Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2021) 29: 616 – 631

doi:10.3906/elk-2002-175

Ensemble learning of multiview CNN models for survival time prediction of brain tumor patients using multimodal MRI scans

Abdela Ahmed MOSSA[1,*], **Ulus ÇEVİK**[2]

[1]Computer Engineering, Faculty of Engineering, Çukurova University, Adana, Turkey
[2]Electrical and Electronics Engineering, Faculty of Engineering, Çukurova University, Adana, Turkey

Received: 28.02.2020 • **Accepted/Published Online:** 10.09.2020 • **Final Version:** 30.03.2021

Abstract: Brain tumors have been one of the most common life-threatening diseases for all mankind. There have been huge efforts dedicated to the development of medical imaging techniques and radiomics to diagnose tumor patients quickly and efficiently. One of the main aims is to ensure that preoperative overall survival time (OS) prediction is accurate. Recently, deep learning (DL) algorithms, and particularly convolutional neural networks (CNNs) achieved promising performances in almost all computer vision fields. CNNs demand large training datasets and high computational costs. However, curating large annotated medical datasets are difficult and resource-intensive. The performances of single learners are also unsatisfactory for small datasets. Thus, this study was conducted to improve the performance of CNN models on small volumetric datasets through developing a DL-based ensemble method for OS classification of brain tumor patients using multimodal magnetic resonance images (MRI). First, we proposed multiview CNNs: OS classifiers based on representing the 3D MRI data as a set of 2D slices along all three planes (axial, sagittal, and coronal) and process them using 2D CNNs. Subsequently, the predicted probabilities by the multiview CNN models were fused using standard machine learning algorithms. The proposed approach was experimentally evaluated on 163 patients obtained from the BraTS'17 training dataset. Our best model achieved an AUC and accuracy values of 0.93 and 92.9%, respectively, on classifying patients with brain tumors into two OS groups, outperforming current state-of-the-art results. In addition, the FLAIR MRI modality yielded the best classification accuracy compared to other MRI modalities. Similarly, axial projections had the best classification performance compared to coronal and sagittal projections. Our findings may provide valuable insights for physicians in advancing treatment planning via noninvasive and accurate prediction of survival using only MRIs at the time of diagnosis.

Key words: Brain tumor, convolutional neural networks, deep learning, ensemble learning, multimodal MRI, survival time prediction

1. Introduction

Brain tumor is one of the most common deadly type of cancers characterized by the uncontrolled growth and spread of abnormal cells in the brain, and the level of aggressiveness and survival time of its patients vary widely among individuals [1]. Gliomas that originate in the glial cells are primary brain tumors, accounting for about 80% of all malignant tumors [2]. Glioblastoma [also called high grade glioma (HGG)] is the most rapidly progressive glioma, responsible for the majority of malignant brain tumors in adults [3] for which the median survival time of patients is 10–15 months [4]. In the United States, five-year survival rate of patients with glioblastoma was 5.6% for the years 2000–2015, and 13,310 new cases were projected for the year 2019 [5].

*Correspondence: amossa@student.cu.edu.tr

Among the top leading deaths caused by cancers, glioblastoma is ranked in first and second for children under the age of 15 and young adults ranging from 15 to 34 years of age, respectively [6]. Thus, early and accurate preoperative survival time prediction of patients with HGG is an important factor to guide future treatment decisions.

Due to the complex mechanical properties of the brain tumor tissues, the conventional survival time predictions of glioma patients using clinical inspections by experienced physicians based on histological examinations following tumor resection and an invasive biopsy are often inaccurate, susceptible to interobserver variability and time-consuming [7, 8]. This results in a necessity for medical imaging techniques, as MRI, that depict the entire glioma region noninvasively. Images generated using MRI provide decision-making information, which supports physicians for evaluating the severity of the disease and monitoring the disease's progression [9, 10].

Fluid attenuation inversion recovery (FLAIR), T2-weighted, T1-weighted, diffusion tensor imaging (DTI), and T1-weighted contrast-enhanced (T1CE) are among the most commonly used MRI modalities in the brain tumor diagnostic field. They are capable of measuring different yet complementary information [11]. In addition, anatomical and pathological viewing in all the three planes (axial, coronal and sagittal) using multiple MR imaging sequences is helpful to the physicians in diagnosing glioma in patients. Therefore, combining the information properly from different modalities and across the three planes may enhance the diagnostic performance. However, when evaluating survival prediction of glioma patients using these multimodal images, which modalities and projections are the most effective, and whether combinations of multiple modalities can improve the performance are still unclear and not investigated very well [12]. MRI also comes at the cost of generating a large number of images per patient, exceeding the capacity of available physicians, particularly in middle- and low-income countries [13]. Consequently, the development of automatic computer assisted diagnosis systems (CAD) for more accurate survival prediction based on the abundant multimodal brain MRI data acquired prior to any invasive examination is currently of great interest [14].

Because of the recent advancements in computational devices and availability of large medical data, studies have highlighted the importance of artificial intelligence (AI) algorithms such as artificial neural networks and machine learning methods for the automatic diagnosis of diseases [15, 16]. Similarly, several AI techniques have been proposed for prediction of overall survival time (OS) of brain tumor patients. These methods can be coarsely classified into three categories. In the first category, handcrafted-based approach, manual or automatic segmentation of glioma are performed in MRI images. Different features like volumetric parameters and histogram will be extracted from the segmented regions. Thereafter, features with better prognostic values will be selected and fused with metadata to train ML algorithms. However, optimizing feature extraction and selection methods are a big challenge in this approach [17–19]. In the second approach, deep learning-based, discriminative features are learned directly from the MRI images [20]. In the third category, hybrid approaches, deep features that are extracted using DL methods, handcrafted features that are extracted from automatic segmented tumor regions, and clinical data are combined to create a feature fusion matrix. This matrix is then used as input to train ML algorithms [21–23].

During the past few years, various 2D and 3D CNN deep learning algorithms have been developed for applications that range from music transcription [24] to computer vision [25]. These emergence of CNN algorithms have fascinated researchers to employ it in medical image analysis [26]. However, this excitement should not overrule concerns raised regarding the application of this deep learning technologies to medicine,

including misdiagnosis and attribution of liability due to errors [27]. Numerous CNN algorithms have also been developed for OS classification of brain tumor patients to improve the performance on multimodal MRI. Nie et al. [28] proposed a hybrid approach using multichannel 3D CNN and SVM to distinguish between long and short OS. Lao et al. [22] reconstructed 3D MRIs into 2D slices in the axial plane, and selected three slices that had the largest tumor core, necrosis and whole tumor area, respectively. The selected slices were then used to fine-tune the pretrained CNN models on ILSVRC-2012 dataset. Similarly, Li and Shen [20] developed a CNN-based framework that takes the four whole brain and tumor segmentation images as an input. In another study [21], authors have extracted handcrafted and DL features from MRI scans to train different ML methods. The best stratification performance was obtained when using DL features extracted by a pretrained AlexNet [29], and trained with a linear discriminant. In their work, tumor regions were extracted from specific axial slices before being fed to AlexNet.

