International Journal of Computer Assisted Radiology and Surgery
https://doi.org/10.1007/s11548-021-02326-z

REVIEW ARTICLE

Deep learning approaches for automated classification and segmentation of head and neck cancers and brain tumors in magnetic resonance images: a meta-analysis study

Samireh Badrigilan[1] · Shahabedin Nabavi[2] · Ahmad Ali Abin[2] · Nima Rostampour[1] · Iraj Abedi[3] · Atefeh Shirvani[3] · Mohsen Ebrahimi Moghaddam[2]

Received: 9 September 2020 / Accepted: 16 February 2021

Abstract

Purpose Deep learning (DL) has led to widespread changes in automated segmentation and classification for medical purposes. This study is an attempt to use statistical methods to analyze studies related to segmentation and classification of head and neck cancers (HNCs) and brain tumors in MRI images.

Methods PubMed, Web of Science, Embase, and Scopus were searched to retrieve related studies published from January 2016 to January 2020. Studies that evaluated the performance of DL-based models in the segmentation, and/or classification and/or grading of HNCs and/or brain tumors were included. Selected studies for each analysis were statistically evaluated based on the diagnostic performance metrics.

Results The search results retrieved 1,664 related studies, of which 30 studies were eligible for meta-analysis. The overall performance of DL models for the complete tumor in terms of the pooled Dice score, sensitivity, and specificity was 0.8965 (95% confidence interval (95% CI): 0.76–0.9994), 0.9132 (95% CI: 0.71–0.994) and 0.9164 (95% CI: 0.78–1.00), respectively. The DL methods achieved the highest performance for classifying three types of glioma, meningioma, and pituitary tumors with overall accuracies of 96.01%, 99.73%, and 96.58%, respectively. Stratification of glioma tumors by high and low grading revealed overall accuracies of 94.32% and 94.23% for the DL methods, respectively.

Conclusion Based on the obtained results, we can acknowledge the significant ability of DL methods in the mentioned applications. Poor reporting in these studies challenges the analysis process, so it is recommended that future studies report comprehensive results based on different metrics.

Keywords Classification · Deep learning · Head & neck tumors · Magnetic resonance imaging · Meta-analysis · Segmentation

Introduction

Worldwide, with more than 600,000 new cases annually, head and neck cancer (HNC) is the sixth most common cancer [1, 2]. It accounts for 300,000 deaths in the world each year [3]. HNCs defined as tumors that develop in the head and neck region which includes tumors of salivary gland, pharynx, larynx, oral cavity, nasal cavity, paranasal sinus, and the thyroid. [4, 5]. Regardless of different histopathology of HNCs, almost 85% of them are squamous cell carcinomas [5, 6]. Currently, HNCs are treated with radiation therapy, chemotherapy, surgery, or a combination of them. It is a complicated process due to large target volume and many radiation-sensitive critical tissues near the target [7, 8]. A dismal 40–50% survival rate, due to development of

✉ Nima Rostampour
nima.rostampour@kums.ac.ir

1 Department of Medical Physics, School of Medicine, Kermanshah University of Medical Sciences, Kermanshah, Iran

2 Faculty of Computer Science and Engineering, Shahid Beheshti University, Tehran, Iran

3 Department of Medical Physics, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran

Published online: 05 March 2021

a second primary cancer or distant metastasis, means that treatment failure occurs in almost half of the patients regardless of rigorous combined treatment [9]; however, the locoregional control of most HNCs is acceptable [10–12].

Brain tumor is one of the other types of neoplasms which can also develop in the head and neck region, but its diagnosis and treatment process are different from HNCs. It accounts for 1.6% of all malignances and 2.5% of all cancer-related deaths [2]. Based on the origin, brain tumors are divided into two groups, primary and metastatic. Gliomas are the most common type of primary brain tumors in adults which based on the clinical criteria and histopathology are graded as I–IV. Grade I (pilocytic astrocytoma) is benign tumors with nearly normal cells in histopathological examinations. Grade II (low grade glioma (LGG)) includes oligodendroglioma, astrocytoma, and mixed oligoastrocytoma. Some LGGs eventually evolve to high grade glioma (HGG) which include grade III and IV. Grade IV (glioblastoma multiform (GBM)) is considered as the most severe type with highest spreading rate [13].

MRI is an effective imaging technique in the diagnosis process of HNCs and brain tumors due to employing non-ionizing radiation and its higher soft-tissue resolution in addition to employing of contrast-enhanced agents as well as acquiring different images by using various imaging parameters [14, 15]. It is the commonly used imaging modality to evaluate HNCs and brain tumors as contrasted to other imaging procedures including positron emission tomography (PET) and computerized tomography (CT) [16–18], and also it can obtain three-dimensional images which provides detailed information about the target [19].

Image segmentation is a task in which a part of a medical image that is of interest to the medical team is automatically separated from the rest of the image contents [20]. Although the utility of manual segmentation is prevalent in the clinical routine currently, more time and clinical practice is required by the radiologists. In order to detect HNCs and brain tumors, several deep learning (DL) models such as U-Net and DeepMedic were introduced [21, 22]. DL is a subset of machine learning and recently exhibited a considerable efficiency as a widespread technique, especially in classification and segmentation drawbacks. Due to some performance limitation, convolutional neural network (CNN) models were rapidly developed in recent years. CNN is a class of DL that requires minimal preprocessing and generally used in evaluating of visual imagery [23]. Feature learning and providing unlimited accuracy are major superiority of CNNs to classify various types and grades of brain tumors compared with conventional machine learning [15].

Medical image classification refers to a concept in which images are categorized into different classes based on, for example, the type of lesion observed in images using a supervised learning method. Once the training process is performed by a set of images, the classifier can be used in subsequent machine-based medical diagnoses [24]. Classification of brain tumors has been carried out using different imaging techniques and machine learning approaches in recent years [25–28]. CNN uses convolution operators in most layers of the networks instead of matrix multiplication and consequently contributes to priority of convolutional networks in resolving drawbacks with great computational values. This capability is considerable due to thousands of images with different types and qualities included in MRI-dataset as well as automatic feature extraction is the other advantage of this method compared with shallow machine learning methods.

Automatic determination of the type of lesion or its related area using the images is essential because it allows medical diagnoses to be made even in the absence of experts and in the shortest possible time. In this work, we aimed to evaluate the performance of the DL models for the segmentation and classification of MRI data of patients with HNCs and brain tumors. Then, we evaluated the performance of the models in both classification and segmentation of HNCs and brain tumors by comparing automated measurements to manual measurements derived from experts. Therefore, the contribution of this study is twofold:

- Analysis of studies related to DL-based MR images segmentation associated with HNCs and brain tumors.
- Analysis of the results of studies related to the DL-based classification and grading of brain tumors in MR images.

The rest of this paper is organized as follows. "Methods and data to review" section describes the search and selection process of the eligible studies and statistical approaches to calculate the performance metrics. The results of this meta-analysis are given in "Search strategy" section. Finally, "Selection criteria and data extraction" section and "Statistical analysis" section are devoted to the discussion and conclusion.

Methods and data to review

Search strategy

After 2016, massive research efforts have been devoted to develop DL algorithms for automated segmentation of tumors of head and neck region from MR images, continuing to date. To identify potentially relevant articles, a title/abstract/keyword search was performed in PubMed, Web of Science, Embase, and Scopus databases until January 2020. Our search strategy employed specific search tips of each database using the search query: ("deep learning" OR "hierarchical neural network" OR "convolutional neural network" OR "deep" OR "learning") AND ("segmentation" or "automated segmentation" OR "classification" OR "automated classification") AND ("head and neck neoplasms" OR " head-and-neck" OR " glioma " OR "brain tumor" OR "brain") AND (MRI OR "magnetic resonance imaging"). In addition, a manually evaluation was done to collect all potentially eligible articles. We did not impose any search limitation based on date of publications.

