Informatics in Medicine Unlocked 43 (2023) 101414

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: www.elsevier.com/locate/imu

Optimizing brain tumor classification through feature selection and hyperparameter tuning in machine learning models

Mst Sazia Tahosin [a], Md Alif Sheakh [a], Taminul Islam [a,b,*], Rishalatun Jannat Lima [a], Mahbuba Begum [c]

[a] Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh
[b] Department of Computer Science, Southern Illinois University Carbondale, IL, USA
[c] Department of Computer Science and Engineering, Mawlana Bhashani Science and Technology University, Tangail, Bangladesh

ARTICLE INFO

Keywords:
Brain tumor classification
Brain cancer
Brain tumor
Machine learning
Hyperparameter tuning
Feature extraction
Feature selection
Homomorphic filtering

ABSTRACT

Accurately classifying brain tumors using images is extremely important for prognosis and treatment planning. In this study, we have developed an optimized approach using machine learning techniques to classify brain tumors. Our method involves preprocessing the images, extracting features, selecting the most significant ones, and tuning the model parameters. We utilized filtering, morphological opening, and normalization techniques to enhance image quality and reduce noise. We have extracted 17 features that capture the characteristics of the tumors and identify the seven most distinguishing features through importance analysis. By employing a range of models such as Random Forest, Support Vector Machines, Extreme Gradient Boosting, K Nearest Neighbors, Categorical Boosting, Extra Trees, and Naive Bayes, we achieve an accuracy of 98.0 % after thorough hyperparameter optimization. This research highlights the impact of the feature selection process, along with model tuning, on maximizing classification performance. This approach provides a framework that enables the diagnosis of brain tumors for enhanced clinical decision-making and patient care.

1. Introduction

The brain is the body's most important organ. The structure manages and coordinates everyday neurological operations as a central hub [1]. The nervous system analyses sensory data from the body to make decisions [2]. The cerebrum, brainstem, and cerebellum comprise the structure. Grey, white, and cerebrospinal fluid define the brain [3]. Brain abnormalities can have profound implications. Congenital errors, traumatic brain traumas, and unregulated central brain cell proliferation can disrupt brain function [4]. Untreated brain abnormalities can cause several physiological problems [5]. Cellular proliferation without regulation can induce brain abnormalities and brain cancer, a global epidemic.

The most common primary brain tumor is meningioma, which accounts for more than 30 % of brain tumors [6]. Gliomas are the most common type of brain tumor and can be low-grade or high-grade [7,8]. Low-grade gliomas grow slowly and are often curable, while high-grade gliomas are more aggressive and can be fatal [9]. Other types of brain tumors are rarely described, but they can be serious. Fig. 1 shows some common types of primary brain tumors. For example, medulloblastoma is a type of brain tumor that often occurs in children. It is a very aggressive tumor that can spread to other parts of the body [10]. So, classifying brain tumors are of utmost importance in facilitating timely diagnosis, formulating effective treatment strategies, and predicting patient outcomes. The exponential progress in medical imaging technology has bestowed upon clinicians an extensive corpus of brain tumor data. Nevertheless, the manual analysis of such intricate data is a laborious process susceptible to errors and subjectivity. To tackle these challenges, there has been considerable interest in utilizing machine learning (ML) models [11] to automate the classification of brain tumors [12].

The extraction and selection of informative features from medical images are critical in developing accurate machine learning models for brain tumor classification. The feature extraction process entails converting unprocessed image data into a comprehensive collection of features that effectively encapsulate pertinent details regarding the tumor's attributes. In contrast, feature selection endeavors to ascertain the subset of features with the highest discriminatory and pertinent

* Corresponding author. Southern Illinois University Carbondale, Carbondale, Illinois, USA
E-mail address: taminul.islam@siu.edu (T. Islam).

https://doi.org/10.1016/j.imu.2023.101414
Received 15 October 2023; Received in revised form 17 November 2023; Accepted 17 November 2023
Available online 24 November 2023

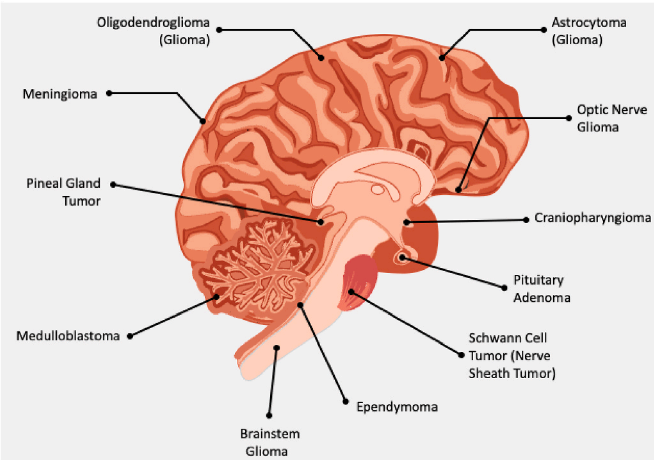

Fig. 1. Some common primary brain tumors.

qualities, thereby contributing substantially to the classification objective.

This paper presents a comparative analysis of feature extraction and feature selection techniques for brain tumor classification using ML models. The main objectives of this study are twofold: Firstly, we aim to evaluate the performance of different feature extraction methods in capturing the distinctive patterns and characteristics of brain tumors. Various techniques, such as wavelet transforms, texture analysis, and deep learning-based approaches, will be explored in this context. Secondly, we investigate the impact of feature selection on brain tumor classification accuracy. To conduct our analysis, we will employ rigorous evaluation metrics such as accuracy, recall, precision, f1 score, and confusion matrix to assess the performance of the ML models [13]. The results of this comparative analysis will provide valuable insights into the effectiveness of various feature extraction and selection techniques for brain tumor classification. Furthermore, it will guide clinicians and researchers in selecting the most appropriate approach when developing ML models for accurate brain tumor diagnosis and classification.

Here are the contributions of our research on this paper:

- Proposed a comprehensive methodology for brain tumor classification combining image preprocessing, feature extraction, feature selection, and machine learning models.
- Utilized Homomorphic Filtering, Morphological Opening, and Normalization in the preprocessing phase to enhance image quality and remove noise.
- Extracted 17 informative features from preprocessed images, capturing essential characteristics relevant to brain tumor classification.
- Utilized feature importance to perform feature selection and reduced dimensionality, selecting the top 7 most discriminative features.
- Conducted hyperparameter tuning to optimize model performance, significantly improving accuracy, precision, recall, and F1 score.
- Demonstrated the importance of feature extraction and selection techniques in improving the efficiency and effectiveness of machine learning models for medical imaging tasks.

2. Literature review

In this section, this research has studied and reviewed some papers related to this topic.

