

## RESEARCH ARTICLE



# Prostate Cancer Survival Prediction and Treatment Recommendation: A Machine Learning Perspective

 OPEN ACCESS

Received: 14-12-2023

Accepted: 08-02-2024

Published: 07-03-2024

Garvita Ahuja<sup>1</sup>, Priyal<sup>1</sup>, Ishleen Kaur<sup>2\*</sup>, Deepali Virmani<sup>1</sup><sup>1</sup> Vivekananda Institute of Professional Studies- Technical Campus, 110034, Delhi, India<sup>2</sup> Sri Guru Tegh Bahadur Khalsa College, University of Delhi, 110009, Delhi, India

## Abstract

**Citation:** Ahuja G, Priyal, Kaur I, Virmani D (2024) Prostate Cancer Survival Prediction and Treatment Recommendation: A Machine Learning Perspective. Indian Journal of Science and Technology 17(11): 1097-1106. <https://doi.org/10.17485/IJST/V17i11.3157>

\* **Corresponding author.**

[ishleen@sgtbkhalsa.du.ac.in](mailto:ishleen@sgtbkhalsa.du.ac.in)

**Funding:** None

**Competing Interests:** None

**Copyright:** © 2024 Ahuja et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment ([iSee](https://www.isee.org/))

**ISSN**

Print: 0974-6846

Electronic: 0974-5645

**Objective:** Prostate cancer, a formidable life-threatening ailment predominantly affecting males, ranks as the third most prevalent global tumor. Its formidable nature arises from the persistent challenges encountered in early detection, often leading to delayed diagnoses and more advanced disease stages. The primary objective is to harness the power of advanced machine learning techniques for the prediction of patient survivability in prostate cancer cases. Furthermore, the study aims to identify a set of treatments pivotal for ensuring positive survival rates. **Methods:** This investigation leverages a comprehensive retrospective dataset comprising 410 cases of prostate cancer, collected from a Cancer Centre in New Delhi. This dataset encompasses vital clinical and treatment attributes. Models, including Artificial Neural Networks (ANN), Adaboost, Random Forest, etc., are thoroughly evaluated. In addition, the Generalized Sequential Pattern (GSP) algorithm is utilized to scrutinize the treatment attributes, thereby uncovering frequent patterns and their correlation with survival rates. **Findings:** The ANN model emerges as the most promising, exhibiting an impressive 84.14% accuracy. The findings stemming from these classification techniques, as well as the insights garnered through sequential mining, underscore the pivotal role of machine learning in the prognostication of prostate cancer. This advancement holds the potential to transform precision medicine and enhance patient care strategies on a global scale. **Novelty:** The study used clinical dataset to predict the survival of cancer patients using neural networks. GSP algorithm is also modified to uncover frequent treatment patterns in patients.

**Keywords:** Artificial Neural Network; Cancer; Machine learning; Sequence mining; Survival analysis

## 1 Introduction

Prostate cancer is a prevalent form of cancer that primarily affects men, ranking as the third most common type of tumor globally. Typically, prostate cancer cases are diagnosed around the age of 66. The early stages of prostate cancer often present with no noticeable symptoms, leading to delayed diagnosis until the disease has

progressed. In India, there has been a noticeable rise in prostate cancer incidence rates, although they are generally lower compared to Western countries. However, what is concerning is that around 85% of prostate cancer cases in India are diagnosed at stage 4, indicating metastatic cancer, which tends to have a less favorable prognosis<sup>(1)</sup>. Prostate cancer ranks among the top five prevalent cancer sites affecting Indian men. Existing techniques like digital rectal exams (DRE) and prostate-specific antigen (PSA) screening exhibit limitation in terms of accuracy and efficiency. Biopsies, though considered definitive, are invasive and can miss over 30% of malignant tumors. Ongoing advancements in diagnostic technologies aim to enhance precision and mitigate these drawbacks<sup>(2)</sup>.

In the medical field, the choice of treatments and medications plays a critical role in determining a patient's survival. The selection of appropriate treatment tailored to an individual's specific circumstances can greatly impact their chances of improvement. Clinicians have explored various medication combinations to analyse survival rates in prostate cancer patients with metastasis<sup>(3)</sup>. The primary aim of this study is to predict the survival outcomes of individuals with metastatic prostate cancer. Additionally, it will analyse the treatment sequences provided to these patients, along with other relevant clinical characteristics. Hormonotherapy treatments are commonly provided to metastatic prostate cancer patients to extend their survival. However, the order or sequence in which subsequent treatments are administered is often overlooked. The sequence of treatments holds significant importance for patients, as certain treatment sequences may yield lower mortality rates for specific individuals compared to alternative approaches.

