Journal of Mathematics and Informatics
Vol. 27, 2024, 39-48
ISSN: 2349-0632 (P), 2349-0640 (online)
Published 17 December 2024
www.researchmathsci.org
DOI:http://dx.doi.org/10.22457/jmi.v27a04248

Journal of
Mathematics and
Informatics

Prediction of Lower-Grade Glioma and Glioblastoma Multiforme Using Machine Learning Models and Optuna-Optimization

Mohammad Raquibul Hossain[1*]***, Md. Jamal Hossain***[2]***, Md. Mijanoor Rahman***[3] **and** ***Mohammad Manjur Alam***[4]

[1,2]Department of Applied Mathematics, Noakhali Science and Technology University, Noakhali-3814, Bangladesh. E-mail: [1]raquib.math@gmail.com, [2]z_math_du@yahoo.com

[3]Department of Mathematics, Mawlana Bhashani Science and Technology University, Santosh, Tangail-1902, Bangladesh. E-mail: mizanmath@yahoo.com

[4]Department of Computer Science and Engineering, International Islamic University Chittagong (IIUC), Chittagong-4318, Bangladesh. E-mail: manjufse@iiuc.ac.bd
[*]Corresponding author

Received 12 November 2024; accepted 16 December 2024

Abstract. Artificial intelligence including machine learning (ML) is considered new electricity for human civilization which started contributing to almost all sectors. In the healthcare sector, the diagnosis of brain gliomas or tumours is a very crucial task for doctors. Traditional practices in this task are laborious, time-consuming and also expensive. In this case, effective ML techniques can be of great assistance to both doctors and patients. This paper presents the findings from the experimentation of 19 ML models for brain glioma prediction using a dataset of UCL ML repository consisting of 839 instances and 23 input features (20 molecular and 3 clinical related to demographics). From the experimented results Optuna-tuned logistic regression was found to outperform other ML models. The study results indicate that ML techniques can have a high potential in medical diagnoses like glioma prediction. Thus, this field needs further research and exploration.

Keywords: Machine learning; Optuna; logistic regression; brain tumor; support vector classifier

AMS Mathematics Subject Classification (2010): 62H30, 62C25

1. Introduction

Brain is a vital human organ and brain tumors as well as brain cancers are life threatening. However, very few reliable diagnostic tools and even less effective treatment options are in current practice. In recent years, the development of medical technologies carries machine learning (ML) power as a technique to enhance glioma detection, diagnosis and prognosis. Brain gliomas are primary brain tumors that arise from neural glial cells. However, Gliomas of two major types, including Lower Grade Gliomas (LGGs) and

Glioblastoma Multiforme (GBM), have differences in prognosis and clinical handling, as well as in their response to treatment. GBMs manifest quickly, whereas LGGs generally take a long time to manifest. GBMs generally show high aggressiveness and poor survival while LGGs are benign in nature. Early and accurate diagnosis of these two types in clinical practice remains critically dependent upon timely intervention and individual approaches to treatment. However, using techniques of ML, it can be used on medical images as well as genetic data and patient outcomes to do earlier detection and personalized treatment plans.

There are various research studies on brain Glioma-type prediction. For glioma grading, [1] used voting-based ensemble ML methods and achieved accuracy of 87.606% for one dataset and 79.668% for another dataset. The study of [2] investigated how well different ML models perform to predict glioma outcome. Gliomas-related explorative reviews were done by [3], [4] and [5] which were mainly focused on ML methods.

In tumor grade prediction using ML, [6] found that random forest model was stable and better performing than logistic regression and support vector machine. To predict GBM prognosis, [7] and to predict health-related life quality outcomes of meningioma, LGG and acoustic neuroma patients, [8] employed ML approaches. The study of [9] investigated ML application for glioma patient survival prediction. Using deep convolutional neural networks, [10] improved prediction of glioma grading. To classify molecular subtype of LGG, [11] employed MRI-based ML approach. The study of [12] focused on tumor classification using ML on patients with special type of GMB. To identify driver mutations in GBM, [13] and to classify ependymoma and GBM, [14] employed ML.

All these works reflects the potential usability of ML methods in medical diagnostics like glioms enhancing with more and large datasets and robust, stable and effective models. Also, with the improvement and availability of internet, computer storage, computing speed and fast as well as more RAMs, generating and using large data is not a big deal now-a-days.