From the abovementioned studies, it can be noticed that there are at least four common issues associated with the existing CNN architectures developed for OS prediction of glioma patients. Primarily, many existing CNN architectures are 3D CNNs which do not consider the insufficient training available datasets. For example, the well-known BraTS 2017–2018 datasets contain only 163 training samples. These numbers are far from enough to reach the full potential of 3D CNNs. Transfer learning, a DL technique that enables the use of a model developed for one task using large-scale image datasets, can be reused as a starting point for a model on another task in datasets with limited size. However, large annotated 3D medical image dataset does not exist, and current transfer learning methods are tailored to 2D images [30]. When volumetric dataset sizes are small, multiview CNN (Mv-CNNs) are designed and trained for 3D shape detection from a collection of their rendered views on 2D images with greater accuracy compared to 3D CNN methods [31]. They can also be customized and used for multislice 2D projection of 3D medical images in order to overcome the above mentioned limitations [32]. However, to the best of our knowledge, no work has been done for OS prediction using Mv-CNNs.

Secondly, without considering the prognostic features on slices in coronal and sagittal projections, most of the existing 2D CNNs are based on slices in the axial projections only. Thirdly, most of the existing CNN networks are also based on segmented tumor regions. However, the information around the tumor region may contain tumor cell infiltrates that contribute to poor prognosis [33]. In addition, tumor segmentation is yet a challenging task, and requires more training times, hindering their application for time-constrained large-scale medical imaging tasks, like preoperative OS prediction. The fourth common issue is that, most existing approaches make predictions based on single classifiers. However, single learners sometimes present poor results for small and imbalanced datasets. Moreover, previous studies in other medical image analysis problems have shown that ensemble learning of multiple CNN models may yield better stratification accuracy than individual CNN models [34–36]. Therefore, it will also be worthwhile to develop OS prediction models based on ensemble learning of multiple CNN algorithms .

To overcome the above described shortcomings, we present a novel framework to discriminate between long and short OS of HGG patients using preoperative multimodal MRI images via a combination of Mv-CNNs and standard ML techniques (Figure 1). Specifically, first we reconstructed all the multimodal 3D MRI scans as a set of 2D projections along all the three planes. We then developed novel Mv-CNN architectures that can perform end-to-end binary classification on 2D projected images. Finally, the probability predictions of the Mv-CNN models were fused using six ML techniques [multilayer perceptron (MLP), logistic regression, support vector machines (SVM), Bayes Net (BN), random forest (RF), and random tree (RT)]. The individual classification performance of each modality and 2D projections were also compared. To the best of our knowledge, this

method is the first work that attempts to differentiate between long and short OS of patients with HGG using an ensemble of multiple CNN models.

The rest of this article has the following structure. Section 2 describes the materials and methods. Experimental results and discussions are reported in Section 3. Finally, Section 4 points out future works and concludes this paper.

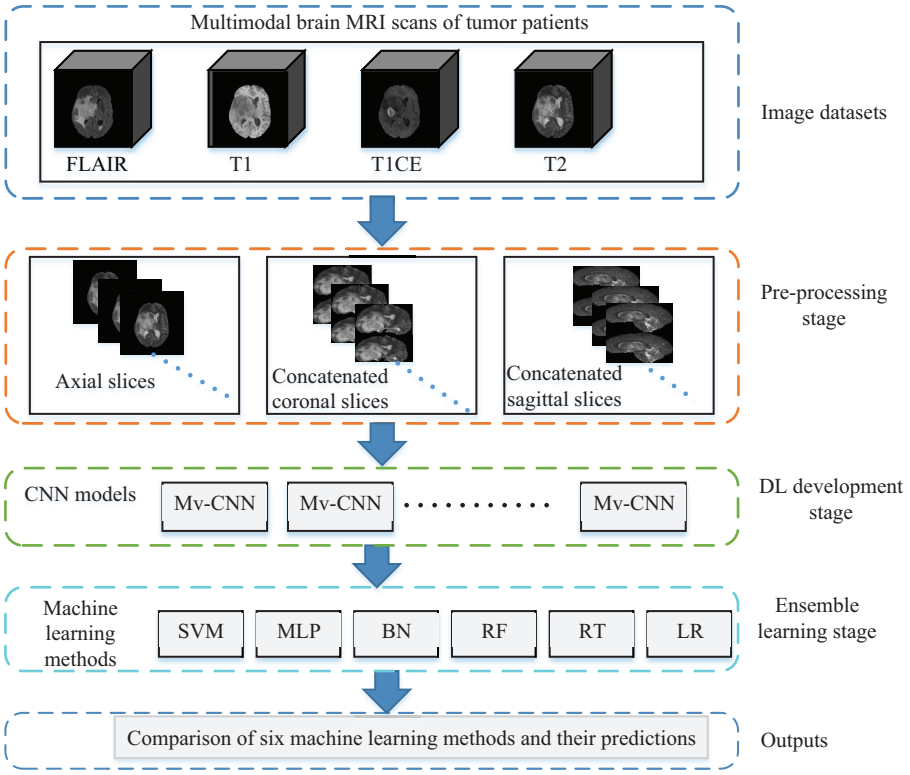

Figure 1. The schematic description of the proposed deep learning based ensemble method for survival time classification, which is achieved in three main stages. (i) Preprocessing. In this stage, 2D images were sliced from volumetric MRIs in the axial, coronal and sagittal planes. Consequently, two consecutive coronal, and two consecutive sagittal slices were concatenated before being fed to the CNN models. (ii) Deep learning (DL) development. In this stage, multiview CNN (Mv-CNN) models that map multiple 2D slices to a prediction probability of the survival time were developed. (iii) Finally, in the ensemble learning stage, the predictions by the Mv-CNN models were fused using the six standard machine learning techniques to obtain a better classification accuracy.

2. Materials and methods

As illustrated in Figure 1, the proposed framework for OS classification consists of three core modules: (1) "preprocessing", (2) "DL development", and (3) "ensemble learning". During preprocessing, we reconstructed the 3D MRI scans as a stack of of 2D images along all the three planes. The second module is the primary building of our prediction system where we designed Mv-CNN architectures that take preprocessed 2D images of volumetric MRI as input and outputs binary classification prediction. We proposed three different variants of Mv-CNN architectures. In the third module, the probability predictions by the Mv-CNN models were fused using various standard ML methods to obtain the final prediction result. Details of the proposed methods are presented in the following sections.