Recent strides in machine learning and deep learning have significantly contributed to refining prostate cancer diagnosis and survival analysis. However, the existing literature reveals gaps in understanding the sequential dynamics of treatments and their influence on survival outcomes. This study aims to bridge these gaps by applying cutting-edge techniques to unravel patterns in treatment sequences and their correlation with survival rates in metastatic prostate cancer patients. In their study, Koo et al. developed a LSTM model to predict prostate cancer survivability and created an online decision-making system using a dataset of 7627 patients<sup>(4)</sup>. Similarly, Thongpim et al. also utilized Cox Regression for their analysis<sup>(5)</sup>. In the study conducted by Doja et al., an array of Machine Learning techniques was employed, leading to the creation of two predictive models - one for the complete dataset and another for separate age groups<sup>(6)</sup>. Using the SEER-Medicare data, Nezhad et al. applied the Deep Active Survival Analysis algorithm to investigate prostate cancer among African-American and white patients, showcasing how their model outperforms traditional approaches in survival analysis<sup>(7)</sup>. Wen et al. conducted a comparative analysis, evaluating the performance of ANN in conjunction with KNN, DT, and SVM to classify patients' survival time using the SEER prostate cancer database<sup>(8)</sup>. Utilizing Shapley additive explanation (SHAP) values, the study by Li et al. applied a gradient-boosted tree model to predict mortality risk in nonmetastatic prostate cancer patients. Analyzing a dataset of 372,808 patients, significant interaction effects between variables, particularly in Gleason  $\geq 8$  disease with PPC  $\geq 50\%$ , were identified, emphasizing the potential of SHAP values to enhance risk stratification in prostate cancer<sup>(9)</sup>. It is imperative to look at the literature of the sequential data mining used in prostate cancer analysis and other healthcare domains. Lee et al. approached the challenge of identifying breast cancer patients who could benefit from chemotherapy using data mining techniques. Through a clustering process, they successfully classified breast cancer patients into distinct groups: Good, Poor, and Intermediate, thereby contributing to the enhancement of survival outcomes<sup>(10)</sup>. Kaur et al. utilized GSP and SPADE algorithms to find the frequent treatment sequences given to the patients. Further they employed time intervals in frequent treatment sequences to evaluate patient survival<sup>(3)</sup>.

Within this study, the authors delve into the analysis of prostate cancer survivability through the utilization of machine learning and deep learning methods, while also employing sequential mining to map the treatments. The paper's structure is as follows: Section 2 covers the research methodology used, Section 4 presents the outcomes derived from the implemented methods and analyses, and finally, Section 5 concludes the study with a summary of the findings.

## 2 Methodology

There are several steps followed during a machine learning modeling process as shown in Figure 1. The first step includes data collection and preprocessing. The data is cleaned and made fit for analysis. Following that, the model is trained on a subset of the dataset. The model is then assessed on the remaining dataset, and its accuracy is computed. However, in this study, the modeling process is going to be two-fold: Survival prediction using Machine Learning on the complete dataset and using Sequential Mining on the treatments. Figure 1 illustrates the methodology of this study.

### 2.1 Data Collection and Preprocessing

A fundamental necessity for conducting a research study is to obtain a suitable dataset. While ample cancer datasets, such as the SEER dataset, are accessible online, this research concentrated on obtaining particular attributes that were not readily present in existing online datasets<sup>(11)</sup>. Additionally, it should be noted that none of the available datasets encompassed data pertaining to

