



Automatic classification of brain magnetic resonance images with hypercolumn deep features and machine learning

Kemal Akyol¹

Received: 11 November 2021 / Accepted: 24 July 2022 / Published online: 23 August 2022
© Australasian College of Physical Scientists and Engineers in Medicine 2022

Abstract

Brain tumours are life-threatening and their early detection is very important in a patient's life. At the present time, magnetic resonance imaging is one of the methods used for detecting brain tumours. Expert decision support systems serve specialist physicians to make more accurate diagnoses by minimizing the errors arising from their subjective opinions in real clinical settings. The model proposed in this study detects important keypoints and then extracts hypercolumn deep features of these keypoints from some convolutional layers of VGG16. Finally, Random Forest and Logistic Regression classifiers are fed with a set of these features. Random Forest classifier offered the best performance with 94.51% accuracy, 91.61% sensitivity, 8.39% false-negative rate, 97.42% specificity, and 97.29% precision using fivefold cross-validation in this study. Consequently, it is thought that the proposed model could contribute to field experts by integrating it into computer-aided brain magnetic resonance imaging diagnosis systems.

Keywords Brain magnetic resonance imaging · Keypoint detection · Hypercolumn deep features · Deep learning

Introduction

Medical imaging analysis plays a vital role in detecting abnormalities. Brain tumour is considered as one of the deadliest diseases [1]. Early detection of brain tumour helps radiologists for an effective prognosis and increases the potentiality of long-term survival [2]. Detection of whether there is a brain tumour is very important in terms of planning the treatment process. Magnetic resonance imaging (MRI) is used to detect the tumour in the brain. Usually, field specialists handle out brain tumour analysis manually. However, MRI screening by field specialist is time-consuming and open to human errors. On the other hand, automatic MRI classification by an expert system decreases the workload of neurologists and helps them make final decisions. Artificial intelligence-based systems that present efficient solutions to biomedical problems have recently attracted researchers. Coronary artery disease [3], eye diseases [4, 5], breast cancer [6–11], brain tumour [12–19], Alzheimer disease [20–27] are just a few of them.

Researchers have proposed many different approaches [15, 28–30] based on image processing and machine learning for brain tumour classification which has attracted widespread attention so far. Classical machine learning approaches use hand-craft feature extraction techniques. Focusing only on low-level or high-level features and relying on handcrafted features are the main problems in the feature extraction phase of traditional machine learning [31]. Expert knowledge is needed in using the handcrafted technique for the features required in solving a machine learning problem. Moreover, detailed knowledge of the task is also required to obtain these features [32]. In addition, intra-class variations and inter-class similarities pose difficulties in obtaining high classification accuracy [33]. Besides these, handcraft feature extraction is expressed as a low-level method in the literature [34]. In line with this reason, good feature extraction is required for successful classification and modelling.

Deep learning architectures, which are sophisticated more than artificial neural networks, provided an opportunity to build models with effective and high performance. They could be successfully applied on the detection and classification of various diseases in the medical field without the need for the feature extraction process. In this context, deep learning-based classification and segmentation studies on brain MRI are quite popular and widespread. For

✉ Kemal Akyol
kakyol@kastamonu.edu.tr

¹ Department of Computer Engineering, Kastamonu University, Kastamonu, Turkey

example, Mehrotra et al. examined the performance of five pre-trained deep learning algorithms on benign and malignant brain tumour images. According to the authors' studies, the AlexNet pre-trained model provided higher performance compared to others [13]. Deepak and Ameer performed Meningioma, Glioma, and Pituitary tumour classification by sending the deep features extracted with GoogleNet pre-trained model as input to the K-Nearest Neighbors and Support Vector Machines [16].

Mohsen et al. classified the brain MRI including normal, glioblastoma, sarcoma, and metastatic bronchogenic carcinoma tumours, using deep neural networks. In the classification step, they used new features by concatenating principal component analysis and discrete wavelet transform features [14]. Kesav and Jibukumar proposed a new architecture for brain tumour classification and tumour-type object detection using region-based convolutional neural networks. In their study, the authors first classified the Glioma and healthy tumour MRI and then identified tumour regions on the Glioma MRI [35].

Tandel et al. proposed a majority voting-based ensemble algorithm to optimize overall classification performance of five convolutional neural networks-based models and also five traditional classifier-based models [36]. Marghalani and Arif identified areas of interest, using the SURF keypoint detection algorithm in their studies. Then, they extracted the features of the areas of interest and built a visual dictionary. Lastly, they classified brain tumour, Alzheimer's disease, and normal brain images, using this visual dictionary [37]. Amin et al. proposed a new method based on fused MRI sequences for brain tumour classification. They used a 23-layer CNN model to classify the segmented images obtained by a global thresholding method [38]. Shafi et al. proposed an ensemble learning method using magnetic resonance images to classify brain tumours or neoplasms and multiple sclerosis. In their study, authors first identified the region of interest about tumour and lesion and obtained the features representing these regions. They conducted majority voting prediction with the support vector machine as the base learner [39].

Swati et al. classified multiclass brain tumours, using transfer learning and block-based fine-tuning with the fivefold cross-validation technique. The authors examined AlexNet, VGG16, and VGG19 pre-trained models, and reported that VGG19 outperformed AlexNet and VGG16 [31]. Paul et al. classified meningioma, glioma, and pituitary tumours from brain images with fully connected and convolutional neural networks models [40]. Abiwinanda et al. applied a simple CNN architecture that includes the convolution, maximum pooling, and smoothing layers followed by fully connected hidden layer to recognize glioma, meningioma, and pituitary [41]. As can be seen, CNN-based studies, specifically transfer learning approaches are widely used in the literature.

Deep learning could perform feature extraction and classification into self-learning, but generally requires a large training dataset for learning. Deep learning application and CNN training from scratch are difficult on small medical datasets [31]. In line with this subject, the analysis of magnetic resonance imaging was carried out, using machine learning and its sub-field deep learning techniques. Moreover, this approach consists of machine learning experiments on the hypercolumn deep features extracted from MRI. First, keypoint detection is conducted on the brain MRI with Oriented FAST and rotated BRIEF (ORB) keypoint detector algorithm [42]. ORB is a computationally efficient keypoint detection algorithm that is less affected by Gaussian image noise [42, 43]. Second, the vectors including hypercolumn deep features are obtained from five different layers of the VGG16 model by taking the positions of the obtained keypoints as reference. Lastly, the feature vectors are sent as input to the random forest (RF) and logistic regression (LR) classifiers. Therefore, the performances of the classifiers on the tumour and normal cases are discussed. In this context, the main contribution of this study is as follows:

- Keypoints are detected on the images in the training set.
- Hypercolumn deep features are extracted from different layers, and then class labelling is performed for each feature set.
- Hypercolumn deep features are sent to RF and LR classifiers. Due to the small number of images in the dataset used in this study, the VGG16 pre-trained architecture is used.
- Finally, a comparative analysis of the different models is carried out for tumour classification.

Material and methodology

The main aim of this study is to propose a new model that consistently and efficiently detects normal and tumour images. Classification experiments include three scenarios within the frame of hold-out and fivefold cross-validation techniques. Scenario A includes the experiments on the VGG16 deep features. Scenarios B and C include the results based on the integration of hypercolumn deep features and traditional classifiers. The hypercolumn deep features are extracted from these points based on the VGG16 architecture, after important keypoints on the image are detected in the Scenario A and Scenario B. All experimental studies in this study were implemented in a 64-bit Windows 10 operating system with Intel @1.80 GHz CPU and 8 GB RAM, and Python 3.6 with Keras 2.3.1 framework.

