Article

A Novel System for Precise Grading of Glioma

Ahmed Alksas [1], **Mohamed Shehata** [1], **Hala Atef** [2], **Fatma Sherif** [2], **Norah Saleh Alghamdi** [3], **Mohammed Ghazal** [4],
Sherif Abdel Fattah [2], **Lamiaa Galal El-Serougy** [2] and **Ayman El-Baz** [1,*]

1   Bioengineering Department, University of Louisville, Louisville, KY 40292, USA
2   Department of Radiology, Faculty of Medicine, Mansoura University, Mansoura 35516, Egypt
3   Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah Bint
    Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
4   Electrical, Computer, and Biomedical Engineering Department, Abu Dhabi University,
    Abu Dhabi 59911, United Arab Emirates
*   Correspondence: aselba01@louisville.edu

Abstract: Gliomas are the most common type of primary brain tumors and one of the highest causes of mortality worldwide. Accurate grading of gliomas is of immense importance to administer proper treatment plans. In this paper, we develop a comprehensive non-invasive multimodal magnetic resonance (MR)-based computer-aided diagnostic (CAD) system to precisely differentiate between different grades of gliomas (Grades: I, II, III, and IV). A total of 99 patients with gliomas (M = 49, F = 50, age range = 1–79 years) were included after providing their informed consent to participate in this study. The proposed imaging-based glioma grading (GG-CAD) system utilizes three different MR imaging modalities, namely; contrast-enhanced T1-MR, T2-MR known as fluid-attenuated inversion-recovery (FLAIR), and diffusion-weighted (DW-MR) to extract the following imaging features: (i) morphological features based on constructing the histogram of oriented gradients (HOG) and estimating the glioma volume, (ii) first and second orders textural features by constructing histogram, gray-level run length matrix (GLRLM), and gray-level co-occurrence matrix (GLCM), (iii) functional features by estimating voxel-wise apparent diffusion coefficients (ADC) and contrast-enhancement slope. These features are then integrated together and processed using a Gini impurity-based selection approach to find the optimal set of significant features. The reduced significant features are then fed to a multi-layer perceptron artificial neural networks (MLP-ANN) classification model to obtain the final diagnosis of a glioma tumor as Grade I, II, III, or IV. The GG-CAD system was evaluated on the enrolled 99 gliomas (Grade I = 13, Grade II = 22, Grade III = 22, and Grade IV = 42) using a leave-one-subject-out (LOSO) and k-fold stratified (with k = 5 and 10) cross-validation approach. The GG-CAD achieved 0.96 ± 0.02 quadratic-weighted Cohen's kappa and $95.8\% \pm 1.9\%$ overall diagnostic accuracy at LOSO and an outstanding diagnostic performance at $k = 10$ and 5. Alternative classifiers, including RFs and SVM$_{lin}$ produced inferior results compared to the proposed MLP-ANN GG-CAD system. These findings demonstrate the feasibility of the proposed CAD system as a novel tool to objectively characterize gliomas using the comprehensive extracted and selected imaging features. The developed GG-CAD system holds promise to be used as a non-invasive diagnostic tool for Precise Grading of Glioma.

Keywords: GG-CAD; MRIs; HOG; GLCM; GLRLM; ADC

1. Introduction

In 2022, brain and spinal cord primary cancerous tumors will be diagnosed in an estimated 25,050 adults in the United States (14,170 male and 10,880 female). Despite the odds of developing these tumors in one's lifetime being less than one percent, they represent 85% to 90% of all primary central nervous system (CNS) tumors. With a high mortality rate among adults in the USA (18,280 deaths estimated in 2022), cancer of the CNS is the 10th leading cause of death [1]. Glioma is the most common primary malignant tumor of the

CNS in adults, demonstrating distinct characteristics. It represents the highest prevalent cerebral tumor, which takes place in the cerebral glial tissues comprising nearly 30% of brain tumors. Its incidence is about 5–10 per 100,000 people every year with serious morbidity and high mortality rates [2]. Despite improvements in medical treatment targeting glioma-specific molecular pathways, the prognosis remains poor. The median survival is less than 15 months following primary diagnosis, and the five-year survival rate is less than 10% [3–5]. The tendency of malignant gliomas to infiltrate the brain parenchyma in a diffuse manner and the particular tumor microenvironment, that may promote glioma development, contribute to their resistance to conventional surgical therapy and other treatment strategies.

According to the classification of the World Health Organization (WHO) [6,7], gliomas are categorized into four grades based on cellular morphology and malignant tumoral behavior. Grade I gliomas are biologically benign with low risk, while WHO Grade II gliomas are considered low-grade gliomas, with benign tendency, however, they have a considerable recurrence rate. WHO Grade III glioma (anaplastic glioma) and Grade IV glioma (glioblastoma) are considered high grade undifferentiated malignant gliomas, with poor prognosis [2,8]. The most aggressive, glioblastoma, has a survival rate of only 6.8% [9]. Gliomas are graded primarily on the histopathology of tissue obtained through surgical biopsy or resection, which shows the malignancy scale of the tumor. Revised criteria for glioma grading were published with the 2016 revision of the WHO classification, taking into account molecular and genetic information along with the histological features of the tumors, aiming at better prediction of tumor behavior, treatment response, and prognosis [10–12]. Precise tumor grading is of immense importance for guiding neuro-oncologists to the proper decisions for tumor treatment planning and consequently better patient prognosis [13,14]. Despite the fact that the biopsy is the reference standard for identifying the grade of gliomas, it is not favorable because of high invasiveness, expense, and its adverse effects such as bleeding and infection. Therefore, many researchers were motivated to investigate imaging techniques for a non-invasive, early, and precise grading of gliomas for a timely management plan [15–24]. In particular, magnetic resonance imaging (MRI) is the most common imaging modality for the diagnosis and assessment of cerebral neoplasms, including gliomas. Conventional MRI can evaluate the anatomy of the tumor regarding location, morphology, multiplicity, and mass-related effects. Advanced MRI submodalities such as diffusion- and perfusion-weighted imaging, as well as magnetic resonance spectroscopy (MRS), have the ability to provide physiological information of brain tumors. They can provide quantitative metrics of tumor cellularity, vascularization, and metabolism before and after treatment [12,25–27]. Particularly, diffusion-weighted MR (DW-MR) has proven its ability to differentiate gliomas by grade, and in determining the prognosis. In general, higher grades are related to reduced apparent diffusion coefficient (ADC), as water diffusion is substantially decreased in the setting of increased tumor cellularity [14]. Perfusion MR imaging techniques including arterial spin labeling (ASL) assess tumor vascularity associated with neoangiogenesis [27]. MRS gives information about the tumor internal biochemical milieu, by assessment of the concentration of different tumoral cell metabolites. In glial brain tumor grading; N-Acetyl Aspartate (NAA) is a neuronal integrity marker, choline (Cho) a cell membrane breakdown and turnover marker while creatinine (Cr) is a metabolism marker. Increased glioma grade is related to increased aggression of tumor cells that is associated with rising in Cho and reduction in both NAA and Cr levels as well as increased Cho/NAA and Cho/Cr ratios [28].

With the recent advances in artificial intelligence, specifically, machine learning algorithms, and the important role they are playing in the early diagnosis, detection, and prediction of different medical conditions [29,30], they have been incorporated in different studies to help with the clinical diagnosis of gliomas [31]. For example, Zhang et al. [16] explored the abilities of texture analysis along with machine learning on multimodal MR images to differentiate between glioma grades. Their study included a total of 120 subjects with gliomas, 28 low-grade gliomas (LGGs) and 92 high-grade gliomas (HGGs) (Grade I =

3, Grade II = 25, Grade III = 29, and Grade IV = 63). They extracted texture features from the aforementioned data and then performed an oversampling technique to balance their data. In their study, a total of 25 machine learning classifiers were incorporated for classification purposes. The optimal classification model achieved the highest accuracy of 96.1% in grading gliomas (Grades: II, III, and IV) using oversampling, while achieving 78.6% accuracy on the original data. Despite their promising results, their model did not investigate any functional or appearance features which could have led to higher performance especially with the unbalanced data. Another recent study by Cho and Park [17] assessed the accuracy for classification of gliomas by characterization of multimodal MRI data from the MICCAI Brain Tumor Segmentation Challenge (BRATs 2015) [32]. They utilized the logistic regression using shape and texture features to classify gliomas (LGG = 54 and HGG = 54). They reported an overall accuracy of 89.8%. Suarez-Garcia et al. [18] investigated the role of multimodal-MRI along with texture features to identify HGG from LGG. A total of 285 subjects were obtained (HGG = 210 and LGG = 75) from the BRATs 2018 challenge. They extracted texture features from the gray level size zone matrix. These features are then fed to multiple linear regression models toward finding the best model to classify gliomas. Their best model was reported to have an accuracy of 91.8% distinguishing HGG from LGG. Banerjee et al. [19] evaluated the abilities of CNN along with multi-sequence MR images to differentiate HGG from LGG. The authors included 746 patients (HGG = 472 and LGG = 274) from The Cancer Imaging Archive (TCIA) [33] and BRATs. The study proposed two fine-tuned ConvNets and evaluated these models using leave-one-subject-out. They reported their highest accuracy of 95% for classification of gliomas. In another recent study, Alis et al. [20] utilized the artificial neural networks (ANN) to differentiate HGG from LGG. The authors assessed the diagnostic accuracy of conventional MR images texture analysis. They extracted first and second order texture features from manually-placed ROIs of 181 subjects (HGG = 97 and LGG = 84). Their proposed pre-trained ANN achieved an accuracy of 88.3% discriminating HGG from LGG. Hsieh et al. [21] evaluated intensity-invariant MR imaging obtained from 107 patients diagnosed with glioma (HGG = 34 and LGG = 73). They extracted texture features and histogram moments from local binary patterns of manually-delineated tumors. These features were then combined in a binary logistic regression classifier to distinguish HGG from LGG. They reported an accuracy of 93%. Six CNN models were developed by Kalaiselvi et al. [22] for the classification of glioma lesions from volumetric MR scans. Their study included 4500 images extracted randomly from BRATs 2013 (30 volumes) and Whole Brain Atlas (WBA) (8 volumes). Their best model successfully discriminated HGG from LGG with an accuracy of 88.91%. Zhuge et al. [23] gathered conventional MR images of 315 patients (HGG = 210 and LGG = 105) from TCIA and BRATs 2018 to develop two deep learning models for distinguishing HGG from LGG. With an accuracy of 97.1%, their proposed 3D ConvNet model achieved the best diagnostic performance. Using multimodal MR images of 285 cases (HGG = 210 and LGG = 75) from the BRATs 2017, Cho et al. [24] assessed the diagnostic performance of radiomic features for discriminating between LGG and HGG. The authors extracted 468 radiomic features and then they selected their optimal features to build three different machine learning classification models. They reported an average accuracy of 88.54% using three different classifiers: logistic regression, SVM, and random forest classifiers.