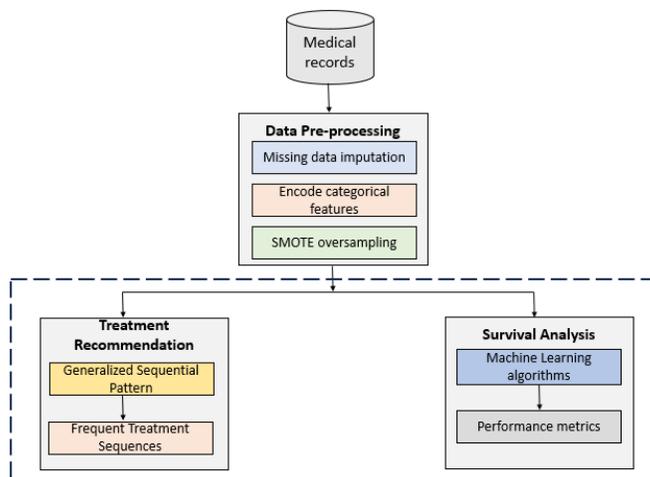


Fig 1. Methodology of the study

Indian patients, while the factors of race and region have been found to influence patient survivability. The dataset was sourced from the Rajiv Gandhi Cancer Institute and Research Centre, New Delhi. Experienced oncologists from the hospital provided guidance for the data collection process. The dataset consists of retrospective cases of prostate cancer patients. It consists of 410 records of prostate cancer patients and 21 features. Features such as, Age, Histology, PSA levels, Gleason Score, etc. are essential to determine the survivability of the patient<sup>(12)</sup>. The class outcome is Survival after 3 Years. Within the dataset under consideration, 61% of the patients did not survive, while 39% of them exhibited a survival duration exceeding three years. The study exclusively included patients diagnosed between 2011 and 2015 to ensure the data’s relevance and contemporaneity. Patients diagnosed prior to this period would have outdated diagnostic methods and treatment procedures, rendering their data obsolete for the study’s objectives. Moreover, the retrieval of 3-year survival information from patients’ records or through direct communication was confined to the data available up to 2015. To uphold the integrity of the analysis, patients who were no longer being monitored or did not continue with the treatment were not included from the study, and their data was not taken into account in any subsequent analyses. Table 1 presents the description of the dataset used for the analysis.

Before proceeding with the analysis, the dataset underwent a rigorous preprocessing phase to guarantee its quality and suitability for further examination. Initial steps involved removing irrelevant features like Height and Weight which were deemed inconsequential to the research objective. Also, all the comorbidities of the patients were combined into a single attribute- Charlson Comorbidity Index. Subsequently, missing values were carefully handled to mitigate any potential biases or distortions in the dataset. To further refine the dataset, a process of feature engineering was employed, including the creation of dummy variables through one-hot encoding for categorical attributes. By expanding the dimensionality of the dataset in this manner, the subsequent analysis could capture the nuanced relationships between the transformed features. Also, superfluous columns associated with the categorical variables and redundant target labels were pruned from the dataset, streamlining it for further exploration and modelling.

Finally, the dataset was processed using SMOTE (Synthetic Minority Over-sampling Technique) balancing technique to balance the dataset. The application of SMOTE to address the data imbalance within prostate cancer data, where the survival rate stands at 61%, holds profound significance in the context of this research. This rebalancing of the dataset is critical for fostering precision and mitigating bias in machine learning model training, ultimately contributing to more informed decision-making and the refinement of patient care strategies within the sphere of prostate cancer management. Following are the number of instances in the dataset before and after SMOTE balancing technique.

- Training dataset (before SMOTE): 328
- Training dataset (after SMOTE): 418
- Survived cases in Training dataset (before SMOTE): 119
- Deceased cases in Training dataset (before SMOTE): 209
- Survived cases in Training dataset (after SMOTE): 209
- Deceased cases in Training dataset (after SMOTE): 209

This comprehensive pre-processing methodology ensured the dataset’s integrity, completeness, and suitability for subsequent research analyses in the field of interest.

**Table 1.** Dataset description

Attributes	Description	Minimum Value	Maximum Value	Mode
<b>Year of diagnosis</b>	Year in which diagnosis was made	2002	2015	2014
<b>Age at diagnosis</b>	Age of the patient at diagnosis	15	94	65
<b>Family History</b>	Instance of cancer in the patient’s family	0	1	1
<b>Histology</b>	Description of cancer cells	NA	NA	Adenocarcinoma
<b>PSA</b>	Prostate Specific Antigen level at the the time of diagnosis	0	23500	NA
<b>Charlson Comorbidity Index</b>	10-year survival prediction based on various comorbidities	2	7	5
<b>T</b>	Tumor stage	2	4	4
<b>N</b>	Nodes stage	NA	NA	N0
<b>Body Mass Index</b>	A measure of body fat based on height and weight	18	30	NA
<b>Gleason Group</b>	The Gleason score gives the severity of the cancer.	1	9	8
<b>ECOG</b>	Performance status of the patient	1	4	2
<b>Nadir PSA</b>	Lowest PSA Level	0	5	2580
<b>PSA doubling time&lt; 6 months</b>	PSA doubling time less than 6 months	NA	NA	Yes
<b>BP</b>	Blood pressure	NA	NA	Yes

## 2.2 Survival analysis

### 2.2.1 Experiment 1: Survival Prediction

This study delved into survival prediction in prostate cancer by utilizing a wide range of machine learning algorithms, which encompassed Artificial Neural Networks (ANN), Adaboost, Gradient Boosting, Random Forest, SVM, and Logistic Regression. Each of these techniques brings unique strengths and capabilities to various classification tasks including varying health related applications<sup>(13-15)</sup>. Table 2 provides a brief summary of the algorithms used along with their hyperparameters in the survival prediction process.

