frontiers in Genetics

ORIGINAL RESEARCH
published: 14 March 2022
doi: 10.3389/fgene.2022.822666

Explanation-Driven Deep Learning Model for Prediction of Brain Tumour Status Using MRI Image Data

*Loveleen Gaur[1], Mohan Bhandari[2], Tanvi Razdan[1], Saurav Mallik[3] and Zhongming Zhao[3,4]**

[1]Amity International Business School, Amity University, Noida, India, [2]Nepal College of Information Technology, Lalitpur, Nepal, [3]Center for Precision Health, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, United States, [4]Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, United States

Cancer research has seen explosive development exploring deep learning (DL) techniques for analysing magnetic resonance imaging (MRI) images for predicting brain tumours. We have observed a substantial gap in explanation, interpretability, and high accuracy for DL models. Consequently, we propose an explanation-driven DL model by utilising a convolutional neural network (CNN), local interpretable model-agnostic explanation (LIME), and Shapley additive explanation (SHAP) for the prediction of discrete subtypes of brain tumours (meningioma, glioma, and pituitary) using an MRI image dataset. Unlike previous models, our model used a dual-input CNN approach to prevail over the classification challenge with images of inferior quality in terms of noise and metal artifacts by adding Gaussian noise. Our CNN training results reveal 94.64% accuracy as compared to other state-of-the-art methods. We used SHAP to ensure consistency and local accuracy for interpretation as Shapley values examine all future predictions applying all possible combinations of inputs. In contrast, LIME constructs sparse linear models around each prediction to illustrate how the model operates in the immediate area. Our emphasis for this study is interpretability and high accuracy, which is critical for realising disparities in predictive performance, helpful in developing trust, and essential in integration into clinical practice. The proposed method has a vast clinical application that could potentially be used for mass screening in resource-constraint countries.

Keywords: LIME, SHAP, XAI, brain tumor, MRI

OPEN ACCESS

Edited by:
Alfonso Maurizio Urso, Institute for High Performance Computing and Networking (ICAR)—Italian National Research Council (CNR), Italy

Reviewed by:
Andrea Tangherloni, University of Bergamo, Italy
Leonardo Rundo, University of Salerno, Italy

****Correspondence:***
Zhongming Zhao
Zhongming.Zhao@uth.tmc.edu

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: *26 November 2021*
Accepted: *17 February 2022*
Published: *14 March 2022*

Citation:
Gaur L, Bhandari M, Razdan T, Mallik S and Zhao Z (2022) Explanation-Driven Deep Learning Model for Prediction of Brain Tumour Status Using MRI Image Data. Front. Genet. 13:822666. doi: 10.3389/fgene.2022.822666

1 INTRODUCTION

According to the world health organization (WHO) world cancer report (2020), cancer is amongst the leading death-causing diseases, ranked second (after cardiovascular disease), accounting for nearly 10 million deaths in 2020 (Sung et al., 2021). Compared to other diagnoses, cancer screening is a different and more complicated public health approach that needs extra resources, infrastructure, and coordination. The WHO recommends the implementation of screening programs when the following conditions are fulfilled (Sung et al., 2021):

1. The efficiency of tool/model/software has been demonstrated

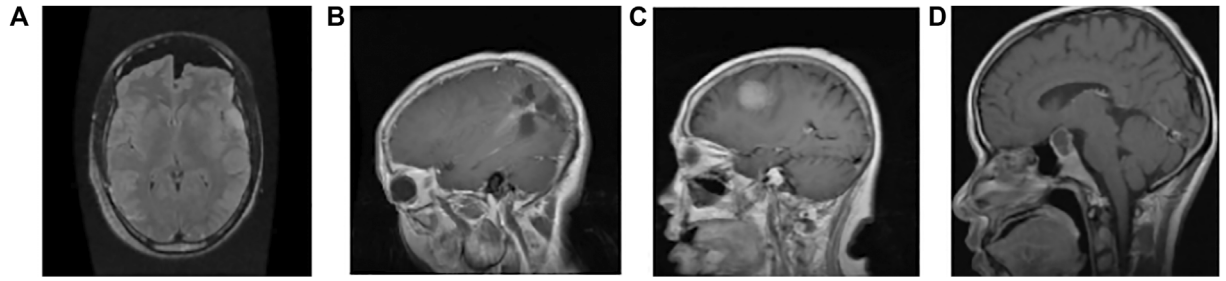

FIGURE 1 | Sample image data of different types of tumours. **(A)** Normal: the intensity of the parenchyma in the brain without any tumour is normal. The ventricular system and cisternal spaces are supposed to be in good working order. There is always no evidence of an intracranial space-occupying lesion (Gaillard, 2021). **(B)** Glioma tumour: gliomas have thick, irregularly enhancing borders of the focal necrotic core with a haemorrhagic component. They are surrounded by vasogenic-type oedema, containing malignant cell infiltration. Intratumoural haemorrhage happens rarely (less than 2%) (Frank, 2021) **(C)** Meningioma tumour: meningiomas are extra-axial tumours arising from meningocytes or arachnoid cap cells of meninges and can be found where meninges exist, as well as in some sites where only rest cells are thought to exist (Gaillard and Rasuli, 2021) **(D)** Pituitary tumour: for pituitary adenomas, minor intra-pituitary lesions appear differently than larger lesions that spread into the suprasellar region and pose various surgical and diagnostic issues. Based on tumour aspects, overall signal qualities can vary (Weerakkody and Gaillard, 2021).

