Machine Learning with Applications 2 (2020) 100003

Contents lists available at ScienceDirect

Machine Learning with Applications

journal homepage: www.elsevier.com/locate/mlwa

# A Transfer Learning approach for AI-based classification of brain tumors

Rajat Mehrotra [a,*], M.A. Ansari [a,1], Rajeev Agrawal [b], R.S. Anand [c]

[a] Department of Electrical Engineering, School of Engineering, Gautam Buddha University, Greater Noida, 201308, India
[b] Department of Electronics & Comm. Engg., GL Bajaj Institute of Technology & Management, Greater Noida, 201306, India
[c] Department of Electrical Engineering, Indian Institute of Technology Roorkee, 247667, India

ARTICLE INFO

*Keywords:*
Image processing
Image classification
Deep Learning
Transfer Learning

ABSTRACT

Classification of Brain Tumor (BT) is a vital assignment for assessing Tumors and making a suitable treatment. There exist numerous imaging modalities that are utilized to identify tumors in the brain. Magnetic Resonance Imaging (MRI) is generally utilized for such a task because of its unrivaled quality of the image and the reality that it does not depend on ionizing radiations. The relevance of Artificial Intelligence (AI) in the form of Deep Learning (DL) in the area of medical imaging has paved the path to extraordinary developments in categorizing and detecting intricate pathological conditions, like a brain tumor, etc. Deep learning has demonstrated an astounding presentation, particularly in segmenting and classifying brain tumors. In this work, the AI-based classification of BT using Deep Learning Algorithms are proposed for the classifying types of brain tumors utilizing openly accessible datasets. These datasets classify BTs into (malignant and benign). The datasets comprise 696 images on T1-weighted images for testing purposes. The projected arrangement accomplishes a noteworthy performance with the finest accuracy of 99.04%. The achieved outcome signifies the capacity of the proposed algorithm for the classification of brain tumors.

## 1. Introduction

Brain cancer (tumor) can be characterized as abnormal and unrestrained development in the synapses. As the individual head is an inflexible and volume constrained, thusly, any startling development in the brain may influence a human capacity; besides, it might swell into various body parts and influence individual capacities (DeAngelis, 2001). As indicated by the WHO, in its cancer report Brain Tumor (BT) represents less than 2% of cancer's in humans; nevertheless, serious bleakness and difficulties are reported (Stewart & Wild, 2014). United Kingdom's cancer research corporation referenced that there are around five thousand two hundred causalities reported every year due to brain diseases and tumors within the skull in the United Kingdom (Brain, other CNS and intracranial tumours statistics | Cancer Research UK. (n.d.), 2020). BTs are arranged under two significant heads; The one that is developed within the brain is termed as a primary brain tumor and correspond to 70% of all BTs on the other hand tumors that swells into the brain from some other body parts are called secondary brain tumors and forms the residual 30%, majority of which are of malignant type (Tandel et al., 2019). The tumor area its type as well as where it is situated in the brain decides the type of treatment required. As a rule, surgery of the brain is deemed as the way of handling tumors (American Brain Tumor Association, Chicago, I.L., U.S.A., 2015). The most frequently occurring brain tumors are of Glioma types that incorporate approximately 30% of all BTs and Central Nervous System, and around 80% of all harmful BTs (Goodenberger & Jenkins, 2012). Amid various clinical advances, MRI gives data about the location and tumor size. Its working depends upon the activity of protons contained in an enormous magnetic field by maneuvering radiofrequency waves and recouping their stable state (Pereira et al., 2016) To decisively separate delicate tissues with high precision MR technology is quite proficient and is increasingly responsive to change in tissue solidity required for pathological consultation. The MRI images are classified into T1-weighted (T1-w) that are constantly utilized in non-invasive brain investigations. Since they portray high contrast and less debris they are referred for anatomy purposes. Whereas, T2-weighted (T2-w) are significant MR slices that are reasonable for perceiving the boundary structures of the medical images (Zimny et al., 2015). The major downside of the T2-weighted arrangement is that tumors in the brain, cerebrospinal fluid (CSF), and Gray Matter (GM) and are bind together. Pathologically, the utilization of these MRI arrangements is

The code (and data) in this article has been certified as Reproducible by Code Ocean: (https://codeocean.com/). More information on the Reproducibility Badge Initiative is available at https://www.elsevier.com/physical-sciences-and-engineering/computer-science/journals.

* Corresponding author.
*E-mail addresses:* rajat.mehrotra@glbitm.ac.in (R. Mehrotra), ma.ansari@gbu.ac.in (M.A. Ansari), rajkecd@gmail.com (R. Agrawal), anandfee@iitr.ac.in (R.S. Anand).
[1] SM IEEE

https://doi.org/10.1016/j.mlwa.2020.100003
Received 13 July 2020; Received in revised form 14 September 2020; Accepted 14 September 2020
Available online 3 October 2020

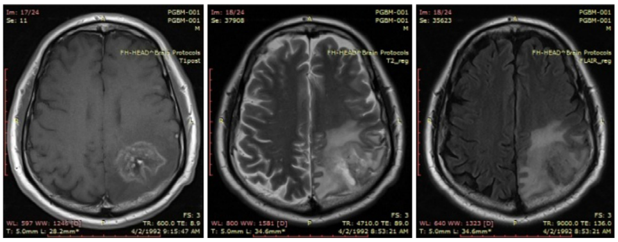

(a) (b) (c)

**Fig. 1.** MR slices, from (a) to (c)- T1, T2, and FLAIR.

fundamental in identifying BTs however these can create few problems in separating tumorous from non-tumorous zones besides grading (Bahadure et al., 2017). Accordingly, to explain the tumor boundary evaluated against a non-tumorous tissue on T1-w and T2-w pictures use a contrast medium are significant. BTs are sometimes confounded as they remain unimproved with contrast enhancement. Subsequently, the FLAIR procedure is utilized along with T2-w to show the unupgraded BTs (Jalab & Hasan, 2019). Fig. 1 shows the types of MR images generally utilized.

Deep Learning (DL) is a type of AI technique that emulates the functioning of an individual brain in data processing and generating prototype useful in making suitable choices. DL calculations make use of various non-linear layers that are well organized for extracting features from an image. The outcome of every ordered layer is the contribution of the following one, and that helps in information deliberation as we dive deep inside the system (Deng & Yu, 2014). Convolutional Neural Network (CNN) is a part of the DL family and is usually utilized in scrutinizing visuals and intended to entail negligible pre-processing (LeCun, 2015). It is motivated by the biological progression of the human brain (Matsugu et al., 2003) and used to deal with information that comes in groups (LeCun et al., 2015). Deep CNN was first used when LeCun et al. (1998) presented a DL network 'LeNet' for document identification in 1998. Many years later, it came up substantially when utilizing a DL network was utilized to identify images by making use of a pre-trained network (PTN) called AlexNet (Krizhevsky et al., 2012). It showcased remarkable outcomes when compared with other systems of that time. Later, its prosperity prompted back to back triumphs of CNNs in the area of DL. The primary points of interest of Convolutional Neural Networks are their ability to learn features and to give boundless precision as opposed to conventional AI techniques by increasing the number of samples used for training and hence leads to a much powerful and precise model (Litjens et al., 2017). In the design of Convolutional Neural Networks, the features are extracted by convolutional filters and as we dive deep, much more intricate features are mined. Extraction of features takes place by convolving small size filters with the patterns of input and thereafter determination of the most distinctive features and hence training the network for classification. Zacharaki et al. (2009) put forward a framework to identify glioma other than a classifying high level and low level utilizing Support vector Machine and k-Nearest Neighbor. They achieved a precision of 85% for multiple classifications and binary identification 88% accuracy is obtained. El-Dahshan et al. (2010) projected a technique to identify 80 BT images both abnormal as well as normal utilizing the Discrete Wavelet Transform (DWT) technique for feature extraction, PCA for feature reduction, and thereafter Artificial Neural Network (ANN) and k-NN to identify images with a precision of 97% and 98% individually. Cheng et al. (2015) presented a technique to upgrade the BT identification by using dilation of image and thereafter by parting them into sub-sections. They utilized three ways for feature extraction; Bag of Words (BOW), Gray Level Co-event Matrix (GLCM), and intensity histogram, and lastly accomplished the finest precision of 91.28%. Ertosun and Rubin (2015) utilized Convolutional Neural Networks to identify grades of glioma as well as high and low grades of glioma. They acquired exactnesses of 71% and 96% separately. Paul et al. (2017) utilized transverse BT images for training and create two principle methods for identification using CNN and accomplished the utmost exactness of 91.43%. Afshar et al. (2019) introduced CapsNet, a capsule network that encapsulates MR images of the brain with the boundary of the coarse tumor to characterize the BT and obtained a precision of 90.89%. Anaraki et al. (2019) put forward a network of two consolidated regulations to identify BT from MR images using Genetic Algorithms and Convolutional Neural Network. They achieved a precision of 90.9% and 94.2% in identifying glioma and its grades respectively. A Wavelet-based Auto Encoder utilizing ANN was presented by Chen et al. (2018) that breaks down the input image into an image with low resolution for classification. These images are then given to CNN as input for reducing complexity in computation without affecting the precision. Shalini et al. (2014) proposed a technique in which the weighted fuzzy system was utilized to separate BT from the image and to increase the segmentation process kernel matrix was utilized. It gave high effectiveness and precision when contrasted with some other widespread techniques. A successful neural system based BT identification procedure was put forward by Damodharan and Raghavan (2015) which concentrated on the segmentation of tissues in the brain. The said technique gave ideal effectiveness and precision in significance to tissues in the brain and segmentation of tumor, extraction of features and identification, etc. (Chelghoum et al., 2020) in their work utilized nine PTN which includes AlexNet, VGG19, GoogleNet, ResNet18, VGG16, ResNet50, ResNet-Inception-v2, ResNet101 and SENet for BT classification through Transfer Learning (TL). Firstly, they customized the three end layers of PTNs so as to acclimatize them to their task of classification. Subsequently, the fully connected layer in their originally taken PTNs is substituted by new layers, in which the types of BT is characterized by its output size. Lastly, they utilized TL based fine-tuned PTN for experimentation with MR data. Their elucidations verify that TL had provided results that are unswerving even with the dataset of small size. Their projected network outperformed the state-of-the-art techniques and achieved an accuracy of 98.71% for the classification of BTs. Rehman et al. (2020) conducted experiments utilizing three PTNs (AlexNet, GoogLeNet, and VGGNet) for classifying BTs such as glioma, meningioma, and pituitary. They then discover the TL methods that are, fine-tuning and freeze by means of MR slices of the BT dataset. They utilized these PTNs to dig out features through the use of TL techniques. Lastly, the classification of features is done through a support vector machine (SVM) and a log-based softmax layer. They have accomplished the utmost exactness of 98.69% through fine-tuned VGG16 network as compared to AlexNet and GoogLeNet. On the other hand in the freeze method of TL, the top most accuracy of 95.77% utilizes the freeze Conv5 layer of AlexNet as contrasted to its other layers as well as to all the architectural layers of VGG16 and GoogLeNet. Their experiment on the classification of BTs had attained the utmost exactness of 98.69% by utilizing a fine-tuned VGG16 network.

