Voprosy Onkologii = Problems in Oncology, 2025. T. 71, № 2
UDC 616.831-006:616-08-07
DOI: 10.37469/0507-3758-2025-71-2-OF-2208

© *Onur Dirican[1], Abbas Ali Husseini[2], Fatma Husseini[3], Serpil Oğuztüzün[4]*

Prognostic Modeling for Brain Cancer Patient Survival using Advanced Supervised Machine Learning Classification Approaches

[1]Istanbul Gelişim University, Department of Pathology Laboratory Techniques, Vocational School of Health Services, İstanbul, Türkiye
[2]Istanbul Gelişim University, Life Science and Biomedical Engineering Application and Research Center, İstanbul, Türkiye
[3]Department of computer engineering, Istanbul University Cerrahpaşa, Istanbul, Turkey
[4]Kırıkkale University, Faculty of Engineering and Natural Sciences, Department of Biology, Kırıkkale, Türkiye

© *О. Дирикан[1], А.А. Хусейни[2], Ф. Хусейни[3], С. Огузтузюн[4]*

Прогностическое моделирование выживаемости пациентов со злокачественными новообразованиями головного мозга с помощью передовых подходов к классификации на основе контролируемого машинного обучения

[1]Стамбульский университет Гелисим, кафедра лабораторных методов патологоанатомии, профессиональная школа здравоохранения, Стамбул, Турция
[2]Стамбульский университет Гелисим, Центр прикладных исследований в области наук о жизни и биомедицинской инженерии, Стамбул, Турция
[3]Стамбульский университет Серрахпаша, кафедра компьютерной инженерии, Стамбул, Турция
[4]Университет Кириккале, факультет инженерных и естественных наук, кафедра биологии, Кырыккале, Турция

Aim. Survival prediction in brain cancer is critical for treatment planning and patient outcomes. This study aimed to develop a prognostic model for brain cancer survival using supervised machine learning approaches. The model integrated demographic, clinical, immunohistochemical, and genomic data.

Materials and Methods. We retrospectively analyzed 149 patients with intracranial tumors who underwent surgery. Demographic and clinical data were systematically collected. Tumor and adjacent tissues underwent histopathological and immunohistochemical analysis for *GST-P*, *GST-T*, *GST-M*, *CYP1A1*, *CYP1B1*, *MDR*, and *p53* expression. Genomic DNA from tumors was analyzed for *GSTM1*, *GSTT1*, and *p53* genotypes. Models were developed using decision tree, Naïve Bayes, and SVM algorithms in Python. Models were compared based on accuracy, precision, sensitivity, and F-measure metrics.

Results. The overall postoperative survival rate was 65 %. Significant differences in protein expression were observed between cancerous and normal tissues for *GST-P*, *GST-T*, *GST-M*, *CYP1A1*, *CYP1B1*, *MDR*, and *p53*. *GST-M1* null genotype was associated with brain tumor development. The decision tree model achieved the highest accuracy (84 %) among models integrating demographic, clinical, immunohistochemical, and genetic data. Precision and sensitivity varied across models, with the decision tree showing acceptable performance.

Conclusion. Decision tree models are effective for predicting brain cancer survival, especially with limited datasets, using demographic, clinical, immunohistochemical, and genotypic variables.

Цель. Прогнозирование выживаемости при раке головного мозга играет ключевую роль в планировании лечения и определении прогноза для пациентов. Целью исследования было разработать модель прогнозирования выживаемости при раке головного мозга с использованием методов машинного обучения с контролем со стороны специалиста. Модель объединила демографические, клинические, иммуногистохимические и генетические данные.

Материалы и методы. Мы провели ретроспективный анализ данных 149 пациентов с внутричерепными опухолями, которые перенесли хирургическое вмешательство. Были систематически собраны демографические и клинические данные. Гистопатологический и иммуногистохимический анализы опухоли и прилегающих тканей проводились для определения экспрессии *GST-P*, *GST-T*, *GST-M*, *CYP1A1*, *CYP1B1*, *MDR* и *p53*. Анализ геномной ДНК опухолей выполнялся для определения генотипов *GSTM1*, *GSTT1* и *p53*. Модели были разработаны с использованием алгоритмов «дерево решений», Naïve Bayes и SVM на языке Python. Мы сравнили модели на основе показателей точности, прецизионности, чувствительности и F-измерения.

Результаты. Общая выживаемость после операции составила 65 %. Наблюдались значительные различия в экспрессии белков между раковыми и здоровыми тканями для *GST-P*, *GST-T*, *GST-M*, *CYP1A1*, *CYP1B1*, *MDR* и *p53*. Отсутствие генотипа GST-M1 было связано с развитием опухоли головного мозга. Модель дерева решений показала наивысшую точность (84 %) среди моделей, объединяющих демографические, клинические, иммуногистохимические и генетические данные. Точность и чувствительность моделей различались, при этом дерево решений продемонстрировало хорошие результаты.