2.1. Dataset

The multimodal preoperative volumetric MRI scans, and OS in days for 163 HGG patients obtained from BraTS'17 public datasets were used to train and evaluate our proposed models [37, 38]. For each patient, four MR imaging modalities were provided: FLAIR, T1CE, T1, and T2. The images size in the dataset was $240 \times 240 \times 155$. OS, defined in days, is calculated from the date of initial diagnosis until tumor-related death. From radiologist reports, patients were categorized into three levels of OS: long-survivors (OS $>$ 15 months), mid-survivors (10 months $<$ OS $<$ 15 months) and short-survivors (OS $<$ 10 months).The prognosis of patients with HGG, often measured by the OS, are generally categorized into two (long vs. short) or three (long, medium and short) groups of survival in order to provide valuable insights in advancing surgical and treatment plan. However, in order to simplify the analysis and due to the availability of limited training samples, the classification task in this paper was to discriminate between short-survival and long-survival (mid-survival or long-survival). We split the dataset into training and validation, where 122 multimodal MRI scans for training, and the remaining 41 were used for validation. We used stratified random sampling for balancing OS distribution between the two groups. Detailed patients' information can be found in Table 1.

Table 1. Demographic and clinical characteristics of patients in the training and validation dataset. The entire dataset was split into two groups: 75% for training, and 25% for validation. No significant differences were observed for demographics and clinical characteristics between the training and validation sets.

Parameters		Training dataset	Validation dataset	Overall dataset
No. of patients		122	41	163
Age (year)	Range	$18.975 - 85.762$	$30.408 - 84.844$	$18.975 - 85.762$
	Mean	60.181	60.774	60.334
	Std. dev.	12.379	11.319	12.083
OS (day)	Range	$5 - 1767$	$30 - 1731$	$5 - 1767$
	Mean	424.645	417.881	422.963
	Std. dev.	346.214	363.787	349.684
	Long(%)	72(59.5)	26(61.9)	98
	Short(%)	49(40.5)	16(38.1)	65

2.2. Preprocessing

One of the key ideas of the approaches proposed in this study was converting the 3D MRI scans into 2D stacked slices in the axial, coronal and sagittal planes. As a result, for every patient, twelve 2D stacked MRI slices were generated from the four MRI modalities defined in Section 2.1. Hereafter, we indicate these twelve 2D stacked slices as FLAIR-axial, FLAIR-coronal and FLAIR-sagittal for axial, coronal and sagittal projection of FLAIR MRI modality, respectively. Similarly, T1-axial, T1-coronal and T1-sagital, T1CE-axial, T1CE-coronal and T1CE-sagittal, and T2-axial, T2-coronal and T2-sagittal for axial, coronal, and sagittal views of T1, T1CE and T2 modalities, respectively. Some beginning and end slices of each MRI scans which do not contain any brain tissue were excluded to avoid processing the background and manage the GPU memory constraint. The intensity pixel of each slice was also rescaled in the range 0 to 255, converted to PNG format, and normalized to have zero mean and unit variance.

The $3D-to-2D$ reconstruction of MRI scans lead each 2D axial, coronal and sagittal projected images to have a shape size of 240×240, 155×240 and 155×240 pixels, respectively. These projected images

were then cropped centrally and resized to 224 × 224, 112 × 224 and 112 × 224 pixel sizes, respectively. Subsequently, each two consecutive coronal slices were concatenated and reshaped to 224 × 224. Similarly, each two consecutive sagittal slices were concatenated and reshaped to 224 × 224 pixel sizes. These allowed us to use pretrained DL models developed for natural images. The concatenation method that we used reduced the number of sagittal and coronal slices of each MRI scan by half. Hence, it reduces the GPU memory requirement and processing time. In addition, we also argue that this concatenation approach enables Mv-CNN models to benefit from the fine-grained details in each sagittal and coronal slices than enlarging them, which usually results in loss of information. Similar concatenation approach was used in [39] where each coronal and sagittal slices at the same position concatenated before fed to the DL model, and achieved promising results in tuberculosis diagnosis. All the preprocessing steps were done using the python programming language and NiBabel package[1]. The total number of axial, coronal and sagittal 2D projected images of the four MRI modalities of the 163 patients after preprocessing is shown in Table 2. The example of reconstructed, cropped and concatenated 2D projected images of a sample patient is shown in Figure 2.

Table 2. Total number of reconstructed and implemented 2D slices from each modality and projection.

Projections	Modalities				
	T2	T1Ce	T1	FLAIR	Total
Axial	22,820	25,265	14,670	22,005	84,760
Sagittal	22,820	26,080	19,560	21,190	89,650
Coronal	29,340	22,820	19,560	24,450	96,170
Total	74,980	74,165	53,790	67,645	270,580

2.3. Architecture and ensemble learning of the proposed CNN models

Standard CNN architectures usually consists of convolutional and pooling layers occurring in alternative fashion to extract higher level discriminative features to represent the original 2D input images followed by dense layers to perform classification [40]. Unlike that of the standard CNN architectures, in the Mv-CNN architectures, multiple multiview 2D images of a 3D object can be provided as a single training example. Hence, through customizing the original Mv-CNN architecture developed for 3D object recognition, in this paper we proposed three CNN architectures: a single column Mv-CNN and two multicolumn Mv-CNN architectures.

2.3.1. Single column Mv-CNN

The single column Mv-CNN architecture that we proposed, inspired by [32], consists of four core parts as shown in Figure 3. (i) The convolutional base of pretrained AlexNet for extracting features, which takes k × 3 × 224 × 224 stacked png images as inputs and outputs k × 256 × 6 × 6 features sizes. 3 indicates the number color channels, and k is the number slices of a 3D MRI scan in one of the three planes obtained after the preprocessing stage. (ii) Global average pooling on top of the convolutional base of AlexNet applied across the spatial dimensions to reduce features to k × 256. (iii) Max pooling layer on top of global average pooling applied across slices to reduce features to 256-dimensional vector, and (iv) Dense layers with sigmoid activation function to map from the computed hidden representation to the output probability prediction. We

[1]Brett M, Hanke M, Marc-Alexandre CÔ, McCarthy P, Cheng C et al. (2018). NiBabel-Access a cacophony of neuro-imaging file formats [online]. Website https://nipy.org/nibabel/ [06.04.2019].