In this study we propose a machine learning approach coupled with the powerful hyperparameter optimization framework “Optuna” to determine which tumors are LGG and which are GBM. Unlike some optimization techniques, Optuna uniquely allows machine learning algorithms to use high predictive power due to its intelligent search of hyper parameter combinations.

2. Materials and methods

This section describes brain Glioma dataset, experimented ML models and evaluation metrics.

The brain tumor or glioma dataset collected from UCI ML repository [15] was created through The Cancer Genome Atlas (TCGA) project by National Cancer Institute (NCI), National Institutes of Health (NIH). It has 839 samples and 24 features include target or outcome feature (i.e., glioma grade LGG or GBM). Each sample has 20 molecular features (which can be mutated or not_mutated) and 3 clinical features (related to the demographics). Table 1 presents description as well as values of feature variables. In this study, 80:20 train-test split ratio was used where in-sample 671 data were used for training the ML models and out-sample 168 data were used to evaluate the performance of the models.

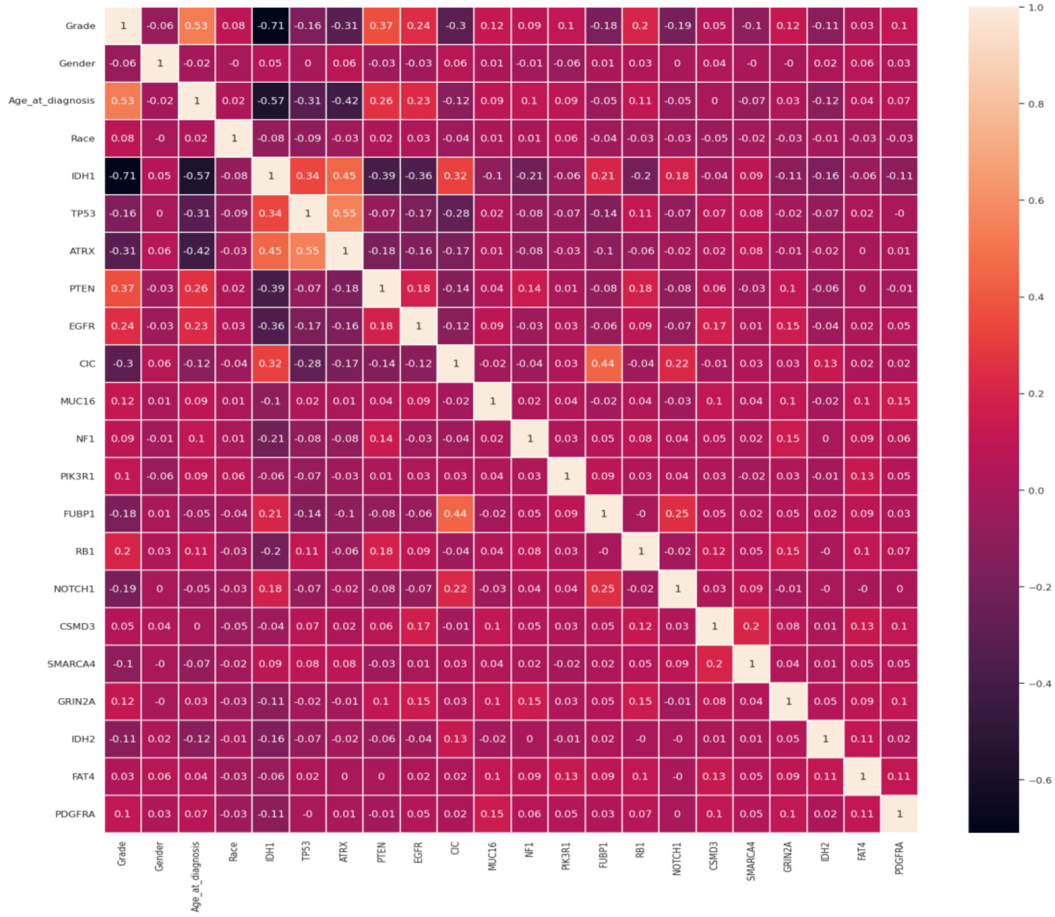
Table 1: Feature description of glioma grade dataset