The general block diagram of the proposed study is shown in Fig. 1. As seen in this block diagram, the class prediction is performed by sending the hypercolumn deep

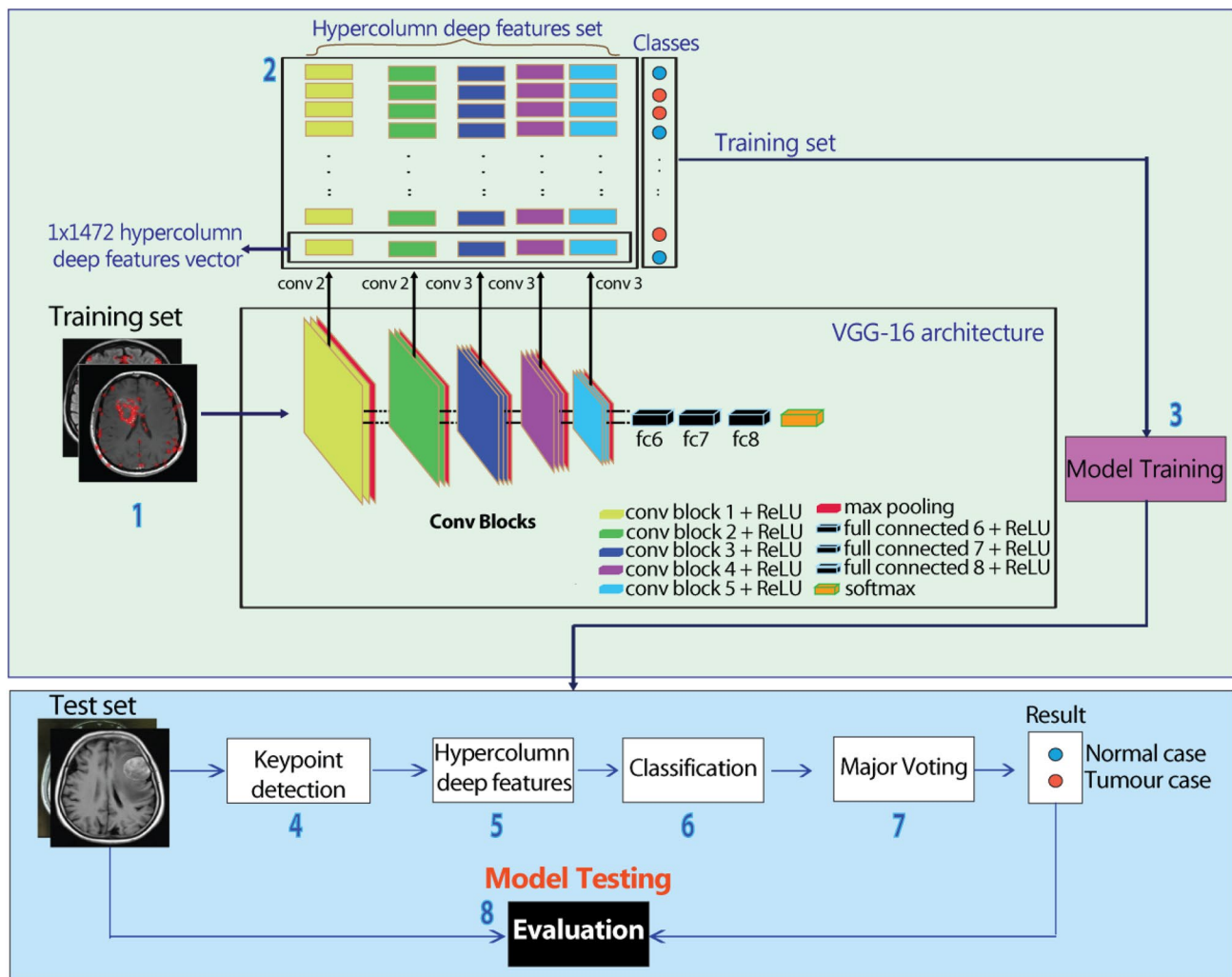


Fig. 1 General block diagram of the proposed model

features as input to the models. RF and LR classifiers predict the target class, using the hypercolumn deep features for each keypoint. Thus, the performances of different classifiers trained on hypercolumn deep features are discussed. Differently from the studies in the literature, the proposed study performs classification for each hypercolumn deep feature obtained from keypoints on the image and lastly votes them. The workflow including the numbers between 1 and 8 in Fig. 1 is as follows:

1. Keypoints on the images in the training set are detected.
2. Hypercolumn deep feature vector is obtained for each keypoint. Also, each vector is labeled with a class.
3. Classifier algorithms are trained on the dataset containing hypercolumn deep feature vectors and their classes.
4. Keypoints are determined for each image in the testing set.

5. Hypercolumn deep feature vector is obtained for each keypoint.
6. Each hypercolumn deep feature vector is classified by trained models.
7. Major voting technique is applied for the final classification of an image.
8. The performances of the classifier models are evaluated for all images in the testing set.

Dataset

In this study, a public brain MRI dataset prepared by Chakrabarty [44] was used. The dataset contains 253 images, including 155 tumours and 98 normal samples. Since the number of normal ones is less than positive samples, the data augmentation technique was applied to images in the normal

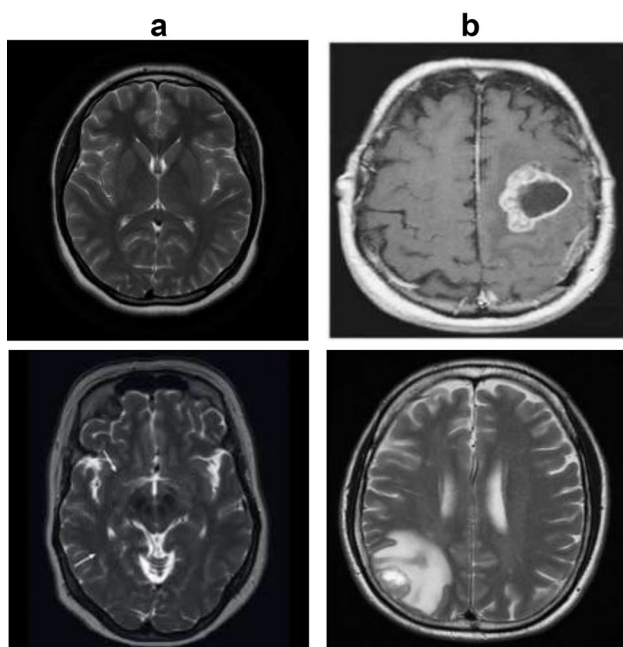


Fig. 2 Samples from each class. From top to bottom: **a** negative, **b** positive

class. So, the dataset was balanced with 155 tumour cases and 155 normal cases. Figure 2 presents sample images.

VGG16 deep learning architecture

VGG16 architecture was developed by Simonyan and Zisserman [45] based on the CNN model which consists of 138 million parameters. VGG16 won the ImageNet Large Scale Visual Recognition Challenge competition held in 2014 with an accuracy score of 92.7%. VGG16 deep learning model was trained on a dataset of 14 million images corresponding to 1000 categories. This model consists of five convolution blocks and fully connected layers [46]. The last layer of each convolution block contains maximum pooling and all hidden layers use the Rectified Linear Unit [47] activation function.

