

An Intelligent Brain Tumor Detection System Using Machine Learning and MRI Images

Mrs. Sangita Kapse

Research Scholar

Department of Electronics & Computer Science,
R.T.M.N.U.

Nagpur, Maharashtra, India

Email: pranalipapse@gmail.com

Dr. Aruna J. Chamatkar

Associate Professor and Head of

Department of Computer Science,
Kamla Nehru Mahavidyalaya

Sakkardara Square Nagpur, Maharashtra, India

Email: aruna.ayush1007@gmail.com

Abstract

Brain tumor detection from magnetic resonance imaging (MRI) is a challenging task that requires accurate and efficient analysis of medical images. This paper presents a machine learning and deep learning-based framework for automated brain tumor detection and classification from MRI images. The proposed methodology involves image preprocessing, noise reduction, image resizing, segmentation, and feature extraction to prepare MRI images for classification. A dataset containing more than 2100 normal and abnormal brain MRI images is considered, and the images are standardized to a size of 224×224 pixels. Convolutional Neural Network (CNN) and transfer learning-based approaches are considered for classifying MRI images into four categories: glioma, meningioma, pituitary tumor, and no tumor. The performance of the developed models is evaluated using standard classification measures such as accuracy, precision, recall, and F1-score. The proposed framework aims to provide an efficient computer-aided approach for brain tumor detection and classification and to support radiologists in achieving faster and more consistent analysis of MRI images. The study demonstrates the potential of machine learning and deep learning techniques for developing intelligent medical image analysis systems.

Keywords: Brain Tumor Detection, MRI, Machine Learning, Deep Learning, Convolutional Neural Network, Image Classification, Medical Image Analysis.

1. INTRODUCTION

Although the traditional diagnosis is based on the visual inspection of medical images to detect various kinds of nodule development in the body, large medical image data makes the manual diagnosis difficult. Furthermore, it is somewhat subjective to identify it owing to interviewer variability. Computer Aided Diagnosis (CAD) system is thus introduced to help the radiologist in the diagnosis process of several diseases including colon cancer, breast cancer, bronchial carcinoma, pathological brain detection, Alzheimer's disease, diabetic retinopathy, congenital heart defect, coronary heart disease and so on [1][2]. The CAD system's primary goal is to detect aberrant indicators early on, which human experts overlook, therefore lowering the incidence of incorrect prediction when the surgeons see medical images. CAD is increasingly used in medical fields like genetic engineering, tailored medical transplants, artificial joint repairing, and robotic operations in recent days [3][4][5]. The value of CAD systems in pathology identification is to improve radiologist performance by means of a reasonably priced sensitivity rate. Four key modules often comprise the computer assisted diagnostic system: picture pre-processing, segmentation of the area of interest (ROI), feature extraction and selection, and lastly ROI classification. Segmentation is the technique of breaking the picture into little pieces to grasp and examine the impacted component

thoroughly. It helps to detect and treat the illness early on more precisely and more effectively locating the contaminated component. Feature extraction, in image or Region of Interest (ROI), is the process of obtaining relevant information or associated characteristics without data loss. Used for decision making, the data gathered from medical images may be in any form—that of statistical characteristics, texture features, wavelet features, shape features, etc. Restructuring the photos helps one to comprehend the data in an understandable manner, hence guiding feature selection. It is also used to choose the finest elements to raise the classifier's performance [6]. Classification method is used to classify the unusual input from the conventional input photos. Conversely, skilled neuroradiologist manually interprets the obtained MRI images to develop the diagnosis. At last, the result of CAD system is matched with the view of neuroradiologist to detect the illness as early as feasible and therefore prevent misprediction of disease identification. By reducing the time needed to examine the intended defect, the CAD system improves the illness diagnostic process [7]. Because of its great precision, dependability, simplicity, and sophisticated characteristics of producing two and three-dimensional images which are utilized to accurately find the damaged sections, the CAD system is regarded as a necessary instrument in the identification of brain anomalies. Generally speaking, brain anomalies including trauma, intracranial pressure, vascular (blood vessels) problems, autoimmune diseases, infections, seizures and neurodegenerative disease are detected using the CAD system [8]. While repeated seizures are characterized by a particular kind of epilepsy, infections in the brain pertain to meningitis, encephalitis, and brain abscess. Furthermore, aggravating epilepsy include head traumas, brain infections, and stroke. Trauma is related to traumatic brain damage and intracerebral problems brought on by unconsciousness and bewilderment.