Ansari et al. (2020) presented a logical strategy for the identification of BTs. They centered around the suppression of noise, extraction of pivotal features dependent on GLCM, and lastly segmentation of BT dependent on Discrete Wavelet Transform (DWT) so as to enhance the output and wither the intricacy. They also performed Morphological activity for noise suppression that is developed because of the segmentation process. Support Vector Machine (SVM) is utilized as a classifier for calculating the precision of BT recognition. Their test results display a characterization precision of 98.91% that shows the viability of their proposed framework. Sultan et al. (2019) in his paper proposed a deep learning model dependent on a CNN to arrange distinctive BT types utilizing two openly accessible datasets. The previous one characterizes tumors into (meningioma, glioma, and pituitary tumor) and the other one separates between the three glioma grades (Grade II, Grade III, and Grade IV). The data incorporate 73 and 233 patients with an image sum of 516 and 3064 on T1-weighted contrast-enhanced images for both

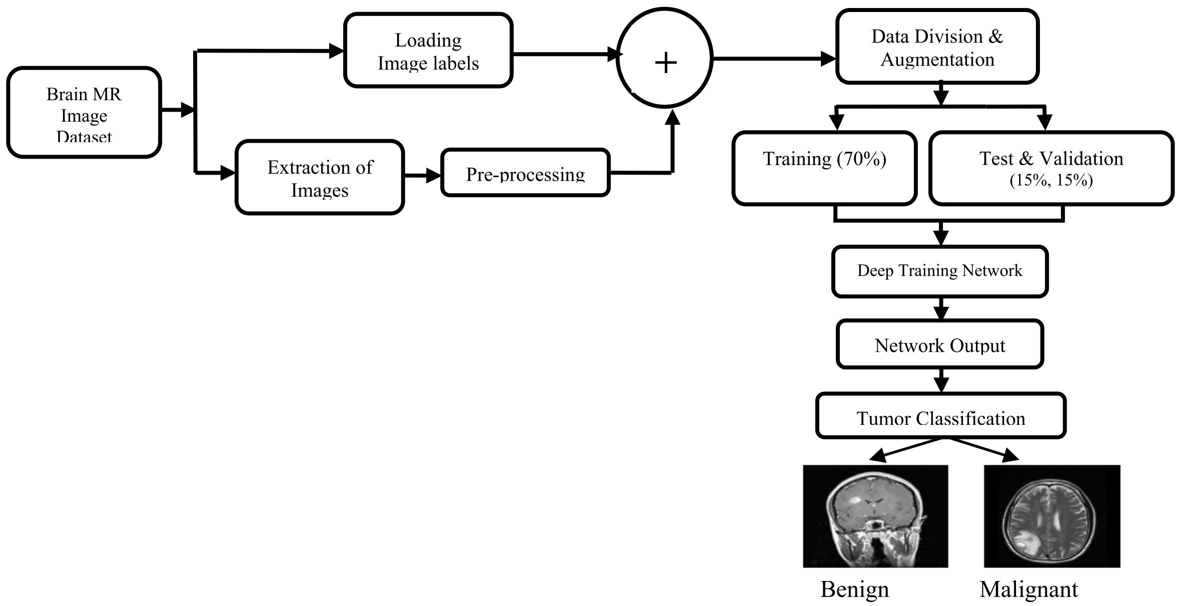



**Fig. 2.** The proposed structure of the Deep Learning system.

the datasets, separately. Their presented system structure accomplishes an exactness of 96.13% and 98.7%, separately. Kumar et al. (2019) presents, a structure dependent on Gray Wolf Optimization and Multi-class Support Vector Machine for the classification of BT in MR images. They showed that the Gray Wolf Optimization technique, utilized for reducing features, works better when contrasted with other calculations, for example, Firefly Algorithm and Particle Swarm Optimization when a Multiclass Support Vector Machine is utilized for the identification of BT into their various kinds. They achieved a classification accuracy of 95.23% with their work. Their work can be upgraded in different perspectives, for example, usage of other AI techniques for characterization. Abiwinanda et al. (2019) utilized five unique designs of CNN and achieved the utmost accuracy on the second design that consists of two convolutional layers, the max-pool layer, and ReLU pursued by hidden 64 neurons. They accomplished an accuracy of 98.51% on training and 84.19% on validation sets, separately. Abir et al. (2018) in their work communicated probabilistic neural system (PNN) for BTs classification. They utilized various pre-processing steps like filtering, resizing, image sharpening, and image contrast enhancement for acquiring features using GLCM. They accomplished the noteworthy precision of 83.33% in their presented work. Bakas et al. (2019) in their investigation surveyed the best-of-the-state Machine Learning strategies which are utilized for BT examination in multi-parametric magnetic resonance imaging scans, all through the end seven cases of the BraTS challenge, i.e., from 2012–18. In particular, they focused primarily on assessing diverse glioma sub-section through pre-operative multi-parametric magnetic resonance imaging scans using segmentation. Secondly in evaluating prospective tumor movement by temperance of longitudinal sub-region development of tumor, away from the utilization of the RECIST/RANO criterion, thirdly in foreseeing the general survival from pre-operative multi-parametric magnetic resonance imaging scans that went through total resection. And at last, they explored the test of recognizing the best Machine Learning calculations for every one of these in hand tasks, bearing in mind that along from being different on each test, the multi-institutional multi-parametric magnetic resonance imaging scans BraTS image set has also been a persistently advancing/developing dataset.

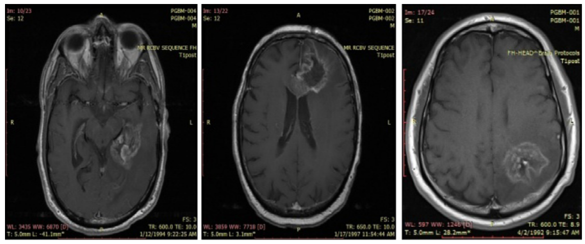
The key contributions of the presented research are the development of a strong and novel method for DL using Transfer Learning (TL) techniques for detecting and classifying BTs by extracting pivotal features on a standard dataset. Secondly to investigate the five unique DL models like AlexNet, GoogLeNet, ResNet50, ResNet101, and SqueezeNet utilizing MR images of BT and apply TL techniques on the given dataset. Thirdly to present an exhaustive evaluation of the performance of significant features that influence the fine-tuning of pretrained systems. And Lastly to perform the investigative comparison among the above-mentioned models using various vital parameters for BT identification. In the presented paper, Transfer Learning-based deep learning algorithms are accessed for identifying BTs into benign and malignant types accurately. The system is enhanced using various designs to attain the best possible structure. The succeeding part of the paper is arranged as; in Section 2, the methodology i.e. line of action is discussed. Section 3 is devoted to the outcome and its discussion followed by Section 4 that gives the conclusion of the paper.