Selection criteria and data extraction

English original articles which met the following criteria were included in the meta-analysis: a) develop a DL algorithm for the following: 1) segmentation of MRI data from patients with HNC and/or brain tumor; 2) classification and grading of MRI data from patients with HNC and/or brain tumor, b) sufficient data regarding the performance of developed algorithms should be available or could be calculated from the raw data (e.g., accuracy, sensitivity, and specificity). The exclusion criteria were as follows: a) reviews, conference abstracts, book chapters, meta-analysis, editorials, duplicate publication, b) lacking sufficient data regarding diagnostic estimates such as true positives (TP), false negatives (FN), true negatives (TN), false positives (FP), sensitivity, and specificity, c) inability to obtain the full text.

The titles and abstracts of the retrieved studies screened independently for valid articles by two reviewers (S. BG., S. N.). In the next step, the following data were extracted from each eligible study: author's name, publication year, type of the cancer, type of DL-based model, TP, FP, FN, and TN. In studies, four labels manually were used to perform tumor segmentation as follows; label 1: necrosis, label 2: edema, label 3: non-enhancing tumor, and label 4: enhancing tumor. The ground truth segmentation was done by experienced neuroradiologists. Commonly, the tumor structure is divided into three regions for clinical applications, as core tumor (necrosis + non-enhancing tumor + enhancing tumor), enhancing tumor, and complete tumor (necrosis + non-enhancing tumor + enhancing tumor + edema).

To evaluate performance of the proposed DL-based methods in automated segmentation of HNC and brain tumor, three publicly available metrics: The Dice score, the specificity, and the sensitivity were extracted for each tumor region.

How to calculate the dice coefficient, the specificity and the sensitivity is given in Eqs. 1 to 3.

$$\text{Dice} = \frac{2\text{TP}}{\text{FP} + 2\text{TP} + \text{FN}} \tag{1}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{FN}}, \tag{3}$$

where TP, TN, FP, and FN show true positive, true negative, false positive and false negative, respectively.

Other widely adopted measures, if calculated, Jaccard similarity coefficient (JSC), Average symmetric surface distance (ASSD), percent match (PM), correspondence ratio (CR), and 95th percentile Hausdorff distance (HD95) were also extracted. Equations 4 to 8 show how these metrics are calculated.

$$\text{JSC} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{4}$$

$$\text{ASSD} = \frac{\text{TN}}{\text{FP} + \text{FN}} \tag{5}$$

$$\text{PM} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{6}$$

$$\text{CR} = \frac{\text{TP} - 0/5 \times \text{FP}}{\text{TP} + \text{FN}} \tag{7}$$

$$\text{HD}(GS) = \max\left\{ \sup_{g \in G} \inf_{s \in S} d(gs) \sup_{s \in S} \inf_{g \in G} d(gs) \right\}, \tag{8}$$

where G and S are related to ground truth and segmented regions, and g and s indicate the points on G and S, respectively.

Apart from these, the effectiveness of tumor grade (HGG, LGG) and image view (axial, sagittal, coronal, fusing) on the segmentation performance was evaluated. The performance of classification methods was evaluated using accuracy, sensitivity, and specificity parameters. Accuracy is also calculated using Eq. 9.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{9}$$

Statistical analysis

Stata software (version 14.2; Stata Corporation, College Station, TX) was used to perform meta-analysis. The segmentation performance measures for each study calculated as pooled Dice, sensitivity, and specificity. Furthermore, positive likelihood ratio (PLR), negative likelihood ratio (NLR), and corresponding 95% confidence intervals (CIs) were calculated for classification of grade and type of brain tumors. The analysis was based on the summary receiver operating characteristic (SROC) curves, and the area under the curve (AUC) is calculated. Fagan's nomogram was applied to interpret clinical utility if DL algorithms diagnosing grade and type of brain tumors.

Results

We identified 1664 relevant articles, of which 1459 were screened after deleting 205 duplicated articles. Of them, 1350 articles excluded by the screening of the little and abstract. Then, we assessed the full-text of the remaining 109 articles for eligibility, of which 30 articles were selected for this meta-analysis [29–57]. The procedure of the literature retrieval and inclusion is presented in Fig. 1. The majority of the studies consisted of patients with brain tumors; however, four studies included nasopharyngeal cancer (NPC) [51, 58–60]. The most brain tumors consisted of patients with LGG and HGG; however, two included LGG [29, 56] and one included HGG [30] only. Most of the studies used available datasets including BRATS, REMBRANDT, TCGA-GBM, and TCGA-LGG; and ten studies relied on local datasets. In total, two studies used BRATS 2012 dataset; eight studies used BRATS 2013 dataset; two studies used BRATS 2014 dataset; 16 studies used BRATS 2015 dataset; five studies used BRATS 2016 dataset; six studies used BRATS 2017 dataset; and eight studies used BRATS 2018 dataset. Four studies used other datasets; one studies used REMBRANDT dataset; one studies used TCGA-GBM,

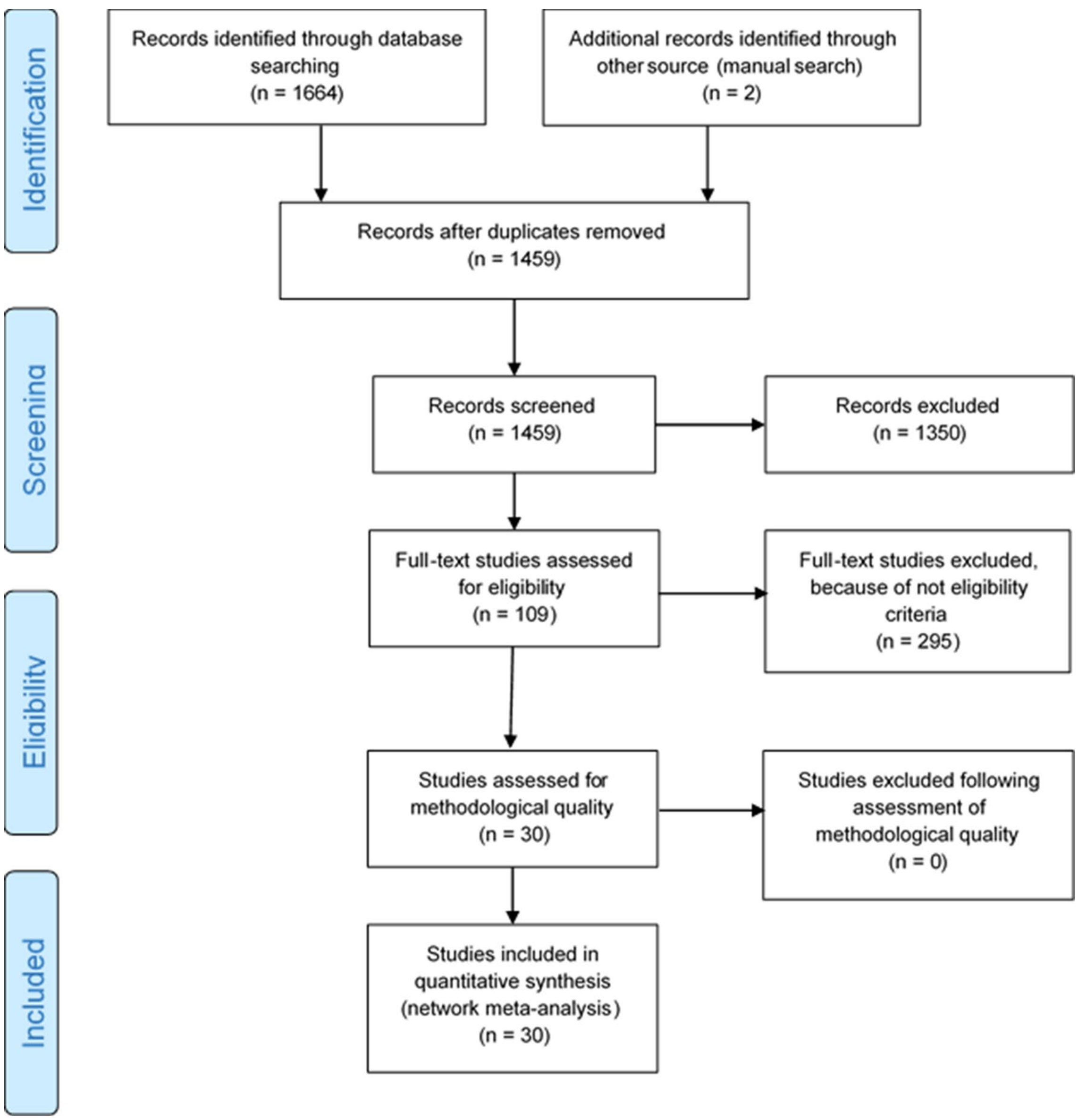

Fig. 1 PRISMA flowchart of the identification of eligible studies

TCGA-LGG, and Figshare datasets. Main characteristics of the included studies are presented in Table 1.