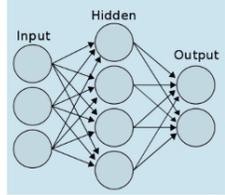
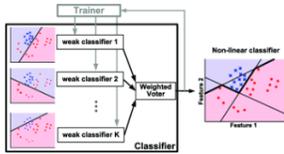
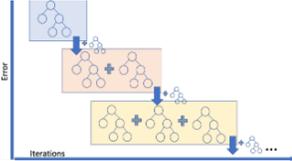
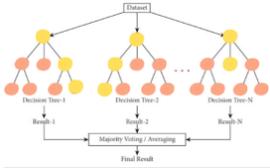
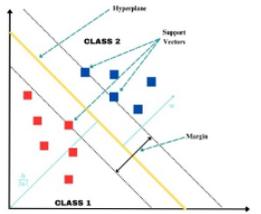
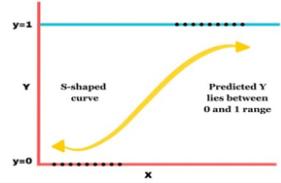
By employing these diverse machine learning techniques, this study aimed to compare their performance and evaluate their efficacy in predicting survival outcomes for prostate cancer patients. The selection of these algorithms provides a comprehensive analysis and allows for a robust assessment of their respective strengths and weaknesses in the context of prostate cancer survival prediction.

### 2.2.2 Experiment 2: Sequential Mining

Sequence mining involves the extraction of commonly occurring sets of sub-sequences from a given sequence. Within the Knowledge Discovery in Databases, data mining plays a crucial role in converting raw data into valuable information and actionable knowledge. Its application in healthcare and personalized medicine holds significant potential, making it especially valuable in these domains<sup>(16)</sup>. As an illustration, by thoroughly analyzing a patient’s comprehensive profile, medical history, results of physical examinations, and accurate diagnosis, and drawing insights from past treatment patterns, novel and highly effective treatment plans can be adeptly recommended<sup>(17)</sup>. Malhotra et al. utilized sequential pattern mining to make predictions about the survival of Glioblastoma cancer patients<sup>(18)</sup>. In their research, Laxminarayan et al. introduced an innovative approach for association rule mining, specifically designed to analyze sequence-valued attributes like sleep-related data<sup>(19)</sup>. Rjeily et al. performed time-based analysis of heart diseases using classification and sequential mining algorithms<sup>(20)</sup>. Uragaki et al. applied time interval sequential pattern mining (TI-SPM) algorithm on electronic medical records (EMRs) to extracted clinical pathways<sup>(21)</sup>.

Generalized sequential pattern mining in the medical field involves the analysis of sequential data to uncover meaningful patterns and associations within patient health records, medical events, or treatment protocols. GSP algorithms applied in the medical domain consider parameters such as support and confidence to determine the significance and reliability of identified patterns. The output of general sequential pattern mining in healthcare consists of discovered patterns or rules that provide insights into the temporal relationships between medical events<sup>(22)</sup>. These patterns can contribute to various applications, such

**Table 2.** Summary of Machine Learning Techniques

Technique	Description	Advantages	Diagram	Hyper parameter
Artificial Neural Network (ANN)	ANN, inspired by the human brain's neural networks, captures complex nonlinear data relationships.	Excellent at capturing complex patterns and relationships. Able to learn from large datasets.		Activation functions: ReLU for hidden layers and sigmoid for the output layer Loss function: Binary Cross entropy. Optimizer: Adam. Number of epochs: 1000. Batch size: 64.
Adaboost	An ensemble method merges multiple weak classifiers to construct a robust prediction model.	Improves accuracy iteratively by focusing on misclassified instances. Leverages collective knowledge of classifiers.		Number of estimators= 200 Random state= 42
Gradient Boosting	An ensemble method that sequentially trains weak learners, focusing on the residual errors of predecessors.	High prediction accuracy. Suitable for complex datasets.		Number of estimators= 200
Random Forest	An ensemble method that aggregates multiple decision trees to make the final prediction.	Reduces overfitting and improves generalization. Useful for handling complex datasets.		Maximum depth= 0
Support Vector Machines (SVM)	Powerful algorithm for complex, high-dimensional data.	Effective in high-dimensional spaces. Can handle nonlinear relationships through kernel functions.		Kernel function: Quadratic
Logistic Regression	A statistical technique used in binary classification tasks. Estimates probabilities based on input variables.	Offers valuable insights into the correlation between predictors and outcomes.		-

as clinical decision support systems, patient monitoring, disease progression modeling, and treatment optimization<sup>(23)</sup>. The GSP algorithm taking dataset  $D$  and an itemset  $I$  is given as below:

$GSP(D, I, minsup) :$

$F \leftarrow \emptyset$

$C^{(l)} \leftarrow \{\emptyset\}$  //Initial prefix tree with single symbols

foreach  $s \in I$  do Add  $s$  as child of  $\emptyset$  in  $C^{(l)}$  with  $sup(s) \leftarrow 0$

$k \leftarrow l$

while  $C^{(k)} \neq \emptyset$  do

    ComputeSupport  $(C^{(k)}, D)$

    foreach leaf  $s \in C^{(k)}$  do

        if  $sup(s) \geq minsup$  then  $F \leftarrow F \cup \{(s, sup(s))\}$

        else remove  $s$  from  $C^{(k)}$

$C^{(k+1)} \leftarrow ExtendPrefixTree(C^{(k)})$

$k \leftarrow k + 1$

    return  $F^{(k)}$

ComputeSupport  $(C^k, D) :$

    foreach  $s_i \in D$  do

```

        foreach  $r \in C^{(k)}$  do

            if  $r < s_i$  then  $sup(r) \leftarrow sup(r) + 1$ 

            ExtendPrefixTree( $C^{(k)}$ ):

                foreach leaf  $r_a \in C^{(k)}$  do

                    foreach leaf  $r_b \in Children(Parent(r_a))$  do

                         $r_{ab} \leftarrow r_a + r_b[k]$ 

                        //prune if there are any infrequent subsequences

                        if  $r_c \in C^{(k)}$  for all  $r_c \supset r_{ab}$ , such that  $|r_c| = |r_{ab}| - 1$  then

                            Add  $r_{ab}$  as child of  $r_c$  with  $sup(r_{ab}) \leftarrow 0$ 

                        if no extensions from  $r_a$  then remove  $r_a$  from  $C(k)$ 

                    return  $C(k)$ 
    
```

### 3 Results and Discussion

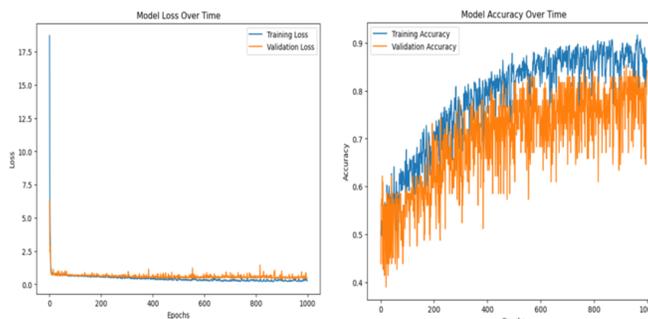
#### 3.1 Survival Prediction

The results for survival prediction using machine learning are summarized in Table 3.

Among the models evaluated, the Artificial Neural Networks achieved the highest accuracy of 0.8414, demonstrating its ability to make accurate predictions on the prostate cancer survival dataset. Furthermore, the study demonstrated a well-balanced trade-off between precision and recall, achieving values of 0.8108 and 0.8333, respectively. The F1 score, which considers both precision and recall, yielded a favorable value of 0.8219, indicating the model's robustness in handling imbalanced class distributions. Additionally, the ROC-AUC score of 0.8405 further confirms the model's strong discriminative capabilities. On the other hand, Logistic Regression demonstrated respectable performance, achieving an accuracy of 0.7682. Nonetheless, the model's precision and recall scores were relatively lower at 0.7428 and 0.7222, respectively, suggesting that it may encounter challenges in achieving a balance between accurately identifying positive instances and avoiding false positives. Thus, the Artificial Neural Networks (ANN) displayed remarkable accuracy, precision, recall, and F1 score, surpassing other techniques and emerging as the most promising candidate for precise survival predictions. The ANN's proficiency in capturing