Baby Pattanaik et al. [14] proposed a feature engineering approach for classifying common brain tumors in MRI scans. They used five machine learning classifiers and extracted handcrafted features from the MRI, enhancing the feature vector's dimension. The Fine KNN classifier performed the best, achieving 91.1 % accuracy and 0.96 AUC. The study's outcome is a potential method, Fine KNN, which could be integrated into low-end devices to improve brain tumor recognition. However, the specific dataset used in the study was not mentioned.

Jaeyong Kang et al. [15] introduced a brain tumor classification method that utilized an ensemble of deep features and machine learning classifiers. They employed transfer learning and pre-trained deep convolutional neural networks to extract deep features from brain MR images. The top-performing features were combined as an ensemble and fed into various classifiers. The study evaluated their approach on three brain MRI datasets and found that the ensemble of deep features significantly improved performance. Support Vector Machine (SVM) with radial basis function (RBF) kernel excelled, especially for large datasets.

Vidyarthi et al. [16] proposed a machine learning methodology for classifying malignant brain tumors multiclass. They utilized real-life datasets with five classes and a vast feature set from six domains to capture essential information from the region of interest. The Cumulative Variance method (CVM) was introduced for feature selection. The proposed approach achieved mean average classification accuracies of 88.43 % (KNN), 92.5 % (mSVM), and 95.86 % (NN) for multi-class prediction. The results indicated that the NN classifier performed best with 95.86 % accuracy using diversified features. The dataset consisted of 660 MR images from 1160 regions marked as malignant and normal by radiologists, which were classified into six classes using the machine learning approach.

Ali et al. [17] introduced an attention-based convolutional neural network for brain tumor segmentation. They used a pre-trained VGG19 network as the encoder and incorporated an attention gate for segmentation noise induction and a denoising mechanism to prevent overfitting. The BRATS'20 dataset, comprising four MRI modalities and one target mask file. The proposed algorithm achieved impressive dice similarity coefficients of 0.83, 0.86, and 0.90 for enhancing core and whole tumors, respectively. The paper showcased qualitative and quantitative results, including sensitivity, specificity, accuracy, and precision, outperforming state-of-the-art methods.

Asiri et al. [18] proposed six machine learning algorithms for brain tumor classification in MRI images, including Random Forest, Naïve Bayes, Neural Networks, CN2 Rule Induction, Support Vector Machine, and Decision Tree. They used a dataset of 253 images, with 98 healthy and 155 tumor-affected images, extracting 2048 features. SVM outperformed other algorithms with 95.3 % accuracy, contributing to an improved brain tumor classification model.

Priyanka Modiya et al. [19] classified brain tumor MRI images using the Grey-level co-occurrence matrix (GLCM) feature extractor and random forest classifier. This study included 245 brain MRI scans, 154 of which had tumors and 91 without. The random forest classifier assessed the accuracy, true positive rate, true negative rate, and confusion matrix-derived false positive and false negative rates of the GLCM features. GLCM feature extraction with appropriate parameters captures brain MRI image textural characteristics with promising accuracy and other performance measures. Chi-square and t-tests validated their methods. This work proposes using GLCM characteristics and a random forest classifier to classify brain tumors.

Senan et al. [20] developed a computer-aided diagnostic system for early brain tumor detection using a mix of deep and machine learning methods. They utilized AlexNet and ResNet-18 with SVM for classification, using an MRI dataset of 3,060 images divided into four classes: three tumors and one normal. The system achieved impressive results, with the AlexNet + SVM hybrid approach performing the best, achieving 95.10 % accuracy, 95.25 % sensitivity, and 98.50 % specificity. This system has the potential to aid doctors in making accurate diagnoses and improving brain tumor patients' survival rates.

Nanmaran et al. [21] investigated image fusion's role in brain tumor classification models using machine learning for personalized medicine. They employed frequency domain techniques to improve edge quality

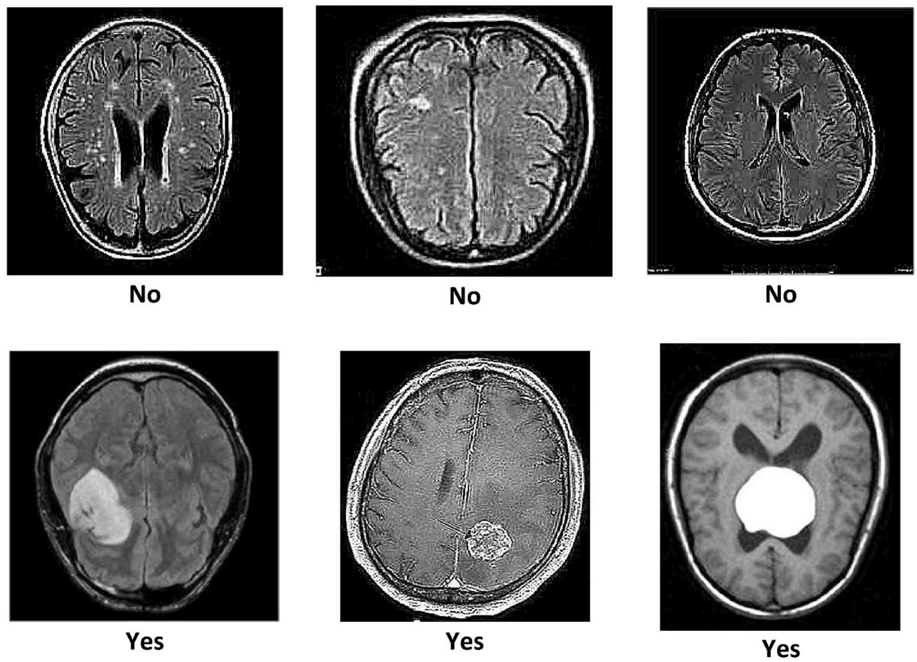

Fig. 2. Some images from dataset.

and used a contrast-limited adaptive histogram equalization technique for image preprocessing. The proposed method combined MRI and SPECT images through discrete cosine transform-based fusion. AI algorithms like SVM, KNN, and decision tree classifiers were tested using fused and individual input images. SVM achieved the highest accuracy of 96.8 %, outperforming KNN and decision tree classifiers. The proposed method's satisfactory results were compared with existing techniques.

Haq et al. [22] proposed a hybrid approach using deep CNN and machine learning classifiers for precise brain MRI tumor segmentation and classification. The model comprises three stages: CNN for learning tumor features, faster region-based CNN for tumor localization, and a deep CNN and machine learning for segmentation and classification. The model achieved high accuracy (98.3 %) and dice similarity coefficient (97.8 %) on brain dataset 1 and similarly on the Figshare dataset [22], with 98.0 % accuracy and 97.1 % DSC. The results show the proposed model surpasses state-of-the-art techniques significantly.

