Research Article
Pamukkale Medical Journal
doi:https://dx.doi.org/10.31362/patd.1528303

Application of machine learning techniques for survival prediction in pediatric malignant non-seminomatous germ cell testicular tumors: a SEER database study

Pediatrik malign non-seminamatöz germ hücreli testis tümörlerinde sağkalım tahmini için makine öğrenme tekniklerinin uygulanması: bir SEER veritabanı çalışması

İncinur Genişol, Batuhan Bakırarar

Posted date:05.08.2024 Acceptance date:03.10.2024

Abstract

Purpose: Childhood testicular cancers constitute 1-2% of all childhood tumors. According to the Surveillance, Epidemiology, and End Results (SEER) database, based on data from 2013 to 2019, the 5-year survival rate is 95.2%. The second most common type of testicular tumor is malignant non-seminomatous germ cell tumor. In recent years, various statistical techniques and extensive databases have been used to obtain information on disease prognosis and survival. In this study, we aimed to develop software using artificial intelligence and machine learning techniques to accurately predict the overall survival of patients with malignant nonseminomatous germ cell testicular tumors.

Materials and methods: Our study included data from 788 patients aged 0-18 diagnosed with malignant nonseminomatous germ cell testicular cancer between January 1975 and December 2019. The main hypothesis of the study was to provide overall survival (OS) in years from the date of diagnosis to the date of death or the last follow-up date for surviving patients. In addition to survival analysis, we also analyzed patient age at diagnosis, race, laterality, year of diagnosis, tumor histological type, T stage, N stage, M stage, tumor size, mortality, and follow-up duration.

Results: The OS was found to be 41.29±0.43 years. The median survival time was 43.21±0.62 years for patients <15 and 40.34±0.52 years for patients aged ≥15. We developed software that enabled the provision of patient-specific survival in addition to OS for all patients.

Conclusion: Recently, artificial intelligence techniques such as machine learning, have shown remarkable advancements compared to other statistical methods. As a result, in this study, we found that the survival rate in pediatric NSCGT was higher if the tumor was diagnosed after 2000, was less than 2 cm in size, and had a T1M0N0 stage yolk sac tumor. We created a 10-year survival prediction model with the results and thought that this model would contribute to the advancement of artificial intelligence studies in prognosis, recurrence and survival analysis.

Keywords: Non-seminomatous germ cell tumors, tescicular tumors, SEER, machine learning, survival.

Genisol I, Bakirarar B. Application of machine learning techniques for survival prediction in pediatric malignant non-seminomatous germ cell testicular tumors: a SEER database study. Pam Med J 2025;18:358-367.

Öz

Amaç: Çocukluk çağı testis kanserleri, tüm çocukluk çağı tümörlerinin %1-2'sini oluşturur. 2013-2019 yılları arasındaki verilere dayanan Surveillance, Epidemiology, and End Results (SEER) veritabanına göre, 5 yıllık sağkalım oranı %95,2 olarak görülmüştür. İkinci en yaygın testis tümörü türü ise malign non-seminomatöz germ hücreli tümördür. Son yıllarda, hastalığın prognozu ve sağkalımı hakkında bilgi elde etmek için çeşitli istatistiksel teknikler ve geniş veritabanları kullanılmaktadır. Bu çalışmada, malign non-seminomatous germ hücreli testis tümörü olan hastaların genel sağkalımını doğru bir şekilde tahmin etmek için yapay zeka ve makine öğrenimi tekniklerini kullanarak yazılım geliştirmeyi amaçladık.

Gereç ve yöntem: Çalışmamız, Ocak 1975 ile Aralık 2019 tarihleri arasında malign non-seminomatous germ hücreli testis kanseri tanısı konmuş 0-18 yaş arası 788 hastanın verilerini içermektedir. Çalışmanın temel hipotezi, tanı tarihinden ölüm tarihi veya sağ kalan hastalar için son takip tarihi itibarıyla genel sağkalımı yıllar olarak gösterebilmekti. Sağkalım yanı sıra, tanı anındaki hasta yaşı, ırk, tümörün yerleşim yeri, tanı yılı, tümör histolojik tipi, T evresi, N evresi, M evresi, tümör büyüklüğü, mortalite ve takip süresi gibi faktörleri de analiz ettik.

Bulgular: Genel sağkalım 41,29±0,43 yıl olarak bulundu. Medyan sağkalım süresi, 15 yaşından küçük hastalar için 43,21±0,62 yıl ve 15 yaş ve üzeri hastalar için 40,34±0,52 yıl olarak bulundu. Tüm hastalar için genel sağkalıma ek olarak, hastaya özgü sağkalım sağlayan bir yazılım geliştirdik.

Sonuç: Son zamanlarda, makine öğrenimi gibi yapay zeka teknikleri, diğer istatistiksel yöntemlere kıyasla dikkate değer ilerlemeler göstermiştir.

