Brain Tumor Detection using Multiple Instance Learning Technique

MSc Research Project
Data Analytics

Diksha Arvind Chaudhary
Student ID: x18184898

School of Computing
National College of Ireland

Supervisor: Dr. Rashmi Gupta

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Diksha Arvind Chaudhary
Student ID:	x18184898
Programme:	Data Analytics
Year:	2020
Module:	MSc Research Project
Supervisor:	Dr. Rashmi Gupta
Submission Due Date:	28/09/2020
Project Title:	Brain Tumor Detection using Multiple Instance Learning Technique
Word Count:	6189
Page Count:	20

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Diksha Arvind Chaudhary
Date:	27th September 2020

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Brain Tumor Detection using Multiple Instance Learning Technique

Diksha Arvind Chaudhary
x18184898

Abstract

Brain tumors are the growth of abnormal cells in the brain, which affects millions of people and causes death. Diagnosis of brain tumor along with patient care is crucial. The proposed research aims to identify tumor in brain MRI scans by an automated model. We use a novel method for the detection of brain tumor, which involves multiple instance learning (MIL) based on attention mechanism. Transfer learning based on pre-trained models such as DenseNet121 and InceptionV3 are applied by utilizing BRATS dataset to compare the results. We test the model using evaluation metrics like accuracy, sensitivity, specificity, precision, recall, and F1-score for improving the results. Impressive results are obtained by the proposed system and the results show that MIL is effective in comparison with other models. The results show that the proposed methods study the region of interest by itself and can detect tumors in brain MRI. The result shows that the proposed model is reliable and can be used in the diagnosis of tumor by neuroradiologists for further patient treatment.

1 Introduction

Brain tumors are the accumulation of abnormal cells that affect a certain area of the brain. The Figure 1 depicts the MRI of a healthy brain on the left and brain containing tumor on right. In Figure 1 we can see the tumor appears like a white block in the MRI image. Brain tumors are among the world's leading causes of death. Benign and malignant are two major types of brain tumours. Benign tumors are noncancerous and known as low-grade gliomas (LGG) whereas malignant are cancerous and are high-grade gliomas (HGG). Low-grade gliomas do not affect the other neighboring normal cells of the brain, but high-grade gliomas affect adjacent cells (Amin et al.; 2020). High-grade gliomas are difficult to detect, and patients with gliomas require urgent attention. Survival rate of patients with high-grade gliomas is around two years, and it requires quick medical treatment. The patients with low-grade gliomas have a survival rate of many years, and therapy is often delayed as they are non-malignant (Amin et al.; 2020). Diagnosis of type and grade of brain tumor along with optimum treatment of patients is essential. Medical imaging plays a critical role in the diagnosis of brain tumors in medical science. Researchers depend on medical imaging for the detection of brain tumors (Febrianto et al.; 2020). Expert radiologists and doctors used the medical imaging for an accurate diagnosis of tumors. Magnetic resonance imaging (MRI) is an imaging technique used in the detection of brain tumors that provides the best visualization of an internal anatomical

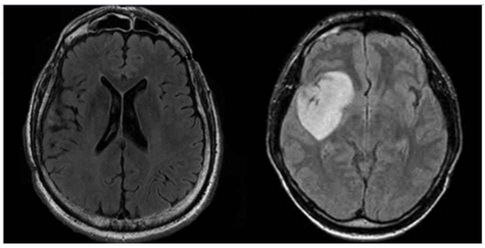

Figure 1: Healthy Brain MRI on Left and Brain Tumor MRI on Right

structure of brain. MRI captures sensitive and specific details of any abnormality in the brain with significant details. Medical imaging helps doctors to analyze and also visualize any abnormalities in the brain which helps in diagnosis (Goswami and Bhaiya; 2013). The Figure 2 shows the visual representation of the BRATS dataset file using anatomical image plotting. The plot shows different parts of brain MRI using axial cut direction; also we enhanced the image quality by changing the contrast level of the image to view the tumor availability in the image. Early detection of brain tumors can help improve the survival

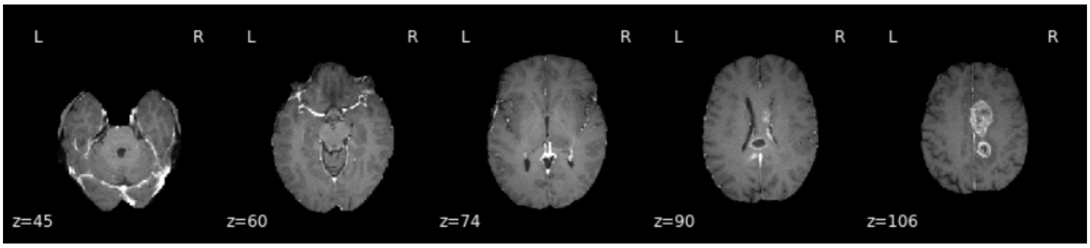

Figure 2: Different parts of MRI scan to detect brain tumor using BRATS Dataset

rate of patients. MRI method is the most preferred method by doctors than radio waves and field gradients, because the evaluation of various body parts is better by using MRI. MRI can detect abnormalities present in brain rather than other methods (Goswami and Bhaiya; 2013). The researchers use different imaging modalities for identifying tumor tissue changes that give different biological information such as FLAIR MRI, T1, T2, MRSI (Menze et al.; 2014). Machine learning algorithms are popular in analysis and aid neuroradiologists in diagnosis and surgery planning. Machine learning algorithms help in interpreting the medical images as stated by Soltaninejad et al. (2018).

The diagnosis of patients with brain tumors remains challenging despite a huge advancement in medical research. Because some gliomas are difficult to segment due to low contrast, we show an example of low contrast MRI in Figure 3. Tumours are often confused by the system because of the presence of tissues in the brain. Various segmentation techniques and multi-class classification methods are present, but still, there are difficulties in the detection of tumors because of variation in the size of tumors and blur borders. Segmentation of MRI images is the major challenge while detection of brain tumors because the tumor can be different in shape, size and location.

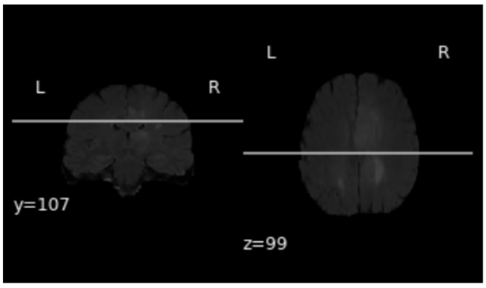

Figure 3: Low Contrast Image of Brain MRI

1.1 Research Question

How well can the Multiple Instance Learning (MIL) detect brain tumors using the BRATS dataset?

1.2 Research Objectives

Table 1: Research Objectives

Obj	Description	Metrics
1	Critically evaluate the different methods focusing on issues and methods used from 2013 to 2020	NA
2	Data Analysis of the images to get insights and overcoming challenges for incorporating Multiple instance learning	NA
3	Designing an architecture and implementing Attention based MIL. Evaluation of results.	Accuracy, Sensitivity, Specificity, Precision, Recall, F1-score
4	Implementing DenseNet121 and evaluating the results	Accuracy, Sensitivity, Specificity, Precision, Recall, F1-score
5	Implementing InceptionV3 and evaluation of results	Accuracy, Sensitivity, Specificity, Precision, Recall, F1-score
6	Comparison of the models and discussing the findings	Accuracy, Sensitivity, Specificity, Precision, Recall, F1-score

The aim of the research is to study and test the state-of-the-art methods for the classification of tumors. The state-of-the-art incorporates the MIL pooling method based on attention mechanism (Ilse et al.; 2018). The state-of-art performed on histopathology datasets for breast cancer detection and colon cancer detectionYousefi et al. (2018). The results outperformed other methods mentioned in the paper and looking at the success of state-of-the-art, the aim of this research is to apply the attention-based MIL technique on the BRATS dataset Menze et al. (2014); Bakas et al. (2018, 2017) for detection of tumor and test the model using different performance metrics. It is still difficult to identify which techniques are most efficient and which algorithm yields better results in identifying the tumors (Menze et al.; 2014). The major challenge is difficulty in classification of tumors because of various factors like variable size, shape, blur borders of tumors and also because of image abnormalities. The motivation of this research is to implement an automated method using deep learning and multiple instance learning for the detection of tumors despite these challenges.