**Table 3.** Performance Metrics of Classification Models

	Accuracy	Precision	Recall	F1 Score	AUC
ANN	<b>0.8414</b>	0.8108	<b>0.8333</b>	<b>0.8219</b>	<b>0.8405</b>
Logistic Regression	0.7682	0.7428	0.7222	0.7323	0.7632
Random Forest	0.7560	0.7666	0.6388	0.6969	0.7433
Decision Tree	0.6707	0.6551	0.5277	0.5846	0.6551
SVM	0.7560	0.7500	0.6666	0.705	0.7463
Adaptive Boosting	0.8170	<b>0.8387</b>	0.7222	0.7761	0.8067
Gradient Boosting	0.7560	0.7666	0.6388	0.6969	0.7433



**Fig 2.** Model Loss Over Time of ANN; (b) Model Accuracy Over Time of ANN

intricate nonlinear relationships and patterns within the data played a crucial role in its successful performance in this task. Figure 2 shows the loss and accuracy of ANN.

Comparing our results with existing literature, our Artificial Neural Network (ANN) model outperforms other methodologies across multiple performance metrics. In terms of accuracy, the ANN model employed achieved an impressive score of 0.8414, surpassing the results obtained by Thongpim’s Cox regression model<sup>(5)</sup>, Koo’s MLP, MLP-5, LSTM models<sup>(4)</sup>, Nezhad’s survival rates at various stages, Doja’s SVM<sup>(6)</sup>, and DT models, as well as Wen’s precision, recall, and F1-score for both classes<sup>(8)</sup>. In our study, the ANN demonstrated exceptional precision (0.8108) and recall (0.8333), resulting in a balanced F1 score of 0.8219. These metrics indicate a notable ability to accurately predict prostate cancer survival. The area under the ROC curve (AUC) for our ANN model further validates its discriminative capabilities with a score of 0.8405. Examining other models in the study, Logistic Regression achieved a respectable accuracy of 0.7682, but its precision (0.7428) and recall (0.7222) were relatively lower, suggesting potential challenges in balancing true positives and false positives. Random Forest, Decision Tree, SVM, Adaptive Boosting, and Gradient Boosting also demonstrated varying levels of performance.

### 3.2 Generalized Sequential Pattern Mining

In this experiment, the Generalized Sequential Pattern (GSP) algorithm was employed to gain invaluable understanding into the survival rates of prostate cancer patients following 3 years of treatment. By applying the GSP algorithm to the ‘First’, ‘Second’, ‘Third’, and ‘Fourth’ treatment columns, they aimed to identify frequent treatment patterns and their corresponding survival rates, thus offering potential treatment combinations to enhance patient outcomes.

Figure 3 represents the top 10 sequences after using GSP. By utilizing the GSP (Generalized Sequential Pattern) algorithm, some valuable insights can be derived from the dataset regarding the survival rates of prostate cancer patients after 3 years. Utilizing the ‘First’, ‘Second’, ‘Third’, and ‘Fourth’ treatment columns, we applied the GSP algorithm to identify frequent treatment patterns.

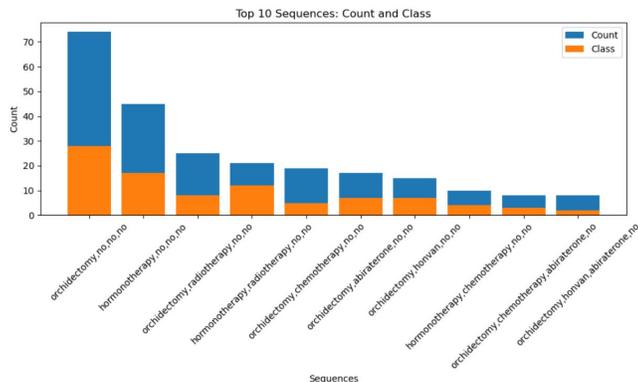


Fig 3. Sequences generated by Generalized Sequential Pattern

- The most frequent treatment pattern observed was ‘orchidectomy, no, no, no’ with a count of 74 patients. The survival rate associated with this pattern was 37.84%, indicating that approximately 37.84% of patients who received this treatment combination survived after 3 years.
- Among the analyzed patterns, ‘hormonotherapy, radiotherapy, no, no’ exhibited the highest survival rate of 57.14%. This pattern suggests that patients who received a combination of hormonotherapy and radiotherapy, without undergoing any other treatments in the ‘Third’ and ‘Fourth’ columns, had the highest likelihood of survival after 3 years.