Rinesh et al. [23] researched brain tumor classification using hybrid machine learning algorithms. They utilized hyperspectral imaging to analyze tumor localization in the brain, employing k-nearest neighbor and k-means clustering with the firefly algorithm for tumor detection. Multilayer feedforward neural networks were used for brain area labeling. The proposed technique outperformed existing methods, showing a higher peak signal-to-noise ratio and lower mean absolute error. The model achieved improved accuracy (96.47 %), sensitivity (96.32 %), and specificity (98.24 %). The dataset had 128 bands with wavelengths from 400 to 1300 nm. The successful hybrid machine learning algorithm for brain tumor classification offering superior results.

Uvaneshwari and Baskar [24] proposed a computer-aided diagnosis model for brain tumor detection and classification using machine learning. Their BTDC-MOML algorithm involves image pre-processing, optimal Kapur's thresholding-based segmentation, LDEP feature extraction, and XG-Boost classification. The method outperformed existing models, achieving an average accuracy of 97.83 %, precision of 95.71 %, recall of 95.63 %, F-score of 95.67 %, and MCC of 94.22 %.

Vijithananda et al. [25] used machine learning to distinguish malignant from benign brain tumors. One hundred ninety-five patients had 995 malignant and 604 benign labeled MRI Apparent Diffusion Coefficient (ADC) images. ANOVA f-test selected demographic and textural features from photos. After hyperparameter adjustment, the Random Forest classifier produced the final model with 90.41 % accuracy. The

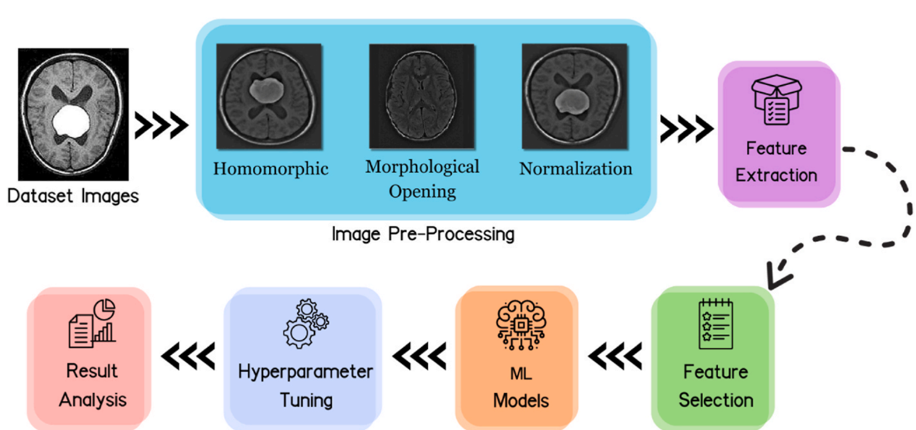

Fig. 3. Our proposed methodology.

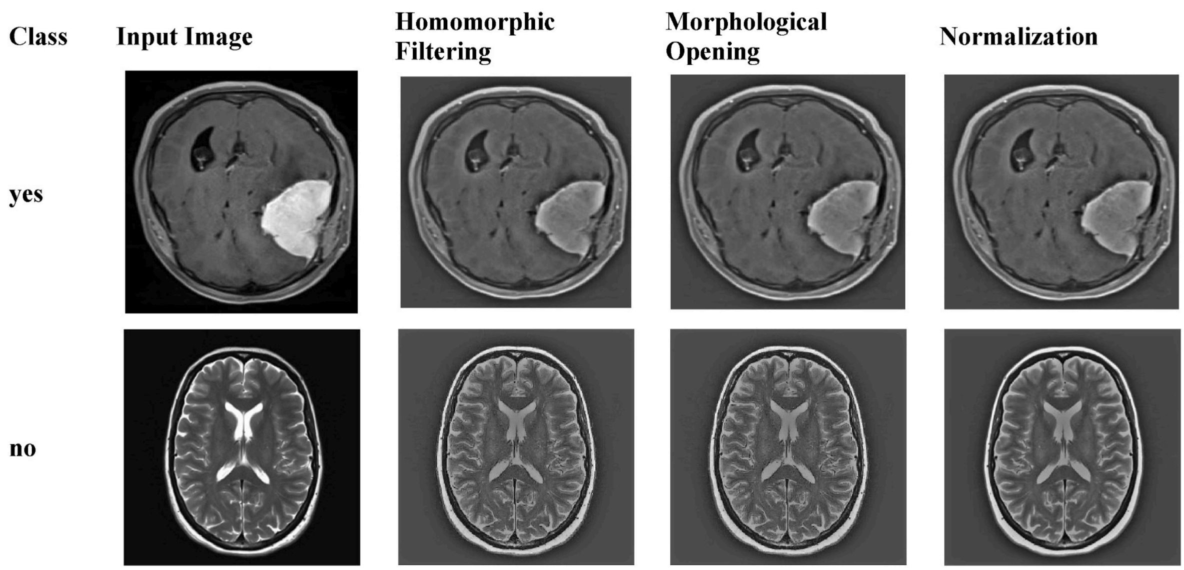

Fig. 4. Steps of image pre-processing.

Table 1
PSNR values of image enhancement techniques on ten random brain tumor images.

Image No.	Gaussian Filtering	Median Filter	Homomorphic Filtering	Morphological Opening	Morphological Closing	Normalization
Image 1	26.5	24.1	**31.2**	**30.8**	26.3	**33.4**
Image 2	29.3	23.7	**32.5**	**30.2**	25.1	**34.2**
Image 3	27.1	22.3	**33.8**	**31.5**	27.9	**35.7**
Image 4	24.7	26.5	**30.4**	**30.8**	23.6	**31.3**
Image 5	25.2	25.4	**30.1**	**31.9**	24.8	**32.6**
Image 6	26.8	21.1	**34.2**	**32.4**	26.5	**34.5**
Image 7	27.9	23.5	**35.5**	**33.2**	28.3	**33.1**
Image 8	23.8	24.3	**30.6**	**30.5**	22.1	**32.8**
Image 9	26.4	23.1	**31.3**	**32.7**	23.4	**30.2**
Image 10	27.2	21.8	**32.1**	**31.6**	24.9	**34.5**

study found that the selected variables (excluding skewness and GLCM homogeneity) are informative for identifying and discriminating brain tumors, and the ML model can help diagnose them.

3. Dataset description

The Brain Tumor MRI Image Dataset that we collect from Kaggle [26] consists of a total of 3000 MRI images, categorized into two classes, “Yes” and “No.” The dataset is visualized in Fig. 2. These images are used for brain tumor detection. The dataset is evenly balanced, with each class containing 1500 images.

