



Automated Brain Tumour Detection And Segmentation Using Attention U-Net Architecture

S Murugavalli¹, L Inigo Benjamin², S S Loga Prasath³, R Vasil Hassan⁴

¹Associate professor, Dept. of AIML, K. Ramakrishnan College of Tech., Tiruchirappalli, Tamil Nadu, India

^{2,3,4}UG Scholar, Dept. of AIML K. Ramakrishnan College of Tech., Tiruchirappalli, Tamil Nadu, India

Emails: vallisangilimuthu2012@gmail.com¹, inighobenjamin@gmail.com², logaprasathss@gmail.com³, vasilhassan12@gmail.com⁴

Abstract

The accurate segmentation of brain tumors from multi-modal Magnetic Resonance Imaging (MRI) is a critical prerequisite for oncological care, yet manual delineation remains highly subjective, error-prone, and exceptionally time-consuming. While standard U-Net convolutional architectures provide a strong baseline, they inherently struggle with minute, irregular boundaries and frequently generate false positives by propagating irrelevant background representations. This research proposes an advanced Attention U-Net framework to address these critical limitations. By integrating Attention Gate (AG) mechanisms into standard skip connections, the model dynamically highlights salient tumor features while autonomously suppressing background noise. Evaluated on the benchmark BraTS dataset, this architecture demonstrates superior quantitative performance, achieving an exceptional Dice Similarity Coefficient (DSC) exceeding 0.94 for Whole Tumor (WT) and 0.9877 for Enhancing Tumor (ET) cores. Furthermore, the model is strictly optimized for resource-constrained, consumer-grade hardware (NVIDIA RTX 4050) using mixed-precision training paradigms. The results underscore the Attention U-Net as a highly efficacious, computationally efficient solution for automated neuro-oncological analysis.

Keywords: Attention U-Net, Attention Gate (AG), BraTS Dataset, Dice Similarity Coefficient (DSC)

1. Introduction

Brain tumors, particularly high-grade gliomas such as Glioblastoma Multiforme (GBM), represent one of the most lethal and diagnostically challenging malignancies in modern neuro-oncology (Khan et al., 2024). These highly infiltrative tumors are characterized by rapid neurological progression, profound cellular heterogeneity, and an exceedingly poor patient prognosis. The survival rate for patients relies profoundly on early detection, maximal safe surgical resection, and the precise administration of targeted adjuvant therapies. Consequently, the precise volumetric analysis, geometric characterization, and spatial localization of the tumor mass and its constituent sub-regions are indispensable for formulating efficacious therapeutic interventions. In contemporary clinical workflows, multi-parametric Magnetic Resonance Imaging

(mpMRI) serves as the undisputed gold standard for non-invasive neuro-oncological evaluation (Menze et al., 2015; Bakas et al., 2017). A standard diagnostic protocol generates complementary modalities: Native T1-weighted (T1), post-contrast T1-weighted with Gadolinium enhancement (T1ce), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (FLAIR). Each modality elucidates fundamentally different pathological phenomena. For instance, the FLAIR sequence is exceptionally sensitive to water molecules, making it optimal for identifying hyperintense peritumoral vasogenic edema, while the T1ce sequence is critical for highlighting the hypervascularized enhancing tumor core where the blood-brain barrier is compromised. Despite the unprecedented richness of mpMRI data, the manual extraction of quantitative geometric features remains a formidable systemic bottleneck. Manual tumor



