Computers in Biology and Medicine 111 (2019) 103345

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Brain tumor classification using deep CNN features via transfer learning

S. Deepak*, P.M. Ameer

Department of Electronics & Communication Engineering, National Institute of Technology, Calicut, India

ARTICLE INFO

Keywords:
Brain tumor
Computer-aided diagnosis
Transfer learning
Convolutional neural network
Support vector machine

ABSTRACT

Brain tumor classification is an important problem in computer-aided diagnosis (CAD) for medical applications. This paper focuses on a 3-class classification problem to differentiate among glioma, meningioma and pituitary tumors, which form three prominent types of brain tumor. The proposed classification system adopts the concept of deep transfer learning and uses a pre-trained GoogLeNet to extract features from brain MRI images. Proven classifier models are integrated to classify the extracted features. The experiment follows a patient-level five-fold cross-validation process, on MRI dataset from figshare. The proposed system records a mean classification accuracy of 98%, outperforming all state-of-the-art methods. Other performance measures used in the study are the area under the curve (AUC), precision, recall, F-score and specificity. In addition, the paper addresses a practical aspect by evaluating the system with fewer training samples. The observations of the study imply that transfer learning is a useful technique when the availability of medical images is limited. The paper provides an analytical discussion on misclassifications also.

1. Introduction

Accurate and prompt diagnosis of brain tumors is essential for implementing an effective treatment of this disease. The choice of a treatment modality depends on the stage of the tumor at the time of diagnosis, the pathological type, and grade of the tumor. Computer-aided diagnosis (CAD) techniques have been assisting neuro-oncologists in numerous ways. CAD applications in neuro-oncology include tumor detection, classification, and grading. CAD-based brain tumor classification into benign and malignant tumors is an extensively researched topic [1]. Grading of glioma, which is a major class of malignant tumors, is another research problem in this domain [2]. The aforementioned CAD systems rely on magnetic resonance imaging (MRI) images of the brain. This is because of the ability of MRI to provide a higher contrast for soft tissues in brain compared to computed tomography (CT) images.

Recent works on computer-aided medical diagnosis provide improved performances owing to the advent of deep learning concepts. Deep learning strategies have been extensively used in the medical image analysis of breast cancer studies [3] and lung cancer diagnosis [4]. Zuo et al. [5] developed a deep learning algorithm for human skin detection, which is a part of dermatology diagnostics. Charron et al. [6] used a deep convolutional neural network (CNN) to monitor brain metastases. More recently, a special class of deep learning, known as deep transfer learning, has been dominating the studies on visual categorization, object recognition and image classification problems [7]. Transfer learning allows the use of a pre-trained CNN model, which was actually developed for another related application. Transfer learning has shown its potential in CAD of medical problems also. Zhou et al. [8] used a pre-trained InceptionV3 model for differentiating benign and malignant renal tumors on CT images. Deniz et al. [9] proposed a classifier for breast cancer on histopathologic images. The authors used a pre-trained VGG-16 model and a fine-tuned AlexNet for extracting features, which were then classified using a support vector machine (SVM). Hussein et al. [10] introduced a learning model for lung tumor characterization and for pancreatic tumor characterization. The learning model was based on knowledge transfer and it had a 3D CNN architecture. The accuracy measures reported in the transfer learning-based algorithms were superior to those obtained using handcrafted engineering. Specifically, transfer learning has gathered attention in applications related to neuro-oncology. Studies were conducted to extract deep features from brain MRI images using pre-trained networks [11,12]. The studies showed the capability of transfer learning to work with smaller datasets. Yang et al. [13] used AlexNet and GoogLeNet in their research work on the grading of glioma from MRI images. In terms of the performance measures observed, GoogLeNet proved superior to AlexNet for the task. Talo et al. [14] achieved remarkable classification performance with deep transfer learning in their work on brain abnormality classification. The authors used ResNet-34 and the experiments included training of modified

* Corresponding author.
E-mail address: deepak_p180039ec@nitc.ac.in (S. Deepak).

https://doi.org/10.1016/j.compbiomed.2019.103345
Received 18 March 2019; Received in revised form 26 June 2019; Accepted 26 June 2019
0010-4825/ © 2019 Elsevier Ltd. All rights reserved.

dense layers, training with data augmentation and fine tuning of a transfer learned model. The experimental results concluded that a deep transfer learned model can be adapted to medical image classification, with minimum pre-processing. Jain et al. [15] used a pre-trained VGG-16 network for diagnosis of Alzheimers disease from MRI. Transfer learning was applied to content-based image retrieval (CBIR) for brain tumors [16]. The evaluation was performed on a publicly available dataset and obtained promising results.