This paper is organized : section 2 presents the related work performed on this research topic. We explain the research method in section 3. In section 4, we discuss the design specification of multiple instance learning model (MIL). The implementation of the state-of the-art and novel approach to detect brain tumor from MRI images is given in section 5 and evaluation of the models in section 6. We conclude this research work in section 7 and propose the future work.

2 Related Work

Brain tumor detection using MRI image classification has gained a lot of attention from researchers since the last few decades. The researchers have implemented different techniques to enhance the detection of brain tumors for diagnosis and classification of tumors. Goswami and Bhaiya (2013) discussed the difficulty in the classification of tumors because of the variance and complexity of the tumors. For better segmentation of the MRI image, researchers have used the unsupervised learning techniques such as Self-Organizing Map (SOM) approach along and Independent Component Analysis (ICA) as image reduction technique (Goswami and Bhaiya; 2013). The accuracy attained by this approach was around 98 percent and suggested neural network techniques provide promising results.

Researchers revealed disagreement in the manual annotation of sub-regions of tumors by implementing quantitative evaluation (Menze et al.; 2014). No single algorithm proved the best in segmenting, and the fusion of several algorithms proved too beneficial. The authors suggested an enormous potential in improvement to achieve better performance in detecting outliers patients by getting additional information such as scans are pre or post surgery and likewise. Naz et al. (2017) presents another review study, providing a brief insight of methods and contributions of different researchers for better results.

The researchers Soltaninejad et al. (2018) identified automated and accurate segmentation of MRI image as a major challenge. They used a novel 3D framework using supervoxel learning method for segmentation in MRI and diffusion tensor images (DTI). They applied the methods to two different datasets, a clinical dataset and BRATS 2013 dataset. This achieved promising results and a close match with expert annotations for different grades of tumors. It suggests addition of features from multimodal images to improve the segmentation accuracy. Azizi et al. (2018) proposed a recurrent neural network for detection of cancer using Temporal enhanced ultrasound (TeUS). The proposed model achieved highest accuracy in separating tumor from benign tissues. Another research by Siar and Teshnehlab (2019), used CNN for detecting brain tumors. A method based on a combination of feature extraction and the accuracy achieved around 99 percent. The results showed that softmax classifier has the best accuracy.

Qureshi et al. (2019) stated the importance of automated brain tumor detection by using DWT and PCA. They suggested that the KNN proved to be more beneficial than the other classifiers, and they can do more advancement for feature extraction and reduction to integrated systems to achieve accurate results. Mohamed Shakeel et al. (2019) proposed a method using machine learning based on neural networks back propagation method that enhances the exactness of the location of the tumor to decrease the chances of deaths. The system is further integrated with wireless imaging sensor that provides warm data of tumor to the radiologists. The results achieved were accurate and better than other classifiers. Amin et al. (2020) proposed a fusion process for structural and texture information. The method used discrete wavelet transform (DWT) which was fed to CNN for differentiating tumors. It is suggested that the work in future for the fusion of different modalities like CT image and PET for analysis of classification results.

Bi et al. (2019) presents the artificial techniques that address the clinical problems related to cancer detection. They discussed the major challenges in accurate detection of brain tumors. Convolutional neural networks (CNN) are used in many studies by Irsheidat and Duwairi (2020), Naser and Deen (2020), Febrianto et al. (2020), Al-Hadidi et al. (2020) in which the results achieved satisfactory accuracy. Irsheidat and Duwairi (2020) implemented an artificial CNN on the model. They analyzed the model using matrix operation and mathematical formulae. Further, the model got the probability of tumors, which attained an accuracy of around 96 percent. Naser and Deen (2020) proposed CNN using U-net and pre-trained VggNet for classifying low-grade gliomas in clinical application. Febrianto et al. (2020) did a comparison of two different classification models and achieved an accuracy of 93 percent. Al-Hadidi et al. (2020) used pixel clustering for segmenting gliomas in the research. The approaches and researches mentioned an enormous potential to expand the approaches in the deep learning field to achieve good accuracy in the classification of brain tumors.

Quellec et al. (2017) proposed a new framework using MIL in medical image and video analysis, in which MIL proved to solve such problems with accurate results. Yousefi et al. (2018) proposed a framework for mass detection in two-dimensional (2D) slices in digital breast tomosynthesis (DBT) data. Carbonneau et al. (2018) discuss the need of multiple instance learning stating the use of benchmarking to overcome the problems by MIL algorithm. They provide the survey results for 16 different MIL algorithms, providing a brief information about the impact and problems associated with the variation in performance of MIL algorithms. The research on MIL led to further study in Alzheimer diseases using brain MRI. Kavitha et al. (2019) presented a deep learning framework using U-net like CNN and MIL for identifying patterns in AD patients. The results proved effective in identifying the patterns in AD scans and medical images. Deep convolutional neural network (DCNN) is used, which learns complex patterns and MIL with randomized trees for classification. The approach performed much better than CAD systems. MIL proved to be effective, and researchers suggested more investigation on the usage of MIL with deep learning for clinical applications.

As mentioned by Bhattacharjee et al. (2020), in the medical field the data obtained is from real-world scenarios and not from experimental results and therefore there is inaccuracy, imbalance, inconsistency in the data. They take the decisions on very little information rather than the entire history of the patients, and therefore they need a robust model for extracting significant information from low-quality data. In the medical field the image is weakly annotated i.e., the image is labeled by a single label benign or malignant. It does not label all pixels and roughly defines the region of interest. We find the Multiple instance learning is a suitable machine learning algorithm in such cases where the data is weakly annotated. If the brain MRI is labeled as malignant, then only a certain portion in the image contains a tumor. It becomes very important for the model to learn which patches i.e. pixels represent the tumor. Multiple instance learning uses bag-level classification where a bag contains many instances whose labels are unknown. The goal of MIL to predict the bag label for an accurate medical diagnosis. We provide a summary Table 2, that reported relevant research studies for this paper.