2. Sufficient resources and facilities to confirm diagnoses and treatments are available
3. The prevalence of the disease is extreme enough to justify the screening

The total prevalence of all central nervous system tumours is 3.9 per 100,000 persons worldwide; the incidence differs with age, gender, race, and region and is extremely frequent in Northern Europe, followed by Australia, the United States, and Canada. Meningioma is the most common one, accounting for 36.8% of all tumours; glioma is the most widespread malignant tumour, accounting for 75% of central nervous system malignant tumours, with a total incidence of six cases per 100,000 people per year. MRI is presently the ideal method for early detection of human brain tumours as it is non-invasive (Spatharou et al., 2021). However, the interpretation of MRI is predominantly centred on the opinions of radiologists.

The advent of convolution neural network (CNN)-based deep learning (DL) provides the basis for imaging-based artificial intelligence (AI) solutions. DL-guided solutions intend to supplement clinical decision making. There are several motives why the proposed architecture is a CNN-based DL architecture. First, it is observed that CNN-based DL is extremely good at lowering the threshold of parameters while maintaining model quality. Second, it does not require human feature engineering because it can automatically extract features from an image. Third, the literature supports the CNN-based DL model by several researchers and that it has achieved good image classification and recognition accuracy. However, it is crucial to observe that very few researchers have applied local interpretable model-agnostic explanation (LIME) and Shapley additive explanation (SHAP) along with CNN. Researchers demonstrated the immense potential of imaging tools to mitigate the heavy burden on medical experts (Wojciech et al., 2017). It further allows devoting additional help in patient care, reducing burnout, and shrinking overall medical costs for patients (Dave et al., 2020). Working on the detection system, Gupta et al. (2016) applied DL algorithms, Resnet50, to distinguish COVID-19 from X-rays to achieve a fully autonomous and speedier diagnosis. With an average COVID-19 detection time of roughly 2.5 s and an average accuracy of 0.97, the authors aimed to minimise the run time to about 2.5 s. Kollias et al. (2018) introduced different performance indicators such as precision, responsiveness, specificity, precision, F1 value, and DL. The results showed a standard accuracy of 92.93% and sensitivity of 94.79% to provide robust identification and detection of COVID-19 in the chest X-ray dataset. In one of the research (Ke et al., 2019), the deep neural network correlation learning mechanism for CT brain tumour detection used palettes of CNN architecture to adjust them to the best possible detection result of ANN. The AISA framework for MRI data analysis demonstrated its application to brain scan data by deriving independent subspaces and extracting texture features. Then, dimensionality is reduced using t-SNE embedding for discriminative classification. Finally, the KNN classification is applied. Despite the immense popularity of DL models in clinical decision making, the lack of interpretability and transparency by algorithm-driven decisions remains the biggest challenge, particularly in medical settings. Although, many researchers (Richard et al., 2020; Zucco et al., 2018) observed various impediments in developing XAI-based clinical decision support systems (CDSS) due to the non-availability of any universal notion of explainability. Our study proposes an explanation-driven DL-based model to predict distinctive subtypes of brain tumours (meningioma, glioma, and pituitary) using an MRI image dataset. We also implemented LIME and Shapley additive explanations to create more transparency in the models while keeping intact a high performance rate. Our study will help the users (medical professionals, clinicians, etc.) in comprehending and efficiently managing the ever-increasing number of trustable and reliable AI partners (Sharma et al., 2020).

Compared to previous models, our model used a dual-input CNN approach to prevail over the classification challenge with inferior-quality images and an accuracy of 94.64% compared to other state-of-the-art models. Previous studies lack explanation, and thus, we used Explainable AI (XAI) algorithms such as LIME and SHAP, which is the differentiating element of this study. We used SHAP to ensure consistency and local accuracy for interpretation as Shapley values

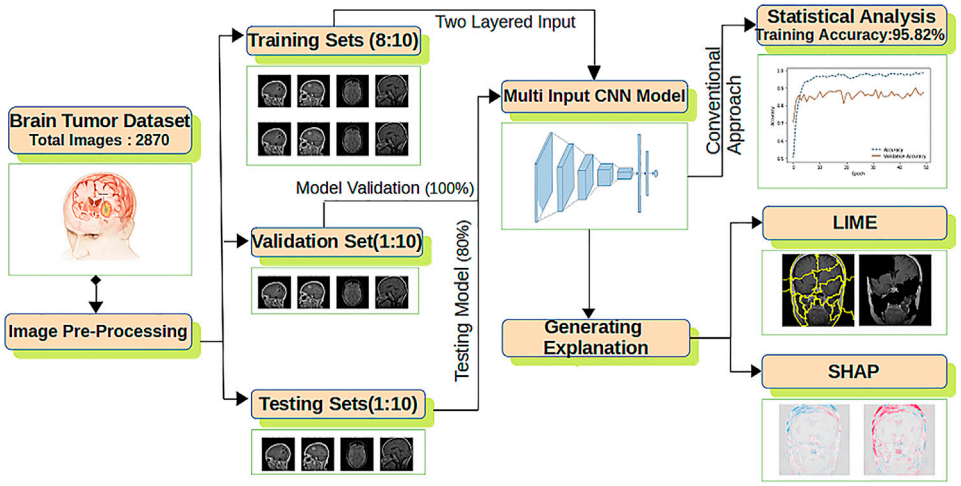

FIGURE 2 | The proposed explanation-driven DL model for prediction of brain tumour status using MRI image data: 2870 MRI images are pre-processed and divided into training, validation, and test sets. Two copies of datasets are fed into a multi-input CNN model to find the training, validation, and test accuracy. The same CNN model was further imposed on LIME and SHAP.