Meta-analysis results

A statistical analysis was conducted using the above data to analyze the performance of DL methods for head and neck tumor segmentation and classification. Seventeen studies provided sufficient information to evaluate the segmentation performance of head and neck tumor subregions using DL models compared with ground truth. Within this group, the performance of all DL model in terms of the pooled Dice score is 0.8965 (confidence intervals (95% CI:) 0.76–0.9994) for complete tumor, 0.7822 (95% CI: 0.56–0.952) for core tumor, and 0.7711 (95% CI: 0.561–0.948) for enhanced tumor (Fig. 2a), the pooled sensitivity is 0.9132 (95% CI: 0.71–0.994) for complete tumor, 0.8103 (95% CI: 0.63–0.92) for core tumor, and 0.79 (95% CI: 0.65–0.9) for enhanced tumor (Fig. 2b), and the pooled specificity is 0.9164 (95% CI: 0.78–1.00) for complete tumor, 0.8314 (95% CI: 0.61–0.93) for core tumor and 0.8105 (95% CI: 0.59–0.92) for enhanced tumor (Fig. 2c).

In addition, the evaluation of the segmentation performance of DL methods in four studies with NPC and one study with LGG and HGG also considered by JSC, ASSD, PM, and CR metrics [40, 51, 58, 60, 61]. In terms of JSC, the DL methods gained an average value of 0.7969 for complete tumor region based on 220 cases. The average PM using 89 patients was 0.879 for complete tumor region. For the CR, the DL methods achieved an average value of 0.804 for complete tumor region based on 130 cases. DL methods achieved an average ASSD of 1.168 and an average HD95 of 6.179 for complete tumor region.

Of the 18 studies, the segmentation performance of DL methods in four studies evaluated by subgrouping grade of glioma tumors. The pooled average Dice score of these four studies was 0.8395 and 0.8794 for LGG and HGG subgroups, respectively. From the pooled two studies, the segmentation sensitivity and specificity were 0.833 and 0.90 for LGG and were 0.8269 and 0.9643 for HGG subgroups, respectively. One study reported 3.4419 and 3.4805 HD95 for LGG and HGG subgroups, respectively.

Of the 18 studies, two studies used four models to segment tumor images slice by slice in axial, coronal, and sagittal views, and results of three different views were fused. For complete tumor, core tumor and enhanced tumor, results of analysis related to dice coefficient, sensitivity, and specificity in different views are tabulated in Table 2. Also, Fig. 3 shows the bar chart of the results. In one another study, the axial, coronal and sagittal view, and the multi-view DL method achieved an average ASSD of 1.361, 1.858, 1.762, and 1.203 for complete tumor, respectively. For the PM and CR, an average value of 0.84 and 0.75 in axial view, 0.745 and 0.64 in coronal view, 0.8 and 0.78 in sagittal, 0.8593 and 0.77 in the multi-view DL method reported for complete tumor, respectively.

The DL methods achieved the best performance for classifying three types of glioma, meningioma, and pituitary tumors with overall accuracies of 96.01%, 99.73%, and 96.58%, respectively. In Fig. 4a, the confusion matrix is showed to check the performance. This matrix gives valuable information about the actual and predicted labels provided by the DL classification methods. Using this information, the sensitivity and specificity of the DL methods for classifying MR brain tumor images were 97.0% and 98.0% for glioma tumors, 94.0% and 98.0% for meningioma tumors, and 96.0% and 98.0% for pituitary tumors, respectively.

For glioma tumors, the AUC for SROC curve was 0.99 (95% CI: 0.98–100) (Fig. 5a). Fagan's nomogram analysis showed that with a pretest probability of 25%, DL as an automated diagnostic method can increase posttest probability of a positive result to 93% while the posttest probability of a negative test was only 1% (Fig. 5b). When setting the pretest probability to 50%, DL increased the probability of a correct detection to 98% while there is only 3% probability of ignoring a glioma patient with a negative result (Fig. 5c). Based on 75% pretest probability, DL increased the probability of a correct detection to 99%, while there is a posttest probability of 10% for ignoring a glioma patient with a negative result (Fig. 5d). The SROC curves and Fagan's nomograms for meningioma and pituitary tumors are shown in Supplementary Material: Fig. S1-2.

Stratification of glioma tumors by high and low grading revealed overall accuracies of 94.32% and 94.23% for the DL methods, respectively (Fig. 4b). In addition, overall estimate achieved a sensitivity and specificity of 97.0% and 95.0% in classifying high-grade glioma respectively, and 95.0% and 98.0% in classifying low-grade glioma, respectively. For high-grade glioma tumors, the AUC for SROC curve was 0.98 (95% CI: 0.96–99) (Fig. 6a). Fagan's nomogram analysis showed that with a pretest probability of 25%, DL as an automated diagnostic method can increase posttest probability of a positive result to 86% while the posttest probability of a negative test was only 1% (Fig. 6b). When setting the pretest probability to 50%, DL increased the probability of a correct detection to 95% while there is only 3% probability of ignoring a high-grade glioma patient with a negative result (Fig. 6c). Based on 75% pretest probability, DL increased the probability of a correct detection to 98%, while there is a posttest probability of 10% for ignoring a high-grade glioma patient with a negative result (Fig. 6d). The SROC curves and Fagan's nomograms for low-grade glioma tumors as shown in Supplementary Material: Fig. S3.

When stratifying by I, II, III, IV grades (Fig. 4c), glioblastoma multiform tumors that are the most common and most malignant brain tumors were classified with

Table 1 Main characteristics of the included studies

Study ID Author, year	Task	Method	Dataset	Remarks
Alex 2017 [29]	Low-grade glioma segmentation	DAE	BRATS 2015	Stacked denoising auto-encoders was pre-trained with HGG images and fine-tuned with LGG images
Alqazzaz 2019 [45]	Brain tumor segmentation	CNN	BRATS 2017	3D CNN SegNet models was used for feature extraction and decision tree was used for classification of the combined features
Amin 2019 [44]	Brain tumor segmentation and classification	CNN (Alex- and Google-net)	BRATS 2013 2014, 2015 and 2016	Initial segmentation of images by optimal thresholding on pixel intensities and classification by using the fine-tuned pre-trained Alex- and Google-nets
Amin 2020 [43]	Glioma and stroke lesion detection	CNN	BRATS 2013, 2014, 2015, 2016 and 2017	Using a CNN consists of 14 layers for glioma and stroke lesion detection
Cahall 2019 [54]	Brain tumor segmentation	CNN (U-Net)	BRATS 2018	Modified U-Net architecture with inception modules was used for multi-scale feature extraction
Chen 2019 [32]	Brain tumor segmentation	CNN	BRATS 2015	CNN-based segmentation networks was extended by adding symmetric masks in several layers
Cui 2018 [42]	Brain gliomas segmentation and classification	CNN	BRATS 2015	A cascaded convolutional neural network consisting of two subnetworks was used for tumor localization and intra-tumor classification
Hussain 2017 [33]	Brain tumor segmentation	CNN	BRATS 2013	Images was normalized and bias field corrected, Deep CNN was used for segmentation, and small false positives were removed using morphological operators
Hussain 2018 [34]	Glioma tumors segmentation	CNN	BRATS 2013–2015	A patch based approach along with an inception module was used for training a deep CNN
Hasan 2018 [35]	Brain tumor segmentation	CNN (U-Net)	BRATS 2017	The U-net model was improved by replacing the de-convolution component with an up-sampled by the Nearest-neighbor algorithm
Khawaldeh 2018 [50]	Grading of glioma tumor	CNN (ConvNet)	Images from the Cancer Imaging Archive (TCIA)	ConvNet was used for classifying brain images into healthy brains, brains with low-grade tumor and brains with high-grade tumor
Kuzina 2019 [41]	Brain tumor segmentation	VAE	BRATS 2018 and MS	Bayesian generative models was used for knowledge transfer in MRI segmentation
Li 2017 [56]	Low-grade glioma segmentation	CNN	BRATS 2013	A multi-pathway CNN and fully connected conditional random field were combined for more accurate recognition of glioma with low contrast