Выводы. Модели дерева решений показали свою эффективность в прогнозировании выживаемости при раке мозга, особенно при работе с ограниченными данными,

Keywords: brain cancer; survival; machine learning; protein expression; genotype

For citation: Onur Dirican, Abbas Ali Husseini, Fatma Husseini, Serpil Oğuztüzün. Prognostic Modeling for Brain Cancer Patient Survival using Advanced Supervised Machine Learning Classification Approaches. *Voprosy Onkologii = Problems in Oncology*. 2025; 71(2): 344-352.-DOI: 10.37469/0507-3758-2025-71-2-OF-2208

✉ Contacts: Abbas Ali Husseini, ahusseini@gelisim.edu.tr
✉ Контакты: Аббас Али Хуссейни, ahusseini@gelisim.edu.tr

благодаря использованию демографических, клинических, иммуногистохимических и генотипических переменных.

Ключевые слова: рак головного мозга; выживаемость; машинное обучение; экспрессия белков; генотип

Introduction

Central Nervous System (CNS) tumors, originating in the brain or spinal cord tissues, pose a significant challenge as they disrupt the coordination and control of bodily functions governed by the CNS [1, 2]. These tumors can be primary, originating within the CNS, or secondary, resulting from metastasis of cancerous cells from other body parts [3]. Different types of CNS tumours can be classified based on the specific cell types affected, including astrocytoma, oligodendroglioma, ependymoma, medulloblastoma, meningioma, etc. [3].

Brain and other CNS tumors are among the most fatal cancers, causing significant morbidity and mortality. [4]. A single center pathology review of 21,622 cases showed that the proportion of CNS tumors was 63.8 % [5]. The global incidence of CNS cancer reached 330,000 cases in 2016, with an age-standardised rate of 4.63 per 100,000 person-years [6]. A recent systematic review and meta-analysis reported a primary CNS tumor prevalence of 3.6 per 100,000 individuals [7]. In the year 2019, a staggering total of 347,992 documented incidences of CNS malignancies were recorded on a global scale. Concurrently, during this temporal span, the lamentable toll of 246,253 lives succumbed to the devastating impact of CNS cancers across the world [8]. The global age-standardized mortality rate in 2019 was 3.05 per 100,000 population [8].

The etiology of CNS cancers, while not yet comprehensively elucidated, manifests as a multifactorial phenomenon, characterized by an intricate interplay of genetic, environmental, and lifestyle factors [9, 10]. Therefore, understanding the complicated interaction of influenced factors is crucial for both the prevention and management of brain cancer. Particularly survival prediction in brain cancer is crucial for guiding treatment decisions and improving patient outcomes. Several studies have highlighted the significance of machine learning techniques in predicting survival in patients with brain tumors. The use of machine learning techniques in predicting survival in patients with brain tumors has shown promising results, offering the potential to improve treatment planning and patient outcomes. These methods leverage medical imaging data and other relevant features [11–13] to provide accurate and robust predictions of patient survival, thereby contributing to the advancement of personalized medicine in brain cancer treatment. However, models based on cancerous cell genotypic and relevant protein expression profiles have not been developed yet. Thus, in the current study, a novel prognostic modeling for brain cancer patient survival using advanced supervised machine learning classification approaches based on the demographic, clinical, and protein expression profile of GST-P, GST-T, GST-M, CYP1A1, CYP1B1, MDR, and p53 genes. Moreover, the model accounts for genotypic variations in GST-M, GST-T, and p53, recognizing their potential roles in the pathogenesis of cancer.

Subjects and Methods

Study Design And Recruitment Of Participants

This modeling inquiry encompassed a cohort of patients afflicted with intracranial tumors who underwent operative procedures at a distinguished neurosurgery clinic within the temporal confines of 2017 and 2019. The archived clinical data pertaining to these patients were subjected to a retrospective evaluation, thereby facilitating an examination of their medical records comprehensively. Subjects who received a diagnosis of intracranial tumors, encompassing gliomas, metastases, meningiomas, and pituitary adenomas, and whose cases were accompanied by the availability of tumor tissue samples, were duly incorporated into the designated dataset.

Certain exclusion criteria were judiciously applied to ensure the internal validity and scientific rigour of the study. Patients with concurrent malignancies unrelated to cerebral pathology, primary tumors outside of intracranial locations, insufficient tumor tissue samples for meaningful analysis, severe co-morbidities such as end-stage renal disease or liver failure, Individuals who had previously received targeted therapeutic interventions explicitly targeting the GST isoenzymes CYP1A1, CYP1B1, MDR and p53 pathways, and cases where informed consent could not be obtained, were systematically excluded from the study.

The final sample comprised a total of 149 subjects with a mean age of 49.44 ± 8.09 years. The

subjects ranged in age from 6 to 83 years, and the gender distribution was 62 females to 87 males, reflecting a diversity of perspectives and making the participant pool representative of the wider population under study. The histological distribution was as follows: gliomas (38.3 %), metastases (27.5 %), meningiomas (22.1 %), and pituitary adenomas (12.1 %). This distribution reflects the expected prevalence of CNS tumors in a surgical cohort, with gliomas being the predominant malignant primary tumor and brain metastases being the most common secondary tumor type.