Variable Name	Description	Value
Grade	Glioma grade category	0 for "LGG"; 1 for "GBM"
Gender	Patient's Gender	0 for "male"; 1 for "female"
Age_at_diagnosis	Age (number of days calculated) during diagnosis with the	Numeric
Race	Patient's Race	0 for "white"; 1 for "black or african American"; 2 for "asian"; 3 for "american indian or alaska native")
IDH1	Isocitrate dehydrogenase (NADP(+))1	0 or 1*
TP53	Tumor protein p53	0 or 1*
ATRX	ATRX chromatin remodeler	0 or 1*
PTEN	Phosphatase and tensin homolog	0 or 1*
EGFR	Eepidermal growth factor receptor	0 or 1*
CIC	Capicua transcriptional repressor	0 or 1*
MUC16	Mucin 16, cell surface associated	0 or 1*
PIK3CA	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha	0 or 1*
NF1	Neurofibromin 1	0 or 1*
PIK3R1	Phosphoinositide-3-kinase regulatory subunit 1	0 or 1*
FUBP1	Far upstream element binding protein 1	0 or 1*
RB1	RB transcriptional corepressor 1	0 or 1*
NOTCH1	Notch receptor 1	0 or 1*
BCOR	BCL6 corepressor	0 or 1*
CSMD3	CUB and Sushi multiple domains 3	0 or 1*
SMARCA4	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4	0 or 1*
GRIN2A	Glutamate ionotropic receptor NMDA type subunit 2A	0 or 1*
IDH2	Isocitrate dehydrogenase (NADP(+)) 2	0 or 1*
FAT4	FAT atypical cadherin 4	0 or 1*
PDGFRA	Platelet-derived growth factor receptor alpha	0 or 1*

* 0 for not_muted; 1 for muted

Figure 1 presents correlation among feature values. Highest positive correlation is correlation between glioma grade outcome and age, i.e., 0.53. Other positively correlated features with glioma grade outcome are PTEN (0.37), EGFR (0.24), RB1 (0.2), MUC16

(0.12), GRIN2A (0.12), PIK3R1 (0.1), PDGFRA (0.1), NF1 (0.09) etc. Highest negative correlation is correlation between glioma grade outcome and IDH1, i.e., -0.71. Other negatively correlated features with glioma grade outcome are ATRX (-0.31), CIC (-0.3), NOTCH1 (-0.19), FUBP1 (-0.18), TP53 (-0.16), IDH2 (-0.11), SMARCA (-0.1) etc.



