Journal of Radiation Research and Applied Sciences 15 (2022) 33–39

Contents lists available at ScienceDirect

Journal of Radiation Research and Applied Sciences

journal homepage: www.journals.elsevier.com/journal-of-radiation-research-and-applied-sciences

MR image normalization dilemma and the accuracy of brain tumor classification model

Majdi Alnowami [a,*], Eslam Taha [a,b], Saeed Alsebaeai [a], Syed Muhammad Anwar [c], Abdulsalam Alhawsawi [a,b]

[a] Department of Nuclear Engineering, Faculty of Engineering, King Abdulaziz University, P.O. Box 80204, Jeddah, 21589, Saudi Arabia
[b] Center for Training & Radiation Prevention, King Abdulaziz University, P.O. Box 80204, Jeddah, 21589, Saudi Arabia
[c] Department of Software Engineering, University of Engineering and Technology, Taxila, 47050, Pakistan

ARTICLE INFO

Keywords:
Artificial intelligence
Brain cancer
DenseNet
MRI
Pre-processing

ABSTRACT

In clinical practice, a detailed medical history and physical exam, which includes a thorough neurological examination, are used to diagnose a brain tumor. The size, form, margin, and texture of the tumor, among other things, can influence the diagnosis of brain tumors. Even tumor types that are pathologically different could have a similar texture and appearance in radiology. Furthermore, carefully reviewing all test results could take a significant amount of time for doctors and radiologists. As a result, more advanced medical technology, such as an automated system to diagnose brain tumors, is required. This study aimed to use artificial neural networks to develop an automated approach for detecting brain tumors in magnetic resonance imaging (MRI) scans. Towards this, about 4314 MRI images were acquired in this study. The data contains four classes: normal healthy brain, brain images having glioma, meningioma, or pituitary tumor. The raw data undergoes several preprocessing steps, and the impact of each preprocessing stage on the model accuracy was evaluated. A Densely Connected Convolutional Network (DenseNet) was trained using three different datasets. Enhancing the MRI image's contrast and normalizing its intensities improve the classification accuracy. It shows that preprocessing steps improved the learning convergence of DenseNet training. The proposed model achieved an accuracy of 96.52%, and the sensitivity and specificity were 98.5% and 82.1%, respectively, using ten-fold cross-validation. Hence, we conclude that specific preprocessing steps significantly enhance the tumor segmentation performance for automated systems when using advanced techniques such as deep learning.

1. Introduction

The brain, which operates with billions of cells, is one of the most complex organs in the human body. A brain tumor occurs when an uncontrolled cell division forms an irregular group of cells outside or within the brain. With 1.8% of the overall number of new cancers globally, it is the 22nd most common cancer. Brain tumors are not as widespread as many other cancers. However, the brain cancer death rate is higher than the number of new cases per year (Siegel, Miller, Fuchs, & Jemal, 2021a). The American Cancer Society estimated 24,530 new brain cancer cases in the United States in 2021. Although brain cancer contributes to only 3% of all deaths caused by different cancer types, 75% of those diagnosed with brain or other nervous system tumors are expected to die. Brain cancer is also considered the deadliest among men below 40 and women below 20 (Siegel, Miller, Fuchs, & Jemal, 2021b; Siegel, Miller, & Jemal, 2018).

The World Health Organization classify brain tumor into 120 types based on tumor origin and behavior of its cells. Tumors are also classified into grades, which indicate tumor aggressiveness and rate of spread (Louis et al., 2007). Brain tumors are graded or classified as benign or low-grade tumors (grades I and II) and high-grade (rate III and IV) or malignant tumors. The low-grade brain tumors are non-progressive (non-cancerous) and considered less aggressive. Malignant tumors, though, grow faster and are cancerous. The source of the cancer cells can be the brain itself (primary malignant tumor) or originate from a different part of the body and travel through the bloodstream to the brain (secondary malignant tumor) (Louis et al., 2007).

A medical history and a physical exam, including a thorough

Peer review under responsibility of The Egyptian Society of Radiation Sciences and Applications.
* Corresponding author.
E-mail address: malnowaimi@kau.edu.sa (M. Alnowami).

https://doi.org/10.1016/j.jrras.2022.05.014
Received 7 March 2022; Received in revised form 11 May 2022; Accepted 27 May 2022
Available online 21 June 2022

neurological examination, are used to diagnose a brain tumor. One or more medical tests, such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Angiography, Skull X-rays, or biopsy, may be included in the medical examination. Total annual net treatment expenses are estimated at $150000 for various cancer groups. However, brain cancer is the most expensive per patient for any cancer category. Along with other types of cancers, brain cancer has the highest total net costs of $135–$210 K per patient (depending on age and gender) (Louis et al., 2007). Hence, the best diagnostic tool is MRI, which gains value for money by a $113,800 QALY (quality-adjusted life year) (Siegel et al., 2018).

Moreover, because an accurate diagnosis will lead to better treatment, precise brain tumor diagnosis is challenging. The size, form, margin, and texture of the tumor, among other things, can influence the diagnosis of brain tumors. Tumors may show similar texture and appearance in an image even if they are pathologically different. In addition, it is time-intensive for doctors and radiologists to review all test findings manually. Thus, improved medical technology in the form of an automated system is required to improve patient care.