Table 1 (continued)

Study ID Author, year	Task	Method	Dataset	Remarks
Li 2019 [40]	Brain tumor segmentation	CNN (U-Net)	BRATS 2015 and 2016	Modifying U-Net architecture by: 1) adding skip connection between the encoding and decoding paths, and 2) adopting an inception module in each block, for tumor segmentation
Ma 2018 [51]	Nasopharyngeal carcinoma segmentation	CNN	Clinical MRI images of 30 NPC patients	A combination of deep CNN and a 3D graph-cut-based method was used for NPC segmentation
Michael Mahesh 2020 [47]	Severity-level categorization of glioma tumors	CNN	BRATS 2012 and 2018	Severity-level of glioma tumors was categorized by a deep CNN based on the features obtained from the images segmented by PSO
Mlynarski 2019 [52]	Brain tumor segmentation	CNN (U-Net)	BRATS 2018	A CNN-based segmentation model was trained using weakly annotated images in addition to fully annotated images
KV AM [48]	Glioma tumor grade identification	CNN (VGG-19)	Clinical images from 20 patients	Noise removal followed by the weighted neighbor distance algorithm and VGG-19 deep CNN were used for the identification of glioma tumor grade
Pereira 2016 [39]	Brain tumor segmentation	CNN	BRATS 2013	Segmentation of gliomas was investigated by using deep CNN with small convolutional kernels
Savareh 2019 [53]	Brain tumor segmentation	CNN	BRATS 2015	The architecture of CNN was enhanced by wavelet transform in brain tumor segmentation
Sultan 2019 [49]	Classification of brain tumor	CNN	Clinical images from 306 patients	One deep CNN was used to classify tumors into meningioma, glioma, and pituitary tumor. The other one differentiates between the three glioma grades
Sun 2019 [38]	Brain tumor segmentation	CNN	BRATS 2018	Ensembles of three different 3D CNN was used for robust segmentation of images through majority rule
Sun 2019 [38]	Brain tumor segmentation	CNN (U-Net)	BRATS 2015	Simple skip connection in U-Net was replaced with encoder adaptation blocks and dense connected fusion blocks was also used in decoder in order to better segmentation
Wang 2019 [55]	Brain tumor segmentation	CNN(U-Net)	BRATS 2018	U-Net was modified by residual blocks nested with dilations in the encoding part and squeeze-and-excitation blocks in both the encoding and decoding parts
Wang 2019 [46]	Brain tumor segmentation	CNN (ResNet)	BRATS 2015 and 2018	A residual network along with a pyramid pool network automatically segment glioma end to end

Table 1 (continued)

Study ID Author, year	Task	Method	Dataset	Remarks
Wu 2019 [37]	Glioma segmentation based on	SVM	BRATS 2017	Super-pixel image was generated based on a simple iterative clustering method. SVM was used to train classification model based on a training set generated by calculating the statistical, texture, curvature and fractal features for each super-pixel
Yang 2019 [36]	Brain tumor segmentation	CNN and RF	BRATS 2015	Small kernels two-path CNN was used for feature extraction and random forest was used as the discriminative model on the learned features
Zhao 2018 [31]	Brain tumor segmentation	CNN	BRATS 2013, 2015 and 2016	An integration of fully CNN and CRFs in a unified framework was used for brain tumor segmentation
Zhuge 2017 [30]	Brain tumor segmentation	HNN	BRATS 2013	An extension from CNN with a deep supervision through holistically nested HNN was trained to learn the multi-scale and multi-level hierarchical appearance representation of the brain tumor in MR images
Deepak 2019 [57]	Brain tumor classification	CNN	Figshare	A pre-trained GoogLeNet was used to extract features from brain MR images

CNN, convolutional neural network; *CRF,* conditional random fields; *CT,* computed tomography; *IDH,* isocitrate dehydrogenase; *DT,* decision tree; *DWI,* diffusion-weighted imaging; *Flair,* fluid attenuated inversion recovery; *GA,* genetic algorithm; *GAN,* generative adversarial network; *H&N,* head and neck; *HGG,* high-grade glioma; *HNN,* nested neural networks; *HOG,* histogram orientation gradient; *KNN,* K-nearest neighbor; *LBP,* local binary pattern; *LGG,* low-grade glioma; *MRI,* magnetic resonance imaging; *OAR,* organ at risk; *PSO,* particle swarm optimization; *RF,* random forests; *ROI,* region of interest; *RTO,* radiation therapy oncology group; *SVM,* support vector machine; $T1$, longitudinal relaxation time; $T2$, transverse relaxation time; *VGG,* visual geometry group

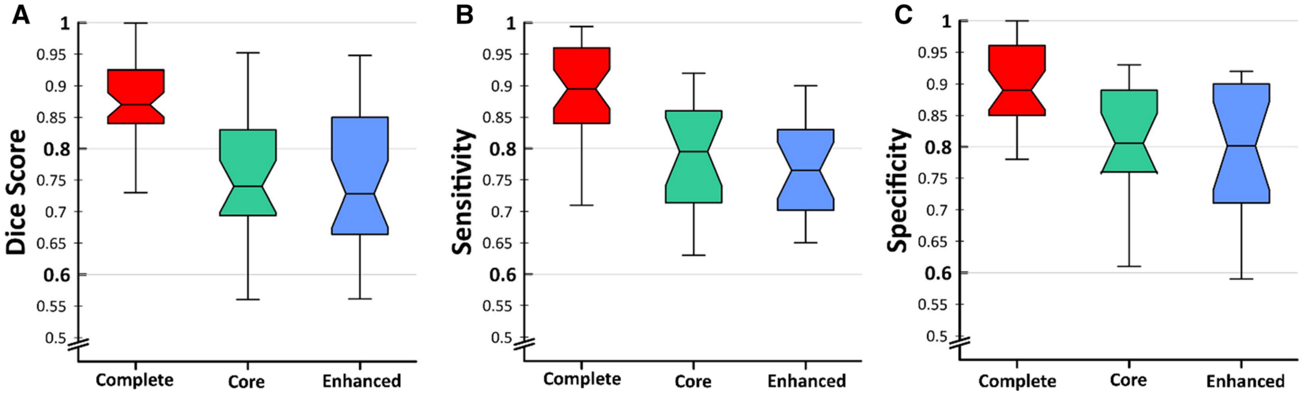

Fig. 2 Box plot results of **a** dice score; **b** sensitivity; and **c** specificity of auto-segmentation versus ground-truth for each tumor region

Table 2 Results of auto-segmentation versus ground-truth for each tumor region in different views

	Dice coefficient				Sensitivity				Specificity			
	Multi-view	Axial	Coronal	Sagittal	Multi-view	Axial	Coronal	Sagittal	Multi-view	Axial	Coronal	Sagittal
Complete tumor	0.831	0.815	0.815	0.797	0.836	0.817	0.785	0.811	0.853	0.834	0.792	0.738
Core tumor	0.73	0.699	0.698	0.702	0.734	0.729	0.77	0.723	0.785	0.76	0.714	0.746
Enhanced tumor	0.63	0.622	0.622	0.589	0.698	0.686	0.663	0.716	0.611	0.611	0.617	0.54