	Grade	Gender	Age_at_diagnosis	Race	IDH1	TP53	ATRX	PTEN	EGFR	CIC	MUC16	NF1	PIK3R1	FUBP1	RB1	NOTCH1	CSMD3	SMARCA4	GRIN2A	IDH2	FAT4	PDGFRA
Grade	1	-0.06	0.53	0.08	-0.71	-0.16	-0.31	0.37	0.24	-0.3	0.12	0.09	0.1	-0.18	0.2	-0.19	0.05	-0.1	0.12	-0.11	0.03	0.1
Gender	-0.06	1	-0.02	-0	0.05	0	0.06	-0.03	-0.03	0.06	0.01	-0.01	-0.06	0.01	0.03	0	0.04	-0	-0	0.02	0.06	0.03
Age_at_diagnosis	0.53	-0.02	1	0.02	-0.57	-0.31	-0.42	0.26	0.23	-0.12	0.09	0.1	0.09	-0.05	0.11	-0.05	0	-0.07	0.03	-0.12	0.04	0.07
Race	0.08	-0	0.02	1	-0.08	-0.09	-0.03	0.02	0.03	-0.04	0.01	0.01	0.06	-0.04	-0.03	-0.03	-0.05	-0.02	-0.03	-0.01	-0.03	-0.03
IDH1	-0.71	0.05	-0.57	-0.08	1	0.34	0.45	-0.39	-0.36	0.32	-0.1	-0.21	-0.06	0.21	-0.2	0.18	-0.04	0.09	-0.11	-0.16	-0.06	-0.11
TP53	-0.16	0	-0.31	-0.09	0.34	1	0.55	-0.07	-0.17	-0.28	0.02	-0.08	-0.07	-0.14	0.11	-0.07	0.07	0.08	-0.02	-0.07	0.02	-0
ATRX	-0.31	0.06	-0.42	-0.03	0.45	0.55	1	-0.18	-0.16	-0.17	0.01	-0.08	-0.03	-0.1	-0.06	-0.02	0.02	0.08	-0.01	-0.02	0	0.01
PTEN	0.37	-0.03	0.26	0.02	-0.39	-0.07	-0.18	1	0.18	-0.14	0.04	0.14	0.01	-0.08	0.18	-0.08	0.06	-0.03	0.1	-0.06	0	-0.01
EGFR	0.24	-0.03	0.23	0.03	-0.36	-0.17	-0.16	0.18	1	-0.12	0.09	-0.03	0.03	-0.06	0.09	-0.07	0.17	0.01	0.15	-0.04	0.02	0.05
CIC	-0.3	0.06	-0.12	-0.04	0.32	-0.28	-0.17	-0.14	-0.12	1	-0.02	-0.04	0.03	0.44	-0.04	0.22	-0.01	0.03	0.03	0.13	0.02	0.02
MUC16	0.12	0.01	0.09	0.01	-0.1	0.02	0.01	0.04	0.09	-0.02	1	0.02	0.04	-0.02	0.04	-0.03	0.1	0.04	0.1	-0.02	0.1	0.15
NF1	0.09	-0.01	0.1	0.01	-0.21	-0.08	-0.08	0.14	-0.03	-0.04	0.02	1	0.03	0.05	0.08	0.04	0.05	0.02	0.15	0	0.09	0.06
PIK3R1	0.1	-0.06	0.09	0.06	-0.06	-0.07	-0.03	0.01	0.03	0.03	0.04	0.03	1	0.09	0.03	0.04	0.03	-0.02	0.03	-0.01	0.13	0.05
FUBP1	-0.18	0.01	-0.05	-0.04	0.21	-0.14	-0.1	-0.08	-0.06	0.44	-0.02	0.05	0.09	1	-0	0.25	0.05	0.02	0.05	0.02	0.09	0.03
RB1	0.2	0.03	0.11	-0.03	-0.2	0.11	-0.06	0.18	0.09	-0.04	0.04	0.08	0.03	-0	1	-0.02	0.12	0.05	0.15	-0	0.1	0.07
NOTCH1	-0.19	0	-0.05	-0.03	0.18	-0.07	-0.02	-0.08	-0.07	0.22	-0.03	0.04	0.04	0.25	-0.02	1	0.03	0.09	-0.01	-0	-0	0
CSMD3	0.05	0.04	0	-0.05	-0.04	0.07	0.02	0.06	0.17	-0.01	0.1	0.05	0.03	0.05	0.12	0.03	1	0.2	0.08	0.01	0.13	0.1
SMARCA4	-0.1	-0	-0.07	-0.02	0.09	0.08	0.08	-0.03	0.01	0.03	0.04	0.02	-0.02	0.02	0.05	0.09	0.2	1	0.04	0.01	0.05	0.05
GRIN2A	0.12	-0	0.03	-0.03	-0.11	-0.02	-0.01	0.1	0.15	0.03	0.1	0.15	0.03	0.05	0.15	-0.01	0.08	0.04	1	0.05	0.09	0.1
IDH2	-0.11	0.02	-0.12	-0.01	-0.16	-0.07	-0.02	-0.06	-0.04	0.13	-0.02	0	-0.01	0.02	-0	-0	0.01	0.01	0.05	1	0.11	0.02
FAT4	0.03	0.06	0.04	-0.03	-0.06	0.02	0	0	0.02	0.02	0.1	0.09	0.13	0.09	0.1	-0	0.13	0.05	0.09	0.11	1	0.11
PDGFRA	0.1	0.03	0.07	-0.03	-0.11	-0	0.01	-0.01	0.05	0.02	0.15	0.06	0.05	0.03	0.07	0	0.1	0.05	0.1	0.02	0.11	1

Figure 1: Correlation matrix of glioma grade data features

There are various ML models used for regression, classification and other purposes. Also, among classification algorithms, some are single benchmark models while others are ensemble methods [16]. Table 2 briefly portrays the ML methods experimented in this study. To get optimized results (i.e., reduce overfitting or underfitting), suitable hyperparameters for the ML models are required to determine by tuning different potential values. In this case, an open source framework “Optuna” [17] is very useful for hyperparameter optimization which employed in this study.