Herein, we propose an automated method for assisting doctors in diagnosis to avoid misdiagnosis and incorrect treatment. Based on brain tumor grades categorized by WHO standards, a novel deep learning-based system is provided for classification. Unlike traditional approaches requiring prior tumor mass segmentation, our convolutional neural network (CNN) methodology does not require region-based preprocessing procedures. Most traditional approaches train their model using an image of the segmented tumor instead of the whole image for many reasons, such as lowering processing time, minimizing data storage, and others. However, this study used the entire image to train the model rather than a segmented tumor to consider other features in the image that may enhance the model's classification accuracy. Several studies have shown that clinical diagnosis using an artificial intelligence (AI)-based system has significant potential towards aiding in brain tumor detection. Such a system could enable a more accurate and reliable diagnosis, with a better indication for prognosis, a less expensive strategy than other diagnostic tests, and speed up the brain tumor diagnosis process.

2. Related work

Nowadays, one of the most promising applications in the clinical sector is the development of AI for healthcare. One of many potential applications is brain cancer detection. The process of brain tumor detection and identification performed by radiologists is impractical, nonreproducible, and time-consuming (Swati et al., 2019). AI offers a practical alternative that overcomes these issues by utilizing traditional machine learning as well as deep learning approaches. Traditional machine learning techniques typically include preprocessing, feature extraction, feature reduction, and classification. Feature extraction can be classified into low-level (global) features and high-level (local) features. Low-level features are effective in describing an image but limited in their representation power (Selvaraj, Selvi, Selvathi, & Gewali, 2007). The limitation is due to the similarity in texture, boundary, shape, and size of different brain tumors. Examples of low-level features include texture features, first-order statistics (e.g., mean, standard deviation and skewness), and second-order statistics derived from gray-level co-occurrence matrix (GLMC) (Selvaraj et al., 2007). High-level features, such as scale-invariant feature transformation (SIFT) and fisher vector (FV), overlook spatial information (Kang, Ullah, & Gwak, 2021). Traditional machine learning techniques don't incorporate high and low-level features but instead focus on one feature type. The classification accuracy depends on feature extraction, which is susceptible to inter-and intra--user variability (Jiang et al., 2013).

Deep learning techniques eliminate the need to manually extract features, as the algorithm itself learns the features utilized in the classification process (Litjens et al., 2017; Ren, He, Girshick, Zhang, & Sun, 2017). Numerous publications discussed brain tumor detection and classification using deep learning algorithms (Chang P. et al., 2018; Liang et al., 2018; Sachdeva, Kumar, Gupta, Khandelwal, & Ahuja, 2013). Sumitra et al. used neural networks to classify subjects as normal or abnormal based on MRI brain images (Sumitra & Saxena, 2013). Their study employed discrete wavelet transformation for feature extraction and principal component analysis for dimensionality reduction. The classification was performed using two neural network techniques: back-propagation and feed-forward artificial neural networks. Although none of the performance metrics for the two neural models were reported, it was concluded that both models could help detect brain tumors.

Zulpe et al. developed a model that combined the Gray-level co-occurrence matrix (GLCM) with a two-layered feed-forward neural network with Levenberg Marquart (LM) nonlinear optimization algorithm to classify four classes of brain tumors. Out of 80 images (20 per class), 56 images were used for training, 16 for validation, and 8 for testing. The model scored 97.5% accuracy (Zulpe & Pawar, 2012). In addition, a model comprising Regularized Kernel-based Fuzzy C-Means Clustering (ARKFCM) segmentation technique with support vector machine (SVM) for feature extraction and artificial neural network (ANN) back-propagation algorithm for classification was suggested by Bhat et al. for brain tumor detection and classification. The model outperformed other models relying on fuzzy C means for segmentation and SVM or ANN for feature extraction and classification, reporting 98%, 78%, and 91.4% sensitivity, specificity, and accuracy, respectively (Thejaswini, Bhat, & Prakash, 2019).

Seetha et al. proposed convolutional neural networks (CNN) instead of conventional classification techniques, such as SVM and deep neural networks, for higher accuracy and less computational time. Two classes of brain images were used: normal and abnormal. The proposed CNN achieved 97.5% training accuracy, much higher than SVM and slightly higher than deep neural networks. Additionally, since CNN does not require features to be extracted before training, it was less computational time than SVM (Seetha & Raja, 2018). Several other researchers employed CNN for brain MRI classification (Balasooriya & Nawarathna, 2017; Çinar & Yildirim, 2020). Deepak et al. used a pre-trained GoogLeNet to extract features from brain MRI images. The extracted features were used with several proven classification models. The proposed classification system achieved a mean accuracy of 98% and performed well with a smaller number of training data (Deepak & Ameer, 2019).

Korfiatis et al. used Residual Deep Convolutional Neural Network on brain MRI images to predict the O6-methylguanine methyltransferase (MGMT) gene status. By employing MGMT, a brain tumor biomarker, as a predictor, segmentation is no longer needed. Three classifications were used: no tumor, methylated MGMT, and non-methylated. The study reported a 94.90% accuracy with a 50-layer neural network model, outperforming other models considered in the research and concluding that biomarkers can be predicted from diagnostic images by deep neural networks (Korfiatis et al., 2017). Another study that incorporated medical images with convolutional neural networks to obtain prognostic and predictive information was carried out by Li et al. MRI images were used to predict the isocitrate dehydrogenase 1 (IDH1) mutation status in their research. In low-grade glioma, the IDH1 gene is a molecular biomarker that reflects more than half the projected value. They evaluated models based on single and multiple MRI modality images. Multiple modality image-based models scored 92.4% accuracy, 94.3% sensitivity, and 86.6% specificity, higher than all single modality-based models (Li, Wang, Yu, Guo, & Cao, 2017).