Table 2: Summary details of the related work for brain tumors detection and MIL

Author(s)	Objectives	Research Design	Keywords	Findings
Goswami and Bhaiya (2013)	Brain tumor diagnostics in image in different phases	Independent Component Analysis (ICA) is used for feature extraction, k-mean clustering method for segmentation	Neural Networks, ICA, SOM, k-means clustering	SOM approach with ICA proved to be efficient with accuracy of 98.6 percent. The dataset used consisted of 70 images
Menze et al. (2014)	Reviews the setup and results applied to 65 MR scans of patients with low grade and high-grade gliomas	Quantitative evaluation of different algorithms	Image segmentation, BRATS, oncology/tumor	Fusion of good algorithms showed the opportunity for further improvement. Dice score ranged from 74 percent to 85 percent
Naz et al. (2017)	A review study of different methods applied on different dataset	Discussed the challenges and issues in segmentation and detection of tumor in brain	Machine learning	A great need for deep machine learning and image processing is identified to save human life. Diagnosis acceleration can be beneficial in terms of time and help to physicians
Soltaninejad et al. (2018)	Analysis of brain images using isotropic and anisotropic components from diffusion tensor imaging (DTI)	Proposed a novel supervoxel learning method using multimodal brain MRI dataset	Brain tumour segmentation, Diffusion tensor imaging, Multimodal MRI, Random forests, Supervoxel Textons	Promising results with dice score 0.84 were achieved which were a close match to the manual expert segmentation
Siar and Teshnehlab (2019)	Using CNN to prompt disease detection and improve quality of life and life expectancy	CNN is applied to brain MRI and accuracy is obtained using classifiers like Radial Basis Function, Decision tree and softmax classifier	Brain tumor, deep neural network, Convolutional neural network, magnetic resonance imaging, feature extraction	The best accuracy in CNN is achieved by using softmax classifier with 99.12 percent accuracy
Qureshi et al. (2019)	Using supervised classification technique using DWT and PCA for detecting type of brain	SVM, k-NN, Naïve Bayes, LDA classifiers are applied to reduced features	Image segmentation, MRI classification, Image Processing, Tumor detection, Feature extraction, Feature reduction	The method proved to be beneficial and rapid. K-NN and LDA classifier are more accurate, RBF SVM proved to be least accurate
Amin et al. (2020)	A fusion process with discrete wavelet transform (DWT) with Daubechies wavelet kernel is used for tumor region	Partial differential diffusion filter (PDDF) is used for filtering. Global thresholding method is applied to CNN for segmentation	Sequences CNN DWT Global thresholding Filter	Fusion of images proved to be better than sequences on datasets
Irsheidat and Duwairi (2020)	A model based on artificial CNN analyzed by matric operation and mathematical formulas for predicting tumor existence	The probability of existence of tumor is obtained by training over MRI	deep learning, convolutional neural networks, image classification, brain tumor detection	The method proved to be beneficial and rapid. K-NN and LDA classifier are more accurate, RBF SVM proved to be least accurate
Naser and Deen (2020)	To achieve medical need using combination of deep learning CNN using U-net and pre-trained Vgg16	Combination of CNN and fully connected classifier for grading of tumor is implemented for classifying LGG in grade II and grade III for clinical applications	Brain tumor, Segmentation, Classification, Grading, Glioma, Deep learning, Magnetic resonance imaging	Dice similarity of 0.84 and accuracy 0.92 achieved by model
Febrianto et al. (2020)	Image classification using deep learning to find whether the tumor is cancerous or not	CNN techniques are applied on brain MRI images for image classification. Two models are implemented and classification results are compared	CNN	Trained CNN mode achieved the accuracy of around 93 percent
Tong et al. (2014)	Propose the use of MIL in detection of Alzheimers Disease (AD)	To overcome the disease label ambiguity, MIL is adapted	Classifical, Structural MRI, MIL, Alzeimers Disease	Classification accuracy of 89 percent is achieved and suggests that MIl can be used in neurodegenerative disease
Quellec et al. (2017)	Reviews strategies for modelling MIVA and MIL specific algorithms	Different experiments and meta-analysis are conducted on medical image and video datasets for MIL discussions	Medical image analysis, medical video analysis, multiple-instance learning (MIL)	MIL proved to ideal solution with more accurate results in many cases for MIVA tasks
Yousefi et al. (2018)	The framework was designed for 2D slices	Patterns were learned using DCNN and MIL was applied for to classify 2D slices	Digital Breast Tomosynthesis (DBT), DCNN, MIL	Framework performed much better than existing CAD systems
Carbonneau et al. (2018)	A survey of different types of MIL problems along with variation in performance	An experiment for comparison of 16 techniques by addressing each category	Multiple instance learning, Weakly supervised learning, Classification	Discussed the necessity of using benchmarking after observing problems that impact MIL algorithms. MIL clustering and regression needs more investigation

Author(s)	Objectives	Research Design	Keywords	Findings
Ilse et al. (2018)	A new attention based mechanism is used as a permutation invariant operator	Weighted average is used as an aggregation function in MIL technique, attention based MIL pooling	Deep MIL, Attention based mechanism	Best results compared to other MIL methods on histopathology dataset.
Heo et al. (2019)	Detecting tuberculosis in chest using deep learning	Feature extraction is done using VGG19, DenseNet121, InceptionV3	Deep learning, CNN, tuberculosis	DenseNet121 achieved accuracy around 0.94 and InceptionV3 around 0.95
Kavitha et al. (2019)	A new framework using U-net and MIL for identifying patterns for AD	Unet like 2D CNN for feature capture and logistic regression with MIL to learn the region of interest	Multi-instance, CNN, regression, attention, multi-class	The method achieved higher performance than conventional methods. Use of MIL efficiently identified the pattern in AD scans
He et al. (2019)	Proposed a new method MIDCN based on multiple instance learning for image classification	Detects bag labels by features and similarities between multiple instances	Lung cancer, Image classification, MIL	The results outperformed the other MIL methods
Al-Hadidi et al. (2020)	Brain tumour segmentation for better efficiency in diagnosis and improved prediction rate for treatment planning	An integrated framework using pixel clustering based on deep learning techniques for segmenting gliomas brain tumours	Brain tumour, CNN, superpixel, image segmentation, Pixel clustering	The approach achieved satisfactory results and can be expanded in broad computer vision approaches
Liang and Zheng (2020)	Deep learning framework for detection of pneumonia in children	Comparison of different CNN architectures is done for medical images	DenseNet121, InceptionV3, pneumonia	Accuracy of DenseNet121 is 0.81 and InceptionV3 is 0.85

The tumor assessment and diagnosis relies on the visual examination and is represented by different computer analyzes. Various work is carried out to identify brain tumors using machine learning and deep learning techniques. In clinical applications, the deep learning techniques show promising results in terms of qualitative analysis for tumors. The use of convolutional neural network resulted successfully in feature extraction from MRI. Multiple instance learning (MIL) has proven ideal for the identification of cancer in weakly annotated medical data. The outcomes of breast cancer, lung cancer detection and detection of Alzheimer's disease using MIL are remarkable. Applying Multiple instance learning to detect brain tumors is the novelty of this research as there is no prior work for brain tumor detection using MIL. After reviewing the literature, the research is inspired to build the MIL framework using convolutional and fully connected layers.

3 Methodology

3.1 Introduction

The research follows the data analytics methodology, which is CRISP-DM for detection of a brain tumor. Manual identification of brain tumors is tedious and time consuming, and having an automated system for identification can be a game-changer. The research focuses on implementation of an automated system to help doctors in diagnosis and further improvement in the rate of survival. As mentioned in related work, deep learning is very effective in detecting brain tumors. In this research, Multiple instance learning approaches and transfer learning based pre-trained models like DenseNet121 and InceptionV3 are applied to detect the tumors. In Figure 4, we illustrate the process flow for the research methodology that comprises related activities which are carried out during the research project. The subsections mentioned further provide details of each activity.