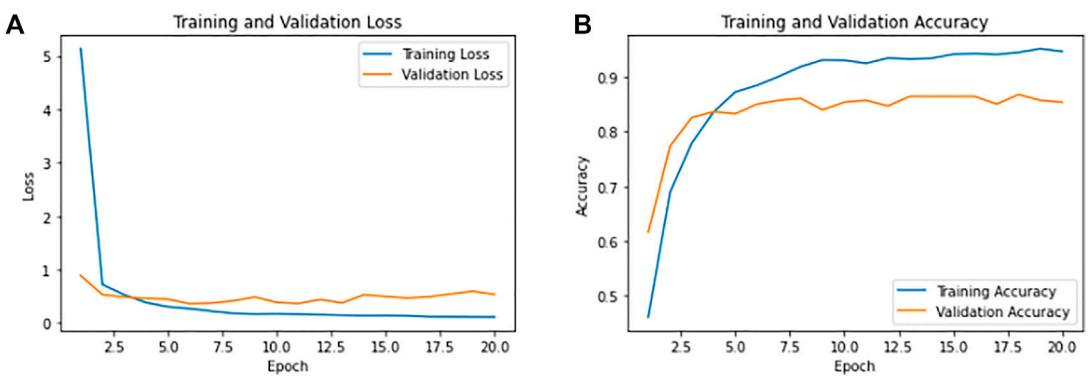

FIGURE 3 | Training and validation results of CNN. **(A)** Shows the overall training loss as 0.1149 and validation loss as 0.53. **(B)** Shows the overall training accuracy as 94.64% and validation accuracy of 85.37%.

examine all potential predictions using all possible combinations of inputs. Conversely, LIME constructs sparse linear models around each prediction to describe how the model operates in the immediate area.

The deep neural network correlation learning mechanism for computed tomography (CT) brain tumour detection used palettes of CNN architecture to adjust them to the best possible detection result of DL. Though the previously suggested models have higher accuracy, they lack explainability, interpretability, and transparency (Abdalla and Esmail, 2018; Khairandish et al., 2021). The proposed model used XAI algorithms such as LIME (Vedaldi and Soatto, 2008) and SHAP as detailed in Algorithm 2.

The contributions in this study are summarised in what follows:

1. We aimed to create an explanation-driven multi-input DL model where SHAP and LIME are used for an in-depth description of results. One set of two input datasets is fed to the convolution layer and one to the fully connected layer.
2. We have achieved high accuracy of (94.64%) brain MRI images compared to other state-of-the-art models.

2 METHODS

2.1 Datasets

In this study, we used the publicly available MRI images (Bhuvaji, 2020). The datasets are annotated into three categories of tumours: glioma tumour, meningioma tumour, and pituitary tumour, along with the normal image. Out of 2,870 total images, 2,296 images of distinct types are used as training sets and the remaining as test sets.

TABLE 1 | K-fold cross-validation results.

Fold	Final validation loss	Final validation accuracy (%)
1	0.011 44	99.5
2	0.017 06	98.47
3	0.021 52	99.13
4	0.009 88	99.34
5	0.005 54	100
6	0.012 98	99.13
7	0.008 74	99.78
8	0.005 33	99.78
9	0.010 18	99.34
10	0.008 8	99.56

2.1.1 Data Pre-Processing

All 512 × 512 × 3 images are resized to 150 × 150 × 3. The images are rearranged for faster convergence and preventing the CNN model from learning the training order. For better classification results, we have introduced Gaussian noise as it improves the learning for DL (Neelakantan et al., 2015) with mean = 0 and standard deviation $10^{0.5}$. **Figure 1** shows a single instance among the categories of tumours from the dataset.

2.2 Proposed Framework

The overall architecture of the model used is shown in **Figure 2** composed of feature extraction, a CNN model, statistical performance measures, and explanation extraction frameworks.

For improved accuracy, two copies of the dataset are fed to the CNN model having an output layer of size 1 × 4 and six hidden layers (Yu et al., 2017). Adam optimiser with its default parameters is applied with the rectified linear unit (ReLU) and softmax as the activation function. The final CNN model is used for statistical accuracy measurement, LIME and SHAP. For LIME explanations, perturbation is calculated, whereas for SHAP, a gradient explainer is applied. The whole process is formalized in Algorithm 1.

Algorithm 1. Explanation-driven multi-input DL model for prediction of brain tumour.

Input: MRI Dataset(Break down in ratio 8:1:1) with size 150x150x3
Epoch: 20
Optimizer: Adam
Kernel Size: 3 x 3
Dropout: 0.2
Filter : Conv2D
for *For every iteration in dual input CNN Model* **do**
    Input one set of dataset to convolution layer
    Input one set of dataset to fully connected layer
    Calculate loss, accuracy, validation loss, validation accuracy
end for
Implement XAI: Implement LIME and SHAP for the model

For the classification task in the proposed explainable model, a CNN with dual-input architecture is used. The CNN is imposed with ReLU as activation in all hidden layers. Compared with the input value and zero value, ReLU is simple to calculate. Furthermore, ReLU has a derivative of either 0 or 1 based on positive or negative input. This feature of ReLU is essential in comparing explainable modules such as LIME and SHAP. Adam optimiser with its default parameter (Kingma and Ba, 2015) is used along with sparse categorical cross entropy; the kernel size is set to 3 × 3.

3 RESULTS

Following the classification process, the performance of CNN models is evaluated based on accuracy and the number of wrong predictions. The curves for the conventional results of CNN are presented in **Figure 3**.

3.1 CNN

The model was iterated for 20 epochs, and during callback in CNN modules, we had monitored the loss with min mode and patience level of three to cross the over-fitting. Achieving the training accuracy of 94.64% and overall test accuracy of 85.37%, the model has 26 wrong predictions with 0.1149 as training loss and 0.53 as validation loss.

Furthermore, to estimate the performance of the CNN model on the configured dataset, K-fold cross validation is performed with K = 10 non-overlapping folds for 20 epochs with a batch size of 128. The test and train sets were split in the ratio of 1:4. The final validation result of the cross fold is shown in **Table 1**. The proposed model has achieved almost 100% training accuracy during cross validation.

