Mobile Networks and Applications
https://doi.org/10.1007/s11036-022-02021-6

Numerical Grad-Cam Based Explainable Convolutional Neural Network for Brain Tumor Diagnosis

Jose Antonio Marmolejo-Saucedo[1] · **Utku Kose**[2]

Accepted: 23 May 2022

Abstract
Since the start of the current century, artificial intelligence has gone through critical advances improving the capabilities of intelligent systems. Especially machine learning has changed remarkably and caused the rise of deep learning. Deep learning shows cutting-edge results in terms of even the most advanced, difficult problems. However, that includes a trade-off in terms of interpretability. Although traditional machine learning techniques employ interpretable working mechanisms, hybrid systems and deep learning models are black-box being beyond our understanding capabilities. So, the need for making such systems understandable, additional methods by explainable artificial intelligence (XAI) has been widely developed in last years. In this sense, this study purposes a Convolutional Neural Networks (CNN) model, which runs a new form of Grad-CAM. As providing numerical feedback in addition to the default Grad-CAM, the numerical Grad-CAM (numGrad-CAM) was used within the developed CNN model, in order to have an explainability interface for brain tumor diagnosis. In detail, the numGrad-CAM-CNN model was evaluated via technical and physicians-oriented (human-side) evaluations. The model provided average findings of 97.11% accuracy, 95.58% sensitivity, and 96.81% specificity for the target brain tumor diagnosis setup. Additionally, numGrad-CAM integration provided 90.11% accuracy according to the other CAM variations in the same CNN model. The physicians used the numGrad-CAM-CNN model gave positive responses in terms of using the model for an explainable (and safe) diagnosis decision-making perspective for brain tumors.

Keywords Grad-cam · Explainable artificial intelligence · CNN · Deep learning · Brain tumor · Medical diagnosis

1 Introduction

Today's modern, brave world is rising over many advanced technological advances. As a result of different industrial phases triggering transformation of both technological tools and the societies, the current world has reached to the era of intelligent systems [1, 2]. It is remarkable that the field of artificial intelligence passed through many different development processes since its first introduction to the scientific audience. Although there are different methods and techniques used in the context of artificial intelligence field, machine learning is known as the main actor on the background of successful applications [3, 4]. Machine learning algorithms, which are essentially based on the logic of providing adaptation to the target training data by optimizing various parameters inside the model, have been affected by intense advances within hardware and software technologies, seen in especially last twenty years [5–7]. More effective techniques having better capabilities to process more challenging data are now accepted under the name of new machine learning: deep learning [8, 9]. For now, neural network-oriented techniques build the main characteristics of deep learning. It is clear that the future has the potential of newer deep learning models out of the neural network formation but the deep learning has been shaped with advanced

✉ Jose Antonio Marmolejo-Saucedo
jmarmolejo@up.edu.mx

Utku Kose
utkukose@sdu.edu.tr

[1] Facultad de Ingenieria, Universidad Panamericana, Augusto Rodin 498, 03920 Ciudad de Mexico, Mexico

[2] Suleyman Demirel University, Isparta, Turkey

neural network models as a result of the needs of different problem areas and alternative solution mechanisms. Here, Convolutional Neural Networks (CNN) has a remarkable popularity among all deep learning techniques with its successful applications in different fields [10, 11]. Various CNN models have been used effectively in image-based problems, especially thanks to integrated feature extraction mechanisms and layer structures designed primarily based on image data [12, 13]. Because of that, CNN models are often used in critical areas such as healthcare, which have been closely linked with artificial intelligence through the historical advances. Among many other healthcare-oriented problem areas, disease diagnosis is widely employing CNN models to get successful outcomes. CNN models are often reported as they provide more sensitive, early detections when compared to physicians [14–16].

Nowadays, CNN is among top intelligent tools to deal with disease diagnosis problems. Although there are alternative deep learning models, CNN has already shown its critical role in especially medical image-based diagnosis problems [17–19]. It seems that CNN and deep learning era takes the outcomes by hybrid machine learning formations some more steps away. However, when examined under the human-side understanding capabilities, it is often discussed that the solution mechanisms by both hybrid machine learning formations and the recent deep learning models are black-box, meaning that we cannot control well enough the decision-making mechanism between input data and the outcomes [20–23]. When we consider the CNN model, the total number of parameters reaching to hundreds and even thousands cause us to lose our tracking control among different mathematical calculations. At this point, it is not possible to understand safety level of the decisions made by the CNN-based intelligent system. Furthermore, it is also not possible to catch any errors as well as bias or understand success and fail borders of the used system [24, 25]. This situation affecting not only present state of the artificial intelligence field but also future advances of the upcoming intelligent systems has enabled researchers to search for effective solutions. Nowadays, these solutions include using integration of interpretable machine learning techniques and building new mathematical components to make black-box models transparent (or white-box). The efforts so far caused a new research area: explainable artificial intelligence (XAI) to rise, and that area has been effective for designing alternative models of deep learning techniques, building explainability interfaces for medical diagnosis applications [26, 27].