Brain tumor classification into subtypes is a more challenging research problem. The related challenges are attributed to the following factors: 1) brain tumors exhibit high variations with respect to shape, size and intensity [17]; 2) tumors from different pathological types might show similar appearances [18]. Among all brain tumors, glioma, meningioma and pituitary tumor account for the highest incidence rates [16]. Cheng et al. [17] worked on the 3-class brain tumor categorization problem with T1-MRI images. This was the first significant work on classification that used the challenging dataset from figshare. The proposed approach relied on the manually delineated tumor border to extract features from the region of interest. In the work, the authors experimented with a wide range of image features and a set of classifier models. The best classification performance was achieved by an SVM model on bag of words (BoW) features. The experiments followed a standard five-fold cross-validation procedure. The performance measures used were specificity, sensitivity and classification accuracy. Ismael and Abdel-Qadar [19] proposed a classification algorithm based on statistical features extracted by discrete wavelet transform (DWT) and a Gabor filter. The features were then used to train a multi-layer perceptron (MLP) classifier. The algorithm was evaluated on the figshare dataset, as used in the previous work [17]. A random division of database images into 70% and 30% was done to form the training set and validation set, respectively.

A CNN-based deep learning model was successfully applied to the considered brain tumor classification problem [20]. The advantage of CNN-based classifier systems is that they do not require manually segmented tumor regions and provide a fully automated classifier. A CNN architecture was designed to extract features from brain MRI in the work by Pashaei et al. [21]. The model had five learnable layers and the size of filters in all the layers was 3x3. The CNN model gave a classification accuracy of 81%. The performance was enhanced by using CNN features with a classifier model from the class of extreme learning machines (ELM). For this work, recall measures for the class pituitary tumor were very high while those of meningioma were very low. This suggests limitations in the classifier's discrimination capability. A modified CNN architecture, known as a capsule network (CapsNet), was used in brain tumor classification by Afshar et al. [22]. CapsNet utilised the spatial relationship between the tumor and its surrounding tissues. Still, the improvement in performance was meagre.

The limitations of the existing works on classification of brain tumors into meningioma, glioma and pituitary tumors are summarized as follows. The performance of state-of-the-art methods is inadequate considering the medical significance of the classification problem. Earlier methods relied on manually delineated tumor regions, prior to classification. This prevented them from being fully automated. The automatic algorithms developed using CNN and its variants could not achieve an influential improvement in performance. Further, the existing methods were tested on the figshare dataset, which is unbalanced (with respect to tumor classes). Hence, performance evaluation using metrics other than accuracy becomes significant. Another fact is that no related work has attended to the concern of data scarcity that can appear in practice.

In this paper, we present an accurate and automatic classification system designed for three pathological types of brain tumor (glioma, meningioma and pituitary tumor). The implementation uses a deep transfer learned CNN model for feature extraction from brain MRI images. The extracted features are classified using proven classifiers. A comprehensive evaluation of the proposed system is then conducted. The proposed system recorded the best classification performance, compared to all the related works, when evaluated on the open dataset from figshare. Also, with a smaller number of training samples the proposed algorithm is found to produce acceptable results. The major contributions of this paper are listed below.

- For the first time, the concept of deep transfer learning is applied to the specific 3-class brain tumor classification problem.
- A substantial improvement in performance is observed when transfer learned deep CNN features are used with proven classifier models.
- The proposed method recorded the best classification accuracy compared to state-of-the-art methods.
- The proposed method gives a considerably good performance even when a smaller number of training samples is used.
- The phenomenon of overfitting with fewer training data and its impact on classification performance are studied.

The remaining part of this paper is organized as follows. In section 2, we present an overview of the concept of transfer learning, followed by a description of the dataset used in the study and the complete framework for the proposed classification algorithm. Section 3 gives details of the experiments performed, presents evaluation results and provides discussion on the results. Section 4 provides the conclusion of this paper.

2. Method

In this work, we applied a pre-trained deep network, the GoogLeNet, to our classification problem via transfer learning.

2.1. Transfer learning in inductive setting

Transfer learning is the technique by which the knowledge gained by an already trained model is used to learn another set of data [23]. Given a source domain ($\mathbf{D}_s$) and a learning task ($\mathbf{T}_s$) in the source domain, a target domain ($\mathbf{D}_t$) and the corresponding learning task ($\mathbf{T}_t$), the objective of transfer learning is to improve the learning in $\mathbf{D}_t$, using the knowledge in $\mathbf{D}_s$ and $\mathbf{T}_s$. Based on the type of task and on the nature of the data available at the source and target domains, different settings are defined for transfer learning [7]. The transfer learning approach is known as inductive transfer learning when labelled data is available in the source and target domains for a classification task [24]. In such a scenario, the domain is represented as $\mathbf{D} = (x_i y_i)\ \forall\ i$, where x_i is the feature vector for the ith training sample and y_i is the class label for the ith training sample. The setting for transfer learning with respect to our application is explained in the following statements.