A. Brain Tumor Segmentation:

The first and most important stage in developing an AI-based diagnostic system for glioma characterisation and treatment management is neoplasm segmentation. Internally, gliomas are characterised by complex pathology, including edoema, necrosis, an enhancing core, and a non-enhancing tumor region. They grow infiltratively and sporadically. The goal of tumor segmentation is to isolate the tumors malignant components from their healthy counterparts in the surrounding area. The end result is the ROI or VOI defined in two or three dimensions. Glioma internal features (including enhancing and non-enhancing characteristics, necrosis, and edoema) shown in a 2D segmented image[9][10]. In order to identify and localise brain tumors, MRI imaging is often used by neuroradiologists and other experts in the field. Although it is technically feasible, manual segmentation is time-consuming and often produces inaccurate results. Research into the creation of trustworthy algorithms for automated segmentation has therefore been a prominent issue in recent years. Manual, semi-automated, and

fully-automated segmentation techniques are categorised according to the level of human input needed[11].

B. Manual segmentation:

In this method, a skilled neuroradiologist manually annotates the tumors ROI. Experts examine multimodal MRI scans slice by slice, using their knowledge and experience to detect tumor characteristics, and then meticulously designating subregions of tumor by hand. This task is very time-consuming, labour-intensive, and susceptible to differences in rating between or among raters. To evaluate the performance of fully-automatic or semi-automatic methods, however, experienced radiologists are still routinely using manual segmentation.

C. Semi-automatic segmentation:

However, there aren't many approaches that combine the advantages of semi-automatic and fully automated segmentation. With little human input, they could be able to repair a segmented tumor mask by hand. Decisions like selecting a ROI starting point, thresholds, pre-processing parameter tweaks, receiving feedback on automated process performance, and making required adjustments are examples of settings that may need human participation[12]. Traditional image segmentation methods, including image expanding or thresholding, are used in semi-automatic methods. While it's true that these methods can reduce time and effort when compared to manual processes, human error may still lead to inaccurate ROI definitions. Therefore, totally automated methods have been the centre of attention in most recent research on segmentation.

D. Automatic segmentation:

Artificial intelligence, machine learning, and deep learning algorithms are the backbone of these methods. Glioma MRIs show a broad variety of tumor variations in terms of size, shape, location, texture, and intensity. A further issue is that the boundaries of tumors are sometimes not well delineated. Multimodal MRI data becomes much more complicated with the 3D representation[13][14]. Strong, repeatable, fully autonomous segmentation requires the right handling of these issues. Generative and discriminative processes are the two main categories into which automatic methods fall. Discriminative approaches use supervised learning strategies to acquire the feature, which helps them understand the relationship between the input and the ground truth image. Instead of using data like normal brain anatomy maps, geographic locations, and tumor area, generative algorithms create probabilistic models.

