showed some improvement at each next epoch. Training loss curves showed some stability in loss between epochs 11th to 15th. Similarly, evaluation loss curves were not showing much stability at all. The average training loss was 0.54 and evaluation loss was 0.80 from epoch 1 to epoch 15th. On completion of training, IOU was calculated as 0.78 at epoch 15th. The trained model on the shear dataset achieved 68% accuracy. Table 2 given below shows the summary of all experiments.

4. DISCUSSION

The test results showed that the experiment performed on image rotation at 180° and 90° datasets achieved the highest test accuracy results among other techniques which indicate that these datasets have many variations (as tumor objects at different positions) in images than the other applied approaches such as image flipping. However, the similarity between the shape and size of the glioma object in dataset images persisted.

The second-best accuracy achieved by the crop & scale experiment indicates that variations in shape were also increased when each annotated object was cropped and scaled up but this dataset had fewer variations than the rotation (180° & 90°) datasets. Similarly, the results obtained by the experiments performed on the original, vertical flip and horizontal flip achieved almost the same accuracies indicating that these datasets did not have many variations as tumor objects at different positions in the images such tumor object positions that appeared in images after applying rotation techniques. The accuracies achieved by experiments performed on the original and shear dataset indicates that the original dataset achieved low accuracy because of the low volume of images in the dataset. However, shear dataset achieved low accuracy due to the significant negative changes appeared in the object shape when the shear effect was applied on the images as it changed the original shape of brain tumor by shifting the upper part of an image to the right side and the lower part to the left side. The lowest accuracy was achieved by the experiments performed on Gaussian blur and noise datasets, which indicates that effects such as blur & noise that were added in the images, disturbed the quality of the images. It also indicates that by adding these effects, the features of the glioma object in the images were changed.

Different research studies and experiments on med-

ical imaging were performed by the researchers using the ML approaches. Shenzhen, Pang et al. (19) identified the cholelithiasis and classified the gallstone from CT scan images using YOLO model. Study results showed that consequently, the highest accuracy was achieved as 92.7% in identifying the granular gallstone in their study. Another study by Joseph Redmon et al. (27) evaluated YOLO model for the prediction of natural objects. The network was trained on ImageNet dataset and the study results showed that background and localization error was 4.7% and 19% respectively, and the correct object detection accuracy was 65.5% with processing speed at 45 frames per second. Table 3 shows a comparison of current study results with prior studies.

5. CONCLUSION

ML is a rapidly growing domain that achieved exceptional success in solving our daily life problems. In this paper, to enhance the low volume of the medical imaging dataset, eight different data augmentation approaches were evaluated by the training state of the art ML model YOLO version 3 on each dataset individually under the same configuration environment. The study results showed that rotate at 180° and rotate at 90° are significantly performed better as data enhancement techniques for medical imaging. However, the crop & scale approach also achieved a good result (83%) but its accuracy was less than the rotation techniques. In the future, other approaches such as synthetic DA using GAN will be explored on MRI scans of medical imaging.

- **Author's Contribution:** Each author gave substantial contribution in acquisition, analysis and data interpretation. Each author had a part in preparing article for drafting and revising it critically for important intellectual content. Each author gave final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved..
- **Conflict of Interest:** None declared.
- **Financial Support and Sponsorship:** Nil.

REFERENCES

1. Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. In Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS). 2014; (2): 3104–3112
2. Taigman Y, Yang M, Ranzato and Wolf. Deep Face: Closing the gap to human-level performance in face verification. In 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2014; 1701–1708.
3. Hinton G, Deng L, Yu D. et al. Deep neural networks for acoustic modelling in speech recognition: The shared views of four research groups. IEEE Signal Processing Magazine (SPM). 2012; 29(6) :82–97.
4. Ronneberger O, Fischer P, Brox. U-Net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention MICCAI. 2015; 234–241.
5. Fu GS, LS. YH. Machine Learning for Medical Imaging. J HealthcEng.2019; 1–2.
6. J, R. You Only Look Once: Unified, Real-Time Object Detection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR).2016; 779–788.
7. Mikołajczyk AGM. Data augmentation for improving deep learning in image classification problem.International Interdisciplinary PhD