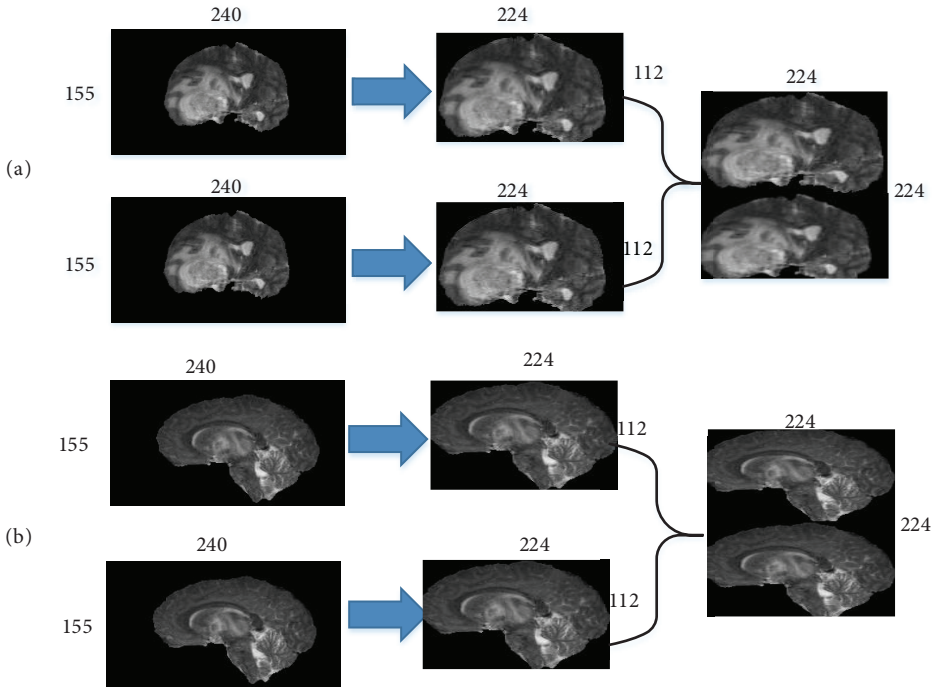

Figure 2. Example of the concatenation procedures performed during preprocessing stage: (a) two consecutive coronal slices having the same shape size (155×240) were sliced from the 3D MRI scans, center cropped and then concatenated, (b) two consecutive sagittal slices having the same shape size (155×240) were sliced from 3D MRI scans, center cropped and then concatenated. The proposed AlexNet-based multiview CNN models used these concatenated images as input to improve the robustness of training and prediction.

used the backpropagation algorithm for training. The binary cross entropy loss function along with adaptive moment estimation (Adam) optimizer were used for optimizing the models [41]. The models were implemented using Python programming language and PyTorch framework [42].

The single column Mv-CNN architecture were trained twelve times using each preprocessed 2D images of the four modalities projected in the three planes. With such trained twelve models, patient-level twelve predicted probabilities were generated. Then, we ensemble these twelve predicted probabilities using different ML methods. The idea is that a more accurate stratification of OS patients with glioma may be obtained when the individual learners trained separately on 2D projected images of multimodal MRI scans are combined together.

2.3.2. Multicolumn Mv-CNN

The proposed multicolumn Mv-CNN architectures can combine information from multiple 2D-projection images within a single CNN architecture (Figure 4). Two multicolumn Mv-CNN architectures named Mc-Mv-CNN-1 and Mc-Mv-CNN-2 were proposed to (i) investigate performances of each modality and 2D projection images, (ii) investigate the learning capability of variant Mv-CNNs, and (iii) verify the effectiveness of the proposed ensemble learning approach on the evaluation of survival classification.

Mc-Mv-CNN-1: The Mc-Mv-CNN-1 architecture comprises three-column subnetworks. Each subnetwork adopts the single column Mv-CNN architecture after discarding the dense layers. Reconstructed and preprocessed axial, coronal and sagittal slices of the 3D MRI are fed into the three parallel Mv-CNN archi-

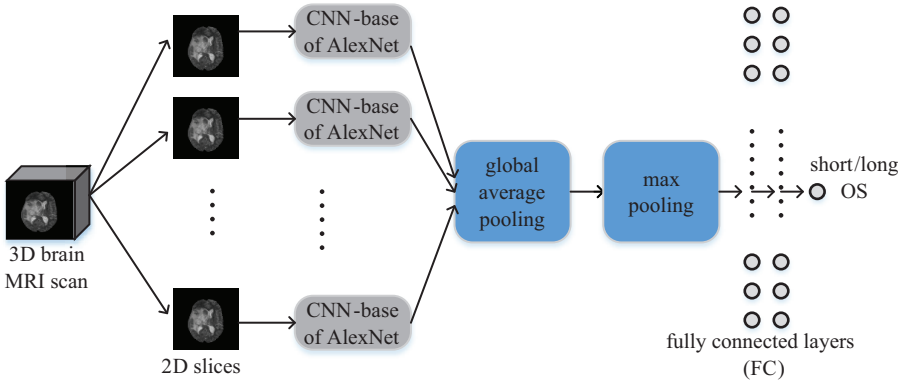

Figure 3. The proposed survival prediction pipeline using the single column Mv-CNN architecture. The 3D brain MRI scans of tumor patients were first decomposed into a stack of 2D slices. Each 2D slice was then passed to the CNN base of AlexNet. The architecture's input was $k \times 3 \times 224 \times 224$, where k is the number of 2D stacked slices obtained after the preprocessing stage, and 3 indicates the number of color channels. Global average and max pooling layers were applied to reduce generated feature maps from the slices and increase compactness of the models. Finally, FC were used on the top of max pooling layer to classify the patient to the short or long survival group.

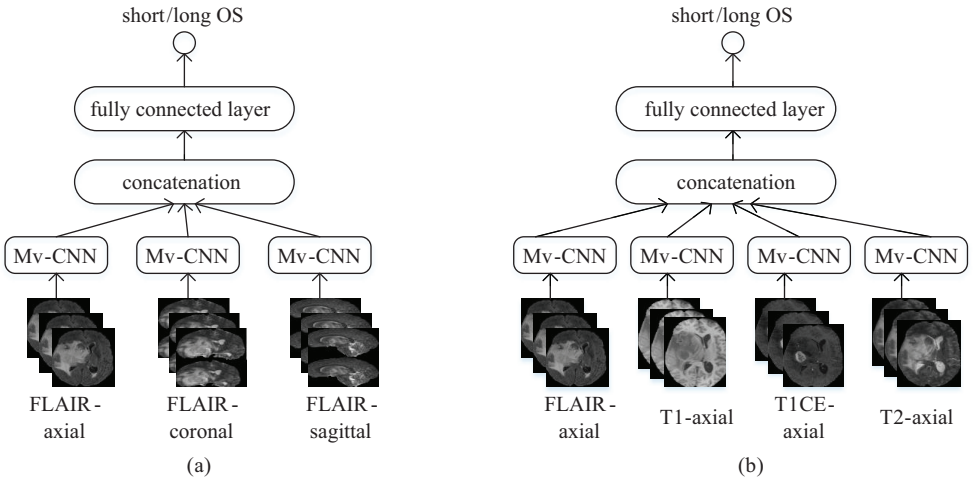

Figure 4. Survival prediction pipelines using the proposed multicolumn Mv-CNN architectures. (a) A three-column Mv-CNN architecture (Mc-Mv-CNN-1) operating on the three 2D projection images of each modality. The figure shows an example for the FLAIR modality. The same architecture were applied for the other three modalities, resulting a patient-level four different prediction probabilities of patients' survival time. (b) A four-column Mv-CNN architecture (Mc-Mv-CNN-2) operating on the 2D images of the four modalities projected in the same plane. The figure shows an example for the axial projection images. The same architecture were applied for coronal and sagittal projections, resulting a patient-level three different prediction probabilities of patients' survival time.