4. Proposed methodology

In this section, our proposed methodology combines image pre-processing, feature extraction, feature selection, and the utilization of machine learning models to achieve robust and reliable results. The preprocessing phase involves Homomorphic Filtering, Morphological Opening, and Normalization to prepare the images for analysis. After that, we performed feature extraction from the preprocessed images and found the crucial features by feature selection. After getting the top features, we applied seven machine learning algorithms to evaluate the performance, and by doing hyperparameter tuning on the machine learning model, we got the best result. Also, we do various analyses of results to choose the best machine learning algorithm for our work. In Fig. 3, our proposed methodology workflow has been visualized.

4.1. Image pre-processing

To enhance the quality and robustness of the input images, we employ the following image preprocessing techniques. It has also been visualized in Fig. 4. As shown in Table 1, we experimentally compared various techniques on sample brain tumor images. We evaluated the image quality through Peak signal-to-noise ratio (PSNR) values to determine the optimal preprocessing approaches of Homomorphic Filtering, Morphological Opening, and Normalization.

4.1.1. Homomorphic filtering

As in our dataset, we have brain cancer MRI (Magnetic Resonance Imaging) images; it may suffer from uneven illumination due to variations in the scanning conditions or tissue properties. Homomorphic filtering is particularly well-suited for addressing illumination variations, separating the image into low-frequency and high-frequency components. By attenuating the low-frequency illumination component, homomorphic filtering enhances the visibility of subtle features and details in the brain images, making it easier for subsequent analysis and detection algorithms to identify relevant patterns indicative of brain cancer [27].

4.1.2. Morphological opening

In brain cancer images, images may be susceptible to noise and artifacts, especially during the imaging acquisition process. Morphological opening is a powerful image processing technique that helps remove noise while preserving important structures in the image. So, using morphological opening, we can effectively reduce noise-induced false positives, thus enhancing the reliability of our brain cancer detection

Table 2
Short description of all features.

Features Name	Description
Area	The total area of the object.
Pa_ratio	The perimeter to area ratio provides information about the object's shape complexity.
Solidity	The ratio of the object area to its convex hull's area indicates how compact it is.
Circularity	A measure of how closely the object resembles a circle.
Equivdiameter	The diameter of a circle provides size-related information.
Convexarea	The area of the convex hull of the object.
Extent	The ratio of the object's area to the bounding box's area provides information about its spatial distribution.
Filled area	After filling holes, the area of the object can be relevant for particular objects.
Major_axis_length	The length of the major axis of the best-fitting ellipse to the object, provides shape-related information.
Mean	The mean intensity value of the object's pixels which may be indicative of its appearance.
Std_deviation	The standard deviation of the pixel intensities provides information about texture variation.
Shannon_entropy	A measure of the object's information content relevant for texture analysis.
Skewness	A measure of the object's asymmetry in pixel intensity distribution.
LBP_energy	A texture descriptor representing the sum of squared Local Binary Pattern (LBP) values.
LBP_entropy	A measure of the texture complexity based on LBP values.
Gabor_energy	A texture feature extracted using Gabor filters indicates specific textures' presence.
Gabor_entropy	A measure of the texture complexity based on Gabor filter responses.

system [28].

4.1.3. Normalization

Normalization plays a significant role in preparing brain cancer images for subsequent feature extraction and classification tasks. Brain cancer images obtained from different imaging devices or protocols may have varying pixel intensity levels, leading to inconsistencies and biases in the extracted features. Normalizing the pixel values to a standardized range ensures that these intensities variations do not influence the subsequent feature extraction process [29].

The PSNR values achieved by different image enhancement techniques on ten randomly selected brain tumor images have been shown in Table 1. Higher PSNR values above 30 dB indicate good image quality, while lower values suggest quality degradation [30,31]. As seen, Homomorphic Filtering, Morphological Opening, and Normalization consistently attain higher PSNR across the images, demonstrating their ability to improve quality. The table provides quantitative evidence through the PSNR metric that these techniques enhance image robustness and reduce noise for more accurate tumor analysis.

4.2. Feature extraction

From each preprocessed image, a set of 17 features is extracted. These features are selected to capture the essential characteristics of the images [32]. In Table 2 the extracted features details have been given:

4.3. Feature selection

To reduce the feature dimensionality and focus on the most discriminative features, we utilize feature importance [33] from a suitable feature selection technique; that is we rank the features based on their importance scores and select the top k features for the subsequent machine learning models.

In our case, we select the seven most relevant features based on their importance scores, which are as follows:

- Circularity
- Mean
- Std_deviation
- Shannon_entropy
- Skewness
- LBP_energy
- LBP_entropy

By using this, we focus on the most influential features while reducing the feature dimensionality, which can improve the efficiency and effectiveness of the subsequent machine learning models in accurately classifying brain cancer based on the selected features.

4.4. Machine learning models

Based on a thorough literature study [34–36], we chose the seven machine learning models, which indicated their extensive use and success in brain tumor image classification tasks, particularly in medical imaging. Previous study has shown that these models are frequently used without significant parameter modifications, highlighting their strong performance and versatility across brain tumor MRI datasets. The similarity in their use indicates those selected models provide accurate and effective solutions for brain tumor classification, confirming the overall findings of prior studies in the field.

- Random Forest (RF): RF is a descriptive clustering technique that works well with high-dimensional image data. By combining multiple decision trees, RF can capture complex visual tumor images [37].
- K-Nearest Neighbors (KNN): KNN was chosen because of its ability to classify new data points based on the visual similarity with the training samples. This non-parametric approach is suitable for studying local imaging patterns associated with different tumor types [38].
- Support Vector Machine (SVM): SVM was introduced because it efficiently handles high-quality data and results. Its high margin hyperplane adaptation makes it suitable for detecting complex linear relationships in tumor images [39].
- XGBoost: XGBoost was chosen for its outstanding performance in domains. It can learn complex visual features useful for the discrimination of images. Embedded systems help eliminate redundancy [40].
- CatBoost: CatBoost was chosen for its expertise in handling categorical input features. It can better represent hierarchical pixel values in an image to improve performance [41].
- Extra Trees: Extra trees were added to reduce gaps and eliminate unnecessary trees. If you have a decision tree with moderate noise, this can reduce the redundancy of image data [42].
- Naive Bayes: Naive Bayes provides a simple initial model. Despite being a simple concept, it often works surprisingly well in practice and is quick to train [43].

These models were selected according to their particular strengths and weaknesses with a focus on their demonstrated success in various areas of their particular strengths and weaknesses with a focus on their demonstrated success in various image classification areas. This extensive methodology ensures that the models comprehensively assess brain tumor classification. The models were chosen for their ability to handle high-dimensional image data, local imaging patterns, complicated visual features, categorical input features, noise reduction, and rapid training. This careful examination represents a planned combination of strategies designed to handle the complex issues of brain tumor classification.

