Dr.K. Ravikiran, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science,
Engineering and Technology
An Open Access Journal

A Hybrid Multi-Task Deep Learning Framework for Brain Tumor Classification and Segmentation Using ResNet-U-Net Architecture

Dr.K. Ravikiran [1], Guruvannagari Rishi [2], Veerabathini Sai Kumar [3], Uliravula Akhil [4]
Department of Information Technology GRIET affiliated with JNTUH Hyderabad, India

Abstract—Accurate identification and definition of brain tu-mors is a key element in effective diagnosis and treatment planning which at the same time is very time consuming and subject to variable results with manual analysis of medical images. We present a hybrid modality multi task deep learning framework for at the same time classifying and segmenting brain tumors which uses MRI and CT scans. We have put together a ResNet-50 encoder with a U-Net based decoder which enables joint learning of spatial and semantic features in a single architecture. Also we have used a dual branch design which at the same time produces pixel level tumor segmentation and at the same time determines tumor types which include Glioma, Meningioma, Pituitary, and Normal cases. To solve for the issues of different imaging modalities we put forth a unified preprocessing pipeline which allows the model to learn modality invariant features. We report we see a classification accuracy of 97.40% and a Dice similarity coefficient of 0.843 with our put forth framework also at the same time reporting that we are able to perform efficient real time inference on CPU based systems. Also we see that multi task learning does in fact improve diagnostic performance which in turn shows off the put forth system's value as a practical tool for clinical decision support.

Index Terms—Brain Tumor Segmentation, Tumor Classifica-tion, Multi-Task Learning, ResNet-50, U-Net, Medical Imaging, Deep Learning, MRI, CT

I. INTRODUCTION

Brain tumors present a very complex and serious challenge in neurology which requires accurate diagnosis and precise tumor definition for proper treatment planning. Imaging tools especially magnetic resonance imaging (MRI) are at the fore-front in terms of identifying tumor features like site, size, and growth patterns. But also here we see that radiologists' interpretation of the images is a very time intensive process which also is very much at the discretion of the radiologist's expertise which in turn leads to delayed clinical decisions and inter-observer variation. [1].

In recent past deep learning has seen great growth in medical image analysis. Convolutional Neural Networks which include in particular the encoder decoder types like U-Net have proved very successful in medical image segmentation tasks as a result of their ability to capture large scale context as well as small scale details of images [4]. Also it has been the case that residual based networks such as ResNet have enabled the training of larger models which in turn have improved on feature representation and classification accuracy [5]. These improvements have in turn made possible the development of automated tools which support clinicians in the detection and analysis of tumors.

Despite these developments, most existing approaches treat tumor segmentation and classification as independent tasks. Segmentation models focus on pixel-level delineation, while classification models predict tumor categories based on global image features. This separation limits the ability of models to leverage the inherent relationship between spatial tumor struc-ture and pathological identity. Multi-task learning (MTL) has emerged as a promising paradigm to address this limitation by enabling shared feature learning across related tasks, leading to improved generalization and performance [6].

Also present that which research is mostly into single modality imaging especially of the MRI variety we see due to the availability of benchmark tools like the Brain Tumor Segmentation (BraTS) challenge [3]. While MRI does a great job at imaging soft tissues it also has it's limitations in the detection of calcifications and hemorrhage which is where the value of CT imaging comes in. We see a large gap in present literature which is the hybrid modality systems that can put MRI and CT data together [14].

Another key issue we see is that of the implementation of deep learning models in practice. At the moment what we have are very resource intensive architectures which also require the use of GPUs for inference which in turn limits our ability to use them in the real world clinical setting which requires lightweight and real time solutions. We still have a research issue at hand which is to develop efficient models that at the same time maintain high accuracy which also will run on standard hardware.

