ANKARA SCIENCE UNIVERSITY, RESEARCHER
Vol. 4, No. 2, Decmber 2024
e-ISSN: 2717-9494
Research Article/Araştırma Makalesi

Detection of Brain Tumor using Boosting Algorithms based on Feature Selection

Ezgi ÖZER[1*],
[1]Department of Computer Engineering, Piri Reis University, İstanbul, Turkey; ORDIC: 0000-0003-1567-2216
* Corresponding Author: eozer@pirireis.edu.tr

Received: 19 May 2024; Accepted: 23 July 2024

Reference/Atıf: E. Özer, “Detection of Brain Tumor using Boosting Algorithms based on Feature Selection”, Researcher, vol. 04, no. 02, pp. 130–140, 2024.

Abstract

Brain tumors are one of the most common causes of death. An early and correct identification of brain tumors is critical for effective therapy. Using artificial intelligence-based software programs instead of traditional methods can provide more accurate results in brain tumor detection. Especially recently, there have been many studies in the detection of diseases based on the processing of medical images. In this study, a novel hybrid algorithm was proposed based on three different feature selection algorithms (univariate feature ranking for classification using chi-square tests (f-chi2), rank the importance of features using ReliefF algorithm (f-Relief), rank features for classification using minimum redundancy maximum relevance algorithm (f-mRMR), and the classic and ensemble learning, respectively based on support vector machine (SVM) with different kernel structures and ensemble learning (EL) with boosting methods, were performed to detect the brain tumor using magnetic resonance imaging (MRI) features. K-fold is used to prevent overfitting. Analysis results show that a 100% accuracy score was achieved in the ensemble-based classifier in the detection of brain tumors with the proposed hybrid method. As a novelty for detecting the tumors, statistics-based feature selection methods are proposed, to help reduce the size and thus reduce complexity in complex network problems. The proposed method suggests a feature selection algorithm that can help reduce the data size in future studies.

Keywords: tumor detection, feature selection, support vector machine, ensemble learning

1. Introduction

A brain tumor is characterized as an unexpectedly large group of brain cells that can significantly affect the central nervous system. Tumorous tissues that do not have physiological functions may form in the human body as a result of abnormal growth of tissues due to uncontrolled proliferation of cells. These types of tumors, which can disrupt the functionality of the brain by increasing its size and pressure, cause different neurological symptoms such as attention deficit, coordination disorder, speech changes, headaches, hallucinations, anxiety, seizures, and memory loss, depending on their location in the brain. As stated by the latest data from the World Health Organization, brain tumors are one of the most common types of cancer death worldwide and can occur at any age. It causes more deaths, especially for under the age of 40 people [1]. The causes of brain tumors are not fully known. They are divided into primary and secondary. Primary brain tumors are benign and do not spread to other parts of the brain. Secondary brain tumors are malignant, they occur when cancer cells spread to the brain from other organs such as the lung or breast [2].

Early detection of brain tumors is critical for treatments. Patients who are diagnosed early on might have a considerably greater survival rate. However, signs of this condition include headaches, vomiting, visual abnormalities or diplopia, tiredness, trouble swallowing, personality or behavior changes, and hand tremors. As a result, neurosurgeons rely heavily on it to identify even the slightest abnormalities in the brain. However, manual controls by radiologists and doctors can be time-consuming and result in incorrect conclusions. It is critical to examine the photos promptly and precisely. Computer-aided automated detection systems have been created to assist radiologists and clinicians in making decisions. Numerous imaging methods exist. The most common techniques for identifying brain cancers are computed tomography (CT), magnetic resonance imaging (MRI), and head biopsies. Latent characteristics are acquired with these imaging methods. The literature has made extensive use of deep learning and machine learning models for brain tumor diagnosis [3][4][5][6][7][8][9][10].

Developing technology and scientific research allow the development of innovative approaches to cancer detection. Biomedical datasets used for cancer detection are often high-dimensional and multivariate, requiring feature selection to extract meaningful patterns from these data and obtain accurate and reliable results. In this context, feature selection is a key step for enabling meaningful information from large and complex biomedical data sets. Moreover, feature selection helps to improve model performance, reduce the computational cost, reduce the risk of overfitting, reduce meaningless features, and increase the interpretability of the model. So, it can contribute to making the model more effective, faster, and generalizable.

This study aims to investigate the effectiveness of feature selection and machine learning algorithms to develop new and advanced methods for brain tumor detection. Considering the data size, it is not enough for the proposed method to only obtain the tumor with high success. At the same time, it is very important to be able to make faster predictions by using fewer variables. This gives a smoother model. This study focuses on how feature selection and classification methods can be used for cancer detection by examining existing literature about brain tumors. These methods can provide high sensitivity, specificity, and accuracy; and have the potential to contribute to early diagnosis and effective treatment in clinical applications. In this study, the data were analyzed, which have brain tumors and non-tumor. Different feature selection algorithms are performed, including, f-chi2, f-Relief, f-mRMR. After feature selection, ensemble learning-based boosting methods and support vector machine-based kernel structure were performed to detect brain tumors. During the training process, k-fold cross-validation is preferred to handle the overfitting. The following diagram shows the algorithm in this study.