tectures, respectively. Each one of the three columns outputs a $256-$dimension hidden representation for the 2D projection images. A fusion layer is then added to integrate the outputs of the three-column vectors into a $768-$dimension vector through concatenation. The model architecture also further incorporated a fully connected layer and a sigmoid activation function to map the computed hidden representations to the OS prediction in the 0 to 1 range. We trained the architecture four times, one using each modality. The architecture depicted in Figure 4a shows an example for FLAIR modality. The same architecture and training were used for the other three MRI modalities, resulting in four different binary probability estimates for each patient's survival time.

Mc-Mv-CNN-2: It is a four-column Mv-CNN architecture operating on the same 2D projection images of the four MRI modalities. Each of the four columns outputs the same $256-$dimension hidden representation for each 2D projection images of the four modalities. Each one of the four columns outputs a $256-$dimension hidden representation for the 2D projection images. A fusion layer is added to integrate the outputs of the four-column vectors into a $1024-$dimension vector through concatenation. The model architecture also further incorporated a fully connected layers and a sigmoid activation function to map the computed hidden representations to the OS prediction in the 0 to 1 range. We trained the architecture three times, one for each projection of the four modalities. The architecture depicted in Figure 4b shows an example for axial projection. The same architecture and training were used for the other two projection, resulting in three different binary probability estimates for each patient's survival time. In the end, we fused predicted probabilities by Mc-Mv-CNN-1 and Mc-Mv-CNN-2 models using the same six different ML methods used during fusing the probability predictions by the single-column Mv-CNN models.

3. Results and discussions

3.1. Performance of the single column Mv-CNN models

The classification performances of each of the twelve single column Mv-CNN models have been evaluated in terms of accuracy and area under the receiver operating characteristics curve (AUC). As shown in table 3 left-side, it is observed that half of the twelve models obtained an AUC value greater than 0.8 on the validation dataset. These demonstrate that the proposed single column Mv-CNN models are able to extract discriminative features from MRI images that can distinguish between long and short OS. In addition, single column Mv-CNN model trained solely with 2D axial projection of FLAIR modality has a better classification performance than the other eleven models trained with the other 2D projection images. In addition, single column Mv-CNN models trained solely with 2D axial projections of T1 and T2 modalities achieved better AUC than models trained solely with coronal and sagittal slices of the corresponding modalities. However, the single column Mv-CNN model trained on T1CE$-$ sagittal projections achieved better AUC performances than the corresponding axial and coronal projections. In summary, the results indicated that 2D axial projections of MRI modalities have more prognostic features compared to the other 2D projections of the corresponding modalities.

3.2. Performance of the ensemble single column Mv-CNN models using the six fusion strategies

With regard to AUC and accuracy, the OS classification results obtained when the individual single column Mv-CNN learners fused using the six ML methods are depicted in Table 3 right-side. The results show that most of the ensemble models yielded quite an impressive stratification performances, ranging from 0.736 to 0.93 and 76.19% to 92.9%, in AUC and accuracy, respectively. All the ensemble models stratification performance, with the exception of RT, achieved greater performance than the individual single column Mv-CNN models, with more than 7.5%, and 4.8% improvement in accuracy and AUC measurements, respectively. Furthermore, SVM presents the highest OS prediction performance achieving 0.93 and 92.9% in AUC and accuracy, respectively. These results demonstrate that the developed system combining individual single column Mv-CNN models into a single framework using classical ML methods could be considered as a promising strategy for OS stratification, and SVM is the best fusion method to ensemble multiple DL models for OS stratification.

3.3. Performance of the multicolumn Mv-CNN models

The OS stratification performances of each of the seven multicolumn Mv-CNN models have been evaluated in terms of accuracy and AUC as shown in Table 4 left-side. Their AUC and accuracy range from 0.68 to 0.86

Table 3. Single column Mv-CNN models performance results on the validation dataset. Left-side: performances of the individual single column Mv-CNN models; right-side: performances of the ensemble twelve single column Mv-CNN models when fused using the six ML techniques. Bold values indicate our best results among 2D projections of each modality, and ensemble models.

Models	AUC	Accuracy(%)
FLAIR-axial	**0.875**	**83**
FLAIR-coronal	0.84	81
FLAIR-sagittal	0.84	76
T1-axial	**0.83**	64
T1-coronal	0.76	67
T1-sagittal	0.82	**76**
T1CE-axial	0.77	67
T1CE-coronal	0.75	67
T1CE-sagittal	**0.83**	**74**
T2-axial	**0.70**	**62**
T2-coronal	0.68	57
T2-sagittal	0.69	60

ML methods	AUC	Accuracy(%)
MLP	0.923	90.5
BayesNet	0.908	85.7
SVM	**0.93**	**92.9**
RF	0.909	85.7
RT	0.736	76.19
Logistic regression	0.882	88.1

and 61.9% to 83.3%, respectively. It is observed that when multicolumn Mv-CNN models trained on solo MRI modalities, the best stratification performance was obtained using FLAIR modality with 0.86 and 78.6% in AUC and accuracy, respectively. This indicates that FLAIR MRI modality has more prognostic values compared to the other three MRI modalities. Similarly, multicolumn Mv-CNN model trained with 2D axial projections has a reasonable better classification performance with 0.82 and 83.3% in AUC and accuracy, respectively.