Table 2 shows the confusion matrix for 287 test images. A total of 7 normal images out of 46, 14 glioma images out of 84, 12 meningioma images out of 77, and 3 pituitary images out of 80 were misclassified.

To validate our model statistically, we performed McNemar's test (Smith et al., 2020). For labels of test data and labels of model prediction under test data, McNemar's test gave a chi-squared value of 42.022 and p value 9.02 x e^{-11}. We can reject the null-hypothesis that both labels perform equally well on the test set, since the p value is smaller than $\alpha = 0.005$.

3.2 SHAP

For each pixel on a predicted image, the scores show its contribution and can be used to explain tumour classification tasks. The Shapley values correspond to each feature for different categories of the tumour according to Algorithm 2.

Algorithm 2. Algorithm to calculate the Shapley values.

Input: Number of iterations M, instance of interest x, feature index j, data matrix X, and Dual input CNN model
for *Every Iteration 1 ..... M* **do**
    Draw random instance z from the data matrix X
    Choose a random permutation o of the feature values
    Order instance x: (x_0 ...., x_j, ... , x_p)
    Order instance z: (z_0 ...., z_j, ... , z_p)
    Construct two new instances:
        with j:x_{-j} = (x_0 ..., x_j,z_{j+1}, ... z_p)
        without j:x_{+j} = (x_0 ..., x_j,z_{j+1}, ... z_p)
    Compute Marginal Distribution:

$$\phi_j^m = \hat{f}(x_{+j}) - \hat{f}(x_{-j}) \tag{1}$$

end for
Compute shapely values:

$$\phi_j(x) = \frac{1}{M}\sum_{m=1}^{M}\phi_j^m \tag{2}$$

TABLE 2 | Confusion matrix for the CNN.

		Actual value			
		Normal	**Glioma**	**Meningioma**	**Pituitary**
Predicted values	Normal	37	8	1	0
	Glioma	7	70	5	2
	Meningioma	0	12	65	0
	Pituitary	0	3	0	77

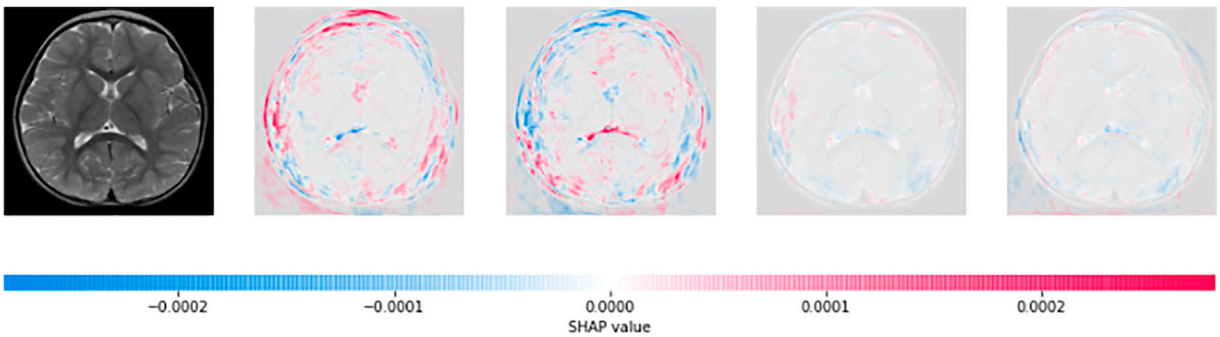

FIGURE 4 | On the basis of Shapley values, we can say that the MRI image is normal.

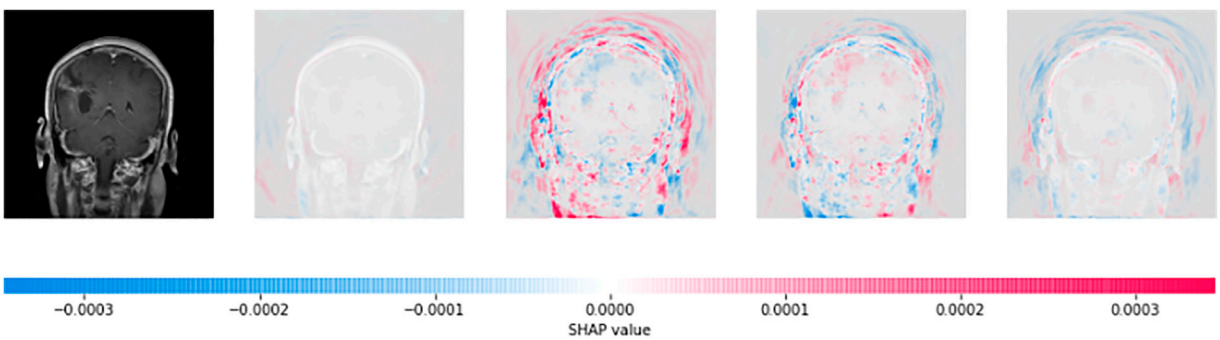

FIGURE 5 | On the basis of Shapley values, we can say that the MRI image holds meningioma tumour.

The CNN model with mathematical behaviour is complicated to interpret directly. Thus, the effect of individual input features on the model's output is clearly explained using SHAP and shown in **Figures 4**, **5**. Positive SHAP values that raise the likelihood of the class are represented by red pixels. In contrast, negative SHAP values that lower the probability of the class are represented by blue pixels. **Figure 4** and **Figure 5** are test images. In contrast, the rest of the figures indicate the normal image and three other categories of tumour: glioma, meningioma, and pituitary tumours in successive order.