## 2. Methodology

The purpose of this investigation is to enhance the exactness of brain MR image identification by using DL algorithms and Transfer Learning (TL) approach. TL is the assignment of utilizing the information given by a pretrained system to learn new models provided by new data. Calibrating a pretrained system with TL is usually a lot quicker and simpler than starting from basic. Utilizing pretrained DL systems empowers us to rapidly learn new jobs. We here investigate the five unique DL models like AlexNet, GoogLeNet, ResNet50, ResNet101, and SqueezeNet utilizing MR images of BT and apply TL techniques on the given dataset. Such pretrained CNN models are utilized to perform TL to extricate features that are visually distinguishable and essential. Finally, the classification of these features is done utilizing the softmax layer. The procedure of this investigation is presented in Fig. 2. It begins with the MR brain image dataset that was gathered and arranged into benign and malignant MR slices. The proposed strategy contains the accompanying stages: pre-processing, data division, and augmentation, DL based extraction of features, lastly the tumor type classification.

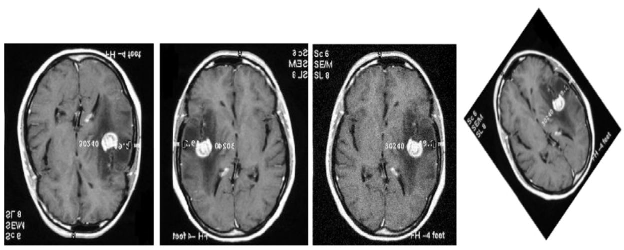
### 2.1. Experimental dataset

Brain MR dataset for evaluation of the projected work is acquired from The Cancer Imaging Archive (TCIA) Public Access repository (Clark et al., 2013). This assortment incorporates datasets from 20 patients with recently identified glioblastoma. We chose T1-weighted images as presented in Fig. 3. It contains a sum of 696 MR images, of which 224 images are labeled as benign and the 472 are malignant images. The image size of every image is set to 225 × 225 in JPG/JPEG format.



*(a)* *(b)* *(c)*

**Fig. 3.** T1-weighted MR images.



(a) (b) (c) (d)

**Fig. 4.** (a) Flipped image (b) Mirrored image (c) Noisy image (d) 45° rotated image.

### *2.2. Image pre-processing*

Before taking care of the images into the proposed structure, its preprocessing is done. The primary step is to downscale the raw MR image from 512 × 512 × 1 pixels into 225 × 225 × 1 pixels to diminish dimensionality calculations and support the system to show a superior outcome in much less time. At that point, image data is jumbled up before parting them to keep up the network to work on an unarranged dataset and forestall concentrating on a small set of the whole dataset. Image data is then separated into three segments; train, test, and validation with 70% for training and 15% for test and validation respectively. Lastly, image augmentation like flipping, mirroring, and rotation is done with the goal to create a surplus of data for the network which is typically utilized to abstain the network from overfitting and augment system toughness (Wong et al., 2016). Along with these image augmentations, some salt noise is also added for a grayscale deformation to the dataset. Fig. 4 represents the image after augmentation that comprises of flipping, mirroring, salt noise addition, and rotation of image by 45 degrees.

### *2.3. Extraction of features through deep learning*

Deep learning networks or the convolutional neural networks are a subcategory of the artificial NN. CNN's different layers along with the pooling layers are utilized as a means for mapping multi-dimensional MR images to yield an ideal outcome after they get trained (Lundervold & Lundervold, 2019). The benefit of using DL is that the system itself figures out how to do the feature extraction while getting trained. DL networks do the feature extraction by their own utilizing their kernels or convolutional filters. Furthermore, the Convolutional Layers (CL) have a bunch of small size filters. These filters are applied to each layer to create a tensor of features. Up to what distance the filter will move in each progression starting with one point onto the next is termed as 'stride'. Practically, one or two-pixel stride do good, more than this decays the output of CNNs altogether (Van der Burgh et al., 2017). For structuring a CNN for a particular assignment, it is necessary to comprehend the prerequisites to be met and in what way the information is given to the system. Additionally, while setting the stride it should be kept in mind that the output should yield a whole number but never a fraction. At times, zero padding's are required if the filter in the convolution layer does not cover all the images given as an input. This is done to keep the equivalent spatial measurements. The tensor of feature maps that are generated through a CL, is determined through a Rectified Linear Unit (RLU) in the activation layer. The RLU is the vastly utilized function for activation in DL systems which is utilized to restrain every negative value to zero in the feature map. The rectified features are connected to the pooling layers for downsizing the dimensions by creating minute non-overlapping areas as input and decide the only value for every region. Average and pooling are the two well-known function, that is often utilized by the pooling layers (Liu et al., 2018). To normalize the feature map, a batch normalization layer is commonly utilized after the activation layer to standardize feature maps. This normalization layer regulates the network and accelerates the training. The final convolution layer is succeeded by the fully connected layer (FC). Fig. 5 shows the DL architecture. The features derived from the last pooling or convolutional layers are converted to a 1-D vector and are linked with at least one dense layer to outline the system's last outcome (Yamashita et al., 2018). Typically the final layer of the system is used to be the Softmax layer which is used for the classification of images and is formulated as:

$$y_z = \frac{e^z}{\sum_{i=1}^{j} e^{z_j}} \tag{1}$$

Lastly, a classification layer is utilized that depends on cross-entropy to assess the loss due to classification and gives the last anticipated clear cut labels for every MR image. The estimation of loss is given by Eq. (4), where 's' indicates the labels given to target and r (x) is the output from the softmax layer.

$$H(s,r) = -\sum_{x} (s(x) * \log(r(x))) \tag{2}$$

We are utilizing and investigating five famous and best in class structures of CNN for identifying BT utilizing MR images of the dataset. The architecture of the proposed models: AlexNet, GoogLeNet, ResNet50, ResNet101, and SqueezeNet is presented in Fig. 5.

#### *2.3.1. Transfer learning based feature extraction using pretrained networks*

Transfer Learning (TL) is an inspiration of surmounting the models of remote learning and using informative knowledge obtained for one assignment to solve the similar ones. It is the assignment of utilizing the information given by a pretrained system to learn new models provided by new data. Calibrating a pretrained system with TL is usually a lot quicker and simpler than starting from basic. Utilizing pretrained DL systems empowers us to rapidly learn new jobs. Various scientists and data experts consider TL as an effective tool that can speed up our development towards AI (Sarkar, 2018). There is an obvious distinction between the conventional methodology used to train machine learning system models, and utilizing a technique obeying TL standards. Conventional learning is disengaged and happens simply on particular assignments or datasets and to train standalone models on them. None of the information is preserved which could be moved on from one task in hand onto the next. In the case of TL, you can use information (weights and features so on) from the last model trained and even handle issues like having less information for the new task. The basic idea of transfer learning is shown in Fig. 6. DL frameworks and models are layered structures that learn various features at various levels. These levels are then ultimately linked to the last layer called a fully connected layer to yield the outcome. Such a layered structure permits us to use a pre-trained system, (e.g. AlexNet, GoogLeNet, etc.) without its last layer as a permanent feature extractor for different applications (Pan & Yang, 2009). For example, if pretrained networks like AlexNet or GoogLeNet etc. are utilized without its last layer i.e. the classification layer, it will assist us with transforming a fresh domain image into a multi-dimensional vector depending upon its hidden states, thus empowering us for feature extraction from a fresh domain, using the information from a pretrained domain task. This is one of the most broadly used strategies for performing TL utilizing DL systems. The