3.4. Performance of the ensemble multicolumn Mv-CNN models using the six fusion strategies

With regard to AUC and accuracy, the OS classification results achieved when the individual multicolumn Mv-CNN models combined using the six ML techniques are illustrated in Table 4 right-side. It is observed that most of the ensemble models obtained quite an impressive performance, ranging from 0.81 to 0.873 and 73.8% to 85.7%, in AUC and accuracy, respectively. In addition, ensemble models using RF and RT fusion methods achieved the highest AUC (0.87) and accuracy (85.7%) than the other fusion methods. These results demonstrate that, performances combined from multimodal MRI scans were better than any single MRI modality with improvement rates of more than 2.4% in accuracy. However, when best results of the ensemble of single column Mv-CNN and multicolumn Mv-CNN models were compared, ensemble of the single column Mv-CNN models achieved better with an increment of 2.4% in accuracy. In summary, the developed ensemble learning of multiple multiview CNN models approach achieved state-of-the-art results, and could also be considered as a promising strategy for any other medical classification problems with limited volumetric medical image datasets.

3.5. Result comparison with some existing works

A comparison of our proposed approach with ten recently published works with state-of-the-art results for brain tumor patients survival time classification is discussed in this section (see also Table 5). To further validate the effectiveness of the proposed approach, we also presented a reference baseline based on limited demographic and

Table 4. Multicolumn Mv-CNN models performance results on the validation dataset. Left-side: performances of the seven multicolumn Mv-CNN models; right-side: performances of the multicolumn Mv-CNN models when fused using the six ML methods. Bold values indicate our best results among modalities, 2D projections and ensemble models.

Models	AUC	Accuracy(%)
FLAIR	**0.86**	**78.6**
T1	0.76	71.4
T1CE	0.68	66.7
T2	0.71	61.9
Axial	**0.82**	**83.3**
Coronal	0.78	71.4
Sagittal	0.69	61.9

ML methods	AUC	Accuracy(%)
MLP	0.846	78.6
BayesNet	0.861	73.8
SVM	0.81	81
RF	**0.873**	81
RT	0.837	**85.7**
Logistic regression	0.827	81

tumor volume related features, and trained with SVM. As shown in Table 5, the baseline achieved a performance of 0.65 and 60% in AUC and accuracy, respectively.

Deep learning algorithms works well when large annotated datasets are available. However, handcrafted based approaches have been superior than deep learning algorithms in datasets with limited sizes. Hence, the majority of state-of-the-art results for OS stratification of brain tumor patients have been obtained based on handcrafted features trained with traditional ML classifiers. For instance, in [19], authors extracted 74 features from automatically segmented tumor regions and subsequently combined with age of patients to train XGBoost classifier. Using the BraTS training dataset, they obtained an accuracy of 73% in stratifying survival into three groups. Using the same dataset, Chato et al. [43] achieved an accuracy of 66.7% using histogram features and SVM classifier, and Sanghani [44] achieved a 3-class classification accuracy of 87% on stratified 5-fold cross-validation using clinical, volumetric, tumor shape, and texture features trained with ML algorithm. Similarily, Chaddad et al. [45] used JIM, GLCM and gene expression texture features extracted from T1CE and FLAIR MRI modalities to train RF classifier. Their method achieved a leave-one-out cross-validation AUC of 0.78. In another work [46], using clinical information (age and sex) and more than 60 features derived from computer-based segmented tumor regions, authors achieved an overall accuracy of 80% in classifying patients into short, medium and long survival.

Chen [47] showed that extracted handcrafted features, such as intensity, shape, texture and wavelet, from manually delineated tumor regions in presurgical axial T1Ce modality, and subsequently combined with clinical data allowed stratifying patients' survival into a low- or high-risk group with an AUC of 0.851. In [12] using SVM classifier and 2D texture features extracted from slices with the largest tumor size that are manually segmented by two experienced radiologists, authors compared the performance of four MRI modalities when used individually and in combination for classifying survival into two groups. The result showed that when

Table 5. Comparison of the classification accuracy and AUC of the proposed ensemble learning approach and the existing models for survival time classification of brain tumor patients based on MRI data.

Experiments	Years	MRI modalities	Survival stratification	Methods	Accuracy	AUC
Upadhaya et al. [48]	2015	T1, T1Ce, T2, FLAIR	short (<14.8 months) vs. high (otherwise)	Handcrafted	90	-
Macyszyn et al. [46]	2016	T1, T1CE, T2, FLAIR, DTI, DSC-MRI	short (<6 months) vs. medium (6−18 months) and high (>18 months)	Handcrafted	80	-
Chato et al. [21]	2017	T1, T1Ce, T2, FLAIR	short (<18 months) vs. high (otherwise)	Deep learning	90	-
Chato et al. [43]	2018	T1, T1CE, T2, FLAIR	short (<10 months) vs. medium (10−15 months) and high (>15 months)	Handcrafted	66.7	-
Liu et al. [12]	2018	T1, T1CE, T2, FLAIR	short (<12 months) vs. high (otherwise)	Handcrafted	80.7	0.79
Sanghani et al. [44]	2018	T1, T1CE, T2, FLAIR	short (<10 months) vs. medium (10−15 months) and high (>15 months)	Handcrafted	87.1	-
Z.A. Shboul [19]	2019	T1,T1CE, T2, FLAIR	short (<10 months) vs. medium (10−15 months) and high (>15 months)	Handcrafted	73	-
Chen et al. [47]	2019	T1CE	short (<12 months)vs. high (otherwise)	Handcrafted	-	0.85
Chaddad et al. [45]	2019	T1CE and FLAIR	short (<12 months) vs. high (otherwise)	Handcrafted	-	0.78
Nie et al. [28]	2019	T1CE,DTI, rs-fMRI	short (<22 months) vs. high (otherwise)	Hybrid	90.7	-
Baseline	2020	T1, T1CE, T2, FLAIR	short (<10 months) vs. high (otherwise)	Handcrafted	60	0.65
This work	2020	T1, T1CE, T2, FLAIR	short (<10 months) vs. high (otherwise)	Deep learning	**92.9**	**0.93**

using only T1Ce and the four MRI modalities separately, both models achieved nearly equal accuracy and AUC value of 80.7%, and 0.79, respectively. Likewise, the authors [48] achieved a classification accuracy of 90% based on heterogeneity textural features extracted from T1Ce and T1 MRI modalities. However, considering only one MRI modality at a time, the best classification accuracy (82.5%) was obtained using T1Ce, followed by FLAIR. In addition, the experimental result indicated that, considering only a single MRI modality at a time, the survival classification performance using only the T1Ce modality is comparable to that of using all the four MRI modalities together. However, in consistent with other study [49], our proposed approach shows that the performances combined from multimodal MRI were superior than using T1Ce alone. In addition, our experimental results show that considering one MRI sequence at a time, the best stratification performance was obtained using FLAIR MRI modality. This indicates that further investigation is needed on the individual and multimodal MRI scans comparisons for glioma patients' survival time prediction.