İncinur Genişol, M.D. Pediatric Surgeon, Dr. Behçet Uz Child Disease and Pediatric Surgery Training and Research Hospital, İzmir, Türkiye, e-mail: incinurgenisol@msn.com (https://orcid.org/0000-0003-3843-5942) (Corresponding Author)

Batuhan Bakırarar, Assoc. Prof. Ankara University Faculty of Medicine Department of Biostatistics, Ankara, Türkiye, e-mail: batuhan_bakirarar@hotmail.com (https://orcid.org/0000-0002-5662-8193)

Sonuç olarak bu çalışmamızda pediatrik NSCGT'de tümörün 2000 yılından sonra teşhis edilmesi, boyutunun 2 cm. den küçük olması, T1M0N0 evre yolk sak tümör olmasının hayatta kalma oranının daha yüksek olduğunu saptadık. Sonuçlarla 10 yıllık sağ kalma tahmin modeli oluşturduk ve bu modelin prognoz, nüks ve hayatta kalma analizinde yapay zeka çalışmalarının ilerlemesine katkı sağlayacagını düşündük.

Anahtar kelimeler: Non-seminomatöz germ hücreli tümör, testis tümörü, SEER, machine learning, sağkalım.

Genişol I, Bakırarar B. Pediatrik malign non-seminamatöz germ hücreli testis tümörlerinde sağkalım tahmini için makine öğrenme tekniklerinin uygulanması: bir SEER veritabanı çalışması. Pam Tıp Derg 2025;18:358-367.

Introduction

Testicular cancer has an incidence of 5.5/100,000 in males under the age of 19 years according to the Surveillance, Epidemiology, and End Results (SEER) database from 2015 to 2019 in the United States [1]. The prevalence of testicular tumors in childhood is estimated at 1-2% [2]. Testicular cancer accounts for 0.5% of all new cancer cases in the U.S. and ranks 25th in cancer incidence. According to the American Cancer Society's 2022 data, 9,190 new cases of testicular cancer were diagnosed in America. Testicular cancer accounted for 0.5% of all new cancer cases in the same year. Additionally, 470 men died from testicular cancer in 2022 [3]. According to the Global Cancer Observatory (GCO) 2020 global cancer data, 6,446 new cases were reported in males under the age of 19 years. Additionally, the 5-year overall survival (OS) rate was reported to be 95.2%, based on data from 2013 to 2019 [3, 4]. The American Joint Committee on Cancer tumor-node-metastasis (TNM) staging system is currently used to determine the prognosis of cancer patients. We believe that developing alternative techniques to the TNM staging system could be beneficial for the management and follow-up of complex diseases such as cancer. Knowing the survival outcomes of malignant diseases such as cancer, can assist clinicians in making appropriate treatment decisions and reducing patient anxiety regarding treatment choices [5].

The second most common malignant testicular tumor among testicular cancers is malignant nonseminomatous germ cell tumor (NSGCT) [3]. In this study, we aimed to develop an artificial intelligence program using machine learning techniques to predict the prognosis and survival of children with NSGCTs using the extensive SEER database. The program aims to assist clinicians in making predictions and decisions more easily.

Material and methods

Data collection and study population

These data, published by the National Cancer Institute, are a compilation of testicular cancer data from 18 population-based SEER cancer registries in the United States. The SEER database contains data from approximately 47.9% of the United States population [6]. The SEER program is used to summarize data from patients' medical records, and it is estimated to capture over 95% of all cancer cases in the surveillance areas. The duration of follow-up was determined in months using the date of diagnosis and whichever of the following events occurred first: 1) date of death, 2) date when the patient was last confirmed to be alive, or 3) the designated cutoff date of December 2019, which was employed in our analysis.

Because patient data were obtained from the SEER database with permission and did not include any personal patient information, the requirement for ethical approval was waived.

The SEER website is an American system that collects tumor data from patients without including their names. It is an open-access database that does not require ethical approval. However, researchers are provided with login credentials to access the data. Access permission has been obtained under the name of the author, Batuhan Bakırarar. We are providing the password and the printout via the email address as an attachment.

A total of 788 patients with malignant NSGCT, aged 0-18 years, diagnosed between January 1975 and December 2019, were included in the study using the ICD O3 site codes C62.1 and C62.9. Patients with in situ tumors were excluded from the study. Only patients who underwent surgery and had first and solitary malignant tumors were included in the study. Only patients

who were diagnosed by histopathological examination at the hospital were included in the study, whereas those diagnosed postmortem or through imaging techniques were excluded. Tumors at all locations of the testes (descending, undescending, and unknown) were included in this study. Patients with "unknown" or "missing" causes of death recorded in the database were excluded. Patients with bilateral testicular tumors or unknown tumor involvement were also excluded from the study. Because of the high amount of missing data in the variables related to surgical management, radiation therapy, and chemotherapy techniques, these variables were not included in the study.

