



OPEN

Explainable ML models for a deeper insight on treatment decision for localized prostate cancer

Jang Hee Han^{1,9}, Sungyup Lee^{2,9}, Byoungwa Lee², Ock-kee Baek², Samuel L. Washington III^{3,4}, Annika Herlemann^{3,5}, Peter E. Lonergan^{3,6,7}, Peter R. Carroll³, Chang Wook Jeong^{1,3,8,9}✉ & Matthew R. Cooperberg^{3,4}

Although there are several decision aids for the treatment of localized prostate cancer (PCa), there are limitations in the consistency and certainty of the information provided. We aimed to better understand the treatment decision process and develop a decision-predicting model considering oncologic, demographic, socioeconomic, and geographic factors. Men newly diagnosed with localized PCa between 2010 and 2015 from the Surveillance, Epidemiology, and End Results Prostate with Watchful Waiting database were included (n = 255,837). We designed two prediction models: (1) Active surveillance/watchful waiting (AS/WW), radical prostatectomy (RP), and radiation therapy (RT) decision prediction in the entire cohort. (2) Prediction of AS/WW decisions in the low-risk cohort. The discrimination of the model was evaluated using the multiclass area under the curve (AUC). A plausible Shapley additive explanations value was used to explain the model's prediction results. Oncological variables affected the RP decisions most, whereas RT was highly affected by geographic factors. The dependence plot depicted the feature interactions in reaching a treatment decision. The decision predicting model achieved an overall multiclass AUC of 0.77, whereas 0.74 was confirmed for the low-risk model. Using a large population-based real-world database, we unraveled the complex decision-making process and visualized nonlinear feature interactions in localized PCa.

Accounting for more than 75% of newly diagnosed men with prostate cancer (PCa)¹, localized PCa exhibits remarkable inter-tumor heterogeneity and risk group diversity²; thus, various treatment options are now being proposed without solid decision criteria.

In fact, none of the possible treatments—such as observation (active surveillance(AS)/watchful waiting(WW)), radical prostatectomy (RP), or radiation therapy (RT)—have been proven superior in terms of cancer control in localized disease³. Thus, an initial treatment decision is often based on the patient's PCa risk stratification and patient's and physician's treatment preference^{4,5}. However, in the real-world setting, decision-making is a complex process that is not only affected by cancer characteristics but also by various patient-level, state/county-level regional, and socioeconomic factors⁶. Thus, a comprehensive approach is strongly needed for patients and physicians. In order to address this issue, there are several decision aids (DAs) for localized PCa patients⁷. However, there is lack of uniformity between these decision aids and their performances⁸.

An explainable machine learning model offers advantages by obtaining a deeper understanding of the internal processes, while the model itself trains or makes decisions and identifies cause-and-effect relationships within the system's inputs and outputs⁹. In this study, we highlight the complex treatment decision process in localized PCa using the Surveillance, Epidemiology, and End Results (SEER) prostate with watchful waiting dataset (SEER/WW) through an explainable machine learning model. Using two-variable interaction plots, our primary aim was to gain a deeper insight into the important features associated with each treatment modality. Our secondary aim was to develop a treatment decision prediction model considering overall features, including oncologic,

¹Department of Urology, Seoul National University Hospital, Seoul, Republic of Korea. ²Electronics and Telecommunications Research Institute (ETRI), Daejeon, Republic of Korea. ³Department of Urology, Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, CA, USA. ⁴Department of Epidemiology and Biostatistics, University of California, San Francisco, CA, USA. ⁵Department of Urology, Ludwig-Maximilians-University of Munich, Munich, Germany. ⁶Department of Urology, St. James's Hospital, Dublin, Ireland. ⁷Department of Surgery, Trinity College, Dublin, Ireland. ⁸Department of Urology, Seoul National University College of Medicine, Seoul, Republic of Korea. ⁹These authors contributed equally: Jang Hee Han, Sungyup Lee and Chang Wook Jeong. ✉email: drboss@snu.ac.kr

geographic (county-level data), demographic, and socioeconomic factors, which were integrated into a web-based platform for use in daily clinical routine.