Data Collection

A comprehensive in-house checklist was used to systematically collect demographic and clinical data from the subjects. This tool facilitated the collection of essential information, including age, sex, smoking and alcohol consumption habits, previous exposure to radiotherapy and chemotherapy, number of previous surgical interventions, brain region involved, resection margins, lesion location and postoperative status. The data were collected retrospectively to ensure a rigorous examination of the subject's background. A comprehensive in-house checklist was used to systematically collect demographic and clinical data from the subjects. This tool facilitated the collection of essential information, including age, sex, smoking and alcohol consumption habits, previous exposure to radiotherapy and chemotherapy, number of previous surgical interventions, brain region involved, resection margins, lesion location and postoperative status. The data were collected retrospectively to ensure a rigorous examination of the subject's background.

Subsequently, tumor tissue removed from the surgical sites by experienced neurosurgeons using standardised procedures was subjected to histopathological examination. These tissue samples were embedded in paraffin for subsequent analysis. Immunohistochemistry was used to carefully evaluate the expression of GST-P, GST-T, GST-M, CYP1A1, CYP1B1, MDR and p53 proteins. Microscopic examination following immunostaining allowed the classification of protein expression into discrete categories, labelled 0, 1, 2 or 3, allowing a precise evaluation of the observed protein expression levels.

Histopathological Examination

Histopathological analysis of cerebral neoplastic tissues provides a valuable means of gaining a comprehensive insight into the cellular and architectural characteristics of the tumor, thereby aiding in accurate diagnosis, categorization and prognosis. This meticulous examination involves the microscopic examination of tissue samples obtained from cerebral tumors, allowing the identification of distinctive morphological features and molecular alterations that contribute to informed therapeutic decisions and prognostic predictions.

The process commenced with the procurement of neoplastic tissue during surgical resection procedures. These tissue samples were subsequently immersed in a 10 % buffered formalin solution, which served the purpose of preserving their structural integrity and preventing decay. Subsequent to fixation, the tissue underwent a series of procedural steps aimed at preparing thin sections suitable for microscopic evaluation. This entailed embedding the tissue in paraffin wax, which facilitated the creation of 4 μm thin sections. Immunohistochemical analysis was employed to uncover the expressions of GST-P, GST-T, GST-M, CYP1A1, CYP1B1, MDR, and p53 proteins in the obtained tissues following their preparation.

Immunohistochemical (IHC) Staining

For immunohistochemistry, the endogenous peroxidase activity was neutralized by immersing the sections in a solution of 1 % hydrogen peroxide (v/v) in methanol for a duration of 10 minutes at room temperature. Subsequently, the sections were rinsed in distilled water for 5 minutes, and the GST-P, GST-T, GST-M, CYP1A1, CYP1B1, MDR, and p53 protein retrieval was carried out in a household pressure cooker for 3 minutes, utilizing a 0.01 M citrate buffer (pH 6.0). After a further rinse with distilled water, the sections were transferred to a solution of 0.05 M Tris-HCl (pH 7.6) containing 0.15 M sodium chloride (TBS). To prevent non-specific background staining, sections were incubated with Super Block (streptavidin/HRP complex [SHP125]; ScyTek Laboratories, USA) for 10 minutes at room temperature. Subsequently, the sections were treated with primary antibodies, diluted at a ratio of 1:1,000 for anti- GSTP, anti-GST-T, anti-GST-M, anti-CYP1A1, anti-CYP1B1, anti-MDR and 1:50 for anti-p53, and left to incubate overnight at 4°C (Anti-GST-M obtained from Boster Biological, Pleasanton, CA, USA; anti-p53 acquired from Santa Cruz Biotechnology Inc., USA). After a 15-minute wash in TBS, the sections were incubated with a biotinylated link antibody, followed by SHP125, at room temperature. The visualization of peroxidase activity in the tissues was achieved using diaminobenzidine. The nuclei were lightly counterstained with hematoxylin, and subsequently, the sections were dehydrated and mounted. The evaluation of tissue nuclei from both the central region of the tumor and the invasive front was conducted separately for each sample, focusing on the presence of nuclear and cytoplasmic staining in tumor epithelial cells. The staining intensity was graded on a scale of 0 (no staining), 1 (poor staining), 2 (moderate staining), or 3 (strong staining).

Genotyping GSTM1 and GSTT1 and p53

Tumor tissues were subjected to extract genomic DNA for further genotypic analysis. Finally genomic DNA was accessible from 143 subjects