Despite the fact that these papers represent the state-of-the-art research studies for grading gliomas using MRIs along with computer-aided diagnostic (CAD) systems, they had some limitations that need to be addressed. Most of these studies [17–24] only focused on differentiating between LGG and HGG instead of grading these tumors into Grades I, II, III, or IV for a proper management plan. In addition, none of the aforementioned studies incorporated any functional features for a better diagnostic performance. Some of the authors [16,18,20,21] incorporated texture features in their work, but they did not investigate any shape or appearance features which could have led to higher performance as well.

To circumvent for these drawbacks, we introduce a novel comprehensive glioma grading CAD (GG-CAD) system, shown in Figure 1. To the best of our knowledge, the GG-CAD system is the first of its kind to integrate novel 3D appearance features, volumetric features, 3D first and second order textural features, with functional features extracted from multimodal MR images to precisely identify the glioma grade as Grade I, II, III, or IV for a proper medical management plan.
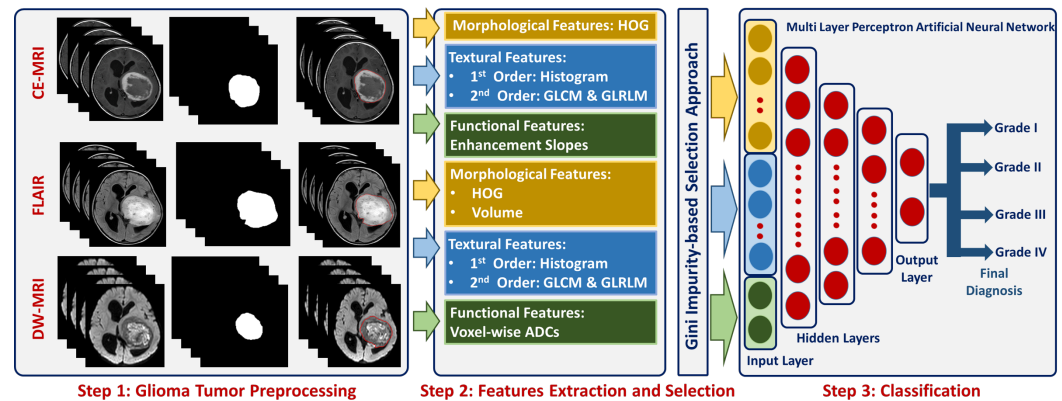

Figure 1. The developed GG-CAD system for accurate grading of glioma tumors

2. Materials

Input Data Description: A total of 99 biopsy confirmed glioma tumors (M = 49 and F = 50) with age range between 1 and 79 years and average of 40.15 ± 19.94 years were included (Grade I = 13, Grade II = 22, Grade III = 22, and Grade IV = 42). All patients, themselves or their parents/legal guardians (for minors with age < 18 years), provided their informed consent to participate in this study. Multimodal MR images, namely; T1-MR with pre- and post-contrast phases, T2-MR (FLAIR), and DW-MR, were acquired at Mansoura University Hospital, Egypt.

Imaging Protocols: The MRI examinations were performed using a 1.5 Tesla scanner (Ingena, Philips medical system, Best, The Netherlands). Examination was done with the participants in the supine position, using a standard eight-channel head coil. With a slice thickness = 3 mm and a matrix = 256×256, the following sequences were performed: axial T1 (TR/TE = 475/15 ms), axial T2 (TR/TE = 1250/100 ms), and axial FLAIR (TR/TE/TI = 8000/140/2000 ms). For the contrast-enhanced T1-MR, post-contrast axial images were acquired after administration of gadolinium-based contrast agents with a dose of 0.1 mmol/kg. The axial DW-MR were performed with multi-section single-shot spin-echo planar imaging (EPI) sequence (TR/TE/NEX = 3000/88/1) using *b*-values of 0 and 1000 s/mm^2. It is worth-mentioning that for all modalities, multiple axial cross-sections were obtained to cover the whole volume and were stored in DICOM format.

Reference Standard Diagnosis: Biopsy remains the reference standard to definitively diagnose gliomas and give a prognosis to determine and guide a treatment/management plan by pathologically testing the abnormal tissues within the brain. In most of the included subjects, excisional biopsy was performed. Multiple cores from the solid enhancing portion of the excised tumor were obtained. This solid portion differs from one tumor to another, and hence the number and locations of the cores differ according to the anatomical position of the lesion which correlates with tumor type and prognosis. For subjects with gliomas in hard-to-reach and/or sensitive areas within the brain that might be damaged by a more extensive procedure, a stereotactic needle biopsy was done. During this, a neurosurgeon drilled a small hole into the skull to allow the insertion of the thin needle. Small pieces from the abnormal tissues were then removed through the needle guided by radiological scanning. In both cases, the extracted specimens are sent to be pathologically tested, then the highest pathological glioma grade was assumed for the whole tumor.

It is worth mentioning that this study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of Mansoura University (MD.20.01.278).

Glioma Tumor Preprocessing: The more accurate the segmentation is, the more precise feature extraction will be. The initial data for each subject are multiple gray-level images at different modalities stored in DICOM format. These DICOM images were transferred to a workstation (extended MR Workspace release 2.6, Philips Medical Systems, B.V., Eindhoven, The Netherlands). For each subject, regions of interest (ROIs) were manually segmented as binary masks. In the case of the T1-MR and DW-MR; the ROIs identify the tumor itself, while for the T2-MR (FLAIR), the ROIs identify the tumor with its surrounding edema. The segmentation process was performed using an in-house software by two radiologists (blinded from each other while performing the segmentation) with more than 10-years of hands-on experience in analyzing medical images. To generate the ground truth segmentation, the common area between the two observers were extracted, and then an expert radiologist with more than 25-years of hands-on experience in medical image analysis decided whether the difference should be considered as a part of the tumor or normal tissues. Meanwhile, to assess and quantify the agreement between the two segmentations, we performed Bland-Altman analysis [34]. The analysis showed a high agreement between the two segmentations with a mean absolute error of 40.556 and (-141.705, 131.867) $\pm 95\%$ confidence interval. The mean-difference plot for the analysis is shown in Figure 2. Finally, 3D glioma objects were constructed and used for extracting distinguishing features.
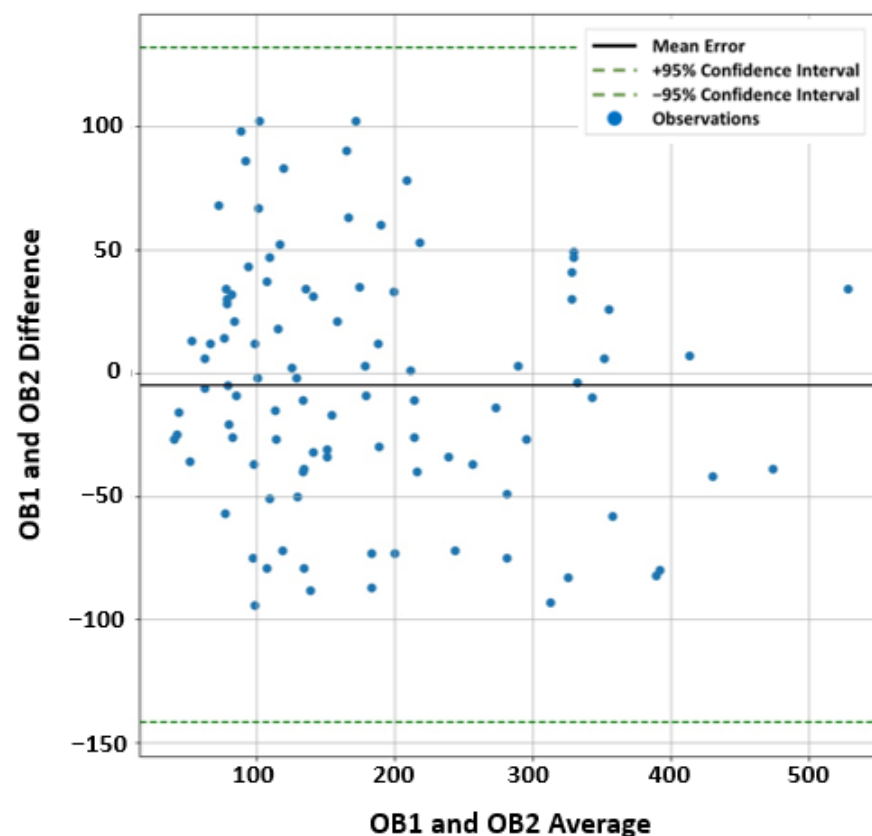

Figure 2. Bland-Altman mean-difference plot showing the agreement between the manual segmentation (extracted ROIs) performed by two observers (OB1 and OB2) blinded-from each other.

3. Methods

The proposed GG-CAD system (Figure 1) can discriminate between different grades of gliomas by performing the following steps: (i) extracting higher order 3D-appearance

features from the segmented glioma tumor by constructing the histogram of oriented gradients (HOG) and estimating the tumor volume from both contrast-enhanced T1-MR and from T2-MR (FLAIR), (ii) calculating first and second order textural features from the segmented tumor based on constructing histogram, gray-level co-occurrence matrix (GLCM), and gray-level run length matrix (GLRLM) by utilizing contrast enhanced T1-MR and T2-MR (FLAIR) imaging, (iii) estimating functional features by estimating 3D ADC maps for the segmented tumor from DW-MR imaging acquired at the b-value of 1000 s/mm^2 and calculating the amount of enhancement between pre- and post-contrast phases of T1-MR imaging, (iv) performing features selection using Gini impurity approach over the integration of all the aforementioned extracted features to get the most significant set of features, and (v) feeding the optimal set of features to a multi-layer perceptron artificial neural networks (MLP-ANN) classification model toward getting the final diagnosis of the tumor as Grade I, II, III, or IV.