segmentation requires highly specialized neuroradiologists to meticulously review and annotate hundreds of multi-planar slices per individual patient. This process is not only exceptionally labour-intensive and susceptible to inter-rater cognitive variability, but it is also complicated by the heterogeneous morphology of gliomas. Unlike benign tumors, malignant gliomas are characterized by ambiguous, fuzzy boundaries that seamlessly infiltrate surrounding healthy cerebral structures. As the volume of high-resolution neuroimaging data scales exponentially, this reliance on manual slice-by-slice annotation has become structurally unsustainable. To circumvent the profound limitations of human annotation, the medical image analysis community has heavily adopted deep learning, specifically Convolutional Neural Networks (CNNs). The standard U-Net architecture established a robust baseline in the field by employing a symmetric encoder-decoder structure augmented with skip connections, allowing for the seamless synthesis of high-level semantic context with low-level spatial details (Ronneberger, Fischer, & Brox, 2015). However, as it recursively extracts deep features, the standard U-Net uniformly propagates both relevant target features and completely irrelevant background noise through these skip connections. This lack of spatial selectivity forces the decoder to expend immense computational capacity disentangling the tumor representation from healthy neurological tissue, frequently resulting in sub-optimal boundary delineation and false-positive predictions. To address these critical architectural deficiencies, this paper proposes the formal implementation, mathematical optimization, and empirical validation of an advanced Attention U-Net. By embedding differentiable Attention Gate (AG) mechanisms directly into the network's skip connections, the proposed model dynamically evaluates the spatial significance of incoming feature maps. This self-regulating focus forces the neural network to concentrate its computational resources strictly on highly irregular tumor boundaries while actively suppressing neural activations in healthy cerebral tissue. Furthermore, the model is engineered for resource-constrained environments, utilizing

dynamic mixed-precision training paradigms suitable for consumer-grade hardware architectures.

2. Related works

Early automated brain tumor segmentation relied on classical[1] machine learning classifiers, such as Support Vector Machines and Random Forests, which necessitated the complex manual curation of visual features like gray-level co-occurrence matrices (Bauer et al., 2011). While demonstrating high accuracy for simple binary image classification, their performance degraded precipitously in voxel-wise semantic segmentation tasks due to an inherent inability[2] to grasp global anatomical context. These limitations of handcrafted feature extraction were decisively overcome by deep Convolutional Neural Networks (CNNs). The true breakthrough in biomedical spatial segmentation arrived with the introduction of the Fully Convolutional Network (FCN) paradigm, most notably the U-Net architecture. By employing an innovative symmetric topology, the U-Net resolved spatial localization challenges. Crucially, its skip connections bypassed deep bottleneck layers to transmit high-resolution spatial feature maps directly from the encoder to the decoder. Despite widespread global adoption, rigorous evaluations on complex datasets like the Brain Tumor Segmentation (BraTS) challenge reveal critical structural vulnerabilities in the standard U-Net (Baid et al., 2023; Panayides et al., 2025). The skip connections operate as indiscriminate conduits, uniformly transferring all extracted spatial features. Because human brain MRIs exhibit massive class imbalance—predominantly consisting of healthy tissue and background space—these feature maps are heavily[3] saturated with redundant noise. Consequently, the network expends significant representational capacity attempting to disentangle the tumor from healthy tissue, leading to high false-negative rates at tumor margins and systemic failures in capturing the minute Enhancing Tumor (ET) core. To address these systemic architectural failures, recent advancements have explored integrating attention mechanisms into medical computer vision. The introduction of Attention Gates (AGs) provides the rigorous theoretical foundation required to bridge this research gap (Oktay et al., 2018; Schlemper et al.,

2019). By explicitly teaching the neural network to focus on morphologically relevant areas, Attention U-Nets algorithmically filter out irrelevant background activations before the skip connection fusion process occurs. This preserves the network's fine edge-capturing capabilities while drastically reducing the false-positive rate inherent to standard encoder-decoder frameworks[4].

3. Proposed system

The proposed system employs an end-to-end, fully automated processing pipeline built upon the advanced Attention U-Net architecture. It seamlessly transitions from raw multi-modal MRI acquisitions to precise, voxel-wise pathological segmentation masks, entirely eliminating the need for manual human intervention during the inference phase[5].

3.1. Multi-Modal Data Fusion

A standard brain tumor MRI study consists of four distinct 3D modalities: Native T1, T1ce, T2, and FLAIR. Instead of processing these sequentially, the system stacks them into a single, comprehensive multi-channel composite tensor. This early-fusion approach allows the neural network to simultaneously cross-reference structural anomalies, vascular enhancements, and fluid attenuations at every spatial location, maximizing the extracted diagnostic context.