Table 2: Description of ML methods used for glioma grade (LGG or GBM) prediction

	Model	Description
1	ABC	ABC is short form for AdaBoost Classifier. It adopts boosting method which iteratively combines weak models. ABC particularly focuses on instances which tend to misclassify.
2	BC	Bagging (or bootstrap aggregating) Classifier (or BC in short form) splits dataset into bootstrapped subsets for training and then combines multiple trained models.
3	CBC	CBC stands for CatBoost Classifier. It optimizes gradient boosting for categorical features. Also, CBC reduces necessity of data preprocessing and it can manage overfitting.
4	DTC	Decision Tree Classifier (abbreviated as DTC) splits data into tree-like pattern for prediction on outcomes. It is easy to interpret while it tends to overfit.
5	ETC	Full form of ETC is ExtraTrees Classifier. It is similar to random forest classifier. However, ETC splits nodes randomly and it is faster. Also, it reduces variance.
6	GBMC	Full form of GBMC is Gradient Boosting Machine Classifier. GMBC corrects errors iteratively occurred in previous models and thus improves accuracy. However, it is computationally expensive.
7	GNBC	GNBC stands for Gaussian Naive Bayes Classifier is a probabilistic model based on Bayes theorem. GNBC assumes that features follow Gaussian distribution and also they independ.
8	HGBC	HGBC is short form for Histogram-Based Gradient Boosting Classifier. It is a type of gradient boosting approach. For efficiency, HGBC turns continuous features into histogram. This algorithm works well for large datasets.
9	kNNC	k-Nearest Neighbors Classifier (or kNNC in short form) classifies an outcome by measuring majority class of its nearest k neighbors.
10	LDA	LDA is abbreviated for Linear Discriminant Analysis. It maximizes ability of class separation by transforming data to lower dimensional space. Assumption in LDA is that classes have same covariance matrix and follow Gaussian distribution.
11	LGBMC	LGBMC stands for Light Gradient Boosting Machine or LightGBM Classifier. Specially designed for large dataset, LGBMC is fast and efficient.
12	LR	LR, i.e., Logistic Regression, is an ML classification algorithm. It classifies outcomes based on probabilities found from sigmoid function.
13	OptunaLR	OptunaLR is short form for logistic regression tuned with optuna, an efficient open source hyper-parameter tuning tool.
14	QDA	Quadratic Discriminant Analysis is abbreviated as QDA. It is similar to LDA. However, in QDA each class is allowed to possess its own covariance matrix.

15	RFC	Random Forest Classifier, abbreviated as RFC, is an ML algorithm which combines multiple decision tree classifiers to improve accuracy as well as reduce overfitting.
16	SC	SC stands for Stacking Classifier. It trains a meta-classifier by combining predictions of multiple models. It gains predictability strength from other classifiers.
17	SoftVC	Soft Voting Classifier, abbreviated as SoftVC, combines predicted probabilities of multiple classifiers by averaging. It can effectively balance multifarious models.
18	SVC	SVC stands for Support Vector Classifier. SVC separates outcomes finding best boundary, i.e., hyper-plane.
19	XGBC	XGBC is abbreviated for eXtreme Gradient Boosting or XGBoost Classifier. XGBC improves performances by enhancing gradient boosting with regularization technique.

The performance metrics frequently used in ML classification problems are i. accuracy, ii. precision, iii. recall and iv. F1-score which are defined as follows:

$$\text{i. Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \qquad \text{ii. Precision} = \left(\frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}\right)$$

$$\text{iii. Recall} = \left(\frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}\right) \qquad \text{iv. } \mathrm{F}_1 = \left(\frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}\right)$$

where True Positive (TP) and True Negative (TN) are correctly predicted positive value and negative value respectively ; False Positive (FP) and False Negative (FN) are incorrectly predicted positive value and negative value respectively.

The most commonly used performance metric of classification problem is accuracy. A receiver operating characteristic (ROC) curve is widely used visual performance tool for binary classification. ROC curve is the graph of true positive rate ($\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP+FN}}$) against false positive rate ($\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP+TN}}$) at each threshold.

Figure 2 presents the flow chart of overall process followed for implementation of machine learning models on brain glioma dataset (from data collection to model evaluation using different performance metrics).