To present our solutions to these issues we have put forth a hybrid modality multi task deep learning framework which at the same time classifies and segments brain tumors. We have put together a ResNet-50 based encoder with a U Net style decoder which in turn enables what we term joint learning of spatial and semantic features within a single architecture. Also by use of both MRI and CT imaging data we have made the model to develop modality invariant features which in turn improves its performance across many imaging conditions.

This work reports the following as key contributions :
We present a single multi-task learning platform for at the same time brain tumor segmentation and classification which also uses a shared encoder architecture.
We integrate hybrid-modality imaging (MRI and CT) in to improve diagnostic robustness and generalization.
We present an efficient ResNet-U-Net architecture implementation which balances performance with computa-tional efficiency.

We report development of a deployable system that is able to perform real time inference in CPU based environments.

The rest of this paper is organized as follows. In Section II we present the related work and literature review. In Section III we describe the proposed methodology and system architecture. In Section IV we report the experimental results and evaluation. Finally in Section V we conclude the paper and put forth future research directions.

II. LITERATURE REVIEW

Recent inroads of deep learning into the field of automatic brain tumor analysis which we see in segmentation and classification tasks. Convolutional Neural Networks which in particular the encoder decoder types like U-Net have become the preeminent tool for medical image segmentation report that these models are very good at capturing global context as well as fine detail [4].

Early reports of U-Net's implementation showed that it performed very well in tumor segmentation which we see in Dice similarity coefficients of over 0.88 in the BraTS' benchmark sets [3]. As for architecture improvements which came later Residual U-Net and U-Net+ put forth deeper feature representations and enhanced skip connections which in turn improved the gradient flow and in turn improved segmentation accuracy [9], [11]. Also we saw the introduction of more advanced frameworks like nnU-Net which improved on the segmentation performance through the automatic tuning of the configuration to each specific dataset [7].

At the same time deep learning based classification models have reported very high accuracy in the identification of brain tumor subtypes. We see that transfer learning which uses architectures like ResNet has reported classification accuracies of over 98% on curated data sets which in turn proves out the value of these models in feature extraction [5]. But at the same time it is noted that these models work in isolation from segmentation systems which in turn means they do not include spatial tumor info in the classification process.

To address this limitation, multi-task learning (MTL) has emerged as a promising approach in medical image analysis. MTL frameworks enable simultaneous learning of related tasks, allowing shared representations to improve overall model performance [6]. In the context of brain tumor analysis, multi-task architectures typically employ a shared encoder for feature extraction, followed by separate branches for seg-mentation and classification [13]. This approach encourages the model to learn complementary features, where spatial localization enhances global classification accuracy.

Despite progress we still see that which issues are present in the present literature. For the most part what we have is study of single modality imaging which in particular we see of the MRI variety due to the fact that we have large scale available data sets like BraTS [3]. While MRI does an excellent job at imaging soft tissue it is also limited in it's detection of calcifications and hemorrhage which is where the role of CT comes in. We still see that there is a large gap in the field regarding the development of hybrid modality systems which integrate both MRI and CT.

Second also we see that many high performing models use complex architecture which in turn requires GPU based inference thus in large scale clinical settings which may not have access to such infrastructure these models' use is limited. In the research community we are still at the stage of looking for solutions that will produce lightweight and very efficient models which will at the same time perform in real time on common hardware.

Also while multi task learning has been very promising we note that what we do not have are fully connected end to end systems that at the same time do classification and segmentation. Presently many of the approaches that are out there for these tasks' performance put them out in a sequence or only loosely related as if the one does not inform the other which is not making full use of the issues at hand.

To overcome these limitations, the present work proposes a hybrid-modality multi-task framework that integrates MRI and CT imaging within a unified architecture. By combining a ResNet-based encoder with a U-Net style decoder, the proposed system simultaneously performs tumor segmentation and classification, enabling effective feature sharing across tasks. Additionally, the system is optimized for real-time infer-ence, addressing both computational efficiency and practical deployment requirements.