Nie et al. [28] proposed a multichannel 3D CNN architecture that extract deep features from multimodal MRI scans, and subsequently combined deep features, demographic and tumor related features to train SVM that classified OS into two groups with an accuracy of 90.7%, and 90.5% on 3-fold and, 10-fold cross-validation, respectively. The paper also justified the importance of using multimodal MRI in predicting OS, and rs-fMRI resulted in the best classification performance among all the single modalities. In another study [21], using pretrained AlexNet for extracting deep features from specific slices that clearly contained the three regions of HGG glioma, a 91% accuracy was obtained for the linear discriminant classifier, and 86.4% for the linear SVM classifier. In summary, the performance of our proposed ensemble learning approach regarding classification accuracy and AUC is superior to several state-of-the-art results in brain tumor patients' OS stratification.

4. Conclusion

In this paper, we have introduced an ensemble learning framework for preoperative survival time classification of brain tumor patients based on fusing multiple deep learning models with different machine learning techniques. Our best model was obtained based on an ensemble of twelve multiview CNN models that were trained using 2D projection images of multimodal MRI images and fused using the SVM machine learning classifier. We achieved state-of-the-art results with AUC of 0.93 on the validation dataset. We only had limited number of training and validation samples to train and validate the classification models. However, we believe more training data will further improve the performance of our proposed approaches. In the future, we would like to increase the number of training samples and explore different training strategies for deep learning models. Moreover, MRI scans may only explain patients' survival outcome partially. In the future, we would like to modify our proposed network architectures to incorporate not only the MRI images but also clinical information of patients.

Acknowledgment

This work was supported by the research fund of the Çukurova University, Adana, Turkey (project number: 10683).

References

[1] Neervoort FW, Van Ouwerkerk WJR, Folkersma H, Kaspers GJL, Vandertop WP. Surgical morbidity and mortality of pediatric brain tumors: a single center audit. Child's Nervous System 2010; 26 (11): 1583-1592.

[2] He Z, Mitteer RA, Mou Y, Fan Y. Multimodality Targeting of Glioma Cells. In: Brem S, Abdullah KG (editors). Glioblastoma. Philadelphia, PA, USA: Elsevier, 2016, pp. 55–72.

[3] Ostrum QT, Gittleman H, Liao P, Rouse C, Chen Y et al. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2007–2011. Neuro-oncology 2014; 16 (s4): iv1-iv63. doi: 10.1093/neuonc/nou223

[4] Van Meir EG, Hadjipanayis CG, Norden AD, Shu HK, Wen PY et al. Exciting new advances in neuro-oncology: the avenue to a cure for malignant glioma. CA: A Cancer Journal for Clinicians 2010; 60 (3): 166-193.

[5] Ostrom QT, Gittleman H, Truitt G, Boscia A, Kruchko C et al. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2011–2015. Neuro-oncology 2018; 20 (s4): iv1-iv86. doi: 10.1093/neuonc/noy131

[6] Castells X, García-Gómez JM, Navarro A, Acebes JJ, Godino Ó et al. Automated brain tumor biopsy prediction using single-labeling cDNA microarrays-based gene expression profiling. Diagnostic Molecular Pathology 2009; 18 (4): 206-218. doi: 10.1097/PDM.0b013e31818f071b

[7] Christakis NA, Smith JL, Parkes CM, Lamont EB. Extent and determinants of error in doctors' prognoses in terminally ill patients: prospective cohort study Commentary: Why do doctors overestimate? Commentary: Prognoses should be based on proved indices not intuition. The British Medical Journal 2000; 320 (7233): 469-473. doi: 10.1136/bmj.320.7233.469

[8] Peleg R, Biderman A. A memorable patient: the scars of the Jewish holocaust. The British Medical Journal 2000; 320 (7233): 473-473. doi: 10.1136/bmj.320.7233.473

[9] Legaz-Aparicio AG, Verdú-Monedero R, Larrey-Ruiz J, Morales-Sánchez J, López-Mir F et al. Efficient variational approach to multimodal registration of anatomical and functional intra-patient tumorous brain data. International Journal of Neural Systems 2017; 27 (06): 1750014. doi: 10.1142/S0129065717500149

[10] Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology 2016; 278 (2): 563-577. doi: 10.1148/radiol.2015151169

[11] Shukla G, Alexander GS, Bakas S, Nikam R, Talekar K et al. Advanced magnetic resonance imaging in glioblastoma: a review. Chinese Clinical Oncology 2017; 6 (4): 40. doi: 10.21037/cco.2017.06.28

[12] Liu Y, Zhang X, Feng N, Yin L, He Y et al. The effect of glioblastoma heterogeneity on survival stratification: a multimodal MR imaging texture analysis. Acta Radiologica 2018; 59 (10): 1239-1246. doi: 10.1177/0284185118756951

[13] Zhang L, Wang H, Li Q, Zhao MH, Zhan QM. Big data and medical research in China. The British Medical Journal 2018; 360: j5910. doi: 10.1136/bmj.j5910

[14] Wang K, Wang Y, Fan X, Wang J, Li G et al. Radiological features combined with IDH1 status for predicting the survival outcome of glioblastoma patients. Neuro-oncology 2015; 18 (4): 589-597. doi: 10.1093/neuonc/nov239

[15] Du W, Zhang M, Ying W, Perc M, Tang K et al. The networked evolutionary algorithm: a network science perspective. Applied Mathematics and Computation 2018; 338: 33-43. doi: 10.1016/j.amc.2018.06.002

[16] Erkaymaz O, Ozer M, Perc M. Performance of small-world feedforward neural networks for the diagnosis of diabetes. Applied Mathematics and Computation 2017; 311: 22-28. doi: 10.1016/j.amc.2017.05.010

[17] Bakas S, Reyes M, Jakab A, Bauer S, Rempfler M et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. arXiv 2018; arXiv:1811.02629 [cs.CV].

[18] Li Q, Bai H, Chen Y, Sun Q, Liu L et al. A fully-automatic multiparametric radiomics model: towards reproducible and prognostic imaging signature for prediction of overall survival in glioblastoma multiforme. Scientific Reports 2017; 7 (1): 1-9. doi: 10.1038/s41598-017-14753-7

[19] Shboul ZA, Alam M, Vidyaratne L, Pei L, Elbakary MI et al. Feature-guided deep radiomics for glioblastoma patient survival prediction. Frontiers in Neuroscience 2019; 1: 13. doi: 10.3389/fnins.2019.00966

[20] Li Y, Shen L. Deep learning based multimodal brain tumor diagnosis. In: Crimi A, Bakas S, Kuijf H, Menze B, Reyes M (editors). Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2017, Lecture Notes in Computer Science. Cham, Switzerland: Springer, 2018, pp. 1-20.