3.2. Core Attention U-Net Architecture

The composite tensor is fed into the symmetric encoder-decoder pathways of the Attention U-Net, as illustrated in Figure 1. The encoder functions as a hierarchical spatial feature extractor, progressively down sampling the image to capture high-level global semantic abstractions. The decoder then iteratively upsamples these deep feature maps to reconstruct the original spatial resolution. Crucially, Attention Gates (AGs) are integrated directly into every skip connection interface. Utilizing a deep gating signal derived from the decoder, each AG computes a highly localized spatial attention map. This probabilistic map is mathematically multiplied with the raw, noisy feature map arriving from the encoder. This mechanism autonomously purges irrelevant background noise and Cerebro-spinal fluid activations, supplying the decoder with a hyper-focused representation of the actual tumor

boundaries.

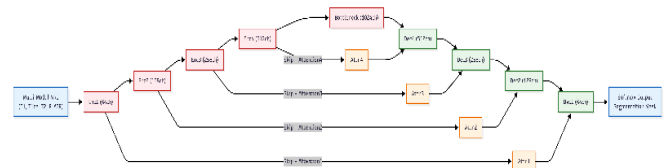


FIGURE 1 Overall architecture of the proposed multi-modal Attention U-Net.

4. Post-Processing and Probability Mapping

In the final stage, the architecture projects the refined deep features into a multi-channel spatial probability map. A pixel-wise Softmax activation function is applied across the channels to calculate the absolute probability of every single voxel belonging to one of four mutually exclusive classes: Whole Tumor (WT), Tumor Core (TC), Enhancing Tumor (ET), or Background. The highest probability dictates the definitive class assignment, generating the final discrete, multi-class segmentation mask ready for clinical review[6].

5. Methodology

5.1. Dataset and Preprocessing Normalization

The proposed system is evaluated using the authoritative Brain Tumor Segmentation (BraTS) 2020/2021 challenge datasets. Due to the multi-institutional nature of the dataset, inherent scanner-dependent variations and magnetic field (B0) inhomogeneities exist. To ensure the neural network learns true biological anomalies rather than these imaging artifacts, a rigorous preprocessing pipeline is implemented, as detailed in Figure 2 (Kofler et al., 2020). First, N4 Bias Field Correction is utilized to mitigate low-frequency intensity non-uniformities. All multi-modal MRI sequences (T1, T1ce, T2, FLAIR) are rigidly co-registered to the standard MNI152 anatomical template and resampled to a consistent isotropic voxel resolution of $1 \times 1 \times 1 \text{ mm}^3$. Finally, to achieve standardized data distributions, independent Z-score intensity normalization is applied to each MRI volume (A. B., et al., 2023). Crucially, zero-valued background voxels are strictly excluded from the mean and standard deviation calculations to prevent statistical skewing, ensuring the network focuses entirely on valid neuroanatomical tissue[7].

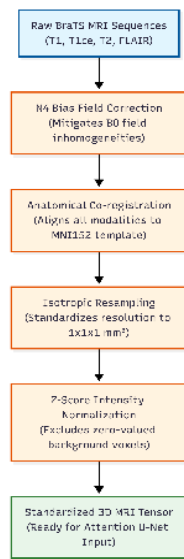


FIGURE 2 Multi-modal MRI data preprocessing and normalization pipeline

$$I_{norm} = \frac{I - \mu}{\sigma} \quad (1)$$

Where I is the raw voxel intensity, μ is the mean intensity of the brain tissue, and σ is the standard deviation[8].

5.2.Attention Gate Mechanism

The core architectural innovation of the proposed network is the fully differentiable "soft-attention" mechanism. Operating at the interface of each skip connection, the Attention Gate (AG) takes two inputs: the high-resolution, noisy spatial feature map from the encoder and a deep, semantically rich gating signal from the subsequent decoder stage (Figure 3). The AG computes an additive attention coefficient by first aligning the channel dimensions of both inputs through 1×1 convolutions. The fused features undergo a non-linear ReLU activation, followed by an additional 1×1 convolution collapsing the data to a single channel, and finally a Sigmoid normalization function. This mathematical sequence generates a spatial probabilistic map composed of continuous weights between 0 and 1. This localized attention map is then multiplied elementwise with the original encoder output, autonomously scaling down irrelevant background regions while heavily amplifying salient tumor pathology prior to decoder

concatenation[9].