Based on the explanations so far, the main objective of this study is to purpose an explainable CNN for brain tumor diagnosis problem. As it is known, CNN models have been often used in diagnosis from medical image data including X-Ray, CT, MRI, ultrasound…etc. [12, 18, 19, 28–32]. For supporting explainability of the verbal decision-makings from such medical images, Class Activation Mapping (CAM) method is used for creating heat maps over input medical image and showing which regions are detected by the model during detection (diagnosis). [33, 34]. The CAM includes different variations currently and the Grad-CAM is among these. However, according to the authors, the Grad-CAM is still open for further updates. So, this study aimed to improve it and use with the CNN, for a remarkable diagnosis problem. The target problem was chosen as brain tumor diagnosis, as it has gained momentum in recent years [35–37].

Pointing the main objective, the motivations of the study can be expressed briefly as follows:

- Building a successful enough CNN model for brain tumor diagnosis over a dataset from the literature.
- Ensuring XAI touch for a CNN model solving brain tumor diagnosis.
- Improving the Grad-CAM by adding numerical explanations over the heat map outputs. That improved Grad-CAM method was named as numGrad-CAM.
- Performing critical evaluations with the physicians to see if the developed numGrad-CAM-CNN model is successful and safe enough for the brain tumor diagnosis problem.

Considering the objective of the study, and the problem scope, the next sections are organized as follows: In the next section, the whole components (CNN, dataset, XAI addition…etc.) in the context of the performed study are explained accordingly. Following to that, the third section provides information regarding the performed application and the obtained findings. Next, the fourth section ensures a general discussion considering outcomes, limitations and any future work suggestions. Finally, the content is ended by discussions regarding conclusions and future work plans thought by the authors.

2 Components and the Solution Approach

This section refers to some essential explanations regarding the main components of the study and the general problem-solving approach. Moving from that, the following paragraphs include explanations for the used CNN model, numGrad-CAM for the XAI touch, and the general application flow, which was done at the heart of the research aiming to achieve both successful and safe diagnosis for brain tumors.

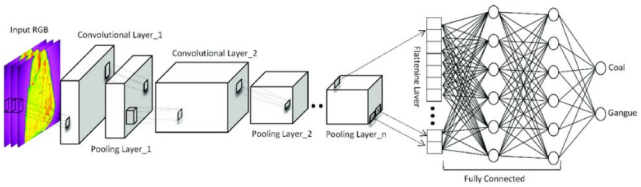

Fig. 1 Typical model structure for the CNN technique [39]

2.1 Convolutional Neural Networks

Convolutional Neural Networks (CNN) is a deep learning technique, which is known best with its applications in image-based data [12–15]. Unlike the artificial neural networks (ANN) technique in the context of traditional machine learning, the CNN models consist of special layers that ensures the mechanism of feature extraction from input data [11, 13, 38]. A typical CNN model, which is ended-up with a fully connected layer structure, may include different type layers such as convolutional layer, pooling layer, and flattening layer [39] (Fig. 1). The essential philosophy of the CNN technique is based on a kind of numerical inspiration from visual perception processes in living organisms, which perceive visual objects in the real world, thanks to feature recognitions. In this context, the convolutional layer allows detection of alternative features from the input data by applying different filters, the pooling layer reduces the detected features to an optimum level, and the flattening layer prepares the appropriate data structure for the fully connected layer, which corresponds to a typical ANN model flow [11–13, 38]. Thus, a neural network model, which does not require any additional feature extraction (as done for the traditional ANN models), is built in the context of CNN model. As different, the CNN model is better in analysis of the input data in detail [12–14]. Due to the inside mechanisms of the CNN models, the training processes take much more time. But because of remarkably better outcomes and the current computing technologies reduce the processing time to reasonable levels, the CNN has become one of the most important techniques that make deep learning widespread.

CNN models run effective mechanisms in terms of reducing computational complexity and producing more successful solutions for more challenging data (especially for image data), as compared to traditional ANN models. However, that process also leads to more comprehensive mathematical flows that require too many parameters to be optimized during the training stage. That causes CNN models to be black-box because it is too difficult to analyze the exact connections between input and output data (More parameters cause complicated mathematical and logical connections making the connections blurred). Considering the field of healthcare, physicians and radiologists can

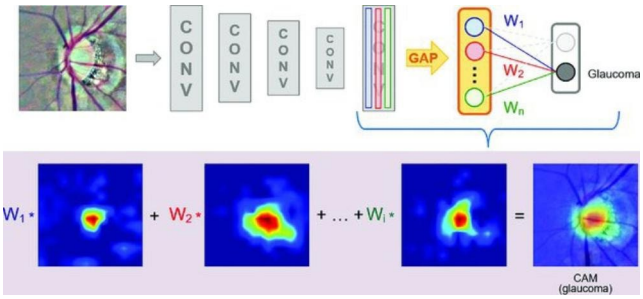

Fig. 2 Use of the CAM method for a medical diagnosis problem [41]

explain how they reach to the diagnosis / detection through a medical image. Similarly, a deep learning model such as CNN should be providing similar feedback to show that it is safe enough as an automated decision-making tool. Such a solution will also enable people to take the necessary actions by understanding the limitations of a deep learning oriented intelligent system. In order to meet this requirement, various explainable artificial intelligence (XAI) methods have been developed to be integrated into CNN models. The Class Activation Mapping (CAM) method is known as one of these XAI solutions.