3.1. Engineering Features

After preprocessing glioma tumors, the structured objects which represent the different subjects should be expressed as distinguishing, standardized, and machine-understandable features. These features have the ability to discriminate between different subjects through showing the learning algorithm how to interpret the characteristics of each object. The quality of these characteristics defines and enhances our machine learning model predictive abilities. Hence, after consulting the medical team, we had agreed on multiple categories of distinctive features that suit the nature of our problem. Below, we are going in depth with the extracted imaging features.

Higher order 3D-Appearance Features: To obtain a sensitive and specific enough GG-CAD system with the ability to differentiate between different grades of glioma tumors, distinctive parametric higher order appearance features were identified. The motivation for using these 3D-appearance features relies on the hypothesis that gliomas with higher grades have more aggressive growth rates, more complex, rough, and irregular shapes than those with lower grades. Therefore, accurate identification, modeling, and extraction of such descriptors is essential towards a precise diagnosis. In the proposed framework, we identified gliomas by HOG as well as the total tumor volume.

3D-HOG: The HOG descriptor is concerned with an object's morphological structure/appearance finding a simplified image representation that only includes the most significant details about the image [35,36]. The HOG descriptor measures how frequent a gradient orientation occurs in a specific area of an object. In the proposed work, we applied a 3D-HOG approach over the volumes from the T2-MR (FLAIR) and pre- and post-contrast T1-MR. The extracted ROIs are resized to a new shape of ($32 \times 32 \times$ Number of slices per volume). So, all the volumes have the same length (X) and width (Y), while the depth (Z) is the different dimension depending on the number of 2D slices of each volume. The resized ROI is passed to Algorithm 1 with the number of bins set to 9 and the number of cells set to 4. Finally, the HOG descriptor is formed. We have 4 cells, each cell has 8 histograms, and each histogram is generated over 9 bins. Hence, the total number of

features are $4 \times 8 \times 9 = 288$ features for each volume. Figure 3 shows the different steps to applying the 3D-HOG approach.

Algorithm 1 Histogram of oriented gradients

Input: ROI V of size $m \times n \times p$ and voxel size $\Delta x \times \Delta y \times \Delta z$, cell size C and number of bins B

Output: The HOG H_{ijk}

1 **foreach** *voxel* (i, j, k) **do**

2 $\quad M_{ijk} := \sqrt{(V_{i+1,jk} - V_{i-1,jk})^2 + (V_{ij+1,k} - V_{ij-1,k})^2 + (V_{ijk+1} - V_{ijk-1})^2}$

3 $\quad \Theta_{ijk} := \cos^{-1} \left(\frac{V_{i+1,jk} V_{i-1,jk} + V_{ij+1,k} V_{ij-1,k} + V_{ijk+1} V_{ijk-1}}{\sqrt{V_{i+1,jk}^2 + V_{ij+1,k}^2 + V_{ijk+1}^2} \sqrt{V_{i-1,jk}^2 + V_{ij-1,k}^2 + V_{ijk-1}^2}} \right)$

4 **foreach** *non-overlapping cell i of size* $C \times C \times p$ **do**

5 $\quad$ **foreach** *bin j* **do**

6 $\quad\quad D_j := [(j-1)\pi/B, j\pi/B)$

7 $\quad\quad H_{ij} := \Sigma \{ M_{rst} | (r,s,t) \in i \land \Theta_{rst} \in D_j \}$

8 $\quad\quad$ **foreach** *k in the 8-neighborhood of i* **do**

9 $\quad\quad\quad H_{ijk} := H_{ij} / \Sigma \{ M_{rst} | (r,s,t) \in k \}$
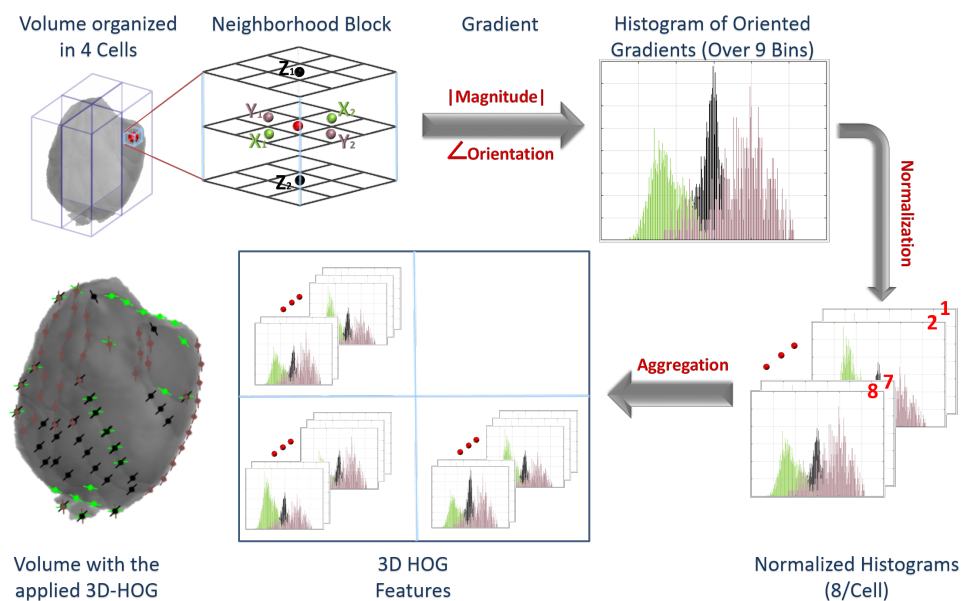
Figure 3. Applying the 3D HOG approach.

Textural Features: To enhance the performance of early differentiation between different grades of glioma tumors, we extracted comprehensive first and second order texture features that give precise description for the heterogeneity/homogeneity of the detected gliomas. We were motivated by the hypothesis that high grade glioma tumors are more heterogeneous than low grade gliomas in terms of textural appearance [37,38]. An illustrative Figure (see Figure 4) showing the texture differences between 4 subjects with different glioma grades demonstrates the feasibility of our hypothesis.

Figure 4. Visualization of texture differences between four gliomas with different grades (**a**) Grade I, (**b**) Grade II, (**c**) Grade III, and (**d**) Grade IV using contrast-enhanced T1-MR (pre and post phases) and T2-MR (FLAIR).
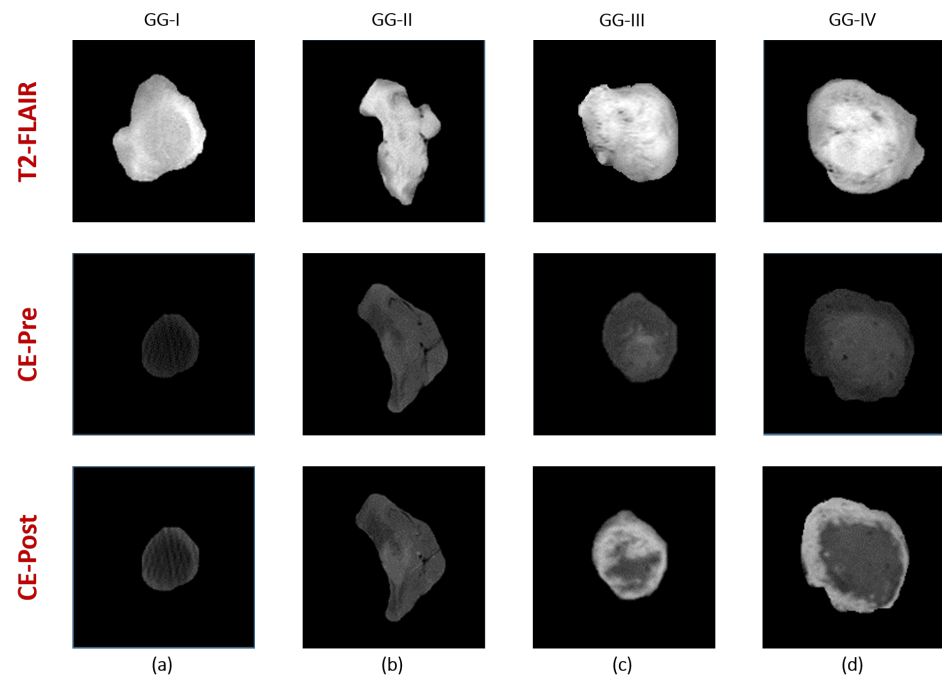
All first order textural features were estimated using a normalized empirical histogram. While first-order textural qualities are useful descriptors, they are susceptible to noise. To grant a better quantification of heterogeneity between different grades of tumors, second order textural features (GLCM and GLRLM) were used [39,40].

GLCM: The GLCM is used to evaluate the spatial interactions between voxels in a neighborhood block (reference and nearby voxels). The GLCM is a bivariate histogram encoding the frequency of neighboring voxel pairs having specific intensities. To construct the GLCM it is first necessary to identify the intensity range of the region of interest and define the level of quantization. Next, one defines the neighborhood system by specifying the spatial relationship between pairs of voxels that are considered neighbors. Each element (i, j) of the GLCM is proportional to the instance of a voxel with intensity i having a neighboring voxel with intensity j, and the matrix is normalized such that all elements sum to unity [39,41,42]. For our application, voxel intensities were 8-bit quantized, producing a 256×256 GLCM. Each voxel's neighbors were all those within $\leq \sqrt{2}$ mm distance, so the resulting GLCM was approximately symmetric, except for the effect of boundary voxels.