3. Materials and methods

3.1. Dataset and preprocessing

More than 4314 brain images were gathered from different open-access repositories in this study (Jun Cheng, 2017; n.dKaggle.). The

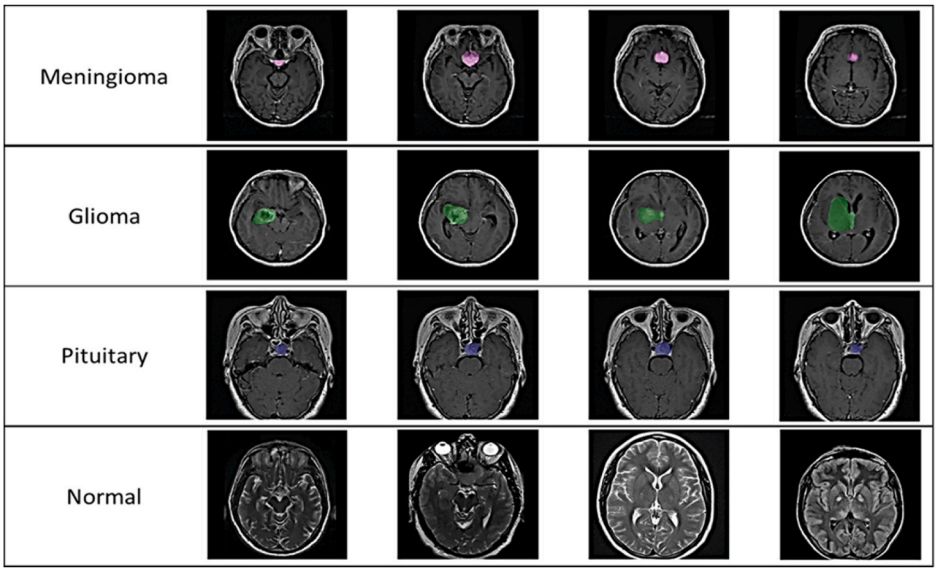

Fig. 1. An example from each class where the highlighted area indicates the tumor location and size. The data contains four brain images classes: Normal healthy brain or brain images with glioma, meningioma, or pituitary tumor.

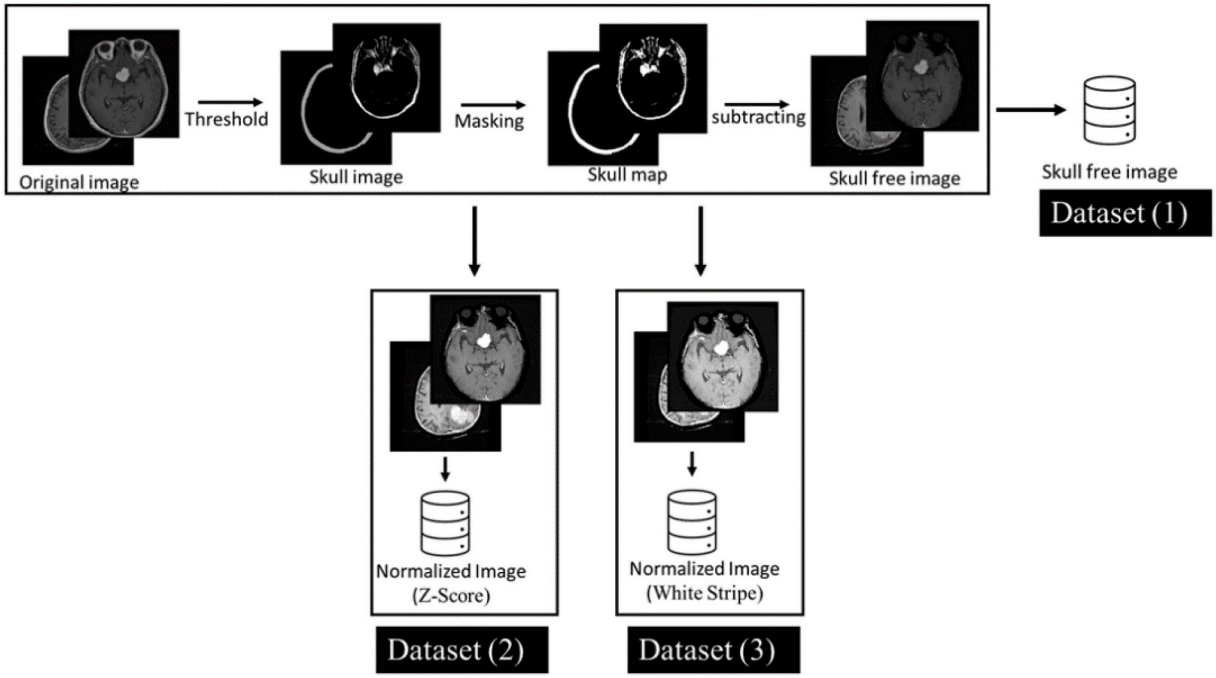

Fig. 2. An illustration of the original (left) and processed image (right).

data was collected at Nanfang Hospital in Guangzhou and General Hospital at Tianjin Medical University, China, between 2005 and 2010. We iterated through each file, automatically extracting and storing the original image and related metadata. All images are two-dimensional and represent 341 patients with multiaxial, coronal, and sagittal views for each patient. The image matrices are 512 × 512 with a pixel size of 0.49 mm. The data contains four classes: Normal healthy or images with glioma, meningioma, or pituitary tumor. Fig. 1 shows an example of each class where the highlighted area indicates the tumor location and size.