The main hypothesis of the study was overall survival (OS), defined as the years from the date of diagnosis until the date of death or last follow-up with surviving patients (censored observations). In addition to survival, the variables selected for analysis included age at diagnosis, race/ethnicity, laterality, year of diagnosis, histological behavior, T, N, and M stages, and tumor size. The patients were divided into two groups based on whether they were diagnosed before or after 2000, with approximately half of the patients in each group. Testicular malignancies are most commonly observed between 15 and 40 years of age [7] and the incidence of extracranial germ cell tumors increases after 15 years of age [2] Therefore, we divided the patients into two groups: those aged <15 years and those aged ≥15 years.

Statistical analysis

The data analysis was performed using SPSS Statistics version 11.5 and R version 4.0.2. Descriptive statistics such as mean±standard deviation and median (minimum-maximum), were used for quantitative variables, while the number of patients (percentage) was used for qualitative variables. To examine the relationship between two qualitative variables, the chi-square test and Fisher's exact test were used. Survival analyses of qualitative variables were conducted using the Kaplan-Meier method, and significant differences between groups were determined using the log-rank test. Statistical significance was set at $p<0.05$. The classification methods used in this study include logistic regression, multilayer perceptron, support vector machine, bagging, and decision tables. The dataset was assessed using 10-fold cross-validation. The accuracy, F1 score, Matthew's Correlation Coefficient (MCC), Precision-Recall Curve (PRC Area), and ROC Area were employed as performance metrics for data mining. The RWeka package in R was used for analysis.

Results

Table 1 presents the demographic data and other descriptive patient and tumor characteristics of 788 patients aged 0-18 from the SEER database. In this study, 396 (50.3%) patients were diagnosed with testicular cancer before 2000, and 392 (49.7%) patients were diagnosed after 2000. Among the patients included in the study, 22.6% were < 15 years of age and 77.4% were ≥15 years of age. Among the patients, 89.0% were White, 2.2% were Black, and 8.8% had other ethnic backgrounds. Of these patients, 48.7% had tumors on the left side and 51.3% had tumors on the right side. Table 1 presents the tumor histologic behavior groups and the TNM stages. Tumor size, mortality status, and follow-up duration are provided in Table 1.

Table 2 presents the comparisons between mortality and other variables. The mortality rate was 3.4% among children younger than 15 years and 8.7% among children aged 15 years and older ($p=0.018$). Among children diagnosed before 2000, 11.1% died, whereas among those diagnosed in or after 2000, 3.8% died ($p<0.001$). The highest mortality rate was observed in children with choriocarcinoma as the histological behavior type, whereas the lowest mortality rate was observed in children with yolk sac tumors ($p<0.001$). In total, 2.4%, 2.7%, and 14.3% of children with stage T1, T2, and T3 died, respectively ($p=0.047$). The highest mortality rate was observed in children with stage N3 disease, whereas the lowest mortality rate was observed in children with stage N0 disease ($p=0.001$). The mortality rate was significantly higher in children at the M1 stage than in those at the M0 stage ($p<0.001$). The mortality rate increased as tumor size increased ($p<0.001$).

Table 1. Descriptive statistics

Variables		
Age (years), n (%)	<15	178 (22.6)
	≥15	670 (77.4)
Race, n (%)	White	697 (89.0)
	Black	17 (2.2)
	Other	69 (8.8)
Year of Diagnosis, n (%)	<2000	396 (50.3)
	≥2000	392 (49.7)
Laterality, n (%)	Left	384 (48.7)
	Right	404 (51.3)
Histologic Behavior, n (%)	Embryonal Carcinoma	169 (21.4)
	Yolk Sac Tumor	115 (14.6)
	Teratoma	65 (8.3)
	Teratocarcinoma	131 (16.7)
	Mixed Germ Cell Tumor	299 (37.9)
	Choriocarcinoma	9 (1.1)
T Stage, n (%)	T1	166 (55.3)
	T2	113 (37.7)
	T3	21 (7.0)
N Stage, n (%)	N0	221 (72.9)
	N1	38 (12.5)
	N2	28 (9.3)
	N3	16 (5.3)
M stage, n (%)	M0	256 (83.9)
	M1	49 (16.1)
Tumor Size (cm), n (%)	≤2	136 (22.7)
	2-5	299 (49.8)
	>5	165 (27.5)
Mortality, n (%)	Alive	729 (92.5)
	Dead	59 (7.5)
Follow-up Time (year)	Mean±SD	18.43±12.49
	Median (Min.-Max.)	17.50 (0.08-44.75)