4.5. Hyperparameter tuning on machine learning models

To optimize the performance of the machine learning models, we conduct hyperparameter tuning using grid search techniques [44]. The

Table 3
Performance of machine learning models.

Model	Accuracy	Precision	Recall	F1 Score
RF	0.92	0.90	0.91	0.89
KNN	0.85	0.88	0.84	0.86
XGBoost	0.92	0.93	0.90	0.91
Extra Tress	**0.93**	**0.92**	**0.93**	**0.92**
CatBoost	0.85	0.87	0.88	0.85
Naïve Bayes	0.69	0.71	0.69	0.70
SVM	0.67	0.68	0.67	0.69

hyperparameter values that yield the best results on the validation set are selected for each model. Table 3 presents the results obtained from our machine-learning models after conducting hyperparameter tuning. The table showcases the performance metrics of each model, including accuracy, precision, recall, and F1-score, based on the selected 7 most relevant features obtained through the feature selection process.

5. Result analysis

Result analysis is the most crucial step in evaluating the effectiveness and performance of our proposed methodology for brain cancer detection. In this section, we present a comprehensive analysis of the results obtained from the machine learning models, focusing on accuracy, precision, recall, and F1-score, along with their significance in brain cancer detection.

$$\text{Accuracy} = \frac{\text{True Positive (TP)} + \text{True Negative (TN)}}{\text{True Positive} + \text{False Positive (FP)} + \text{True Negative} + \text{False Negative (FN)}} \tag{1}$$

Accuracy measures the overall correctness of our classification model. It is calculated as the ratio of correctly classified samples (both true positives and true negatives) to the total number of samples. Here is the formula [45]:

Precision represents the proportion of true positive predictions among all positive predictions made by the model and is computed using the formula [46]:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \tag{2}$$

Recall assesses the model's ability to correctly identify positive cases from the total number of true positive cases and is calculated as [47]:

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{3}$$

The F1 score balances precision and recall by taking their harmonic mean, providing a single metric to assess the model's performance [48]:

$$\text{F1 score} = 2 * \left(\frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}\right) \tag{4}$$

5.1. ML models performance analysis

In this section, we observe varying degrees of accuracy, precision, recall, and F1 score, which is present in Table 3. Among the models, Extra Trees demonstrated the highest accuracy of 0.93, along with excellent Precision (0.92), Recall (0.93), and F1 Score (0.92). This indicates that the Extra Trees model correctly classifies brain cancer cases while minimizing false positives and negatives. After that, Random Forest (RF) and XGBoost got the second-highest accuracy of 92 %. K-Nearest Neighbors (KNN) and CatBoost achieve moderate accuracy at 85 % and demonstrate good precision and recall. However, Naïve Bayes and Support Vector Machine (SVM) show relatively lower accuracy at 69 % and 67 %, respectively. While SVM exhibits decent precision, all models present comparable recall values. Based on these results, we can

Table 4
Machine learning models performance using hyperparameter tuning.

Model	Hyperparameter Tuning	Range	Best	Accuracy	Precision	Recall	F1 Score
RF	n_estimators	100, 300, 500	100	0.92	0.9	0.91	0.89
	max_depth	None, 5, 10, 20	None				
	min_samples_split	2, 5, 10	2				
	min_samples_leaf	1, 2, 2004	1				
KNN	n_neighbors	3, 5, 7, 9	3	0.92	0.88	0.84	0.86
	p	Manhattan distance (p = 1) Euclidean distance (p = 2)	1				
	weights	uniform, distance	distance				
XGBoost	learning_rate	0.01, 0.1, 0.3	0.3	0.92	0.93	0.9	0.91
	max_depth	3, 5, 7	7				
	n_estimators	100, 200, 300	200				
	gamma	0, 0.1, 0.2	0				
Extra Tress	n_estimators	100, 200, 300	300	**0.98**	**0.97**	**0.98**	**0.96**
	max_depth	None, 5, 10, 20	None				
	min_samples_split	2, 5, 10	2				
	min_samples_leaf	1, 2, 4	1				
CatBoost	iterations	100, 200, 300	300	0.92	0.87	0.88	0.85
	learning_rate	0.01, 0.1, 0.3	0.3				
	depth	3, 5, 7	7				
	l2_leaf_reg	1, 3, 5	3				
Naïve Bayes	No			0.69	0.71	0.69	0.7
SVM	Regularization parameter (C)	0.1, 1, 10	10	0.67	0.68	0.67	0.69
	kernel	linear, radial basis function (rbf)	rbf				
	gamma	scale, auto	scale				
	degree	2, 3	2				