3. Results

Table 2 presents the performance of 19 ML models used to predict glioma grade using 80:20 train-test split. OptunaLR (i.e., Logistic Regression optimized using optuna) outperformed other methods with recall score of 91.38%, accuracy of 90.48%, precision of 90.39% and F1-score of 90.41%. The least performing method was QDA with recall score of 64.43% and accuracy of 58.93%. It was evident that simple linear models performed comparatively better than sophisticated ensemble and nonlinear models. Figure 3 and Figure 4 present the ROC curves and accuracy bar chart respectively for the ML models.

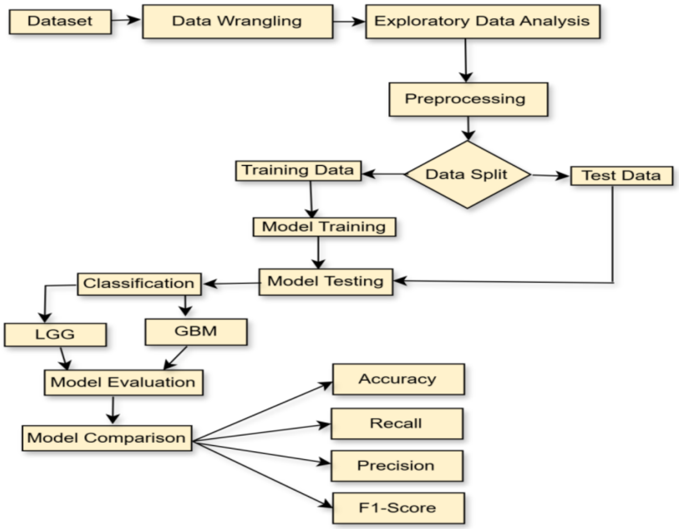


Figure 2: Flowchart of ML technique for glioma grade prediction

Table 2: ML models performance for glioma grade (LGG or GBM) prediction

	Model	Precision	Recall	F1	Accuracy
1	OptunaLR	90.39	91.38	90.41	90.48
2	SVC	90.09	91.05	89.84	89.88
3	LDA	90.09	91.05	89.84	89.88
4	LR	89.88	90.86	89.82	89.88
5	DTC	88.9	89.83	88.64	88.69
6	SC	87.87	88.75	87.99	88.1
7	ABC	87.22	88.04	87.37	87.5
8	SoftVC	88.25	88.99	87.48	87.5
9	GBMC	86.57	87.34	86.75	86.9
10	kNNC	86.51	87.15	86.72	86.9
11	CBC	86.02	86.82	86.17	86.31
12	XGBC	84.18	84.9	84.34	84.52
13	HGBC	83.5	84.01	83.67	83.93
14	LGBMC	82.88	83.3	83.04	83.33
15	RFC	82.29	82.41	82.35	82.74
16	ETC	80.95	79.73	80.14	80.95
17	BC	77.14	76.07	76.41	77.38
18	GNBC	81.7	78.87	75.4	75.6
19	QDA	75.36	64.43	56.05	58.93

