Irish Interdisciplinary Journal of Science & Research (IIJSR)
Volume 8, Issue 2, Pages 86-93, April-June 2024

Brain Tumor Malignancy Prediction Using Machine Learning Techniques

Seethalakshmi B.*

*Department of Artificial Intelligence and Data Science, Sri Shanmuga College of Engineering and Technology, Sankari, Tamilnadu, India. Email: seethakrish1719@gmail.com**

DOI: https://doi.org/10.46759/IIJSR.2024.8210

Article Received: 09 March 2024 *Article Accepted: 11 May 2024* *Article Published: 14 May 2024*

ABSTRACT

It is crucial to detect cancer early in order to save many lives. If a brain tumor is detected at a higher grade, it is often one of the most prevalent and severe malignant tumor disorders, with a very short predicted life. The differences in tumor size, form, and location present a significant obstacle to the detection of brain tumors. This survey aims to provide researchers with a thorough literature review on magnetic resonance imaging (MRI)-based brain tumor detection. The MRI images were classified using nine machine learning algorithms: Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbor (KNN), Naïve Bayes (NB), Decision Tree (DT) classifier, Random Forest classifier, XGBoost classifier, Stochastic Gradient Descent (SGD) classifier, and Gradient Boosting classifier. The ML algorithms have been compared and contrasted.

Keywords: Brain tumor segmentation; Computer-aided diagnostics; Deep learning; Machine learning; MRI.

1. Introduction

Since brain tumors are the leading cause of cancer-related mortality in children and adults under 40, it is imperative to encourage early diagnosis. As a result, methods for accelerating the early diagnosis of brain cancers must be developed. Early brain tumor diagnosis means a quicker response to treatment, which raises patient survival rates. It would be ideal to have a system that can identify, locate, and categorize brain tumors automatically. Machine learning has become increasingly popular in nearly every area of decision-making and can be effectively applied to the identification and categorization of brain tumors.

The aim of this work is to investigate the application of machine learning (ML) classification algorithms to identify brain tumors from brain MRI images and to differentiate between different types of brain tumors, such as gliomas, meningiomas, and pituitary tumors. For the diagnosis of brain tumors, a computer-aided categorization method is more trustworthy. A few phases make up the suggested scheme: gathering data; preparing it (labeling data, pre-processing images); classifying it using improved machine learning techniques; and lastly, comparing the models that have been put into practice.

1.1. Types of Brain Tumors

The brain stem, cerebrum, and cerebellum are the three primary regions of the brain. The cerebellum, the brain's second biggest region, controls all of the body's motor functions, including walking, balance, posture, and overall motor coordination. It is attached to the brain stem and situated behind the brain. The cerebellum and cerebrum include internal white matter, extremely thin gray matter outer cortex, and tiny yet deeply positioned quantities of gray matter. The spinal cord and brainstem are connected. It is located at the base of the brain. The brainstem controls all essential body functions, such as motor, sensory, cardiac, repositories, and reflexes. The midbrain, pons, and medulla oblongata make up its three structural elements. An unplanned proliferation of brain cells is referred to in medicine as a brain tumor. Scientists have classified many brain tumor types depending on the location of the origin (primary or secondary) and additional contributing factors, as well as the type of tissue involved and whether the tumors are malignant or benign. Brain tumors were classified by the World Health Organization (WHO) into 120 different types. This classification, which goes from less aggressive to more aggressive, is based on the origin and behavior of the cell. Grades I through IV represent the least malignant and most malignant tumor types, respectively.

Grade I: These tumors grow slowly and do not spread rapidly. These are associated with better odds for long-term survival and can be removed almost completely by surgery. An example of such a tumor is grade 1 pilocyticastrocytoma.

Grade II: These tumors also grow slowly but can spread to neighboring tissues and become higher grade tumors. These tumors can even come back after surgery. Oligodendroglioma is a case of such a tumor.

Grade III: These tumors develop at a faster rate than grade II, and can invade the neighboring tissues. Surgery alone is insufficient for such tumors, and post-surgical radiotherapy or chemotherapy is recommended. An example of such a tumor is anaplastic astrocytoma.