3.3 LIME

A total of 150 perturbations are used. Random ones and zeros are produced and formed into a matrix, with perturbations as rows and superpixels as columns. A superpixel is ON if it is 1, and it is OFF if it is 0. The length of the displayed vector represents the number of superpixels in the image. The test image is perturbed based on the perturbation vector and predefined superpixels (Vedaldi and Soatto, 2008). The final perturbed image is shown in **Figure 6C** for normal test image under consideration and in **Figure 7C** for test image under consideration with meningioma tumour, which shows the portion of the image having a major role for classification.

The CNN model is utilised to generate the explanation using LIME. **Figure 6A**is a normal image, and **Figure 7A** is under the meningioma category. The classification produces a vector of 2,870 probabilities for each category accessible in the CNN model. The quick-shift segmentation method is used to create superpixels. 22 superpixels are generated for **Figure 6A** and

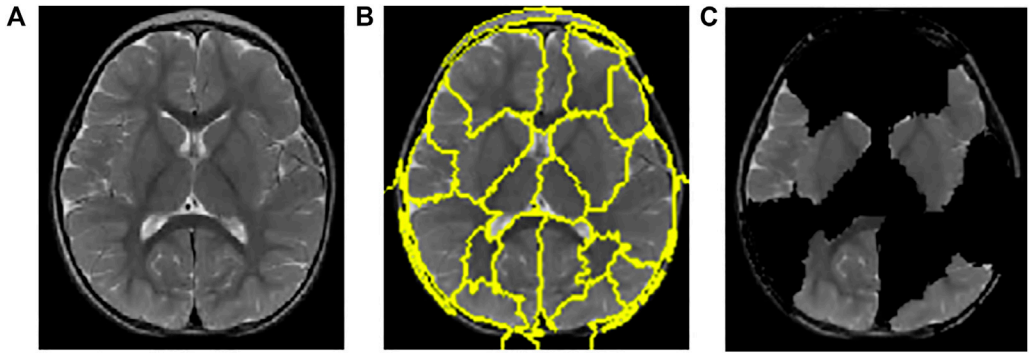

FIGURE 6 | Interpretations generated by LIME for a normal image. **(A)** Sample of the normal image from the test image. **(B)** Superpixels generated from a sample of the normal image from test image quick-shift segmentation to create perturbations. **(C)** Final perturbed image for the normal image.

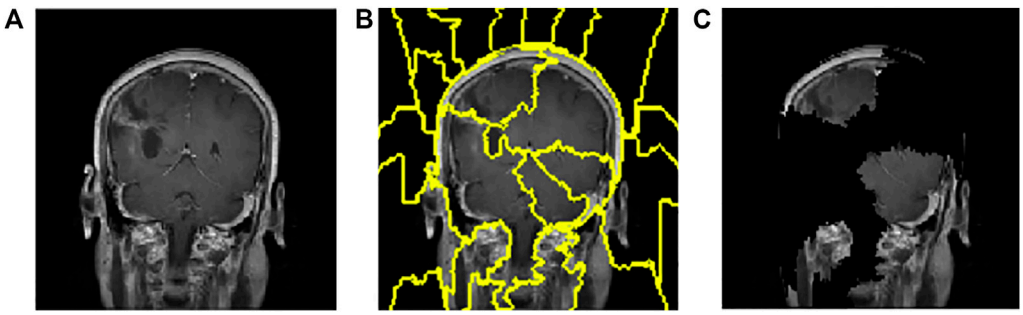

FIGURE 7 | Interpretations generated by LIME for meningioma tumour. **(A)** Sample of meningioma tumour from the test image. **(B)** Superpixels generated from quick-shift segmentation to create perturbations. **(C)** Final perturbed image showing meningioma tumour.

TABLE 3 | Brain tumour detection using traditional ML methods.

Authors	Algorithm	Dataset	Accuracy (%)	XAI
Martinez et al. (2020)	Random Forest	BraTs Dataset	76	No
Minz and Mahobiya (2017)	Adaboost Classifier	BraTs Dataset	89.90	No
Abdalla and Esmail (2018)	Back-Propagation Network	MRI Images	99	No
Asodekar and Gore (2019)	Random Forest	BraTs Dataset	81.90	No
Asodekar and Gore (2019)	SVM	BraTs Dataset	78.57	No
Proposed model	Dual-Input CNN	MRI Images	94.64	Yes

shown in **Figure 6B**, and 24 superpixels are calculated for **Figure 7A** and shown in **Figure 7B**.

4 DISCUSSION

4.1 Comparison of the Proposed Feature Extraction Methods Using Traditional Machine learning (ML) Methods

We compare the proposed feature extraction methods to traditional ML methods. The comparative results are presented in **Table 3**. Minz and Mahobiya (2017) pre-processed the MICCAI BraTS dataset to eliminate noise and employed the GLCM (gray-level co-occurrence matrix) for feature extraction and classification boosting (Adaboost). An MRI was used to extract 22 characteristics. The Adaboost classifier is utilised for classification, and the suggested system achieves a maximum accuracy of 89.90%. Abdalla and Esmail (2018) executed a computer-aided detection system after collecting the MRI images. They processed the image before implementing the back-propagation algorithm and extracted the features using Haralick's features based on the spatial gray-level dependency matrix (SGLD). The results were 99%, but the study could not focus on the explainable section in the training images. A comparative study between support vector machine (SVM)

TABLE 4 | Brain tumour detection using other state-of-the-art models.