GLRLM: Aside from determining the occurrence frequency of GLCM-represented voxel pairs, GLRLM examines voxel runs to determine the connectivity of voxels. GLRM encodes the number of times a run of n consecutive voxels occurs with the same intensity. As with the GLCM, we first define the gray level range and its quantization $(0, \ldots, 255$ in 8 bits), which fixes the number of rows in the GLRLM. The column dimension is the maximum extent of the region of interest in the dimension where runs are being observed. Then element (i, j) of the GLRLM is the relative instance of a run of j consecutive voxels all having intensity i [40–42]. Because of the difference between slice spacing and pixel spacing in the MRI data, we constructed one GLRLM for voxel runs in the x and y directions, within the same MRI slice, and another for runs in the z direction, i.e., across slices at the same (x, y) location in each slice. See Tables A1–A5 for a detailed description and formulas of the extracted textural features.

Functional Features: Functionality of gliomas has direct effects on the imaging which can lead to enhancing the abilities of the model in identifying the grade of glioma tumors.

Therefore, we investigated two different functional features from T1-MR and DW-MR to capture the functionality aspects of different glioma tumors.

Contrast-enhancement slopes (T1-MR): Hyperenhancement (contrast-enhancement) can characterize the functionality of a given glioma tumor. The contrast-enhancement slope can be estimated between pre- and post-contrast phases of T1-MR. The enhancement leads to remarkable changes in the gray values. These temporal changes enable the construction of contrast-enhancement slope. This slope is calculated by getting the gray-level intensity change rate over the time period between the two phases [43,44]. HGG tumors might demonstrate higher and faster slopes than those of LGG tumors.

ADCs (DW-MR): The diffusion-weighted MR signal is affected by diffusion of water through the tissue and also by capillary perfusion. Pathological tissues, such as gliomas, are likely to have unusual diffusivity properties relative to surrounding healthy tissue. Hence, well-known functional parameters called apparent diffusion coefficients (ADCs) can be estimated from different gradient field strength and duration (*b*-value). Higher *b*-values produce greater attenuation of the DW-MR signal intensity. The voxel-wise ADC values [45] at a certain *b*-value can be calculated as shown in Figure 5. Subsequently, we set the *b*-value to 1000 s/mm² and calculated ADC accordingly for each voxel within the segmented tumor. Using this set of ADCs directly as a descriptive feature is problematic, since every segmented region includes a different number of voxels. Therefore, we standardized the length of the descriptor by binning the ADC measurements and calculating their cumulative distribution function, or CDF (Figure 5).
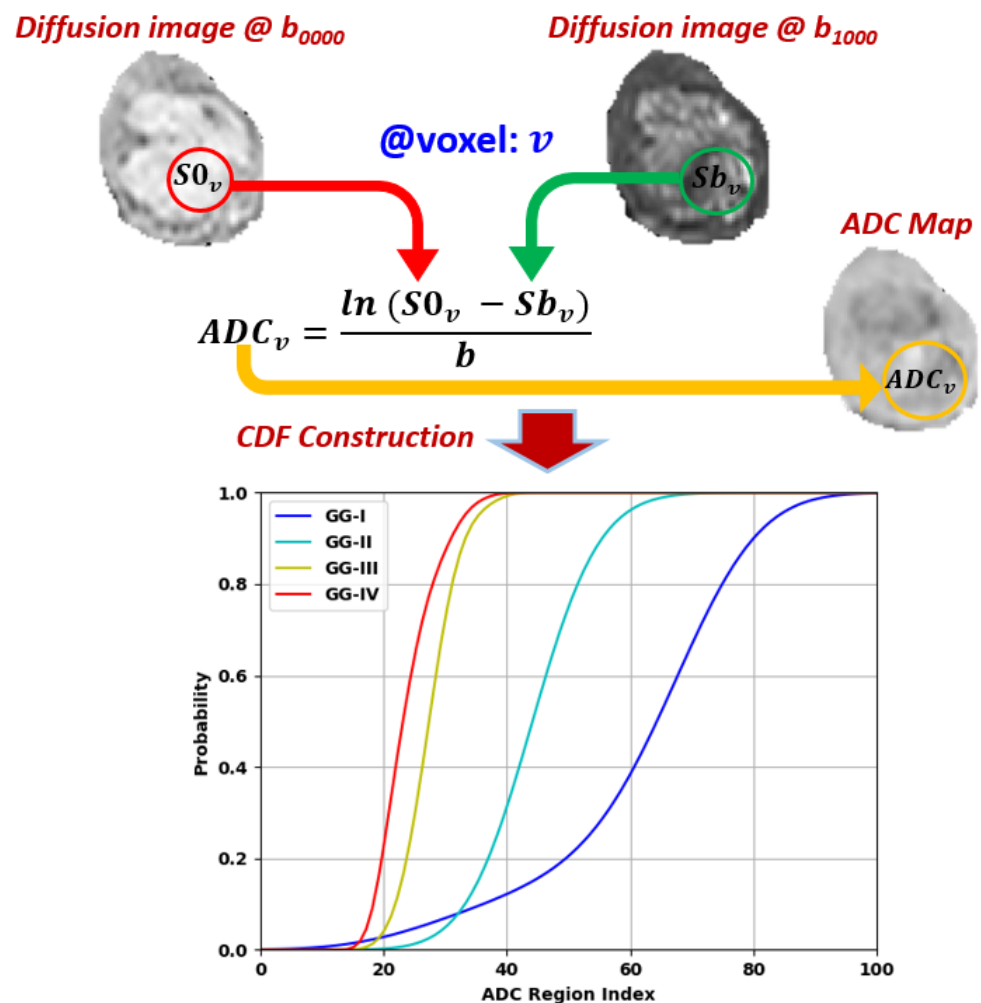

Figure 5. Calculations of voxel-wise apparent diffusion coefficients (ADC) for gliomas and the cumulative distribution functions (CDFs) of four gliomas with different grades.

3.2. Classification and Hyper-Parameters Tuning

After deciding on the extracted discriminating features: appearance, textural, and functional for all glioma tumors from three different MR modalities, namely; contrast-enhanced T1-MR, T2-MR (FLAIR), and DW-MR, we proceeded with the classification process to differentiate between the different glioma grades by utilizing MLP-ANN classification models.

MLP-ANNs are well known machine learning classifiers that mainly include three types of layers: an input layer, one or more hidden layers, and an output layer, each with a large number of activation/processing units called nodes/neurons. The network is structured assuring that a full connection occurs between adjacent layers. The MLP-ANN can partition the input space of a feature into arbitrarily complex regions using non-linear activation functions used by neurons. In order to minimize the loss function, the MLP-ANN primarily employs a supervised backpropagation learning technique, where the connection weights and additive biases can be updated in the training phase using gradient descent methods [46].

Hyper-Parameters Tuning: To obtain the optimal set of MLP-ANN hyper-parameters, an in-depth grid search algorithm was employed. In the grid search, a grid containing different combinations of hyper-parameters is set up and then the MLP-ANN is trained/tested on each of the possible combinations searching for the best diagnostic performance. During the search process the accuracy of diagnosis was used as a metric for optimization purposes. In order to avoid using some hyper-parameters leading to a good performance on the training data but not so good with the test data, the hyper-parameters optimization is implemented using leave-one-subject-out (LOSO) cross-validation on all of the extracted features from all subjects within the dataset. Finally, the following parameters were approved for the proposed GG-CAD system using MLP-ANN (trainfcn: trainlm, max epochs = 500, hidden layers: $hl_1 = 200$, $hl_2 = 100$, $hl_3 = 50$, goal = 0, max validation failure = 6, min gradient = 10^{-7}, training gain (μ): initial $\mu = 0.001$, μ decrease factor = 0.1, μ increase factor = 10, $\mu_{max} = 10^{10}$).

3.3. Engineering Features Selection

Typically, a features selection approach is utilized to select the optimal relevant features from a wide cohort of prospective features. This technique yields m features from a set of n options, where $m < n$, and m is the smallest set of significant and important features. Here, we applied a Gini impurity-based selection [41].

Gini Impurity-based Selection: Besides their reputation as one of the robust machine learning classifiers, random forests can be used as feature selectors. This is due to the fact that random forests' tree-based techniques naturally rely on how effectively the purity of the node is strengthened. This indicates an impurity deterioration across all trees, which is called Gini impurity. The nodes with the highest drop in impurity are found at the beginning of the trees, while the nodes with the lowest impurity drop are found at the end. By trimming trees below a certain node or a certain impurity threshold, an optimal subset of the most significant features can be selected. The process of adopting this selection strategy are shown in Algorithm 2. To discover the best set of features to employ for the learning process, we used the Gini impurity-based technique on the entire set of the integrated features. Here, searching for the optimal impurity threshold value to be used, we applied three alternative scenarios for the selection process through using three different values of impurity threshold. Using LOSO cross-validation approach, each reduced features set is utilized (trained/tested) on the MLP classification model. Then, we compared the diagnostic performance of the developed GG-CAD system using all the extracted features with using the three reduced sets of features. From Table 1, we can find that using an impurity threshold of 0.001 gave the best performance and that a set of 332 selected features is the optimal set to be used for the proposed model. Table 2 shows the number of selected

components from each group of features. Details of the selected features for the final model are shown in Table A6.

Algorithm 2 Gini-Impurity-based Selection Process
1　　Prepare the integrated features set.
2　　Model and train a Random Forests classifier.
3　　Set the impurity threshold.
4　　Determine the features with the highest significance.
5　　Build the final set that includes only these selected features.

Table 1. Diagnostic performance (in terms of quadratic-weighted Cohen's kappa and accuracy of the proposed GG-CAD system using the integrated features vs. three different sets of selected features. Let m, IT, GG-I, GG-II, GG-III, and GG-IV denote the number of features, impurity threshold, Glioma Grades I, II, III, and IV, respectively.