GoogLeNet is defined as a deep network with the number of learnable layers being 22 [25]. It has two convolutional layers, two pooling layers and nine inception modules, and a fully connected layer. Again, an inception module has six convolution layers and a pooling layer. The module consists of filters of sizes 1x1, 3x3 and 5x5 in its architecture [13]. Filters with different kernel sizes are expected to capture different patterns in data. The feature maps, corresponding to different filters, are concatenated at the output of each module. Moreover, 1x1 convolutions are performed prior to convolutions by larger kernels. They are meant for dimension reduction in computations. GoogLeNet was the winner in the Imagenet large scale visual recognition challenge (ILSVRC) held in 2014. It was trained using 1.2 million natural images in Imagenet. Each image in Imagenet belonged to one of the 1000 defined classes. Training samples in Imagenet define the source domain and the source task is a 1000-class classification problem.

We define our task as the classification of brain tumors into three types. T1-contrast enhanced (CE) MRI images from figshare [26] form our dataset and define the target domain. Fig. 1 demonstrates the basic

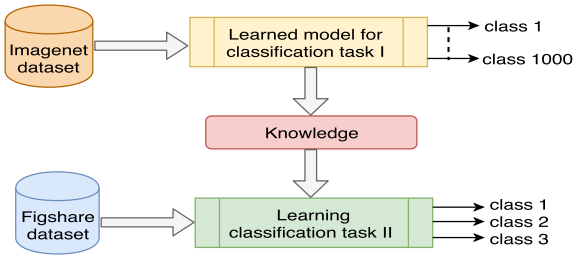


Fig. 1. Transfer learning setting.

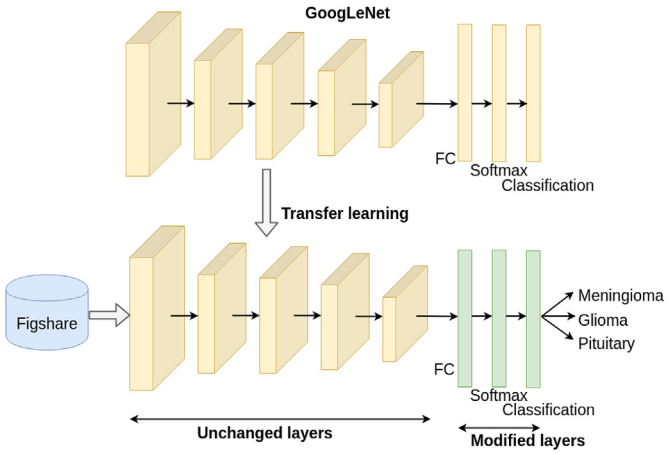


Fig. 2. Modifying GoogLeNet for the application.

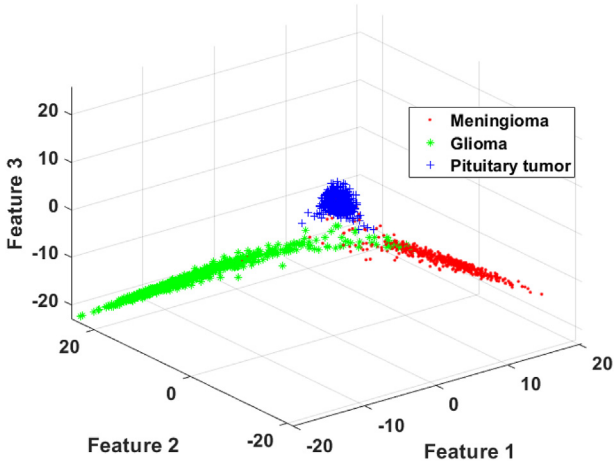


Fig. 3. Scatter plot of features at FC layer.

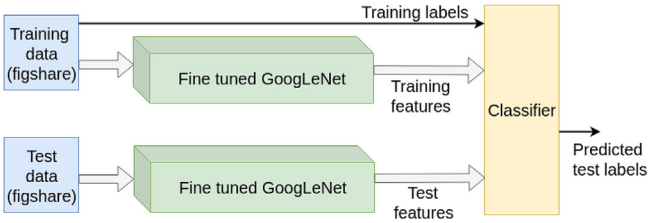


Fig. 4. Overall framework for brain tumor classification.

framework of the transfer learning approach as applied to our work. We modified the last three layers of GoogLeNet so as to adapt it to the target domain. The fully connected (FC) layer in the original GoogLeNet was removed. Instead, a new FC layer with an output size of three was inserted. The softmax layer, following the FC layer, and the cross-entropy-based classification layer at the output were replaced with new ones. The modified deep learning model is shown in Fig. 2. Then, we performed a fine tuning of the modified GoogLeNet by training it with MRI images from figshare. The learning factors for weights and bias at the FC layer were set to a high value of 10. The idea was to make the network learn abstract high-level features specific to the target domain. The low-level features were expected to be learned by the pre-trained layers from the original GoogLeNet. The transfer learned and fine-tuned deep CNN could then be used for experiments using MRI data from the figshare dataset.