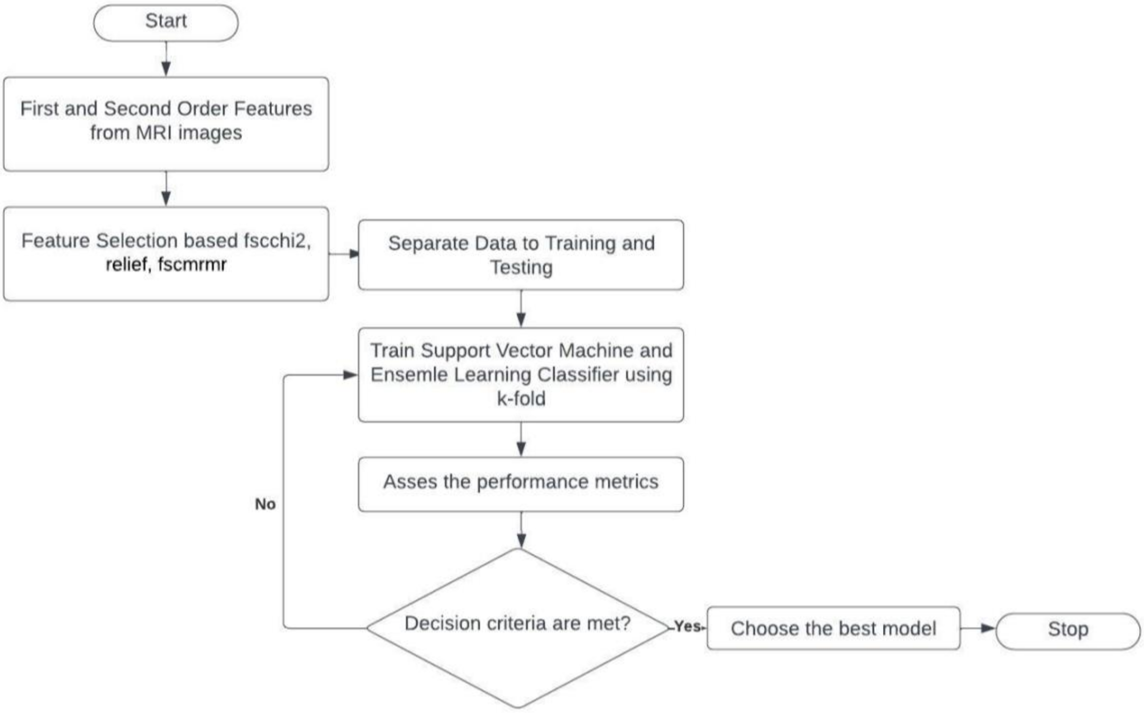

Figure 1: The proposed hybrid algorithm

2. Motivation and Overview

Brain tumor detection techniques are the process of identifying the presence or absence of a tumor using MRI image databases. The goal of tumor detection is usually an MRI image labeled as normal or abnormal. Factors such as the increasing number of patients, tumors of different shapes and sizes, and the fact that they can be found in different parts of the brain and make the process of diagnosing tumors more complicated for experts. It takes a long time for experts to manually diagnose brain tumors. Early diagnosis of the tumor plays an important role in increasing treatment opportunities and survival rates of patients. For this reason, the use of computer-aided systems to reduce the time spent by experts in the early diagnosis of brain tumors, and making improvements in this field have become the main focus of many studies. Brain tumor detection in MRI images can be done using traditional machine learning methods or deep learning techniques. In recent years, remarkable advances have been made in numerous studies aimed at processing medical images and diagnosing brain tumors. Vani et al. [3] proposed a machine learning-based, Support Vector Machine (SVM) method to classify brain tumors. They stated that they predicted brain tumors positively and negatively with an accuracy of 82% and 81.48%, respectively. Mohsen et al. [4] proposed a new method to classify brain tumors using deep learning methods and the Discrete Wavelet Transform (DWT) model. In experimental studies, they achieved 93.94% accuracy with this model [4]. Shahzadi et al. [5] used the CNN-based hybrid model construct to detect brain tumors. Feature extraction and classification used the Long Short-Term Memory (LSTM) structure together with the AlexNet and VggNet CNN models. In the study, they achieved 71% accuracy with AlexNet-LSTM and 84% accuracy with VGGNet-LSTM. Swati et al. [6] proposed transfer learning for the multiclass classification of brain tumors. For this purpose, AlexNet used CNN models VGG16 and VGG19. In experimental studies, AlexNet achieved accuracy rates of 89.95%, 94.65%, and 94.82% in VGG16 and VGG19 models, respectively. Nayak et al. (2022) performed an algorithm based on DWT and DCNN. In experimental studies, the brain tumors were detected with accuracy, as 97% [9]. Rammurthy, and Mahesh [11] proposed a method Whale Harris Hawks optimization (WHHO) based on deep convolutional neural network (DCNN). Their method gave an 81.6% accuracy for detecting tumors. Pendela et al. [12] proposed an approach based on Exponential Deer Hunting Optimization (ExpDHO), Shepard convolutional neural network (ShCNN), deep convolutional neural network (DCNN), Exponential weighted moving average (EWMA) and they detected the tumor with 91.7% accuracy.