Selection	m	Kappa	Accuracy%				
			GG-I	GG-II	GG-III	GG-IV	Overall
All features	1099	0.94 ± 0.04	95.4 ± 5.5	95.8 ± 4.2	89.4 ± 5.7	96.4 ± 1.5	94.6 ± 2.3
IT = 0.0010	332	0.96 ± 0.02	95.9 ± 3.8	96.7 ± 4.2	91.2 ± 4.2	97.6 ± 2.5	95.8 ± 1.9
IT = 0.0015	128	0.90 ± 0.08	87.7 ± 7.3	91.8 ± 6.3	90.0 ± 6.0	97.5 ± 2.7	93.3 ± 3.9
IT = 0.0020	61	0.86 ± 0.04	81.5 ± 8.8	87.3 ± 3.4	87.6 ± 7.1	94.9 ± 3.2	89.8 ± 3.0

Table 2. Extracted features taxonomy with their various types and counts before and after features selection. Let n and m denote number of features before and after selection, respectively.

Features	MR Modality	n	m
Higher order 3D-Appearance Features			
3D-HOG + Volume	T2-MR (FLAIR)	289 (288 + 1)	82 (81 + 1)
3D-HOG	T1-MR (Pre-contrast)	288	58
3D-HOG + Volume	T1-MR (Post-contrast)	289 (288 + 1)	46 (45 + 1)
Textural Features			
Histogram + GLCM + GLRLM	T2-MR (FLAIR)	44 (26 + 6 + 12)	30 (18 + 4 + 8)
Histogram + GLCM + GLRLM	T1-MR (Pre-contrast)	44 (26 + 6 + 12)	35 (21 + 5 + 9)
Histogram + GLCM + GLRLM	T1-MR (Post-contrast)	44 (26 + 6 + 12)	37 (22 + 6 + 9)
Functional features			
CDFs of ADCs (b_{1000} s/mm^2)	DW-MR	100	43
Contrast-enhancement Slope	T1-MR	1	1
Integrated Features			
Integrated	All	1099	332

3.4. Experiments

To validate the performance of the proposed GG-CAD system, the following experiments were performed:

Experiment I: To assess the final diagnostic performance of the proposed GG-CAD system, the optimal subset of features that was selected using the Gini impurity approach, were fed to a hyper-tuned MLP-ANN classification model.

It is worth mentioning that the diagnostic performance of the GG-CAD system was evaluated using a LOSO and k-fold (with k = 5 and 10) stratified cross-validation approach on the aforementioned data. For the LOSO, all data except one subject are used to train the

classification model. Before the next iteration, the classification model is reinitialized, and the observation that was previously left out of the training data is included in the training data, leaving the next subject out for testing purposes. This method is performed 99 times (the included subjects in the study), with the training and testing samples being 98 and 1, respectively, at each iteration. For the k-fold stratified cross-validation, a subset $\frac{1}{k} \times 100\%$ of the data is chosen at random and saved for testing, while the rest ($\frac{1-k}{k} \times 100\%$) is utilized for training. In the next iteration, the classification model is reinitialized, and the subjects from the previous iteration are included in the training, leaving the next $\frac{1}{k} \times 100\%$ group of subjects for testing. This procedure is performed k times.

Stratification was important to eliminate any opportunity for bias and variance during the execution of k-fold cross-validation. The stratification technique not only achieves randomization, but it also assures that original distribution of subjects over different classes in the overall data set is maintained in the training/testing sets. To achieve stratification in our case, 13% of the training/testing sets were derived from Grade I, 22% from Grade II, 22% from Grade III, and 43% from Grade IV. These randomly stratified k-fold cross-validation approaches (with $k = 5$ and 10) were utilized to ensure the that the proposed model is robust and not prone to overfitting.

Experiment II: To show the added value of each group of extracted features and highlight the importance of the features integration process, we measured monothetic classifier performance using each individual group of features in turn.

Experiment III: Finally, to appreciate the diagnostic performance obtained by the developed GG-CAD system, we applied two different approaches from the literature [20,22] on our dataset (N = 99) and with the intended classification problem of glioma grading (GG-I vs. GG-II vs. GG-III vs. GG-IV) for a fair comparison. Then, we compared the final diagnostic results obtained by the developed GG-CAD system with those obtained by the two different approaches.

Evaluation Metrics: Using quadratic-weighted Cohen's kappa [47] and accuracy as the evaluation metrics during the classification, all experimental results (Section 4) were documented in terms of mean $\pm$ standard deviation over 15 times of repetition.

4. Results

Experiment I: To evaluate the classification abilities of the hyper-tuned MLP-ANN model utilized in the developed GG-CAD system, the diagnostic results using the selected optimal set of features along with MLP-ANN were obtained. To demonstrate the generalization ability and reproducibility of the GG-CAD model, a favorable comparison is performed using the three aforementioned validation approaches. As documented in Table 3, the proposed GG-CAD system shows high and robust diagnostic performance using the three different validation approaches. The confusion matrix (in terms of accuracy) of the proposed model (utilizing MLP-ANN uisng LOSO cross-validation over the selected features) is shown in Figure 6.

Table 3. Diagnostic performance (in terms of quadratic-weighted Cohen's kappa and accuracy) of the proposed GG-CAD system using a LOSO and k-fold (with $k = 5$ and 10) stratified cross-validation approach on the optimal set of features. Let GG-I, GG-II, GG-III, and GG-IV denote Glioma Grades I, II, III, and IV, respectively.

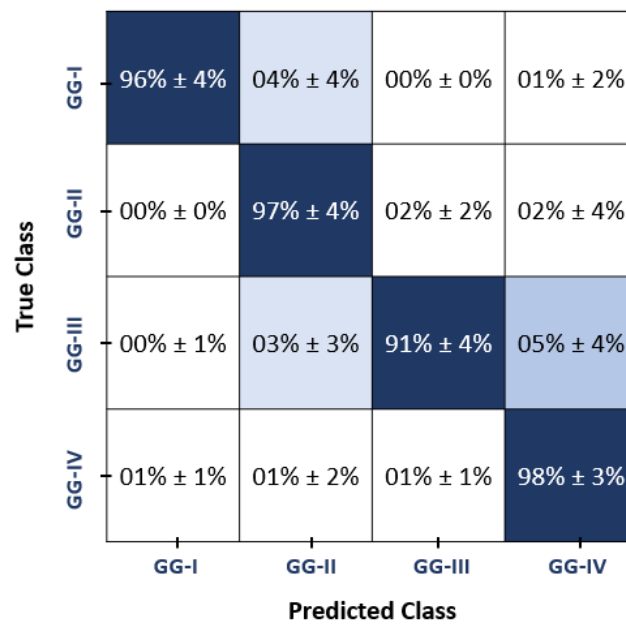| Approach	Kappa	Accuracy%				
		GG-I	GG-II	GG-III	GG-IV	Overall
LOSO	0.96 $\pm$ 0.02	95.9 $\pm$ 03.8	96.7 $\pm$ 04.2	91.2 $\pm$ 04.2	97.6 $\pm$ 02.5	95.8 $\pm$ 1.9
10-Fold	0.91 $\pm$ 0.02	94.9 $\pm$ 06.1	93.0 $\pm$ 03.3	87.0 $\pm$ 05.7	95.2 $\pm$ 01.5	92.9 $\pm$ 01.8
5-Fold	0.90 $\pm$ 0.03	89.7 $\pm$ 06.1	93.6 $\pm$ 05.5	80.9 $\pm$ 05.3	93.7 $\pm$ 01.7	90.3 $\pm$ 02.4

Figure 6. Confusion matrix of the final proposed MLP-ANN model, showing the accuracy confusion between different grades. Let GG-I, GG-II, GG-III, and GG-IV represent the four glioma grades.

Experiment II: The performance of the GG-CAD system was then evaluated using the individual features along with hyper-tuned MLP-ANN classification models. To highlight the advantage of integrating these individual features, we compared the diagnostic performance of the proposed GG-CAD model (fusion of extracted features) with these individual models. With a kappa of 0.96 ± 0.02 and an overall accuracy of $95.8\% \pm 1.9\%$, the diagnostic performance of the GG-CAD system using the integrated model outperformed all other individual classification models as documented in Table 4. This can be justified by noting that integrating different quantitative features characterizes different aspects of the tumor such as, appearance, texture, and functionality.

Table 4. Diagnostic performance (in terms of quadratic-weighted Cohen's kappa and accuracy) of the proposed GG-CAD system using the integrated features vs. individual ones using MLP-ANN. Let GG-I, GG-II, GG-III, and GG-IV denote Glioma Grades I, II, III, and IV, respectively. The MLP-ANN column shows the structure/architecture of the MLP-ANN models.

MR Modality	Kappa	Accuracy%					MLP-ANN
		GG-I	GG-II	GG-III	GG-IV	Overall	
Higher order 3D-Appearance Features							
T2-MR (FLAIR)	0.85 ± 0.03	93.3 ± 7.4	84.6 ± 8.4	73.0 ± 6.7	94.4 ± 2.4	87.3 ± 2.2	(50, 25)
T1-MR (Pre)	0.71 ± 0.06	83.6 ± 7.4	79.7 ± 5.5	67.6 ± 3.7	92.9 ± 1.7	83.1 ± 1.9	(50, 25)
T1-MR (Post)	0.78 ± 0.05	87.7 ± 3.8	78.2 ± 6.0	66.1 ± 4.9	93.3 ± 1.3	83.2 ± 1.6	(50, 25)
Textural Features							
T2-MR (FLAIR)	0.82 ± 0.02	87.7 ± 7.3	72.1 ± 6.0	60.9 ± 4.9	93.7 ± 2.2	80.8 ± 1.2	(25, 10)
T1-MR (Pre)	0.79 ± 0.04	79.5 ± 8.7	76.7 ± 6.4	63.9 ± 5.6	91.6 ± 2.4	80.5 ± 1.8	(25, 10)
T1-MR (Post)	0.78 ± 0.05	86.2 ± 4.2	81.2 ± 2.3	70.3 ± 6.0	92.2 ± 1.1	84.1 ± 2.5	(25, 10)
Functional Features							
DW-MR & T1-MR	0.79 ± 0.05	82.6 ± 7.7	78.8 ± 5.2	64.6 ± 6.5	91.8 ± 1.7	81.6 ± 2.5	(50, 25)
Integrated Features							
GG-CAD (All)	0.96 ± 0.02	95.9 ± 3.8	96.7 ± 4.2	91.2 ± 4.2	97.6 ± 2.5	95.8 ± 1.9	(200, 100, 50)

Experiment III: Finally, to highlight the advantages of the developed GG-CAD system, in-depth comparisons with different approaches have been performed. As documented in Table 5, the diagnostic performance of the developed CAD system outperformed both the aforementioned approaches for glioma tumor grading.