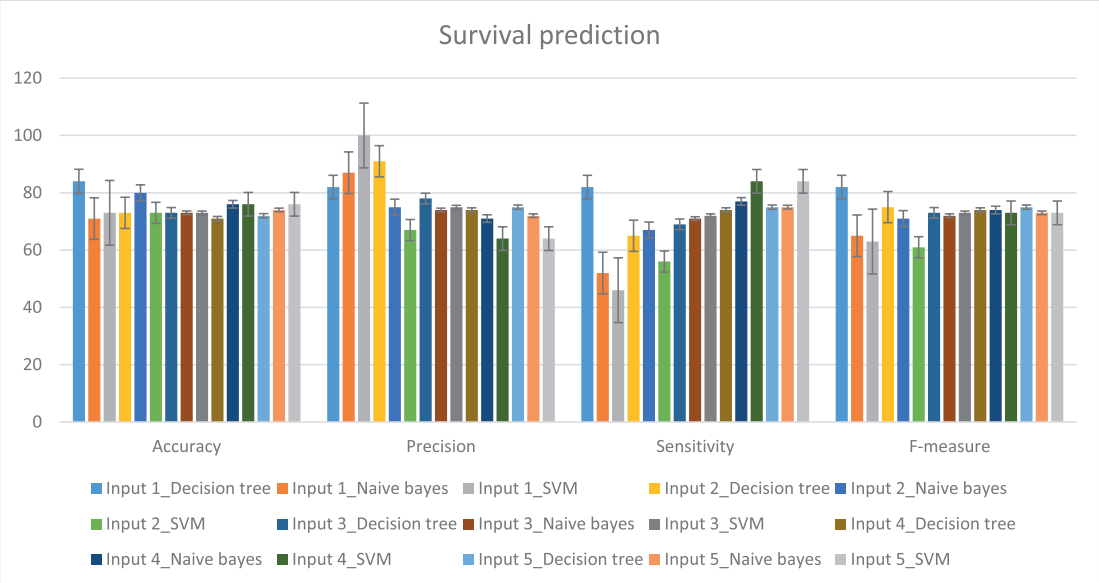

Fig. 1. The performance metrics of models adjusted to test survival prediction for brain cancer patients

to examine GSTM1 and GSTT1 via melting-curve analysis-based qPCR method to determine the deletion status of the GST-M1 and GST-T1 gene regions. In addition, genomic DNA from 120 brain tumor tissues and 47 normal tissues was used for sequence analysis to distinguish the SNP genotype of the p53 gene region exon4 codon72 (Figure 1), where the Arg/Pro change associated with the guanine/cytosine base conversion occurs, affecting the phenotypic expression of the tumor suppressor p53 gene due to a point mutation in this gene region [14,15].

Melting Curve Analysis by qPCR

In our study, qPCR method based on melting-curve analysis was applied to examine the deletion status of GST-M and GST-T gene regions. During this application, the Roche Lightcycler 480 qPCR system was used. Bio-Rad SSO Advanced Universal SYBR Green Supermix was used. GST-M and GST-T primer sequences used in the study; for GST-T1; 5-CAAGTCCCAGAGCACCTCACCTC-3' (NM_000853) Forward, 5-GTGTGCATCATTCT-CATTGTGGCTT-3' (NM_000853) Reverse, for GST-M1; 5'-TGCATTCGTTCATGTGACAG-TATTCT-3' (NM000561) Forward, 5'-GAGAG-GAGACCGGGCACTCA-3' (NM000561) Reverse [16]. In addition, this primer sequence was synthesized in the laboratory in specialized columns with CPG (Controlled Pore Glass) using the oligo-synthesis method and its purification was obtained from the C18 column and the sample collector with the reverse phase HPLC (Agilent) system. The controls were confirmed by the electrophoresis gel method in terms of both the graphs in the chromatographic system and the base sizes[17]. Mixing ratios prepared at the pre-PCR stage prior to the qPCR system; SYBR Green PCR Master Mix (Power) 5uL; forward primer 0.4uL; reverse primer 0.4uL; cDNA 1uL (100ng/uL); water (DN-Ase/RNAse free) 3.2uL; total volume was 10µL. In addition, the melting curve analysis program used in the qPCR system; 3 min at 98oC. 1 loop; each cycle of the denaturation process was 10 seconds at 95oC and 15 seconds at 60oC, 40 cycles of amplification and finally 0.3oC/second (ramp rate) from 65oC to 95oC.

SNPs Analysis by Sanger Sequence

It is known that the Arg/Pro change that occurs with the guanine/cytosine base conversion in the exon4 codon72 gene region, which is known to affect the phenotypic expression of the p53 tumor suppressor gene, is caused by the point mutation that occurs in this gene region. In order to clarify the situation of this point mutation in the patient group of our study, the position of the gene region that we amplified by polymerase chain reaction was analyzed for point mutations by sequence analysis method (Applied Biosystems-3130XL, P53 primers= NM000546.6) [14,15,18] .

Feature Selection and Classification Algorithms

Five categories of input features were selected for the development of a high-metric prediction model. The categories of input characteristics are summarised in Table 1. Demographic profile includes age, gender, smoking and alcohol consumption habits. Clinical profile includes previous exposure to radiotherapy and chemotherapy, number of previous surgical procedures, affected brain region, resection margins and lesion location. The protein expression levels of GST-P, GST-T, GST-M, CYP1A1, CYP1B1, MDR and p53 genes in tumor and adjacent healthy tissue were included in the protein expression profile. Gene dosage and deletion junction information related to GSTM1 and

Table 1. The input features categories for constructing models

Inputs	Demographic profile	Clinic profile	Protein expression	Genotypic variation
Input 1	√	√	√	√
Input 2	√	√	√	
Input 3	√	√		√
Input 4	√		√	√
Input 5		√	√	√

