Data Mining in Brain Tumor Detection (Literature Review and Methodology)

By: Tariere Timitimi, Ana Sytina, Tahmidul Haque

Literature Review

Brain tumor detection is critical for improving patient outcomes, and recent advancements in deep learning and data mining offer promising tools for enhancing diagnostic accuracy, efficiency, and accessibility.

Deep learning models, especially Convolutional Neural Networks (CNNs), have shown significant potential in automating the detection process by learning complex features directly from medical images. Abdusalomov et al. refined the YOLOv7 model for accurate detection of meningioma, glioma, and pituitary gland tumors, achieving an overall accuracy of 99.5%. This refined model incorporates image enhancement techniques to improve the visual representation of MRI scans and uses data augmentation to enhance the training dataset. The inclusion of a Convolutional Block Attention Module (CBAM) improves feature extraction, while the Spatial Pyramid Pooling Fast+ (SPPF+) layer and Bi-directional Feature Pyramid Network (BiFPN) enhance sensitivity and multi-scale feature fusion. The enhanced YOLOv7 model demonstrates its potential as a decision-making tool for experts in diagnosing brain tumors and its usefulness in monitoring and detecting brain tumors using MRI. Additionally, Saeedi et al. used 3264 MRI brain images to develop a new 2D CNN and a convolutional auto-encoder network, achieving optimal accuracy of approximately 95% to 96%. Furthermore, the principles of quantum rotation matrices have been applied to traditional algorithms for feature selection. Bilal et al.'s Q-BGWO-SQSVM model, which uses quantum-infused techniques, achieved top-tier results in accuracy, sensitivity, specificity, precision, F1 Score, and Matthews Correlation Coefficient (MCC) across diverse medical image datasets.

Image pre-processing techniques play a crucial role in improving the accuracy of brain tumor detection in MRI images. These techniques aim to remove noise and extraneous elements from the images, with common steps including noise reduction, grayscale conversion, and smoothing and sharpening procedures. Some techniques also apply global thresholding, adaptive thresholding, Sobel filters, and high-pass filters to improve image quality, as well as normalization to convert brain images into intensity brain images using a min–max normalization rule. To improve the readability of low-resolution MRI images, a three-stage image preparation strategy can be used.

Various data mining and machine learning algorithms also contribute to brain tumor detection. Saeedi et al. showed that:

- Support Vector Machines (SVMs) classify brain tumors with high accuracy when combined with other techniques.
- Extreme Learning Machines (ELM) apply Regularized Extreme Learning Machine (RELM) classification to identify and classify tumor types.
- K-Nearest Neighbors (KNN) is used for brain tumor classification and has high precision and recall rates.
- Random Forest (RF) classifiers demonstrate effectiveness in classifying different types of tumors.
- Multi-layer Perceptron (MLP) classifiers are also used, though with lower accuracy rates compared to other methods.

Despite these advancements, there are still challenges to address. One such challenge is detecting small-size brain cancers, though integrating SPPF+ and BiFPN components helps focus on localized tumors. Data diversity can be increased through data augmentation strategies when available data is low. Future studies should incorporate a diverse collection of clinically relevant brain lesions and additional imaging modalities to enhance models' segmentation accuracy. Moreover, challenges remain in the accurate diagnosis of tumors due to the radiologist's experience and potential fatigue, which computational intelligence-oriented techniques can assist with. Future research should focus on developing deep neural networks that are robust, simple, and have less execution time for a more rapid and accurate diagnosis.

Methodology

This study will work on creating a predictive model by integrating data mining techniques and deep learning algorithms to extract and analyze features indicative of brain tumor presence in MRI images. The following sections detail the procedures for data collection and preprocessing, exploratory data analysis, model development and training, model evaluation and optimization, and the final documentation of results.

Data Collection and Preprocessing

Using the Crystal Clean: Brain Tumors MRI Dataset acquired from Kaggle, we will form the basis of our analysis. During this phase, the dataset is carefully examined to ensure it is comprehensive and structured for subsequent analysis. Data preprocessing involves handling missing values, normalizing pixel intensities, and performing data augmentation where necessary. Standardization and normalization techniques are employed to ensure that the input features are on comparable scales.

Exploratory Data Analysis

During exploratory data analysis, we use a variety of data mining tools to identify underlying patterns, trends, and correlations within the dataset. This involves generating summary statistics and visual representations. This phase provides insights into the distribution of tumor types, the variability of imaging features, and potential outliers. These insights help the feature selection process and assist in understanding the complex nature of brain tumor imaging data.

Model Development and Training

The core of this research lies in the development of several predictive models using a combination of traditional machine learning methods and advanced deep learning techniques. The models under consideration include Decision Trees, Support Vector Machines (SVMs), and Neural Networks. In order to apply the research done by Abdusalomov et al. and Saeedi et al., we can apply a 2D CNN to extract and classify features from the MRI images and a YOLOv7 model for localization and detection. The combination of these two models enhances the feature extraction and classification processes, which results in a more accurate identification of brain tumors. During model development, the dataset is partitioned into training, validation, and test sets. The training process involves tuning of hyperparameters to optimize performance. Cross-validation is used to ensure the robustness of the model, while dropout and regularization techniques are applied to prevent overfitting.

Model Evaluation and Optimization

Upon training the predictive models, the performance is evaluated using key metrics such as accuracy, precision, recall, and F1-score where the diagnostic ability is assessed by comparing the model's predictions against the ground truth labels from the dataset. By completing this step, we are able to measure the effectiveness of the models and identify the most efficient algorithm for early brain tumor detection. We can also use these evaluations to tune model parameters to improve sensitivity and specificity.

Final Phase

The final phase of this research involves a report on the methodology, experimental setup, results, and conclusions where a discussion of the findings, the limitations of the study, and recommendations for future research are presented.

References

Abdusalomov, A. B., Mukhiddinov, M., & Whangbo, T. K. (2023). Brain tumor detection based on deep learning approaches and magnetic resonance imaging. Cancers, 15, 4172. https://doi.org/10.3390/cancers15164172

World Health Organization. (2021). Artificial intelligence is changing the health sector. *WHO Consultation Towards the Development of Guidance on Ethics and Governance of Artificial Intelligence for Health.* Retrieved from https://www.jstor.org/stable/resrep35680.7

Saeedi, S., Rezayi, S., Keshavarz, H., & Niakan Kalhori, S. R. (2023). MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques. BMC Medical Informatics and Decision Making, 23, 16. https://doi.org/10.1186/s12911-023-02114-6

Manjunathan, N., & Gomathi, N. (2025). A comparative analysis of the health monitoring process using deep learning methods for brain tumor. *Measurement: Sensors, 37*(101807). https://doi.org/10.1016/j.measen.2025.101807

Crystal Clean: Brain Tumors MRI Dataset. (n.d.). Www.kaggle.com. https://www.kaggle.com/datasets/mohammadhossein77/brain-tumors-dataset/data

Hemanth, G., Janardhan, M., & Sujihelen, L. (n.d.). Design and implementing brain tumor detection using machine learning approach. IEEE Xplore. Retrieved from https://ieeexplore.ieee.org/

Santos, A., Malheiros, B., Cavalheiros, C., & Olivera, D. (n.d.). A data mining system for providing analytical information on brain tumors to public health decision makers. ScienceDirect. Retrieved from https://www.sciencedirect.com/