2.2 From Grad-CAM to the Numerical Grad-CAM

In this study, the default Grad-CAM method was improved with feedback mechanism through numerical values. In order to understand that, the methods of default Class Activation Mapping (CAM), and Gradient-weighted CAM (Grad-CAM) should be both introduced briefly.

2.2.1 Default CAM method

The CAM allows weighted feature maps (in the deep learning model) to be gathered together in line with different class results to create activation maps [33, 34, 40]. Eventually, these maps are visualized to the user side with a heat map interface where the intense (red color) visual regions point more focused pixels by the model (so less focused regions are with lighter colors towards green and / or blue). Figure 2 represents an applied CAM for a Glaucoma diagnosis application [41].

2.2.2 Grad-CAM and the numGrad-CAM methods

Several variations of the CAM method were developed in time. Gradient-weighted CAM (Grad-CAM) is one of these variations as including differences from the default CAM method. The CAM requires global average pooling (GAP) to create feature maps. At the final stage, the weights for feature maps are set thanks to the fully-connected layer. As different, Grad-CAM weights the maps by using alpha

values from gradients. In this way, Grad-CAM has a more architecture-independent solution according to the (default) CAM [42, 43]. The Grad-CAM goes through three steps to get the resulting heat maps [44]:

- **Step 1**: Considering a convolutional layer and each class in the model (c), the gradient of the class (y_c) related to feature map activations (A^k) are calculated: $\frac{\partial y_c}{\partial A^k}$
- **Step 2**: Alpha values for the class c, and the map k were calculated with global average pooling and the gradients (Step 1):

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y_c}{\partial A^k} \tag{1}$$

- **Step 3**: In order to get final heat map, a weighted sum of the feature maps is calculated (considering the ReLU function):

$$L_{Grad\_CAM}^c = ReLU\left(\sum_k \alpha_k^c A^k\right) \tag{2}$$

Although Grad-CAM comes with an architecture-independent solution with heat maps to show effects of different image regions on the output, there is still open ways to improve capabilities. For example, the heat map can be supported with numerical explanations in order to improve explanation ground for humans. This may be also a faster explainability approach to improve capabilities of the known CAM and Grad-CAM methods. This study added a new calculation step to divide heat rates in different (additional) detection regions to reflect how strong each different region supports the associated explanation outcome (Each region corresponds to the value of 100, so sum of different focus level values in a single region is 100). Because there is a total of k feature maps as default, the rates were shown in user-defined, x different focus areas (x_area). For example, considering three focus areas, the corresponding focus areas may be classified as low, medium, high. That is actually done by finding average gradient for x focus areas and normalizing each value (Eq. 3). As a visual mechanism, the values are located over the heat maps by using x circles with the rates.

For each numGrad-CAM focus area value:

$$norm\left(\frac{\sum_j \frac{\partial y_c}{\partial A^k}}{x\_area}\right) \tag{3}$$

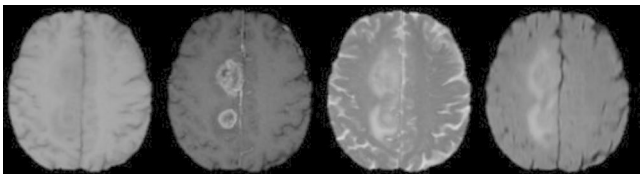

Fig. 3 Sample brain MRI images from the BRATS 2017 dataset [45, 46]

2.3 Brain Tumor Diagnosis with NumGrad-CAM-CNN and BRATS Dataset

In order to set-up a remarkable research process, this study aimed to diagnose brain tumors. As the dataset, the study included 2017 version of the BRATS dataset, which was used in the Multimodal Brain Tumor Segmentation Challenge [45, 46]. The dataset briefly consists of 285 MRI images with image scans regarding each patient / person are presented with T1, T1ce, T2 and Flair brain images (Fig. 3). In the study, a total of 200 images from the BRATS 2017 dataset were used accordingly. By combining numGrad-CAM and the CNN model, 170 MRI images were used during the training stage while the remaining 30 images were included in the test stage. The performance of the numGrad-CAM-CNN model was evaluated with average findings from 20 independent runs.

The designed numGrad-CAM-CNN model was organized by considering the following details:

- There is a total of 7 layers to build the whole CNN model.
- The first layer was arranged as the input layer, and the second layer was created as a convolution process using 8 convolution filters.
- The convolution layer was combined with the pooling layer (4×4 in size).
- The pooling layer was supported by a maximum pooling layer.
- The fourth layer was created as a convolution layer (employing 10 filters).
- The fifth layer was created as a pooling layer with maximum pooling, with a 3×3 dimension.
- The CNN model was ended with a fully connected layer of 75 artificial neurons.
- XAI mechanism of the CNN was supported with the numGrad-CAM, as proposed in this study.