Convolution and maximum pooling of the VGG16 present the feature vector representing the image. Five convolution blocks are followed by fully connected layers where feature vectors are transformed into 1-dimensional arrays. The soft-max function in the last layer of this architecture is used to classify binary and multi-class data. This function is an activation function used to determine the most likely class [48].

Hypercolumn deep features

In the CNN architecture, all features extracted from convolutional layers have a role in the classification, and the output layer represents the final features to be used in classification.

Hypercolumn deep features are obtained by combining intermediate features extracted from the layers of the CNN architecture into a single vector [49]. Hypercolumn is a vector of activation outputs obtained from CNN layers for any pixel in the image [50]. With the hypercolumn deep feature extraction technique, more accurate prediction is performed using the features that represent previous layers. Many studies using this technique such as Alzheimer's disease classification [51], tumour region detection [52], brain magnetic resonance imaging classification [50] are available in the literature.

Performance metrics

There are two classes in the dataset as tumour (positive) and normal (negative). The confusion matrix obtained in two-class classifications includes true positive (TP), true negative (TN), false positive (FP), and false negative (FN) basic criteria. Here, the TP refers to the number of positive images that were correctly classified, while the TN refers to the number of correctly classified negative images. In addition, the FN gives the number of misclassified positive images while the FP gives the number of misclassified negative images. The performances of models are measured using different metrics such as accuracy (Acc), sensitivity (Sen), false negative rate (FNR), specificity (Spe), and precision (Pre). Sen called as True Positive Rate (TPR) refers to the ratio of positive samples correctly classified to all positive samples, while Spe refers to the ratio of negative samples correctly classified to all negative samples. FNR is the ratio of misclassified tumour cases to all tumour cases. Accuracy indicates the general classification performance by calculating the ratio of correctly classified samples out of all samples. Precision gives the ratio of positive samples correctly classified to all cases predicted positively. These metrics were given in between Eqs. 1, 4, respectively.

$$Sen(TPR) = \frac{TP}{TP + FN} \quad (1)$$

$$FNR = \frac{FN}{FN + TP} \quad (2)$$

$$Spe = \frac{TN}{TN + FP} \quad (3)$$

$$Acc = \frac{TN + TP}{TP + FN + TN + FP} \quad (4)$$

$$Pre = \frac{TP}{TP + FP} \quad (5)$$

Fig. 3 Model performance evaluation technique: a) hold-out b) fivefold cross-validation

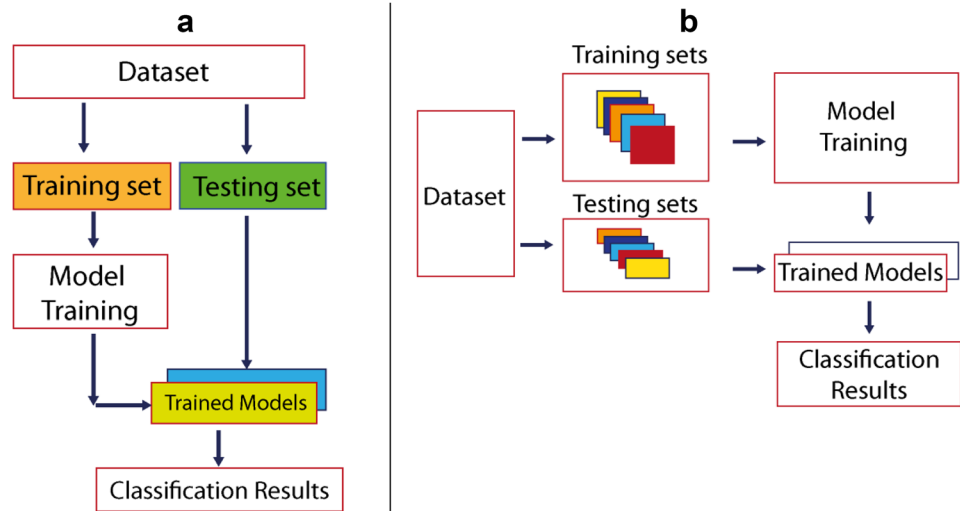


Table 1 Information about the train and test sets

Original dataset	Train (80%)	Test (20%)	Total
Classes			
Normal cases	124	31	155
Tumour cases	124	31	155
Total	248	62	310

Model training

Before model training, firstly, data augmentation techniques were applied to some images in the negative class to eliminate the unbalanced distribution problem on the dataset and improve the model performance. Image rotation between 5 and 15 degrees were applied to images in the negative class, and then 57 images were randomly selected from the augmented sample images subset. Therefore, each class consists of 155 images.

In this study, machine learning experiments were carried out by utilizing hold-out and fivefold cross-validation techniques on the balanced dataset. The hold-out technique involves the performances of traditional classifiers on the deep features extracted from pre-trained VGG16 model based on the transfer learning approach (Scenario A) and the performances of them on the hypercolumn deep features (Scenario B). With the hold-out technique, as shown in Fig. 3a, the dataset was split into 80:20 ratios for training and testing sets, respectively. So, 248 images were reserved for the training of the models and 62 images for the testing. Table 1 summarized the sub-dataset distribution according to hold-out technique.

In the last scenario (Scenario C) which is the fivefold cross-validation technique ($k = 5$), as shown in Fig. 3b,

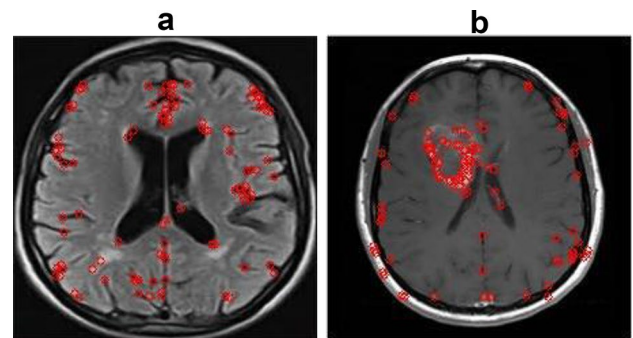


Fig. 4 Keypoints detected on the sample images for each class; From left to right; **a** normal case, **b** tumour case. (red points are keypoints detected on the image)

while 80% of the dataset, which is $k - 1$ parts of the dataset for each fold, was reserved for the model training, the remaining part (20%) was reserved for model testing. Experiments that include training and testing of the models were carried out five-times in Scenario C in this context. In Scenarios B and C, important keypoints required for training of the models were detected on the train set utilizing the ORB detector. Figure 4 shows the keypoints detected on some sample images. Hypercolumn deep feature vectors regarding these keypoints were extracted from certain convolutional layers given in Table 2. All layers in the CNN architecture could be used to extract features, but it causes system resource problems. As a result of try-and-error experiments with combinations of different layers, these layers were selected to extract hypercolumn deep features. Feature maps were upscaled to 224×224 when extracting these features. All hypercolumn feature vectors extracted from tumour and normal images are assigned to the related class.