2.2. Proposed classification framework

The proposed algorithm makes use of the modified and fine-tuned GoogLeNet to learn the features of brain MRI images with tumors. (The extracted features using modified GoogLeNet will be referred to as deep CNN features in the rest of this paper.) Features extracted from layers near to the fully connected layer give good discriminative representations for trained images. Fig. 3 shows the scatter plot of features obtained at the FC layer of the transfer learned GoogLeNet. The transfer learned model has a softmax classifier layer, which can classify image data from figshare into three tumor classes. However, an improved performance is expected when deep CNN features are provided to well-established pattern classifiers [27]. In the proposed work, the deep CNN features were tested on two sets of classifier models other than the softmax classifier within deep CNN. These are the SVM and k-nearest neighbours (KNN) classifiers. The choice of SVM for our experiments was a result of motivation from previous works where SVM had shown good performance in classifying CNN features. In cases when data points are not easily separable by SVM planes, a density estimation based classifier could be more useful. Following this thought, KNN was considered for our experiments. Fig. 4 shows the overall framework for the proposed classification system. Deep CNN features of the training data and test data, and class labels of the training data form inputs to the classifier model. The model predicts and outputs class labels for the test data.

3. Experiments and results

We implemented the proposed classification model in MATLAB 2018b on a computer having specifications of 32 GB RAM and Intel E3-1245v6 @3.70 GHz CPU.

3.1. Dataset and pre-processing

The dataset from figshare is openly available [26] and is commonly used for evaluating classification and retrieval algorithms [16]. It is a collection of 3064 brain MRI images from 233 patients, diagnosed with one of the three brain tumors (meningioma, glioma and pituitary tumors). The images belong to the T1-CE MRI modality and include coronal, sagittal and axial views. It contains 1426 brain MRI images with glioma (corresponding to 89 patients), 708 images for meningioma (corresponding to 82 patients) and the remaining 930 images correspond to cases of pituitary tumor (belonging to 62 patients). The images are available as *.mat* files and the size of each image is 512x512. GoogLeNet was originally designed for RGB colour images, with an input layer of size 224x224x3. The MRI images in the dataset were pre-processed in the following manner (Fig. 5). They were normalized in intensity values. A min-max normalization technique was followed to scale the intensity values between 0 and 1. They were resized to

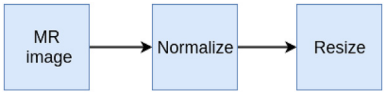


Fig. 5. Data preprocessing steps.

224x224. Because MRI images are greyscale images, three channels were then created by replicating the greyscale values three times. The evaluation procedure for the designed system on the figshare dataset followed a patient-level five-fold cross-validation. The entire dataset of 233 patients was divided into five disjoint subsets. The divided subsets were of approximately equal size. One subset was selected as the test set while the rest formed the training set. This process was repeated such that every subset formed the test set once. Such a division of the dataset was to ensure that the data of a particular patient was not simultaneously present in the test set and in the training set.

3.2. Classifier settings

The performance of an image classification system depends on the combination of image features and the classifier model. With regard to the classifier models used in the final stage of the proposed brain tumor classification system, there are three distinct experimental settings.

1. Transfer learned deep CNN model with its softmax classifier, as a stand-alone system.

The modified GoogLeNet was trained using the training set (after pre-processing). The hyperparameters of the network were heuristically adjusted so as to facilitate the convergence of the loss function during training (Fig. 6). *Adam* was the chosen optimizer, considering its good learning rate and the parameter-specific adaptive nature of the learning rates. For Adam, the initial learning rate was chosen as 0.0003. The choice of a high value might prevent the loss function from converging and could cause overshoots. And a very small value of the learning rate increases the training time. The mini-batch size was set to 30. The choice is a compromise between the speed of training (a larger batch size means faster training) and the computational requirements (limit set by the computer specifications). Also, a very large batch size adversely affects the model quality. Cross-entropy is the loss function used as it gives the measure of the closeness of the predicted and actual distributions. A higher learning rate is desirable at the modified FC layer so as to learn the MRI image specific features. So, a learning factor of 10 is set. The number of epochs was limited to 10, considering the occurrence of overfitting. The hyperparameter settings of our experiment are listed in Table 1.

2 Deep CNN features with SVM classifier.

We extracted features from the pooling layer, placed after the final inception module of the modified GoogLeNet. The features were then classified using SVM. We used a multi-class SVM with an error-correcting output code (ECOC) model. A one-vs-all strategy was used for multi-class classification. There were three binary SVM learners, each with a *linear* kernel. Other parameters of SVM are given in Table 1.