3 Brain Tumor Diagnosis Applications

After development of the numGrad-CAM-CNN model, the purposed solution approach was evaluated with the brain

tumor diagnosis applications. The applications included use of the BRATS 2017 dataset and evaluations were done by considering technical performance of the model and the human-side feedback. The related application phase and the evaluation works are explained under the following sub-sections:

3.1 Applications and the Evaluations

As it was mentioned before, 170 MRI images (from the BRATS 2017 dataset) were used during the training stage while the remaining 30 images were included in the test stage. Before passing to the human-side evaluations, technical performance of the numGrad-CAM-CNN model was evaluated with average findings from 20 independent runs. Here, some similar neural network-oriented models were also employed for a general comparative evaluation. The performance was measured by using three metrics frequently employed in similar works [47, 48]:

$$Accuracy = \frac{\text{True\_Positives} + \text{True\_Negatives}}{\text{True\_Positives} + \text{False\_Positives} + \text{True\_Negatives} + \text{False\_Negatives}} \quad (4)$$

$$Sensitivity = \frac{\text{True\_Positives}}{\text{True\_Positives} + \text{False\_Negatives}} \quad (5)$$

$$\text{Specifitivty} = \frac{\text{True\_Negatives}}{\text{True\_Negatives} + \text{False\_Positives}} \quad (6)$$

In addition to the technical performance evaluation, the accuracy findings regarding integration to the CNN model respectively with the new numGrad-CAM, Default CAM [33, 34], (Default) GradCAM [44], GradCAM++ [49], and Score-CAM [50] were all compared.

Except from the technical evaluations, the human-side evaluations were done with a total of 15 physicians. Firstly, the physicians were asked to evaluate the heat maps created by the numGrad-CAM-CNN model against the related brain images. Secondly, they were asked to provide feedback to some survey statements regarding the proposed solution / model in terms of performance and explainability. Finally, they were also asked to evaluate the numGrad-CAM with other CAM alternatives over the same CNN model and the BRATS 2017 dataset setup.

3.2 Obtained Findings

The developed numGrad-CAM-CNN model was run through 20 independent runs and compared with some other models (from the literature), which were previously used the BRATS 2017 dataset. The chosen models were kept as designed originally from their works but the evaluation for them was done at the same training-test data separation as well as 20 independent runs. In this comparative evaluation, the XAI perspective by the numGrad-CAM was not evaluated but a successful enough performance by the CNN-oriented model was awaited to pass further steps of XAI evaluations. Table 1 provides average findings from the comparative evaluation (The best values are in bold style).

Table 1 Average technical performance findings in the context of the comparative evaluation

Model	Accuracy*	Sensitivity*	Specificity*
numGrad-CAM-CNN *(This Study)*	97.11%	95.58%	96.81%
Cascaded CNN [51]	95.70%	92.61%	94.80%
Dense Fully CNN [52]	96.26%	93.19%	94.67%
Densely Connected 3D CNN [53]	97.33%	**95.79%**	97.07%
Hybrid CNN-kNN [54]	92.19%	92.67%	94.09%
Parallel Deep CNN [55]	95.66%	92.19%	94.90%
Multimodal PixelNet [56]	94.83%	93.17%	93.11%
DCT-CNN-ELM-PLS [57]	**98.22%**	95.71%	**97.19%**

* The best values are in bold style.

Table 2 Comparative findings regarding different variations of the CAM method

Method	Accuracy*
numGrad-CAM integrated CNN (This Study)	92.11%
Default CAM [33, 34] integrated CNN	90.57%
(Default) Grad-CAM [44] integrated CNN	92.06%
GradCAM++ [49] integrated CNN	**93.48%**
Score-CAM [50] integrated CNN	93.36%

* The best value is in bold style

As it can be seen from Table 1, the numGrad-CAM-CNN model was not the best solution but the findings were in top three, as dominating the rest of five models. So, further steps of XAI evaluation were taken as a result of positive outcomes by the numGrad-CAM-CNN model for the designed BRATS 2017 applications.

As the other technical evaluation, different CAM method variations (Default CAM, GradCAM, GradCAM++, and Score-CAM) were integrated into the same CNN model and run for the target test images. In order to get a comparative finding, accuracy of each method (integration) was compared. The accuracy was calculated in the test brain images including tumors and by comparing heat maps over the tumors with the average tumor selections (average pixel points) done by the physicians over the medical imaging software. Table 2 provides the obtained findings in this manner (The best value is in bold style).