Table 2 The layers of VGG16 used for hypercolumn deep features extracting

Layer no	Layer name	Explain
2	block1_conv2	Convolutional layer
5	block2_conv2	Convolutional layer
9	block3_conv3	Convolutional layer
13	block4_conv3	Convolutional layer
17	block5_conv3	Convolutional layer

Dynamically obtaining the hypercolumn feature vector with the ORB keypoint detector has a limitation for CNN models because CNN-based models get input in $m \times n$ sizes which m and n indicate the width and height of the image or matrix, respectively. The fact that detecting different keypoints from each image is a handicap to obtain $m \times n$ matrix structure. For example, 19 keypoints can be detected from any image, while 189 keypoints from another image. Here, the numbers 19 and 189 are given symbolically. In this case, 19×1472 hypercolumn deep features are extracted from the first image, while 189×1472 hypercolumn deep features are extracted from the other image. 1472 is the dimension of the feature vector formed by combining the features extracted from different layers in VGG16 for any keypoint. Therefore, a limitation is available for the feature vector, which is expected as a matrix. It is inevitable to make manipulations on the feature vector to overcome this problem. Since the input data for CNN must be a matrix in $m \times n$ sizes, there must be an equal number of keypoints for each image. This problem can only be overcome by:

- Selecting randomly p keypoints from each image which has many keypoints
- Filling the missing parts of the $m \times n$ matrix with zeros

Both approaches have disadvantages. While the first one causes the loss of meaningful information, the second one causes overfitting. In the proposed approach, hypercolumn deep features extracted with reference to the keypoints detected from each image in the training set were matched with the class of the image. Thus, dynamically detecting a different number of keypoint from the images is not a problem. In other words, hypercolumn deep features are dynamically extracted and target class information is assigned for each hypercolumn deep feature. So, the proposed approach in this study overcomes the disadvantages mentioned above. After hypercolumn deep features were extracted from images in the training dataset, traditional classifier-based models were trained on these features. After the model was built, the keypoint detection on the test dataset and the extraction of hypercolumn deep features from the keypoints were carried out. Then, each of the hypercolumn deep features was

Table 3 Some parameters settings of the traditional classifiers

Classifier	Parameters with values
RF	$n_estimators = 100$, criterion = 'gini', max_ features = 'auto'
LR	solver = 'lbfgs', max_iter = 100, multi_ class = 'auto'

classified separately, and then, taking into account the hypercolumn deep features and class matches, the class of image was assigned to the class with the highest number of labels with the major voting approach. In the case of an equal number of positive and negative classifications, dynamically keypoint detection is performed again on the image. As a result of the classifications made for each point, an equal number of positive and negative classifications were not encountered in the experiments. Some parameters of the RF and SVM classifiers used with their default parameter settings are listed in Table 3.

Results

The performances of models were tested using 3 different scenarios listed below:

- Scenario A. Results obtained on the VGG16 deep features with hold-out technique.
- Scenario B. Results obtained on the hypercolumn deep features with hold-out technique.
- Scenario C. Results obtained on the hypercolumn deep features with fivefold cross-validation technique.

In Scenario A, the classification performances of the RF and LR models on the deep features extracted from the VGG16 architecture based on the transfer learning approach were summarized with confusion matrices in Fig. 5. Confusion matrices could also be presented by coloring as here. There are many examples of this in the literature. There is a transition from light to dark in direct proportion to the number of mistaken classifications in the confusion matrices presented in this study. As can be seen in the confusion matrices, the RF misclassified 5 of 31 images in the normal class as the tumour, while LR misclassified 9 of 31 images in the normal class as the tumour. In addition, RF correctly classified 26 of 31 images with tumour while LR correctly classified 23 of them.

Scenario B presents the classification performances of RF and LR classifiers on the hypercolumn deep features extracted from the images in the test dataset. Figure 6 summarized the confusion matrices obtained by the classifiers

in the experiments. As seen in here, RF has less misclassification compared to LR. While RF misclassified 3 of 31 images in the normal class, LR misclassified 4 of 31 images in the normal class. In addition, the RF classifier correctly classified 30 of 31 images with tumour while misclassifying 1 of them. LR classifier correctly classified 29 of 31 images with tumour while misclassified 2 of them. Table 4 summarizes the performance of the models on the hypercolumn deep features and VGG16 deep features within the frame of the hold-out technique.

Within the frame of Scenario C, the confusion matrices obtained by RF and LR classifiers for each fold were presented in Fig. 7. As can be seen in the confusion matrices between Fold 1 and Fold 5, the RF classifier showed better performance than LR in the proposed approach. Table 5 summarized the performance of the models on the

hypercolumn deep features within the frame of the fivefold cross-validation technique. The average Acc, Sen, FNR, Spe, and Pre values of RF and LR classifiers were presented in bold in this table.

The performances of the classifiers on the hypercolumn deep features in Scenarios B and C are better than the VGG16 deep features achieved with Scenario A. Overall results show that hypercolumn deep features improved the RF and LR accuracies to 93.55% from 83.87% and to 90.32% from 72.58 when compared to VGG16 deep features within the frame of the hold-out technique. In addition, hypercolumn deep features in the fivefold cross-validation improved the RF accuracy to 94.51% from 93.55% when compared to hypercolumn deep features with the hold-out technique. Moreover, the FNRs of the classifiers used in this study were also examined in detail. Figure 8 demonstrates the FNR

Fig. 5 Confusion matrices of the classifiers on the VGG16 deep features with hold-out technique

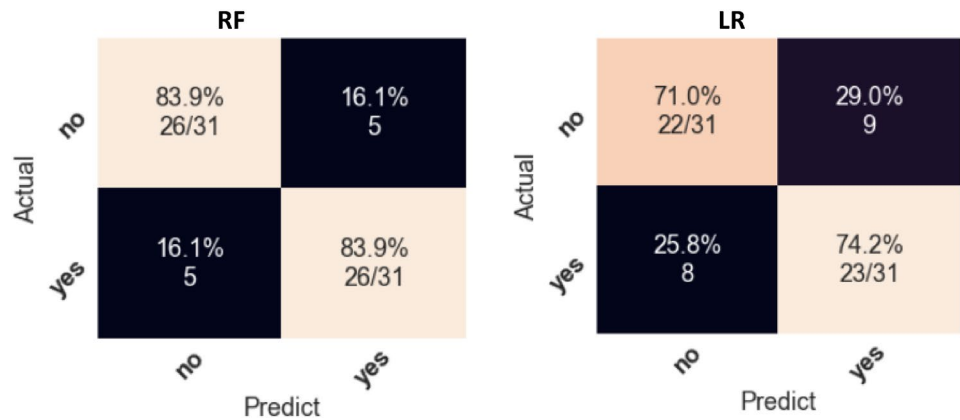


Fig. 6 Confusion matrices of the classifiers on the hypercolumn deep features with hold-out technique

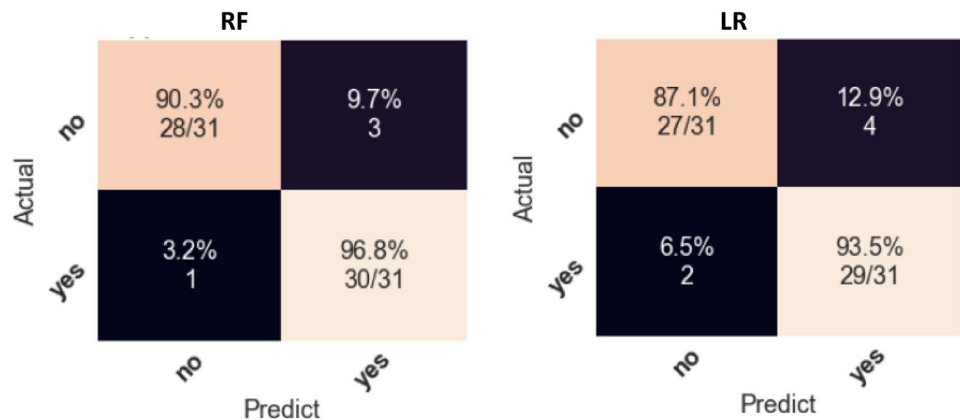


Table 4 Classification results for hypercolumn deep features and VGG16 deep features with hold-out technique