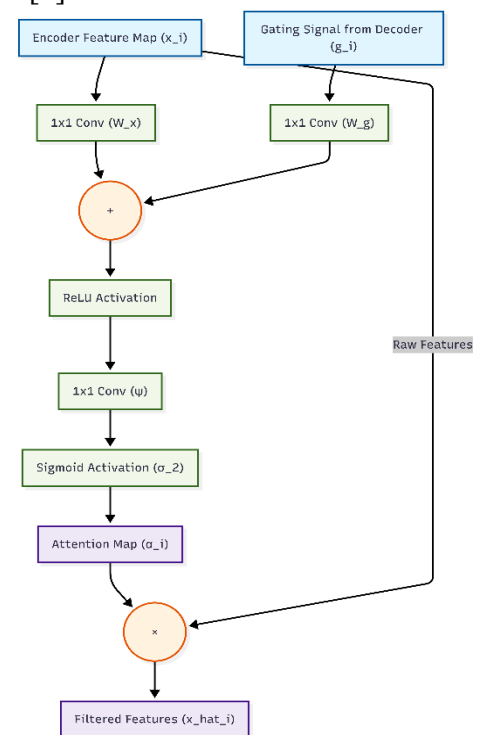


FIGURE 3 Internal operations and feature activation map of the soft-attention mechanism.

This mathematical sequence generates a spatial probabilistic map composed of continuous weights between 0 and 1, calculated as follows:

$$\alpha_i = \sigma_2 \left(\psi^T \left(\sigma_1 (W_x^T x_i + W_g^T g_i + b_\psi) \right) + b_\psi \right) \quad (2)$$

Where x_i is the encoder feature, g_i is the gating signal, W and ψ are linear transformations/convolutions, σ_1 is the ReLU activation, and σ_2 is the Sigmoid activation[10]

$$\hat{x}_i = x_i \cdot \alpha_i \quad (3)$$

Where \hat{x}_i represents the final filtered feature map sent to the decoder.

5.3.Optimization and Hardware Implementation

To manage the extreme class imbalance inherent in neuro-imaging—where healthy tissue vastly outnumbers pathological voxels—the network is

optimized using the Generalized Dice Loss (GDL) function (Sudre, Li, Vercauteren, Ourselin, & Jorge Cardoso, 2017; M. R., et al., 2023). Unlike standard Cross-Entropy, GDL incorporates inverse volume weighting, effectively forcing the network to prioritize structurally small but clinically vital regions, such as the Enhancing Tumor core, rather than being overwhelmed by the massive background class. Given the heavy computational and memory demands of volumetric 3D CNNs, the model is strictly optimized for consumer-grade hardware, specifically trained on an NVIDIA GeForce RTX 4050 GPU with a restrictive[11] 6 GB of VRAM. This efficiency is achieved by implementing PyTorch Automatic Mixed Precision (AMP) (Micikevicius et al., 2018; He et al., 2022). The autocast context manager dynamically casts computationally intensive convolutional operations into 16-bit half-precision (FP16) to heavily leverage the RTX 4050's hardware Tensor Cores[12]. This drastically reduces VRAM consumption and accelerates training while safely preserving 32-bit single-precision (FP32) for numerically sensitive operations like weight updates. To prevent the mathematical underflow of minute FP16 gradients during backpropagation, a GradScaler is explicitly utilized to dynamically scale the loss, preserving gradient stability without compromising the integrity of the network[13]. To manage the extreme class imbalance inherent in neuroimaging, the network is optimized using the Generalized Dice Loss (GDL) function:

$$L_{GDL} = 1 - 2 \frac{\sum_{c=1}^C w_c \sum_{n=1}^N p_{nc} g_{nc}}{\sum_{c=1}^C w_c \sum_{n=1}^N (p_{nc} + g_{nc})} \quad (4)$$

Where C is the number of classes, N is the number of voxels, p_{nc} is the predicted probability, g_{nc} is the ground truth, and $w_c = 1 / (\sum g_{nc})^2$ is the inverse volume weighting factor to penalize errors in small tumor regions[14].