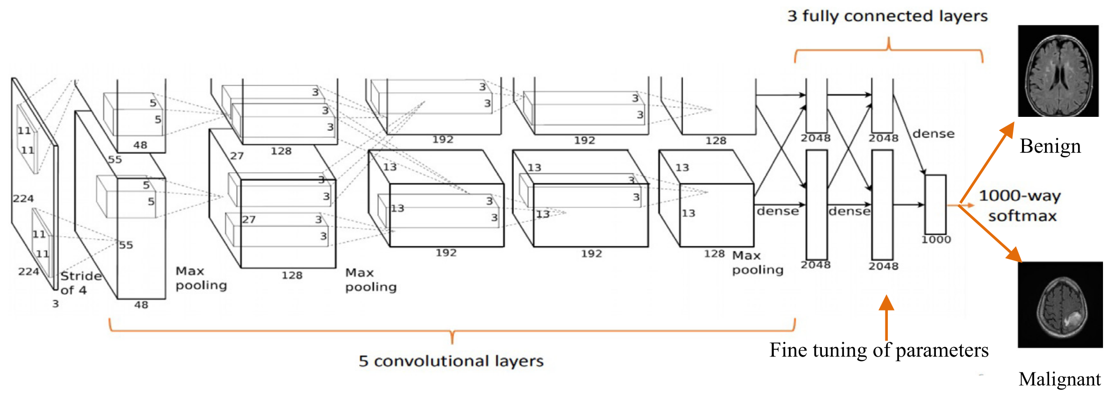



**Fig. 5(a).** AlexNet architecture (Krizhevsky et al., 2012).

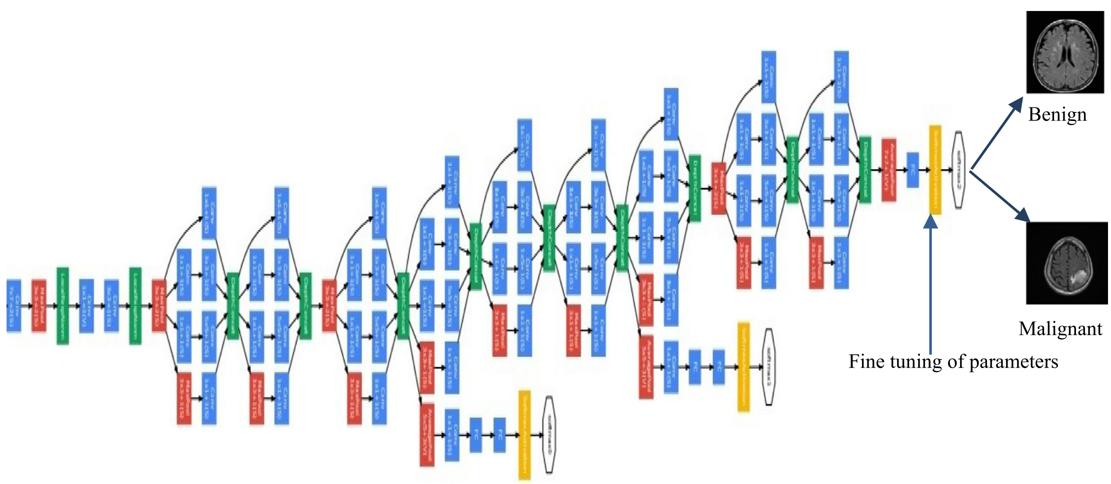



**Fig. 5(b).** GoogleNet architecture (Szegedy et al., 2015).

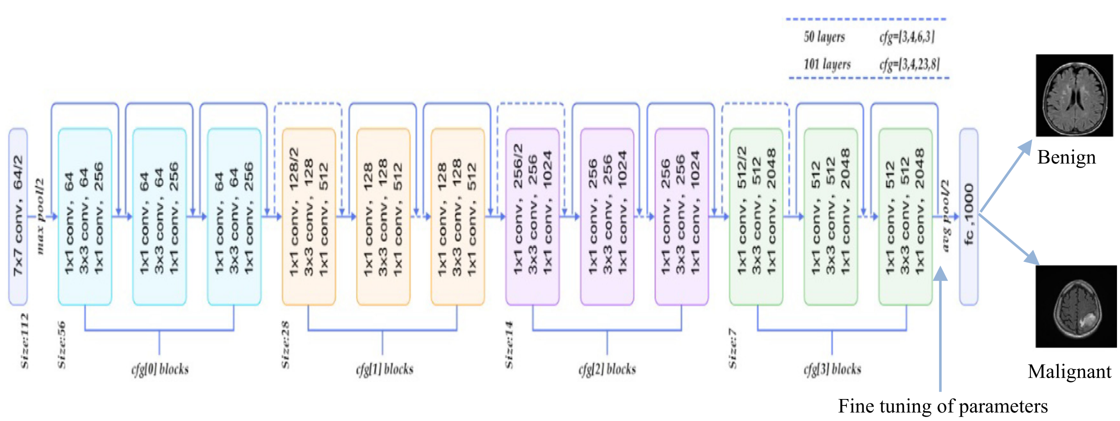



**Fig. 5(c).** ResNet architecture (Deep Learning: ResNet - CodeProject, n.d.).

extraction of visually distinguishable features is done by utilizing fine-tuning for every pretrained network (PTN). Fine-tuning of TL is utilized to build the proficiency and effectiveness of a Convolutional Neural Network by supplanting the last layers of the network. In this situation, the weights of CNN are instated from the top of the PTN as opposed to supplanting and retraining the entire architecture of the classifier. This situation works by moving the weights of the PTN from source to objective dataset. The basic operation is to replace the softmax layer of the PTN and supplant it with a new softmax layer that is significant to the proposed task. In this paper, utilizing the mentioned CNN architectures, the neurons in the objective dataset are put in place of the last fully connected layer.

### 2.3.2. *Regularization and optimization techniques*

To avoid overfitting while training a regularization function is utilized. It implies using a function solver appropriately for stopping network overfitting. Numerous methods have been utilized to prevent overfitting while preprocessing as well as training stages. Initially, image data is augmented to stop overfitting (Ciregan et al., 2012). Subsequently, diverse system structures are tried to ward off system irregularity. Thereafter, dropout layers are utilized, Fig. 7, to expel

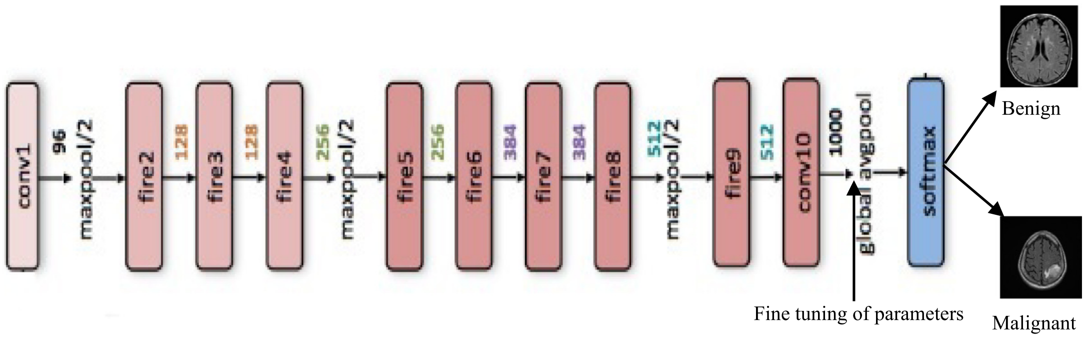



**Fig. 5(d).** SqueezeNet architecture (Fragoulis et al., 2016).

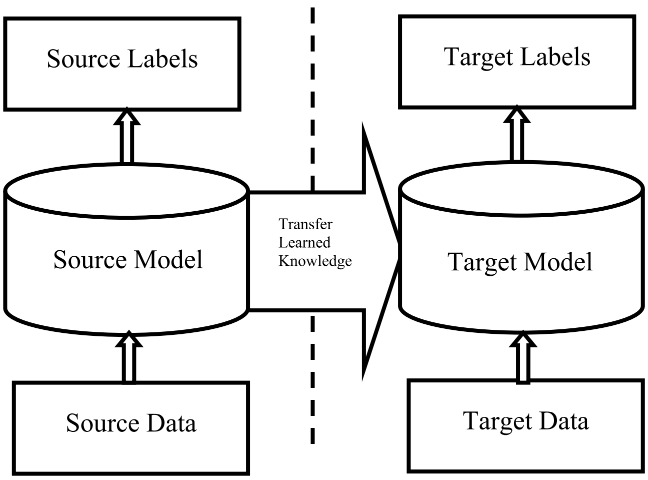



**Fig. 6.** Transfer learning approach.

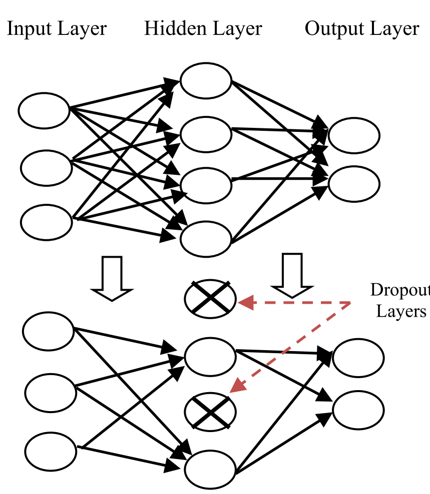



**Fig. 7.** Dropout layer example with 50% dropout probability.

hidden weighs stochastically (Srivastava et al., 2014; Szegedy et al., 2015). Lastly "early stop method" is utilized in certain cases that guards the system while it is getting trained and validated and ends the training before completion of all epochs to avoid the system being overfitted (Goodfellow et al., 2016).

The majority of DL algorithms make use of some kind of optimization techniques for either maximizing or minimizing a function $f(x)$ by varying $x$. This function is termed as an objective function. However, once the function is minimized it is called the cost function or loss(error) function. Gradient descent is a technique to optimize an objective function $J(\theta)$ categorized by a model's constraint $\theta \in \mathbb{R}^d$ by revising it in the reverse direction of the objective function $\nabla_\theta J(\theta)$ w.r.t. to the parameters. '$\eta$' which is the learning rate gives the step size taken to achieve the (local) minimum. SGDM is a technique that accelerates the descent in an appropriate path and reduces its oscillations. This is done by adding $\gamma$ of the previous step update vector to the present update vector.

$$\left.\begin{aligned} v_t &= \gamma v_{t-1} + \eta \nabla_\theta J(\theta) \\ \theta &= \theta - v_t \end{aligned}\right\} \tag{3}$$

RMSprop is an adaptive learning technique. It is developed to answer the drastically thinning learning rate problem. It makes use of an exponentially decaying average to dispose of history so that it can get converged quickly once a convex bowl is found. It is, in fact, similar to Adadelta and is given as:

$$\left.\begin{aligned} E\left|g^2\right|_t &= \gamma E\left|g^2\right|_{t-1} + (1-\gamma) g_t^2 \\ \theta_{t+1} &= \theta_t - \frac{\eta}{\sqrt{E\left|g^2\right|_t + \in}} g_t \end{aligned}\right\} \tag{4}$$

where: $\mathrm{E}\left|g^2\right|_t$ is the running average; $g_t = \nabla_\theta J(\theta)$

Adam is a technique that calculates the adaptive learning rates of every parameter. Besides accumulating an exponentially decaying average of previously squared gradients $v_t$ it also stores the exponentially decaying average of previous gradients $m_t$, as in momentum:

$$\left.\begin{aligned} m_t &= \beta_1 m_{t-1} + (1-\beta_1) g_t \\ v_t &= \beta_2 v_{t-1} + (1-\beta_2) g_t^2 \end{aligned}\right\} \tag{5}$$

where $m_t$ and $v_t$ are approximations of the first (the mean) and the second moment (the variance) of the gradients correspondingly.