E. Principal Component Analysis (PCA):

Mitigating the burden of needless characteristic calculations and memory storage is crucial for streamlining tumor categorization. It is possible to analyse new data with fewer computational resources by using principal component analysis (PCA), which successfully reduces the data's dimensionality. Principal component analysis (PCA) is an excellent tool for reducing the dimensionality of a dataset with several interrelated variables while preserving a large amount of variability. The data is reorganised into a set of variables, with their variances or significance levels used to determine their order of importance. This approach achieves three objectives simultaneously: first, it makes the input vector components orthogonal and uncorrelated. Second, it sorts the created orthogonal components by their contribution to the data's variability[15]

II. BRAIN TUMOR IDENTIFICATION

Imaging technology enables doctors to observe inside parts of the human body, therefore enabling correct analysis. Ultrasound and magnetic resonance imaging (MRI) replaced X-ray imaging in computed tomography (CT) scanners as radiologists find certain illnesses in the human body more valuable. Each of the many different cells that make up a body has a particular characteristic[16]. Growing within the human body, the cells will advance then break apart to create new cells. The human body cannot run as it should without these fresh cells. Normal cells inside the human body may die and new cells replace those old or damaged cells when they grow

old or seen as broken. Sometimes when new cells develop and of course break or the old cells do not die as they must, similarly if the body also does not require a new cell, the process goes wrong. Under such conditions, if those extra pointless cells proliferate unchecked and result in control loss, this provides more cells—often creating enormous amounts of tissue—increasing momentum and direction. This is known as a tumour. With integrated 55-99 billion neurons covering a vast region of the nerves, the brain may be regarded as the main support for the human central nervous system. These tumors are engulfing parts of the brain that influence human body sensitivity. Its proximity and strong spreading potential make the therapy very composite and dangerous[17].

One uses the term tumour in a sense as a proportional word for a neoplasm that seems to be enlarged. Because of unusual twists of events or cell development, neoplasm is an aberrant mass of tissue. Figure 1. (a) displays the route plan for fundamental internal structure flow chart of the brain tumour detection. Figure 1. (b) displays the brain tumour existence along with its identifying sample type. Either solid or fluid-filled, this tumour might be Still, tumour does not define the disease. Although ailment is by definition hazardous, a tumour might be generous (not dangerous), pre-harmful (precancerous), threatening (cancer-causing), or can speak to an injury with no carcinogenesis potential. Tumors may start in the brain or malignant development somewhere else in the body can go to the brain. Manifestations include disorientation, convulsions, loss of equalization, new or gradually solid migraines, and blurred vision. Sometimes side effects may not manifest themselves. Medical treatments, radiation, and chemotherapy are included into medications[18].