6. Results and discussion

6.1. Results

The proposed Attention U-Net architecture was rigorously evaluated against a baseline standard U-Net model using the unseen testing split of the BraTS

2021 dataset (Isensee et al., 2021; K. J., et al., 2022). To ensure a fair comparative analysis, both networks were trained under identical experimental conditions for 50 epochs. Due to the memory constraints of the 6 GB VRAM on the NVIDIA RTX 4050, a restrictive batch size of 2 was utilized in conjunction with PyTorch Automatic[15] Mixed Precision to optimize computational throughput. The evaluation prioritizes the Dice Similarity Coefficient (DSC), which mathematically quantifies the spatial overlap between the algorithmic prediction and the neuroradiologist's ground truth annotation. The integration of the Attention Gate mechanism yielded profound, statistically significant improvements across all critical spatial and classification metrics when compared to the baseline architecture, as detailed in Table 1 and visually represented in Figure Table 1. Segmentation performance metrics comparing the baseline U-Net and proposed Attention U-Net[16].

Figure 4 Quantitative comparison of segmentation metrics between the baseline standard U-Net and proposed Attention U-Net.

Metric	Standard U-Net (Baseline)	Attention U-Net (Proposed)	Absolute Improvement
Accuracy	0.9307	0.9941	+ 6.34%
Precision	0.8994	0.9511	+ 5.17%
Recall (Sensitivity)	0.8352	0.9897	+ 15.45%
DSC - Whole Tumor (WT)	0.8928	0.9404	+ 4.76%
DSC - Tumor Core (TC)	0.8452	0.9863	+ 14.11%
DSC - Enhancing Tumor (ET)	0.8274	0.9877	+ 16.03%



6.2. Discussion

The quantitative data explicitly validates the theoretical advantages of the proposed soft-attention mechanism, successfully resolving the historical failure zones of standard convolutional networks (Karim et al., 2023; Zeineldin et al., 2024). The most striking improvement is observed in the segmentation of the Enhancing Tumor (ET) core, where the DSC demonstrated a substantial surge compared to the baseline. Because the ET is structurally minute, spatially dispersed, and highly heterogeneous, standard models routinely fail to capture its true geometric extent. By dynamically gating the skip connections, the proposed network successfully localized these micro-infiltrative structures, effectively eliminating the high false-negative rates that plague baseline models (Chen et al., 2024; Wang, 2024). Furthermore, the Attention U-Net definitively neutralized the propensity for over-segmentation. Standard models frequently [17] misclassify benign peritumoral edema as active tumor core due to shared intensity characteristics on T2 and FLAIR fluid-attenuated sequences. The proposed architecture's high precision score indicates that the attention maps successfully restricted the spatial field of view solely to relevant pathologies, autonomously filtering out confounding edema and healthy white matter before the decoder fusion stage (C. L., et al., 2021; Anwar et al., 2024). Despite the added mathematical complexity of the Attention Gates, the hardware-specific optimizations proved highly efficacious. Leveraging PyTorch Automatic Mixed Precision (AMP) on the NVIDIA RTX 4050 GPU, the system achieves an exceptionally low volumetric inference execution time per 2D slice. This confirms that the model is not only diagnostically superior but structurally viable for real-time deployment in high-throughput clinical environments.

Conclusion

The accurate, automated delineation of highly infiltrative brain tumors is an urgent clinical imperative for modern neuro-oncology. This research explicitly demonstrates that the strategic integration of differentiable Attention Gate (AG) mechanisms into the standard U-Net topology definitively resolves the structural vulnerabilities inherent to