The images went through two different stages (Fig. 2). In the first stage, an image augmentation technique was employed to increase the number of images by altering the existing dataset to generate an artificial dataset. The augmentation process was utilized by randomly rotating and flipping the images. The primary purpose of augmenting the original data is to increase the data size and improve the learning ability of the deep learning algorithm. Then, to prepare the images for the machine learning models and train the model faster, all images were resized to 224 × 224 pixels, the size of the neural network's input layer. The non-brain tissue was then removed from the images in the second stage by implementing a skull extraction. The skull stripping process was done by implementing the procedure proposed by Roslan(Roslan, Jamil, & Mahmud, 2010). At this stage, the original dataset was augmented, resized, cleaned, and skull-free. The resulted dataset was then duplicated, and one set was saved separately as the first version, dataset-1, and the second set was passed to the next stage.

Table 1
The number of images used for each class and each partition.

Class	No. of images				
	Original	Augmented	Total	Training	Verifying
Normal	1250	250	1500	1050	450
Glioma	1426	74	1500	1050	450
Meningioma	708	792	1500	1050	450
Pituitary	930	570	1500	1050	450
Total	4314	1686	6000	4200	1800

MRI images generated via different scanners, settings, or times may lead to significant intensity fluctuations, affecting the classification accuracy. Therefore, intensity normalization is an essential preprocessing step. In the second stage, intensity normalization was carried out using two approaches, WhiteStripe and Z-Score (Carré et al., 2020). The Z-Score technique normalizes the image's pixels intensities by subtracting the image mean intensity value (μ_I) from each pixel intensity I (x) and dividing the result by the standard deviation (σ_I). However, in the White-Stripe technique, we subtract the mean intensity value of the normal-appearing white matter (NAWM) (μ_{ws}) from each pixel intensity I(x) and divide the result by the standard deviation of the NAWM (σ_{ws}). Finally, to evaluate the impact of the intensity normalization technique, two more versions of the data were created. Dataset-2 contains all the dataset-1 images after applying the Z-Score normalization technique. However, dataset-3 includes all the dataset-1 images after using the White-Stripe normalization technique. Note, all datasets have the same images but have undergone different processing steps.

3.2. Data partitioning

Supervised machine learning algorithms often involve dividing data into several portions for training, verifying, and testing classifiers. In this study, four distinct datasets were randomly generated. The dataset includes four labeled datasets for the classification model (Normal healthy brain and brain images with glioma, meningioma, or pituitary tumor). Each class has 1500 images, the original dataset plus the augmented dataset. Then, the organized data set was randomly partitioned into 70% and 30% for the training and testing set, respectively. Table 1 summarizes the number of images used for each class and partition.

3.3. Network architecture

In general, any deep learning model consists of two stages, training and testing. The model extracts and examines the features that lead to the best classification accuracy in the training phase. While in the testing phase, the selected features in the training phase are used to examine the model classification accuracy in unseen images (testing data). Although several types of deep learning model architecture for classification exist, they all share the same training and testing stages. The significant difference between them is the depth and the architect of the model used.