Table 2. Distribution of demographic and clinical characteristics of participants

Characteristics	Total n (%)
Demographic characteristics	
Gender	
Female	58 (40.6 %)
Male	85 (59.4 %)
Age (year)	
< 60	93 (65 %)
≥ 60	50 35 %)
Smoking	
Yes	43 (30 %)
No	100 (70 %)
Alcohol	
Yes	14 (9.8 %)
No	129 (90.2 %)
Clinical characteristics	
Radiotherapy	
Yes	55 (38.5 %)
No	88 (61.5 %)
Chemotherapy	
Yes	32 (22.3 %)
No	111 (77.6 %)
Lesion localization	
Frontal	40 (28 %)
Parietal	4 (2.8 %)
Cerebellar	15 (10.5 %)
Temporal	16 (11.2 %)
Other	68 (47.5 %)
Post-operation status	
Alive	93 (65 %)
Exitus	50 (35 %)

GSTT1 genes and Arg/Pro change that occurs with the guanine/cytosine base conversion located in the p53 exon4 codon72 gene region were defined as genotypic variation input.

Model Construction

Three models were constructed to predict the survival of brain tumor patients after surgery. Each category of input was then used as input to classifiers. A total of 15 models were constructed, each trained with randomly selected input parameters from 75 % of the patients. The models were then tested with the remaining 25 % of the patients' parameters. All algorithms were developed using Python version 3.8.0. The performance of the models was compared for model accuracy, precision, sensitivity and F-measure metrics.

Results

The descriptive data show that the mean age was 49.44 ± 8.09 years. Of the total number of patients, 85 were male (59.4 %) and 58 were female (40.6 %). Among the patients, 30 % had a history of smoking, while 70 % had never smoked. Only 9.8 % reported alcohol consumption, with the majority (90.2 %) abstaining. Regarding treatment, 38.4 % received radiotherapy and 22.3 % received chemotherapy. Lesion analysis showed the highest frequency in the frontal region (28 %), followed by the temporal region (11.2 %), cerebellar region (10.5 %) and parietal region (2.8 %). The remaining cases (47.5 %) showed various tumor localizations. The overall postoperative survival rate was 65 %. Table 2 provides a comprehensive overview of patient demographics, treatment history, and clinical profile. Table 3 presents the

Table 3. Distribution of protein expression levels between patients and healthy tissue

Protein marker	Patinets				Healthy				P value
	0	1	2	3	0	1	2	3	
GST-Pi	113	31	5	0	138	10	1	0	0.00038
CYP1A1	139	10	0	0	141	8	0	0	0.61006
CYP1B1	80	62	7	0	133	14	2	0	< 0.00001
GST-M	96	41	11	0	129	16	3	0	< 0.00001
MDR	62	65	22	0	139	9	1	0	< 0.00001
GST-TETA	61	42	25	21	137	8	3	1	< 0.00001
P53	117	32	0	0	145	1	0	0	< 0.00001

Table 4. GSTM1, GSTT1 and p53 genotyping in cancer tissue

Genotyping	Observed	Expected	p-value
GSTM1 present (1/1)	44	27	P < 0.001 χ^2 = 39.756
GSTM1 present (1/0)	33	70	
GSTM1 null (0/0)	66	46	
Total	143		
GSTT1 present (1/1)	92	91	p = 0.846 χ^2 = 0.335
GSTT1 present (1/0)	44	46	
GSTT1 null (0/0)	7	6	
Total	143		
P53 Wild type (Arg/Arg)	31	18	p = 0.281 χ^2 = 2.536
P53 Heterozygous (Arg/Pro)	70	23	
P53 Mutant type (Pro/Pro)	19	6	
Total	120		

frequency distribution of immunohistochemical staining categories for the protein markers examined in the study in both brain tumor and adjacent normal tissue.

The protein expression of GST-P, GST-T, GST-M, CYP1A1, CYP1B1, MDR and p53 genes was examined and showed significant differences in expression levels between cancerous and normal tissues ($p \leq 0.05$). GSTP expression in cancer tissue correlated with CYP1A1, CYP1B1, MDR and GST-T ($P < 0.05$), whereas no correlation was observed in normal tissue. CYP1A1 protein expression correlated with CYP1B1 and MDR in healthy tissue ($p<0.05$), but the correlation disappeared in tumor tissue. Similarly, CYP1B1 expression correlated with MDR, GSTT and p53 in healthy tissue ($p<0.05$), but the correlation was lost in cancer cells. GSTM expression correlated with p53 in normal tissue ($p=0.03$), but the correlation was lost in cancer cells. MDR and GSTT expression levels changed with p53 expression ($p=0.02$) in cancerous tissue. In our immunohistochemical analysis, we observed significant differences in the expression levels of GST-P, GST-T, GST-M, CYP1A1, CYP1B1, MDR and p53 between different tumor types.

Specifically: Gliomas showed high GST-P and p53 expression, consistent with their aggressive nature and the role of oxidative stress and apoptotic dysregulation in gliomagenesis.

Metastatic tumors showed significantly high CYP1A1 and MDR expression, suggesting enhanced detoxification and drug resistance mechanisms.