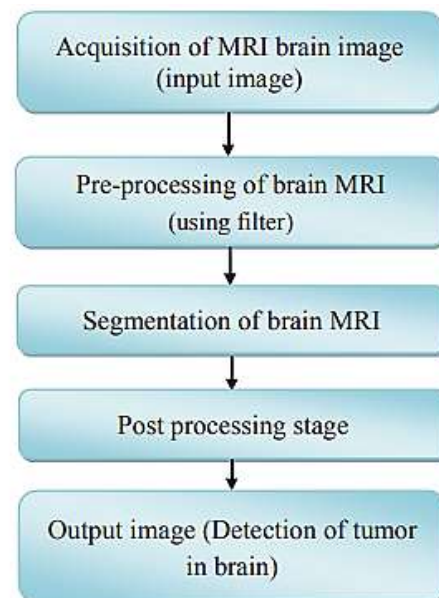


Fig. 1. (a) Basic structure flowchart of brain tumor identification

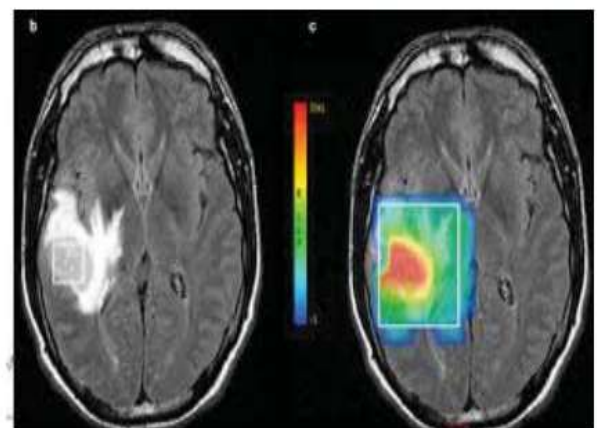


Fig. 1. (b) Presence and identification of brain tumor

III. PROPOSED METHODOLOGY

In our system, we make use of real-time data including MRI scans taken from many hospitals and internet archives. We standardize the picture widths to 224x224 pixels to guarantee uniformity. Every picture is carefully preprocessed upon capture to make it ready for study. Image accession in image processing is the process of obtaining a picture from a dataset for processing. Since without a picture no processing is impossible and the obtained image is completely naked, it is the initial phase of the workflow sequence. We then re-use the picture following the train route from the original device. Our work uses an MRI dataset consisting of over 2100 images, including normal and aberrant brain scans. We use sigma filtering to lower noise interference therefore improving the clarity of these photographs. This procedure includes pixel analysis within a defined region and variation smoothing out over a certain threshold. We want to increase picture quality and precision by using sigma filtering, therefore enabling more exact tumour identification[19].

Processing the photos comes first; then, segmentation follows. Here, segmentation using thresholding is in progress. Thresholding is fundamentally meant to ease the visual data processing. Thresholding bethinks a really common segmentation method used to distinguish the thing thought as a foreground from its surroundings. Here we are segmenting using binary thresholding. Every pixel in binary thresholding lives under the same threshold value. Should the pixel intensity value be less than the threshold, it sets to 0 (black); should not be so, it sets to 255 (white). Segmentation is the procedure of separating the tumour from normal tissues by cutting one picture into several pieces. Segmentation techniques satisfy the study of size, volume, position, texture, and

form of the picture by allowing one to identify or decide the aberrant section from the image[20][21].

Especially in the field of medical imaging analysis including brain MRI scans, feature extraction is a crucial first step in fully exploiting massive data sets. We may greatly reduce the inherent duplication in these datasets by choosing and aggregating variables into relevant features. This lowers not only the data processing pipeline but also improves the machine learning algorithm learning time. In our method, we extract characteristics from the obtained brain MRI images by use of morphological procedures. These processes enable us to emphasize different structures and trends in the images, therefore offering important new perspectives on the fundamental features of the brain tissue. A feature vector is the changed data—now expressed as a smaller collection of instructive characteristics. This vector captures the fundamental data required for further model construction and analysis[7]. Training the model comes second after photos have been transformed into N-D arrays. CNN and transfer learning approaches may be used to teach the model before training as this is a classification issue and the data consists of images. Five distinct ideas are used in the proposed system. Regarding precision and resilience, they turned out to be the finest[10][12]. Once the model is taught and examined, each fresh unseen picture may be utilized to forecast the kind of tumour contained within it using the same learnt model. The same preprocessing philosophy will be followed on fresh unknown images, then fed through these trained models, and lastly the prediction if the MRI shows a pituitary tumor, glioma tumor, meningioma tumor or none at all[14]. Only these four classes are now accessible; other classes may also be added using the current data to widen the scope of the project.

A. Dataset Classification

For Machine Learning-based Brain Tumor Detection, datasets are typically classified based on various attributes such as modality, image type, tumor type, and labeling technique. Following is a classification of datasets commonly used for the same