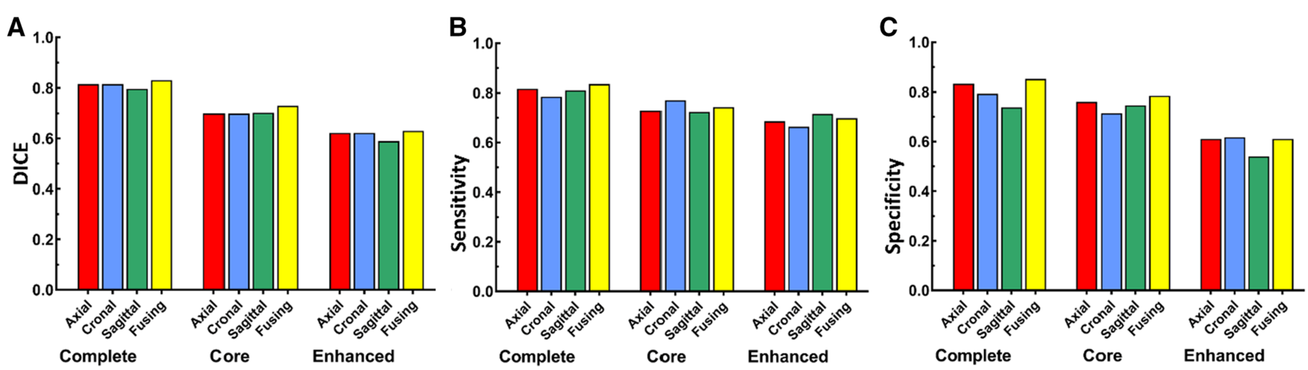

Fig. 3 Bar chart results of **a** dice score; **b** sensitivity; and **c** specificity of auto-segmentation versus ground-truth for each tumor region in axial, coronal, sagittal, and fusing views

an excellent sensitivity and specificity of 97.61% and 96.23%, and a total accuracy of 96.58%. Classifying grade III glioma revealed 94.65% accuracy, 84.28% sensitivity, and 98.08% specificity. For grade II, the accuracy, sensitivity, and specificity were 95.14%, 90.36% and 96.8%, respectively. Accuracy of 99.76%, sensitivity of 98.99%, and specificity of 99.79% were obtained to classify grade I glioma. In the normal cases, almost all cases have identified correctly with high accuracy of 99.72%.

Discussion

In this meta-analysis, of the performance of DL-based algorithms in the segmentation and classification of MRI images in cases with HNCs and/or brain tumors were evaluated. Of the 109 related articles, 30 eligible studies were considered for this meta-analysis. Based on the analyzes performed on the selected studies, for the segmentation

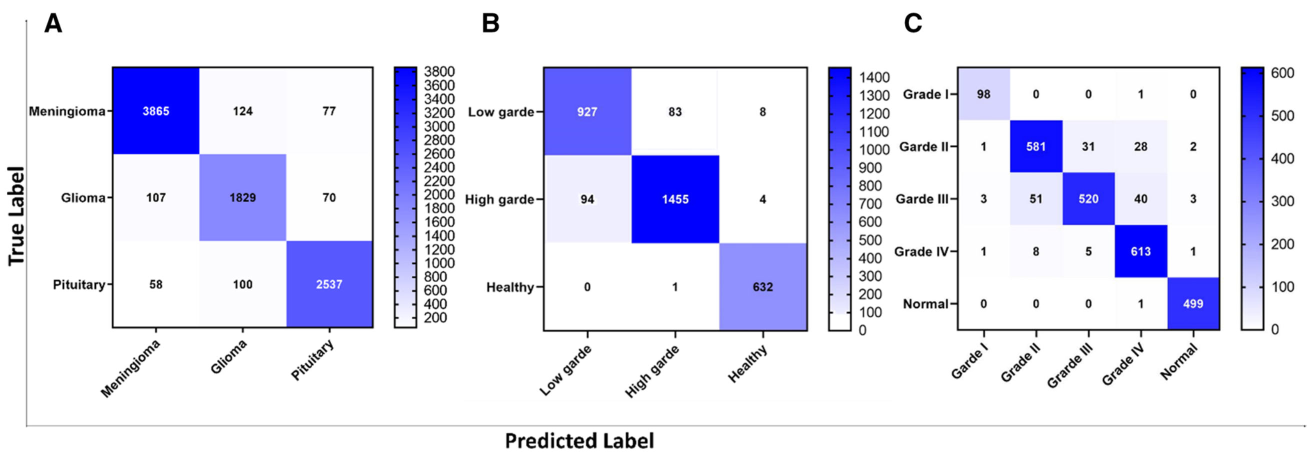

Fig. 4 Confusion matrices for classification of tumor by **a** type; **b** low-grade (I and II grade) and high-grade (III and IV grade); and **c** I, II, III, and IV grades of auto classification versus manual classification

of HNCs and brain tumors, the dice coefficient, sensitivity, and specificity were examined, and the results show the high capability of DL-based methods in the complete tumor, tumor core, and enhanced tumor. The performance of DL models for the segmentation of HNCs and brain tumors is reduced from complete tumor to tumor core and enhanced tumor, respectively. Besides, the efficiency of segmentation was evaluated for complete tumor region based on JSC, ASSD, PM, and CR metrics in four studies. These parameters have not been reported in other studies, but can be used to evaluate the efficiency of segmentation methods.

The study also examined the performance of segmentation methods using DL in determining the grade of glioma based on different metrics. The necessary information for this analysis was not available in all studies; therefore, this issue has been investigated in some studies. The studies were compared based on the views of images used for segmentation, and the results were reported based on the dice coefficient, sensitivity and specificity. The results mainly show that the values obtained for different views are close and in descending order for the complete tumor, tumor core, and enhanced tumor, respectively.

The accuracy for classifying glioma, meningioma, and pituitary tumors by DL methods is more than 96%. These values for the sensitivity and specificity of the DL methods are also more than 93% for tumor-type classification. The values of accuracy, sensitivity, and specificity were also examined for the classification of tumor grades, which indicate the significant ability of automated methods based on DL. Due to the considerable ability of DL methods in the studied fields, some of the challenges and capabilities of these methods will be reviewed in the following.

Based on extensive studies, DL approaches have revolutionized the various fields of medical image processing and analysis. In many applications related to the classification and segmentation of medical images, DL approaches work better than other traditional methods [62, 63]. DL also facilitates the process of image analysis by automatically extracting features from images instead of extracting handcrafted features. However, one of the major challenges in applying these approaches is the need for large amounts of data due to the multiplicity of parameters to be learned in deep neural networks. This large amount of data must also be labeled, which requires a time-consuming process to produce the labels. To address these challenges, researchers in the field of machine learning have made efforts to use data augmentation techniques or transfer learning.

Data augmentation refers to the concept in which the number of images is increased using various approaches such as translational changes, rotation, mirroring, optical transforms, and other methods of image processing, or methods based on generative adversarial networks (GANs) [64]. Transfer learning is also a concept that uses a neural network that has been previously trained for similar applications, and this network must then be fine-tuned using available data [65]. Using domain adaptation approaches can also greatly solve the problem of the need for labeled data [66]. Therefore, using these methods can improve the performance of DL approaches.

Conclusion

In this meta-analysis, some studies related to the use of DL methods in segmentation and classification of HNCs and brain tumors in MRI images were evaluated. Based on the obtained results, we can acknowledge the significant ability of DL methods in the mentioned applications. Paying attention to the strategies such as data augmentation,