Results

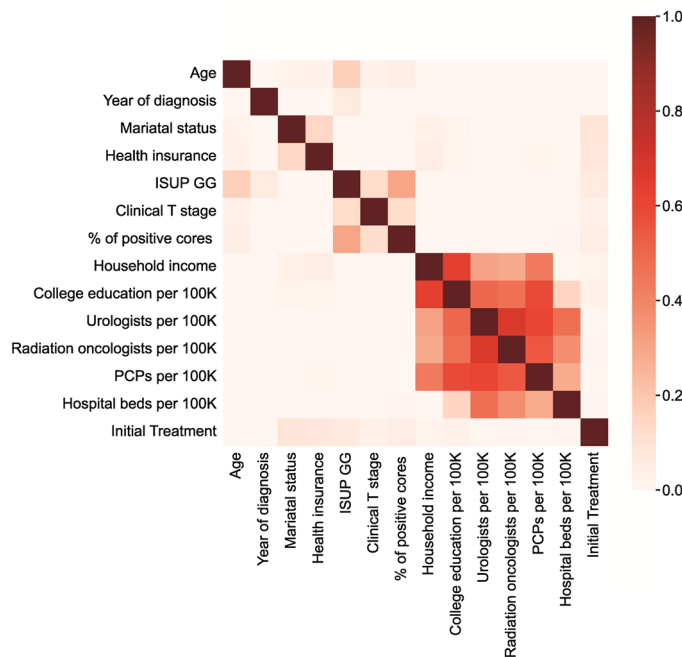
We identified 255,837 men with newly diagnosed localized PCa who met the inclusion and exclusion criteria (Supplementary Fig. 1). Among these, 26,389 (10.3%) underwent AS/WW, 86,714 (33.9%) underwent RP, 76,919 (30.8%) underwent RT, and 63,815 (24.9%) underwent other/unknown treatments, including androgen deprivation therapy (ADT), combined ADT and RT, etc. In the low-risk cohort model (patients with clinical T stage T1c and T2a, Gleason grade group 1, and PSA ≤ 10 ng/mL, age < 80), 79,633 patients were included. Among them, 17,553 (22%) underwent AS/WW and 62,080 (78%) underwent other treatments such as androgen deprivation therapy or focal therapy.

Entire cohort. Regional factors, such as the SEER registry, state/county, average level of education, average number of healthcare providers (urologists, radiation oncologists, primary care practitioners), and health facilities showed correlative associations (Fig. 1A). The decision-predicting machine-learning model achieved fair discrimination in the test set, with an overall multiclass AUC of 0.77 (Fig. 1B). The easiest treatment to distinguish among them was AS/WW (AUC of 0.84), whereas the most difficult to distinguish was RT (AUC of 0.72). RP treatment showed intermediate discrimination performance (AUC of 0.78).

Identification of the important features per treatment. The contribution and effect of each feature (as measured by SHAP scores) on the chosen outcome are presented in Fig. 2. The decision to perform AS/WW was mostly affected by the ISUP grade group (GG), followed by the positive core percentage (PPC), clinical T stage, state/county, and year of diagnosis. Those for RP were age, ISUP GG, PSA, and PPC. For RT, ISUP GG, age, state/county, and year of diagnosis were the important features.

Interaction of features in a dependence plot. Within the AS/WW group, for the T1c stage, low PPC patients chose AS/WW, whereas T2 stage patients tended to choose AS/WW when PPC was high. Similarly, for the same ISUP GG, groups with a differing PSA level (low vs. high) showed different decision-making (Supplementary Fig. 2). For the RP group, the recently diagnosed patients chose RP when ISUP GG was high, whereas previously diagnosed patients showed the opposite trend (the ISUP GG low group chose RP). In addition, there was a race/ethnic difference in RP treatment decisions, where white men tended to choose RP when PPC was high, while those with other races avoided RP (Supplementary Fig. 3). For the RT group, ISUP GG1 preferred not to choose RT, while ISUP GG2-5 preferred RT. PSA was a distinct oncological feature that affected different decisions regarding RT treatment. The older group avoided undergoing RT when PSA was high and chose AS/WW instead, while the younger group preferred undergoing RT when PSA was high. Furthermore, T1 and T2