3. Methodology

3.1. Univariate feature ranking for classification using chi-square tests (f-chi2)

Univariate feature ranking for classification using chi-square tests (f-chi2) is a statistical technique. It involves evaluating the relevance of individual features for classification using chi-square tests to evaluate the significance of each feature (variable) in a dataset for categorical result prediction [13]. To determine which features are the most discriminative for a given classification problem, univariate feature ranking using chi-square tests might be a useful step in the feature selection process It is noteworthy that although this method is instructive, taking feature interactions into account could enhance the feature selection process, particularly in datasets with greater complexity [14].

$$\chi^2(x_k, y_i) = \frac{N(AD - CB)^2}{(A + C)(B + D)(A + B)(C + D)} \tag{1}$$

In Equality 1, $\boldsymbol{x_k}$ belongs to the kth feature and $\boldsymbol{y_i}$ is the *i*th class. A is the number of samples in $\boldsymbol{y_i}$ that contain the feature $\boldsymbol{x_k}$. B is the number of samples that contain the feature $\boldsymbol{x_k}$ in other classes. C is the number of records in the class $\boldsymbol{y_i}$ that do not contain the feature $\boldsymbol{x_k}$. D is the number of samples that do not contain the feature $\boldsymbol{x_k}$ in other classes and N is the total number of samples in the dataset.

3.2. Rank the importance of features using ReliefF algorithm (f-Relief)

The ReliefF (Relief) algorithm is designed to rank the importance of predictors in a dataset. Relief-based algorithms like ReliefF and RReliefF are employed as feature selection techniques to prioritize predictors (features) within a dataset. They are based on the concept of feature weighting. ReliefF is a deterministic version, while RReliefF introduces randomization for enhanced efficiency [15][16]. When significant features are present among noisy or irrelevant ones, these algorithms are very helpful in recognizing the former. Features' relevance and redundancy are assessed using the feature selection algorithms ReliefF and RReliefF. To identify which traits are important, they consider the distance between instances and their class labels [17]. The following algorithm gives the main idea behind the Relief method [18][19].

Algorithm 1: Relief Method

```
Input: M learning instances $x_k$ described by N features; sampling parameter m
Output: for each feature $F_i$ a quality weight $-1 \le W[i] \le 1$
for i=1 to N do $W[i] = 0.0$
   end for
for l=1 to m do
   randomly pick an instance $x_k$
   find its nearest hit $x_H$ and nearest miss $x_M$
   for i=1 to N do
      $W[i] = W[i] - diff(i, x_k, x_H)/\mathrm{m} + \mathrm{diff}\ (i, x_k, x_{m-M})/\mathrm{m}$
   end for
end for
return (W)
```

3.3. Rank features for classification using minimum redundancy maximum relevance algorithm (f- MRMR)

The Minimum Redundancy Maximum Relevance algorithm emphasizes the redundancy among the characteristics that are chosen as well as their relevance to the target variable [20]. The goal of f-mRMR is to rank features based on their relevance to the target variable while minimizing redundancy among selected features (Ding and Peng, 2005). The following algorithm shows the basic idea behind f-mRMR [21]:

$$argmax\ _{j\epsilon\Omega_S}\big(f(x_j, y) - \frac{1}{|S|^2}\sum_{l\epsilon S} g(x_j, x_l)\big) \tag{2}$$

Algorithm 2: f-mRMR Method

```
$S \longleftarrow \emptyset$
$add\ x_i = argmax\ _{j\epsilon\Omega_S} f(x_j, y) to\ S$
for t=1:k-1 do
add the feature that satisfies Eq. xx to S
end for
return S
```

In Equality 2, the first term and second term maximize the relevance condition and minimizes the redundancy condition, respectively. In this algorithm, x_i is the ith feature in X, $g(x_j, x_l)$ is the function that returns the redundancy between two features x_j and x_l. $f(x_j, y)$ is the function that returns the relevance between a feature x_j and class labels y. S is the selected feature. Ω is all features. Ω_S shows the candidate's features. k is the number of selected features.