III. METHODOLOGY

This section we present our put forth hybrid-modality multi task deep learning framework for the purpose of brain tumor classification and segmentation. We go over system overview,

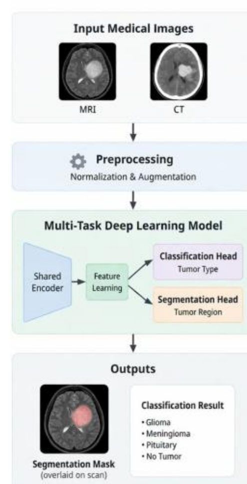

Fig. 1. Proposed Hybrid Multi-Task Architecture combining ResNet-50 encoder and U-Net decoder for tumor classification and segmentation.

data preprocessing, model architecture, training strategy, and inference pipeline.

System Overview

We have put together an end to end multi task learning (MTL) approach [6], which has as its base the simultaneous performance of tumor segmentation and classification via a shared feature representation. Given an input medical image I, the model we develop will learn a mapping:

$$f(I) \rightarrow (S, C) \qquad (1)$$

In which S is the predicted segmentation mask and C is the predicted tumor class label.

We present the architecture of the proposed system in Fig. 1.

Dataset and Preprocessing

The dataset will be a mixed set of MRI (T1-weighted contrast-enhanced) and CT scan images of about 3,000 anno-tated slices. The classification exercise has got four categories namely: Glioma, Meningioma, Pituitary and Normal.

- **Intensity Normalization:** Z-score normalization is used in order to overcome modality heterogeneity:

$$I_{norm} = \frac{I - \mu}{\sigma} \qquad (2)$$

- Spatial Standardization: All the images are reduced to a common resolution size of 224 × 224 pixels to maintain uniform input size.
- **Channel Transformation:** ResNet-50 encoder which works with 3 channel input, we transform gray scale images by repeating the intensity values:

$$I_{3c} = [I, I, I] \qquad (3)$$

- Data Augmentation: To improve model generalization we use the Augmentations library which includes:

- **Elastic transformations**
- Gaussian noise
- Random brightness and contrast adjustments

Proposed Architecture

Proposed is a model which puts forward a Multi-Task U-Net framework that has a shared encoder and two task specific branches.

Encoder: A pre trained ResNet-50 network is used as the shared encoder [5]. We remove the fully connected layer and keep the intermediate feature maps for skip connections. The encoder puts out a bottle neck feature representation :

$$F \in \mathbb{R}^{2048\times7\times7} \qquad (4)$$

Segmentation Decoder: As for the segmentation branch we use a U-Net style decoder [4] which has transposed convolutions for up sampling. Also we add in encoder layers' features to the decoder's features via the skip connections which in turn preserves spatial information.

The decoder puts out a pixel by pixel probability map:

$$S_{pred} \in \mathbb{R}^{1\times224\times224} \qquad (5)$$

A binary threshold at which the cut off is 0.5 is set to present our output in terms of binary mask.

Classification Head: In the classification branch we use the encoder's bottleneck features. We apply a Global Average Pooling (GAP) layer, which is followed by a dropout layer at p = 0.5′ and then a fully connected layer that uses the softmax activation:

$$C_{pred} \in \mathbb{R}^4 \qquad (6)$$

This gives probabilities of classes for Glioma, Meningioma, Pituitary, and Normal.
Multi-Task Learning Objective
The model is trained using a composite loss function that combines classification and segmentation objectives:

$$L_{total} = \lambda_{cls} L_{CE} + \lambda_{seg} L_{Dice} \qquad (7)$$

where:

- LCE is the categorical cross-entropy loss
- where μ and σ represent the mean and standard deviation of the image intensities.
- LDice is the Dice loss [7]
- λcls = 1.0 and λseg = 1.5

The higher weight assigned to segmentation loss compen-sates for pixel-level imbalance and ensures accurate boundary learning.