Meningiomas and pituitary adenomas, which are benign tumors, generally had lower expression levels of apoptotic and detoxification markers, but showed variable GST-M and GST-T expression, likely indicating tumor subtype-specific metabolic activity. These findings suggest that histological tumor variants have distinct molecular signatures and highlight the value of incorporating IHC markers into predictive modelling of survival outcomes.

Gene dosage and deletion junction information associated with patient GSTM1 and GSTT1 genotyping were described (Table 4). The GSTM1 genotype distribution was not in Hardy-Weinberg equilibrium in tumor tissue ($\chi^2=39.756$, $p<0.001$), indicating that the GSTM1 null genotype may be associated with brain tumor development. The genotype frequencies were 30.8 %, 23.1 % and 44.3 % for GSTM1 1/1, GSTM1 1/0 and GSTM1 0/0, respectively. However, the GSTT1 genotype distribution was in Hardy-Weinberg equilibrium ($\chi^2=0.335$, $p=0.846$) and the genotype frequencies were 64.3 %, 30.8 % and 4.9 % for GSTT1 1/1, GSTT1 1/0 and GSTT1 0/0, respectively. On the other hand, the most frequent genotype of p53 codon 72 polymorphism was Arg / Pro (heterozygous) followed by Arg /Arg (wild type) in both tumor and healthy tissues (Table 4). Furthermore, the genotype distribution of GSTM1, GSTT1 and P53 exon 4 codon 72 did not show significant differences between patients with respect to demographic and clinical characteristics.

The performance metrics of the models, including accuracy, precision and sensitivity, were evaluated (Fig. 1). The highest accuracy (84 %) was achieved when demographic, clinical, immunohistochemical and genetic information were used to train the decision tree model, followed by Naïve Bayes (80 %) when genotypic variables were eliminated. However, the accuracy of the SVM model was consistently ≤ 76 % in all cases. Despite these results, the overall deviation in accuracy was only 3.5 %. This deviation was 9.9 % and 11.3 % for precision and sensitivity, respectively.

Although there was high precision for Decision Tree (82 %), Naïve Bayes (87 %) and SVM (100 %) when Input 1 was used to train the algorithms, sensitivity was significantly low for Naïve Bayes and SVM. In general, there was significant variation in the performance of precision and sensitivity. However, the F-score was calculated to mitigate the heterogeneity of the precision and sensitivity parameters. Finally, the decision tree, when trained with input 1, showed an acceptable level of accuracy, precision and sensitivity.

Discussion

The study dataset includes demographic, clinical, immunohistochemical and genotypic variables. This information is important for developing a machine learning approach to survival prediction as it provides valuable context and characteristics of the patient population. This information can be used to identify patterns and trends in the data, which can then be used to create more accurate and effective machine learning models.

The most commonly used machine learning algorithms for creating predictive models for brain tumor patients include methods such as support vector machines (SVM), random forest, gradient boosting, decision trees and neural networks. These algorithms have been widely used to develop predictive models for the survival of brain cancer patients, particularly in cases of glioblastoma. Researchers have used these algorithms to analyse clinical, imaging and genomic data to predict how long brain tumor patients may survive [19,20]. In addition, machine learning and deep learning techniques have been used to predict overall survival in brain tumor patients using MRI images [13]. These algorithms have been shown to be effective in predicting various types of cancer, including brain cancer, and have the potential to improve diagnosis, prognosis, and quality of life for brain tumor patients [21]. In particular, the benefits of using machine learning approaches highlight the potential to improve prognostic modelling for brain cancer patients, ultimately contributing to more accurate predictions and personalised treatment approaches [21].

In the current endeavor, the decision tree classifier stands out as being particularly effective when all features are considered as input. It has demonstrated remarkable performance, achieving an F-measure of 82 % and an accuracy rate of 84 %. Accuracy, a widely used metric, reflects the proportion of correct predictions relative to total predictions and provides a concise summary of a model's effectiveness. However, relying solely on accuracy may prove inadequate for assessing the overall robustness of a model. To ascertain the model's resilience, it is imperative to extend the evaluation beyond accuracy alone. The true-positive rate emerges as a more dependable predictor in this context. Low precision and sensitivity can lead to an excess of false positives and false negatives, respectively. Precision and sensitivity, which measure the accuracy and completeness of a model, play a central role in decision making related to survival prediction and diagnosis. Models characterized by high precision and low sensitivity, or those with both low precision and low sensitivity, lack the necessary parameters to make informed decisions. Consequently, the F score, which represents the harmonic mean of precision and sensitivity, provides a comprehensive approach to addressing both concerns. This single score combines sensitivity and precision with a factor that controls their relative importance. This ensures a more nuanced and comprehensive assessment of a model's performance in predicting survival outcomes [22].