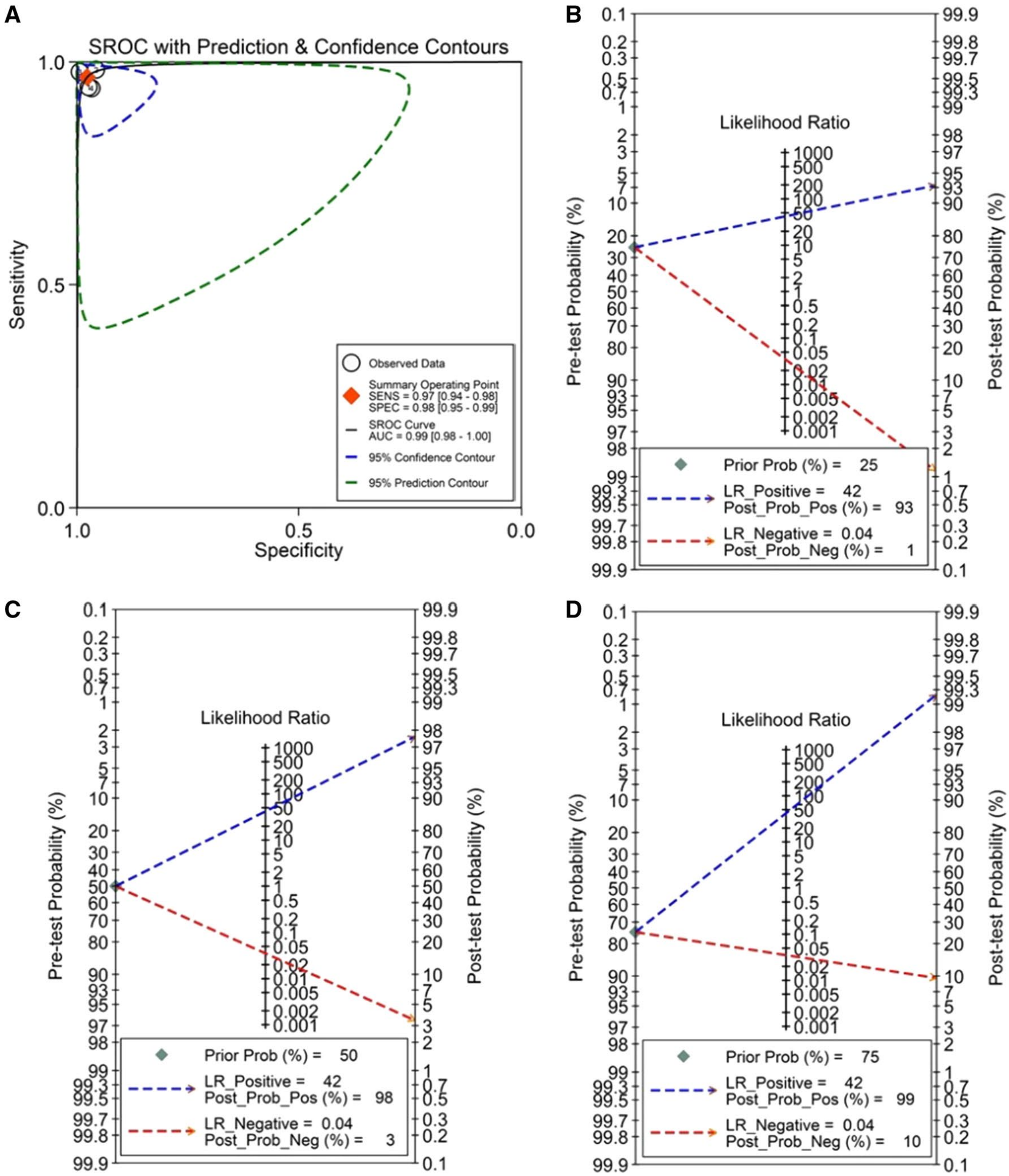

Fig. 5 **a**SROC curve for diagnostic accuracy of DL in discrimination of glioma from meningioma and pituitary tumors; Fagan's nomogram with **b** 25% pretest probability; **c** 50% pretest probability; **d** 75% pre-test probability. Summary receiver operating characteristic (SROC), sensitivity (SENS), specificity (SPEC), area under the curve (AUC), probability (Prob), likelihood ratio (LR)

transfer learning, and domain adaptation can increase the performance of DL-based segmentation and classification methods, and handle challenges related to DL such as limited annotated data, overfitting, class imbalance and so on. We do believe that this paper can help researchers interested in this field to choose a suitable DL model and method of defining the study methodology and show the main challenges of this field and ways to deal with them. Poor reporting in these studies challenges the analysis process, so it is recommended that future studies report comprehensive results based on different metrics.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11548-021-02326-z.

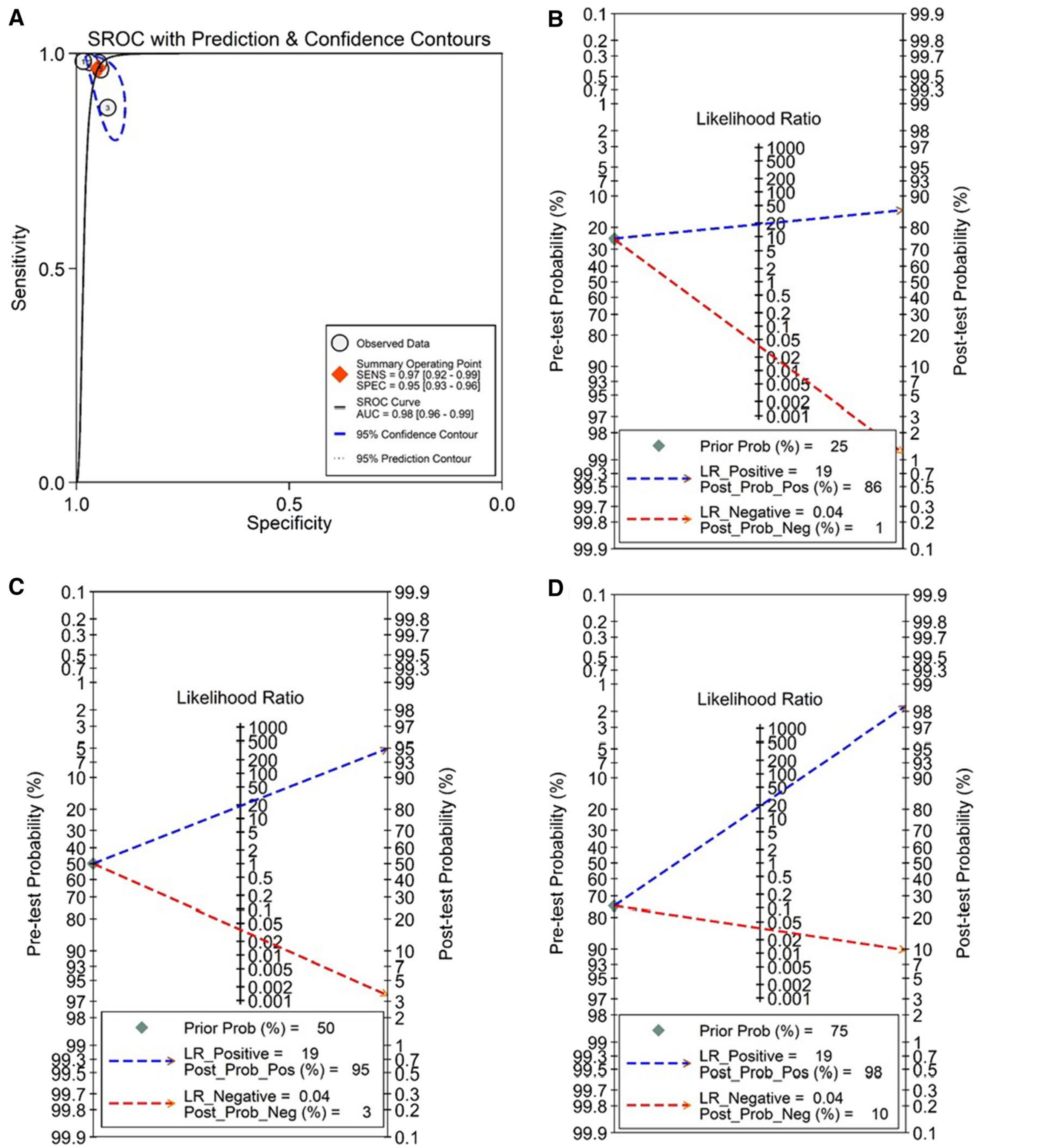

Fig. 6 **a** SROC curve for diagnostic accuracy of DL in discrimination of high-grade glioma from low-grade glioma tumors; Fagan's nomogram with **b** 25% pretest probability; **c** 50% pretest probability; **d** 75% pretest probability. Summary receiver operating characteristic (SROC), sensitivity (SENS), specificity (SPEC), area under the curve (AUC), probability (Prob), likelihood ratio (LR)

Acknowledgements The authors wish to acknowledge Kermanshah University of Medical Sciences, Kermanshah, Iran, for their support, cooperation and assistance throughout the period of study.

Funding This work was supported by the Kermanshah University of Medical Sciences [Grant numbers 990688].