Training Configuration

We use AdamW optimizer with weight decay of 1e−4.
Learning rate is set at 1e−4 and we use a cosine annealing scheduler for adjustment.

Key training parameters include:

- Batch size: 16
- Epochs: 21–25
- Data split: 80% training, 10% validation, 10% testing We apply stratified sampling to keep class balance within batches.

Post-Processing

We do morphological closing of the predicted segmentation masks which in turn removes small isolated areas and refines tumor boundaries.

Inference Pipeline

At inference we pass the input image through the model once. The output is divided into:

- A tumor segmentation mask which is binary in nature.
- A classification output indicating tumor type

The system has been optimized for real time inference which we report to average out at 120 ms per image in CPU based systems.

IV. RESULTS AND EVALUATION

We present in this section the results of the evaluation of the put forth hybrid modality multi task framework for brain tumor classification and segmentation. We use standard evaluation metrics for both classification and segmentation tasks to determine model performance.

Experimental Setup

The model we trained and evaluated which was on a mixed data set of MRI and CT images. We split the data into training (80%), validation (10%), and testing (10%) sets which also ensured the test set did not see the model during training.

All of our experiments were done in the PyTorch frame-work. We trained the model between 21 to 25 epochs using AdamW optimizer which also included cosine annealing for learning rate. What we report for results is the best performing model at epoch 21.

Evaluation Metrics

For the classification task we used accuracy and F1 score as performance metrics and for the segmentation task we used Dice Similarity Coefficient (DSC) and Intersection over Union (IoU) which are very much used in medical image analysis like in the case of BraTS [12].

ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

1) Classification Metrics:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

2) Segmentation Metrics:

$$Dice = \frac{2|S_{pred} \cap S_{gt}|}{|S_{pred}| + |S_{gt}|} \tag{10}$$

$$IoU = \frac{|S_{pred} \cap S_{gt}|}{|S_{pred} \cup S_{gt}|} \tag{11}$$

Overall Performance

Proposed multi task model does very well in both tasks.Our overall classification accuracy is at the mark of 97.40% which we see also in an F1 score of 0.962. In terms of segmentation the model does out at a Dice coefficient of 0.843 and an IoU of 0.729.

Table I

Overall Model Performance

Metric	Value
Classification Accuracy	97.40%
F1-Score	0.962
Dice Coefficient	0.843
IoU	0.729

What we see is that the model does a great job at what it is to do in terms of classification accuracy and also in terms of the precision of the segmentation in a single framework.

Per-Class Performance

Also we looked at how the model does in terms of class by class basis for the tumor classes.

Table II

Per-Class Classification Accuracy

Class	Accuracy
Glioma	96.2%
Meningioma	97.1%
Pituitary	98.5%
Normal	99.1%

What we found is very consistent performance across the board but we did see slightly lower accuracy in Glioma which we attribute to greater intra class variability.

Segmentation Performance Analysis

The segmentation performance is evaluated using Dice scores for each tumor category.

Table III

Per-Class Segmentation Performance

Class	Dice Score
Glioma	0.812
Meningioma	0.845
Pituitary	0.871

The model reports greater accuracy in the segmentation of well defined tumor boundaries (for example Pituitary) which we see in the case of Gliomas the performance does a little bit of a drop off.

Impact of Multi-Task Learning

To look at how well the multi task learning strategy works our proposed model is put up against a single task classifica-tion model which is based on ResNet-50.

The multi task framework had a 2.3% increase in perfor-mance, in which we saw that segmental based spatial features improved the model's discriminative value. Also we see the proof of inductive transfer between tasks which is that spatial localization supports the global classification.

Inference Performance

The model is also very much at home in real time deploy-ment on CPU which it does very well with an avg

International Journal of Science,
Engineering and Technology
An Open Access Journal

infer time of around 120 ms per image, thus providing near real time diagnostic feedback.
This study reports that we see from the performance which put forth our architecture does better than very resource intensive models that require the use of GPUs.