SD: Standard Deviation, Min: Minimum, Max: Maximum

Table 2. The relationship between qualitative variables and mortality

Variables		Mortality				*p* value	Test value
		Alive		Dead			
		n	%	n	%		
Age (years)	<15	172	96.6	6	3.4	0.018[a]	5.625
	≥15	557	91.3	53	8.7		
Race	White	642	92.1	55	7.9	0.549[a]	1.199
	Black	16	94.1	1	5.9		
	Other	66	95.7	3	4.3		
Year of Diagnosis	<2000	352	88.9	44	11.1	<0.001[a]	15.092
	≥2000	377	96.2	15	3.8		
Laterality	Left	353	91.9	31	8.1	0.543[a]	0.371
	Right	376	93.1	28	6.9		
Histologic Behavior	Embryonal Carcinoma	159	94.1	10	5.9	<0.001[a]	88.185
	Yolk Sac Tumor	109	94.8	6	5.2		
	Teratoma	59	90.8	6	9.2		
	Teratocarcinoma	122	93.1	9	6.9		
	Mixed Germ Cell Tumor	279	93.3	20	6.7		
	Choriocarcinoma	1	11.1	8	88.9		
T Stage	T1	162	97.6	4	2.4	0.047[b]	-
	T2	110	97.3	3	2.7		
	T3	18	85.7	3	14.3		
N Stage	N0	217	98.2	4	1.8	0.001[b]	-
	N1	35	92.1	3	7.9		
	N2	27	96.4	1	3.6		
	N3	12	75.0	4	25.0		
M Stage	M0	254	99.2	2	0.8	<0.001[b]	-
	M1	39	79.6	10	20.4		
Tumor Size (cm)	≤2	133	97.8	3	2.2	<0.001[a]	20.381
	2-5	291	97.3	8	2.7		
	>5	146	88.5	19	11.5		

a: Chi-square test, b: Fisher-exact test

Table 3 presents the survival analysis results. The mean overall survival (OS) for the patients was 41.29±0.43 years. In addition, age, histologic behavior, T, N, and M stages, and tumor size were significant risk factors (p<0.05). The mean survival time was found to be 43.21±0.62 years for patients aged <15 and 40.34±0.52 years for patients aged ≥15 (p=0.017). The highest mean survival time was found in patients with histologic behavior type yolk sac tumors, whereas the lowest mean survival time was observed in patients with histologic behavior type choriocarcinoma (p<0.001).

The mean survival times for patients with T1, T2, and T3 stages were found to be 15.54±0.19, 15.54±0.22, and 12.00±0.98 years, respectively (p=0.009). When evaluated according to the N stage, the lowest mean survival was observed in patients with N3 stage, whereas in terms of M stage, the lowest mean survival was observed in patients with M1 stage (p<0.001 and p<0.001, respectively). The mean survival for children with tumor size <2 cm was found to be 36.1±0.47 years, while this duration was 35.17±0.35 years for children with tumor size between 2-5 cm and 32.45±0.92 years for children with tumor size >5 cm (p<0.001) (Table 3).

Table 3. Kaplan-Meier analysis results for factors affecting mortality

Variables		Survival					*p* value	Test value
		1 year (%)	5 year (%)	10 year (%)	Survival Time			
					Mean±SE	Median±SE		
Overall Survival (years)		96.9	93.5	93.2	41.29±0.43	-	-	-
Age (years)	<15	98.8	97.0	96.3	43.21±0.62	-	0.017*	5.664[a]
	≥15	97.2	92.5	92.3	40.34±0.52	-		
Race	White	96.8	93.3	92.9	40.72±0.47	-	0.569	1.128[a]
	Black	94.1	94.1	94.1	42.15±2.52	-		
Laterality	Other	98.6	95.3	95.3	40.39±1.05	-	0.498	0.459[a]
	Left	96.0	93.0	92.3	40.12±0.63	-		
	Right	97.7	94.0	94.0	41.55±0.58	-		
Histologic Behavior	Embryonal Carcinoma	97.0	95.7	95.7	41.46±0.80	-	<0.001*	243.187[a]
	Yolk Sac Tumor	98.2	95.4	95.4	41.54±0.91	-		
	Teratoma	95.4	92.0	92.0	40.27±1.76	-		
	Teratocarcinoma	96.9	93.1	93.1	40.44±0.93	-		
	Mixed Germ Cell Tumor	99.0	94.6	93.7	39.31±0.70	-		
	Choriocarcinoma	22.2	11.1	11.1	3.77±2.95	0.58±0.12		
T Stage	T1	99.3	98.0	97.1	15.54±0.19	-	0.009*	9.415[a]
	T2	-	97.8	97.8	15.54±0.22	-		
	T3	89.7	84.4	84.4	12.00±0.98	-		
N Stage	N0	99.1	98.0	98.0	15.46±0.14	-	<0.001*	27.812[a]
	N1	97.1	94.0	94.0	14.90±0.61	-		
	N2	96.3	96.3	96.3	14.96±0.53	-		
	N3	93.3	76.4	61.1	10.03±1.64	-		
M Stage	M0	-	99.6	99.6	15.80±0.08	-	<0.001*	41.253[a]
	M1	89.4	80.1	77.4	12.81±0.87	-		
Tumor Size (cm)	<2	98.5	97.7	97.7	36.10±0.47	-	<0.001*	19.293[a]
	2-5	99.3	97.8	97.4	35.17±0.35	-		
	>5	96.9	91.0	90.1	32.45±0.92	-		