These findings highlight the potential significance of the treatment combination ‘hormonotherapy, radiotherapy, no, no’ in improving the survival outcomes for prostate cancer patients after 3 years. The insights gained through the GSP algorithm not only contribute to understanding the efficacy of specific treatment combinations but also provide valuable information regarding the complex relationships between treatments and survival rates in prostate cancer patients. The research findings have implications for tailoring personalized treatment approaches for prostate cancer and advancing precision medicine.

In the realm of Generalized Sequential Pattern (GSP) mining, our study extends beyond existing literature. While previous works have explored survival patterns, our GSP algorithm provides nuanced insights by associating specific treatment patterns with survival rates. These findings contribute to a more detailed understanding of the intricate relationships between treatments and survival outcomes in prostate cancer patients. Our research has broader implications for tailoring personalized treatment approaches and advancing precision medicine in prostate cancer care.

### 4 Conclusion

In the culmination of this research, advanced machine learning methodologies were skillfully employed to unravel the intricate dynamics of survival prediction among 410 metastatic prostate cancer patients. The standout performer, without question, is the Artificial Neural Networks (ANN), which emerged as the frontrunner with unparalleled accuracy (0.8414), recall (0.8333), F1 score (0.8219), and a robust ROC-AUC Score (0.8405). This unequivocally positions the ANN as a beacon of precision in forecasting survival outcomes in the context of prostate cancer. The incisive application of the Generalized Sequential Pattern (GSP) algorithm delved deep into the labyrinth of treatment patterns and their intricate pattern with survival rates. The revelation of the ‘hormonotherapy AND radiotherapy’ combination wielding the highest survival rate not only underscores its clinical significance but also advocates for its pivotal role in shaping future patient outcomes.

As we chart the course forward, the trajectory of this research extends into broader horizons. Scaling up to larger and more diverse datasets, enriched with a comprehensive array of clinical and demographic variables, is paramount. This broader canvas ensures not only the validation and generalization of our current findings but also enhances the adaptability of the predictive models. It’s worth noting that while our current dataset was real-time, modest in size, the prospect of expanding this paradigm to encompass a spectrum of cancers or diseases is tantalizing. This not only amplifies the impact of our predictive models but also offers a transformative roadmap for the personalized treatment landscape. In a bid to fortify the arsenal of survival analysis, the prospect of incorporating more sophisticated algorithms looms large. The evolving frontier of advanced machine learning technologies, coupled with real-time data integration, promises not just refinement but a paradigm shift in treatment strategies and, ultimately, an elevation in patient care across a multitude of medical domains.