3.4. Ensemble Learning

Ensemble learning is a method of combining multiple models and typically aims to improve the predictive performance of the model. Numerous algorithms have been developed for training ensemble classifiers. Bagging, boosting, voting, and stacking algorithms are the most common approaches. Among them, the boosting algorithm is a method for obtaining strong classifiers to produce strong classifiers with low training error from weak classifiers [22][23]. The following algorithm gives the Adaboosting:

Algorithm 3: Adaboosting Method

```
Input: $D = \{d_1, d_2, \ldots, d_n\}$ and $d_i = (x_i, y_i)\ x_i \epsilon X$ and $y_i \epsilon \{y_1, y_2, \ldots, y_M\}$
M: maximum number of classifiers
Output: O is the classifier, $O: X \to \{y_1, y_2, \ldots, y_M\}$
Initialize the weights $w_i^1 = \frac{1}{N}\ i \epsilon \{1, 2, \ldots, N\}$
For $m = 1$ to M, calculate the following metrics:
$Error_m = \sum_{i=1}^{N} w_i^m h(-y_i O_M(x_i))$ ($Error_m$ is the weighted error for $O_m$)
$a_m = \frac{1}{2} log \frac{1-error_m}{error_m}$ ($a_m$ is the weight of weak learners)
$v_i^m = w_i^m \exp(-a_m y_i O_m(x_i))$ (Updating the weights)
$S_m = \sum_{j=1}^{N} v_j$ and $w_i^{m+1} = v_i^m / S_m$ (Normalizing the weights)
$O(x) = sign \sum_{j=1}^{M} a_j O_j(x)$ (Calculating the last classifier)
```

3.5. Support Vector Machine

The fundamental principle behind support vector machines (SVMs) is to maximize the distance between a discrete hyperplane and the data in an N-dimensional space. In where, N is the number of features, to identify a hyperplane that effectively separates the data from each other. The margin is the length of time between the decision border and the closest training sample [24]. It is introduced in the context of statistical learning theory and structural risk minimization. This approach is highly effective in resolving numerous common problems including density estimation, function estimation, and nonlinear classification. It additionally acts as a guide for numerous new advancements in kernel-based techniques [23]. Kernel functions are utilized to create a convex quadratic problem, where the global optimal value is the quadratic function's minimal value, in cases when the dataset has a nonlinear structure that is not linearly separable. The effectiveness of kernel functions varies depending on the nature of the issue. The following table shows the kernel functions and their formula (Bishop, 2006):

Table 1: Kernel function and formula

Kernel Function	Formula
Linear	$K(x_i, x_j) = \langle x_i, x_j \rangle$
RBF	$K(x_i, x_j) = \exp\left(-\gamma \left\lVert x_i - x_j \right\rVert^2\right)$
Gaussian	$K(x_i, x_j) = exp\left(-\frac{\left\lVert x_i - x_j \right\rVert^2}{2\sigma^2}\right)$
Polynomial	$K(x_i, x_j) = (1 + \langle x_i, x_j \rangle)^1$
Quadratic	$K(x_i, x_j) = (1 + \langle x_i, x_j \rangle)^2$
Qubic	$K(x_i, x_j) = (1 + \langle x_i, x_j \rangle)^3$

3.6. Performance Metrics

Evaluation metrics provide information about the success of the prediction, by comparing the predictions obtained from the model with the actual results. They are needed to measure which of the created models will give better results. In this work, the following evaluation criteria were used: the model's accuracy, recall, precision, F1 Score, and G Mean [25][26][27][28]. In table 2, it can be found that these metrics' mathematical formulas.

Table 2: The evaluation metrics

Evaluation Metrics	Formula
Accuracy	$(TP + TN)/(TP + TN + FP + FN)$
Sensitivity	$TP/(TP + FN)$
Specificity	$TN/(TN + FP)$
Precision	$TP/(TP + FP)$
F1_Score	$2 * (Precision * Sensitivity)/(Precision + Sensitivity)$
G-Mean	$\sqrt{Precision * Sensitivity}$

In this table, TP, TN, FP and FN are described as follows:

TP; when both the expected and actual values are positive
TN; when both the predicted value and the actual value are negative.
FP (Type 1 Error); when the prediction is positive but the actual is negative.
FN (Type 2 Error); when the prediction is negative but it actually is positive.

4. Analysis

4.1. Dataset

The dataset includes five first-order features (mean, variance, standard deviation, skewness, kurtosis) and eight texture features (contrast, energy, angular second moment, (ASM), entropy, homogeneity, dissimilarity, correlation, coarseness) from the 3762 images, the number of having tumor and without tumor is 1683 and 2079, respectively. Detailed information can be found at Bohaju, 2020 [29]. Following Figure shows the images with tumor and without tumor.

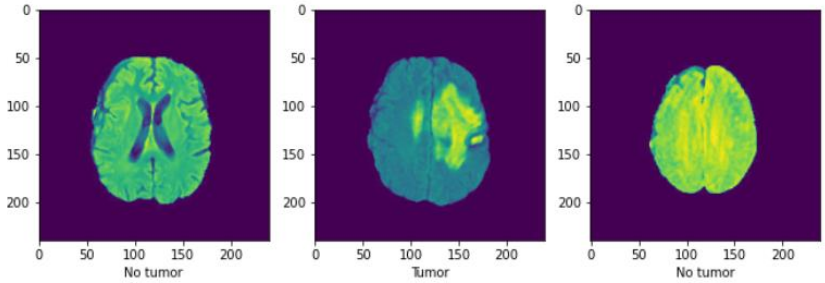

Figure 2: Sample of 3 brain images with labels

4.2. Results

Initially, 13 features were used in the study. With the proposed feature selection methods, the number of features was reduced, and resulted in a reduction in model complexity. During the training process, different hyperparameters were performed to increase the classifiers' performance. To reduce overfitting, k-fold cross-validation was performed during the training processes.