Decision trees are easy to understand and interpret, making them valuable for identifying the most important features for survival prediction in brain tumor patients. They can also handle non-linear relationships between features and outcomes, which can be beneficial when dealing with complex datasets related to brain cancer. Additionally, decision trees have been shown to achieve high accuracy and efficiency in predicting survival in patients with glioblastoma multiforme, a type of brain tumor [23,24]. On the other hand, SVM classification has been reported to achieve high accuracy in predicting cancer outcomes, including brain cancer, making it a valuable tool for survival prediction. It is effective in handling complex proteomic, genomic, and imaging data, which is crucial for accurate survival prediction in brain cancer patients. SVM has also been found to be the optimum classification method, with high precision, in predicting cancer diseases using machine learning approaches [25]. In some cases, simpler machine learning methods like Naive Bayes have been reported to substantially outperform more complex algorithms in cancer prediction and prognosis [26,27].

The primary constraint of this study was the sample size, consisting of 149 patients. Diseases such as brain cancer pose a perpetual challenge due to the difficulty in accessing cell-based data. Consequently, it is imperative for models to demonstrate effectiveness and efficiency in handling limited data. The decision tree, naive bayes and SVM algorithms are particularly useful for small sample sizes[28–30].

Conclusion

Decision trees offer high accuracy and acceptable precision and sensitivity in limited datasets

compared to SVM and naive Bayes for predicting survival in brain cancer when using demographic, clinical, immunohistochemical and genotypic variables. However, each algorithm has its own strengths and the choice of algorithm may depend on the specific characteristics of the dataset and the goals of the prognostic modelling task.

Conflict of interest
The authors declare no conflict of interest.

Funding
The work was performed without external funding.

Compliance with the rules of bioethics
The Ethics Committee of Istanbul Gelişim University approved this study (Ethics Code: 2024-06-71, dated 12.01.2024). Informed consent forms were signed by all participants before they were included in the study.

Authors' contributions
Onur Dirican collected the materials, data and/or performed analysis and critical review.
Abbas Ali Husseini suggested the idea for publication, developed the research design, collected data and/or performed analysis, reviewed publications on the topic of the article, drafted the manuscript, provided supervision and critical review.
Fatma Husseini developed the research design, collected data and/or performed analysis.
Serpil Oğuztüzün collected the materials, performed supervision and critical review.
All authors approved the final version of the article before submission for publication and agreed to take responsibility for all aspects of the work, including appropriate review and resolution of any issues relating to the accuracy or integrity of any part of the work.

ЛИТЕРАТУРА / REFERENCES

1. PDQ Adult Treatment Editorial Board. Adult Central Nervous System Tumors Treatment (PDQ®): Patient Version. 2023. In: PDQ Cancer Information Summaries. Bethesda (MD): National Cancer Institute (US); 2023.
2. Wang H., Zheng Q., Lu Z., et al. Role of the nervous system in cancers: a review. *Cell Death Discov*. 2021; 7: 76.-DOI: 10.1038/s41420-021-00450-y.
3. Louis D.N., Perry A., Wesseling P., et al. The 2021 WHO classification of tumors of the central nervous system: a summary. *Neuro Oncol*. 2021; 23: 1231-51.-DOI: 10.1093/neuonc/noab106.
4. Miller K.D., Ostrom Q.T., Kruchko C., et al. Brain and other central nervous system tumor statistics, 2021. *CA Cancer J Clin*. 2021; 71: 381-406.-DOI: 10.3322/caac.21693.
5. Chen L., Zou X., Wang Y., et al. Central nervous system tumors: a single center pathology review of 34,140 cases over 60 years. *BMC Clin Pathol.* 2013; 13: 14.-DOI: 10.1186/1472-6890-13-14.
6. Patel A.P., Fisher J.L., Nichols E., et al. Global, regional, and national burden of brain and other CNS cancer, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* 2019; 18: 376-93.-DOI: 10.1016/S1474-4422(18)30468-X.
7. Salari N., Ghasemi H., Fatahian R., et al. The global prevalence of primary central nervous system tumors: a systematic review and meta-analysis. *Eur J Med Res*. 2023; 28: 39.-DOI: 10.1186/s40001-023-01011-y.
8. Fan Y., Zhang X., Gao C., et al. Burden and trends of brain and central nervous system cancer from 1990 to 2019 at the global, regional, and country levels. *Archives of Public Health*. 2022; 80: 209.-DOI: 10.1186/s13690-022-00965-5.
9. Jaiswal J., Shastry A.H., Ramesh A., et al. Spectrum of primary intracranial tumors at a tertiary care neurological institute: A hospital-based brain tumor registry. *Neurol India*. 2016; 64: 494-501.-DOI: 10.4103/0028-3886.181535.
10. Naser R.K.A., Hassan A.A.K., Shabana A.M., Omar N.N. Role of magnetic resonance spectroscopy in grading of primary brain tumors. *The Egyptian Journal of Radiology and Nuclear Medicine.* 2016; 47: 577-84.-DOI: 10.1016/j.ejrnm.2016.03.011.
11. Pálsson S., Cerri S., Poulsen H.S., et al. Predicting survival of glioblastoma from automatic whole-brain and tumor segmentation of MR images. *Sci Rep*. 2022; 12: 19744.-DOI: 10.1038/s41598-022-19223-3.
12. Gaur L., Bhandari M., Razdan T., et al. Explanation-driven deep learning model for prediction of brain tumour status using MRI image data. *Front Genet*. 2022; 13.-DOI: 10.3389/fgene.2022.822666.
13. di Noia C., Grist J.T., Riemer F., et al. Predicting survival in patients with brain tumors: current state-of-the-art of ai methods applied to MRI. *Diagnostics*. 2022; 12: 2125.-DOI: 10.3390/diagnostics12092125.
14. Piña-Sánchez P., Hernández-Hernández D.M., Taja-Chayeb L., et al. Polymorphism in exon 4 of TP53 gene associated to HPV 16 and 18 in Mexican women with cervical cancer. *Medical Oncology*. 2011; 28: 1507-13.-DOI: 10.1007/s12032-010-9599-8.
15. Huszno J., Grzybowska E. TP53 mutations and SNPs as prognostic and predictive factors in patients with breast cancer (Review). *Oncol Lett.* 2018.-DOI: 10.3892/ol.2018.8627.
16. Girault I., Lidereau R., Bièche I. Trimodal GSTT1 and GSTM1 genotyping assay by real-time PCR. *Int J Biol Markers*. 2005; 20: 81-6.
17. Song L.-F., Deng Z.-H., Gong Z.-Y., et al. Large-scale de novo oligonucleotide synthesis for whole-genome synthesis and data storage: challenges and opportunities. *Front Bioeng Biotechnol.* 2021; 9.-DOI: 10.3389/fbioe.2021.689797.
18. Słomiński B., Skrzypkowska M., Ryba-Stanisławowska M., et al. Associations of TP53 codon 72 polymorphism with complications and comorbidities in patients with type 1 diabetes. *J Mol Med*. 2021; 99: 675-83.-DOI: 10.1007/s00109-020-02035-1.
19. Charlton C.E., Poon M.T.C., Brennan P.M., Fleuriot J.D. Development of prediction models for one-year brain tumour survival using machine learning: a comparison of accuracy and interpretability. *Comput Methods Programs Biomed*. 2023; 233: 107482.-DOI: 10.1016/j.cmpb.2023.107482.
20. Cruz J.A., Wishart D.S. Applications of machine learning in cancer prediction and prognosis. *Cancer Inform*. 2007; 2: 59-77.
21. Zhang B., Shi H., Wang H. Machine learning and ai in cancer prognosis, prediction, and treatment selection: a critical approach. *J Multidiscip Healthc.* 2023; 16: 1779-91.-DOI: 10.2147/JMDH.S410301.
22. Ghadiri F., Husseini A.A., Öztaş O. A machine-learning approach for nonalcoholic steatohepatitis susceptibility estimation. *Indian Journal of Gastroenterology*. 2022; 41: 475-82.-DOI: 10.1007/s12664-022-01263-2.