	Classifier	Acc	Sen	FNR	Spe	Pre
VGG16 deep features	RF	83.87	83.87	16.13	83.87	83.87
	LR	72.58	74.19	25.81	70.97	71.88
Hypercolumn deep features	RF	93.55	96.77	3.23	90.32	90.91
	LR	90.32	93.55	6.45	87.1	87.88

Bold values indicate the best results in Table 4

Fig. 7 Confusion matrices of the classifiers on the hypercolumn deep features with fivefold cross-validation technique

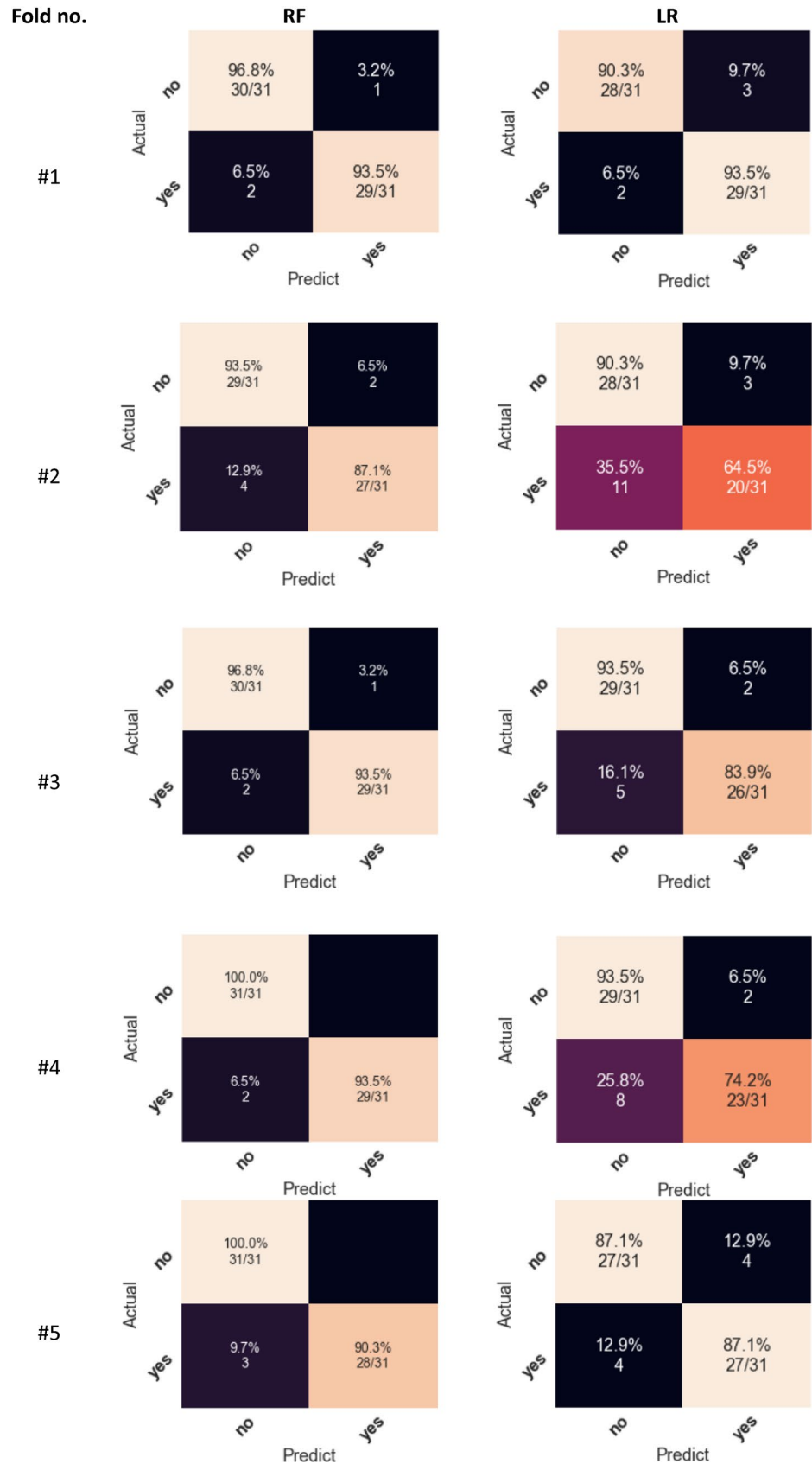


Table 5 Fivefold cross-validation results on the hypercolumn deep features

Fold no	Classifier	Acc	Sen	FNR	Spe	Pre
1	RF	95.16	93.55	6.45	96.77	96.67
	LR	91.94	93.55	6.45	90.32	90.62
2	RF	90.32	87.1	12.9	93.55	93.1
	LR	77.42	64.52	35.48	90.32	86.96
3	RF	95.16	93.55	6.45	96.77	96.67
	LR	88.71	83.87	16.13	93.55	92.86
4	RF	96.77	93.55	6.45	100.0	100.0
	LR	83.87	74.19	25.81	93.55	92.0
5	RF	95.16	90.32	9.68	100.0	100.0
	LR	87.1	87.1	12.9	87.1	87.1
Average	RF	94.51	91.61	8.39	97.42	97.29
	LR	85.81	80.65	19.35	90.97	89.91

Bold values indicate the best results in Table 5

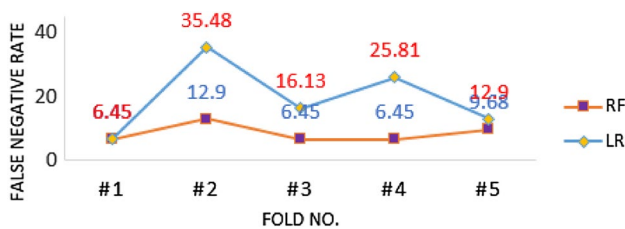


Fig. 8 FNRs of the classifiers

values of the classifiers with the fivefold cross-validation technique. Accordingly, the LR misclassified more tumor cases as normal in each experiment except fold #1 compare to RF. These experiments show that the proposed approach is feasible and effective for classifying brain tumours.

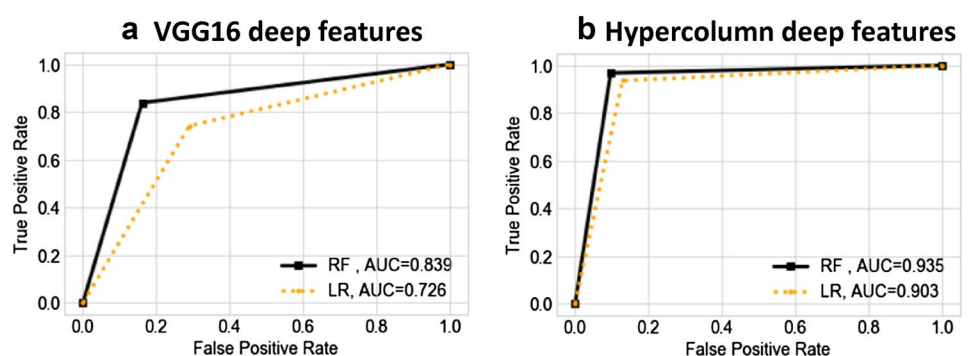
The Receiver operating characteristic (ROC) curve shows the relationship between true positive rate (TPR) and false positive rate (FPR). TPR also called Sensitivity gives the ratio of correctly predicted positives to all positives. FPR is the ratio of falsely predicted negatives to all negative data. The area under the curve (AUC) value represents the area under the ROC curve. If this value is close to 1, it indicates that the model presented high success in separating

the classes, and close to 0 indicates that the model presented low success [53]. While Fig. 9 presents the ROC curves with AUC values that show the performances of RF and LR algorithms on the hypercolumn deep features and VGG16 deep features Fig. 10, demonstrates the ROC curves with AUC values obtained by utilizing the fivefold cross-validation. According to the results, the performances of the classifiers on the hypercolumn deep features are quite high when compared to the VGG16 deep features. When compared the Scenario A and Scenario B, it can be seen that the AUC values of RF are better than LR. In addition, 0.945 average AUC value of RF and 0.858 average AUC value of LR confirm the results of fivefold cross-validation. This indicates that the RF has good classification performance.