SE: Standard error, a: Log Rank test

The gain ratio attribute (GAR) method was used for variable selection to determine variable importance. This method examines the importance of variables and their contributions to the dataset. Variables deemed insignificant using the gain ratio method and considered to lack clinical relevance were excluded from the analysis. As a result, eight variables (seven independent and one dependent) were retained in the dataset. These variables included histological behavior, M stage, age, tumor size, N stage, race, T stage, and status. The percentage of variable importance with respect to the dependent variable, mortality, is shown in Figure 1.

The performance metrics for the machine learning methods used in the prediction model are presented in Table 4. After evaluating the performance metrics, we determined that logistic regression provided the best predictive model. After logistic regression, the following methods were used: support vector machine, bagging, decision table, and multilayer perceptron.

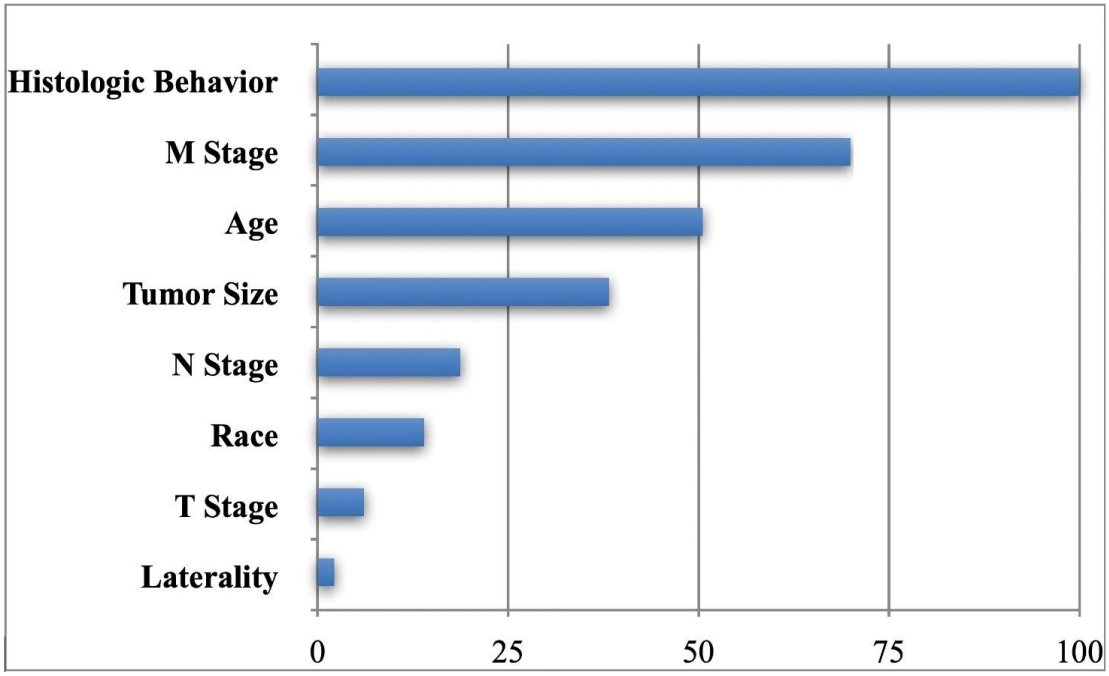

Figure 1. Variable importance for mortality

Table 4. Using machine learning methods for mortality prediction

Methods	Performance Criteria				
	Accuracy	F-measure	MCC	PRC Area	ROC Area
Logistic Regression	0.934	0.911	0.332	0.892	0.647
Multilayer Perceptron	0.926	0.906	0.243	0.755	0.919
Support Vector Machine	0.934	0.911	0.332	0.878	0.567
Bagging	0.934	0.911	0.332	0.862	0.524
Decision Table	0.931	0.907	0.287	0.851	0.420

MCC: Matthew's Correlation Coefficient, PRC Area: Precision-Recall Curve Area, ROC Area: Receiver Operating Characteristic Curve Area

Based on the best-performing machine learning method from Table 4, logistic regression, a decision support system software was developed. The variables included in the decision support system were determined based on various important factors, including histological behavior, M stage, age, tumor size, N stage, race, and T stage. The outputs of the decision support system that calculates survival based on these variables are shown in Figure 2. An overall view of the decision support system is shown in Figure 2(a), a screenshot of the survival prediction for a sample patient input is presented in Figure 2(b), and a screenshot of the mortality prediction for a sample patient input is displayed in Figure 2(c).