### 2.3.3. Metrics for assessment

The viability of the proposed BT detection and identification framework is assessed by calculating seven significant results that are utilized to check the effectiveness of the classifier. The proposed framework performance is calculated as follows:

(i) Accuracy: It is the ability of a system to determine the type of BT correctly and is given by:

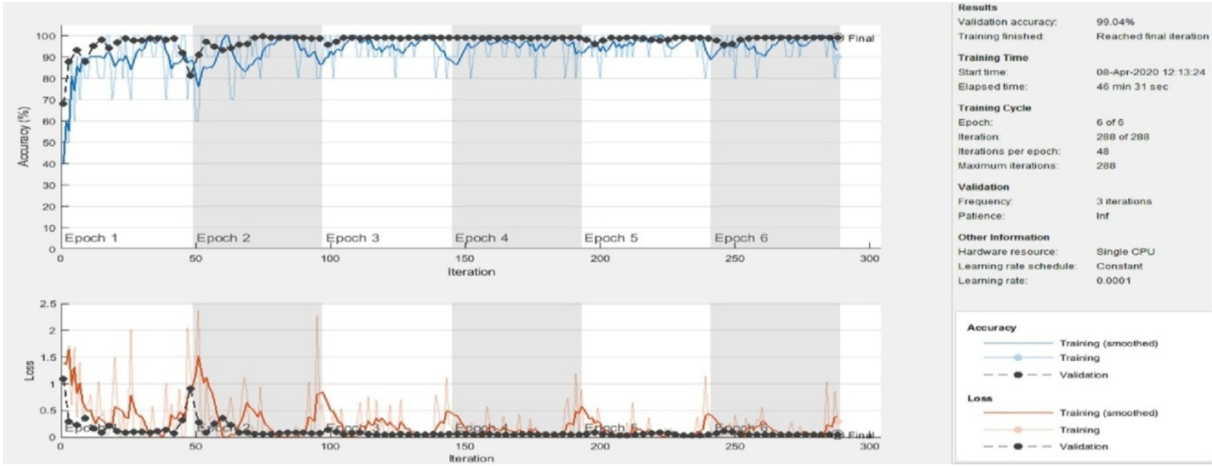
$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{6}$$

(ii) Specificity: It is the ability of a system to precisely identify the genuine BT and is calculated as:

$$Specificity = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}} \tag{7}$$

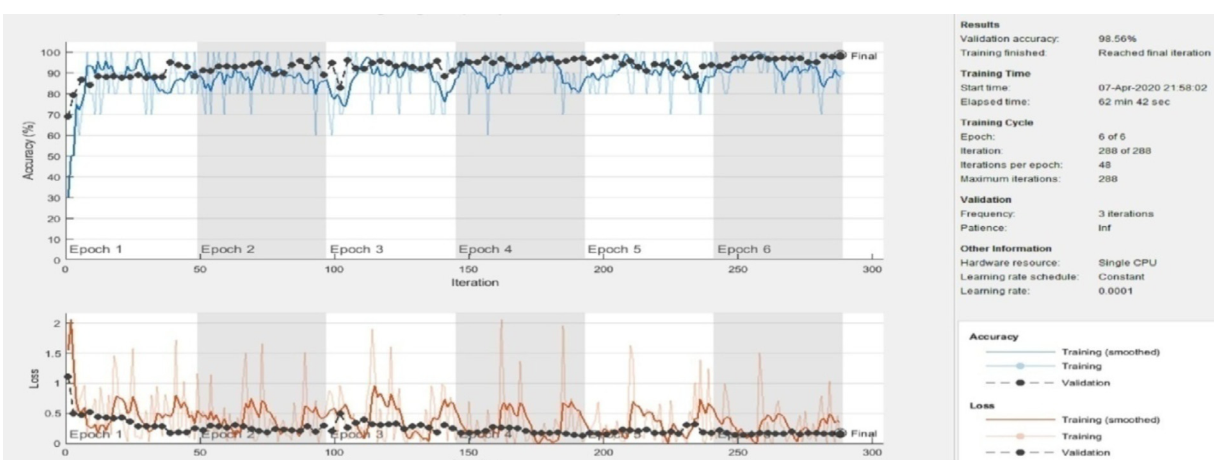
(iii) Sensitivity: It takes into account the capacity of a model to correctly classify the BT and is measured as:

$$Sensitivity = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{8}$$




(a). Finetuned AlexNet


(b). Finetuned GoogleNet




(c). Finetuned SqueezeNet

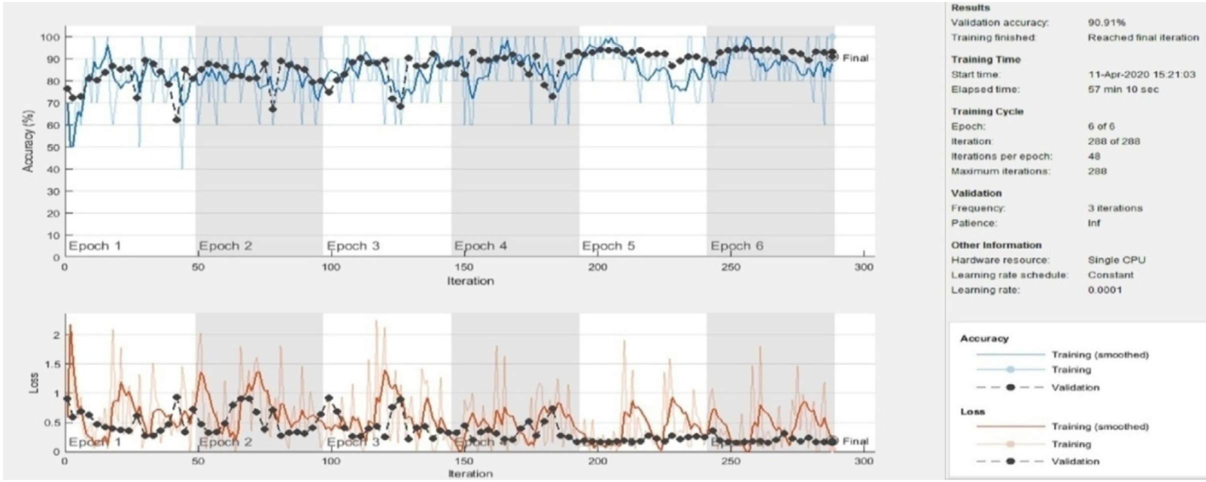
**Fig. 8.** Training progress and loss of different pretrained networks using SGDM optimizer.

(iv) Precision: It is defined as the proximity of the two measured values to each other and is given by

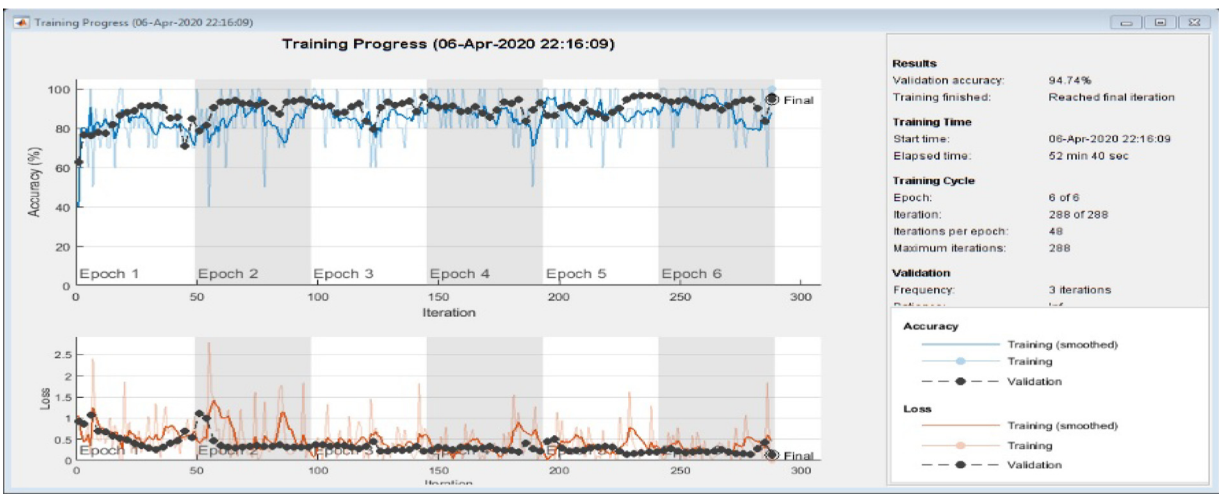
$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{9}$$