[21] Chato L, Latifi S. Machine learning and deep learning techniques to predict overall survival of brain tumor patients using MRI images. In: 2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE); Washington, DC, USA; 2017. pp. 9-14.

[22] Lao J, Chen Y, Li ZC, Li Q, Zhang J et al. A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. Scientific Reports 2017; 7 (1): 1-8. doi: 10.1038/s41598-017-10649-8

[23] Sun L, Zhang S, Chen H, Luo L. Brain Tumor Segmentation and Survival Prediction Using Multimodal MRI Scans with Deep Learning. Frontiers in Neuroscience 2019; 13: 810. doi: 10.3389/fnins.2019.00810

[24] Thickstun J, Harchaoui Z, Foster DP, Kakade SM. Invariances and data augmentation for supervised music transcription. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); Calgary, AB, Canada; 2018. pp. 2241-2245.

[25] CireşAn D, Meier U, Masci J, Schmidhuber J. Multi-column deep neural network for traffic sign classification. Neural Networks 2012; 32: 333-338. doi: 10.1016/j.neunet.2012.02.023

[26] Li R, Zhang W, Suk HI, Wang L, Li J et al. Deep learning based imaging data completion for improved brain disease diagnosis. In: Golland P, Hata N, Barillot C, Hornegger J, Howe R (editors). Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014. Lecture Notes in Computer Science, Vol. 8675. Cham, Switzerland: Springer, 2014, pp. 305-312.

[27] Perc M, Ozer M, Hojnik J. Social and juristic challenges of artificial intelligence. Palgrave Communications 2019; 5 (1): 1-7. doi: 10.1057/s41599-019-0278-x

[28] Nie D, Lu J, Zhang H, Adeli E, Wang J et al. Multi-channel 3D deep feature learning for survival time prediction of brain tumor patients using multi-modal neuroimages. Scientific Reports 2019; 9 (1): 1-4.

[29] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Communications of the ACM 2017; 60 (6): 84-90.

[30] Gupta V, Demirer M, Bigelow M, Little KJ, Candemir S et al. Performance of a deep neural network algorithm based on a small medical image dataset: incremental impact of 3D-to-2D reformation combined with novel data augmentation, photometric conversion, or transfer learning. Journal of Digital Imaging 2020; 33 (2): 431-438.

[31] Su H, Maji S, Kalogerakis E, Learned-Miller E. Multi-view convolutional neural networks for 3d shape recognition. In: Proceedings of the IEEE international Conference on Computer Vision; Santiago, Chile; 2015. pp. 945-953.

[32] Bien N, Rajpurkar P, Ball RL, Irvin J, Park A et al. Deep-learning-assisted diagnosis for knee magnetic resonance imaging: development and retrospective validation of MRNet. PLoS Medicine 2018; 15 (11): e1002699.

[33] Choi Y, Ahn KJ, Nam Y, Jang J, Shin NY et al. Analysis of heterogeneity of peritumoral T2 hyperintensity in patients with pretreatment glioblastoma: prognostic value of MRI-based radiomics. European Journal of Radiology 2019; 120: 108642.

[34] Winzeck S, Mocking SJ, Bezerra R, Bouts MJ, McIntosh EC et al. Ensemble of convolutional neural networks improves automated segmentation of acute ischemic lesions using multiparametric diffusion-weighted MRI. American Journal of Neuroradiology 2019; 40 (6): 938-945. doi: 10.3174/ajnr.A6077

[35] Zhang B, Qi S, Monkam P, Li C, Yang F et al. Ensemble learners of multiple deep CNNs for pulmonary nodules classification using CT images. IEEE Access 2019; 7: 110358-110371. doi: 10.1109/ACCESS.2019.2933670

[36] Xiao Y, Wu J, Lin Z, Zhao X. A deep learning-based multi-model ensemble method for cancer prediction. Computer Methods and Programs in Biomedicine 2018; 153: 1-9. doi: 10.1016/j.cmpb.2017.09.005

[37] Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K et al. The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Transactions on Medical Imaging 2014; 34 (10): 1993-2024. doi: 10.1109/TMI.2014.2377694

[38] Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M et al. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Scientific Data 2017; 4: 170117. doi: 10.1038/sdata.2017.117

[39] Mossa AA, Yibre AM, Çevik U. Multi-view CNN with MLP for diagnosing tuberculosis patients using CT scans and clinically relevant metadata. CLEF Working Notes 2019; 2380: 9-12.

[40] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE 1998; 86 (11): 2278-2324.

[41] Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv 2014; arXiv:1412.6980 [cs.LG].

[42] Paszke A, Gross S, Chintala S, Chanan G, Yang E et al. Automatic differentiation in PyTorch. In: 31st Conference on Neural Information Processing Systems; Long Beach, CA, US; 2017. pp. 1-20.

[43] Chato L, Chow E, Latifi S. Wavelet transform to improve accuracy of a prediction model for overall survival time of brain tumor patients based on MRI images. In: 2018 IEEE International Conference on Healthcare Informatics (ICHI); New York, NY, USA; 2018. pp. 441-442.

[44] Sanghani P, Ang BT, King NK, Ren H. Overall survival prediction in glioblastoma multiforme patients from volumetric, shape and texture features using machine learning. Surgical Oncology 2018; 27 (4) :709-714. doi: 10.1016/j.suronc.2018.09.002

[45] Chaddad A, Daniel P, Desrosiers C, Toews M, Abdulkarim B. Novel radiomic features based on joint intensity matrices for predicting glioblastoma patient survival time. IEEE journal of Biomedical and Health Informatics 2018; 23 (2): 795-804. doi: 10.1109/JBHI.2018.2825027

[46] Macyszyn L, Akbari H, Pisapia JM, Da X, Attiah M et al. Imaging patterns predict patient survival and molecular subtype in glioblastoma via machine learning techniques. Neuro-oncology 2015; 18 (3): 417-425. doi: 10.1093/neuonc/nov127

[47] Chen X, Fang M, Dong D, Liu L, Xu X et al. Development and validation of a MRI-based radiomics prognostic classifier in patients with primary glioblastoma multiforme. Academic Radiology 2019; 26 (10): 1292-1300. doi: 10.1016/j.acra.2018.12.016

[48] Upadhaya T, Morvan Y, Stindel E, Le Reste PJ, Hatt M. A framework for multimodal imaging-based prognostic model building: preliminary study on multimodal MRI in glioblastoma multiforme. Innovation and Research in BioMedical Engineering 2015; 36 (6): 345-350. doi: 10.1016/j.irbm.2015.08.001

[49] Nie D, Zhang H, Adeli E, Liu L, Shen D. (2016) 3D deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients. In: International Conference On Medical Image Computing and Computer-Assisted Intervention; Athens, Greece; 2016. pp. 212-220.