Table 2 proves that the proposed numGrad-CAM method ensures near accuracy to the GradCAM++, and Score-CAM, which both seem better than the other Default CAM, and GradCAM methods. That's because the main

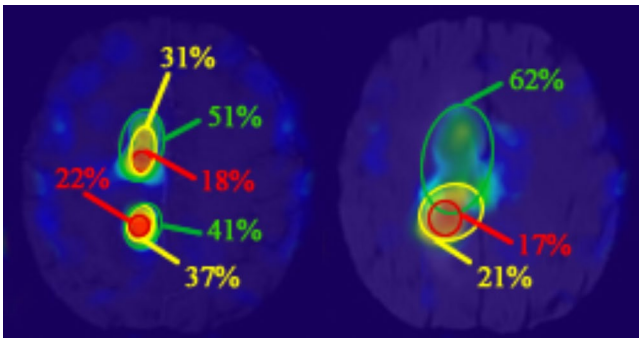

Fig. 4 XAI views by the proposed numGrad-CAM method inside the CNN model

Table 3 Scores by the physicians for the heat map views over the 30 test brain images

Physician No.	Correct (3)	No-Idea (2)	Wrong (1)	Average Score
1	20	6	4	5.60
2	23	4	3	5.73
3	25	3	2	5.80
4	26	3	1	5.80
5	25	3	2	5.80
6	22	3	5	5.80
7	22	5	3	5.67
8	21	5	4	5.67
9	25	3	2	5.80
10	27	2	1	5.87
11	28	1	1	5.93
12	21	4	5	5.73
13	22	5	3	5.67
14	24	3	3	5.80
15	26	2	2	5.87

A total of 30 brain images was evaluated by the physicians.

improvement of the numGrad-CAM is with the addition of numerical views since the background actually comes from the (Default) Grad-CAM. At this point, a better evaluation was done with the direct evaluation by the physicians, as explained under the next paragraphs.

In the context of the first human-side evaluation, the heat maps created by the numGrad-CAM-CNN model, for 30 test brain images, were scored by 15 physicians. They were wanted to score each image views by considering three different scales: correct (3), no-idea (2), wrong (1). According to the scores, the average feedback scores were considered in order to evaluate success of the numGrad-CAM. Figure 4 represents sample XAI views by the numGrad-CAM (with three focus areas), and Table 3 shows the findings obtained from the scorings by the physicians.

As the Fig. 4 shows, the new numGrad-CAM method successfully focuses on especially tumor regions and reflects numerical values by improving the explainability view for humans. Additionally, Table 3 points that the numGrad-CAM-CNN model focuses on the true regions (of the MR images) successful enough, according to the average scores by each physician.

The second evaluation was with a total of eight survey statements asking feedback about the performance and the explainability of the numGrad-CAM-CNN model. Table 4 provides details regarding the survey statements and the received feedback on the Likert Scale.

As the Table 4 shows, the physicians were positive about performance and explainability aspects of the developed numGrad-CAM-CNN model. They found diagnosis performance of the model successful enough for both detection and speed level. Additionally, they found XAI addition by the numGrad-CAM effective and even thought using the model for alternative diagnosis applications.

The physicians were also asked to give feedback for the use of different CAM variations (including the proposed numGrad-CAM), as an evaluation for their XAI support. That was done by asking three questions and getting responses for each of them. Table 5 represents the findings for that final evaluation work.

According to the majority of the physicians (as Table 5 shows), the XAI approach by the numGrad-CAM is better than the others for ensuring a good, stable and easy to understand view for the brain tumor problem in this study.

Table 4 Findings for the survey asking about the performance and the explainability of the numGrad-CAM-CNN model

St. No.	Statement	Totally Agree (5)	Agree (4)	No-Idea (3)	Disagree (2)	Totally Disagree (1)	Average
1	"The diagnosis performance of the model for brain tumors is acceptable."	9	4	2	0	0	4.47
2	"The diagnosis speed of the model for brain tumors is acceptable."	8	4	2	1	0	4.27
3	"I think this model is safe enough to be used for decision-making applications."	10	3	2	0	0	4.53
4	"I do not want to use this model for brain tumor diagnosis."	0	0	1	5	9	1.47
5	"The model provides a good explainability for understanding its safety."	10	3	2	0	0	4.53
6	"The numerical views over the heat maps improve the XAI perspective."	11	2	2	0	0	4.60
7	"I think this model should be used for alternative diagnosis problems."	8	4	2	1	0	4.27
8	"By using this model, I can diagnose brain tumors and track the safety of the solution mechanism."	10	1	3	1	0	4.33

Table 5 Findings for the questions asked about comparison of the CAM methods

Physician No.	Which CAM method had a better explainable view?	Which CAM method did you find more stable?	Which CAM method was easier to understand about the solution by CNN model?
1	numGrad-CAM	numGrad-CAM	numGrad-CAM
2	Score-CAM	Score-CAM	GradCAM++
3	Score-CAM	Score-CAM	Score-CAM
4	numGrad-CAM	(Default) Grad-CAM	numGrad-CAM
5	GradCAM++	GradCAM++	numGrad-CAM
6	numGrad-CAM	numGrad-CAM	numGrad-CAM
7	numGrad-CAM	numGrad-CAM	numGrad-CAM
8	Score-CAM	(Default) Grad-CAM	numGrad-CAM
9	numGrad-CAM	Score-CAM	(Default) Grad-CAM
10	GradCAM++	GradCAM++	GradCAM++
11	(Default) Grad-CAM	(Default) Grad-CAM	numGrad-CAM
12	numGrad-CAM	numGrad-CAM	numGrad-CAM
13	numGrad-CAM	numGrad-CAM	numGrad-CAM
14	GradCAM++	Score-CAM	numGrad-CAM
15	numGrad-CAM	Score-CAM	Score-CAM

The closest competitive method was Score-CAM, and the Default CAM could not find a place in three different responses by the physicians.