Grade IV: These tumors are the most aggressive and are highly spreadable. They may even use blood vessels for rapid growth. Glioblastoma multiforme is such a type of tumor

2. Literature survey

A lot of research work has been done in the field of Artificial Intelligence (AI) and Machine Learning (ML) A Comparative Study of Enhanced Machine Learning Algorithms for Brain Tumor Detection and Classification application in the field of medical imaging. Noreen et al. have proposed the use of two pre-trained deep learning models i.e. Inception-v3 and DensNet201 for developing a multi-level feature extraction and concatenation method for the early detection of brain tumors and their classification. At first, they have extracted the features from different Inception modules from the pre-trained Inception-v3 model. Subsequently, the softmax classifier was given those features to classify the brain tumors. Second, they have extracted features from different DensNet blocks using a pre-trained DensNet201. In order to classify the brain tumors, they concatenated the features and fed them to the softmax classifier. The three classifications of brain tumors included in the dataset they used are publicly available. In terms of brain tumor identification and classification, their suggested methodology has surpassed all current machine learning (ML) and deep learning (DL) models, yielding remarkable results.

The decision tree classification method was employed by Naik and Patel to identify and categorize brain tumors from MRI pictures. They employed the textural feature extraction technique and the median filtering process in the pre-processing stage to extract the features. Their suggested model has demonstrated enhanced efficacy when juxtaposed with conventional image mining techniques. Their findings have been juxtaposed with the outcomes of the Naïve Bayesian classification method. The decision tree classification algorithm has achieved a precision of 100%, Sensitivity of 93%, Specificity of 100% and Accuracy of 96%.

Sarhan has demonstrated a method for classifying brain cancers in MRI images using computer-aided detection (CAD). The Discrete Wavelet Transform (DWT) has been used to extract the characteristics from the brain MRI

Irish Interdisciplinary Journal of Science & Research (IIJSR)
Volume 8, Issue 2, Pages 86-93, April-June 2024

images. The input MRI image has been classified using a CNN after the retrieved features have been applied. Overall accuracy has been produced by his suggested method.

Tandel et al. have presented a Convolutional Neural Network (CNN) based transfer-learning AI paradigm for classifying brain tumors from MRI data. Six distinct machine learning (ML) classification methods have been used as benchmarks for the transfer-learning-based CNN model: Decision Tree, Linear Discrimination, Naive Bayes, Support Vector Machine, K-nearest neighbor, and Ensemble. When it comes to multiclass brain tumour grading, their suggested model has shown to be highly helpful and has produced superior outcomes than the other ML models.

In their study, Mohsen et al. suggested creating a Deep Neural Network (DNN) classifier for the categorization of brain tumors using 66 brain MRI images representing four different types of brain tumors: normal, glioblastoma, sarcoma, and metastatic bronchogenic carcinoma tumors. Principal components analysis (PCA) and feature extraction have been performed by combining the classifier with DWT. With an average recall of 0.97, average precision of 0.97, average F-Measure of 0.97, average area under the ROC curve (AUC) of 0.984 of all four classes (normal, glioblastoma, sarcoma, and metastatic bronchogenic carcinoma tumors), the DNN classifier produced incredibly good results.

Rehman et al. have carried out three investigations employing three convolutional neural network architectures (AlexNet, GoogLeNet, and VGGNet) to classify brain cancers, including meningioma, glioma, and pituitary. Then, using MRI slices from the Figshare dataset of brain tumors, they investigated transfer learning strategies, i.e., freeze and finetune. To enhance the dataset samples, decrease the likelihood of over-fitting, and generalize the results, they have employed data augmentation techniques to the MRI pictures. In terms of classification and detection, the suggested fine-tuned VGG16 architecture has achieved the maximum accuracy, reaching 98.69%.