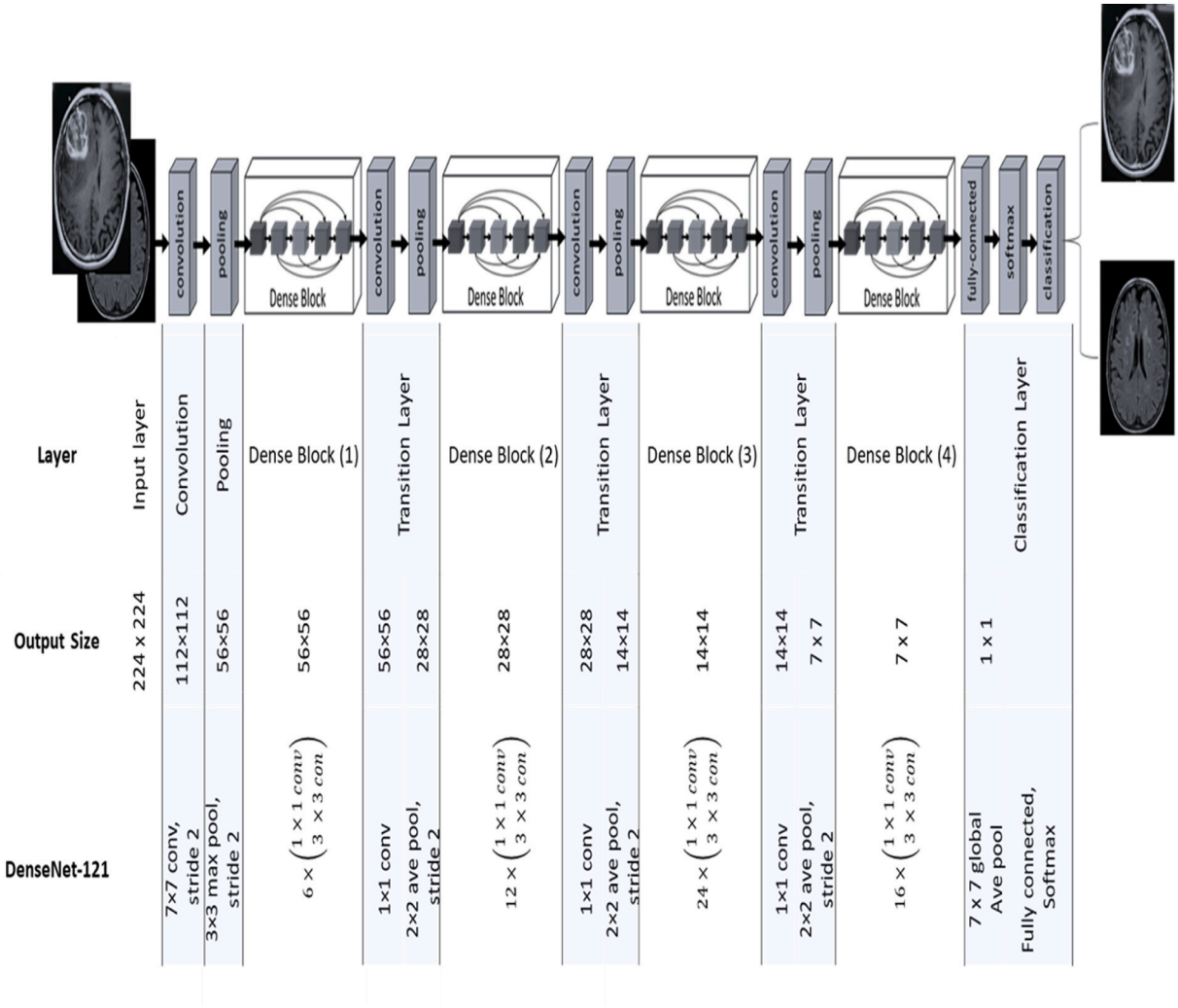

Fig. 3. A diagram of a deep DenseNet and the network architecture.

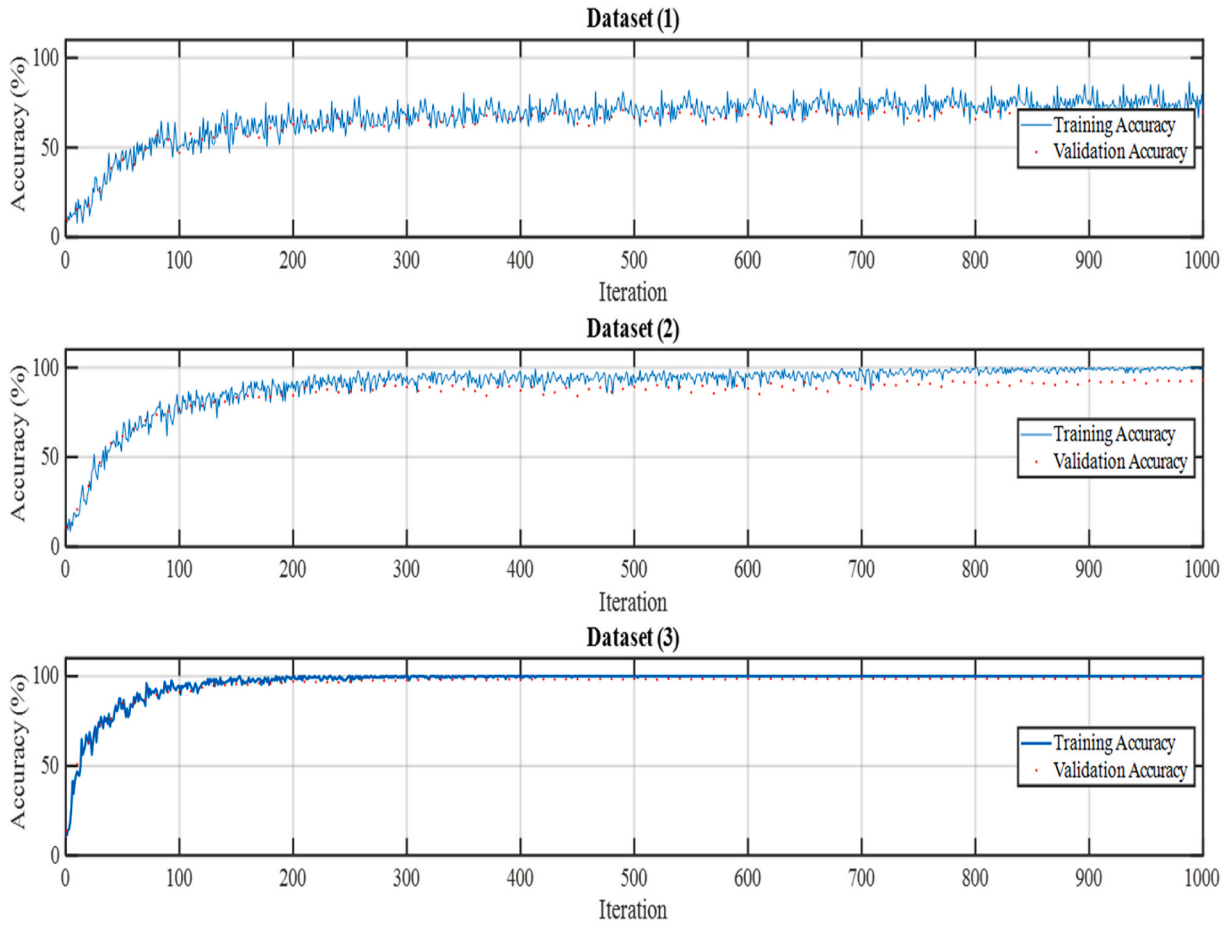

Fig. 4. Training progress for DenseNets, the first dataset was the original data without contrast enhancement and intensity normalization. In the second dataset, only the intensity of all the images was normalized. The third dataset images passed through the whole preprocessing stages.

Recent studies investigate the advancement of using a deeper network (Huang, Liu, Van Der Maaten, & Weinberger, 2017). It shows that a deeper network could learn more complex features of the input data and produce superior results. However, several authors highlighted that using a deeper network comes with a cost degradation error of deeper convolutional networks (Chang K. et al., 2018).