## References

- 1) Rawla P. Epidemiology of prostate cancer. 2019. Available from: <https://doi.org/10.14740/wjon1191>.
- 2) Shariat SF, Roehrborn CG. Using biopsy to detect prostate cancer. *Reviews in urology*. 2008;10(4):262–262. Available from: <https://pubmed.ncbi.nlm.nih.gov/19145270/>.
- 3) Kaur I, Doja MN, Ahmad T. Time-range based sequential mining for survival prediction in prostate cancer. *Journal of Biomedical Informatics*. 2020;110:103550–103550. Available from: <https://doi.org/10.1016/j.jbi.2020.103550>.
- 4) Koo KC, Lee KS, Kim S, Min C, Min GR, Lee YH, et al. Long short-term memory artificial neural network model for prediction of prostate cancer survival outcomes according to initial treatment strategy: development of an online decision-making support system. *World Journal of Urology*. 2020;38(10):2469–2476. Available from: <https://doi.org/10.1007/s00345-020-03080-8>.
- 5) Thongpim N, Choksuchat C, Bejrananda T, Matayong S. On Predicting Survival Opportunities for Prostate Cancer by COX Regression in PSU Patients Data. *2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*. 2020;p. 775–778. Available from: <https://doi.org/10.1109/ECTI-CON49241.2020.9158318>.
- 6) Doja MN, Kaur I, Ahmad T. Age-specific survival in prostate cancer using machine learning. 2020. Available from: <https://doi.org/10.1108/DTA-10-2019-0189>.
- 7) Nezhad MZ, Sadati N, Yang K, Zhu D. A Deep Active Survival Analysis approach for precision treatment recommendations: Application of prostate cancer. *Expert Systems with Applications*. 2019;115:16–26. Available from: <https://doi.org/10.1016/j.eswa.2018.07.070>.
- 8) Wen H, Li S, Li W, Li J, Yin C. Comparison of Four Machine Learning Techniques for the Prediction of Prostate Cancer Survivability. *2018 15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*. 2018;p. 112–116. Available from: <http://dx.doi.org/10.1109/ICCWAMTIP.2018.8632577>.
- 9) Li R, Shinde A, Liu A, Glaser S, Lyou Y, Yuh B, et al. Machine Learning–Based Interpretation and Visualization of Nonlinear Interactions in Prostate Cancer Survival. *JCO Clinical Cancer Informatics*. 2020;4(4):637–646. Available from: <https://pubmed.ncbi.nlm.nih.gov/32673068/>.
- 10) Lee YJ, Mangasarian OL, Wolberg WH. Survival-time classification of breast cancer patients. *Computational Optimization and Applications*. 2003;25:151–166. Available from: <https://doi.org/10.1023/A:1022953004360>.
- 11) SEER Cancer Statistics Review. *National Cancer Institute, Bethesda*. 1975.
- 12) Cruz JA, Wishart DS. Applications of Machine Learning in Cancer Prediction and Prognosis. *Cancer Informatics*. 2006;2. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2675494/>.
- 13) Qayyum A, Qadir J, Bilal M, Al-Fuqaha A. Secure and Robust Machine Learning for Healthcare: A Survey. *IEEE Reviews in Biomedical Engineering*. 2021;14:156–180. Available from: <https://doi.org/10.1109/RBME.2020.3013489>.
- 14) Kaur I, Ahmad T. A cluster-based ensemble approach for congenital heart disease prediction. 2024. Available from: <https://doi.org/10.1016/j.cmpb.2023.107922>.
- 15) Kaur I, Doja MN, Ahmad T. An Empirical Analysis of Survival Predictors for Cancer Using Machine Learning. In: *Advances in Intelligent Systems and Computing*;vol. 1. Springer Singapore. 2022;p. 203–212. Available from: [https://doi.org/10.1007/978-981-16-2594-7\\_17](https://doi.org/10.1007/978-981-16-2594-7_17).
- 16) Holzinger A. Trends in interactive knowledge discovery for personalized medicine: cognitive science meets machine learning. *IEEE Intelligent Informatics Bulletin*. 2014;15(1):6–14. Available from: <https://graz.elsevierpure.com/en/publications/trends-in-interactive-knowledge-discovery-for-personalized-medicine>.
- 17) Campbell T, Srivastava S. Exploration of Classification Techniques as a Treatment Decision Support Tool for Patients with Uterine Fibroids. *International Workshop on Data Mining for HealthCare Management*. Available from: <https://conservancy.umn.edu/handle/11299/215828>.
- 18) Malhotra K, Navathe SB, Chau DH, Hadjipanayis C, Sun J. Constraint based temporal event sequence mining for Glioblastoma survival prediction. *Journal of Biomedical Informatics*. 2016;61:267–275. Available from: <https://doi.org/10.1016/j.jbi.2016.03.020>.
- 19) Laxminarayan P, Alvarez SA, Ruiz C, Moonis M. Mining Statistically Significant Associations for Exploratory Analysis of Human Sleep Data. *IEEE Transactions on Information Technology in Biomedicine*. 2006;10(3):440–450. Available from: <https://doi.org/10.1109/titb.2006.872065>.
- 20) Rjeily CB, Badr G, Hassani HE, E AA. Medical data mining for heart diseases and the future of sequential mining in medical field. 2019. Available from: [https://doi.org/10.1007/978-3-319-94030-4\\_4](https://doi.org/10.1007/978-3-319-94030-4_4).
- 21) Uragaki K, Hosaka T, Arahori Y, Kushima M, Yamazaki T, Araki K, et al. Hajjarm El Hassani A, Andres E. Medical data mining for heart diseases and the future of sequential mining in medical field. In: *In2016 IEEE Symposium on Computers and Communication (ISCC)*. IEEE. 2016;p. 71–99. Available from: <https://doi.org/10.1109/ISCC.2016.7543708>.
- 22) Lu J, Hales A, Rew D, Keech M, Fröhlingdsdorf C, Mills-Mullett A, et al. Data Mining Techniques in Health Informatics: A Case Study from Breast Cancer Research. In: *Information Technology in Bio- and Medical Informatics*;vol. 6. Springer International Publishing. 2015;p. 56–70. Available from: [https://doi.org/10.1007/978-3-319-22741-2\\_6](https://doi.org/10.1007/978-3-319-22741-2_6).
- 23) Teo MY, Rathkopf DE, Kantoff P. Treatment of Advanced Prostate Cancer. *Annual Review of Medicine*. 2019;70(1):479–499. Available from: <https://doi.org/10.1146/annurev-med-051517-011947>.