Table 5. The final diagnostic performance (in terms of quadratic-weighted Cohen's kappa and accuracy) for grading the tumors into (GG-I, GG-II, GG-III, and GG-IV) by using (a) the proposed CAD system, (b) approach by Alis et al. [20], and (c) approach by Kalaiselvi et al. [22].

Model	Kappa	Accuracy%				
		GG-I	GG-II	GG-III	GG-IV	Overall
GG-CAD	0.96 ± 0.02	95.9 ± 03.8	96.7 ± 04.2	91.2 ± 04.2	97.6 ± 02.5	95.8 ± 01.9
Alis [20]	0.79 ± 0.05	62.1 ± 16.5	77.3 ± 06.4	66.7 ± 02.1	90.5 ± 01.9	78.5 ± 04.6
Kalaiselvi [22]	0.82 ± 0.05	77.4 ± 15.0	78.8 ± 05.4	67.0 ± 08.2	90.5 ± 02.1	80.9 ± 03.5

5. Discussion

In adults, gliomas have the highest rates of prevalence and incidence among brain tumors and cause significant mortality and morbidity. They can occur anywhere in the central nervous system but primarily occur in the brain and arise in the glial tissue namely astrocytes, oligodendrocytes, and ependymal cells. Based on histological characteristics, the WHO traditionally classifies gliomas into four different grades . Grade I are solid and non-infiltrative tumors (pilocytic astrocytomas), while grades II–IV are diffuse infiltrating gliomas. Each grade of the glioma has a different treatment plan, which increases the importance of accurate grading of gliomas. Despite its demerits, surgical biopsy remains the gold standard for grading gliomas. Therefore, studying the potential for MRI-based grading of gliomas using CAD systems as a replacement for biopsy has been largely investigated with an ongoing interest [16–24].

In this study, the proposed GG-CAD system showed high diagnostic abilities in differentiation between different grades of glioma tumors. This precise tumor grading can guide neuro-oncologists to a proper decision for tumor treatment planning and better prognosis. Conventional MRI can assess the anatomy of tumor and advanced MRI modalities can provide quantitative assessment of tumor metabolic features, tumor cellularity, and vascularity. Hence, three different MR imaging modalities, namely; contrast-enhanced T1-MR, T2-MR (FLAIR), and DW-MR were utilized to provide different aspects of features. Using the most significant sets of the extracted discriminating features along with one of the most powerful machine learning classification techniques like MLP-ANN is proven to be highly efficient in determining the malignancy grade of a given glioma tumor.

In this work, the extracted ROIs from the images of different modalities were stacked and rendered in 3D objects representing the subjects. These 3D objects comprise multiple voxels representing the tumor itself (in contrast-enhanced T1-MR and DW-MR) or the tumor and its surrounding edema (in T2-MR (FLAIR)). Numerous histopathological parameters control the signal strength of each voxel displaying a specific gray-scale value. As a result, 3D arrays of gray-scale values in ROIs may reveal sophisticated geometric patterns that are specific to tumor grades, even if they are visually unrecognizable.

Hence, texture analysis was performed in our research. Texture analysis gives an effective description of how the gray-level of each voxel in a particular area affects the overall distribution of the values of voxels. These texture data have shown a significant impact on the performance of classification systems in a variety of studies [16–18,20,21,24].

The developed model utilized first and second order texture features using different methods and algorithms of texture analysis. First-order features or histogram-based features express the histogram representing the distribution of image intensity, which describe how the intensity signals of voxels are distributed over the tumor. So, these features neglect the spatial orientation and relationship between voxels. Second-order features are statistical

relationships between adjacent voxels, or groups of voxels, in terms of intensity levels. Intratumoral heterogeneity is quantified by these features. These features are derived from quantitatively describing matrices encoding precise spatial relationships between voxels in a specific area in the tumor. We've incorporated both GLCM and GLRLM in our research [39,40]. While the GLCM shows how frequently two intensity levels appear in adjacent voxels within the tumor's object, the GLRLM encodes the size of homogeneous runs in the same object.

In addition to the role of texture analysis in specifying the grade of gliomas, the severity of a glioma determines the 3D-appearance and shape of both the tumor and the surrounding edema. Higher grade gliomas appear to be much more complex and have sharper edges than these lower grade gliomas. This motivated us to measure 3D-appearance features to capture the potential shape differences between different grades of gliomas. Additionally, the volumes of the tumor itself and the tumor with the surrounding edema are also calculated.

One of the most beneficial aspects was to capture the functionality of different grades of gliomas that might be a key point towards achieving our goal. Functionality of gliomas has an effect on the imaging which can lead to enhancing the abilities of the model towards identifying the grade of glioma tumors. 3D DW-MR of each glioma was acquired using $b = 0$ and $b = 1000$ s/mm^2). ADCs at the nonzero b-value were then calculated to capture the functionality differences between subjects with different grades. In contrast-enhanced T1-MR, the contrast-enhancement changes which differ according to the severity of the tumor and this difference is also captured. Contrast-enhancement slopes detecting the changes in gray-level distribution between pre- and post-contrast phases are estimated to quantify the enhancement variations between the different grades.

In the classification process, we incorporated all of the features with machine learning models. Most of these characteristics differed significantly amongst glioma grades, however there was still a significant overlap. Even when the most appropriate MR sequence has been utilized, such diversity negatively affects the ability of using a single feature class to accurately identify the glioma grade. On the other hand, using an optimal set of significant features selected from the combination of all features provided a better feature representation to detect the grades of different gliomas. With high classification performance, the developed GG-CAD system discriminated between the different grades of gliomas using the optimal set of features selected from the integration of all extracted features. The obtained results demonstrate the diagnostic abilities of the proposed model as well as the clinical utility of our methods when combined with MR imaging in the computer-aided diagnosis of brain malignancies. These results are reported in Tables 1, 3–5. Moreover, Figure 6 shows the confusion matrix of the final proposed model.

6. Conclusions and Future Work

To sum up, the developed GG-CAD system integrates and selects the optimal appearance, textural, and functional features, which demonstrated an impressive diagnostic performance (kappa = 0.96 ± 0.02 and overall accuracy = 95.8% ± 1.9%) using a MLP-ANN classification model. The viability of integrating different significant features representing diverse elements of the glioma tumor characteristics, namely; appearance, texture, and functionality is demonstrated by these findings. This study, however, is limited by the relatively small data size. We are currently collecting a larger data cohort to investigate the abilities of different deep learning (e.g., CNN and stacked-autoencoders) in segmenting glioma tumors, extracting best discriminating features, and identifying the glioma grade in a fully-automated way. Having a larger cohort of data will also enable us to study the effect of different patients' age groups on the precise grading of gliomas. Our future work will include correlating the pathological diagnosis with the treatment response (e.g., no response, partial response, complete response, or progressive response). The success of such findings will lead to building a computer-aided prediction (CAP) system that will be

able to predict the treatment response in an objective way to identify the best treatment plan for each glioma patient. That is known as a personalized medicine treatment plan.

Author Contributions: Conceptualization, A.A., M.S., H.A., N.S.A., S.A.F. and A.E.-B.; methodology, A.A., M.S., H.A., F.S., N.S.A., S.A.F. and A.E.-B.; software, A.A., M.S. and A.E.-B.; validation, A.A., M.S., H.A., N.S.A., M.G., S.A.F., L.G.E.-S. and A.E.-B.; formal analysis, A.A., M.S., H.A., F.S., N.S.A., M.G., S.A.F., L.G.E.-S. and A.E.-B.; investigation, A.A., M.S., and A.E.-B.; resources, N.S.A., M.G., S.A.F., L.G.E.-S. and A.E.-B.; data curation, H.A., F.S., S.A.F. and A.E.-B.; writing—original draft preparation, A.A., M.S., H.A., M.G. and A.E.-B.; writing—review and editing, A.A., M.S., F.S., N.S.A., S.A.F., L.G.E.-S., M.G. and A.E.-B.; visualization, A.A., M.S., S.A.F. and A.E.-B.; supervision, S.A.F. and A.E.-B.; project administration, A.E.-B.; funding acquisition, N.S.A. and A.E.-B. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of Mansoura University (MD.20.01.278).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data could be made available after acceptance upon a reasonable request to the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

In this appendix, we are going to detail the mathematical formulas used to extract the textural features and the definitions of these features are also given. Therefore, the following basic notations will be used:

Notation

- N: The largest possible run length.
- N_p: Gray-level intensity levels.
- N_g: Normalized intensity levels.
- n_r: The overall count of runs.
- n_p: The overall count of pixels.
- (i,j): Row and column index, respectively.
- P(i): Gray-level intensity values.
- g(i): Normalized gray-level intensity values.
- ϵ: Arbitrarily small, positive number
- GLCM: The gray-level Co-occurrence Matrix.
- CM: The normalized GLCM.
- CM(i,j): Element at position (i,j) in the normalized GLCM.
- RL: The gray-level Run Length Matrix.
- RL(i,j): Element at position (i,j) in the GLRLM.
- μ_x: The marginal rows mean.
- μ_y: The marginal columns mean.
- σ_x: The marginal rows standard deviation.
- σ_y: The marginal columns standard deviation.

Table A1. First order textural features.

Textural Feature	Definition
Mean (μ)	Average MR signal intensity (T1-weighted or T2-weighted) of the object of interest.
Variance	Second central moment of MR signal intensity distribution.
Skewness	Asymmetry of signal intensity distribution.
Kurtosis	Tail weight of signal intensity distribution.
Entropy	Metric of randomness of the intensity distribution.
CDFs	Cumulative sum of the relative frequency histogram.
Percentiles	Inverse function of the CDF.

Table A2. Second order textural features.