(v) Matthews correlation coefficient (MCC): It is denoted as a measure to evaluate the superiority of classification and is given by:

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{10}$$




(d). Finetuned ResNet50




(e). Finetuned ResNet101

**Fig. 8.** (*continued*).

(vi) Error Rate: It is measured as the ratio of all false predictions to the whole of the objective dataset and formulated as

$$Error\ Rate = \frac{FP + FN}{TP + TN + FP + FN} \tag{11}$$

(vii) F1Score: It is used to quantify the accuracy of a test and is given by:

$$F1\ Score = \frac{2 * (\text{Precision} * \text{Specificity})}{(\text{Precision} + \text{Specificity})} \tag{12}$$

where **TP** is the predicted positive cases that are positive. **TN** is the predicted negative case which is negative. **FN** is the predicted negative cases that are actually positive. Such cases are termed as type two errors. **FP** is the predicted positive cases that are in the actual negative. Such cases are termed as type one error.

## 3. Results & discussions

A Transfer Learning-based deep learning algorithms are accessed is this paper for precisely identifying BT types into benign and malignant. The system is trained using different deep learning pretrained networks like AlexNet, GoogLeNet, SqueezeNet, ResNet50, and ResNet101 to access the system's best accuracy. The image identification is done by using softmax layers of pretrained networks by fine-tuning the features. Here the trained visually distinguishable features from every DL networks are adjusted to the objective dataset and BT classification is done by the softmax layer by instating the quantity of neurons to two classes. These fine-tuned parameters are not self-trained, and therefore it is fundamental to set the optimized parameters as per the outcome of trained MR images for performance enhancement. The system is trained multiple times for all the above-pretrained networks using different well-acclaimed optimizers like Stochastic gradient descent with momentum (SGDM), Root means square propagation (RMSProp), and Adaptive moment estimation (ADAM) for training the network to attain the best possible trained system. Fig. 8 represents the training progress and loss of pretrained networks used in our projected system. Investigative results of comparison of finetuned CNN architectures on original Image sets and augmented Image sets are shown in Tables 1, 2, 3, 4, 5.

### 3.1. Matrix of confusion

A confusion matrix is a chart that is frequently used to depict how the particular classification network is performing on a given test dataset for which the actual values are already known. Fig. 9 exhibits the matrix of confusion which consolidates the performance of the system for the classification of brain tumors. In the given matrix $X$-axis is the Target class while the $Y$-axis shows the Output class. The expression for Specificity, Accuracy, Sensitivity, precision, etc. is shown in Table 6 which contains the values derived from the matrix of

**Table 1**
Investigative results of finetuned CNN architecture-Alexnet.

| Tested parameters | Original image sets | | | Augmented image sets | | |
|---|---|---|---|---|---|---|
| | SGDM | RMSProp | ADAM | SGDM | RMSProp | ADAM |
| Batch size | 7 | 10 | 7 | 10 | 10 | 8 |
| Maximum epochs | 6 | 5 | 8 | 6 | 6 | 7 |
| InitialLearnRate | 1.00E−04 | 1.00E−04 | 1.00E−04 | 1.00E−04 | 1.00E−04 | 1.00E−04 |
| Training time (min/s) | 36:45 | 38:40 | 34:15 | 46:31 | 62:41 | 51:31 |
| Accuracy | 95.9 | 96.5 | 95.9 | 99.04 | 99.04 | 98.56 |

**Table 2**
Investigative results of finetuned CNN architecture-Googlenet.

| Tested parameters | Original image sets | | | Augmented image sets | | |
|---|---|---|---|---|---|---|
| | SGDM | RMSProp | ADAM | SGDM | RMSProp | ADAM |
| Batch size | 10 | 10 | 10 | 10 | 10 | 10 |
| Maximum epochs | 5 | 6 | 8 | 5 | 6 | 6 |
| InitialLearnRate | 1.00E−04 | 1.00E−04 | 1.00E−04 | 1.00E−04 | 1.00E−04 | 1.00E−04 |
| Training time (min/s) | 51:51 | 56:37 | 58:39 | 95:31 | 110:13 | 111:58 |
| Accuracy | 93.53 | 90 | 95.88 | 96.65 | 98.09 | 97.13 |

**Table 3**
Investigative results of finetuned CNN architecture-Squeezenet.

| Tested parameters | Original image sets | | | Augmented image sets | | |
|---|---|---|---|---|---|---|
| | SGDM | RMSProp | ADAM | SGDM | RMSProp | ADAM |
| Batch size | 10 | 8 | 10 | 10 | 10 | 10 |
| Maximum epochs | 7 | 6 | 5 | 6 | 7 | 6 |
| InitialLearnRate | 1.00E−04 | 1.00E−04 | 1.00E−04 | 1.00E−04 | 1.00E−04 | 1.00E−04 |
| Training time (min/s) | 25:54 | 27:59 | 28:8 | 62:42 | 58:44 | 63:5 |
| Accuracy | 93.53 | 90 | 93.53 | 98.56 | 96.17 | 96.65 |

**Table 4**
Investigative results of finetuned CNN architecture-Resnet50.

| Tested parameters | Original image sets | | | Augmented image sets | | |
|---|---|---|---|---|---|---|
| | SGDM | RMSProp | ADAM | SGDM | RMSProp | ADAM |
| Batch size | 10 | 10 | 7 | 10 | 10 | 10 |
| Maximum epochs | 6 | 7 | 6 | 6 | 6 | 6 |
| InitialLearnRate | 1.00E−04 | 1.00E−04 | 1.00E−04 | 1.00E−04 | 1.00E−04 | 1.00E−04 |
| Training time (min/s) | 26:49 | 30:50 | 30:33 | 57:10 | 61:50 | 57:18 |
| Accuracy | 94.71 | 97.06 | 85.29 | 90.91 | 95.69 | 94.74 |

**Table 5**
Investigative results of finetuned CNN architecture-Resnet101.

| Tested parameters | Original image sets | | | Augmented image sets | | |
|---|---|---|---|---|---|---|
| | SGDM | RMSProp | ADAM | SGDM | RMSProp | ADAM |
| Batch size | 10 | 8 | 10 | 10 | 10 | 10 |
| Maximum epochs | 6 | 5 | 6 | 6 | 6 | 8 |
| InitialLearnRate | 1.00E−04 | 1.00E−04 | 1.00E−04 | 1.00E−04 | 1.00E−04 | 1.00E−04 |
| Training time (min/s) | 27:35 | 31:1 | 34:48 | 52:40 | 75:50 | 64:20 |
| Accuracy | 93.53 | 92.35 | 97.65 | 94.74 | 94.74 | 93.3 |

confusion for AlexNet. The maximum values of Specificity, Accuracy, Sensitivity, precision, etc. are written in bold. Table 7 gives the training results of different pretrained DL networks using optimizers mentioned in the table along with training time. From the above-tabulated results, it is evident that the pretrained network-AlexNet has given the best results in minimum computational time as compared to the rest of the mentioned networks. It is clear from the given training curves that AlexNet has achieved 99.04% accuracy after completing full iterations.

### 3.2. Finetuned network architecture and hyper parameter optimization

In the current section, various network system designs and architectures implicated in designing the best system are introduced. Table 8 presents various parameters tried before arriving at the fine-tuned system that gives the best outcome. Fig. 10 represents the comparison among the PTNs based on training time. A comparison of the training results of different pretrained networks using different network optimizers is shown in Fig. 11.