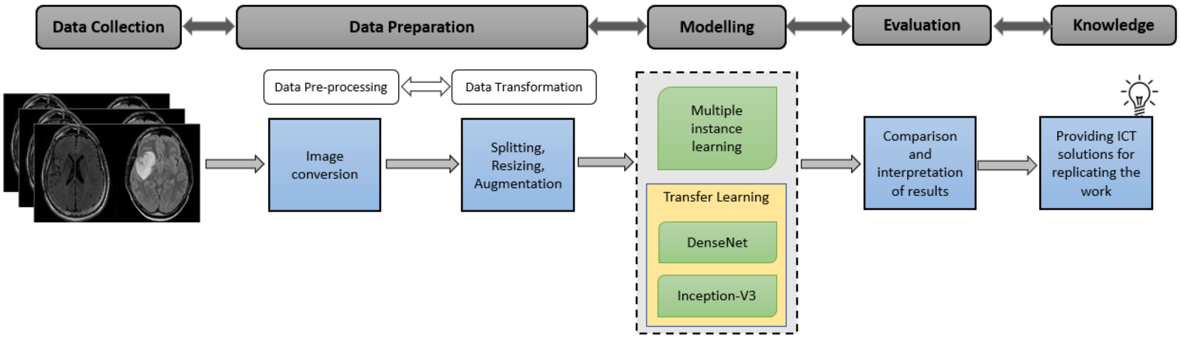

Figure 4: Project design process flow

3.2 Data Collection

We conduct the research on large dataset provided by "Multimodal Brain Tumor Image Segmentation Challenge (BRATS)" [1] (Menze et al.; 2014) (Bakas et al.; 2017) (Bakas et al.; 2018). Because of the variable appearance and form of tumor, segmentation of brain tumors from multi-modal imaging data is one of the challenging tasks in medical imaging research. Data set has the following volumes: T1 MRI, T1 MRI, T2 MRI and T2 FLAIR MRI. Data collection comprises MRI scans, and experts and professional neuroradiologists manually labeled the images in the dataset as shown in Figure 5. The whole tumor is represented as the yellow patch visible in Figure 5 labeled as FLAIR. We represent the core tumor as the red patch visible in Figure 5 labeled as T2 FLAIR. The enhanced structures of tumor are visible in light blue, and the core is visible in green. It also shows a combined segmentation of sub-regions: non-enhancing solid core (red), edema (yellow), necrotic core (green) and enhancing core (blue) [2]. The image patches from top left corner shows tumor structures in different modalities and the right image shows the final combined label for it. Figure 5 is taken from Amin et al. (2020).

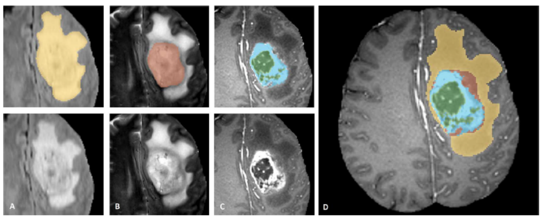

Figure 5: Expert radiologists' manual annotation seen.

3.3 Data Preparation

3.3.1 Image conversion

The images obtained in data collection are in NIfTI (.nii.gz) format which is neuroimaging file format. These are 3D images of brain MRI and image dimensions are 240 X 240 X

[1] http://braintumorsegmentation.org/
[2] https://sites.google.com/site/braintumorsegmentation/home/brats2015

155. In python NiBabel library supports handling of neuroimaging file formats. In the data preparation step it was mandatory to convert the nii.gz images to .png or .jpeg format for avoiding challenges in the modelling phase. The initial preprocessing and data exploratory analysis was done on .nii.gz file using NiBabel library. There were following challenges that were faced while handling nii.gz file. The image is 240 X 240 X 155 dimensions, there are 155 different slices for an image of .nii.gz format since it is a 3D image. Converting these images to numpy arrays in python is demanding for model building. The image generator in keras does not support the .nii.gz file directly and changing the image generator code would be tedious and complex. So, converted the nifti images to jpg format. Using mathematical operations (Bohaju; 2020) extracted the features of the BRATS2015 dataset and converted to .jpeg format. The converted images contain 3,764 files and a csv file containing the image id and class (healthy or tumor).

3.4 Modelling

The data which is pre-processed is used by three different models in this step. Multiple Instance Learning approach is the novelty of this research for detecting brain tumors. Transfer learning based on pre-trained models like DenseNet and InceptionV3 are implemented using the processed data.

3.4.1 Pre-trained models (Transfer Learning)

The literature review proved that convolutional neural networks have been very effective in terms of detection of brain tumors (Siar and Teshnehlab; 2019; Naser and Deen; 2020; Febrianto et al.; 2020). Adding more layers and creating dense structure can improve the accuracy and thus we used transfer learning in the research. Transfer learning is using pre-trained models that are already trained on some datasets such as ImageNet and its specialization can be used to classify image as healthy or containing tumor. We show the concept of transfer learning in Figure 6. The pre-trained models are already trained on weights of ImageNet and are used to detect tumors.

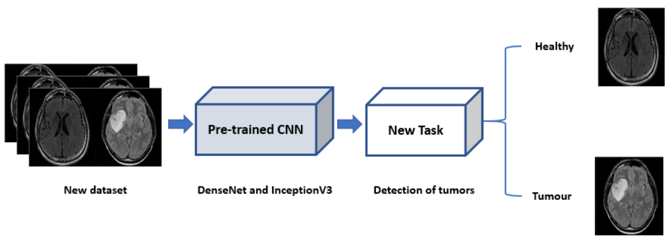

Figure 6: Transfer Learning in the research.

- **DenseNet121 :** We selected a deeper, denser and accurate DenseNet121 network in the research (Huang et al.; 2019). DenseNet shows continuous improvement in precision, with an increase in the number of parameters without over-fitting and degrading of efficiency. Most interesting advantage of DenseNet is the increase in feature reuse, which makes the model very efficient. Thus, DenseNet was considered in this research after some modifications. The pre-trained DenseNet121 model was trained on ImageNet weight in the research.

- **InceptionV3 :** We adapted inceptionV3 in this research after modifications. The pre-trained model InceptionV3 was loaded with weights on ImageNet. InceptionV3 is known for high quality results along with low computational cost even on relatively small dataset (Szegedy et al.; 2016).

3.4.2 Multiple Instance Learning (MIL)

Image classification and segmentation are two vital parameters in the medical field. Multiple instance learning can do both of the tasks by handling weakly labeled data. MIL uses bags of instances which comprise a label and does the medical diagnosis by predicting the label of the bag. MIL proved to be an ideal solution in medical image analysis (Quellec et al.; 2017; Kavitha et al.; 2019) and also beneficial in identifying patterns in medical images and scan. Figure 7 shows the architecture of Multiple instance learn-