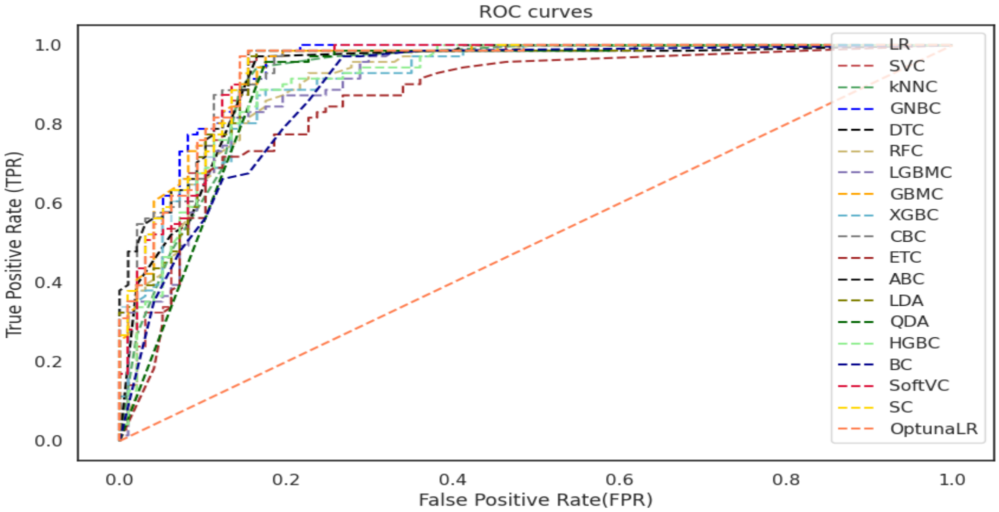


Figure 3: ROC curves of ML models for glioma grade prediction

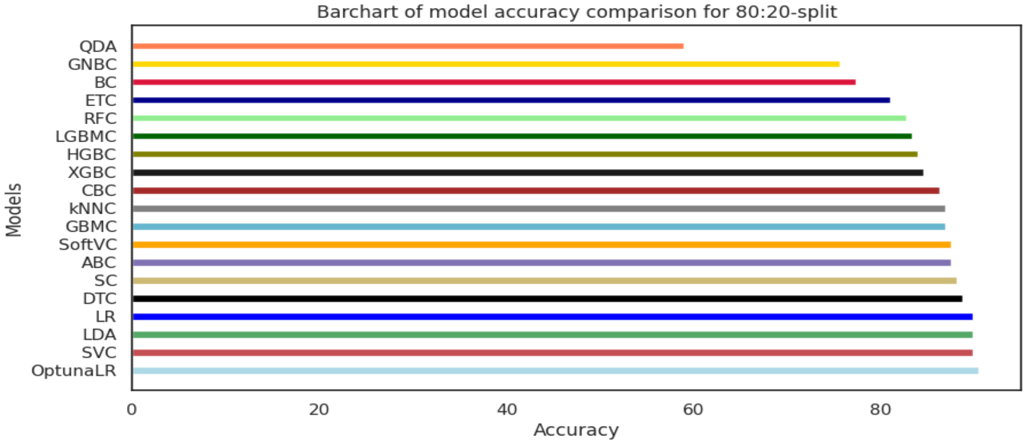


Figure 4: Bar chart of ML models accuracy for glioma grade prediction

4. Discussion

If machine learning is the engine, then data is the fuel. To get high performance, both ML model and data need to work in synergy. For the experimented dataset of brain glioma prediction, linear models namely logistic regression, linear discriminant analysis (LDA), support vector classifier (SVC) with linear kernel and Optuna-tuned logistic regression (OptunaLR) produced better results than other models. Even LDA and SVC produced same result (91.05% recall score and 89.88% accuracy) while OptunaLR outperformed all the models with 91.38% recall score and 90.48% accuracy. The outperformance of simple linear models is potentially due to intrinsic linearity characteristic of the data. However, outperformance of OptunaLR has two reasons- one is inherent linearity in data and another is best selection of LR hyperparameters using Optuna framework. Nonlinear model quadratic discriminant analysis (QDA) performed very poorly with 64.43% recall score and 58.93% accuracy. Even the ensemble methods like random forest (with recall score 82.41% and accuracy 82.74%) performed poorly. Also, all the boosting methods (e.g., LGBMC and XGBC) fall behind simple linear models in prediction performance. These results show that with more and quality data, ML models (at least of some types) can potentially best-fit (overcoming underfitting and overfitting). However, further research is necessary to achieve more accuracy with large and more datasets to find robust stable model.

The limitation of this study is that due to inadequate data availability, potential effectiveness of ML models for brain glioma may not be generalized with high degree of confidence. Also, medical sector is crucial. Therefore, practitioners in this field require reliable tools to rightly investigate and correctly diagnose the patients. However, works like this study are expected to inspire similar investigations to move forward current state of brain glioma prediction and other health diagnostic predictions which will contribute producing medical ML technologies.

5. Conclusion

This study mainly focused on predicting two primary brain tumors or glioma (LGG and GBM) using machine learning models on brain glioma dataset of 839 instances and 23 input features (20 molecular and 3 clinical). Of the 19 experimented ML models, OptunaLR (or logistic regression optimized by hyperparameter tuning with Optuna framework) was found to outperform other models with recall score of 91.38% and accuracy of 90.48%. Also, linear models were better than nonlinear and ensemble methods. The results reflect that using larger and appropriate features, medical diagnosis like brain glioma prediction can be done quickly and more accurately using ML approaches which can ease the tasks of doctors and health service providers. However, more researches need to be done using more and large datasets along with robust and stable ML models. Future works include application of ML or deep learning techniques in other health related sub-fields.