4 Discussion

According to the findings obtained for the evaluation of numGrad-CAM-CNN model, it was found that the brain tumor diagnosis can be done successfully with a direct CNN model. Furthermore, such model can be supported with an improved XAI method, in order to support medical image-based applications. As it is too critical to ensure safe and explainable enough Deep Learning solutions for today's advanced medical diagnosis problems, CAM method and its variations may give an effective view for image-based applications. This study proves that it is needed to employ model integrated methods to provide visual feedback over input image-data. In detail, the study also proves that such a model of numGrad-CAM-CNN can be effectively, and efficiently used specifically for brain tumor diagnosis. The technical evaluations showed three different evaluation findings for the developed CNN model, and all these findings were above 95%, which is a good rate for such a challenging problem associated with finding tumors in MRI images. The study used the BRATS 2017 dataset for the research purpose. This dataset is a widely used component for developing Machine / Deep Learning models to diagnose brain tumors. Of course, that's also a limitation for the current study that further studies may be done with alternative brain tumor datasets. Also, alternative Deep Learning models may be used as this study used only CNN technique. It is thought that alternative models may be designed with bigger CNN models or hybrid solutions including more Deep Learning techniques or synthesis of traditional Machine Learning techniques with Deep Learning ground.

When the study is thought in terms of provided XAI perspective with the numGrad-CAM, it is remarkable that the study provided recent findings to prove that brain tumor diagnosis can be made safely for automated decision-making. The physicians took part in this study generally showed positive feedback for using CAM method (including all variations) to get some automated solution acting like them (i.e. looking at some regions of MR data, focusing more on the regions with tumors). It is also remarkable that the improvement (numerical views) over the CAM / Grad-CAM made the new numGrad-CAM a better alternative for the physicians. Human-side evaluations showed that the improvement in this study affected the XAI approach of the CAM better.

The study has a limitation with the application on only brain tumor diagnosis but the findings show that different medical diagnosis applications through alternative medical images can be done by using the same numGrad-CAM method. As the numGrad-CAM method provides additional numerical views, it became a better alternative for the physicians in this study. That may be tested in further studies and the formed numGrad-CAM-CNN model can be used widely in the context of mobile devices, and Internet of Things (IoT) environment, by employing cloud-oriented components. So, further research may include more number of physicians, experts, and medical staff to gather more data on evaluation.

5 Conclusions and Future Work

This study proposed an explainable Convolutional Neural Networks (CNN) model, which is using a new Grad-Class Activation Mapping (CAM) method for the brain tumor diagnosis problem. In detail, the widely used CNN technique was built with specific architecture model to ensure good enough diagnosis for the BRATS 2017 dataset, and the explainability level of the model was improved thanks to the explainable artificial intelligence (XAI) perspective done via numerically updated Grad-CAM. Called as numGrad-CAM, the updated method provides an additional visual way to show numerical focus levels of each heat map regions. As forming the numGrad-CAM-CNN model accordingly, the developed approach was applied for the brain tumor diagnosis problem (considering BRATS 2017 dataset), and the model was evaluated in terms of both technical and explainability perspective. In addition to the technical, comparative evaluations for the diagnosis process, the model was analyzed by physicians to see if it is acceptable for decision-making support. The physicians attended to several tests to give feedback for explainability capabilities of the numGrad-CAM-CNN model and wanted to decide if it is better than alternative CAM integrations. Physicians generally think that the model is successful at tumor brain diagnosis and providing enough information to understand if its outcomes are as a chance or safe enough for further applications. It seems also that the numerical information provided over the heat map (numGrad-CAM) images make it easier to track effects of heat map regions over the focus points by the CNN model.

According to the obtained findings, the developed solution provided positive outcomes regarding the brain tumor diagnosis targeted in this study. Of course, there are also limitations considering open scope of brain tumor datasets and deep learning models to develop. As a result of the positive state caught in this study, the authors have already planned to go towards the mentioned limitations in future works. Also, future works include integration of the numGrad-CAM-CNN model to different platforms (i.e. mobile platforms / devices) and different diagnosis applications (i.e. cancer, eye diseases). There are also more future works including use of different XAI methods and compare them with the built numGrad-CAM method. Finally, more works with physicians, medical staff and alternative human-based tests will be done for the same numGrad-CAM-CNN model formation.

Research Data Policy and Data Availability Statements All data generated or analysed during this study are included in this published article (and its supplementary information files).

Compliance with Ethical Standards

Conflict of Interest The authors declare that there is no conflict of interest.