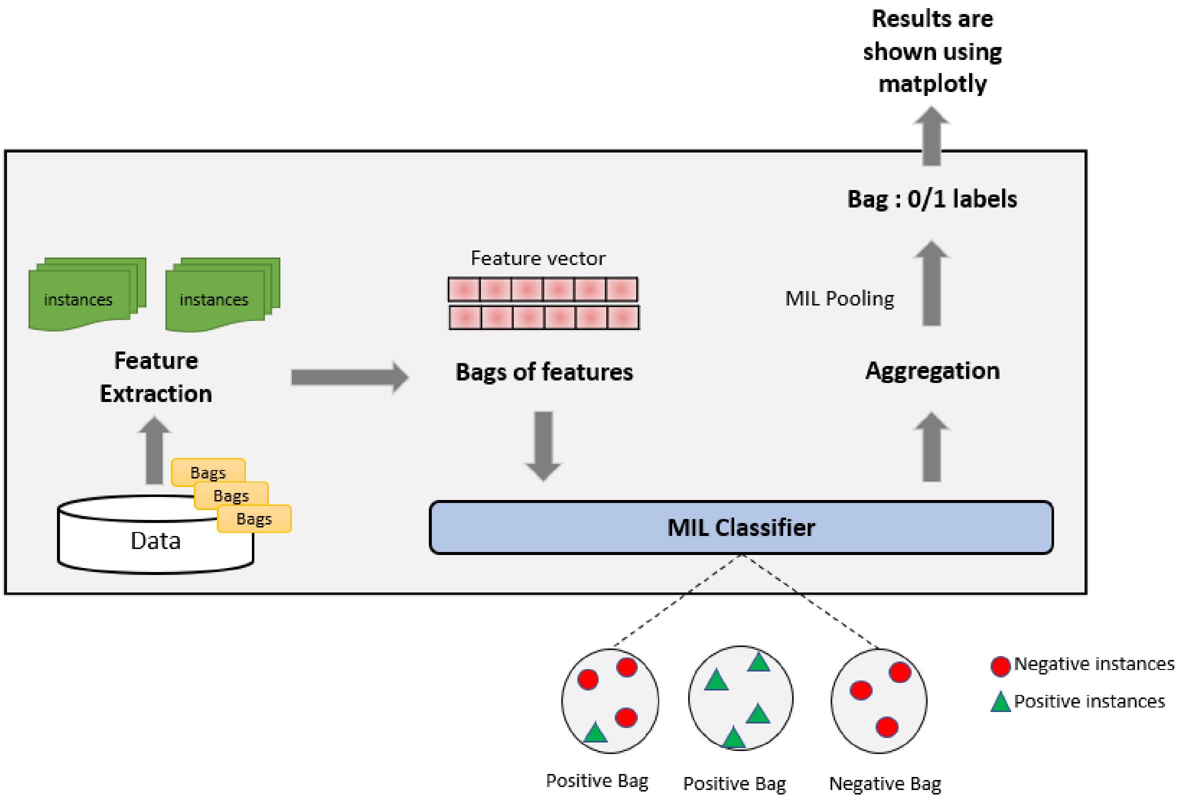

Figure 7: Design architecture - MIL

ing. Every image is a bag of instances, and each bag has only one label assigned to it. The goal of MIL is to identify the label of a bag which is a medical diagnosis Ilse et al. (2018). The feature extraction and transformations are done using convolutional layers and fully connected layers using neural networks. Bags contain different instances and it also shows a feature vector as bags of feature in the Figure 7 before feeding it to MIL classifier. It transforms the instances to a low dimension, creating a feature vector. The feature vector is fed to the MIL classifier, which distinguishes it as positive and negative bags. We consider the bag probability negative if all instances in the bag are negative and we consider the bag probability positive, even if a single instance is positive.

The bag probability is set using permutation invariant function, which is applied to embedded instances. The problem addressed in MIL is no dependency between the instances and we can carry it out using either instance level approach or embedding level approach (Ilse et al.; 2018). The embedding approach is used where the individual labels are not known and using instance level approach can lead to insufficient training. In embedding approach joint representation of bag and do not lead to additional bias.

Ilse et al. (2018) mentioned the disadvantages of max operator and mean operator in MIL pooling methods. The research also further suggested the use of attention based MIL pooling using a weighted average of instances calculated by neural networks. The attention mechanism has proved very accurate in predicting the region of interest (ROI) and thus provides a high performance which is inline as mentioned in the state-of-art (Ilse et al.; 2018).

3.5 Evaluation

We used confusion matrix to evaluate the model and calculated the evaluation metrics such as accuracy, loss, sensitivity, specificity, and F1-score. Sensitivity (True positive rate) is calculated to evaluate the ratio of tumor images classified as tumors and specificity (True negative rate) for correctly identifying the healthy images. The evaluations are further discussed in section 6.

4 Design Specification

The MIL framework addresses the issues of no dependency between the instances inside a bag. The method uses the general procedure for modeling bag label probability i.e. bag score function, which is calculated by the fundamental theorem of symmetric function. There are three steps involved in the fundamental theorem of symmetric function, i) instances are transformed into low dimension embedding ii) aggregation using permutation invariant function iii) final transformation of bag prediction.

4.1 Neural Network

In most of the MIL scenarios, we consider the instances as features, but in image classification, we require the extraction of features. Therefore, a class of transformations for feature extraction is done using parameterised functions of neural networks which perform three steps involved in the fundamental theorem of symmetric function. In both an instance-based and embedding-based approach, the parameterised functions transform the instances to low dimensional. The usage of the neural networks increases the flexibility of the model and also optimizes the unconstrained objective function. We train the complete model using the back-propagation flexibly.

4.2 Attention based MIL Pooling

MIL operators are differential and can be easily used in the architecture of neural networks. The max operator and mean operators are good for the instance-based approach but are inappropriate for embedded-instance based approach. We use a more adaptive pooling operator in the research which achieves better performance results. We base the MIL pooling on an attention mechanism which uses the weighted average of the instances. The low dimensional embedding (weighted average) is used from the weights from the neural network. The summation of weights should be equal to one so it can be invariant to the bag size. Thus, for proper gradient flow between the positive and negative values, a hyperbolic tangent (tanh) is used in MIL pooling. Further, it helps to recognize the differences between the instances.

The final representation of the bags is highly effective and adaptive using attention mechanisms (Ilse et al.; 2018). The attention mechanism for image captioning is widely used in medical imaging (Ilse et al.; 2018; Quellec et al.; 2017).

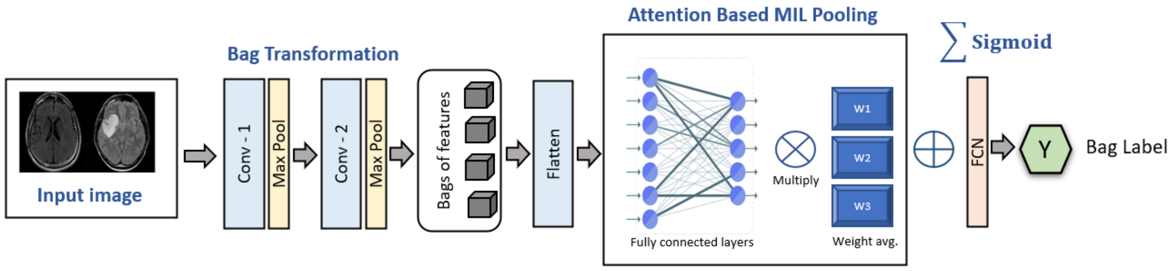

Figure 8: Deep MIL architecture using attention based mechanism

We show the architecture in Figure 8. The convolution layer is notated as conv, max pooling layer as max pool, and fully connected layer as FCN. We get the feature extraction after a set of convolutional and max pool layers as shown in the figure. The attention-based MIL pooling comprises a fully connected layer which is multiplied by the average weights got from neural networks output highlighted in the blue box in the architecture. In the end, an aggregation function sigmoid with one neuron is used to predict the bag label.

5 Implementation

We discuss the overall implementation of the research in this section. The initial setup, data transformation steps, implementation of the model are discussed in this section along with the tools used.

5.1 Environment Setup

The machine configuration used for the implementation was Windows 10 with a 64-bit operating system and 8 GB of RAM with NVIDIA GeForce MX250 graphics. We used Python programming language for the implementation.

5.2 Data Transformation

The transformation of data is a process in which we extract the data according to the model requirement. The data collected comprised a folder containing 3,762 images and a .csv (comma-separated values) file comprising the image name and its category (0 - Non tumor and 1 - tumor). We created two separate folders based on the tumor category from the data frame by comparing the image name of the file with the image name in the data frame. We perform the file operations using a high-level file operations library available in python known as shutil [3].

Splitting of files in test, train and validation was performed differently for transfer learning and MIL as both the models had different requirements. The split ratio was 70 percent train data, 25 percent validation, and 5 percent test images in a case of transfer learning. For MIL the splitting of images in train bags, test bags and validation bags was done using K-fold cross validation by sklearn model selection library [4].