3 Deep CNN features with KNN classifier.

Features were extracted in a manner similar to the previous setting. A classification experiment was then performed with the KNN classifier. The main parameters of KNN include k, the number of nearest neighbours and the distance metric. We chose the value for k as 49, the square root of the number of samples in the training set as per our cross-validation settings. A lower value of k can make the system susceptible to noise and overfitting. A higher value means more computations. In addition, data imbalances with respect to classes may dominate the results if k is chosen high. We used Euclidean distance as the distance metric.

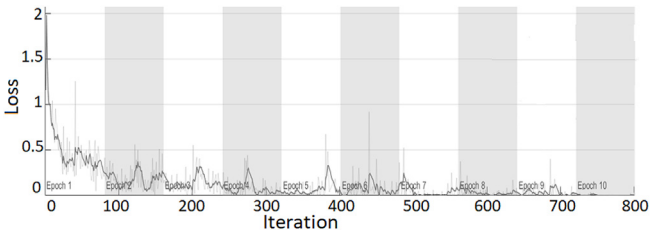


Fig. 6. Loss function of the transfer learned model during training.

Table 1
Experimental parameters.

Model	Parameters	Settings (values)
Transfer learned deep network	initial learning rate	0.0003
	mini-batch size	30
	learning algorithm	Adam cross-entropy
	loss function	10
	maximum epochs	10
	learning factor @ FC layer	
SVM	model sub-type	ECOC hinge
	loss function	one-vs-all
	coding	SVM linear
	learner	L2
	kernel	BFGS
	regularization	
	solver	
KNN	Number of neighbours, k distance	49
		Euclidean

We conducted the experiments five times, and each experiment followed a five-fold cross-validation process. The average of the results after five trials is presented in a mean±standard deviation format.

3.3. Performance metrics and evaluation

Several performance measures are defined for the standard evaluation of a classifier. Classification accuracy is the most extensively used quality index. Accuracy, in classification, is defined as the ratio of the number of correctly classified samples to the total number of data samples. The classification accuracies obtained in our experiments are the following.

- The classification accuracy of the deep transfer learned (stand-alone) model is 92.3±0.7%.
- The accuracy with SVM on deep CNN features is 97.8±0.2%.
- The accuracy with KNN on deep CNN features is 98.0±0.4%.

The data reveals that a superior performance is achieved when SVM or KNN is used to classify the deep CNN features.

Classification accuracy is an effective measure to characterize the performance when the test dataset contains an equal number of samples from each class. However, the dataset considered for the discussed classification problem is an unbalanced dataset. This necessitates further evaluation of the proposed system with more performance indices. We used confusion matrices to study the performance of our tumor classification system. A confusion matrix summarizes correct and incorrect classifications in a tabular form. Table 2 shows a sample confusion matrix for the SVM classifier, obtained during our experimentation. (M, G and P refer to meningioma, glioma and pituitary

Table 2
Confusion matrix for SVM classifier on deep CNN features (where M, G, and P refer to meningioma, glioma and pituitary tumor, respectively).

		Predicted		
		M	G	P
Actual	M	684	11	13
	G	30	1394	2
	P	9	0	921

Table 3
Class-specific evaluation of brain tumor classifier.

Tumor type	Precision	Recall	Specificity
Meningioma	94.7±0.8	96.0±0.5	98.4±0.2
Glioma	99.2±0.3	97.9±0.2	99.4±0.3
Pituitary tumor	98.0±0.7	98.9±0.2	99.1±0.1

tumor, respectively.)

From a confusion matrix, different metrics can be derived to indicate the classifier's performance, specific to each tumor class. Essential metrics are precision, recall (or sensitivity) and specificity and are calculated using the relations given below Eq 1.

$$Precision = \frac{TP}{TP + FP}$$
$$Recall = \frac{TP}{TP + FN}$$
$$Specificity = \frac{TN}{TN + FP} \quad (1)$$

where, TP, FP, TN and FN are the number of classified cases of true positives, false positives, true negatives and false negatives, respectively. Table 3 presents the category-specific performance of the proposed system when the SVM classifier was used with deep CNN features. The specificity values for all the classes are high. This is an indication of correctly identifying samples without a particular disease. The harmonic mean of precision and recall gives another important statistical measure of classification called the F-score, for each class. As there is an imbalance among the three classes, the metric called the average F-score (F_{avg}) is computed by the relation given below Eq 2.

$$F_{avg} = \frac{2}{3}\sum_{c=1}^{3}\frac{Precision_c * Recall_c}{Precision_c + Recall_c} \quad (2)$$

The calculated value of F_{avg} for our system with SVM classifier is 0.97.