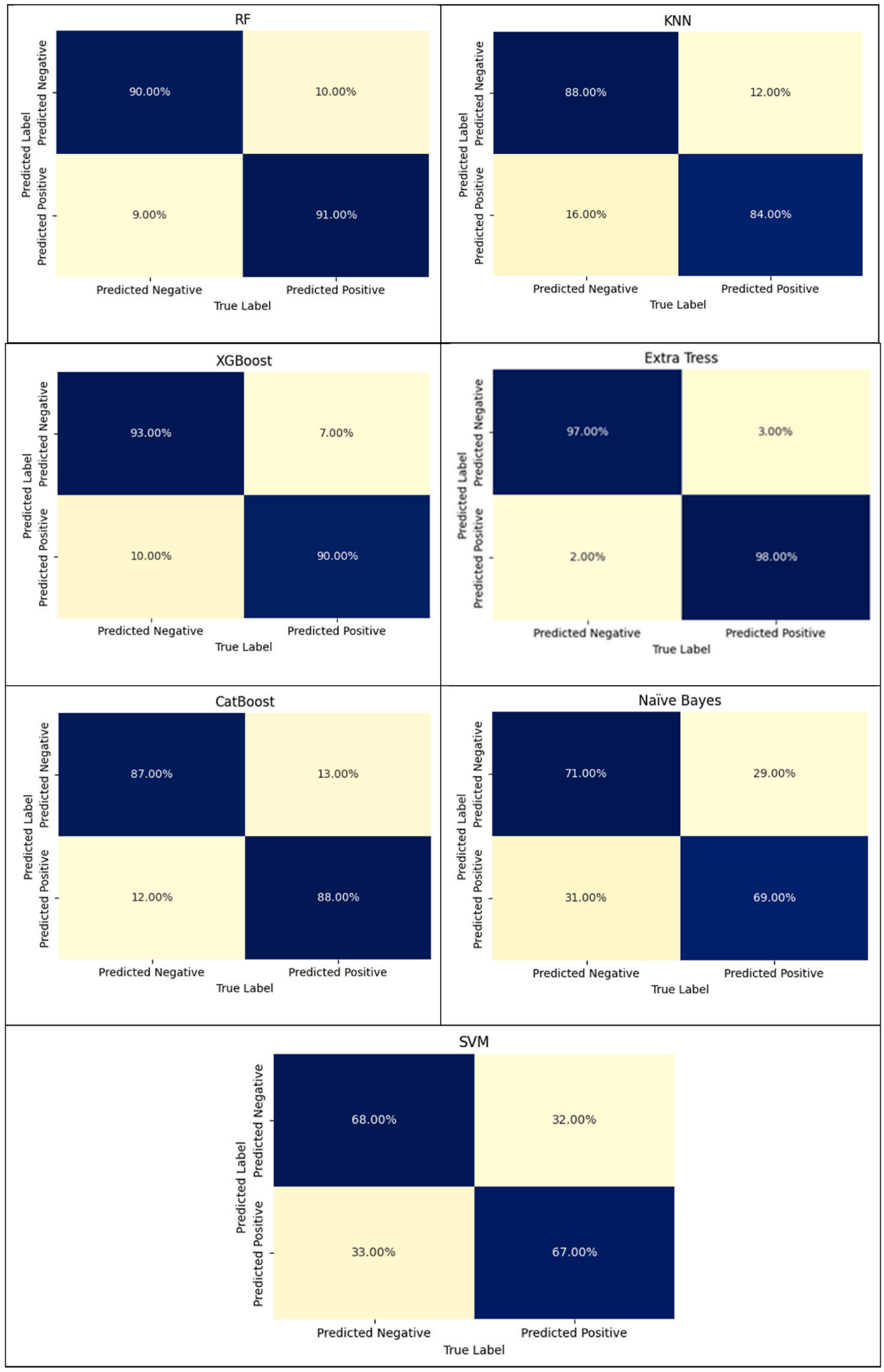

Fig. 5. Confusion matrix of all machine learning model.

conclude that Extra Trees stands out as the best-performing model for accurately detecting brain cancer from the selected features obtained through the feature selection process.

5.2. Performance analysis after hyperparameter tuning

After hyperparameter tuning in Table 4, we evaluated the models' performance using the metrics of Accuracy, Precision, Recall, and F1 Score. Among the models, again, Extra Trees demonstrated outstanding performance with an accuracy of 0.98, along with high precision (0.97), recall (0.98), and F1 Score (0.96). This indicates that the Extra Trees model, after hyperparameter tuning, performs better in correctly classifying brain cancer cases. Similarly, XGBoost also performed well, with an accuracy of 0.94 and high precision (0.96), recall (0.95), and F1 Score (0.95). The Random Forest model also showed commendable results with an accuracy of 0.95 and a balanced precision, recall, and F1 Score of 0.93. KNN and CatBoost achieved respectable accuracy scores of 0.92 and 0.94, respectively, along with reasonably high precision, recall, and F1 Scores. However, SVM and Naïve Bayes exhibited comparatively lower performance, with SVM achieving an accuracy of 0.79 and Naïve Bayes at 0.69. In summary, the hyperparameter tuning process significantly improved the model's performance. Extra Trees is the best-performing model for brain cancer detection, closely followed by XGBoost and Random Forest. These results provide valuable insights

Table 5
Comparison with other works.

Ref	Contribution	Dataset	Model	Accuracy
[14]	Feature engineering for brain tumor classification in MRI scans	Figshare	SVM, KNN, NB, DT, and Ensemble ML	91.1 % (KNN)
[16]	Machine learning methodology for multiclass classification of malignant brain tumors	SMS Medical College Jaipur, Rajasthan, India	KNN, mSVM, NN	95.86 % (NN)
[18]	Brain tumor classification in MRI images using six machine learning algorithms	Kaggle	RF, NB, Neural Networks, CN2 Rule Induction, SVM, DT	95.3 % (SVM)
[24]	Computer-aided diagnosis model for brain tumor detection and classification using machine learning	Not mentioned	BTDC-MOML (XG-Boost)	97.83 % (XGBoost)
This Work	A comprehensive and effective methodology for brain tumor classification	Kaggle	RF, KNN, SVM, XGBoost, CatBoost, Extra Trees, Naive Bayes	**98.00 % (Extra Trees)**

into selecting the most effective machine learning model and its hyperparameter configuration for accurate brain cancer detection.

5.3. Confusion matrix

The confusion matrix is a vital tool in evaluating the performance of our brain tumor classification model. It helps us understand how well the model has classified the tumor cases into true positives, true negatives, false positives, and false negatives. The high number of TP and TN, along with low FP and FN values, demonstrates the excellent performance of the Extra Trees model in accurately detecting brain cancer, making it a powerful tool for reliable brain tumor classification. The confusion matrixes are represented in Fig. 5.

6. Discussion

The discussion section of our research highlights the significance of our proposed methodology for brain tumor classification. Combining image preprocessing, advanced feature extraction, feature selection, and machine learning models, we achieved remarkable results in distinguishing brain cancer cases from non-cancerous instances. Applying Homomorphic Filtering, Morphological Opening, and Normalization played a crucial role in enhancing image quality and removing noise, leading to more reliable results. Feature importance and hyperparameter tuning further improved the model's performance, with Extra Trees emerging as the best-performing model with an impressive accuracy of 0.98. In Table 5, we compare our results with other works, showcasing the superiority of our methodology in accurately detecting brain cancer. Our research highlights the potential of machine learning in aiding clinicians with precise and reliable brain tumor diagnosis, offering promising prospects for real world clinical applications and improved patient outcomes.

7. Conclusion and future work

Brain tumors proliferate, and the chance of life is reduced compared with other types of tumors. Discovering and classifying brain tumors is essential for the best care and prognosis for the patient. This research shows a complete and helpful way to classify brain tumors using image preprocessing, feature extraction, feature selection, and machine learning models. The proposed approach, which combines Homomorphic Filtering, Morphological Opening, and Normalization, makes it much easier to identify the difference between brain cancer cases and cases that are not dangerous. By figuring out which features are the most important and then tuning the model's performance using hyperparameters, we found that Extra Trees was the best-performing model, with very high accuracy. The result shows the importance of feature extraction and selection methods in medical image analysis. It also shows how machine learning could help doctors make accurate and reliable brain tumor diagnoses. We want to improve the generalizability of our method by adding more advanced imaging modalities and larger datasets in future work. Looking into ensemble techniques and deep learning architectures could also help accurately classify brain tumors. This would open the door for real-world clinical uses and could lead to better patient outcomes.

Funding

This research received no external funding.

Availability of data

The brain tumor MRI image dataset [26] is publicly available.