3. Types of Machine Learning Algorithms

A separative hyper-plane created the supervised machine learning classification technique known as Support Vector Machine (SVM). Finding the optimal method for data segregation is the SVM's primary goal. As a result, the frontier that best divides the two classes is SVM. A supervised machine learning approach called logistic regression is used to forecast a binary result given a set of independent factors. Finding the optimal model to explain the relationship between an outcome and a set of predictor variables is the primary goal of logistic regression. KNN is a supervised machine learning method. Problems involving binary classification are resolved using it. By computing the distance between a given data point and the other points, KNN makes predictions about whether the provided data point belongs to a specific class or not. The data point in question is a member of the class whose members are closest to it. The number of points to be chosen in the neighborhood of the specified data point is denoted by K in a KNN. A supervised machine learning algorithm, Naïve Bayes (NB) is mostly utilized for binary classification. Its foundation is the Bayes theorem, which makes the assumption that the predictors are independent. The NB classifier makes the assumption that a feature's presence in a class is independent of any other feature's presence. A supervised learning technique called Decision Tree (DT) is used to address binary classification issues. DTs anticipate the value of a target variable and learn from basic decision rules deduced from the data features. An ensemble machine learning algorithm called Random Forest creates several DTs before combining them. As a result, it generates more accurate findings. Over-fitting is a possibility with DTs if the dataset is too big. Thus, Random Forest is utilized to prevent data from being overfit. Problems with regression and classification can both be resolved with Random Forest. By changing the values of a function's parameters or coefficients, the effective optimization process known as stochastic gradient descent (SGD) optimizes the cost function. SGD Classifier uses. One of the boosting algorithms in the family is Extreme Gradient Boosting, or XGBoost. It is a proficient execution of the supervised learning technique known as the Gradient Boosted Trees algorithm. This method of ensemble machine learning makes predictions using the Gradient Boosting framework. Boosting is a method for group learning. It creates a model with higher accuracy by combining predictors of lower accuracy. Gradient boosting produces a robust and highly accurate model by having the predictor itself rectify the mistakes caused by the predecessors.

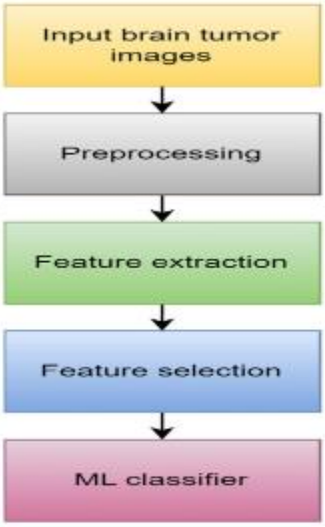

Figure 1. Process Flow for the Suggested Model

3.1. Data Acquisition

We can collect brain cancer images using several imaging modalities such as MRI, CT, and PET. This technique effectively visualizes aberrant brain tissues.

3.2. Preprocessing

In the medical field, preprocessing is a critical step. Typically, preprocessing is when noise removal or enhancement in photos takes place. Image quality is greatly reduced by medical noise, rendering them ineffective for diagnosis. The preprocessing step needs to be efficient enough to remove as much noise as possible from medical images without compromising crucial image elements. Many techniques are used to accomplish this process, such as image scaling, cropping, histogram equalization, median filter filtering, and image adjustment.

3.3. Feature extraction

The process of converting images into features based on several image characteristics in the medical field is known as feature extraction. These features carry the same information as the original images but are entirely different.

This strategy improves classifier accuracy, reduces overfitting risk, allows users to interpret data, and speeds up training. The numerous types of features include texture, contrast, brightness, shape, gray level co-occurrence matrix (GLCM), Gabor transforms, wavelet-based features, 3D Haralick features, and a histogram of local binary patterns (LBP).

3.4. Feature selection

The technique attempts to arrange the features in ascending order of importance or relevance, with the top features being mostly employed in classification. As a result, multiple feature selection techniques are needed to reduce redundant information to discriminate between relevant and nonrelated features, such as PCA, genetic algorithm (GA), and ICA.