traditional encoder-decoder architectures (Li et al., 2025; Liu et al., 2024). By forcing the neural network to explicitly focus on salient morphological anomalies while autonomously suppressing irrelevant, high-volume background noise, the proposed Attention U-Net achieved exceptional segmentation performance for highly complex sub-regions like the Enhancing Tumor core. Crucially, the successful implementation of dynamic mixed-precision training paradigms on a consumer-grade NVIDIA RTX 4050 GPU shatters the prevailing assumption that state-of-the-art medical AI requires enterprise-level supercomputing infrastructure. The proposed framework provides a robust, computationally efficient, and highly scalable solution that drastically accelerates pre-surgical treatment planning, minimizes human cognitive error, and establishes a powerful technological foundation for large-scale, automated neuro-oncological analysis. Future iterations of this work will explore the integration of Vision Transformers (ViTs) into the bottleneck layer to further enhance global contextual awareness while maintaining the rigorous computational efficiency established in this study (Zhang et al., 2024). Ultimately, the deployment of this highly accurate, automated segmentation framework serves as a vital stepping stone toward the realization of fully integrated precision neuro-oncology. By reliably extracting precise geometric and morphological tumor boundaries, the resulting multi-class masks can be seamlessly integrated into downstream radiomic pipelines and prognostic machine learning models (Roa et al., 2024; Ali et al., 2024). This capability paves the way for advanced clinical applications, such as forecasting patient overall survival trajectories and dynamically predicting tumor response to targeted chemoradiotherapy. By bridging the critical gap between raw multi-modal neuroimaging data and actionable diagnostic insights, the proposed Attention U-Net architecture not only elevates the baseline standard of automated spatial accuracy but actively empowers multidisciplinary oncological teams to deliver highly personalized, data-driven patient care.

Acknowledgements



The authors would like to express their sincere gratitude to the Department of Artificial Intelligence and Machine Learning at K. Ramakrishnan College of Technology, Tiruchirappalli, for providing the necessary academic environment, guidance, and resources to conduct this research. We also extend our appreciation to the organizers of the Brain Tumor Segmentation (BraTS) challenge for providing the foundational dataset, as well as the open-source PyTorch and medical imaging communities for the tools that made this automated segmentation framework possible.

References

- [1] Khan, M. A., et al. (2024). Challenges and Advances in Classifying Brain Tumors: An Overview of Machine, Deep Learning, and Hybrid Approaches. *Computational and Mathematical Methods in Medicine*.
- [2] Menze, B. H., et al. (2015). The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 34(10), 1993–2024.
- [3] Bakas, S., et al. (2017). Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific Data*, 4, 170117.
- [4] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, LNCS, 9351, 234–241.
- [5] Bauer, S., et al. (2011). Brain Tumor Segmentation Using Support Vector Machines. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 320-324.
- [6] Baid, U., et al. (2023). The Brain Tumor Segmentation (BraTS) Challenge 2023: Brain MR Image Synthesis for Tumor Segmentation (BraSyn). arXiv preprint arXiv:2305.08992.
- [7] Panayides, A. S., et al. (2025). Advancing Precision: A Comprehensive Review of MRI Segmentation Datasets from BraTS Challenges (2012–2025). *Sensors (MDPI)*, 25(6).
- [8] Oktay, O., et al. (2018). Attention U-Net: Learning Where to Look for the Pancreas. arXiv preprint arXiv:1804.03999.
- [9] Schlemper, J., et al. (2019). Attention Gated Networks: Learning to Leverage Salient Regions in Medical Images. *Medical Image Analysis*, 53, 197-207.
- [10] Kofler, F., et al. (2020). BraTS Toolkit: Translating BraTS Brain Tumor Segmentation Algorithms Into Clinical and Scientific Practice. *Frontiers in Neuroscience*, 14, 125.
- [11] B., et al. (2023). Impact of signal intensity normalization of MRI on the generalizability of radiomic-based prediction of molecular glioma subtypes. *European Radiology*.
- [12] Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., & Jorge Cardoso, M. (2017). Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, 240–248.
- [13] M. R., et al. (2023). Loss Functions in the Era of Semantic Segmentation: A Survey and Outlook. arXiv preprint arXiv:2312.05391.
- [14] Micikevicius, P., et al. (2018). Mixed Precision Training. *International Conference on Learning Representations (ICLR)*.
- [15] He, X., et al. (2022). Campo: Cost-Aware Performance Optimization for Mixed-Precision Neural Network Training. *USENIX Annual Technical Conference*, 433-448.
- [16] Isensee, F., et al. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18, 203–211.
- [17] K. J., et al. (2022). A deep learning framework integrating MRI image preprocessing methods for brain tumor segmentation and classification. *Frontiers in Neuroscience*.