Authors	Algorithm	Dataset	Accuracy (%)	XAI
Shahzadi et al. (2018)	CNN with LSTM	MRI Images	84	No
Hemanth et al. (2019)	CNN	MRI Images	91	No
Avsar and Salcin (2019)	R-CNN	MRI Images	91.66	No
Ranjbarzadeh et al. (2021)	C-CNN	BraTs Dataset	92.03	No
Khairandish et al. (2021)	CNN–SVM	MRI Images	98.49	No
Proposed model	Dual-Input CNN	MRI Images	94.69	Yes

and random forest (RF) classified benign and malignant tumours. First, the brain tumour's region of interest was determined for feature extraction, and then, features were calculated. Shape characteristics were obtained and utilised to classify benign and malignant tumours. According to the authors, RF (81.90%) outperformed the SVM (78.57%). By combining principal component analysis (PCA), KSVM, and GRB kernels, Arora and Ratan (2021) established a unique technique for categorisation of MRI brain images using discrete wavelet transform (DWT). The experiment was carried out with four different kernels. The findings demonstrate that combining DWT, PCA, KSVM, and the GRB kernel yields the highest accuracy compared to other methodologies. The results show that the time it takes to classify a segmented picture significantly decreases, which might be a watershed moment in the medical profession for tumour diagnosis. Martinez et al. (2020) worked on the FLAIR images on the BRATS 2015 training dataset; it is used to restructure and increase data attributes that lead to a pixel-based classifier. The U-net suggested method performs a semantic segmentation with a precision of 76%, which increases by 23% compared to the random forest classifier with synthetic minority oversampling technique (SMOTE) class balancing algorithm.

4.2 Comparison of the Proposed Method With the Other State-of-the-Art Methods

This section compares our dual-input CNN model with other state-of-the-art models. The results are compared in **Table 4**. After several data-collection and pre-processing steps such as average filtering segmentation, the DL model was implemented by researchers (Hemanth et al., 2019). In comparison to existing approaches such as conditional random field (89%), SVM (84.5%), and genetic algorithm (GA) (83.64%), the research represents overall performance and comparative output on the brain MRI images. In contrast to existing algorithms, the suggested CNN (91%) produces improved results. The TensorFlow library was used to construct a DL method called faster R-CNN in the work of Avsar and Salcin (2019), and the classifier algorithm was trained and tested using a publicly available dataset of 3,064 MRI brain pictures (708 meningiomas, 1,426 gliomas, and 930 pituitary gland tumours) from 233 patients. The quicker RCNN algorithm has been demonstrated to attain 91.66% accuracy, which is exceptional compared to past work on the same dataset. Ranjbarzadeh et al. (2021) proposed a cascaded convolutional neural network (C-ConvNet/C-CNN). A simple but effective cascade, the CNN model, has been suggested to extract local and global characteristics in two methods, with different extraction patches in each. Those patches were chosen to feed the network that their centre was located inside this area after extracting the tumour's predicted location using a sophisticated pre-processing strategy. As a result of removing a high number of insignificant pixels from the picture in the pre-processing stage, the computing time and ability to generate quick predictions for categorising the clinical image are reduced. The results were compared to other algorithms. Still, the CNN model achieved the highest accuracy (92.03%) on the whole Dice score (mean) and the highest precision (97.12%) on the core sensitivity score (mean). Khairandish et al. (2021) made use of a hybrid model of CNN and SVM in phrases of classification, type, and threshold-based segmentation in terms of detection to classify benign and malignant tumours in brain MRI images. This hybrid CNN–SVM is rated as having an overall accuracy of 98.49%. Still, their study does not show evidence for manipulating low-quality images and XAI. Shahzadi et al. (2018) proposed a CNN cascade with a long short-term memory (LSTM) network for classifying 3D brain tumour MRIs into HG and LG glioma. The features from the pre-trained VGG-16 were retrieved and fed into an LSTM network for learning high-level feature representations. The components extracted from VGG-16 had a classification accuracy of 84%, higher than that of those extracted from AlexNet and ResNet, 71%. Isola et al. (2018) investigated conditional adversarial networks as a general-purpose solution for image-to-image translation challenges by using a 1,616 PatchGAN. The PatchGAN 70 × 70 reduces these distortions and improves scores slightly. It is observed that scaling to the full 286 × 286 ImageGAN does not significantly improve the visual quality of the findings and results in a considerably lower FCN-score, indicating that conditional adversarial networks are a promising option for many image-to-image translation tasks, especially those involving highly structured graphical outputs. Milletari et al. (2016) proposed an approach to 3D image segmentation based on a volumetric, fully convolutional neural network. The CNN is trained end-to-end on MRI volumes depicting the prostate and predicts segmentation for the whole volume at once. The training was performed on 50 MRI volumes, and the relative manual ground truth annotation was obtained from the PROMISE2012 challenge dataset. The novel objective function was to optimise during training based on the dice overlap coefficient between the predicted segmentation and the

ground truth annotation. Han et al. (2020) proposed an unsupervised medical anomaly detection generative adversarial network (MADGAN). This two-step method uses GAN-based multiple adjacent brain MRI slice reconstruction to detect brain anomalies at various stages on multi-sequence structural MRI. MADGAN can detect anomaly on T1 scans at a very early stage, mild cognitive impairment (MCI), with area under the curve (AUC) 0.727, and anomaly detection (AD) at a late stage with AUC 0.894, while detecting brain metastases on T1c scans with AUC 0.921. On multi-sequence MRI, the model may accurately detect the accumulation of subtle anatomical abnormalities and hyper-intense enhancing lesions, such as (particularly late stage) AD and brain metastases, as the first unsupervised varied disease diagnosis. Baur et al. (2020) presented a novel method towards unsupervised AD in brain MRI by embedding the modelling of healthy anatomy into a CycleGAN-based style-transfer task, which is trained to translate healthy brain MRI images to a simulated distribution with lower entropy and vice versa. By filtering high-frequency, low-amplitude signals from lower entropy samples during training, the resulting model suppresses anomalies in reconstructing the input data at test time. The method outperforms the state-of-the-art method in various measures and can deal with high-resolution data, a current pitfall of autoencoder (AE)-based methods. Castiglioni et al. (2021) concentrated on the issues that must be addressed to create AI applications as clinical decision support systems in a real-world setting. A narrative review with a critical appraisal of publications published between 1989 and 2021 was conducted. According to the study, biomedical and healthcare systems are among the most significant domains for AI applications, with medical imaging being the most suited and promising domain. Clarification of specific challenging points facilitates the development of such systems and their translation to clinical practice. Barragán-Montero et al. (2021) showcased the technological pillars of AI, as well as the state-of-the-art methods and their implementation to medical imaging. This review offered an overview of AI, emphasising medical imaging analysis demonstrating the potential of the state-of-the-art ML and DL algorithms to automate and enhance several aspects of clinical practice.