Discussion

Table 6 summarizes the results of this study and other studies in the literature. As seen in the ‘Method’ column from this table, more studies were conducted with classical deep learning approaches. For example, Deepak and Ameer obtained 97.1% accuracy using deep features and SVM classification with the fivefold cross-validation technique [16].

Fig. 9 ROC results with hold-out technique; **a** VGG16 deep features, **b** Hypercolumn deep features



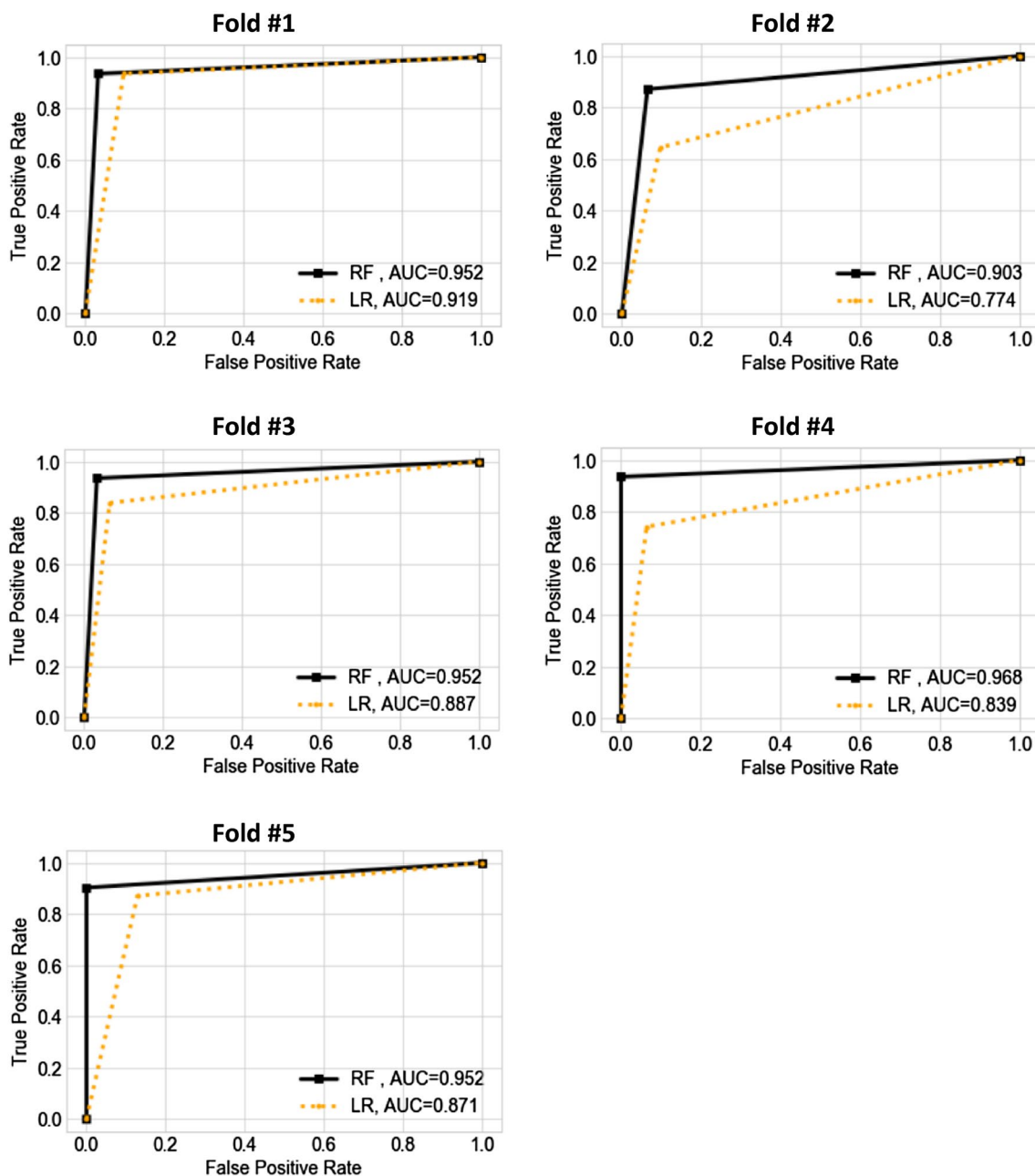


Fig. 10 ROC results for hypercolumn deep features with fivefold cross-validation technique

Swati et al. classified three brain tumours using transfer learning and block-based fine-tuning. They achieved 94.82% accuracy with the fivefold cross-validation technique [31]. Paul et al. achieved 91.43% accuracy using fully connected and convolutional neural networks models with the fivefold cross-validation technique [40]. Abiwinanda et al. obtained 84.19% accuracy with CNN fine-tuned model using the hold-out validation technique [41]. Arı and Hanbay [54] achieved 97.18% classification accuracy with the deep learning-based model on the cranial MR images. But, they used quite a few images with nine patients and seven patients for training

and testing, respectively with hold-out validation technique in the study. Toğaçar et al. [50] achieved 96.77% accuracy in their study based on the selection of important ones from hypercolumn deep features. They were applied the feature selection technique with the recursive feature elimination technique on the hypercolumn deep features obtained with pre-trained models. Then, they classified the hypercolumn deep features with the SVM classifier. But, they divided the dataset into 70% training and 30% test set with the hold-out technique which is called simple validation in the study. Different from the study of Toğaçar et al. [50], this study

Table 6 Comparison of proposed model and related studies

Study	Method	Evaluation method	# of class	Acc (%)
Deepak and Ameer [16]	Deep CNN-SVM	Fivefold cross-validation	3	97.1
Swati et al. [31]	VGG-19	Fivefold cross-validation	3	94.82
Paul et al. [40]	Deep learning	Fivefold cross validation	3	91.43
Abiwinanda et al. [41]	CNN + Hyper-parameters optimization	Hold-out	3	84.19
Arı and Hanbay [54]	Deep learning, the extreme learning machine local receptive fields	Hold-out	2	97.18
Toğaçar et al. [50]	Hyper column technique with convolutional neural network and feature selection method	Hold-out	2	96.77
The proposed study	Hyper-column deep features and random forest	Hold-out, Fivefold cross-validation	2	93.55, 94.51

#Bold values indicate the results obtained in this study

proposes a novel approach including keypoint detection, hypercolumn deep features, and machine learning experiments validated with the fivefold cross-validation technique to classify the brain MRI. The model proposed in this study showed very successful classification accuracy with 94.51%.