23. Payabvash S., Aboian M., Tihan T., Cha S. Machine learning decision tree models for differentiation of posterior fossa tumors using diffusion histogram analysis and structural MRI findings. *Front Oncol*. 2020; 10: 71.-DOI: 10.3389/fonc.2020.00071.
24. Nematollahi M., Jajroudi M., Arbabi F., et al. The Benefits of decision tree to predict survival in patients with glioblastoma multiforme with the use of clinical and imaging features. *Asian J Neurosurg*. 2018; 13: 697-702.-DOI:-DOI: 10.4103/ajns.AJNS_336_16.
25. Shaikh F.J., Rao D.S. Prediction of cancer disease using machine learning approach. *Mater Today Proc.* 2022; 50: 40-7.-DOI: 10.1016/j.matpr.2021.03.625.
26. Cruz J.A., Wishart D.S. Applications of machine learning in cancer prediction and prognosis. *Cancer Inform*. 2006; 2: 117693510600200.-DOI: 10.1177/117693510600200030.
27. Wu Y., Guo Y., Ma J., et al. Research progress of gliomas in machine learning. *Cells.* 2021; 10: 3169.-DOI: 10.3390/cells10113169.
28. Minghui M., Chuanfeng Z. Application of support vector machines to a small-sample prediction. *Advances in Petroleum Exploration and Development.* 2015; 10: 72-5.-DOI: 10.3968/7830.
29. Chen H., Hu S., Hua R., Zhao X. Improved naive Bayes classification algorithm for traffic risk management. *EURASIP J Adv Signal Process.* 2021; 2021: 30.-DOI: 10.1186/s13634-021-00742-6.
30. Morgan J., Dougherty R., Hilchie A., Carey B. Sample Size and Modeling Accuracy with Decision Tree Based Data Mining Tools. *Acad Inf Manag Sci J*. 2003; 6.

Received / 08.11.2024
Reviewed / 21.02.2024
Accepted for publication / 20.03.2025

Author's information / ORCID ID

Onur Dirican / ORCID ID: https://orcid.org/0000-0003-0511-6611.
Abbas Ali Husseini / ORCID ID: https://orcid.org/0000-0002-8861-7106.
Fatma Husseini / ORCID ID: https://orcid.org/0009-0008-7724-3736.
Serpil Oğuztüzün / ORCID ID: https://orcid.org/0000-0002-5892-3735.