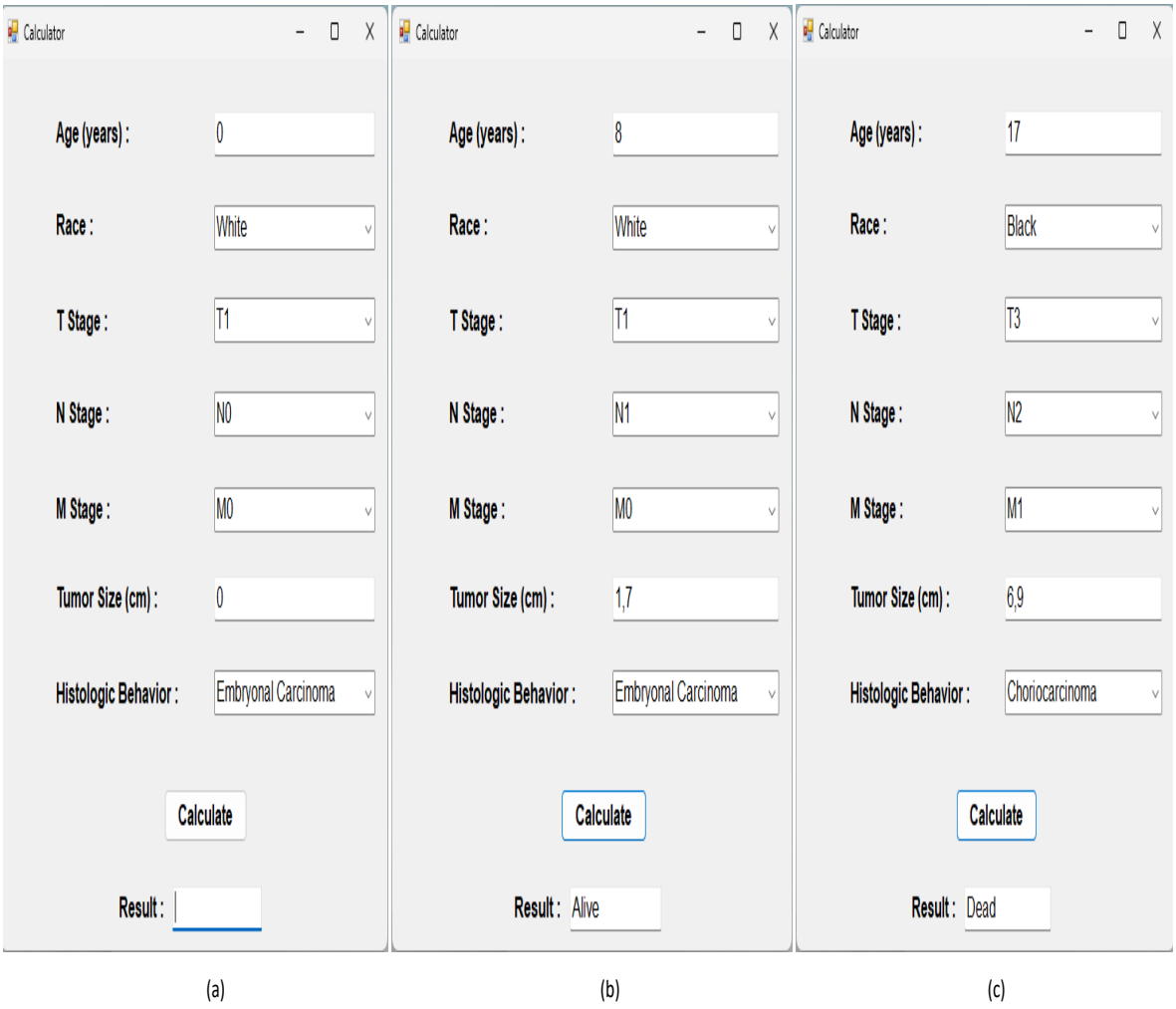

Figure 2. Output screenshots of the decision support system

(Link to download the application: https://drive.google.com/file/d/125nYPDH5b2sMovHBNjOxe3EeJwQKxd3u/view?usp=sharing)

Discussion

In this study, we used a machine learning technique to determine the survival rates of a nonseminomatous germ cell subgroup of testicular cancers, which are rare in pediatric patients. We developed a 10-year survival prediction model for pediatric patients with NSGCT. This model is presented as a software solution at the end of the study. We initially examined the demographic data of cases available in the SEER database. Only pediatric patients were included in this study. Therefore, our study is a novel and comprehensive study that specifically included a pediatric age group (0-18 years). Studies have consistently shown that the age group of 15-39 years represents the adolescent and young adult population and is most commonly affected by testicular cancer [8, 9]. Similarly, our study supports these findings, as patients aged 15 years accounted for 77.4% of the study population.