3.5. ML algorithms

The following methodologies of ML are implemented:

1. Logistic Regression: A type of supervised machine learning algorithm used for classification tasks that estimates the probability of an outcome based on several independent variables.

2. Decision Tree: A type of supervised machine learning algorithm used for classification and regression tasks that uses a tree-like graph to represent decisions and their possible outcomes.

3. Random Forest: A type of supervised machine learning algorithm used for classification and regression tasks that uses multiple decision trees to make predictions.

4. Naïve Bayes: A type of supervised machine learning algorithm used for classification tasks based on Bayes' theorem that assumes that the features in the data are independent of each other. 5. AdaBoost Algorithm: A type of supervised machine learning algorithm used for classification and regression tasks that combine several weak learners to form a strong learner.

6. CNN (Convolutional Neural Network): A type of artificial neural network used for image recognition and processing that is composed of multiple layers of neurons that analyze and process data from images.

7. ANN (Artificial Neural Network): A type of supervised machine learning algorithm used for classification and regression tasks that is composed of multiple layers of neurons that analyze and process data.

4. Discussions

TP, TN, FP, FN are terms commonly used in the field of statistics and machine learning to describe the performance of a binary classification model.

• TP (True Positive): Refers to the number of cases where the model predicted the positive class correctly, i.e., the case was actually positive, and the model predicted it as positive.

• TN (True Negative): Refers to the number of cases where the model predicted the negative class correctly, i.e., the case was actually negative, and the model predicted it as negative.

• FP (False Positive): Refers to the number of cases where the model predicted the positive class incorrectly, i.e., the case was actually negative, but the model predicted it as positive.

• FN (False Negative): Refers to the number of cases where the model predicted the negative class incorrectly, i.e., the case was actually positive, but the model predicted it as negative.

These terms are important for evaluating the performance of a binary classification model, and are used to calculate metrics such as accuracy, precision, recall, and F1 score.

- True positive (TP) = the number of cases correctly identified as patient
- False positive (FP) = the number of cases incorrectly identified as patient
- True negative (TN) = the number of cases correctly identified as healthy
- False negative (FN) = the number of cases incorrectly identified as healthy

After the evaluation of the test scores, it has been concluded that Gradient Boosting is the best classifier among all the other ML classifiers that have been used. Also, multi-class classification has been performed on a different dataset comprising of brain MRI images of glioma, meningioma, pituitary and no tumor using SVM, KNN, Random Forest and XGBoost classifier. The ML algorithms have been compared based on accuracy, recall, precision, F1-score, AUC-ROC score and it has been observed that XGBoost classifier has exhibited the best results. In future, one of the most important improvements that can be made is adjusting the architecture so that it can be used during brain surgery, for classifying and accurately locating the tumor. Detecting the tumors in the operating theatre can be performed in real-time conditions; thus, in that case, the improvement would also involve adapting the network architecture to a 3D system. By keeping the network architecture simple, detection in real time can be made possible.

Declarations

Source of Funding

This study did not receive any grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing Interests Statement

The authors declare no competing financial, professional, or personal interests.

Consent for publication

The authors declare that they consented to the publication of this study.

References

Ankit Ghosh & Alok Kole (2021). A comparative study of enhanced Machine Learning Algorithms for Brain Tumor Detection & Classification. IEEE.

Javaria Amin, Muhammad Sharif, Anandakumar Haldorai, Mussarat Yasmin, & Ramesh Sundar Nayak (2022). Brain tumor detection and classification using machine learning: a comprehensive survey. Complex & Intelligent Systems, 8: 3161–3183. https://doi.org/10.1007/s40747-021-00563-y.

Janki Naik & Sagar Patel (2014). Tumor Detection and Classification using Decision Tree in Brain MRI. International Journal of Computer Science and Network Security, 14(6): 87–91.

P.P. Bhattacharya, Alok Kole, Tanmay Maity & Ananya Sarkar (2014). Neural Network Based Energy Efficiency Enhancement in Wireless Sensor Networks. International J. of Applied Engineering Res., 9(22): 11807–11818.