The following tables include results obtained full model and feature selection model using different boostings and kernel structures. In the tables below, NoF, ACC, SPEC, PREC, SENS, F_M, G_M, LOSS are descriped as number of features, accuracy, specifity, precision, sensitiviy, F1 Score, geometric mean, respectively. The classification results are given in the following tables:

Table 3: Ensemble Learning K fold=2

		Testing Data								Training Data	
Model		NoF	ACC	SPEC	PREC	SENS	F_M	G_M	LOSS	ACC	LOSS
Gentle Boost	full model	13	0,99	0,99	0,99	0,99	0,99	0,99	0,01	0,98	0,02
	fscchi2	7	0,99	0,99	0,99	0,99	0,99	0,99	0,01	0,98	0,02
	relieff	***4***	***0,99***	***0,99***	***0,99***	***0,99***	***0,99***	***0,99***	***0,02***	***0,97***	***0,03***
	f-mRMR	3	0,98	0,98	0,98	0,98	0,98	0,98	0,02	0,96	0,04
Ada Boost	full model	13	0,99	0,99	0,99	0,99	0,99	0,99	0,01	0,98	0,02
	fscchi2	4	0,99	0,99	0,99	0,99	0,99	0,99	0,02	0,97	0,03
	relieff	4	0,99	0,99	0,99	0,99	0,99	0,99	0,01	0,97	0,03
	f-mRMR	***3***	***0,99***	***0,99***	***0,99***	***0,99***	***0,99***	***0,99***	***0,02***	***0,96***	***0,04***
Logit Boost	full model	13	0,99	0,99	0,99	0,99	0,99	0,99	0,01	0,98	0,02
	fscchi2	4	0,98	0,98	0,98	0,98	0,98	0,98	0,02	0,97	0,03
	relieff	***5***	***0,99***	***0,99***	***0,99***	***0,99***	***0,99***	***0,99***	***0,02***	***0,98***	***0,02***
	f-mRMR	6	0,99	0,99	0,99	0,99	0,99	0,99	0,01	0,98	0,02

From Table 3, instead of use all features, it can be seen the best results are obtained using Gentle Boost with relieff, AdaBoost with f-mRMR and Logit Boost with relieff. The number of features are 4, 3, and 5, respectively. When k-fold=2, AdaBoost with f-mRMR gave the least complex result.

Table 4: Support vector machine K fold=2.

		Testing Data								Training Data	
Model		NoF	ACC	SPEC	PREC	SENS	F_M	G_M	LOSS	ACC	LOSS
Linear	full model	13	0,96	0,96	0,96	0,96	0,96	0,96	0,04	0,96	0,04
	fscchi2	7	0,96	0,96	0,96	0,96	0,96	0,96	0,04	0,96	0,04
	relieff	7	0,96	0,96	0,96	0,96	0,96	0,96	0,04	0,96	0,04
	f-mRMR	***6***	***0,94***	***0,94***	***0,94***	***0,94***	***0,94***	***0,94***	***0,05***	***0,94***	***0,06***
RBF	full model	13	0,97	0,97	0,97	0,97	0,97	0,97	0,03	0,96	0,04
	fscchi2	4	0,98	0,98	0,98	0,98	0,98	0,98	0,02	0,98	0,02
	relieff	***4***	***0,98***	***0,98***	***0,98***	***0,98***	***0,98***	***0,98***	***0,02***	***0,98***	***0,02***
	f-mRMR	2	0,97	0,97	0,97	0,97	0,97	0,97	0,03	0,96	0,04
Polynomial	full model	13	0,97	0,97	0,97	0,97	0,97	0,97	0,03	0,96	0,04
	fscchi2	***4***	***0,98***	***0,98***	***0,98***	***0,98***	***0,98***	***0,98***	***0,02***	***0,97***	***0,03***
	relieff	3	0,97	0,97	0,97	0,97	0,97	0,97	0,03	0,97	0,03
	f-mRMR	4	0,97	0,97	0,97	0,97	0,97	0,97	0,03	0,96	0,04
Quadratic	full model	13	0,97	0,97	0,97	0,97	0,97	0,97	0,03	0,97	0,03
	fscchi2	7	0,98	0,98	0,98	0,98	0,98	0,98	0,03	0,97	0,03
	relieff	***2***	***0,97***	***0,97***	***0,97***	***0,97***	***0,97***	***0,97***	***0,03***	***0,97***	***0,03***
	f-mRMR	5	0,97	0,97	0,97	0,97	0,97	0,97	0,03	0,97	0,03
Qubic	full model	13	0,98	0,98	0,98	0,98	0,98	0,98	0,02	0,96	0,04
	fscchi2	4	0,97	0,97	0,97	0,97	0,97	0,97	0,03	0,97	0,03
	relieff	***3***	***0,97***	***0,97***	***0,97***	***0,97***	***0,97***	***0,97***	***0,03***	***0,97***	***0,03***
	f-mRMR	5	0,98	0,98	0,98	0,98	0,98	0,98	0,03	0,97	0,03

From Table 4, instead of use all features, it can be seen the best results are obtained using Linear Kernel with f-mRMR, RBF Kernel with relieff, Polynomial Kernel with fscchi2, Quadratic Kernel with relieff, and Qubic Kernel with relieff. The number of features are 6, 4, 4, 2, and 3, respectively. When k-fold=2, Quadratic Kernel with relieff, gave the least complex result having 2 features.