5 CONCLUSION AND FUTURE DIRECTION

Using an explanation-driven dual-input CNN model for finding if a particular MRI image is subjected to a tumour or not, the proposed study achieved an accuracy of 94.64%. A brain MRI image dataset is used to train and test the proposed CNN model, and the same model was further imposed to SHAP and LIME algorithms for an explanation. Our experiment utilised two dataset copies as input for better feature extraction, one in the convolution layer and another in the fully connected layer. However, any attempt to remove any features decreased the prediction model's overall performance; hence, no augmentation was carried out. The proposed model is a locally interpreted model with a model-agnostic explanation, shapely explained to describe the results for ordinary people more qualitatively.

In future, classification algorithms with higher accuracy and better optimiser can be used and imposed on XAI. For better clinical issues, the research may be replicated and applied to other XAI algorithms such as GradCAM. Furthermore, like the most recent advances on computing capacity, neuroimaging technologies, and digital phenotyping tools (Ressler and Williams, 2020), algorithms to imitate natural occurrences can be used on heterogeneous datasets for medical imaging modalities, electronic health record engines, multi-omics studies, and real-time monitoring (Rundo et al., 2019).

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. These data can be found here at https://www.kaggle.com/sartajbhuvaji/brain-tumor-classification-mri.

ETHICS STATEMENT

Written informed consent was not obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

Conceptualisation of the research topic and writing of the original draft were carried out by LG, MB, and TR. Resource collection and design of the methodology and code were carried out LG and MB. Project administration was conducted by LG, SM, and ZZ. Result validation was performed by LG, MB, SM, and ZZ. Final draft and revisions were made by SM and ZZ. Finally, the fund was acquired by ZZ.

FUNDING

ZZ was partially supported by the Cancer Prevention and Research Institute of Texas (CPRIT 180734). The funders did not participate in the study design, data analysis, decision to publish, or preparation of the manuscript.

REFERENCES

Abdalla, H. E. M., and Esmail, M. Y. (2018). "Brain Tumor Detection by Using Artificial Neural Network," in 2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), 1–6. doi:10.1109/iccceee.2018.8515763

Arora, P., and Ratan, R. (2021). "Development of a Novel Approach for Classification of Mri Brain Images Using Dwt by Integrating Pca, Ksvm and Grb Kernel," in Proceedings of Second International Conference on Smart Energy and Communication. doi:10.1007/978-981-15-6707-0_13

Asodekar, B., and Gore, S. A. (2019). *Brain Tumor Classification Using Shape Analysis of Mri Images*.

Avşar, E., and Salçin, K. (2019). Detection and Classification of Brain Tumours from Mri Images Using Faster R-Cnn. *Teh. Glas. (Online)* 13, 337–342. doi:10.31803/tg-20190712095507

Barragán-Montero, A., Javaid, U., Valdés, G., Nguyen, D., Desbordes, P., Macq, B., et al. (2021). Artificial Intelligence and Machine Learning for Medical Imaging: A Technology Review. *Physica Med.* 83, 242–256. doi:10.1016/j.ejmp.2021.04.016

Baur, C., Graf, R., Wiestler, B., Albarqouni, S., and Navab, N. (2020). "Steganomaly: Inhibiting Cyclegan Steganography for Unsupervised Anomaly Detection in Brain Mri," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*. Editors A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, et al. (Cham: Springer International Publishing), 718–727. doi:10.1007/978-3-030-59713-9_69

Bhuvaji, S. (2020). *Brain Tumor Classification-Mri. Vol. 2*. doi:10.34740/kaggle/dsv/1183165

Castiglioni, I., Rundo, L., Codari, M., Di Leo, G., Salvatore, C., Interlenghi, M., et al. (2021). Ai Applications to Medical Images: From Machine Learning to Deep Learning. *Physica Med.* 83, 9–24. doi:10.1016/j.ejmp.2021.02.006

Dave, D., Naik, H., Singhal, S., and Patel, P. (2020). *Explainable AI Meets Healthcare: A Study on Heart Disease Dataset. Vol. abs/2011.03195*.

Di Muzio, B., and Gaillard, F. (2021). *Normal Brain Mri: Radiology Case*. doi:10.53347/rID-42777

Frank, G. (2021). *Glioblastoma: Radiology Reference Article*. doi:10.53347/rID-4910

Gaillard, F., and Rasuli, B. (2021). *Meningioma*. doi:10.53347/rID-1659

Gupta, M., Rao, P., and Rajagopalan, V. (2016). "Brain Tumor Detection in Conventional Mr Images Based on Statistical Texture and Morphological Features," in 2016 International Conference on Information Technology (ICIT), 129–133. doi:10.1109/icit.2016.037

Han, C., Rundo, L., Murao, K., Noguchi, T., Shimahara, Y., Milacski, Z., et al. (2020). Madgan: Unsupervised Medical Anomaly Detection gan Using Multiple Adjacent Brain Mri Slice Reconstruction. *BMC Bioinformatics*. In Press.