Ethical approval

Not Applicable.

CRediT authorship contribution statement

Mst Sazia Tahosin: Conceptualization, Data curation, Formal analysis, Writing – original draft, Validation, Writing – review & editing. **Md Alif Sheakh:** Data curation, Formal analysis, Methodology, Resources, Writing – original draft, Visualization, Writing – review & editing. **Taminul Islam:** Conceptualization, Methodology, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing. **Rishalatun Jannat Lima:** Formal analysis, Resources, Validation, Visualization. **Mahbuba Begum:** Project administration, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Not Applicable.

References

[1] Shipp S. The brain circuitry of attention. Trends Cognit Sci 2004;8:223–30. https://doi.org/10.1016/j.tics.2004.03.004.
[2] Gaiseanu F. Info-relational cognitive operability of the posterior cingulate cortex according to the informational model of consciousness. Int J Psychol Brain Sci 2020;5:61–8. https://doi.org/10.11648/j.ijpbs.20200504.12.
[3] Cannon TD, Van Erp TGM, Huttunen M, Lönnqvist J, Salonen O, Valanne L, et al. Regional gray matter, white matter, and cerebrospinal fluid distributions in schizophrenic patients, their siblings, and controls. Arch Gen Psychiatr 1998;55: 1084–91. https://doi.org/10.1001/ARCHPSYC.55.12.1084.
[4] Fidaleo M, Cavallucci V, Pani G. Nutrients, neurogenesis and brain ageing: from disease mechanisms to therapeutic opportunities. Biochem Pharmacol 2017;141: 63–76. https://doi.org/10.1016/J.BCP.2017.05.016.
[5] Szatmari S, Illigens BMW, Siepmann T, Pinter A, Takats A, Bereczki D. Neuropsychiatric symptoms in untreated Parkinson's disease. Neuropsychiatric Dis Treat 2017;13:815–26. https://doi.org/10.2147/NDT.S130997.