Table 5: Ensemble Learning kfold=10.

		Testing Data								Training Data	
Model		NoF	ACC	SPEC	PREC	SENS	F_M	G_M	LOSS	ACC	LOSS
Gentle Boost	full model	13	1,00	1,00	1,00	1,00	1,00	1,00	0,00	0,98	0,02
	fscchi2	2	0,99	0,99	0,99	0,99	0,99	0,99	0,01	0,96	0,04
	relieff	5	1,00	1,00	1,00	1,00	1,00	1,00	0,01	0,98	0,02
	f-mRMR	4	1,00	1,00	1,00	1,00	1,00	1,00	0,01	0,98	0,02
AdaBoost	full model	13	1,00	1,00	1,00	1,00	1,00	1,00	0,00	0,98	0,02
	fscchi2	6	1,00	1,00	1,00	1,00	1,00	1,00	0,01	0,98	0,02
	relieff	6	1,00	1,00	1,00	1,00	1,00	1,00	0,01	0,98	0,02
	f-mRMR	5	1,00	1,00	1,00	1,00	1,00	1,00	0,01	0,98	0,02
Logit Boost	full model	13	1,00	1,00	1,00	1,00	1,00	1,00	0,00	0,99	0,01
	fscchi2	4	1,00	1,00	1,00	1,00	1,00	1,00	0,01	0,97	0,03
	relieff	2	0,99	0,99	0,99	0,99	0,99	0,99	0,01	0,95	0,05
	f-mRMR	2	0,99	0,99	0,99	0,99	0,99	0,99	0,01	0,96	0,04

From Table 3, instead of use all features, it can be seen the best results are obtained using Gentle Boost with f-mRMR, AdaBoost with f-mRMR and LogitBoost with fscchi2. The number of features are 4, 5 and 4, respectively. When k-fold=10, Gentle Boosting with f-mRMR gave the least complex result.

Table 6: Support vector machine, kfold=10.

		Testing Data								Training Data	
Model			NoF	ACC	SPEC	PREC	SENS	F_M	G_M	LOSS	ACC
Linear	full model	13	0,96	0,96	0,96	0,96	0,96	0,96	0,04	0,96	0,04
	fscchi2	5	0,96	0,96	0,96	0,96	0,96	0,96	0,04	0,96	0,04
	relieff	***3***	***0,95***	***0,95***	***0,95***	***0,95***	***0,95***	***0,95***	***0,05***	***0,95***	***0,05***
	f-mRMR	6	0,95	0,95	0,95	0,95	0,95	0,95	0,05	0,95	0,05
RBF	full model	13	0,98	0,98	0,98	0,98	0,98	0,98	0,02	0,97	0,03
	fscchi2	4	0,98	0,98	0,98	0,98	0,98	0,98	0,02	0,97	0,03
	relieff	***2***	***0,98***	***0,98***	***0,98***	***0,98***	***0,98***	***0,98***	***0,03***	***0,97***	***0,03***
	f-mRMR	5	0,98	0,98	0,98	0,98	0,98	0,98	0,02	0,97	0,03
Polynomial	full model	13	0,99	0,99	0,99	0,99	0,99	0,99	0,02	0,97	0,03
	fscchi2	***4***	***0,98***	***0,98***	***0,98***	***0,98***	***0,98***	***0,98***	***0,02***	***0,97***	***0,03***
	relieff	5	0,98	0,98	0,98	0,98	0,98	0,98	0,02	0,97	0,03
	f-mRMR	5	0,98	0,98	0,98	0,98	0,98	0,98	0,02	0,97	0,03
Quadratic	full model	13	0,98	0,98	0,98	0,98	0,98	0,98	0,02	0,97	0,03
	fscchi2	***5***	***0,98***	***0,98***	***0,98***	***0,98***	***0,98***	***0,98***	***0,03***	***0,97***	***0,03***
	relieff	2	0,97	0,97	0,97	0,97	0,97	0,97	0,14	0,87	0,13
	f-mRMR	6	0,98	0,98	0,98	0,98	0,98	0,98	0,02	0,97	0,03
Qubic	full model	13	0,99	0,99	0,99	0,99	0,99	0,99	0,01	0,97	0,03
	fscchi2	***4***	***0,98***	***0,98***	***0,98***	***0,98***	***0,98***	***0,98***	***0,02***	***0,97***	***0,03***
	relieff	3	0,97	0,97	0,97	0,97	0,97	0,97	0,05	0,95	0,05
	f-mRMR	5	0,98	0,98	0,98	0,98	0,98	0,98	0,02	0,97	0,03

From Table 6, instead of use all features, it can be seen the best results are obtained using Linear Kernel with relieff, RBF Kernel with relieff, Polynomial Kernel with fscchi2, Quadratic Kernel with fscchi2, and Qubic Kernel with fscchi2. The number of features are 6, 4, 4, 2, and 3, respectively. When k-fold=2, Quadratic Kernel with relieff, gave the least complex result having 2 features.