This study employs Densely Connected Convolutional Networks (DenseNets). DenseNets links each layer feature to the next layer in a feed-forward approach to eliminate the risk of the degradation error (Huang et al., 2017). Fig. 3 shows a diagram of a deep DenseNet with four dense blocks. The model has 58 layers and was composed of four dense blocks and the network architecture. Each layer incorporated a convolutional filter, a rectified linear unit (ReLU) activation layer, and a batch normalization (BN) layer, where the transition layers consist of a Convolution and pooling layer.

3.4. Training

During training, the learning rate was carefully regulated using a stochastic descent technique(Huang et al., 2017). An initial training rate of 0.0001 was chosen at the start of training and then decreased by a factor of 0.2 every five epochs with a total of 80 epochs, using mini-batches of 50 observations at each iteration. As the training continued, the training rate changed, allowing for a smaller and more precise search for the best value.

3.5. Evaluation metrics

All labeled images were randomly divided into two subsets: 70% training, 30% validation, and testing. 10-fold cross-validation was applied to evaluate the model performance, the data was split into ten sets, and each set contained the same number of images per class. Randomly, nine sets were used for training, and the remaining set was used for testing. This process loops ten times, with a different subset of the data being evaluated each time. The following are the performance metrics for the proposed work:

- Accuracy: The proportion of true results among the total number of cases studied. Accuracy= $(TN+TP)/(TN+FN+TP+FP)$
- Sensitivity (Se) also be referred to as the recall, hit rate, or TPR (True Positive Rate): The percentage of negatives that are mistakenly recognized as positives: Se = TP/(TP+FN)
- Specificity (Sp) also be referred to as True Negative Rate (TNR): The percentage of true negatives that are appropriately identified as such: Sp = TN/(TN+FP)

Where True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN).

4. Results

As mentioned in the preprocessing data section, three different datasets were used to examine the effect of the preprocessing stage on the overall accuracy of the learning model. A 10-fold cross-validation was applied to evaluate the model performance. Fig. 4 shows the training progress for DenseNets. It presents the training accuracy of each individual mini-batch. The result reflected an accuracy gain in the model performance when the intensity normalization step is applied, dataset-2

Table 2
Sensitivities and specificities for each class in the dataset-3 on ten-fold cross-validation. The table shows the result for each fold and the average across all tests.

cross-validation	Classification accuracy										Average
	cv 1	cv 2	cv 3	cv 4	cv 5	cv 6	cv 7	cv 8	cv 9	cv 10	
Dataset (1)	66.79	73.05	65.93	71.07	81.93	75.88	59.19	74.94	84.13	68.09	**72.10**
Dataset (2)	87.01	91.39	78.84	88.48	94.02	82.03	79.51	93.21	91.79	83.91	**87.02**
Dataset (3)	95.29	96.64	97.63	94.06	98.72	95.96	96.62	97.75	97.93	94.57	**96.52**

Table 3
Sensitivity and specificity for each class in the dataset-3 on ten-fold cross-validation. The table shows the result for each fold and the average across all tests.

Evaluation Metrics	Class	cross-validation result										Average
		cv 1	cv 2	cv 3	cv 4	cv 5	cv 6	cv 7	cv 8	cv 9	cv 10	
Sensitivity	**Normal**	97.65	94.57	96.13	94.26	94.26	97.53	98.23	98.05	98.59	97.71	**96.70**
	Glioma	96.66	94.34	93.26	94.21	93.19	96.16	97.67	95.45	96.66	97.84	**95.54**
	Meningioma	92.76	91.19	94.02	89.05	91.69	95.28	95.90	88.78	96.26	94.03	**92.90**
	Pituitary	96.48	95.82	89.89	92.51	86.48	90.68	90.10	83.32	88.27	88.68	**90.22**
Specificity	**Normal**	92.57	88.52	94.84	95.82	95.78	90.64	93.59	95.67	94.25	95.80	**93.75**
	Glioma	96.53	94.01	95.41	89.59	89.62	97.14	94.81	91.36	95.05	99.95	**94.35**
	Meningioma	98.81	97.88	87.30	93.15	91.92	91.58	96.78	85.73	95.20	88.76	**92.71**
	Pituitary	94.46	87.21	94.91	94.85	92.66	96.97	95.37	90.64	90.37	93.15	**93.06**

Table 4
Comparison of our proposed work and existing Classifiers in terms of performance metrics.

Author	Year	Type of DL	Data size	Evaluation metrics
Our Proposed	2021	Neural Network (NN)	720	Sensitivity (98.5%) Specificity (82.1%) Accuracy (96.52%)
P Thejaswini et al. (Thejaswini, Bhat, & Prakash, 2019)	2020	Artificial Neural Network (ANN)	94	Sensitivity (98%) Specificity (78%) Accuracy (91.4%)
Ken Chang et al. (Chang et al., 2018a)	2019	Residual CNN (ResNet34)	406	Accuracy 85.7%
Peter Chang et al. (Chang et al., 2018b)	2019	CNN	256	Accuracy 83%
Sen Liang et al. (Liang et al., 2018)	2018	Multimodal 3D Dense Net Multimodal 3D Dense Net with Transfer learning	167	Accuracy 84.60% Accuracy 91.40%
L. Han et al. (Han & Kamdar, 2018)	2018	Convolutional Recurrent Neural Network (CRNN)	5235	Accuracy 0.62
Chenjie Ge et al. (Ge, Gu, Jakola, & Yang, 2018)	2018	2D-CNN	285 159	Accuracy 90.87% Accuracy 89.39%
Zeju Li et al. (Li et al., 2017)	2017	CNN	151	Accuracy 92%
P. Korfiatis et al. (Korfiatis et al., 2017)	2017	ResNet50 ResNet36 ResNet18	155	Accuracy 94.90% Accuracy 80.72% Accuracy 76.75%

and dataset-3. The result agreed with the initial assumption about the importance of the MRI image's intensity normalization in the classification accuracy. Also, the white-strip technique converges faster than the Z-Score normalization technique to the highest solution. It shows that using dataset-3 to train the model improved the learning convergence.