We perform data augmentation techniques like geometric image transformation using cv2 library in python. We rotate images at the center with an angle of 90 degree using cv2 getrotationmatrix2D and warpAffine functions. We also flipped the images horizontally and vertically using cv2 flip function. We then pass this augmented data to the MIL model in the model building phase. The data augmentation for the pre-trained model is done in image generator function by specifying rotation range = 90, width and height shift, horizontal flip and re-scale parameters.

5.3 Transfer Learning

We passed the transformed data to transfer learning architectures DenseNet121 and InceptionV3. We discuss the implementation in following sub-sections 5.3.1 and 5.3.2.

5.3.1 DenseNet121

Dense convolutional network (DenseNet) comprises a dense connectivity pattern that takes into account the feature maps of all preceding convolutional blocks Huang et al. (2019). ImageDataGenerator function was used for data augmentation before model building. The DenseNet model from Keras application [5], was used with the DenseNet121 function which uses pre-trained weights on ImageNet. For feature extraction, we applied the global average pooling mode to the last convolutional block. The activation function used was softmax. Adam optimizer with learning rate 0.0001 was used and model.fit_generator() was used for executing the model.

5.3.2 InceptionV3

InceptionV3 pre-trained model is available in keras applications [6]. The model was fine-tuned using the pre-trained weights of ImageNet. We used similar parameters in both DenseNet121 and InceptionV3 as discussed earlier.

5.4 Implementation of Multiple Instance Learning

We used an attention-based mechanism for implementing MIL. We stored the data on google drive. We mount the drive using google colab platform and we extracted data during run-time. The process of data loading was carried out using glob function and stored as a positive and negative path. All the paths created by the test and train indexes used kfold function from sklearn library in python (Ilse et al.; 2018). The value of n_fold

[3]https://docs.python.org/2/library/shutil.html
[4]https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html
[5]https://keras.io/api/applications/densenet/
[6]https://keras.io/api/applications/inceptionv3/

= 2 for cross validation evaluation. We generate the dataset in the form of a dictionary with keys as train and test. We considered this dataset as train_bags and test_bags.

The function generate_batch read the image using cv2 library and converts the image to a numpy array. This converted image was appended with the file name, label and stored in batches. The two convolution layers were used to extract the features of instances with batch_size = 36 and batch_size = 48 and filter size 4 X 4 and 3 X 3, respectively. The activation function ReLU and max-pooling layer of filter size (2,2) was used. The neural network further comprises two fully connected layers with activation function ReLU and layer regularization for optimization with weight_decay = 0.005. MIL attention layers were calculated by dot product and low dimensional mechanism (average of weights). We extract these weights from the neural network. The calculation further comprises a hyperbolic tangent (tanh) for proper gradient between positive and negative values.

We used the fully connected layer with sigmoid activation function as the last layer. We fed the instance features to this layer, which calculates the instance score by multiplying the weighted average. The MIL pooling layer aggregates this score into bag score (Wang et al.; 2018). We calculated bag accuracy and bag loss for the evaluation of the model. We calculated the bag accuracy using the ground truth of the bag and prediction of the bag. We calculated the bag loss by using binary cross-entropy loss for the bag prediction. The backend function from the keras library is used to perform these calculations.

6 Evaluation

In this section, we discuss a comprehensive analysis of transfer learning models and attention-based Multiple instance learning (MIL). After the successful implementation of the models, it is necessary to check the performance of the model on a validation set of images and test images. We take different evaluation metrics into consideration in this research. We explain the results and implications of tuned parameters using plots for accuracy and loss for each model. MIL is the novelty of this research and it was evaluated and compared with other models using validation accuracy, test accuracy, sensitivity and specificity. We explain the evaluation metrics for 100 epochs in this section.

6.1 Baseline Experiment : Transfer Learning

6.1.1 DenseNet121 - 100 epochs

Images were provided as training and validation images to the DenseNet121 model. The model was trained using 3,009 train images and 564 validation images. We trained the model using pre-trained weights from ImageNet and 100 epochs. The results of DenseNet121 on the dataset for classification of tumors resulted in good accuracy with low loss as shown in Figure 9. The test accuracy achieved by the model was 0.98 and validation accuracy was 0.96. The loss curve in the plot was low approaching zero and the accuracy of validation was going up with an increase in the number of epochs.

6.1.2 InceptionV3

The results of InceptionV3 achieved an accuracy of 0.94 on test data and validation accuracy of 0.93.

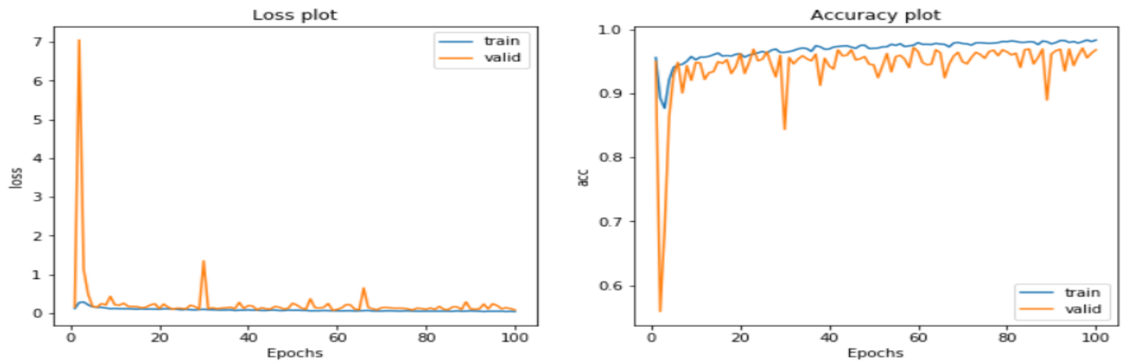

Figure 9: Accuracy and loss plots using DenseNet121

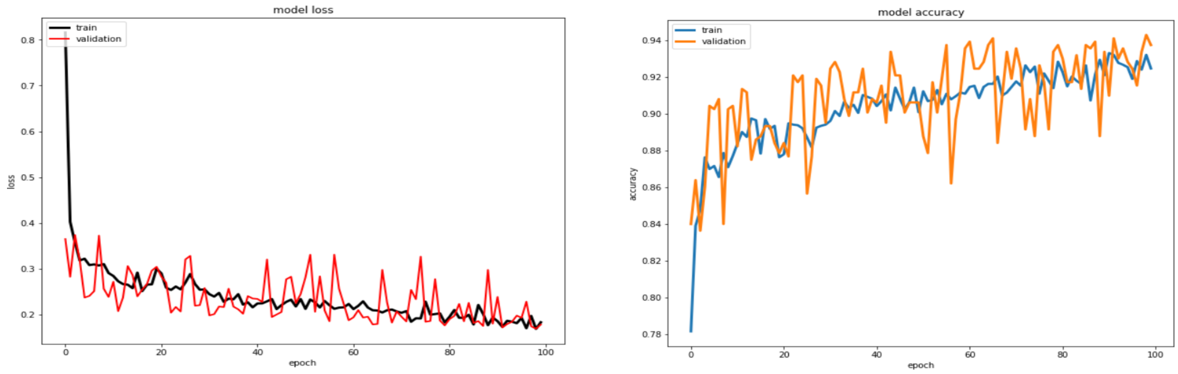

Figure 10: Accuracy and loss plots using InceptionV3

We show the accuracy and loss plot for the InceptionV3 model in the figure. The curve obtained was not smooth enough, but we can observe that the loss was decreasing and accuracy was going up as the number of epochs were increased. We can infer it from the plots that the accuracy in the detection of tumors was good for using the InceptionV3 model.