5. Discussion and Conclusion

By identifying and ranking the characteristics that are most likely to provide useful information for a classification task, this method improves the interpretability and accuracy of the model. In this study, thirteen extracted features from images are used. As a novelty for detecting the tumors, the statistics-based feature selection methods are proposed, to help reduce the size and thus reduce complexity in complex network problems. From Table 3 to 6, the results obtained from full model and 3 different feature selection methods, ensemble learning with different boosting, and different kernel structures for k fold 2 and 10 are compared. In these tables, the best results are shown in italics. From the results, ensemble learning based on boosting methods gives better results than SVM based on kernel structures. The proposed hybrid algorithm-based feature selection and boosting approaches give good results in classifying data obtained via MRI from patients with and without brain tumors.

Among the feature selection algorithms, the f-chi2, f-Relief, f-mRMR, were used to select the features from all features. To detect the brain tumor, classified machine learning algorithms were performed which are Support vector machine-based kernel and Ensemble Learning based Boosting. Some results were obtained when the K fold was 2 and 10, respectively. The best result was obtained using f-mRMR feature selection method and ensemble learning, when k-fold was 10. The proposed hybrid algorithm reduced the model complexity, by decreasing the number of features. Comparative results of some studies in the literature are given in the table below:

Table 7: The performances of existing approaches in the literature

Source	**Classes**	**Method**	**Accuracy**
Vani et al., 2017 [3]	Tumor/No Tumor	SVM	82%
Mohsen et al., 2018 [4]	Tumor/No Tumor	DWT, DL	93.94%
Shahzadi et al., 2018 [5]	Tumor/No Tumor	AlexNet and VggNet, CNN	84%
Ghahfarrokhi and Khodadadi, 2020 [7]	Tumor/No Tumor	LE, ApEn, FD, GLCM, DWT, SVM, KNN	98,90%
Al-Saffar and Yildirim, 2021 [8]	No Tumor/HGG/LGF	LDI-Means, MI, SVD, SVM, MLP, MES	91,02%
Nayak et al., 2022 [9]	Tumor/No Tumor	DWT, SDA-DA CNN	97%
Isunuri and Kakarla, 2023 [10]	No Tumor/HGG/LGF	EfficientNetB4, EMCMA	98,35%
The Proposed work	*Tumor/No Tumor*	*f-chi2, f-Relief, f-mRMR, EL*	*100%*

LE: Lyapunov Exponent
ApEn: Approximate Entropy
FD: Fractal Dimension
GLCM: Gray-Level Co-occurrence Matrix
DWT: Discrete Wavelet Transform
SVM: Support Vector Machine
KNN: K-Nearest Neighbors
LDI-Means: Local Difference in Intensity - Means
MI: Mutual Information
SVD: Singular Value Decomposition
MLP: Multi-Layer Perceptron
MES: Multiple Eigenvalues Selection
SDA-DA CNN: Deep Autoencoder based Spectral Data Augmentation
EMCMA: Multi-Path Convolution Network with Multi-Head Attention
HGG: High Grade Gliomas
LGF: Low Grade Fliomas

Declaration of Interest

The author declares that there is no conflict of interest.

Financial Support

This work was supported in part by the Scientific and Technological Research Council of TURKEY (TUBITAK) [grant number 1059B141900679].

Declaration of ethical standards

The author(s) of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.

References

[1] World Health Organization Report (WHO), 2022. https://www.who.int

[2] Doolittle, N.D., "State of the science in brain tumor classification. Paper presented at the Seminars in oncology nursing", 2004.

[3] Vani, N., Sowmya, A., and Jayamma, N., "Brain Tumor Classification using Support Vector Machine. International Research Journal of Engineering and Technology (IRJET)", 4(7), 792-796, 2017.

[4] Mohsen, H, El-Dahshan E.S.A., El-Horbaty E.S.M., and Salem A.B.M., "Classification using Deep Learning Neural Networks for Brain Tumors, Future Computing and Informatics Journal", 3(1), 68-71, 2018.

[5] Shahzadi I, Tang T. B, Meriadeau F., and Quyyum, A., "CNN-LSTM: Cascaded framework for brain tumor classification. IEEE EMBS Conference on Biomedical Engineering and Sciences", IECBES; 3-6 December, Sarawak, Malaysia: IEEE, 633-637, 2018.

[6] Swati, Z.N.K., Zhao, Q., Kabir, M., Ali, F., Ali, Z., Ahmed, S., and Lu, J., "Brain tumor classification for MR images using transfer learning and fine-tuning", Computerized Medical Imaging and Graphics, 75, 34-46, 2019.