Competing Interests The authors did not receive support from any organization for the submitted work. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. West DM, Allen JR (2020) Turning Point: Policymaking in the Era of Artificial Intelligence. Brookings Institution Press
2. Li D, Du Y (2017) Artificial Intelligence with Uncertainty. CRC Press
3. Janiesch C, Zschech P, Heinrich K (2021) Machine learning and deep learning. Electron Markets 31(3):685–695
4. Kose U, Watada J, Deperlioglu O, Saucedo JAM (2022) Computational Intelligence for COVID-19 and Future Pandemics. Springer
5. Plasek A (2016) On the cruelty of really writing a history of machine learning. IEEE Ann Hist Comput 38(4):6–8
6. Ouyang W, Mueller F, Hjelmare M, Lundberg E, Zimmer C (2019) ImJoy: an open-source computational platform for the deep learning era. Nat Methods 16(12):1199–1200
7. Zhang Y, Ni Q (2020) Recent advances in quantum machine learning. Quantum Eng 2(1):e34
8. Kelleher JD (2019) Deep Learning. MIT press
9. Dargan S, Kumar M, Ayyagari MR, Kumar G (2020) A survey of deep learning and its applications: a new paradigm to machine learning. Arch Comput Methods Eng 27(4):1071–1092
10. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, Farhan L (2021) Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. J Big Data 8(1):1–74
11. Dhillon A, Verma GK (2020) Convolutional neural network: a review of models, methodologies and applications to object detection. Progress in Artificial Intelligence 9(2):85–112
12. Ahmed KB, Goldgof GM, Paul R, Goldgof DB, Hall LO (2021) Discovery of a generalization gap of convolutional neural networks on COVID-19 X-rays classification. IEEE Access 9:72970–72979
13. Sharma N, Jain V, Mishra A (2018) An analysis of convolutional neural networks for image classification. Procedia Comput Sci 132:377–384
14. David DS, Saravanan D, Jayachandran A (2020) Deep Convolutional Neural Network based Early Diagnosis of multi class brain tumour classification system. Solid State Technology 63(6):3599–3623
15. Xu S, Liu C, Zong Y, Chen S, Lu Y, Yang L, Zhang C (2019) An early diagnosis of oral cancer based on three-dimensional convolutional neural networks. IEEE Access 7:158603–158611
16. Janghel RR, Rathore YK (2021) Deep convolution neural network based system for early diagnosis of Alzheimer's disease. IRBM 42(4):258–267
17. Mohapatra S, Swarnkar T, Das J (2021) Deep convolutional neural network in medical image processing. Handbook of Deep Learning in Biomedical Engineering. Academic Press, pp 25–60
18. Kose U, Alzubi J (2021) Deep Learning for Cancer Diagnosis. Springer
19. Sarvamangala DR, Kulkarni RV (2021) Convolutional neural networks in medical image understanding: a survey.Evolutionary Intelligence,1–22
20. Gaur M, Faldu K, Sheth A (2021) Semantics of the black-box: Can knowledge graphs help make deep learning systems more interpretable and explainable? IEEE Internet Comput 25(1):51–59