Brain tumor detection using machine learning relies on labeled datasets to train and evaluate predictive models. Labeled datasets provide ground truth annotations, enabling supervised learning techniques to classify and segment brain tumors accurately. These datasets typically contain medical imaging data such as MRI, CT scans, or histopathology slides. The classification of datasets is based on factors such as modality, tumor type, number of images, and labeling methods. Publicly available information like BraTS, TCGA, and OASIS are utilized in brain tumor detection research and development. Private hospital information are sometimes used for real clinical application but may be access-restricted. Datasets might be binary classification (tumor or non-tumor) or multi-class classification (different types of tumors). Some datasets also support segmentation tasks, where the tumor region is precisely outlined inside medical images. MRI-based data, such as BraTS, provide various imaging modalities (T1, T1c, T2, FLAIR) for full analysis. Histopathology data consist of microscopic images that help in tumor type classification based on cells. Image size in a dataset affects model generalization; larger datasets provide superior performance to deep learning models. Annotation quality matters as incorrect labels can produce inferior model predictions and unstable diagnostic tools. Certain data sets contain additional metadata, i.e., clinical background and patient data, to achieve maximum predictability. Preprocessing procedures like normalization, data augmentation, and noise filtering are employed for cleaning data sets. Dataset imbalance, where a class dominates others, may affect model performance and is handled through resampling techniques. Cross-validation techniques are applied for measuring the performance of a model and for preventing overfitting. Hybrid data sets combining MRI, CT, and histopathology information have a more robust approach to tumor classification. Consistent machine learning model benchmarking allows for comparison of studies.

Open-access data sets encourage researcher collaboration and the advancement of medical AI. Continual developments of high-quality labeled data sets are essential to driving brain tumor detection and treatment planning

IV. CONCLUSION

The application of labeled datasets for brain tumor detection with machine learning has greatly improved diagnostic accuracy and efficiency. Machine learning algorithms, especially deep learning methods like CNNs, have shown excellent performance in tumor classification and segmentation based on labeled medical imaging data. Publicly accessible datasets such as BraTS, TCGA, and OASIS have been instrumental in advancing research through the availability of high-quality annotated images. Even with these developments, issues of dataset imbalance, the quality of annotation, and model generalization remain at the forefront. The combination of hybrid machine learning methods and multimodal datasets (MRI, CT, and histopathology) presents potential solutions for enhancing prediction accuracy. Also, creating more varied and highly labeled datasets can further improve the robustness of AI-based diagnostic models. Future research must concentrate on enhanced data labeling methods, including real-world clinical datasets, and investigate federated learning methods to preserve privacy while sharing global data resources. The ongoing improvement in labeled datasets coupled with sophisticated machine learning models promises a significant prospect for the detection of early and precise brain tumors, finally benefiting patient prognosis and treatment options.

V. REFERENCES

- [1]. W. El Elhola and K. Bozed, "Performance Analysis of Brain Tumor Detection Based on Gradient Boosting Machine -CNN model," 2022 IEEE 2nd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA), Sabratha, Libya, 2022, pp. 544-547, doi: 10.1109/MI-STA54861.2022.9837738.
- [2]. P. Saravanan, S. Saravanakumar and G. Shanmugarathinam, "A Survey on Brain Tumor Prediction with

Various Machine Learning Approaches," 2023 International Conference on Computer Science and Emerging Technologies (CSET), Bangalore, India, 2023, pp. 1-4, doi: 10.1109/CSET58993.2023.10346944.

[3]. Irmak, E 2021, „Multi-classification of brain tumor MRI images using deep convolutional neural network with fully optimized framework“, Iran J Sci Technol Trans Electr Eng, vol. 45, no. 1, pp. 1015-1036.

[4]. Guoli Song, Tian Shan, Min Baod, Yunhui Liud, Yiwen Zhao & Baoshi Chen 2021, Automatic brain tumour diagnostic method based on a back propagation neural network and an extended set-membership filter“, Computer Methods and Programs in Biomedicine, vol. 208, no. 78, P. 106188

[5]. Balakumaresan Ragupathy & Manivannan Karunakaran 2020, A fuzzy logic-based meningioma tumor detection in magnetic resonance brain images using CANFIS and U-net CNN classification“, International Journal of Imaging Systems and Technology, vol. 31, no. 1, pp.123-129.