[7] Ghahfarrokhi, S.S., and Khodadadi, H., "Human brain tumor diagnosis using the combination of the complexity measures and texture features through magnetic resonance image", Biomedical Signal Processing and Control, 61, 2020. https://doi.org/10.1016/j.bspc.2020.102025.

[8] Al-Saffar, Z.A., and Yildirim, T., "A hybrid approach based on multiple Eigenvalues selection (MES) for the automated grading of a brain tumor using MRI", Computer Methods and Programs in Biomedicine, 2021, 201, 105945, ISSN 0169-2607, https://doi.org/10.1016/j.cmpb.2021.105945.

[9] Nayak, D.R., Padhy, N., Mallick, P.K., and, Singh. A., "A deep autoencoder approach for detection of brain tumor images, Computers and Electrical Engineering", 2022, 102, 108238, ISSN 0045-7906, https://doi.org/10.1016/j.compeleceng.2022.108238.

[10] Isunuri, V.B., and Kakarla, J., "EfficientNet and multi-path convolution with multi-head attention network for brain tumor grade classification, Computers and Electrical Engineering", 2023, 108, 108700, ISSN 0045-7906, https://doi.org/10.1016/j.compeleceng.2023.108700.

[11] D. Rammurthy, P. Mahesh, "Whale Harris Hawks optimization based deep learning classifier for brain tumor detection using MRI images," Journal of King Saud University-Computer and Information Sciences, vol. 34, no. 6, pp. 3259-3272, 2020.

[12] Pendela Kanchanamala, Revathi K.G., M. "Belsam Jeba Ananth, Optimization-enabled hybrid deep learning for brain tumor detection and classification from MRI", Biomedical Signal Processing and Control, Volume 84, 2023, 104955, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2023.104955.

[13] Hastie, T., Tibshirani, R., and Friedman, J., "The Elements of Statistical Learning Data Mining", Inference, and Prediction, 2007, Second Edition, Springer.

[14] Hardle, W.K., and Simar, L., Applied Multivariate Statistical Analysis, Fourth Edition, 2015, Springer.

[15] Kononenko, I., Šimec, E., and Robnik-Šikonja, M., "Overcoming the Myopia of Inductive Learning Algorithms with RELIEFF", Applied Intelligence **7**, 1997, 39-55. https://doi.org/10.1023/A:1008280620621.

[16] Robnik-Šikonja, M., and Kononenko, I,. "Theoretical and Empirical Analysis of ReliefF and RReliefF. Machine Learning", 2003, 53, 23-69. https://doi.org/10.1023/A:1025667309714

[17] Blum, A.L., and Langley. P., "Selection of Relevant Features and Examples in Machine Learning Artificial Intelligence", 1997, 97 (1), 245-271.

[18] Liu H., and Motoda, H. (Ed.)., Computational Methods of Feature Selection, Chapman and Hall/CRC Data Mining and Knowledge Discovery Series, Kononenko, I and Sikonja, M., 2008, Chapter 9- Non-Myopic Feature Quality Evaluation with (R) ReliefF, New York.

[19] Urbanowicz, R.J., Meeker, M., La Cava, W., Olson, R.S., Moore, J.H., “Relief-based feature selection: Introduction and review, Journal of Biomedical Informatics”, 2018, 85, 189-203, ISSN 1532-0464, https://doi.org/10.1016/j.jbi.2018.07.014.

[20] Peng, H., Long, F., and Ding, C., "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," in IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(8), 1226-1238, doi: 10.1109/TPAMI.2005.159.

[21] EL-Manzalawy, Y., Hsieh, T.Y., and Shivakumar, M., “Min-Redundancy and Max-Relevance Multi-View Feature Selection for Predicting Ovarian Cancer Survival using Multi-Omics Data”, BMC Med Genomics, 2018, 11(3). doi.org/10.1186/s12920-018-0388-0

[22] Polikar, R., Ensemble learning. Ensemble machine learning, 2012, 10th ed. Boston, Springer, 1-34.

[23] Ozer, E., “Early Diagnosis of Epileptic Seizures over EEG Signals using Deep Learning Approach”, Mimar Sinan Fine Arts University, 2023, Institute of Science and Technology, PhD Thesis.

[24] Cortes, C., and Vapnik, V., Support-Vector Networks. Kluwer Academic Publishers, 1995.

[25] Van Rijsbergen, C.J., Information Retrieval, London: Butterworths, 1979

[26] Fawcett, T., “An introduction to ROC analysis”, Pattern Recognition Letters, 2006, 27(8), 861-874.

[27] Hastie, T., Tibshirani, R., and Friedman, J., The Elements of Statistical Learning, Springer, 2009.

[28] Powers, D.M., “Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation”, Journal of Machine Learning Technologies, 2011, 2(1), 37-63.

[29] Bohaju, J., Brain Tumor Data set, Kaggle, 2020. doi.org/10.34740/KAGGLE/DSV/1370629