Conclusion

Detection of brain tumour is very important in terms of the fast and planned treatment process and increasing the survival rate of the patient. In this study, a novel model that classifies normal and brain tumour cases was proposed. This model extracts hypercolumn deep features from specific layers of the VGG16 deep learning architecture for the images and classifies them with the RF classifier. Major voting technique reaches the final decision regarding the classifications for each hypercolumn deep feature vector dynamically extracted from each image. The performance of the proposed model was validated with hold-out and fivefold cross-validation techniques. Handcrafted features are not used and minimum pre-processing is required in the experiments in the proposed study. The proposed approach offered the best performance with 94.51% accuracy, 91.61% sensitivity, 97.42% specificity, and 97.29% precision on the hypercolumn deep features. The proposed model could play an important role in the development of expert systems to be used in real clinical environments, taking into account its high classification accuracy. Such an expert system to be developed using a low-cost system based on deep learning architecture will help specialist physicians to make a more accurate diagnosis by minimizing errors arising from their subjective opinions.

A model with high generalization capacity on the images obtained from different resources is planned for future work. Also, using other deep architectures that could achieve a better performance in computer vision in future studies is aimed. In addition, the studies based on the different deep

learning architectures on state-of-the-art images obtained by deep dream and blurry image techniques are among the future studies.

Acknowledgements Author would like to thank Chakrabarty [44] to provide the public brain MRI dataset.

Funding Not applicable.

Data availability This study uses public dataset provided by Chakrabarty [44].

Declarations

Conflict of interest The author declares that he has no conflict of interest.

References

1. Sharif M, Amin J, Raza M, Yasmin M, Satapathy SC (2020) An integrated design of particle swarm optimization (PSO) with fusion of features for detection of brain tumor. *Pattern Recogn Lett* 129:150–157. <https://doi.org/10.1016/J.PATREC.2019.11.017>
2. Sharif MI, Li JP, Naz J, Rashid I (2020) A comprehensive review on multi-organs tumor detection based on machine learning. *Pattern Recogn Lett* 131:30–37. <https://doi.org/10.1016/J.PATREC.2019.12.006>
3. JayaSree M, Koteswara Rao L (2020) Survey on—Identification of coronary artery disease using deep learning. *Mater Today Proc.* <https://doi.org/10.1016/j.matpr.2020.09.526>
4. Badar M, Haris M, Fatima A (2020) Application of deep learning for retinal image analysis: a review. *Comput Sci Rev.* <https://doi.org/10.1016/j.cosrev.2019.100203>
5. Perdomo O, Rios H, Rodríguez FJ, Otálora S, Meriaudeau F, Müller H, González FA (2019) Classification of diabetes-related retinal diseases using a deep learning approach in optical coherence tomography. *Comput Method Programs Biomed* 178:181–189. <https://doi.org/10.1016/j.cmpb.2019.06.016>
6. Liu T, Huang J, Liao T, Pu R, Liu S, Peng Y (2021) A hybrid deep learning model for predicting molecular subtypes of human breast cancer using multimodal data. *IRBM.* <https://doi.org/10.1016/j.irbm.2020.12.002>

7. Prasath Alias Surendhar S, Vasuki R (2021) Breast cancers detection using deep learning algorithm. *Mater Today Proc.* <https://doi.org/10.1016/j.matpr.2020.11.600>
8. Gangaputra S, Lovato JF, Hubbard L, Davis M, Esser BA, Ambrosiud WT, ACCORD Eye Research Group (2013) Comparison OF standardized clinical classification with fundus photograph GRADING for the assessment of diabetic retinopathy and diabetic macular Edema severity. *Retina* 33(7):1393–1399. <https://doi.org/10.1097/IAE.0b013e318286c952>
9. Pang T, Wong JHD, Ng WL, Chan CS (2020) Deep learning radiomics in breast cancer with different modalities: overview and future. *Exp Syst Appl.* <https://doi.org/10.1016/j.eswa.2020.113501>
10. Singla C, Sarangi PK, Sahoo AK, Singh PK (2020) Deep learning enhancement on mammogram images for breast cancer detection. *Mater Today Proc.* <https://doi.org/10.1016/j.matpr.2020.10.951>
11. Houssein EH, Emam MM, Ali AA, Suganthan PN (2020) Deep and machine learning techniques for medical imaging-based breast cancer: a comprehensive review. *Exp Syst Appl.* <https://doi.org/10.1016/j.eswa.2020.114161>
12. Saranya C, Priya JG, Jayalakshmi P, Pavithra EH (2021) Brain tumor identification using deep learning. *Mater Today Proc.* <https://doi.org/10.1016/j.matpr.2020.11.555>
13. Mehrotra R, Ansari MA, Agrawal R, Anand RS (2020) A Transfer Learning approach for AI-based classification of brain tumors. *Machine Learn Appl* 2:100003. <https://doi.org/10.1016/j.mlwa.2020.100003>
14. Mohsen H, El-Dahshan E-SA, El-Horbaty E-SM, Salem A-BM (2018) Classification using deep learning neural networks for brain tumors. *Future Comput Info J* 3(1):68–71. <https://doi.org/10.1016/j.fcij.2017.12.001>
15. Saba T, Sameh Mohamed A, El-Affendi M, Amin J, Sharif M (2020) Brain tumor detection using fusion of hand crafted and deep learning features. *Cogn Syst Res* 59:221–230. <https://doi.org/10.1016/j.cogsys.2019.09.007>
16. Deepak S, Ameer PM (2019) Brain tumor classification using deep CNN features via transfer learning. *Comput Biol Med* 111:103345. <https://doi.org/10.1016/j.combiomed.2019.103345>
17. Hashemzahi R, Mahdavi SJS, Kheirabadi M, Kamel SR (2020) Detection of brain tumors from MRI images base on deep learning using hybrid model CNN and NADE. *Biocybern Biomed Eng* 40(3):1225–1232. <https://doi.org/10.1016/j.bbe.2020.06.001>
18. Gupta A, Verma A, Kaushik D, Garg M (2020) Applying deep learning approach for brain tumor detection. *Mater Today Proc.* <https://doi.org/10.1016/j.matpr.2020.10.063>
19. Srikanth B, Venkata Suryanarayana S (2021) Multi-Class classification of brain tumor images using data augmentation with deep neural network. *Mater Today Proc.* <https://doi.org/10.1016/j.matpr.2021.01.601>
20. Li H, Habes M, Wolk DA, Fan Y (2019) A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal magnetic resonance imaging data. *Alzheimer's Dementia* 15(8):1059–1070. <https://doi.org/10.1016/j.jalz.2019.02.007>
21. Abuhmed T, El-Sappagh S, Alonso JM (2021) Robust hybrid deep learning models for Alzheimer's progression detection. *Knowl-Based Syst* 213:106688. <https://doi.org/10.1016/j.knosys.2020.106688>
22. Bringas S, Salomón S, Duque R, Lage C, Montaña JL (2020) Alzheimer's Disease stage identification using deep learning models. *J Biomed Info* 109:103514. <https://doi.org/10.1016/j.jbi.2020.103514>
23. Puente-Castro A, Fernandez-Blanco E, Pazos A, Munteanu CR (2020) Automatic assessment of Alzheimer's disease diagnosis based on deep learning techniques. *Comput Biol Med* 120:103764. <https://doi.org/10.1016/j.combiomed.2020.103764>
24. Raza M, Awais M, Ellahi W, Aslam N, Nguyen HX, Le-Minh H (2019) Diagnosis and monitoring of Alzheimer's patients using classical and deep learning techniques. *Expert Syst Appl* 136:353–364. <https://doi.org/10.1016/j.eswa.2019.06.038>
25. Zhou Y, Lu Y, Pei Z (2021) Intelligent diagnosis of Alzheimer's disease based on internet of things monitoring system and deep learning classification method. *Microprocess Microsyst* 83:104007. <https://doi.org/10.1016/j.micpro.2021.104007>
26. Bi X, Li S, Xiao B, Li Y, Wang G, Ma X (2020) Computer aided Alzheimer's disease diagnosis by an unsupervised deep learning technology. *Neurocomputing* 392:296–304. <https://doi.org/10.1016/j.neucom.2018.11.111>
27. An N, Ding H, Yang J, Au R, Ang TFA (2020) Deep ensemble learning for Alzheimer's disease classification. *J Biomed Inform* 105:103411. <https://doi.org/10.1016/j.jbi.2020.103411>
28. Ramkumar G, Thandaiah Prabu R, Phalguni Singh N, Maheswaran U (2021) Experimental analysis of brain tumor detection system using Machine learning approach. *Mater Today Proc.* <https://doi.org/10.1016/j.matpr.2021.01.246>
29. Sabitha V, Nayak J, Reddy PR (2021) MRI brain tumor detection and classification using KPCA and KSVM. *Mater Today Proc.* <https://doi.org/10.1016/J.MATPR.2021.01.714>
30. Kaplan K, Kaya Y, Kuncan M, Ertunç HM (2020) Brain tumor classification using modified local binary patterns (LBP) feature extraction methods. *Med Hypotheses* 139:109696. <https://doi.org/10.1016/j.mehy.2020.109696>
31. Swati ZNK, Zhao Q, Kabir M, Ali F, Ali Z, Ahmed S, Lu J (2019) Brain tumor classification for MR images using transfer learning and fine-tuning. *Comput Med Imaging Graph* 75:34–46. <https://doi.org/10.1016/J.COMPIMMAG.2019.05.001>
32. Shoeibi A, Ghassemi N, Alizadehsani R, Rouhani M, Hosseini-Nejad H, Khosravi A, Nahavandi S (2021) A comprehensive comparison of handcrafted features and convolutional autoencoders for epileptic seizures detection in EEG signals. *Expert Syst Appl* 163:113788. <https://doi.org/10.1016/J.ESWA.2020.113788>
33. Ghazal S, Qureshi WS, Khan US, Iqbal J, Rashid N, Tiwana MI (2021) Analysis of visual features and classifiers for fruit classification problem. *Comput Electron Agric* 187:106267. <https://doi.org/10.1016/J.COMPAG.2021.106267>
34. Zhen X, Shao L, Maybank SJ, Chellappa R (2016) Handcrafted vs. learned representations for human action recognition. *Image Vis Comput* 55:39–41. <https://doi.org/10.1016/J.IMAVIS.2016.10.002>
35. Kesav N, Jibukumar MG (2021) Efficient and low complex architecture for detection and classification of brain tumor using RCNN with Two channel CNN. *J King Saud Univ.* <https://doi.org/10.1016/J.JKSUCI.2021.05.008>
36. Tandel GS, Tiwari A, Kakde OG (2021) Performance optimisation of deep learning models using majority voting algorithm for brain tumour classification. *Comput Biol Med* 135:104564. <https://doi.org/10.1016/J.COMPBIOMED.2021.104564>
37. Marghalani BF, Arif M (2019) Automatic classification of brain tumor and Alzheimer's disease in MRI. *Proc Comput Sci* 163:78–84. <https://doi.org/10.1016/J.PROCS.2019.12.089>
38. Amin J, Sharif M, Gul N, Yasmin M, Shad SA (2020) Brain tumor classification based on DWT fusion of MRI sequences using convolutional neural network. *Pattern Recogn Lett* 129:115–122. <https://doi.org/10.1016/J.PATREC.2019.11.016>
39. Shafi ASM, Rahman MB, Anwar T, Halder RS, Kays HME (2021) Classification of brain tumors and auto-immune disease using ensemble learning. *Info Med Unlock* 24:100608. <https://doi.org/10.1016/J.IMU.2021.100608>
40. Paul JS, Plassard AJ, Landman BA, Fabbri D (2017) Deep learning for brain tumor classification. *Med Imaging* 2017:10137. <https://doi.org/10.1117/12.2254195>