Textural Feature	Definition
GLCM	
Contrast	Average intensity difference between a voxel and its neighbors.
Dissimilarity	Extent to which neighboring voxels differ in intensity.
Homogeneity	Large-scale uniformity of voxel intensities.
Angular second moment (ASM)	Local uniformity of voxel intensities.
Energy	The square root of the ASM.
Correlation	Linear dependency between neighboring voxels.
Gray-level non-uniformity (GLN)	Dissimilarity of gray-level values within the object.
High gray-level run emphasis (HGLRE)	Incidence of runs of high intensity.
Long run emphasis (LRE)	Distribution of long runs.
Long run high gray-level emphasis (LRHGLE)	Distribution of long runs of high intensity.
Long run low gray-level emphasis (LRLGLE)	Distribution of long runs of low intensity.
Low gray-level run emphasis (LGLRE)	Incidence of runs of low intensity.
Run entropy (RE)	Degree of randomness in the run lengths.
Run length non-uniformity (RLN)	Degree of inhomogeneity in the run lengths.
Run percentage (RP)	Homogeneity of homogeneous runs.
Short run emphasis (SRE)	Distribution of short runs.
Short run high gray-level emphasis (SRHGLE)	Distribution of short runs of high intensity.
Short run low gray-level emphasis (SRLGLE)	Distribution of short runs of low intensity.

Table A3. First order Histogram textural features formulas.

Feature	Formula	
Skewness	$$\frac{\frac{1}{N_p}\sum_{i=1}^{N_p}\left(P(i)-\mu\right)^3}{\left(\sqrt{\frac{1}{N_p}\sum_{i=1}^{N_p}\left(P(i)-\mu\right)^2}\right)^3}$$	(A1)
Kurtosis	$$\frac{\frac{1}{N_p}\sum_{i=1}^{N_p}\left(P(i)-\mu\right)^4}{\left(\frac{1}{N_p}\sum_{i=1}^{N_p}\left(P(i)-\mu\right)^2\right)^2}$$	(A2)
Entropy	$$-\sum_{i=1}^{N_g} g(i)\log_2\left(g(i)+\epsilon\right)$$	(A3)

Table A4. Second order GLCM textural features and their associated formulas.

Feature	Formula	
Contrast	$\sum_{i=0}^{N_g} \sum_{j=0}^{N_g} (i-j)^2 CM(i,j)$	(A4)
Dissimilarity	$\sum_{i=0}^{N_g} \sum_{j=0}^{N_g} |i-j| CM(i,j)$	(A5)
Homogeneity	$\sum_{i=0}^{N_g} \sum_{j=0}^{N_g} \dfrac{CM(i,j)}{1+(i-j)^2}$	(A6)
ASM	$\sum_{i=0}^{N_g} \sum_{j=0}^{N_g} \left(CM(i,j)\right)^2$	(A7)
Energy	$\sqrt{ASM}$	(A8)
Correlation	$\dfrac{\sum_{i=0}^{N_g} \sum_{j=0}^{N_g} CM(i,j) ij - \mu_x \mu_y}{\sigma_x(i)\sigma_y(j)}$	(A9)

Table A5. Second order GLRLM textural features and their associated formulas.

Feature	Formula	
Gray-Level Non-Uniformity (GLN)	$\dfrac{\sum_{i=0}^{N_g} \left(\sum_{j=0}^{N-1} RL(i,j)\right)^2}{n_r}$	(A10)
High Gray-Level Run Emphasis (HGLRE)	$\dfrac{\sum_{i=0}^{N_g} \sum_{j=1}^{N-1} RL(i,j) i^2}{n_r}$	(A11)
Long Run Emphasis (LRE)	$\dfrac{\sum_{i=0}^{N_g} \sum_{j=0}^{N-1} RL(i,j) j^2}{n_r}$	(A12)
Long Run High Gray-Level Emphasis (LRHGLE)	$\dfrac{\sum_{i=0}^{N_g} \sum_{j=0}^{N-1} RL(i,j) i^2 j^2}{n_r}$	(A13)
Long Run Low Gray-Level Emphasis (LRLGLE)	$\dfrac{\sum_{i=0}^{N_g} \sum_{j=0}^{N-1} \frac{RL(i,j) j^2}{i^2}}{n_r}$	(A14)
low gray-level run emphasis (LGLRE)	$\dfrac{\sum_{i=0}^{N_g} \sum_{j=0}^{N-1} \frac{RL(i,j)}{i^2}}{n_r}$	(A15)
Run Entropy (RE)	$-\sum_{i=0}^{N_g} \sum_{j=0}^{N-1} RL(i,j) \log_2 RL(i,j) + \epsilon)$	(A16)
Run Length Non-Uniformity (RLN)	$\dfrac{\sum_{j=0}^{N-1} \left(\sum_{i=0}^{N_g} RL(i,j)\right)^2}{n_r}$	(A17)
Run Percentage (RP)	$\dfrac{n_r}{n_p}$	(A18)
Short Run Emphasis (SRE)	$\dfrac{\sum_{i=0}^{N_g} \sum_{j=0}^{N-1} \frac{RL(i,j)}{j^2}}{n_r}$	(A19)
Short Run High gray-level Emphasis (SRHGLE)	$\dfrac{\sum_{i=0}^{N_g} \sum_{j=0}^{N-1} \frac{RL(i,j) i^2}{j^2}}{n_r}$	(A20)
Short Run Low gray-level Emphasis (SRLGLE)	$\dfrac{\sum_{i=0}^{N_g} \sum_{j=0}^{N-1} \frac{RL(i,j)}{i^2 j^2}}{n_r}$	(A21)

Table A6. Details of the selected features for the final model.

		Higher Order 3D-Appearance Features
T2-MR (FLAIR)		81 HOG-Components + Volume
T1-MR (Pre-contrast)		58 HOG-Components
T1-MR (Post-contrast)		45 HOG-Components + Volume
		Textural Features
T2-MR (FLAIR)	**Histogram**	Mean (μ), Variance, Skewness, Kurtosis, Entropy, 6 CDFs-Components, 7 Percentiles-Components
	GLCM	Contrast, Dissimilarity, Homogeneity, ASM
	GLRLM	HGLRE, LRE, LRHGLE, LGLRE, RE, RLN, SRE, SRHGLE
T1-MR (Pre-contrast)	**Histogram**	Mean (μ), Variance, Skewness, Kurtosis, Entropy, 8 CDFs-Components, 8 Percentiles-Components
	GLCM	Contrast, Dissimilarity, Homogeneity, Energy, Correlation
	GLRLM	GLN, LRE, LRLGLE, LGLRE, RE, RLN, RP, SRE, SRLGLE
T1-MR (Post-contrast)	**Histogram**	Mean (μ), Variance, Skewness, Kurtosis, Entropy, 8 CDFs-Components, 9 Percentiles-Components
	GLCM	Contrast, Dissimilarity, Homogeneity, ASM, Energy, Correlation
	GLRLM	HGLRE, LRE, LRHGLE, LRLGLE, LGLRE, RE, RLN, SRHGLE, SRLGLE
		Functional features
DW-MR		43 CDFs-Components
T1-MR		Contrast-enhancement Slope

References

1. Cancer.Net Editorial Board. Brain Tumor: Statistics. 2022. Available online: https://www.cancer.net/cancer-types/brain-tumor/statistics (accessed on 16 September 2022).
2. Gutta, S.; Acharya, J.; Shiroishi, M.; Hwang, D.; Nayak, K. Improved Glioma Grading Using Deep Convolutional Neural Networks. *Am. J. Neuroradiol.* **2021**, *42*, 233–239. [CrossRef] [PubMed]
3. Qian, M.; Chen, Z.; Guo, X.; Wang, S.; Zhang, Z.; Qiu, W.; Qi, Y.; Zhang, S.; Xu, J.; Zhao, R.; et al. Exosomes derived from hypoxic glioma deliver miR-1246 and miR-10b-5p to normoxic glioma cells to promote migration and invasion. *Lab. Investig.* **2021**, *101*, 612–624. [CrossRef] [PubMed]
4. Ferlay, J.; Soerjomataram, I.; Dikshit, R.; Eser, S.; Mathers, C.; Rebelo, M.; Parkin, D.M.; Forman, D.; Bray, F. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* **2015**, *136*, E359–E386. [CrossRef]
5. Ostrom, Q.T.; Gittleman, H.; Farah, P.; Ondracek, A.; Chen, Y.; Wolinsky, Y.; Stroup, N.E.; Kruchko, C.; Barnholtz-Sloan, J.S. CBTRUS statistical report: Primary brain and central nervous system tumors diagnosed in the United States in 2006–2010. *Neuro-oncology* **2013**, *15*, ii1–ii56. [CrossRef] [PubMed]
6. Louis, D.N.; Perry, A.; Wesseling, P.; Brat, D.J.; Cree, I.A.; Figarella-Branger, D.; Hawkins, C.; Ng, H.; Pfister, S.M.; Reifenberger, G.; et al. The 2021 WHO classification of tumors of the central nervous system: A summary. *Neuro-oncology* **2021**, *23*, 1231–1251. [CrossRef] [PubMed]
7. McNamara, C.; Mankad, K.; Thust, S.; Dixon, L.; Limback-Stanic, C.; D'Arco, F.; Jacques, T.S.; Löbel, U. 2021 WHO classification of tumours of the central nervous system: A review for the neuroradiologist. *Neuroradiology* **2022**, *64*, 1919–1950. [CrossRef]
8. Wang, Q.; Lei, D.; Yuan, Y.; Zhao, H. Accuracy of magnetic resonance imaging texture analysis in differentiating low-grade from high-grade gliomas: Systematic review and meta-analysis. *BMJ Open* **2019**, *9*, e027144. [CrossRef] [PubMed]
9. National Brain Tumor Society. About Brain Tumor: Types and Statistics. 2022. Available online: https://braintumor.org/brain-tumors/about-brain-tumors/brain-tumor-types/astrocytoma/ (accessed on 16 September 2022).
10. Wesseling, P.; Capper, D. WHO 2016 Classification of gliomas. *Neuropathol. Appl. Neurobiol.* **2018**, *44*, 139–150. [CrossRef]
11. Miller, J.J.; Wick, W. What's New in Grade II and Grade III Gliomas? *Semin. Neurol.* **2018**, *38*, 041–049.
12. Iv, M.; Bisdas, S. Neuroimaging in the Era of the Evolving WHO Classification of Brain Tumors, From the AJR Special Series on Cancer Staging. *Am. J. Roentgenol.* **2021**, *271*, 3–15. [CrossRef]
13. Citak-Er, F.; Firat, Z.; Kovanlikaya, I.; Ture, U.; Ozturk-Isik, E. Machine-learning in grading of gliomas based on multi-parametric magnetic resonance imaging at 3T. *Comput. Biol. Med.* **2018**, *99*, 154–160. [CrossRef] [PubMed]
14. Razek, A.A.K.A.; El-Serougy, L.G.; Abdelsalam, M.A.; Gaballa, G.M.; Talaat, M.M. Multi-parametric arterial spin labelling and diffusion-weighted magnetic resonance imaging in differentiation of grade II and grade III gliomas. *Pol. J. Radiol.* **2020**, *85*, e110. [CrossRef] [PubMed]
15. Xie, Y.; Zaccagna, F.; Rundo, L.; Testa, C.; Agati, R.; Lodi, R.; Manners, D.N.; Tonon, C. Convolutional Neural Network Techniques for Brain Tumor Classification (from 2015 to 2022): Review, Challenges, and Future Perspectives. *Diagnostics* **2022**, *12*, 1850. [CrossRef]