Declarations

Conflicts of interest The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

1. Fitzmaurice C, Allen C, Barber RM, Barregard L, Bhutta ZA, Brenner H, Dicker DJ, Chimed-Orchir O, Dandona R, Dandona L, Fleming T, Forouzanfar MH, Hancock J, Hay RJ, Hunter-Merrill R, Huynh C, Hosgood HD, Johnson CO, Jonas JB, Khubchandani J, Kumar GA, Kutz M, Lan Q, Larson HJ, Liang X, Lim SS, Lopez AD, MacIntyre MF, Marczak L, Marquez N, Mokdad AH, Pinho C, Pourmalek F, Salomon JA, Sanabria JR, Sandar L, Sartorius B, Schwartz SM, Shackelford KA, Shibuya K, Stanaway J, Steiner C, Sun J, Takahashi K, Vollset SE, Vos T, Wagner JA, Wang H, Westerman R, Zeeb H, Zoeckler L, Abd-Allah F, Ahmed MB, Alabed S, Alam NK, Aldhahri SF, Alem G, Alemayohu MA, Ali R, Al-Raddadi R, Amare A, Amoako Y, Artaman A, Asayesh H, Atnafu N, Awasthi A, Saleem HB, Barac A, Bedi N, Bensenor I, Berhane A, Bernabé E, Betsu B, Binagwaho A, Boneya D, Campos-Nonato I, Castañeda-Orjuela C, Catalá-López F, Chiang P, Chibueze C, Chitheer A, Choi JY, Cowie B, Damtew S, Das Neves J, Dey S, Dharmaratne S, Dhillon P, Ding E, Driscoll T, Ekwueme D, Endries AY, Farvid M, Farzadfar F, Fernandes J, Fischer F, Gebremedhin TT, Gebru A, Gopalani S, Hailu A, Horino M, Horita N, Husseini A, Huybrechts I, Inoue M, Islami F, Jakovljevic M, James S, Javanbakht M, Jee SH, Kasaeian A, Kedir MS, Khader YS, Khang YH, Kim D, Leigh J, Linn S, Lunevicius R, El Razek HMA, Malekzadeh R, Malta DC, Marcenes W, Markos D, Melaku YA, Meles KG, Mendoza W, Mengiste DT, Meretoja TJ, Miller TR, Mohammad KA, Mohammadi A, Mohammed S, Moradi-Lakeh M, Nagel G, Nand D, Le Nguyen Q, Nolte S, Ogbo FA, Oladimeji KE, Oren E, Pa M, Park EK, Pereira DM, Plass D, Qorbani M, Radfar A, Rafay A, Rahman M, Rana SM, Søreide K, Satpathy M, Sawhney M, Sepanlou SG, Shaikh MA, She J, Shiue I, Shore HR, Shrime MG, So S, Soneji S, Stathopoulou V, Stroumpoulis K, Sufiyan MB, Sykes BL, Tabarés-Seisdedos R, Tadese F, Tedla BA, Tessema GA, Thakur JS, Tran BX, Ukwaja KN, Uzochukwu BSC, Vlassov VV, Weiderpass E, Wubshet Terefe M, Yebyo HG, Yimam HH, Yonemoto N, Younis MZ, Yu C, Zaidi Z, Zaki MES, Zenebe ZM, Murray CJL, Naghavi M (2017) Global, Regional, and National Cancer Incidence, Mortality Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-years for 32 Cancer Groups, 1990 to 2015 A Systematic Analysis for the Global Burden of Disease Study. JAMA oncology 3(4):524–548
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A (2018) Global cancer statistics 2018 GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a Cancer J Clin 68(6):394–424. https://doi.org/10.3322/caac.21492
3. Argirion I, Zarins KR, Defever K, Suwanrungruang K, Chang JT, Pongnikorn D, Chitapanarux I, Sriplung H, Vatanasapt P, Rozek LS (2019) Temporal changes in head and neck cancer incidence in thailand suggest changing oropharyngeal epidemiology in the region. J Global oncol 5:1–11. https://doi.org/10.1200/jgo.18.00219
4. Mehanna H, Paleri V, West CM, Nutting C (2010) Head and neck cancer–Part 1: epidemiology, presentation, and prevention. BMJ (Clinical Research Ed) 341:c4684. https://doi.org/10.1136/bmj.c4684
5. Harrison LB, Sessions RB, Hong WK (2009) Head and neck cancer: a multidisciplinary approach. Lippincott Williams & Wilkins
6. Konings H, Stappers S, Geens M, De Winter BY, Lamote K, van Meerbeeck JP, Specenier P, Vanderveken OM, Ledeganck KJ (2020) A literature review of the potential diagnostic biomarkers of head and neck neoplasms. Front Oncol 10:1020
7. Kies MS, Bennett CL, Vokes EE (2001) Locally advanced head and neck cancer. Curr Treat Options Oncol 2(1):7–13
8. Al-Sarraf M (2002) Treatment of locally advanced head and neck cancer: historical and critical review. Cancer Control 9(5):387–399
9. Vries N, van de Heyning P, Leemans CR (2013) Leerboek keel-neus-oorheelkunde en hoofd-halschirurgie. Bohn Stafleu van Loghum Springer Media
10. Yeh S-A Radiotherapy for head and neck cancer. In: Seminars in plastic surgery, 2010. vol 2. Thieme Medical Publishers, p 127
11. Baxi SS, Pinheiro LC, Patil SM, Pfister DG, Oeffinger KC, Elkin EB (2014) Causes of death in long-term survivors of head and neck cancer. Cancer 120(10):1507–1513
12. Ferlito A, Shaha AR, Silver CE, Rinaldo A, Mondin V (2001) Incidence and sites of distant metastases from head and neck cancer. ORL 63(4):202–207
13. Claus EB, Walsh KM, Wiencke JK, Molinaro AM, Wiemels JL, Schildkraut JM, Bondy ML, Berger M, Jenkins R, Wrensch M (2015) Survival and low-grade glioma: the emergence of genetic information. Neurosurg Focus 38(1):E6
14. Drevelegas A (2002) Embryonal tumors. In: Ayache N, Duncan J (eds) Imaging of brain tumors with histological correlations. Springer, pp 147–154
15. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, Van Der Laak JA, Van Ginneken B, Sánchez CI (2017) A survey on deep learning in medical image analysis. Med Image Anal 42:60–88
16. Brown RW, Cheng Y-CN, Haacke EM, Thompson MR, Venkatesan R (2014) Magnetic resonance imaging: physical principles and sequence design. Wiley
17. Yasmin M, Mohsin S, Sharif M, Raza M, Masood S (2012) Brain image analysis: a survey. World Appl Sci J 19(10):1484–1494
18. Yasmin M, Sharif M, Masood S, Raza M, Mohsin S (2012) Brain image enhancement-A survey. World Appl Sci J 17(9):1192–1204
19. Nida N, Sharif M, Khan MUG, Yasmin M, Fernandes SL (2016) A framework for automatic colorization of medical imaging. IIOAB J 7:202–209
20. Román KL-L, Ocaña MIG, Urzelai NL, Ballester MÁG, Oliver IM (2020) Medical Image Segmentation Using Deep Learning. In: Deep Learning in Healthcare. Springer, pp 17–31
21. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Chen Y-W, Jain LC (eds) International conference on medical image computing and computer-assisted intervention, 2016. Springer, pp 424-432
22. Kamnitsas K, Ledig C, Newcombe VF, Simpson JP, Kane AD, Menon DK, Rueckert D, Glocker B (2017) Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med Image Anal 36:61–78
23. LeCun Y (2015) LeNet-5, Convolutional neural networks. http://yann.lecun.com/exdb/lenet 20 (5):14
24. Wang W, Liang D, Chen Q, Iwamoto Y, Han X-H, Zhang Q, Hu H, Lin L, Chen Y-W (2020) Medical image classification using deep learning. In: Chen Y-W, Jain LC (eds) Deep learning in healthcare: paradigms and applications. Springer, Cham, pp 33–51. https://doi.org/10.1007/978-3-030-32606-7_3
25. Cheng J, Huang W, Cao S, Yang R, Yang W, Yun Z, Wang Z, Feng Q (2015) Enhanced performance of brain tumor classification via tumor region augmentation and partition. PLoS ONE 10(10):e0140381
26. Ertosun MG, Rubin DL Automated grading of gliomas using deep learning in digital pathology images: A modular approach with ensemble of convolutional neural networks. In: AMIA Annual Symposium Proceedings, 2015. American Medical Informatics Association, p 1899
27. Afshar P, Plataniotis KN, Mohammadi A Capsule networks for brain tumor classification based on MRI images and coarse tumor boundaries. In: ICASSP 2019–2019 IEEE International Conference