**Table 6**
Confusion matrix parameters for the AlexNet.

| Parameters | Proposed model | |
|---|---|---|
| Type of tumor | Benign | Malignant |
| TP | 66 | 141 |
| TN | 141 | 66 |
| FP | 1 | 1 |
| FN | 1 | 1 |
| Error rate | 0.009569 | 0.009569 |
| F1-score | 0.985075 | 0.992958 |
| MCC | 0.556423 | 0.810051 |
| Specificity | 0.992958 | 0.985075 |
| Sensitivity | 0.985075 | 0.992958 |
| Accuracy | 0.985043 | 0.993043 |
| Precision | 0.985075 | 0.992958 |
| Overall accuracy | 99.04% | |

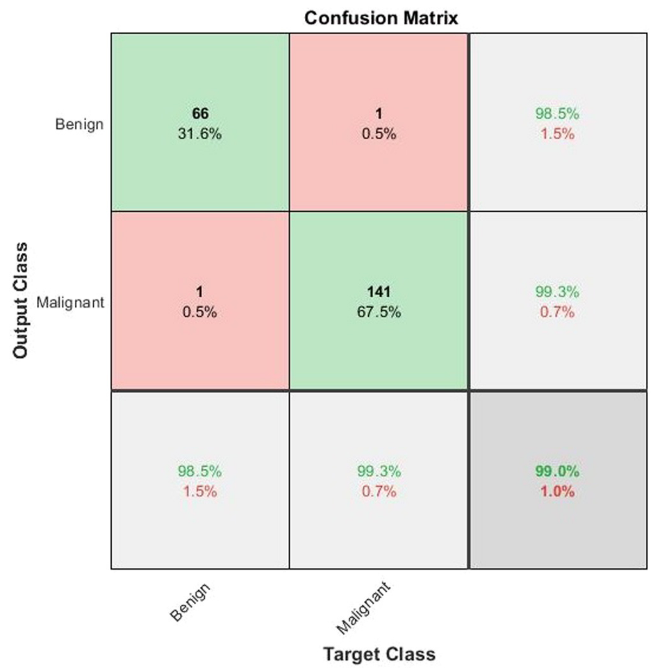



**Fig. 9.** Confusion matrix for the AlexNet.

**Table 7**
Training results of pretrained DL network.

| Algorithms | Optimizer | Accuracy | Benign Acc | Malignant Acc | Training time |
|---|---|---|---|---|---|
| ResNet 101 | SGDM | 94.74% | 96.70% | 94.00% | 52 m 40 s |
| ResNet 50 | | 90.91% | 96.20% | 89.20% | 57 m 10 s |
| GoogleNet | | 96.65% | 90.50% | 100% | 95 m 31 s |
| AlexNet | | **99.04%** | 98.50% | 99.30% | **46 m 31 s** |
| SqueezeNet | | 98.56% | 95.70% | 100% | 62 m 42 s |
| ResNet 101 | ADAM | 93.30% | 98.20% | 91.60% | 64 m 20 s |
| ResNet 50 | | 94.74% | 86.80% | 99.20% | 57 m 18 s |
| GoogleNet | | 97.13% | 93.00% | 99.30% | 111 m 58 s |
| AlexNet | | **98.56%** | 97.10% | 99.30% | **51 m 31 s** |
| SqueezeNet | | 96.65% | 95.50% | 97.20% | 63 m 5 s |
| ResNet 101 | RMSProp | 94.74% | 86.80% | 99.20% | 75 m 50 s |
| ResNet 50 | | 95.69% | 90.30% | 98.50% | 61 m 50 s |
| GoogleNet | | 98.09% | 94.40% | 100.00% | 110 m 13 s |
| AlexNet | | **99.04%** | 98.50% | 99.30% | 62 m 41 s |
| SqueezeNet | | 96.17% | 96.80% | 95.90% | **58 m 44 s** |

**Table 8**
Parameters fine-tuned for achieving the highest accuracy.

| Factors | Values |
|---|---|
| No. of convolutional + RLU Layers | 1, 2, 3, 4 |
| No. of dropout layer | 1, 2, 3 |
| Maximum iterations | 40, 80, 120, 150, 200, 288 |
| Types of pooling layers | Maximum and average pooling |
| Optimizers used | SGDM, ADAM, RMSProp |
| DL based pretrained networks tested | ResNet50, ResNet101, GoogLeNet, AlexNet, SqueezeNet |
| Mini batch size taken | 4, 8, 10, 12, 16 |
| Initial learning rate | 0.01, 0.001, 0.0001 |
| Learning rate drop factor | 0.1, 0.2, 0.3, 0.5 |
| Maximum epochs | 5, 6, 7, 8 |

### 3.3. Comparison with state of the art techniques

This paper gives a framework for selecting the pretrained network for applying transfer learning to classify brain tumors into benign and malignant types. These systems perform the self-extraction of features depending upon their network architecture. The last few layers of a pre-trained network are well adjusted for performing a new classification task. The image size taken is 225 × 225 in JPG/JPEG configuration. For executing the proposed work MATLAB R2018a is used. From among the presented Pre-Trained Convolutional Neural Networks (PT-CNN), AlexNet gives the best accuracy of 99.04%. Table 9 represents various related State of the Art (SoA) techniques.

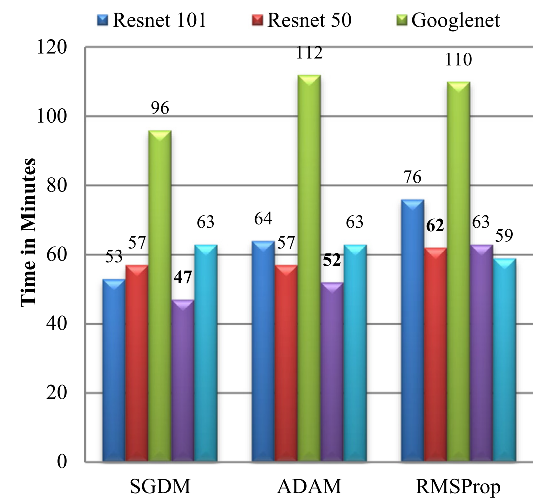



**Fig. 10.** Comparison among the PTNs based on training time in minutes.

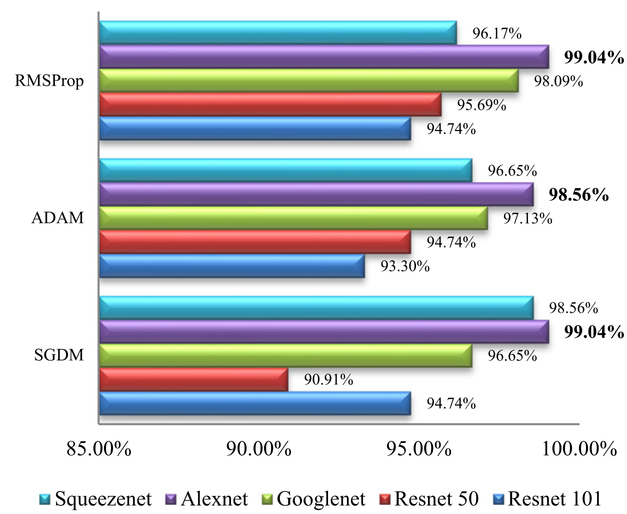



**Fig. 11.** Comparison of the training results of different pretrained networks.

## 4. Conclusion

In the presented paper, Deep Learning based Pre-Trained Convolutional Neural Network framework is utilized for classifying MRI images of BT into benign and malignant types. The paper investigates various pretrained networks for improving the MR image classification of BT using transfer learning techniques. In light of the system execution results shown in Table 2, it is evident that transfer learning through AlexNet gives the utmost performance of 99.04% among all the PTNs used in this study. We observationally came across a few discoveries. To begin with, the training results of different pretrained DL networks show that the performance of a PTN heavily depends on the type of optimizer chosen. It affects the accuracy and most importantly the time is taken to train the network. Next, it is observed that among all the pretrained networks GoogLeNet has taken the maximum time for training the network with all the given optimizers with the highest of 112 min using ADAM. Lastly, The average time taken for training

**Table 9**
State of the art techniques.

| Authors | Techniques used | Type of classification | Dataset used | Accuracy |
|---|---|---|---|---|
| Cheng et al. (2015) | SVM and KNN | Multi | T1-weighed CE-MRI, 708 meningiomas, 1426 gliomas, and 930 pituitary tumors | 91.28% |
| Paul et al. (2017) | CNN | Multi | T1-weighed CE-MRI, 208 meningioma, 492 glioma, and 289 pituitary tumor images | 91.43% |
| Afshar et al. (2019) | CNN | Multi | T1-weighed CE-MRI, 708 meningiomas, 1426 gliomas, and 930 pituitary tumors | 90.89% |
| Anaraki et al. (2019) | GA-CNN | Multi | T1-weighed CE-MRI, 708 meningiomas, 1426 gliomas, and 930 pituitary tumors | 94.20% |
| Bahadure et al. (2017) | BWT+SVM | Binary | T2-weighted brain images, 67 normal and 134 abnormal | 95.00% |
| Ansari et al. (2020) | DWT+PCA+GLCM+SVM | Binary | T1-weighed CE-MRI, 140 tumor affected, 60 normal | 98.91% |
| Sultan et al. (2019) | CNN | Multi | T1-weighed CE-MRI, 708 meningiomas, 1426 gliomas, and 930 pituitary tumors | 96.13% |
| Kumar et al. | GWO+M-SVM | Multi | T1-weighed CE-MRI, 248 meningiomas, 12 gliomas, and 55 pituitary tumors | 95.23% |
| Abiwinanda et al. | CNN | Multi | T1-weighed CE-MRI, 708 meningiomas, 1426 gliomas, and 930 pituitary tumors | 84.19% |
| **Proposed system** | PT-CNN(AlexNet) | **Binary** | T1-weighed MRI, 224 Benign and 472 malignant tumor | **99.04%** |

using a particular optimizer is the highest of 74 min with RMSProp and lowest of 63 min with SGDM optimizer. Even though the dataset is not large enough, augmentation of image data has done fairly well to yield superior outcomes and consequently conquer this issue. Our proposed system has accomplished the most superior accuracy of 99.04% utilizing the given dataset. In the future work can be done using the larger dataset to further improve the accuracy and trying to minimize the time taken to train the network by using advanced processors.