16. Zhang, X.; Yan, L.F.; Hu, Y.C.; Li, G.; Yang, Y.; Han, Y.; Sun, Y.Z.; Liu, Z.C.; Tian, Q.; Han, Z.Y.; et al. Optimizing a machine learning based glioma grading system using multi-parametric MRI histogram and texture features. *Oncotarget* **2017**, *8*, 47816. [CrossRef]

17. Cho, H.h.; Park, H. Classification of low-grade and high-grade glioma using multi-modal image radiomics features. In Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju, Korea, 11–15 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 3081–3084.

18. Suárez-García, J.G.; Hernández-López, J.M.; Moreno-Barbosa, E.; de Celis-Alonso, B. A simple model for glioma grading based on texture analysis applied to conventional brain MRI. *PLoS ONE* **2020**, *15*, e0228972. [CrossRef] [PubMed]

19. Banerjee, S.; Mitra, S.; Masulli, F.; Rovetta, S. Glioma classification using deep radiomics. *SN Comput. Sci.* **2020**, *1*, 209. [CrossRef]

20. Alis, D.; Bagcilar, O.; Senli, Y.; Isler, C.; Yergin, M.; Kocer, N.; Islak, C.; Kizilkilic, O. The diagnostic value of quantitative texture analysis of conventional MRI sequences using artificial neural networks in grading gliomas. *Clin. Radiol.* **2020**, *75*, 351–357. [CrossRef] [PubMed]

21. Hsieh, K.L.C.; Chen, C.Y.; Lo, C.M. Quantitative glioma grading using transformed gray-scale invariant textures of MRI. *Comput. Biol. Med.* **2017**, *83*, 102–108. [CrossRef]

22. Kalaiselvi, T.; Padmapriya, T.; Sriramakrishnan, P.; Priyadharshini, V. Development of automatic glioma brain tumor detection system using deep convolutional neural networks. *Int. J. Imaging Syst. Technol.* **2020**, *30*, 926–938. [CrossRef]

23. Zhuge, Y.; Ning, H.; Mathen, P.; Cheng, J.Y.; Krauze, A.V.; Camphausen, K.; Miller, R.W. Automated glioma grading on conventional MRI images using deep convolutional neural networks. *Med. Phys.* **2020**, *47*, 3044–3053. [CrossRef]

24. Cho, H.h.; Lee, S.h.; Kim, J.; Park, H. Classification of the glioma grading using radiomics analysis. *PeerJ* **2018**, *6*, e5982. [CrossRef] [PubMed]

25. Zeng, Q.; Dong, F.; Shi, F.; Ling, C.; Jiang, B.; Zhang, J. Apparent diffusion coefficient maps obtained from high b value diffusion-weighted imaging in the preoperative evaluation of gliomas at 3T: Comparison with standard b value diffusion-weighted imaging. *Eur. Radiol.* **2017**, *27*, 5309–5315. [CrossRef] [PubMed]

26. Abdel Razek, A.A.K.; Alksas, A.; Shehata, M.; AbdelKhalek, A.; Abdel Baky, K.; El-Baz, A.; Helmy, E. Clinical applications of artificial intelligence and radiomics in neuro-oncology imaging. *Insights Into Imaging* **2021**, *12*, 152. [CrossRef] [PubMed]

27. Abdel Razek, A.A.K.; Talaat, M.; El-Serougy, L.; Gaballa, G.; Abdelsalam, M. Clinical applications of arterial spin labeling in brain tumors. *J. Comput. Assist. Tomogr.* **2019**, *43*, 525–532. [CrossRef] [PubMed]

28. Aggarwal, A.; Das, P.K.; Shukla, A.; Parashar, S.; Choudhary, M.; Kumar, A.; Kumar, N.; Dutta, S. Role of multivoxel intermediate TE 2D CSI MR spectroscopy and 2D echoplanar diffusion imaging in grading of primary glial brain tumours. *J. Clin. Diagn. Res. JCDR* **2017**, *11*, TC05. [CrossRef]

29. Chang, K.V.; Özçakar, L. What Can Artificial Intelligence Do for Pain Medicine? *Asia Pac. J. Pain* **2022**, *1*, 1.

30. Gao, W.; Zhang, X.; Yuan, J.; Li, D.; Sun, Y.; Chen, Z.; Gu, Z. Application of Medical Imaging Methods and Artificial Intelligence in Tissue Engineering and Organ-on-a-Chip. *Front. Bioeng. Biotechnol.* **2022**, *10*, 1557. [CrossRef]

31. Tabatabaei, M.; Razaei, A.; Sarrami, A.H.; Saadatpour, Z.; Singhal, A.; Sotoudeh, H. Current status and quality of machine learning-based radiomics studies for glioma grading: A systematic review. *Oncology* **2021**, *99*, 433–443. [CrossRef]

32. Menze, B.H.; Jakab, A.; Bauer, S.; Kalpathy-Cramer, J.; Farahani, K.; Kirby, J.; Burren, Y.; Porz, N.; Slotboom, J.; Wiest, R.; et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* **2014**, *34*, 1993–2024. [CrossRef]

33. Clark, K.; Vendt, B.; Smith, K.; Freymann, J.; Kirby, J.; Koppel, P.; Moore, S.; Phillips, S.; Maffitt, D.; Pringle, M.; et al. The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository. *J. Digit. Imaging* **2013**, *26*, 1045–1057. [CrossRef]

34. Altman, D.G.; Bland, J.M. Measurement in medicine: The analysis of method comparison studies. *J. R. Stat. Soc. Ser. D (Stat.)* **1983**, *32*, 307–317. [CrossRef]

35. McConnell, R.K. Method of and Apparatus for Pattern Recognition. US Patent 4,567,610, 28 January 1986.

36. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; IEEE: Piscataway, NJ, USA, 2005; Volume 1, pp. 886–893.

37. Soni, N.; Priya, S.; Bathla, G. Texture analysis in cerebral gliomas: A review of the literature. *Am. J. Neuroradiol.* **2019**, *40*, 928–934. [CrossRef] [PubMed]

38. Skogen, K.; Schulz, A.; Dormagen, J.B.; Ganeshan, B.; Helseth, E.; Server, A. Diagnostic performance of texture analysis on MRI in grading cerebral gliomas. *Eur. J. Radiol.* **2016**, *85*, 824–829. [CrossRef] [PubMed]

39. Haralick, R.M. Statistical and structural approaches to texture. *Proc. IEEE* **1979**, *67*, 786–804. [CrossRef]

40. Gallowy, M. Texture analysis using gray level run length. *Comput. Graph Image Process* **1975**, *4*, 172–179. [CrossRef]

41. Alksas, A.; Shehata, M.; Saleh, G.A.; Shaffie, A.; Soliman, A.; Ghazal, M.; Khelifi, A.; Khalifeh, H.A.; Razek, A.A.; Giridharan, G.A.; et al. A novel computer-aided diagnostic system for accurate detection and grading of liver tumors. *Sci. Rep.* **2021**, *11*, 13148. [CrossRef]

42. Alksas, A.; Shehata, M.; Saleh, G.A.; Shaffie, A.; Soliman, A.; Ghazal, M.; Khalifeh, H.A.; Razek, A.A.; El-Baz, A. A novel computer-aided diagnostic system for early assessment of hepatocellular carcinoma. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10-15 January 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 10375–10382.

43. Shehata, M.; Alksas, A.; Abouelkheir, R.T.; Elmahdy, A.; Shaffie, A.; Soliman, A.; Ghazal, M.; Abu Khalifeh, H.; Salim, R.; Abdel Razek, A.A.K.; et al. A comprehensive computer-assisted diagnosis system for early assessment of renal cancer tumors. *Sensors* **2021**, *21*, 4928. [CrossRef]

44. Shehata, M.; Alksas, A.; Abouelkheir, R.T.; Elmahdy, A.; Shaffie, A.; Soliman, A.; Ghazal, M.; Khalifeh, H.A.; Razek, A.A.; El-Baz, A. A New Computer-Aided Diagnostic (CAD) System for Precise Identification of Renal Tumors. In Proceedings of the 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), Nice, France, 13–16 April 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1378–1381.

45. Ayyad, S.M.; Badawy, M.A.; Shehata, M.; Alksas, A.; Mahmoud, A.; Abou El-Ghar, M.; Ghazal, M.; El-Melegy, M.; Abdel-Hamid, N.B.; Labib, L.M.; et al. A New Framework for Precise Identification of Prostatic Adenocarcinoma. *Sensors* **2022**, *22*, 1848. [CrossRef]

46. Basheer, I.A.; Hajmeer, M. Artificial neural networks: Fundamentals, computing, design, and application. *J. Microbiol. Methods* **2000**, *43*, 3–31. [CrossRef]

47. Cohen, J. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychol. Bull.* **1968**, *70*, 213. [CrossRef]