on Acoustics, Speech and Signal Processing (ICASSP), 2019. IEEE, pp 1368–1372
28. Anaraki AK, Ayati M, Kazemi F (2019) Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms. Biocybernet Biomed Engg 39(1):63–74
29. Alex V, Vaidhya K, Thirunavukkarasu S, Kesavadas C, Krishnamurthi G (2017) Semisupervised learning using denoising autoencoders for brain lesion detection and segmentation. J Med Imag 4(4):041311
30. Zhuge Y, Krauze AV, Ning H, Cheng JY, Arora BC, Camphausen K, Miller RW (2017) Brain tumor segmentation using holistically nested neural networks in MRI images. Med Phys 44(10):5234–5243
31. Zhao X, Wu Y, Song G, Li Z, Zhang Y, Fan Y (2018) A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. Med Image Anal 43:98–111
32. Chen H, Qin Z, Ding Y, Tian L, Qin Z (2019) Brain tumor segmentation with deep convolutional symmetric neural network. Neurocomputing 392:305–313
33. Hussain S, Anwar SM 2017 Majid M Brain tumor segmentation using cascaded deep convolutional neural network. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, pp 1998–2001
34. Hussain S, Anwar SM, Majid M (2018) Segmentation of glioma tumors in brain using deep convolutional neural network. Neurocomputing 282:248–261
35. Hasan SK, Linte CA 2018 A modified U-Net convolutional network featuring a Nearest-neighbor Re-sampling-based Elastic-Transformation for brain tissue characterization and segmentation. In: 2018 IEEE Western New York Image and Signal Processing Workshop (WNYISPW), IEEE, pp 1–5
36. Yang T, Song J, Li L (2019) A deep learning model integrating SK-TPCNN and random forests for brain tumor segmentation in MRI. Biocybernet Biomed Eng 39(3):613–623
37. Wu Y, Zhao Z, Wu W, Lin Y, Wang M (2019) Automatic glioma segmentation based on adaptive superpixel. BMC Med Imag 19(1):1–14
38. Sun L, Zhang S, Chen H, Luo L (2019) Brain tumor segmentation and survival prediction using multimodal MRI scans with deep learning. Front Neurosci 13:810
39. Pereira S, Pinto A, Alves V, Silva CA (2016) Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Trans Med Imag 35(5):1240–1251
40. Li H, Li A, Wang M (2019) A novel end-to-end brain tumor segmentation method using improved fully convolutional networks. Comput Biol Med 108:150–160
41. Kuzina A, Egorov E, Burnaev E (2019) Bayesian generative models for knowledge transfer in mri semantic segmentation problems. Front Neurosci 13:844
42. Cui S, Mao L, Jiang J, Liu C, Xiong S (2018) Automatic semantic segmentation of brain gliomas from MRI images using a deep cascaded neural network. J Healthcare Eng 2018:14. https://doi.org/10.1155/2018/4940593
43. Amin J, Sharif M, Anjum MA, Raza M, Bukhari SAC (2020) Convolutional neural network with batch normalization for glioma and stroke lesion detection using MRI. Cogn Syst Res 59:304–311
44. Amin J, Sharif M, Yasmin M, Saba T, Anjum MA, Fernandes SL (2019) A new approach for brain tumor segmentation and classification based on score level fusion using transfer learning. J Med Syst 43(11):326. https://doi.org/10.1007/s10916-019-1453-8
45. Alqazzaz S, Sun X, Yang X, Nokes L (2019) Automated brain tumor segmentation on multi-modal MR image using SegNet. Comput Vis Media 5(2):209–219. https://doi.org/10.1007/s41095-019-0139-y
46. Wang Y, Li C, Zhu T, Zhang J (2019) Multimodal brain tumor image segmentation using WRN-PPNet. Comput Med Imag Graph 75:56–65
47. Michael Mahesh K, Arokia Renjit J (2020) Multiclassifier for severity-level categorization of glioma tumors using multimodal magnetic resonance imaging brain images. Int J Imag Syst Technol 30(1):234–251
48. Am KV, Rajendran V (2019) Glioma tumor grade identification using artificial intelligent techniques. J Med Syst 43(5):113
49. Sultan HH, Salem NM, Al-Atabany W (2019) Multi-classification of brain tumor images using deep neural network. IEEE Access 7:69215–69225
50. Khawaldeh S, Pervaiz U, Rafiq A, Alkhawaldeh RS (2018) Noninvasive grading of glioma tumor using magnetic resonance imaging with convolutional neural networks. Appl Sci 8(1):27
51. Ma Z, Wu X, Song Q, Luo Y, Wang Y, Zhou J (2018) Automated nasopharyngeal carcinoma segmentation in magnetic resonance images by combination of convolutional neural networks and graph cut. Exp Therapeut Med 16(3):2511–2521
52. Mlynarski P, Delingette H, Criminisi A, Ayache N (2019) Deep learning with mixed supervision for brain tumor segmentation. J Med Imag 6(3):034002
53. Savareh BA, Emami H, Hajiabadi M, Azimi SM, Ghafoori M (2019) Wavelet-enhanced convolutional neural network: a new idea in a deep learning paradigm. Biomed Eng Biomed Technik 64(2):195–205
54. Cahall DE, Rasool G, Bouaynaya NC, Fathallah-Shaykh HM (2019) Inception modules enhance brain tumor segmentation. Front Comput Neurosci 13:44
55. Wang L, Wang S, Chen R, Qu X, Chen Y, Huang S, Liu C (2019) Nested dilation networks for brain tumor segmentation based on magnetic resonance imaging. Front Neurosci 13:285
56. Li Z, Wang Y, Yu J, Shi Z, Guo Y, Chen L, Mao Y (2017) Low-grade glioma segmentation based on CNN with fully connected CRF. J Healthcare Eng 2017:12. https://doi.org/10.1155/2017/9283480
57. Deepak S, Ameer P (2019) Brain tumor classification using deep CNN features via transfer learning. Comput Biol Med 111:103345
58. Ma Z, Zhou S, Wu X, Zhang H, Yan W, Sun S, Zhou J (2019) Nasopharyngeal carcinoma segmentation based on enhanced convolutional neural networks using multi-modal metric learning. Phys Med Biol 64(2):025005
59. Wang Y, Zu C, Hu G, Luo Y, Ma Z, He K, Wu X, Zhou J (2018) Automatic tumor segmentation with deep convolutional neural networks for radiotherapy applications. Neural Process Lett 48(3):1323–1334
60. Li Q, Xu Y, Chen Z, Liu D, Feng S-T, Law M, Ye Y, Huang B (2018) Tumor segmentation in contrast-enhanced magnetic resonance imaging for nasopharyngeal carcinoma: deep learning with convolutional neural network. BioMed Res Int 2018:7. https://doi.org/10.1155/2018/9128527
61. Lin L, Dou Q, Jin Y-M, Zhou G-Q, Tang Y-Q, Chen W-L, Su B-A, Liu F, Tao C-J, Jiang N (2019) Deep learning for automated contouring of primary tumor volumes by MRI for nasopharyngeal carcinoma. Radiology 291(3):677–686
62. Haque IRI, Neubert J (2020) Deep learning approaches to biomedical image segmentation. Info Med Unlock 18:100297
63. Hesamian MH, Jia W, He X, Kennedy P (2019) Deep learning techniques for medical image segmentation: achievements and challenges. J Digit Imag 32(4):582–596
64. Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. J Big Data 6(1):60
65. Pan SJ, Yang Q (2009) A survey on transfer learning. IEEE Trans Knowl Data Eng 22(10):1345–1359
66. Wang M, Deng W (2018) Deep visual domain adaptation: a survey. Neurocomputing 312:135–153