Various methods are available to predict the survival and prognosis of cancer patients [10-13]. In the study conducted by Srivastava et al. [11], for survival analysis of non-seminomatous mediastinal and testicular germ cell cancer, the researchers utilized the Cox proportional hazard model and Kaplan-Meier curves based on the SEER dataset. According to the study, the 5-year cancer-specific survival rates were 88.23% for individuals aged <19 years, 82.22% for those aged 19-30, and 74.13% for those aged >30 [11]. In our study, we conducted a survival analysis of patients aged <15 and ≥15 years. The results indicated that the 1-year, 5-year, and 10-year survival rates for patients aged < 15 years were 98.8%, 97.0%, and 96.3%, respectively. For patients aged 15 years,

the corresponding survival rates were 97.2%, 92.5%, and 92.3%, respectively. Moreover, the 1-year, 5-year, and 10-year cancer-specific survival rates for all cases were 96.9%, 93.5%, and 93.2%, respectively. The higher observed survival rates may be attributed to the focus of our study on testicular germ cell cancer. In a study by Wu et al. [14] that utilized the SEER database, the 3-year and 5-year survival rates for mediastinal malignant germ cell cancer were 63.1% and 61.2%, respectively.

There was no significant difference in tumor laterality among patients. In our literature review, we were unable to find specific data regarding the laterality of NSGCTs in pediatric patients. However, a study conducted by Berney et al. [15] reported that malignant NSGCTs were more frequently observed on the right side in individuals aged 60 years.

In our study, malignant NSGCTs were observed more frequently in Caucasians. However, no significant difference was observed in mortality between White and Black individuals. In a study conducted by Li et al. [16] on adult patients, a higher mortality rate was observed among Black individuals. This study suggests that the underlying reasons for this disparity could be attributed to factors such as hormones, dietary habits, cryptorchidism, and familial risk factors. However, the exact cause for this difference remains unclear [16].

According to the data from the SEER database, we identified mixed germ cell tumors as the most common subtype of malignant NSGCTs. According to the literature, yolk sac tumors are the most common subtype of NSGCT in children under 2 years of age (approximately 10-44% of cases) [17, 18]. Mixed germ cell tumors are the most frequently observed subtype in adolescents. We believe that the reason for the most frequent occurrence of mixed germ cell tumors in this study was that 77.4% of tumors were observed in adolescents. In our study, the highest mortality rate was observed in cases with the histological behavior of choriocarcinoma, whereas the lowest mortality rate was observed in cases of yolk sac tumors. Curto et al. [19] reported a 100% survival rate in yolk sac tumors. Our findings are consistent with those in the literature.

In our study, we observed disease-related deaths in 11.1% of patients diagnosed before 2000, whereas the rate decreased to 3% in patients diagnosed in 2000 and thereafter. We believe that advancements in medical technology as well as improvements in radiotherapy and chemotherapy may have contributed to this decrease. However, because we did not have access to a database regarding chemotherapy administered before or after surgery, we could not make a definitive conclusion regarding its impact on the observed outcomes.

In our study, the average lifespan for children with tumor sizes smaller than 2 cm was found to be 36.10±0.47 years, while it was 35.17±0.35 years for tumor sizes between 2-5 cm and 32.45±0.92 years for tumor sizes larger than 5 cm. In the study conducted by Song et al. [20], a correlation between larger radiological dimensions of tumors and malignancy was observed. Similarly, our study supports the finding that as the tumor size increases, the mortality risk also increases.

In conclusion, this study evaluated the factors influencing survival in pediatric NSCGT, which is a rare urological cancer. We examined the relationship between survival and factors such as patient age, race, histological tumor type, TNM stage, year of diagnosis, and tumor size. We found that if the tumor was diagnosed after 2000, had a size smaller than 2 cm, was diagnosed before the age of 15 years, and was a yolk sac tumor with T1N0M0 staging, the survival rate would be higher. We believe that our 10-year survival prediction model and the software created for this model will contribute to the advancement of artificial intelligence studies in prognosis, recurrence, and survival analysis.

Funding: None.

Author contributions: İ.G. and B.B. constructed the main idea and hypothesis of the study. İ.G. and B.B. developed the theory and arranged/edited the material and method section. İ.G. has done the evaluation of the data in the results section. Discussion section of the article was written by İ.G., İ.G. and B.B. reviewed, corrected and approved. In addition, all authors discussed the entire study and approved the final version.

Conflict of interest: No conflict of interest was declared by the authors.