After the training stage, the trained model was used to classify the test data. Table 2 presents the ten-fold cross-validation and the average accuracy for each dataset. The results show that dataset-3 enhanced the model performance and significantly improved classification accuracy. The model showed the highest accuracy, with an average accuracy of 96.52%.

Table 3 presents sensitivities and specificities for each class in dataset-3 on ten-fold cross-validation. The table demonstrates the result for each fold and the average across all tests. The proposed model performance was compared against the latest publications. However, this comparison may not be genuinely valid as different datasets were used. The primary purpose is to highlight the overall view of the results obtained in this study and other studies in this area. Table (4) shows a list of DL models used to classify brain tumors based on the MRI images. The table stated the author, the year of publication, the model used, the data size used to train the model, and the evaluation metric result. Some of the publications include the Accuracy, Sensitivity, and Specificity of the classification model. Sensitivity describes the model's ability to predict true positives in each given category is sensitivity. It presents a performance comparison of proposed and current classifiers, demonstrating that the proposed work outperforms all current classifiers. The proposed model achieved the highest accuracy, with an approximate accuracy of 96.52% and the sensitivity and specificity are 98.5% and 82.1%, respectively (Table 4).

5. Discussion and conclusions

This study aimed to build an automated method for classifying MRI images of brain tumors using artificial neural networks and examine the effect of the preprocessing stages on the overall accuracy of the learning model. An algorithm was successfully developed and tested using publicly available brain tumor MRI images. About 4314 MRI images were gathered for this investigation. The data included images of several forms of brain tumors, four classes: normal healthy or images with glioma, meningioma, or pituitary tumor, and three orthogonal planes (sagittal, axial, and coronal). The images were prepared in the preprocessing stage to train the convolutional neural network. The preprocessing stages included augmentation, resizing, skull stripping, and normalization. Two different approaches were implemented to normalize the MRI image's intensities, the Z-Score normalization technique and the white-strip technique. In addition, three different datasets

were generated to examine the effect of the normalization step and technique.

The first step was to investigate whether normalizing the image's intensity would affect the learning model. Then the paper examines the influence on the accuracy of the learning model when a different type of normalization is utilized. Ten-fold cross-validation was applied in all tests. This study used Densely Connected Convolutional Networks (DenseNets). The model has 58 layers and is composed of four dense blocks. The study has shown that: 1) classification model performance is substantially improved when performing an intensity normalization preprocessing step; 2) classification accuracy is affected by the normalization method used. There was a significant difference in the classification model accuracy when the white-strip technique was used. Since each normalization method uses a diffident approach to normalize the image pixel's intensity, these normalization approaches enable a different set of activation features in the deep learning model and eventually affect the training accuracy. The white-strip normalization method enhanced the white matter contrast, and that improved the classification model performance.

Although utilizing the white-strip normalization method positively impacted the model accuracy, it requires more processing time and is more challenging to perform. Moreover, the normalization quality relies on the white matter segmentation as the white-strip technique normalizes images based on the intensity values of the normal-appearing white matter (NAWM).

Acknowledgment

The project was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah under grants no Project no: MG/34/17. The authors, therefore, acknowledge with thanks DSRtechnical and financial support.

References

Balasooriya, N. M., & Nawarathna, R. D. (2017). A sophisticated convolutional neural network model for brain tumor classification. In *2017 IEEE International Conference on Industrial and Information Systems (ICIIS)* (pp. 1–5). IEEE. https://doi.org/10.1109/ICIINFS.2017.8300364.

Carré, A., Klausner, G., Edjlali, M., Lerousseau, M., Briend-Diop, J., Sun, R., et al. (2020). Standardization of brain MR images across machines and protocols: Bridging the gap for MRI-based radiomics. *Scientific Reports, 10*(1), Article 12340. https://doi.org/10.1038/s41598-020-69298-z

Chang, K., Bai, H. X., Zhou, H., Su, C., Bi, W. L., Agbodza, E., et al. (2018a). Residual convolutional neural network for the determination of IDH status in low- and high-grade gliomas from MR imaging. *Clinical Cancer Research, 24*(5), 1073–1081. https://doi.org/10.1158/1078-0432.CCR-17-2236

Chang, P., Grinband, J., Weinberg, B. D., Bardis, M., Khy, M., Cadena, G., et al. (2018b). Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas. *American Journal of Neuroradiology, 39*(7), 1201–1207. https://doi.org/10.3174/ajnr.A5667

Cheng, J. (2017). *Brain tumor dataset.* https://doi.org/10.6084/m9.figshare.1512427.v5

Çinar, A., & Yildirim, M. (2020). Detection of tumors on brain MRI images using the hybrid convolutional neural network architecture. *Medical Hypotheses, 139*, Article 109684. https://doi.org/10.1016/j.mehy.2020.109684

Deepak, S., & Ameer, P. M. (2019). Brain tumor classification using deep CNN features via transfer learning. *Computers in Biology and Medicine, 111*, Article 103345. https://doi.org/10.1016/j.compbiomed.2019.103345

Ge, C., Gu, I. Y.-H., Jakola, A. S., & Yang, J. (2018). Deep learning and multi-sensor fusion for glioma classification using multistream 2D convolutional networks. In *2018 40th annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 5894–5897). IEEE. https://doi.org/10.1109/EMBC.2018.8513556.