3.4. Smaller training data

The original training set, used for validation of one test set, constitutes 80% of the images in the figshare dataset. The theory of transfer learning is recognized as a solution to the problem of scarcity of data to train a deep CNN [13,23]. We tested the capability of our system to perform under the condition when the availability of training data is limited:

1. By randomly selecting 70% of the original training set (i.e. 56% of the total images in figshare)
2. Using 50% of the original training set (i.e. 40% of the total images in figshare)
3. Using 25% of the original training set (i.e. 20% of the total images in figshare)

We trained the transfer learned model with the smaller amount of data. The test set was then provided to the model. The features were extracted and then classified using SVM. Table 4 presents the corresponding performance measures. In addition to the overall

Table 4
Performance with reduced training data.

Training data	Accuracy	AUC (%)		
		Meningioma	Glioma	Pituitary
Full	97.8±0.2	99.5	99.9	99.7
70%	97.1±0.2	99.4	99.7	99.8
50%	95.7±0.5	98.7	99.1	99.7
25%	93.3±0.6	97.8	98.9	99.2

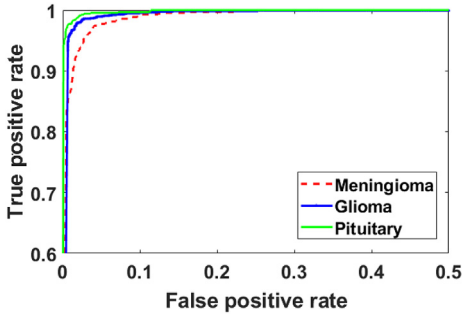


Fig. 7. ROC curves of classifier.

classification accuracy, we used receiver operating characteristic (ROC) curves during analysis. Fig. 7 represents the ROC curves obtained for the case when 50% of the training data is used.

Ideally, the area under the curve (AUC) of ROC is unity. AUC values for the three classes of tumors are shown in Table 4. The values indicate that the reduction in size of the training data has not impaired the system performance significantly. The observation has an advantage in practical scenarios. This is because the number of training samples available could be limited due to the unavailability of more medical data. Another obvious advantage of the smaller training set is a shorter training time. In our experiments, the training time for deep CNN was decreased from 2hr 40min to 55min when the number of training samples was reduced to 25% of the original.

3.5. Comparison with related works

We compared the performance of our method with all the existing methods on the specific 3-class classification problem of brain tumors. Table 5 provides a broad comparison based on classification accuracy as a metric. The comparison shows that our method surpasses all the state-of-the-art methods. The third column in the table defines the portion of the entire dataset used in training. The proposed method recorded the best result when 80% of the data samples was used for training. Instead, we present results for the case when 56% of the dataset is used for training. This is to illustrate the performance of our method with much smaller training data in comparison to related works. The table contains only accuracy as a performance metric because it is the common metric that is used in all the related works. In fact, the proposed work is better than the state-of-the-art methods in terms of all the metrics. Table 6 provides a more detailed comparison. Based on sensitivity and specificity measures, the proposed method is superior to the works [17,19] that used these measures. The proposed method shows an improvement over another method [21] in terms of the F_{avg} score also.

3.6. Regarding misclassifications

Based on the performance evaluation and detailed analysis, the following inferences about the system are made. The accuracy of the system improved when SVM or KNN was used instead of the classification layer within the transfer learned model. This meant that some of

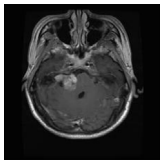
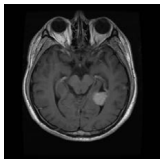
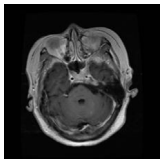
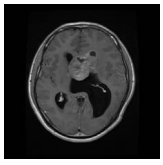
Table 5
Related works & comparison using figshare dataset.

Work	Method	Training data	Accuracy
Jun Cheng [17]	BoW-SVM	80%	91.28%
Ismael [19]	DWT-Gabor-NN	70%	91.90%
Pashaei [21]	CNN-ELM	70%	93.68%
Nyoman [20]	CNN	-	84.19%
Afshar [22]	CapsNet	-	90.89%
Proposed	deep CNN-SVM	56%	97.1%

Table 6
Comparison using other metrics (where M, G, and P refer to meningioma, glioma and pituitary tumor, respectively).

Work	Precision			Recall			Specificity			*F_avg*
	M	G	P	M	G	P	M	G	P	
JunCheng [17]	-	-	-	86.0	96.4	87.3	95.5	96.3	95.3	-
Ismael [19]	-	-	-	86.9	95.1	91.2	96.0	96.3	95.7	-
Pashaei [21]	94.5	91.0	98.3	76.8	97.5	100	-	-	-	0.93
Proposed	94.7	99.2	98.0	96.0	97.9	98.9	98.4	99.4	99.1	0.97

Table 7
Classifier predictions.