6.2 Newly Proposed Experiment : Multiple Instance Learning

Multiple instance learning is very efficient in learning from the instances and assigning bag labels for binary classification. We show the accuracy and loss plot for 100 epochs below. The key attributes used for evaluation of MIL is bag accuracy and bag loss. The bag accuracy computes the accuracy of a single bag and bag loss is computed using binary cross-entropy loss. We use ground truth of the bag and the predicted score of the bag for both. The MIL method achieved validation accuracy at 1 and test accuracy at 1 using BRATS dataset (Bohaju; 2020). The model learns from the instance with accuracy at 1 on the second epoch, which makes the model highly effective and fast. The loss decreases and tends to zero at early epochs. We used the early stopping callback in the code with patience 20, so if the model achieves the desired output and there were no further changes in the loss, the model stops. This optimization method makes the method fast.

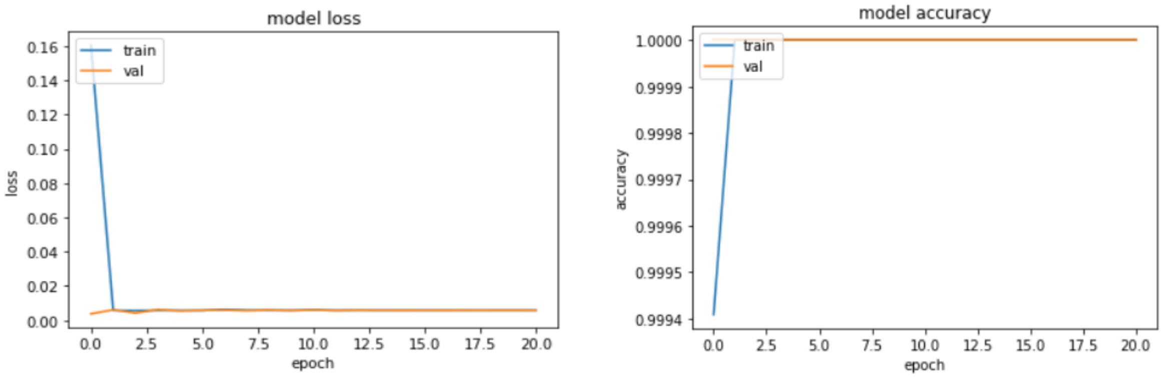

Figure 11: Accuracy and loss plots using MIL

K-fold cross validation approach was used for further evaluation of the model which makes the model results more generalized. Two-fold validation was performed in the model and we provide the result details in the Table Table 3.

Table 3: Two-fold cross validation

n_fold	runtime	loss	validation acc	test acc
k=1	11.31	0.00381	1	1
k=2	11.15	0.00371	1	1

The research project was related to the medical field and therefore, sensitivity and specificity along with accuracy define the reliability of the models. Sensitivity for this research specifies the ratio of tumor images classified as tumors. We show the sensitivity and specificity of the models obtained in Figure 12. The sensitivity achieved by DenseNet121, InceptionV3 and MIL is 0.51,0.95 and 1 respectively as shown in Figure 12. The specificity achieved by models are 0.40, 0.94 and 1 respectively. MIL performs better than other models by comparing the evaluation metrics.

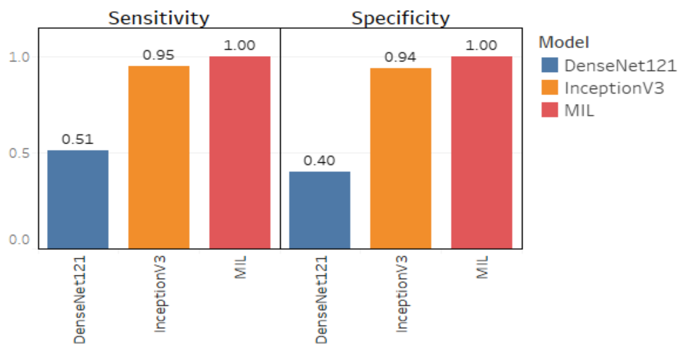

Figure 12: Sensitivity and Specificity of the models

6.3 Discussion

The research aims to evaluate the performance of the newly proposed approach comprising attention-based MIL technique after effective results obtained in the medical field discussed in related work. This approach was used in breast cancer and colon cancer detection, which we chose as state-of-the-art for this research (Yousefi et al.; 2018). We carried the comparison between the transfer learning models and MIL out for solving the problem of the detection of brain tumors from brain MRI (Ilse et al.; 2018; Kavitha et al.; 2019; Bhattacharjee et al.; 2020; Heo et al.; 2019; Liang and Zheng; 2020). The models are evaluated and we provide results in Table 4.

We compared the models using different evaluation measures such as accuracy, precision, recall, F1-score. The accuracy of DenseNet121 performed better than InceptionV3 on the dataset with the value of 0.96 as validation accuracy and 0.97 test accuracy. The residuals reuse from the previous layer yields constant improvement in the accuracy in DenseNet121 (Huang et al.; 2019). Considering other evaluation metrics such as precision, 47 % of time DenseNet121 correctly predicts tumor and 94 % of time InceptionV3 correctly identifies tumors. Similarly, recall and F1-score for DenseNet121 is 0.47 and InceptionV3 is 0.94. The large number of parameters and module based features of InceptionV3 makes it more effective for classification of tumor (Szegedy et al.; 2016). The novel approach MIL based on attention mechanism showed outstanding results with 1 accuracy for both validation and test data. MIL can learn the weakly supervised data more precisely than the transfer learning models for brain tumor detection. As per the state-of-the-art for this research, the accuracy achieved was around 0.95 for breast cancer and colon cancer detection data (Yousefi et al.; 2018).

Table 4: Comparison of models based on evaluation metrics

Model	Val acc	Test acc	Precision	Recall	F1-score
DenseNet121	0.96	0.97	0.47	0.47	0.47
InceptionV3	0.93	0.94	0.94	0.94	0.94
MIL	**1**	**1**	**1**	**1**	**1**

Sensitivity and specificity comparison shown in Figure 12 shows that all three models performed well for the detection of brain tumors. Looking at the performance metrics, the attention-based MIL performed better than the other two models with sensitivity and specificity of value 1. Also precision, recall and F1-score were calculated on the models and the results show that the MIL method is the best technique compared to others since MIL method adopted attention based MIL pooling technique that is highly effective and adaptive considering average weights of instances from convolutional layers. The time taken by the model to train is another important attribute for performance. We compiled all the models for 100 epochs to be consistent and better model comparison, DenseNet121 is slow compared to other models which took around 30150 secs for training data, InceptionV3 is fast which took 3025 secs and exceptional performance was achieved by MIL with 1320 secs shown in Table 5.

Table 5: Comparison of models based on training time

Model	Data	Training data	Training time (secs)
DenseNet121	3764	3009	30105
InceptionV3	3764	3009	3025
MIL	**3764**	**3384**	**1320**

After analyzing all metrics, MIL performance is impressive as it easily learns complex patterns in the instances. MIL can detect the tumor in brain MRI with high accuracy and less training time.