Han, L., & Kamdar, M. R. (2018). MRI to MGMT: Predicting methylation status in glioblastoma patients using convolutional recurrent neural networks. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing, 23*, 331–342. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/29218894.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2261–2269). IEEE. https://doi.org/10.1109/CVPR.2017.243.

Jiang, J., Wu, Y., Huang, M., Yang, W., Chen, W., & Feng, Q. (2013). 3D brain tumor segmentation in multimodal MR images based on learning population- and patient-specific feature sets. *Computerized Medical Imaging and Graphics, 37*(7–8), 512–521. https://doi.org/10.1016/j.compmedimag.2013.05.007

Kaggle. (n.d.). Brain Tumor MRI Dataset. Retrieved December 23, 2019, from https://www.kaggle.com/sartajbhuvaji/brain-tumor-classification-mri?select=Training.

Kang, J., Ullah, Z., & Gwak, J. (2021). MRI-based brain tumor classification using ensemble of deep features and machine learning classifiers. *Sensors, 21*(6), 2222. https://doi.org/10.3390/s21062222

Korfiatis, P., Kline, T. L., Lachance, D. H., Parney, I. F., Buckner, J. C., & Erickson, B. J. (2017). Residual deep convolutional neural network predicts MGMT methylation status. *Journal of Digital Imaging, 30*(5), 622–628. https://doi.org/10.1007/s10278-017-0009-z

Liang, S., Zhang, R., Liang, D., Song, T., Ai, T., Xia, C., et al. (2018). Multimodal 3D DenseNet for IDH genotype prediction in gliomas. *Genes, 9*(8), 382. https://doi.org/10.3390/genes9080382

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., et al. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis, 42*, 60–88. https://doi.org/10.1016/j.media.2017.07.005

Li, Z., Wang, Y., Yu, J., Guo, Y., & Cao, W. (2017). Deep Learning based Radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma. *Scientific Reports, 7*(1), 5467. https://doi.org/10.1038/s41598-017-05848-2

Louis, D. N., Ohgaki, H., Wiestler, O. D., Cavenee, W. K., Burger, P. C., Jouvet, A., et al. (2007). The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathologica, 114*(2), 97–109. https://doi.org/10.1007/s00401-007-0243-4

Ren, S., He, K., Girshick, R., Zhang, X., & Sun, J. (2017). Object detection networks on convolutional feature maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(7), 1476–1481. https://doi.org/10.1109/TPAMI.2016.2601099

Roslan, R., Jamil, N., & Mahmud, R. (2010). Skull stripping of MRI brain images using mathematical morphology. In *2010 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)* (pp. 26–31). IEEE. https://doi.org/10.1109/IECBES.2010.5742193.

Sachdeva, J., Kumar, V., Gupta, I., Khandelwal, N., & Ahuja, C. K. (2013). Segmentation, feature extraction, and multiclass brain tumor classification. *Journal of Digital Imaging, 26*(6), 1141–1150. https://doi.org/10.1007/s10278-013-9600-0

Seetha, J., & Raja, S. S. (2018). Brain tumor classification using convolutional neural networks. *Biomedical and Pharmacology Journal, 11*(3), 1457–1461. https://doi.org/10.13005/bpj/1511

Selvaraj, H., Selvi, S. T., Selvathi, D., & Gewali, L. (2007). Brain MRI slices classification using least squares support vector machine. *International Journal of Intelligent Computing in Medical Sciences & Image Processing, 1*(1), 21–33. https://doi.org/10.1080/1931308X.2007.10644134

Siegel, R. L., Miller, K. D., Fuchs, H. E., & Jemal, A. (2021a). Cancer statistics, 2021. *CA: A Cancer Journal for Clinicians, 71*(1), 7–33. https://doi.org/10.3322/caac.21654

Siegel, R. L., Miller, K. D., Fuchs, H. E., & Jemal, A. (2021b). Cancer statistics, 2021. *CA: A Cancer Journal for Clinicians, 71*(1), 7–33. https://doi.org/10.3322/caac.21654

Siegel, R. L., Miller, K. D., & Jemal, A. (2018). Cancer statistics, 2018. *CA: A Cancer Journal for Clinicians, 68*(1), 7–30. https://doi.org/10.3322/caac.21442

Sumitra, N., & Saxena, R. (2013). Brain tumor classification using back propagation neural network. *International Journal of Image, Graphics and Signal Processing, 5*, 45–50.

Swati, Z. N. K., Zhao, Q., Kabir, M., Ali, F., Ali, Z., Ahmed, S., et al. (2019). Brain tumor classification for MR images using transfer learning and fine-tuning. *Computerized Medical Imaging and Graphics, 75*, 34–46. https://doi.org/10.1016/j.compmedimag.2019.05.001

Thejaswini, P., Bhat, B., & Prakash, K. (2019). Detection and classification of tumour in brain MRI. *International Journal of Engineering and Manufacturing, 9*(1), 11–20. https://doi.org/10.5815/ijem.2019.01.02

Zulpe, N., & Pawar, V. (2012). GLCM textural features for brain tumor classification. *International Journal of Computer Science Issues (IJCSI), 9*(3), 354.