[6] Stiller CA, Bayne AM, Chakrabarty A, Kenny T, Chumas P. Incidence of childhood CNS tumours in Britain and variation in rates by definition of malignant behaviour: population-based study. BMC Cancer 2019;19:1–15. https://doi.org/10.1186/S12885-019-5344-7/FIGURES/2.
[7] Pollack IF. The role of surgery in pediatric gliomas. J Neuro Oncol 1999;42: 271–88. https://doi.org/10.1023/A:1006107227856/METRICS.
[8] Abrigo JM, Fountain DM, Provenzale JM, Law EK, Kwong JSW, Hart MG, et al. Magnetic resonance perfusion for differentiating low-grade from high-grade gliomas at first presentation. Cochrane Database Syst Rev 2018;2018. https://doi.org/10.1002/14651858.CD011551.PUB2/MEDIA/CDSR/CD011551/IMAGE_N/NCD011551-TST-001.PNG.
[9] Claus EB, Walsh KM, Wiencke JK, Molinaro AM, Wiemels JL, Schildkraut JM, et al. Survival and low-grade glioma: the emergence of genetic information. Neurosurg Focus 2015;38:E6. https://doi.org/10.3171/2014.10.FOCUS12367.
[10] Fults DW, Taylor MD, Garzia L. Leptomeningeal dissemination: a sinister pattern of medulloblastoma growth. J Neurosurg Pediatr 2019;23:613–21. https://doi.org/10.3171/2018.11.PEDS18506.
[11] Sheakh MdA, Sazia Tahosin Mst, Hasan MM, Islam T, Islam O, Rana MM. Child and maternal mortality risk factor analysis using machine learning approaches. In: 2023 11th international symposium on digital forensics and security (ISDFS), IEEE; 2023. p. 1–6. https://doi.org/10.1109/ISDFS58141.2023.10131826.
[12] kumar DrTV. Classification of brain cancer type using machine learning. J Artificial Intell Capsule Networks 2019. https://doi.org/10.36548/JAICN.2019.2.006. 2019.
[13] Islam T, Kundu A, Islam Khan N, Chandra Bonik C, Akter F, Jihadul Islam M. Machine learning approaches to predict breast cancer: Bangladesh perspective. Smart Innovat, Syst Technol 2022;302:291–305. https://doi.org/10.1007/978-981-19-2541-2_23/COVER.
[14] Pattanaik B, Anitha K, Rathore S, Biswas P, Sethy P, Behera S. Brain tumor magnetic resonance images classification based machine learning paradigms. Współczesna Onkologia 2022;26:268–74. https://doi.org/10.5114/wo.2023.124612.
[15] Kang J, Ullah Z, Gwak J. MRI-based brain tumor classification using ensemble of deep features and machine learning classifiers. Sensors 2021;21:2222. https://doi.org/10.3390/S21062222.
[16] Vidyarthi A, Agarwal R, Gupta D, Sharma R, Draheim D, Tiwari P. Machine learning assisted methodology for multiclass classification of malignant brain tumors. IEEE Access 2022;10:50624–40. https://doi.org/10.1109/ACCESS.2022.3172303.
[17] Ali TM, Nawaz A, Ur Rehman A, Ahmad RZ, Javed AR, Gadekallu TR, et al. A sequential machine learning-cum-attention mechanism for effective segmentation of brain tumor. Front Oncol 2022;12:873268. https://doi.org/10.3389/FONC.2022.873268/BIBTEX.
[18] Asiri AA, Khan B, Muhammad F, Ur Rahman S, Alshamrani HA, Alshamrani KA, et al. Machine Learning-Based Models for Magnetic Resonance Imaging (MRI)-Based Brain Tumor Classification n.d. https://doi.org/10.32604/iasc.2023.032426 .
[19] Modiya P, Vahora S. Brain tumor detection using transfer learning with dimensionality reduction method. Int J Intell Sys Appl Eng 2022;10:201–6.
[20] Senan EM, Jadhav ME, Rassem TH, Aljaloud AS, Mohammed BA, Al-Mekhlafi ZG. Early diagnosis of brain tumour MRI images using hybrid techniques between deep and machine learning. Comput Math Methods Med 2022;2022. https://doi.org/10.1155/2022/8330833.
[21] Nanmaran R, Srimathi S, Yamuna G, Thanigaivel S, Vickram AS, Priya AK, et al. Investigating the role of image fusion in brain tumor classification models based on machine learning algorithm for personalized medicine. Comput Math Methods Med 2022;2022. https://doi.org/10.1155/2022/7137524.
[22] Haq EU, Jianjun H, Huarong X, Li K, Weng L. A hybrid approach based on deep CNN and machine learning classifiers for the tumor segmentation and classification in brain MRI. Comput Math Methods Med 2022:2022. https://doi.org/10.1155/2022/6446680.
[23] Rinesh S, Maheswari K, Arthi B, Sherubha P, Vijay A, Sridhar S, et al. Investigations on brain tumor classification using hybrid machine learning algorithms. J Healthc Eng 2022:2022. https://doi.org/10.1155/2022/2761847.
[24] Uvaneshwari M, Baskar M. Computer-Aided Diagnosis Model Using Machine Learning for Brain Tumor Detection and Classification n.d. https://doi.org/10.32604/csse.2023.035455.
[25] Vijithananda SM, Jayatilake ML, Hewavithana B, Gonçalves T, Rato LM, Weerakoon BS, et al. Feature extraction from MRI ADC images for brain tumor classification using machine learning techniques. Biomed Eng Online 2022;21: 1–21. https://doi.org/10.1186/S12938-022-01022-6/FIGURES/6.
[26] Br35H: Brain Tumor Detection 2020 | Kaggle n.d. https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection (accessed July 31, 2023).
[27] Yugander P, Tejaswini CH, Meenakshi J, Kumar KS, Varma BVNS, Jagannath M. MR image enhancement using adaptive weighted mean filtering and homomorphic filtering. Procedia Comput Sci 2020;167:677–85. https://doi.org/10.1016/J.PROCS.2020.03.334.
[28] Rajagopal R. Glioma brain tumor detection and segmentation using weighting random forest classifier with optimized ant colony features. Int J Imaging Syst Technol 2019;29:353–9. https://doi.org/10.1002/IMA.22331.
[29] Alnowami M, Taha E, Alsebaeai S, Muhammad Anwar S, Alhawsawi A. MR image normalization dilemma and brain tumor classification model accuracy. J Radiat Res Appl Sci 2022;15:33–9. https://doi.org/10.1016/J.JRRAS.2022.05.014.
[30] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: 3rd international conference on learning representations, ICLR 2015 - conference track proceedings; 2014.
[31] Gurusamy R, Subramaniam V. A machine learning approach for MRI brain tumor classification. Comput Mater Continua (CMC) 2017;53:91–108.
[32] Hussain L, Saeed S, Awan IA, Idris A, Nadeem MSA, Chaudhry Q-A. Detecting brain tumor using machines learning techniques based on different features extracting strategies. Current Med Imag Formerly Current Med Imag Rev 2019;15:595–606. https://doi.org/10.2174/1573405614666180718123533.
[33] Pugalenthi R, Rajakumar MP, Ramya J, Rajinikanth V. Evaluation and classification of the brain tumor MRI using machine learning evaluation and classification of the brain tumor MRI using machine learning technique. CEAI 2019;21:12–21.
[34] Alkayyali Z, Idris S. Applied SA-N-J of T and, 2023 undefined. A systematic literature review of deep and machine learning algorithms in cardiovascular diseases diagnosis. JatitOrg 2023:28.
[35] Jekel L, Brim WR, von Reppert M, Staib L, Cassinelli Petersen G, Merkaj S, et al. Machine learning applications for differentiation of glioma from brain metastasis—a systematic review. Cancers 2022;14:1369. https://doi.org/10.3390/CANCERS14061369/S1.
[36] Senan EM, Jadhav ME, Rassem TH, Aljaloud AS, Mohammed BA, Al-Mekhlafi ZG. Early diagnosis of brain tumour MRI images using hybrid techniques between deep and machine learning. Comput Math Methods Med 2022;2022. https://doi.org/10.1155/2022/8330833.
[37] Anitha R, Siva Sundhara Raja D. Development of computer-aided approach for brain tumor detection using random forest classifier. Int J Imaging Syst Technol 2018;28:48–53. https://doi.org/10.1002/IMA.22255.
[38] Kumar DM, Satyanarayana D, Prasad MNG. MRI brain tumor detection using optimal possibilistic fuzzy C-means clustering algorithm and adaptive k-nearest neighbor classifier. J Ambient Intell Humaniz Comput 2021;12:2867–80. https://doi.org/10.1007/S12652-020-02444-7/FIGURES/9.
[39] Chen B, Zhang L, Chen H, Liang K, Chen X. A novel extended Kalman filter with support vector machine based method for the automatic diagnosis and segmentation of brain tumors. Comput Methods Progr Biomed 2021;200:105797. https://doi.org/10.1016/J.CMPB.2020.105797.
[40] Tseng C-J, Tang C. An optimized XGBoost technique for accurate brain tumor detection using feature selection and image segmentation. Healthcare Anal 2023;4: 100217. https://doi.org/10.1016/J.HEALTH.2023.100217.
[41] Nair PC, Gupta D, Devi BI, Kanjirangat V. Building an explainable diagnostic classification model for brain tumor using discharge summaries. Procedia Comput Sci 2023;218. https://doi.org/10.1016/J.PROCS.2023.01.182. 2058–70.
[42] Bandyopadhyay SK, Goyal V, Dutta S. Artificial-intelligence-based diagnosis of brain tumor diseases. Artificial Intell Cancer Diagnosis and Prognosis 2022;3. https://doi.org/10.1088/978-0-7503-3603-1CH2. 2-1-2–17.
[43] Kaur G, Oberoi A. Novel approach for brain tumor detection based on Naïve Bayes classification. Adv Intell Syst Comput 2020;1042:451–62. https://doi.org/10.1007/978-981-32-9949-8_31/COVER.
[44] Belete DM, Huchaiah MD. Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results. Int J Comput Appl 2021; 44. https://doi.org/10.1080/1206212X.2021.1974663. 875–86.
[45] Islam MdT, Ahmed T, Raihanur Rashid ABM, Islam T, Rahman MdS, Tarek Habib Md. Convolutional neural network based partial face detection. In: 2022 IEEE 7th international conference for convergence in technology (I2CT), IEEE; 2022. p. 1–6. https://doi.org/10.1109/I2CT54291.2022.9825259.
[46] Islam T, Hosen MA, Mony A, Hasan MT, Jahan I, Kundu A. A proposed Bi-LSTM method to fake news detection. In: 2022 international conference for advancement in technology, ICONAT 2022; 2022. https://doi.org/10.1109/ICONAT53423.2022.9725937.
[47] Islam T, Kundu A, Lima RJ, Hena MH, Sharif O, Rahman A, et al. Review analysis of ride-sharing applications using machine learning approaches : Bangladesh perspective. New York: CRC Press; 2023. https://doi.org/10.1201/9781003253051-7.
[48] Hasan M, MstS Tahosin, Farjana A, Sheakh MdA, Hasan MM. A harmful disorder: predictive and comparative analysis for fetal anemia disease by using different machine learning approaches. In: 2023 11th international symposium on digital forensics and security (ISDFS), IEEE; 2023. p. 1–6. https://doi.org/10.1109/ISDFS58141.2023.10131838.