7 Conclusion and Future Work

The study aimed to detect a tumor in brain MRI using Multiple instance learning (MIL) approach based on the attention-based mechanism. We also used transfer learning techniques like DenseNet121 and InceptionV3 (Heo et al.; 2019; Liang and Zheng; 2020) for a comparison of the results with the newly proposed model (Yousefi et al.; 2018; Ilse et al.; 2018; Kavitha et al.; 2019; Bhattacharjee et al.; 2020). The ability and consistency of MIL are studied by applying an attention-based mechanism. The dataset is converted from NiFTi (.nii.gz) format to jpg format by using mathematical operations to overcome the challenges faced during the experiment implementation. The accuracy achieved by the model is exceptional and reliable compared to other transfer learning methods. The strength of the proposed method is that the model learns of the instances rapidly, which leads to high accuracy at early epochs. The limitation of the model is that the image conversion is required when the dataset contains images of .nii.gz format.

The future work involves overcoming the dataset challenge confronting the proposed work. The performance can be improved by applying the model to other publicly available large brain tumor datasets. We can use this research work presented as the groundwork for other research in the medical field. There is a tremendous scope for further work to build a reliable model on multiple instance learning in the medical field.

Acknowledgement

I am very grateful to Dr. Rashmi Gupta, my supervisor, for excellent support and valuable feedbacks throughout my research project. I sincerely thank her for constantly inspiring and motivating me for improving my research.

References

Al-Hadidi, M. R., AlSaaidah, B. and Al-Gawagzeh, M. Y. (2020). Glioblastomas brain tumour segmentation based on convolutional neural networks., *International Journal of Electrical & Computer Engineering (2088-8708)* **10**.

Amin, J., Sharif, M., Gul, N., Yasmin, M. and Shad, S. A. (2020). Brain tumor classification based on dwt fusion of mri sequences using convolutional neural network, *Pattern Recognition Letters* **129**: 115–122.

Azizi, S., Bayat, S., Yan, P., Tahmasebi, A., Kwak, J. T., Xu, S., Turkbey, B., Choyke, P., Pinto, P., Wood, B. et al. (2018). Deep recurrent neural networks for prostate cancer detection: analysis of temporal enhanced ultrasound, *IEEE transactions on medical imaging* **37**(12): 2695–2703.

Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J. S., Freymann, J. B., Farahani, K. and Davatzikos, C. (2017). Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features, *Scientific data* **4**: 170117.

Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Shinohara, R. T., Berger, C., Ha, S. M., Rozycki, M. et al. (2018). Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge.

Bhattacharjee, K., Pant, M., Zhang, Y.-D. and Satapathy, S. C. (2020). Multiple instance learning with genetic pooling for medical data analysis, *Pattern Recognition Letters* .

Bi, W. L., Hosny, A., Schabath, M. B., Giger, M. L., Birkbak, N. J., Mehrtash, A., Allison, T., Arnaout, O., Abbosh, C., Dunn, I. F. et al. (2019). Artificial intelligence in cancer imaging: clinical challenges and applications, *CA: a cancer journal for clinicians* **69**(2): 127–157.

Bohaju, J. (2020). Brain tumor.
URL: *https://www.kaggle.com/dsv/1370629*

Carbonneau, M.-A., Cheplygina, V., Granger, E. and Gagnon, G. (2018). Multiple instance learning: A survey of problem characteristics and applications, *Pattern Recognition* **77**: 329–353.

Febrianto, D., Soesanti, I. and Nugroho, H. (2020). Convolutional neural network for brain tumor detection, *MS&E* **771**(1): 012031.

Goswami, S. and Bhaiya, L. K. P. (2013). Brain tumour detection using unsupervised learning based neural network, *2013 International Conference on Communication Systems and Network Technologies*, IEEE, pp. 573–577.

He, K., Huo, J., Shi, Y., Gao, Y. and Shen, D. (2019). Midcn: A multiple instance deep convolutional network for image classification, *Pacific Rim International Conference on Artificial Intelligence*, Springer, pp. 230–243.

Heo, S.-J., Kim, Y., Yun, S., Lim, S.-S., Kim, J., Nam, C.-M., Park, E.-C., Jung, I. and Yoon, J.-H. (2019). Deep learning algorithms with demographic information help to detect tuberculosis in chest radiographs in annual workers' health examination data, *International journal of environmental research and public health* **16**(2): 250.

Huang, G., Liu, Z., Pleiss, G., Van Der Maaten, L. and Weinberger, K. (2019). Convolutional networks with dense connectivity, *IEEE Transactions on Pattern Analysis and Machine Intelligence* .

Ilse, M., Tomczak, J. M. and Welling, M. (2018). Attention-based deep multiple instance learning, *Conference Proceeding* .

Irsheidat, S. and Duwairi, R. (2020). Brain tumor detection using artificial convolutional neural networks, *2020 11th International Conference on Information and Communication Systems (ICICS)*, IEEE, pp. 197–203.

Kavitha, M., Yudistira, N. and Kurita, T. (2019). Multi instance learning via deep cnn for multi-class recognition of alzheimer's disease, *2019 IEEE 11th International Workshop on Computational Intelligence and Applications (IWCIA)*, IEEE, pp. 89–94.

Liang, G. and Zheng, L. (2020). A transfer learning method with deep residual network for pediatric pneumonia diagnosis, *Computer methods and programs in biomedicine* **187**: 104964.

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R. et al. (2014). The multimodal brain tumor image segmentation benchmark (brats), *IEEE transactions on medical imaging* **34**(10): 1993–2024.

Mohamed Shakeel, P., Tobely, T. E. E., Al-Feel, H., Manogaran, G. and Baskar, S. (2019). Neural network based brain tumor detection using wireless infrared imaging sensor, *IEEE Access* **7**: 5577–5588.

Naser, M. A. and Deen, M. J. (2020). Brain tumor segmentation and grading of lower-grade glioma using deep learning in mri images, *Computers in Biology and Medicine* p. 103758.

Naz, S., Hameed, I. A. et al. (2017). Automated techniques for brain tumor segmentation and detection: A review study, *2017 International Conference on Behavioral, Economic, Socio-cultural Computing (BESC)*, IEEE, pp. 1–6.

Quellec, G., Cazuguel, G., Cochener, B. and Lamard, M. (2017). Multiple-instance learning for medical image and video analysis, *IEEE reviews in biomedical engineering* **10**: 213–234.

Qureshi, A., Khan, K. B., Haider, H. A., Khawaja, R. and Yousuf, M. (2019). An ensemble classification-based methodology applied to mri brain images for tumor detection and segmentation, *International Conference on Intelligent Technologies and Applications*, Springer, pp. 606–615.

Siar, M. and Teshnehlab, M. (2019). Brain tumor detection using deep neural network and machine learning algorithm, *2019 9th International Conference on Computer and Knowledge Engineering (ICCKE)*, IEEE, pp. 363–368.

Soltaninejad, M., Yang, G., Lambrou, T., Allinson, N., Jones, T. L., Barrick, T. R., Howe, F. A. and Ye, X. (2018). Supervised learning based multimodal mri brain tumour segmentation using texture features from supervoxels, *Computer methods and programs in biomedicine* **157**: 69–84.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z. (2016). Rethinking the inception architecture for computer vision, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826.

Tong, T., Wolz, R., Gao, Q., Guerrero, R., Hajnal, J. V., Rueckert, D., Initiative, A. D. N. et al. (2014). Multiple instance learning for classification of dementia in brain mri, *Medical image analysis* **18**(5): 808–818.

Wang, X., Yan, Y., Tang, P., Bai, X. and Liu, W. (2018). Revisiting multiple instance neural networks, *Pattern Recognition* **74**: 15–24.

Yousefi, M., Krzyżak, A. and Suen, C. Y. (2018). Mass detection in digital breast tomosynthesis data using convolutional neural networks and multiple instance learning, *Computers in biology and medicine* **96**: 283–293.