References

1. SEER*Explorer Application. Available at: https://seer.cancer.gov/statistics-network/explorer/application.html?site=67&data_type=1&graph_type=3&compareBy=race&chk_race_1=1&rate_type=2&hdn_sex=2&advopt_precision=1&advopt_show_ci=on&hdn_view=1#resultsRegion1. Accessed April 2, 2023
2. Ahmed HU, Arya M, Muneer A, Mushtaq I, Sebire NJ. Testicular and paratesticular tumours in the prepubertal population. *Lancet Oncol.* 2010;11(5):476-483. doi:10.1016/S1470-2045(10)70012-7
3. Facts About Testicular Cancer | Testicular Cancer Statistics. Available at: https://www.cancer.org/cancer/testicular-cancer/about/key-statistics.html. Accessed March 26, 2023
4. Cancer today. Available at: http://gco.iarc.fr/today/home. Accessed April 2, 2023
5. Turabian J. Prognosis-based medicine-The importance of psychosocial factors: Conceptualization from a case of acute pericarditis. *Trends Gen Pract.* 2018;1(1):1-5. doi:10.15761/TGP.1000101
6. SEER Incidence Database - SEER Data & Software. SEER. Available at: https://seer.cancer.gov/data/index.html. Accessed April 17, 2023
7. McGlynn KA, Trabert B. Adolescent and adult risk factors for testicular cancer. *Nat Rev Urol.* 2012;9(6):339-349. doi:10.1038/nrurol.2012.61
8. Cancer statistics for adolescents and young adults, 2020 - Miller - 2020 - CA: A Cancer Journal for Clinicians - Wiley Online Library. Available at: https://acsjournals.onlinelibrary.wiley.com/doi/10.3322/caac.21637. Accessed April 20, 2023.
9. Key Statistics for Childhood Cancers. Available at: https://www.cancer.org/cancer/cancer-in-children/key-statistics.html. Accessed April 20, 2023
10. Statistical Methods in the Cancer Surveillance Research Program (SRP).
11. Srivastava A, Patel HV, Koehne E, et al. Survival of nonseminomatous germ cell tumors in pediatric patients and young adults - A stage group stratified analysis. *Urol Oncol.* 2022;40(4):169.e1-169.e12. doi:10.1016/j.urolonc.2021.12.012
12. Mao W, Wu J, Kong Q, Li J, Xu B, Chen M. Development and validation of prognostic nomogram for germ cell testicular cancer patients. *Aging.* 2020;12(21):22095-22111. doi:10.18632/aging.104063
13. Li M, Wang J, Li J, et al. Develop and validate nomogram to predict cancer-specific survival for patients with testicular yolk sac tumors. Front Public Health. 2022;10. Available at: https://www.frontiersin.org/articles/10.3389/fpubh.2022.1038502. Accessed April 10, 2023.
14. Wu P, Yang Y, Yu Z, Zhao L, Feng S. Clinical Features and Survival Outcomes in Children and Adolescents With Malignant Mediastinal Germ Cell Tumors Based on Surveillance, Epidemiology, and End Results Database Analysis. *J Surg Res.* 2023;288:362-371. doi:10.1016/j.jss.2023.03.018
15. Berney DM, Warren AY, Verma M, et al. Malignant germ cell tumours in the elderly: a histopathological review of 50 cases in men aged 60 years or over. *Mod Pathol Off J U S Can Acad Pathol Inc.* 2008;21(1):54-59. doi:10.1038/modpathol.3800978
16. Li Y, Lu Q, Wang Y, Ma S. Racial differences in testicular cancer in the United States: descriptive epidemiology. *BMC Cancer.* 2020;20(1):284. doi:10.1186/s12885-020-06789-2
17. Caballero Mora FJ, Muñoz Calvo MT, García Ros M, et al. [Testicular and paratesticular tumors during childhood and adolescence]. *An Pediatr Barc Spain 2003.* 2013;78(1):6-13. doi:10.1016/j.anpedi.2012.05.018
18. Geng R, Zheng Z, Lin Y, et al. Clinical characteristics and prognostic factors of male yolk sac tumor: a Surveillance, Epidemiology, and End Results program study. *World J Urol.* 2021;39(4):1211-1217. doi:10.1007/s00345-020-03311-y
19. Curto ML, Lumia F, Alaggio R, et al. Malignant germ cell tumors in childhood: Results of the first Italian cooperative study “TCG 91”. *Med Pediatr Oncol.* 2003;41(5):417-425. doi:10.1002/mpo.10324
20. Song G, Xiong GY, Fan Y, et al. The role of tumor size, ultrasonographic findings, and serum tumor markers in predicting the likelihood of malignant testicular histology. *Asian J Androl.* 2019;21(2):196-200. doi:10.4103/aja.aja_119_18