- Workshop. Institute of Electrical and Electronics Engineers Inc.2018; 117–122.
8. Andersson EBR. Evaluation of Data Augmentation of MR Images for Deep Learning. 2018. Retrieved from: <https://lup.lub.lu.se/student-papers/search/publication/8952747>. Accessed on: 09-March-2019
9. Medium. (n.d.). Retrieved from: Data Augmentation | How to use Deep Learning when you have Limited Data–Part 2: <https://medium.com/nanonets/how-to-use-deep-learning-when-you-have-limited-data-part-2-data-augmentation-c26971dc8ced> Accessed on 09-March-2019
10. Zhang YD, Dong Z, Zhang YD, Dong Z. Special Issue on “Medical Imaging; Image Processing II.” Technologies. 2018; 6(2): 39.
11. Cancer. (n.d.). Retrieved from Definition of tumor–NCI Dictionary of Cancer Terms–National Cancer Institute. Retrieved from: <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/tumor> Accessed on: 15-May-2019
12. WebMD. (n.d.). Retrieved from Brain Tumors–Signs and Symptoms Not to Ignore. Retrieved from: <https://www.webmd.com/cancer/brain-cancer/brain-tumors-in-adults#1> Accessed on: 20-Jun-2019
13. TCIA. (n.d.). Retrieved from TCIA – The Tumor Cancer Imaging Archive. Retrieved from: <https://www.cancerimagingarchive.net/> Accessed on: 20-Jun-2019
14. Jonathan Hui. Real-time Object Detection with YOLO, YOLOv2 and now YOLOv3. Retrieved from: https://medium.com/@jonathan_hui/real-time-object-detection-with-yolo-yolov2-28b1b93e2088. Accessed on: 23-Jun-2019
15. Towards Data Science. (n.d.). Evolution of Object Detection and Localization Algorithms. Retrieved from:<https://towardsdatascience.com/evolution-of-object-detection-and-localization-algorithms-e241021d8bad>. Accessed on: 23-Jun-2019
16. Towards Data Science. (n.d.). Retrieved from YOLO–You only look once, real time object detection explained. Retrieve from: <https://towardsdatascience.com/yolo-you-only-look-once-real-time-object-detection-explained-492dc9230006>. Accessed on: 23-Jun-2019
17. Medium. (n.d.). Retrieved from Real-time Object Detection with YOLO, YOLOv2 and now YOLOv3. Retrieved from: https://medium.com/@jonathan_hui/real-time-object-detection-with-yolo-yolov2-28b1b93e2088. Accessed on: 25-Jun-2019
18. Zhu Z, Albadawy E, Saha A, Zhang J, Harowicz MR, Mazurowski MA. Deep learning for identifying radiogenomic associations in breast cancer. *Comput Biol Med*. 2019; 109: 85–90.
19. Pang Pang S, Ding T, Qiao S, Meng F, Wang S, Li P, et al. A novel YOLOv3–arch model for identifying cholelithiasis and classifying gallstones on CT images. *Song T, editor. PLoS One*. 2019; 14(6): e0217647. Available from: <http://dx.plos.org/10.1371/journal.pone.0217647>
20. Antropova N, Huynh BQ, Giger ML. A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. *Med Phys*.2017; 44(10): 5162–5171. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28681390>
21. Suk HI, Lee SW, Shen D, Alzheimer’s Disease Neuroimaging Initiative. Deep ensemble learning of sparse regression models for brain disease diagnosis. *Med Image Anal*. 2017; 37: 101–113.
22. Liu M, Zhang J, Adeli E, Shen D. Landmark-based deep multi-instance learning for brain disease diagnosis. *Med Image Anal*. 2018; 43: 157–168.
23. Paul R, Hawkins SH, Balagurunathan Y, Schabath MB, Gillies RJ, Hall LO, et al. Deep Feature Transfer Learning in Combination with Traditional Features Predicts Survival Among Patients with Lung Adenocarcinoma. *Tomogr (Ann Arbor, Mich)*. 2016; 2(4): 388–395.
24. Liu J, Wang D, Lu L, Wei Z, Kim L, Turkbey EB, et al. Detection and diagnosis of colitis on computed tomography using deep convolutional neural networks. *Med Phys*. 2017; 44(9): 4630–4642.
25. Xue Y, Zhang R, Deng Y, Chen K, Jiang T. A preliminary examination of the diagnostic value of deep learning in hip osteoarthritis. He H, editor. *PLoS One*. 2017; 12(6). Available from: <http://dx.plos.org/10.1371/journal.pone.0178992>
26. Islam MT, Aowal A, TahseenMinhaz A, Ashraf K. Abnormality Detection and Localization in Chest X–Rays using Deep Convolutional Neural Networks. 2017. Available from: <https://arxiv.org/pdf/1705.09850.pdf>
27. Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; 2017; 6517–6525. Available from: <http://ieeexplore.ieee.org/document/8100173/>
28. Bosse S, M. D. (2016). A deep neural network for image quality assessment. *IEEE Computer Society In: Proceedings–International Conference on Image Processing*. 2016; 3773–3777.
29. Supervisely. (n.d.). Retrieved from Supervisely–Web platform for computer vision. Annotation, training and deploy. Retrieved from: <https://supervise.ly/> Accessed on: 07-July-2019.
30. Darknet. (n.d.). Retrieved from Darknet: Open Source Neural Networks in C. Retrieved from: <https://pjreddie.com/darknet/> Accessed on: 07-July-2019.