Sample Image	Actual	Softmax	SVM
	Meningioma	Pituitary tumor	Pituitary tumor
	Meningioma	Pituitary tumor	Meningioma
	Pituitary tumor	Meningioma	Pituitary tumor
	Glioma	Glioma	Glioma

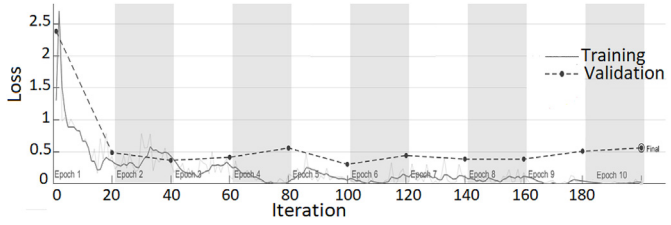


Fig. 8. Training and validation loss on training the model with 25% of training set.

the classifications, which went wrong with the softmax-based classifier (of the deep learning model), were correctly classified by the SVM- and KNN-based classifiers. Table 7 shows a few of the sample instances. From the confusion matrix (Table 2) and the calculated class-specific metrics (Table 3), we find that most of the misclassifications pertain to the class meningioma. This finding can be attributed to the fact that there were fewer samples from this class in the dataset and that no class-specific data augmentation was used to balance the dataset. Another aspect that our work concentrated on was the handling of smaller amounts of training data. We noticed a relative drop in performance (Table 4) with the reduction in training samples. This suggests that the discriminative power of features extracted using the transfer learned deep CNN is affected. This effect was further analysed by studying the variation of training and validation losses against iterations. Fig. 8 shows a sample instance where 25% of the training data was used. In this case, we find that the training loss decreases, whereas the validation loss increases after 60 iterations. This behaviour indicates that the phenomenon of overfitting took place. This is the reason for the lower classification accuracy because the model has learned specific training images without gaining a good generalization capability. This, in turn, means that the model complexity is greater compared to the available training samples. Table 8 represents the confusion matrix for the SVM classifier which was provided with features from the deep transfer learning model trained with 25% of the training set. It is used to study misclassifications as a consequence of overfitting. The class meningioma had the smallest number of samples in the training set and was the most affected class in terms of misclassifications. Overfitting could be avoided either by stopping the training at an earlier stage or by data augmentation. This aspect suggests scope for work in the future.

Table 8
Confusion matrix for SVM classifier on deep CNN features obtained after training with 25% of train data (where M, G, and P refer to meningioma, glioma and pituitary tumor, respectively).

		Predicted		
		M	G	P
Actual	M	614	61	33
	G	65	1358	3
	P	29	8	893

4. Conclusion

This paper presents an accurate and fully automatic system, with minimum pre-processing, for brain tumor classification. The proposed system applied the concept of deep transfer learning to extract features from brain MRI images. The features were used with proven classifier models for an improved performance. The system recorded the best classification accuracy compared to all the related works. The performance was evaluated using other metrics also, to ascertain the robustness of the system. Moreover, the system showed a good performance with a smaller number of training samples. Despite the achievements reported in this paper, several improvements remain possible: First, to the relatively poor performance of the transfer learned model as a stand-alone classifier. Second, there was considerable misclassification of samples from the class meningioma. Third, the phenomenon of overfitting with smaller training data was observed. Future research in the domain shall address these issues, possibly using data augmentation and further tuning of the transfer learned model.

Conflicts of Interest

No conflict of interest

References

[1] S. Kumar, C. Dabas, S. Godara, Classification of brain MRI tumor images: a hybrid approach, Procedia Comput. Sci. 122 (2017) 510–517.