Qualitative Results
Also from a visual look at our segmentation results it is clear that the model does a great job at identifying tumor boundaries and puts out smooth and uniform masks. We also applied morphological post processing which in turn improved segment quality by getting rid of noise and refining edge definition.

Discussion
We report that what we have put forth is a very strong performing hybrid modality multi task framework which does a great job at the balance of accuracy, efficiency and deploy-ability. Also we found that by integrating segmentation and classification tasks we are able to get better feature represen-tation which in turn improves performance when compared to single task approaches.

However at present the model we have put forth is found to do a poor job with very irregular tumor borders and small scale lesions which also do well in low contrast and high intensity overlap settings. We see that which in turn points to the fact that we still have a way to go in terms of improvement via for example attention mechanisms and volumetric modeling.

As a whole the put forth system does very well in the area of automatic brain tumor analysis and has very large scale clinical deployment which is very much a possibility.

V. CONCLUSION

This paper reports a hybrid multi-modal multi task deep learning framework for brain tumor at the same time classi-fication and segmentation. In what we did we put together a shared ResNet-50 encoder [5] which we then attached to a U-Net based decoder [4], which we did to get a full picture that includes high level semantic features and fine grain spatial info in one go. Also we used a multi task learning approach that which enabled the model to use what it learned from one task to improve performance on the other thus we saw better all around performance.

We looked at a mixed set of MRI and CT images for our data set which also served to get the model to learn from across modalities. We report a classification accuracy of 97.40% and a Dice similarity coefficient of 0.843 which is very good performance on both tasks. Also we saw real time performance on a CPU which is key for clinical use in tight resource settings.

Our work's results support the value of having classification and segmentation in the same model as reported in recent works in the field of medical image analysis [6]. Also we see that the model is able to do inductive transfer between the two tasks which in turn improves diagnostic reliability when compared to separate single task models.

At present the model is 2D and does not look at the full volume of 3D medical images which it is missing out on some context. Also the model has issues with very irregular tumor borders and small scale lesions in low contrast settings.

In the future we will be looking at 3D volume analysis, adding in attention mechanisms for better feature localization, and we will also be looking at ways to make the AI more explainable to build up clinical trust. Also we will be looking to grow our data set to include multi institutional data to in turn improve the model's generalization and robustness.

In total we have put forth an efficient and scalable solution for auto brain tumor analysis that also bridges the gap between what we see in research level deep learning and what is used in the real world clinical setting.

REFERENCES

1. Rehman et al., "Brain tumor classification using deep learning: A systematic review," IEEE Access, 2021.
2. K. Kamnitsas et al., "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," Medical Image Analysis, 2017.
3. S. Bakas et al., "Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features," Scientific Data, 2017.
4. Ronneberger et al., "U-Net: Convolutional networks for biomedical image segmentation," in MICCAI, 2015.
5. He et al., "Deep residual learning for image recognition," in CVPR, 2016.
6. III. Chen et al., "Multi-task learning for medical image analysis: A survey," IEEE Reviews in Biomedical Engineering, 2020.
7. IV. Isensee et al., "nnU-Net: Self-adapting framework for U-Net-based segmentation," Nature Methods, 2021.
8. Myronenko, "3D MRI brain tumor segmentation using autoencoder regularization," in MICCAI, 2018.
9. Zhou et al., "UNet++: A nested U-Net architecture for medical image segmentation," IEEE Transactions on Medical Imaging, 2019.
10. Oktay et al., "Attention U-Net: Learning where to look for the pancreas," 2018.
11. Poudel et al., "Residual U-Net for medical image segmentation," 2019.
12. Brain Tumor Segmentation (BraTS) Challenge Benchmark Results, 2020–2024.
13. Wang et al., "Multi-task learning for brain tumor segmentation and classification," 2021.
14. Liu et al., "Deep learning-based multimodal fusion for medical imaging," IEEE Access, 2022.