21. Yang G, Ye Q, Xia J (2022) Unbox the black-box for the medical explainable ai via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond. Inform Fusion 77:29–52
22. Deperlioglu O, Kose U, Gupta D, Khanna A, Giampaolo F, Fortino G (2022) Explainable framework for Glaucoma diagnosis by image processing and convolutional neural network synergy: Analysis with doctor evaluation. Future Generation Computer Systems 129:152–169
23. Kenny EM, Ford C, Quinn M, Keane MT (2021) Explaining black-box classifiers using post-hoc explanations-by-example: The effect of explanations and error-rates in XAI user studies. Artif Intell 294:103459
24. Arrieta AB, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, Herrera F (2020) Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inform Fusion 58:82–115
25. Meske C, Bunde E, Schneider J, Gersch M (2022) Explainable artificial intelligence: objectives, stakeholders, and future research opportunities. Inform Syst Manage 39(1):53–63
26. Gunning D, Stefik M, Choi J, Miller T, Stumpf S, Yang GZ (2019) XAI—Explainable artificial intelligence. Sci Rob 4(37):eaay7120
27. Angelov PP, Soares EA, Jiang R, Arnold NI, Atkinson PM (2021) Explainable artificial intelligence: an analytical review.Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 11(5), e1424
28. Dong Y, Pan Y, Zhang J, Xu W (2017), July Learning to read chest X-ray images from 16000 + examples using CNN. In *2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)* (pp. 51–57). IEEE
29. Fu Y, Mazur TR, Wu X, Liu S, Chang X, Lu Y, Yang D (2018) A novel MRI segmentation method using CNN-based correction network for MRI-guided adaptive radiotherapy. Med Phys 45(11):5129–5137
30. Rajagopalan N, Narasimhan V, Vinjimoor K, Aiyer J (2021) Deep CNN framework for retinal disease diagnosis using optical coherence tomography images. J Ambient Intell Humaniz Comput 12(7):7569–7580
31. Lei Y, He X, Yao J, Wang T, Wang L, Li W, Yang X (2021) Breast tumor segmentation in 3D automatic breast ultrasound using Mask scoring R-CNN. Med Phys 48(1):204–214
32. Kundu R, Basak H, Singh PK, Ahmadian A, Ferrara M, Sarkar R (2021) Fuzzy rank-based fusion of CNN models using Gompertz function for screening COVID-19 CT-scans. Sci Rep 11(1):1–12
33. Wang P, Kong X, Guo W, Zhang X (2021) Exclusive Feature Constrained Class Activation Mapping for Better Visual Explanation. IEEE Access 9:61417–61428
34. Ornek AH, Ceylan M (2021) Explainable Artificial Intelligence (XAI): Classification of Medical Thermal Images of Neonates Using Class Activation Maps.Traitement du Signal, *38*(5)
35. Abd-Ellah MK, Awad AI, Khalaf AA, Hamed HF (2019) A review on brain tumor diagnosis from MRI images: Practical implications, key achievements, and lessons learned. Magn Reson Imaging 61:300–318
36. Naeem A, Anees T, Naqvi RA, Loh WK (2022) A Comprehensive Analysis of Recent Deep and Federated-Learning-Based Methodologies for Brain Tumor Diagnosis. J Personalized Med 12(2):275
37. Naseer A, Yasir T, Azhar A, Shakeel T, Zafar K (2021) Computer-aided brain tumor diagnosis: performance evaluation of deep learner CNN using augmented brain MRI. *International Journal of Biomedical Imaging*, *2021*
38. Kose U, Deperlioglu O, Alzubi J, Patrut B (2021) Deep Learning Architectures for Medical Diagnosis. Deep Learning for Medical Decision Support Systems. Springer, Singapore, pp 15–28
39. Alfarzaeai MS, Niu Q, Zhao J, Eshaq RMA, Hu E (2020) Coal/gangue recognition using convolutional neural networks and thermal images. IEEE Access 8:76780–76789
40. Jalwana MA, Akhtar N, Bennamoun M, Mian A (2021) CAMERAS: Enhanced resolution and sanity preserving class activation mapping for image saliency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 16327–16336)
41. Ko, Y. C., Wey, S. Y., Chen, W. T., Chang, Y. F., Chen, M. J., Chiou, S. H., … Lee,C. Y. (2020). Deep learning assisted detection of glaucomatous optic neuropathy and potential designs for a generalizable model. Plos One, 15(5), e0233079
42. Phan TMN, Nguyen HT (2021) Clinical Decision Support Systems for Pneumonia Diagnosis Using Gradient-Weighted Class Activation Mapping and Convolutional Neural Networks. Soft Computing: Biomedical and Related Applications. Springer, Cham, pp 81–92
43. Sun Y, Dai S, Li J, Zhang Y, Li X (2019) Tooth-marked tongue recognition using gradient-weighted class activation maps. Future Internet 11(2):45
44. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision* (pp. 618–626)
45. Lloyd CT, Sorichetta A, Tatem AJ (2017) High resolution global gridded data for use in population studies. Sci Data 4(1):1–17
46. Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J.,… Van Leemput, K. (2014). The multimodal brain tumor image segmentation benchmark(BRATS). *IEEE Transactions on Medical Imaging*, *34*(10), 1993–2024
47. Wong HB, Lim GH (2011) Measures of diagnostic accuracy: sensitivity, specificity, PPV and NPV. *Proceedings of Singapore Healthcare*, *20*(4), 316–318
48. Pinchi V, Pradella F, Vitale G, Rugo D, Nieri M, Norelli GA (2016) Comparison of the diagnostic accuracy, sensitivity and specificity of four odontological methods for age evaluation in Italian children at the age threshold of 14 years using ROC curves. Med Sci Law 56(1):13–18
49. Chattopadhay A, Sarkar A, Howlader P, Balasubramanian VN (2018) Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 839–847). IEEE
50. Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., … Hu, X. (2020). Score-CAM:Score-weighted visual explanations for convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 24–25)
51. Wang G, Li W, Ourselin S, Vercauteren T (2019) Automatic brain tumor segmentation based on cascaded convolutional neural networks with uncertainty estimation. Front Comput Neurosci 13:56
52. Shaikh M, Anand G, Acharya G, Amrutkar A, Alex V, Krishnamurthi G (2017) Brain tumor segmentation using dense fully convolutional neural network. International MICCAI brainlesion workshop. Springer, Cham, pp 309–319
53. Chen L, Wu Y, DSouza AM, Abidin AZ, Wismüller A, Xu C (2018) MRI tumor segmentation with densely connected 3D CNN. In *Medical Imaging 2018: Image Processing* (Vol. 10574, p. 105741F). International Society for Optics and Photonics
54. Srinivas B, Rao GS (2019) A hybrid CNN-KNN model for MRI brain tumor classification. Int J Recent Technol Eng (IJRTE) ISSN 8(2):2277–3878
55. Abd-Ellah MK, Awad AI, Hamed HF, Khalaf AA (2019) Parallel deep CNN structure for glioma detection and classification via brain MRI Images. In *2019 31st International Conference on Microelectronics (ICM)* (pp. 304–307). IEEE

56. Islam M, Ren H (2017) Multi-modal pixelnet for brain tumor segmentation. International MICCAI Brainlesion Workshop. Springer, Cham, pp 298–308
57. Khan MA, Ashraf I, Alhaisoni M, Damaševičius R, Scherer R, Rehman A, Bukhari SAC (2020) Multimodal brain tumor classification using deep learning and robust feature selection: A machine learning application for radiologists. Diagnostics 10(8):565