[2] G. Mohan, M.M. Subashini, MRI based medical image analysis: survey on brain tumor grade classification, Biomed. Signal Proces. 39 (2018) 139–161.
[3] M. Yousefi, A. Krzyżak, C.Y. Suen, Mass detection in digital breast tomosynthesis data using convolutional neural networks and multiple instance learning, Comput. Biol. Med. 96 (2018) 283–293.
[4] Y. Gu, X. Lu, L. Yang, B. Zhang, D. Yu, Y. Zhao, T. Zhou, Automatic lung nodule detection using a 3D deep convolutional neural network combined with a multi-scale prediction strategy in chest CTs, Comput. Biol. Med. 103 (2018) 220–231.
[5] H. Zuo, H. Fan, E. Blasch, H. Ling, Combining convolutional and recurrent neural networks for human skin detection, IEEE Signal Process. Lett. 24 (3) (2017) 289–293.
[6] O. Charron, A. Lallement, D. Jarnet, V. Noblet, J.B. Clavier, P. Meyer, Automatic detection and segmentation of brain metastases on multimodal MR images with a deep convolutional neural network, Comput. Biol. Med. 95 (2018) 43–54.
[7] L. Shao, F. Zhu, X. Li, Transfer learning for visual categorization: a survey, IEEE Trans. Neural Netw. Learn. Syst. 26 (5) (2015) 1019–1034.
[8] L. Zhou, Z. Zhang, Y.C. Chen, Z.Y. Zhao, X.D. Yin, H.B. Jiang, A deep learning-based radiomics model for differentiating benign and malignant renal tumors, Transl, Oncol. 12 (2) (2019) 292–300.
[9] E. Deniz, A. Şengür, Z. Kadiroğlu, Y. Guo, V. Bajaj, Ü. Budak, Transfer learning based histopathologic image classification for breast cancer detection, Health Inf. Sci. Syst. 6 (1) (2018) 18.
[10] S. Hussein, P. Kandel, C.W. Bolan, M.B. Wallace, U. Bagci, Lung and pancreatic tumor characterization in the deep learning era: novel supervised and unsupervised learning approaches, IEEE Trans. Med. Imaging (2019), https://doi.org/10.1109/TMI.2019.2894349.
[11] R. Liu, L.O. Hall, D.B. Goldgof, M. Zhou, R.A. Gatenby, K.B. Ahmed, Exploring deep features from brain tumor magnetic resonance images via transfer learning, IEEE International Joint Conference on Neural Networks, IJCNN), 2016, pp. 235–242.
[12] K.B. Ahmed, L.O. Hall, D.B. Goldgof, R. Liu, R.A. Gatenby, Fine-tuning convolutional deep features for MRI based brain tumor classification, International Society for Optics and Photonics Medical Imaging 2017: Computer-Aided Diagnosis, 10134 2017, p. 101342E.
[13] Y. Yang, L.F. Yan, X. Zhang, Y. Han, H.Y. Nan, Y.C. Hu, X.W. Ge, Glioma grading on conventional MR images: a deep learning study with transfer learning, Front. Neurosci. 12 (2018).
[14] M. Talo, U.B. Baloglu, U.R. Acharya, Application of deep transfer learning for automated brain abnormality classification using MR images, Cogn. Syst. Res. 54 (2019) 176–188.
[15] R. Jain, N. Jain, A. Aggarwal, D.J. Hemanth, Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images, Cogn. Syst. Res. (2019), https://doi.org/10.1016/j.cogsys.2018.12.015.
[16] Z.N.K. Swati, Q. Zhao, M. Kabir, F. Ali, A. Zakir, S. Ahmad, J. Lu, Content-based brain tumor retrieval for MR images using transfer learning, IEEE Access 7 (2019) 17809–17822.
[17] J. Cheng, W. Huang, S. Cao, R. Yang, W. Yang, Z. Yun, Q. Feng, Enhanced performance of brain tumor classification via tumor region augmentation and partition, PLoS One 10 (10) (2015) e0140381.
[18] J. Cheng, W. Yang, M. Huang, W. Huang, J. Jiang, Y. Zhou W. Chen, Retrieval of brain tumors by adaptive spatial pooling and Fisher vector representation, PLoS One 11 (6) (2016) e0157112.
[19] M.R. Ismael, I. Abdel-Qader, Brain tumor classification via statistical features and back-propagation neural network, IEEE International Conference on Electro/Information Technology, EIT, 2018, pp. 0252–0257.
[20] N. Abiwinanda, M. Hanif, S.T. Hesaputra, A. Handayani, T.R. Mengko, Brain tumor classification using convolutional neural network, Springer World Congress on Medical Physics and Biomedical Engineering, 2018, pp. 183–189.
[21] A. Pashaei, H. Sajedi, N. Jazayeri, Brain tumor classification via convolutional neural network and extreme learning machines, IEEE 8th International Conference on Computer and Knowledge Engineering, ICCKE, 2018, pp. 314–319.
[22] P. Afshar, K.N. Plataniotis, A. Mohammadi, Capsule networks for brain tumor classification based on MRI images and course tumor boundaries, IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2019, pp. 1368–1372.
[23] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, C. Liu, A survey on deep transfer learning, Springer International Conference on Artificial Neural Networks, 2018, pp. 270–279.
[24] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (10) (2010) 1345–1359.
[25] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, A. Rabinovich, Going deeper with convolutions, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.
[26] Figshare brain tumor dataset, https://doi.org/10.6084/m9.figshare.1512427.v5 , Accessed date: December 2018.
[27] D.X. Xue, R. Zhang, H. Feng, Y.L. Wang, CNN-SVM for microvascular morphological type recognition with data augmentation, J. Med. Biol. Eng. 36 (6) (2016) 755–764.