Chinmaya Kumar Pradhan, Shariar Rahaman, Md. Abdul Alim Sheikh, Alok Kole & Tanmoy Maity (2018). EEG Signal Analysis Using Different Clustering Techniques. In Proc. International Conference on Emerging Technologies in Data Mining and Information Security, Kolkata, West Bengal, Pages 99–105.

Ahmad M. Sarhan (2020). Detection and Classification of Brain Tumor in MRI Images Using Wavelet Transform and Convolutional Neural Network. Journal of Advances in Medicine and Medical Research, 32(12): 15–16.

Arshia Rehman, Saeeda Naz, Muhammad Imran Razzak, Faiza Akram & Muhammad Imran (2020). A Deep Learning-Based Framework for Automatic Brain Tumors Classification Using Transfer Learning. Circuits, Systems, and Signal Processing, Springer, 39: 757–775.

Khan, M.A., Lali, I.U., Rehman, A., Ishaq, M., Sharif, M., Saba, T., et al. (2019). Brain tumor detection and classification: A framework of marker-based watershed algorithm and multilevel priority features selection. Microsc Res Tech., 82: 909–922.

Johnson, D.R., Guerin, J.B., Giannini, C., Morris, J.M., Eckel, L.J., & Kaufmann, T.J. (2017). 2016 updates to the WHO brain tumor classification system: what the radiologist needs to know. Radiographics, 37: 2164–2180.

Amin, J., Sharif, M., Anjum, M.A., Raza, M., & Bukhari, S.A.C. (2020). Convolutional neural network with batch normalization for glioma and stroke lesion detection using MRI. Cogn Syst Res., 59: 304–311.

Sharif, M.I., Li, J.P., Naz, J., & Rashid, I. (2020). A comprehensive review on multi-organs tumor detection based on machine learning. Pattern Recogn Lett., 131: 30–37.

Amin, J., Sharif, M., Yasmin, M., & Fernandes, S.L. (2020). A distinctive approach in brain tumor detection and classification using MRI. Pattern Recogn Lett., 139: 118–127.

Haseeb Ur Rehman & Sohaib Masood (2021). Brain Tumor Classification using Deep Learning Methods. International Journal of Information Systems and Computer Technologies, 1(1).

Mukul Aggarwal, Amod Kumar Tiwari, M. Partha Sarathi & Anchit Bijalwan (2023). An early detection and segmentation of Brain Tumor using Deep Neural Network. BMC Medical Informatics and Decision Making, 23: 78. https://doi.org/10.1186/s12911-023-02174-8.

Hareem Kibriya, Rashid Amin, Asma Hassan Alshehri, Momina Masood, Sultan S. Alshamrani & Abdullah Alshehri (2022). A Novel and Effective Brain Tumor Classification Model Using Deep Feature Fusion and Famous Machine Learning Classifiers. Hindawi Computational Intelligence and Neuroscience, Article ID 7897669.

Al-shamasneh, A., & Obaidellah, U. (2017). Artificial intelligence techniques for cancer detection and classificiation. Eur Sci J., Pages 342–370.

Alam, S., Rahman, M., & Hossain, M.A. (2019). Automatic human brain tumor detection in MRI image using template-based K means and improved fuzzy C means clustering algorithm. Big Data Cogn Comput., 3(2): 1–18. https://doi.org/10.3390/bdcc3020027.

Al-ayyoub, M., Alabed-alaziz, A., & Darwish, O. (2012). Machine learning approach for brain tumor detection. In Proceedings of the 3rd international conference on information and communication systems, Pages 1–4. https://doi.org/10.1145/2222444.2222467.

Devi, N., & Bhattacharyya, K. (2018). Automatic brain tumor detection and classification of grades of astrocytoma. In International Conference on Computing and Communication Systems, Springer.

Dong, H., Yang, G., Liu, F., Mo, Y., Guo, Y., & Heart, N. (2017). Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks. In Valdés Hernández M, González-Castro V (eds) Medical image understanding and analysis, MIUA 2017, Cham, Springer.