[6]. Jasmine Hephzipaha, J & Thirumurugan, P 2020, „Performance analysis of meningioma brain tumor detection system using feature learning optimization and ANFIS classification method“, IETE Journal of Research, vol. 1, no. 3, pp. 1-12

[7]. Selvapandian, AK & Manivannan 2018, „Performance analysis of meningioma

[8]. brain tumor classifications based on gradient boosting classifier“, International Journal of Imaging Systems and Technology, vol. 28, no. 4, pp. 1-15

[9]. Angulakshmi, M & Lakshmi Priya, GG 2018, „Brain tumour segmentation from MRI using superpixels based spectral clustering“, Journal of King Saud University - Computer and Information Sciences, vol. 32, no. 10, pp.1-19.

[10]. Soltaninejad, M, Yang, G, Lambrou, T, Allinson, N, Timothy L Jones, Thomas R Barrick, Franklyn A Howec & Xujiong Ye 2018, „Supervised learning based multimodal MRI Brain Tumour segmentation using texture features from supervoxels“, Computer Methods and Programs in Biomedicine, vol. 157, no. 1, pp. 69-84

[11]. Sauwen, N, Marjan Acou, Diana M Sima & Jelle Veraart 2017, „Multi orientation analysis of non-regression of varying the linguistic algorithms with various genetic pattern analysis“, BMC Medical Imaging, Springer, vol. 17, no. 1, pp. 121-129

[12]. Alfonse, M&Salem, A-BM 2016, An automatic classification of brain tumors through MRI using support vector machine“, Egyptian Computer Sci J, vol. 40, pp. 11-21

[13]. Athira, VS 2015, „Brain tumor detection and segmentation in MR images using GLCM an adaboost classifier“, IJSRSET, vol. 1, no. 3, pp.1-7.

[14]. Nabizadeh, N, John, N & Wright Clinton 2014, „Histogram-based gravitational optimization algorithm on single MR modality for automatic brain lesion detection and segmentation“, Expert Systems with Applications, vol. 41, no. 5, pp. 7820-7836.

[15]. P. Shanthakumar and P. Ganeshkumar, „Computer aided brain tumor detection system using watershed segmentation techniques“, International Journal of Imaging Systems and Technology, Vol. 25(4): pp. 297- 301, 2015.

[16]. E. Dandil et al., „Computer-Aided Diagnosis of Malign and Benign Brain Tumors on MR Images“, Advances in Intelligent Systems and Computing, vol 311, Springer, 2015.

[17]. S. R. Telrandhe, A. Pimpalkar and A. Kendhe, „Detection of Brain Tumor from MRI Images by Using Segmentation & SVM“, World Conference on Futuristic Trends in Research and Innovation for Social Welfare, pp. 1-6, 2016.

[18]. A. Goel and V. P. Vishwakarma, „Feature Extraction Technique Using Hybridization of DWT and DCT for Gender Classification“, Ninth International Conference on Contemporary Computing (IC3), pp. 1-6, 2016.

[19]. D. Somwanshi, A. Kumar, P. Sharma and D. Joshi, „An Efficient Brain Tumor Detection from MRI Images using Entropy Measures“, International Conference on Recent Advances and Innovations in Engineering (ICRAIE), pp. 1-5, 2016.

[20]. S. Pereira, A. Oliveira, V. Alves and C. A. Silva, „On Hierarchical Brain Tumor Segmentation in MRI using Fully Convolutional Neural Networks: A preliminary study“, IEEE 5th Portuguese Meeting on Bioengineering (ENBENG), pp. 1-4, 2017.

[21]. V. Shreyas and V. Pankajakshan, „A deep learning architecture for brain tumor segmentation in MRI images“, IEEE 19th International Workshop on Multimedia Signal Processing (MMSP), pp. 1-6, 2017.

[22]. S. K. Shil, F. P. Polly, M. A. Hossain, M. S. Iftekhar, M. N. Uddin and Y. M. Jang, „An Improved Brain Tumor Detection and Classification Mechanism“, International Conference on Information and Communication Technology Convergence (ICTC), pp. 54-57, 2017.