## CRediT authorship contribution statement

**Rajat Mehrotra:** Investigation, Conceptualization, Methodology, Software, Data curation, Writing - original draft. **M.A. Ansari:** Supervision, Software, Validation, Writing - review & editing. **Rajeev Agrawal:** Supervision, Software, Validation, Writing - review & editing. **R.S. Anand:** Supervision, Quality check.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References


Abir, T. A., Siraji, J. A., Ahmed, E., & Khulna, B. (2018). Analysis of a novel MRI based brain tumour classification using probabilistic neural network (PNN). *International Journal of Science and Research Science Engineering Technology*, *4*(8), 65–79.

Abiwinanda, N., Hanif, M., Hesaputra, S. T., Handayani, A., & Mengko, T. R. (2019). Brain tumor classification using convolutional neural network. In *World congress on medical physics and biomedical engineering 2018* (pp. 183–189). Singapore: Springer.

Afshar, P., Plataniotis, K. N., & Mohammadi, A. (2019). Capsule networks for brain tumor classification based on MRI images and coarse tumor boundaries. In *ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 1368–1372). IEEE.

American Brain Tumor Association, Chicago, IL, USA (2015). Surgery. [Online]. Available: http://www.abta.org/secure/surgery.pdf.

Anaraki, A. K., Ayati, M., & Kazemi, F. (2019). Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms. *Biocybernetics and Biomedical Engineering*, *39*(1), 63–74.

Ansari, M. A., Mehrotra, R., & Agrawal, R. (2020). Detection and classification of brain tumor in MRI images using wavelet transform and support vector machine. *Journal of Interdisciplinary Mathematics*, 1–12.

Bahadure, N. B., Ray, A. K., & Thethi, H. P. (2017). Image analysis for MRI based brain tumor detection and feature extraction using biologically inspired BWT and SVM. *International journal of biomedical imaging*, 2017.

Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Takeshi Shinohara, R., Berger, C., Ha, S. Min., Rozycki, M., Prastawa, M., Alberts, E., Lipkova, J., Freymann, J., Kirby, J., Bilello, M., Fathallah-Shaykh, H. M., Wiest, R., Kirschke, J., .... LladóLlad, X. (2019). Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. In *Sandra gonzlez-vill, Vol. 124*.

Brain, other CNS and intracranial tumours statistics | Cancer Research UK. (n.d.). (2020). from https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/brain-other-cns-and-intracranial-tumours?_ga=2.245798211.613033350.1594394057-1698240228.1594394057 (Retrieved July 10, 2020).

Chelghoum, R., Ikhlef, A., Hameurlaine, A., & Jacquir, S. (2020). Transfer learning using convolutional neural network architectures for brain tumor classification from MRI images. http://dx.doi.org/10.1007/978-3-030-49161-1_17.

Chen, T., Lin, L., Zuo, W., Luo, X., & Zhang, L. (2018). Learning a wavelet-like auto-encoder to accelerate deep neural networks. In *Thirty-second AAAI conference on artificial intelligence*.

Cheng, J., Huang, W., Cao, S., Yang, R., Yang, W., Yun, Z., . . ., & Feng, Q. (2015). Enhanced performance of brain tumor classification via tumor region augmentation and partition. *PLoS One*, *10*(10), Article e014038.

Ciregan, D., Meier, U., & Schmidhuber, J. (2012). Multi-column deep neural networks for image classification. In *2012 IEEE conference on computer vision and pattern recognition* (pp. 3642–3649). IEEE.

Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., ., & Tarbox, L. (2013). The cancer imaging archive (TCIA): maintaining and operating a public information repository. *Journal of Digital Imaging*, *26*(6), 1045–1057.

Damodharan, S., & Raghavan, D. (2015). Combining tissue segmentation and neural network for brain tumor detection. *International Arab Journal of Information Technology (IAJIT)*, *12*(1).

DeAngelis, L. M. (2001). Brain tumors. *New England Journal of Medicine*, *344*(2), 114–123.

Deep Learning using Python + Keras (Chapter 3): ResNet - CodeProject. (n.d.). (Retrieved August 25, 2020) from https://www.codeproject.com/Articles/1248963/Deep-Learning-using-Python-plus-Keras-Chapter-Re.

Deng, L., & Yu, D. (2014). Deep learning: methods and applications. *Foundations and Trends in Signal Processing*, *7*(3–4), 197–387.

El-Dahshan, E. S. A., Hosny, T., & Salem, A. B. M. (2010). Hybrid intelligent techniques for MRI brain images classification. *Digital Signal Processing*, *20*(2), 433–441.

Ertosun, M. G., & Rubin, D. L. (2015). Automated grading of gliomas using deep learning in digital pathology images: A modular approach with ensemble of convolutional neural networks. In *AMIA annual symposium proceedings, Vol. 2015* (p. 1899). American Medical Informatics Association.

Fragoulis, N., Pothos, V. K., Kastaniotis, D., & Theodorakopoulos, I. (2016). A fast, embedded implementation of a convolutional neural network for image recognition -revisited. august. http://dx.doi.org/10.13140/RG.2.1.1778.9681.

Goodenberger, M. L., & Jenkins, R. B. (2012). Genetics of adult glioma. *Cancer Genetics*, *205*(12), 613–621.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Jalab, H. A., & Hasan, A. (2019). Magnetic resonance imaging segmentation techniques of brain tumors: A review. *Archive Neuroscience*, *6*.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).

Kumar, A., Ansari, M. A., & Ashok, A. (2019). A hybrid framework for brain tumor classification using grey wolf optimization and multi-class support vector machine. *International Journal of Recent Technology and Engineering*, *8*(3), 7746–7752. http://dx.doi.org/10.35940/ijrte.C6315.098319.

LeCun, Y. (2015). Lenet-5, convolutional neural networks. URL: http://yann.lecun.com/exdb/lenet,20(5),14.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, *42*, 60–88.

Liu, J., Pan, Y., Li, M., Chen, Z., Tang, L., Lu, C., & Wang, J. (2018). Applications of deep learning to MRI images: A survey. *Big Data Mining and Analytics*, *1*(1), 1–18.

Lundervold, A. S., & Lundervold, A. (2019). An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift für Medizinische Physik*, *29*(2), 102–127.

Matsugu, M., Mori, K., Mitari, Y., & Kaneda, Y. (2003). Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural Networks*, *16*(5–6), 555–559.

Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359.

Paul, J. S., Plassard, A. J., Landman, B. A., & Fabbri, D. (2017). Deep learning for brain tumor classification. In A. Krol & B. Gimi (Ed.), *Medical imaging 2017: Biomedical applications in molecular, structural, and functional imaging, Vol. 10137* (p. 1013710). SPIE, http://dx.doi.org/10.1117/12.2254195.

Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Transactions on Medical Imaging*, *35*(5), 1240–1251.

Rehman, A., Naz, S., Razzak, M. I., Akram, F., & Imran, M. (2020). A deep learning-based framework for automatic brain tumors classification using transfer learning. *Circuits, Systems, and Signal Processing*, *39*(2), 757–775. http://dx.doi.org/10.1007/s00034-019-01246-3.

Sarkar, D. (2018). A comprehensive hands-on guide to transfer learning with real-world applications in deep learning.

Shalini, R., Muralidharan, V., & Varatharaj, M. (2014). MRI Brain tumor segmentation using kernel weighted fuzzy clustering. *International Journal of the Engineering Research and Technology*, *3*(4), 121–125.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, *15*(1), 1929–1958.

Stewart, B. W., & Wild, C. P. (2014). *World cancer report 2014* (p. 630). Lyon, France: IARC. IARC Nonserial Publ.

Sultan, H. H., Salem, N. M., & Al-Atabany, W. (2019). Multi-classification of brain tumor images using deep neural network. *IEEE Access*, *7*, 69215–69225.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ., & Rabinovich, A. (2015). Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).

Tandel, G. S., Biswas, M., Kakde, O. G., Tiwari, A., Suri, H. S., Turk, M., ., & Madhusudhan, B. K. (2019). A review on a deep learning perspective in brain cancer classification. *Cancers*, *11*(1), 111.

Van der Burgh, H. K., Schmidt, R., Westeneng, H. J., de Reus, M. A., van den Berg, L. H., & van den Heuvel, M. P. (2017). Deep learning predictions of survival based on MRI in amyotrophic lateral sclerosis. *NeuroImage: Clinical*, *13*, 361–369.

Wong, S. C., Gatt, A., Stamatescu, V., & McDonnell, M. D. (2016). Understanding data augmentation for classification: when to warp?. In *2016 international conference on digital image computing: techniques and applications (DICTA)* (pp. 1–6). IEEE.

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, *9*(4), 611–629.

Zacharaki, E. I., Wang, S., Chawla, S., Soo Yoo, D., Wolf, R., Melhem, E. R., & Davatzikos, C. (2009). Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, *62*(6), 1609–1618.

Zimny, A., Neska-Matuszewska, M., Bladowska, J., & Sąsiadek, M. J. (2015). Intracranial lesions with low signal intensity on T2-weighted MR images–review of pathologies. *Polish Journal of Radiology*, *80*, 40.