Hemanth, G., Janardhan, M., and Sujihelen, L. (2019). "Design and Implementing Brain Tumor Detection Using Machine Learning Approach," in 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), 1289–1294. doi:10.1109/ICOEI.2019.8862553

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2018). *Image-to-image Translation with Conditional Adversarial Networks*.

Ke, Q., Zhang, J., Wei, W., Damasevicius, R., and Wozniak, M. (2019). Adaptive Independent Subspace Analysis of Brain Magnetic Resonance Imaging Data. *IEEE Access* 7, 12252–12261. doi:10.1109/ACCESS.2019.2893496

Khairandish, M. O., Sharma, M., Jain, V., Chatterjee, J. M., and Jhanjhi, N. Z. (2021). A Hybrid Cnn-Svm Threshold Segmentation Approach for Tumor Detection and Classification of Mri Brain Images. *Irbm*. doi:10.1016/j.irbm.2021.06.003

Kingma, D. P., and Ba, J. (2015). "Adam: A Method for Stochastic Optimization," in 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings. Editors Y. Bengio and Y. LeCun.

Kollias, D., Tagaris, A., Stafylopatis, A., Kollias, S., and Tagaris, G. (2018). Deep Neural Architectures for Prediction in Healthcare. *Complex Intell. Syst.* 4, 119–131. doi:10.1007/s40747-017-0064-6

Martinez, E., Calderon, C., Garcia, H., and Arguello, H. (2020). "Mri Brain Tumour Segmentation Using a Cnn over a Multi-Parametric Feature Extraction," in 2020 IEEE Colombian Conference on Applications of Computational Intelligence (IEEE ColCACI 2020), 1–6. doi:10.1109/ColCACI50549.2020.9247926

Milletari, F., Navab, N., and Ahmadi, S.-A. (2016). "V-net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," in 2016 Fourth International Conference on 3D Vision (3DV). doi:10.1109/3dv.2016.79

Minz, A., and Mahobiya, C. (2017). "Mr Image Classification Using Adaboost for Brain Tumor Type," in 2017 IEEE 7th International Advance Computing Conference (IACC), 701–705. doi:10.1109/IACC.2017.0146

[Dataset] Neelakantan, A., Vilnis, L., Le, Q. V., Sutskever, I., Kaiser, L., Kurach, K., et al. (2015). *Adding Gradient Noise Improves Learning for Very Deep Networks*.

Pembury Smith, M. Q. R., Ruxton, G. D., and Ruxton, G. D. (2020). Effective Use of the Mcnemar Test. *Behav. Ecol. Sociobiol.* 74. doi:10.1007/s00265-020-02916-y

Ranjbarzadeh, R., Bagherian Kasgari, A., Jafarzadeh Ghoushchi, S., Anari, S., Naseri, M., and Bendechache, M. (2021). Brain Tumor Segmentation Based on Deep Learning and an Attention Mechanism Using MRI Multi-Modalities Brain Images. *Sci. Rep.* 11. doi:10.1038/s41598-021-90428-8

[Dataset] Ressler, K. J., and Williams, L. M. (2020). Big Data in Psychiatry: Multiomics, Neuroimaging, Computational Modeling, and Digital Phenotyping. *Neuropsychopharmacol.* 46, 1–2. doi:10.1038/s41386-020-00862-x

Richard, A., Mayag, B., Talbot, F., Tsoukias, A., and Meinard, Y. (2020). Transparency of Classification Systems for Clinical Decision Support. *Inf. Process. Manage. Uncertainty Knowledge-Based Syst.* 1239, 99–113. doi:10.1007/978-3-030-50153-2_8

Rundo, L., Militello, C., Vitabile, S., Russo, G., Sala, E., and Gilardi, M. C. (2019). A Survey on Nature-Inspired Medical Image Analysis: A Step Further in Biomedical Data Integration. *Fi* 171, 345–365. doi:10.3233/FI-2020-1887

Shahzadi, I., Tang, T. B., Meriadeau, F., and Quyyum, A. (2018). "Cnn-lstm: Cascaded Framework for Brain Tumour Classification," in 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), 633–637. doi:10.1109/iecbes.2018.8626704

Sharma, Y., Verma, A., Rao, K., Eluri, V., Verma, A., Rao, K., et al. (2020). *'reasonable Explainability' for Regulating Ai in Health*.

Spatharou, A., Hieronimus, S., and Jenkins, J. (2021). *Transforming Healthcare with Ai: The Impact on the Workforce and Organizations*.

Sun, Y., Zhu, L., Wang, G., and Zhao, F. (2017). Multi-input Convolutional Neural Network for Flower Grading. *J. Electr. Comput. Eng.* 2017, 1–8. doi:10.1155/2017/9240407

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA A. Cancer J. Clin.* 71, 209–249. doi:10.3322/caac.21660

Vedaldi, A., and Soatto, S. (2008). "Quick Shift and Kernel Methods for Mode Seeking," in *Computer Vision – ECCV 2008*. Editors D. Forsyth, P. Torr, and A. Zisserman (Berlin, Heidelberg: Springer Berlin Heidelberg), 705–718. doi:10.1007/978-3-540-88693-8_52

Weerakkody, Y., and Gaillard, F. (2021). *Pituitary Adenoma: Radiology Reference Article*. doi:10.53347/rID-11024

Wojciech, S., Thomas, W., and Robert, M. K. (2017). *Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models*.

Zucco, C., Liang, H., Fatta, G. D., and Cannataro, M. (2018). "Explainable Sentiment Analysis with Applications in Medicine," in 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 1740–1747. doi:10.1109/BIBM.2018.8621359

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Gaur, Bhandari, Razdan, Mallik and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.