***Acknowledgements*.** We like to thank the referees for theirs suggestions leading to the improvements of the initial manuscript.

Author's Contributions: All authors are equally contributed.

Conflicts of interest. The authors declare no conflicts of interest.

REFERENCES

1. E. Tasci, Y. Zhuge, H. Kaur, K. Camphausen, and A. V. Krauze, Hierarchical voting-based feature selection and ensemble learning model scheme for glioma grading with clinical and molecular characteristics, *Int. J. Mol. Sci.*, 23(22) (2022) 14155.
2. L. Pasquini *et al.*, AI and high-grade glioma for diagnosis and outcome prediction: do all machine learning models perform equally well?, *Front. Oncol.*, 11 (2021) 601425.
3. Y. Wu, Y. Guo, J. Ma, Y. Sa, Q. Li, and N. Zhang, Research progress of gliomas in machine learning, *Cells*, 10(11) (2021) 3169.
4. K. R. Bhatele and S. S. Bhadauria, Machine learning application in glioma classification: review and comparison analysis, *Arch. Comput. Methods Eng.*, 29(1) (2022) 247–274.

5. R. C. Bahar *et al.*, Machine learning models for classifying high-and low-grade gliomas: a systematic review and quality of reporting analysis, *Front. Oncol.*, 12 (2022) 856231.
6. M. Gao, S. Huang, X. Pan, X. Liao, R. Yang, and J. Liu, Machine learning-based radiomics predicting tumor grades and expression of multiple pathologic biomarkers in gliomas, *Front. Oncol.*, 10 (2020) 1676.
7. Y. Kim, K. H. Kim, J. Park, H. I. Yoon and W. Sung, Prognosis prediction for glioblastoma multiforme patients using machine learning approaches: Development of the clinically applicable model, *Radiother. Oncol.*, 183 (2023) 109617.
8. R. Karri, Y.-P. P. Chen, and K. J. Drummond, Using machine learning to predict health-related quality of life outcomes in patients with low grade glioma, meningioma, and acoustic neuroma, *Plos One*, 17(5) (2022) e0267931.
9. H. Chen, C. Li, L. Zheng, W. Lu, Y. Li, and Q. Wei, A machine learning-based survival prediction model of high grade glioma by integration of clinical and dose-volume histogram parameters, *Cancer Med.*, 10(8) (2021) 2774–2786.
10. S. Gutta, J. Acharya, M. Shiroishi, D. Hwang, and K. Nayak, Improved glioma grading using deep convolutional neural networks, *Am. J. Neuroradiol.*, 42(2) (2021) 233–239,.
11. L. H. T. Lam *et al.*, Molecular subtype classification of low-grade gliomas using magnetic resonance imaging-based radiomics and machine learning, *NMR Biomed.*, 35(11) (2022) e4792.
12. R. Hodeify, N. Yu, M. Balasubramaniam, F. Godinez, Y. Liu, and O. Aboud, Metabolomic profiling and machine learning models for tumor classification in patients with recurrent IDH-Wild-type glioblastoma: a prospective study, *Cancers*, 16(22) (2024) 3856.
13. M. Pandey, P. Anoosha, D. Yesudhas, and M. M. Gromiha, Identification of potential driver mutations in glioblastoma using machine learning, *Brief. Bioinform.*, 23(6) (2022) 451.
14. L. Chen *et al.*, Machine learning-based nomogram for distinguishing between supratentorial extraventricular ependymoma and supratentorial glioblastoma, *Front. Oncol.*, 14 (2024).
15. Glioma Grading Clinical and Mutation Features, [Online]. Available: https://archive.ics.uci.edu/dataset/759/glioma+grading+clinical+and+mutation+features+dataset
16. P. Majumdar, *Mastering Classification Algorithms for Machine Learning: Learn how to Apply Classification Algorithms for Effective Machine Learning Solutions*. BPB Publications, 2023.
17. T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, Optuna: A Next-generation Hyperparameter Optimization Framework, in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.