AI-Powered Brain Tumor Diagnostic Tool

1. Data Preprocessing Pipeline

The data preprocessing phase involved using Keras' ImageDataGenerator to prepare and augment the brain tumor image dataset. Various transformations such as rotation (up to 30 degrees), zoom (up to 30%), width and height shifts, shear transformations, and horizontal flips were applied to enhance model robustness and prevent overfitting. All images were resized to 224×224 pixels and normalized by rescaling pixel values to the [0,1] range. The dataset was organized into training and testing directories with four classes: glioma, meningioma, pituitary, and no tumor.

2. Model Architecture

The model was built using **Transfer Learning** with the ResNet50 architecture pretrained on ImageNet. The initial layers of ResNet50 were frozen, and deeper layers were fine-tuned to adapt to the medical imaging domain. On top of the base model, a global average pooling layer was added, followed by dropout layers to reduce overfitting, a dense layer with ReLU activation, and a final softmax layer for multi-class classification across the four tumor types.

3. Model Training and Optimization

The model was compiled using the **Adam optimizer** with a learning rate of 0.0001 and trained using categorical cross-entropy loss. Key callbacks were used during training:

- **ModelCheckpoint** to save the best model based on validation accuracy,
- **EarlyStopping** to prevent overfitting by halting training if no improvement was seen, and
- **ReduceLROnPlateau** to decrease the learning rate when the model plateaued on validation loss.
  The model was trained for 25 epochs with a batch size of 32.

4. Model Evaluation

Post training, the model was evaluated on the test dataset containing 1311 images. It achieved an overall testing accuracy of 86.88%, demonstrating strong generalization to unseen data. During training, the model reached an accuracy of approximately 94.25% on the training set, indicating effective learning while avoiding overfitting due to regularization techniques such as dropout and data augmentation.

The model also performed well in terms of precision, recall, and F1-score across all four tumor classes: glioma, meningioma, pituitary, and no tumor. Notably, the 'no tumor' and 'pituitary' classes showed exceptional results with F1-scores exceeding 0.90. This performance reflects the model’s robustness and suitability for medical imaging tasks related to brain tumor classification.

5. API Specifications

To enable real-time predictions, a REST API was developed using frameworks like **Flask or FastAPI**. The API accepts an input image, preprocesses it in the same way as training data, and returns the predicted tumor class with its confidence score.

6. Limitations

The model's performance may degrade on extremely low-resolution or noisy images, or on data from different imaging sources not seen during training. It may also struggle with tumors that exhibit characteristics of multiple classes. Additional validation on larger and more diverse datasets is needed before clinical deployment.

7. Deployment

The trained model and API were **containerized using Docker** to ensure consistent deployment across environments. The complete application was then **deployed on Hugging Face Spaces with GPU support**, enabling fast, real-time predictions via a browser-accessible interface. This platform also facilitates easy sharing, collaboration, and testing.