41. Abiwinanda N, Hanif M, Hesaputra ST, Handayani A, Mengko TR (2019) Brain tumor classification using convolutional neural network. *IFMBE Proc* 68(1):183–189. https://doi.org/10.1007/978-981-10-9035-6_33
42. Rublee E, Rabaud V, Konolige K, Bradski G (2011) ORB: an efficient alternative to SIFT or SURF. In 2011 international conference on computer vision. *IEEE*. <https://doi.org/10.1109/ICCV.2011.6126544>
43. Vinay A, Deshpande AR, Pranathi BS, Jha H, Murthy KNB, Natarajan S (2018) Effective descriptors based face recognition technique for robotic surveillance systems. *Proc Comput Sci* 133:968–975. <https://doi.org/10.1016/J.PROCS.2018.07.080>
44. Chakrabarty N (2019) Brain MRI images for brain tumor detection. Kaggle. <https://www.kaggle.com/navoneel/brain-mri-images-for-brain-tumor-detection/metadata>. Accessed 6 Mar 2021
45. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*
46. Mishra NK, Singh P, Joshi SD (2021) Automated detection of COVID-19 from CT scan using convolutional neural network. *Biocybern Biomed Eng* 41(2):572–588. <https://doi.org/10.1016/J.BBE.2021.04.006>
47. Nair V, Hinton GE (2010) Rectified linear units improve restricted boltzmann machines. *Proceedings of the 27th international conference on international conference on machine learning*. <https://doi.org/10.5555/3104322.3104425>
48. Saravanan TM, Karthiha K, Kavinkumar R, Gokul S, Mishra JP (2022) A novel machine learning scheme for face mask detection using pretrained convolutional neural network. *Mater Today Proc* 58:150–156. <https://doi.org/10.1016/J.MATPR.2022.01.165>
49. Hariharan B, Arbelaez P, Girshick R, Malik J (2017) Object instance segmentation and fine-grained localization using hypercolumns. *IEEE Trans Pattern Anal Mach Intell* 39(4):627–639. <https://doi.org/10.1109/TPAMI.2016.2578328>
50. Toğaçar M, Cömert Z, Ergen B (2020) Classification of brain MRI using hyper column technique with convolutional neural network and feature selection method. *Expert Syst Appl* 149:113274. <https://doi.org/10.1016/j.eswa.2020.113274>
51. Toğaçar M, Cömert Z, Ergen B (2021) Enhancing of dataset using deepdream, fuzzy color image enhancement and hypercolumn techniques to detection of the Alzheimer's disease stages by deep learning model. *Neural Comput Appl*. <https://doi.org/10.1007/s00521-021-05758-5>
52. Wijethilake N, Islam M, Ren H (2020) Radiogenomics model for overall survival prediction of glioblastoma. *Med Biol Eng Compu* 58(8):1767–1777. <https://doi.org/10.1007/s11517-020-02179-9>
53. Fawcett T (2006) An introduction to ROC analysis. *Pattern Recogn Lett* 27(8):861–874. <https://doi.org/10.1016/J.PATREC.2005.10.010>
54. Ari A, Hanbay D (2018) Deep learning based brain tumor classification and detection system. *Turk J Elec Eng Comp Sci*. <https://doi.org/10.3906/elk-1801-8>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.