(A) Kendall Correlation



(B) ROC AUC of Entire Cohort

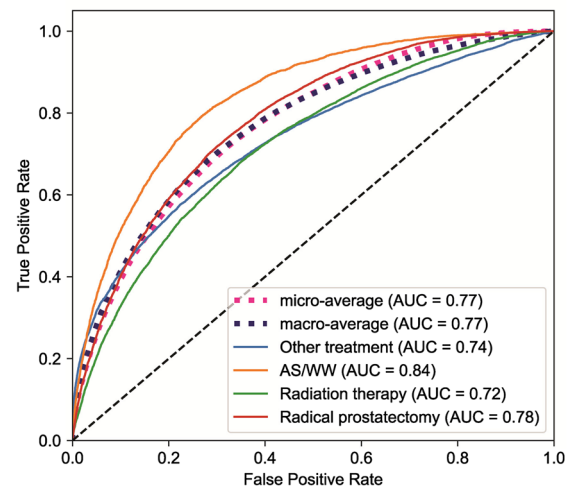
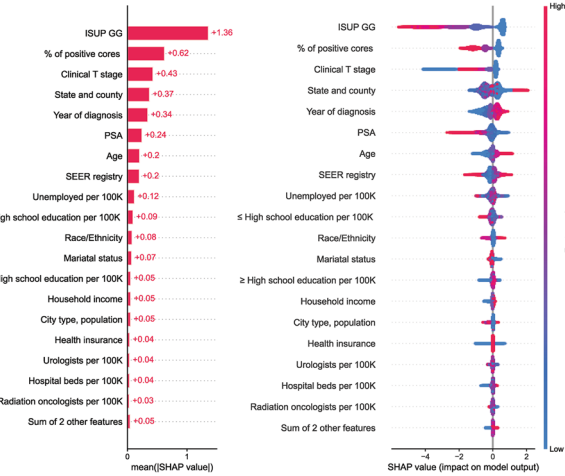
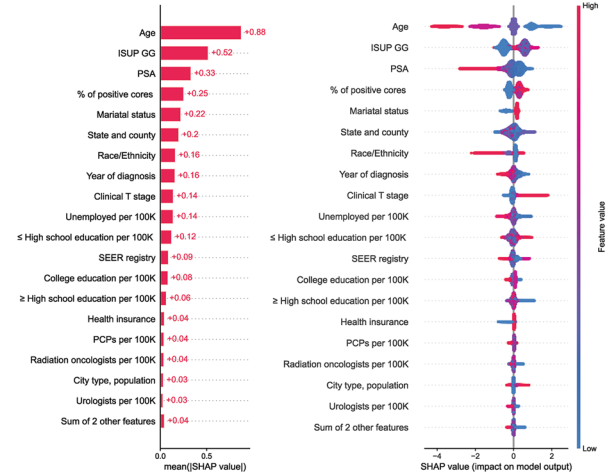


Figure 1. (A) Correlation between categorical features by Kendall method in entire localized prostate cancer cohort. (B) A decision prediction model using the entire cohort's receiver operating characteristics curve and multiclass area under the curve (AUC).

(A) AS/WW in Entire Cohort



(B) Radical prostatectomy in Entire Cohort



(C) Radiation therapy in Entire Cohort

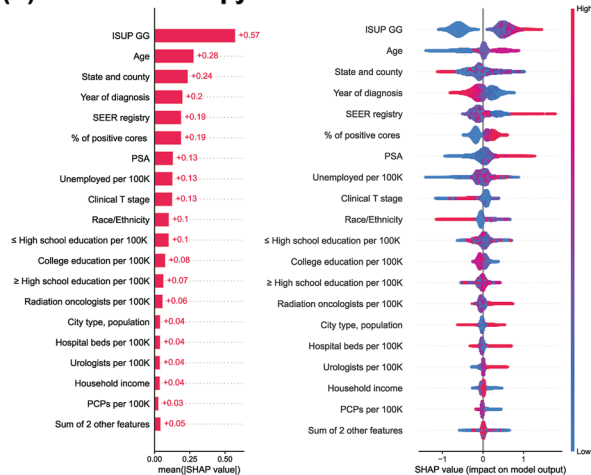


Figure 2. Global feature importance plot and Beeswarm plot using SHAP value for each treatment decision (A) AS/WW (B) RP (C) RT in the entire localized prostate cancer cohort.

stage patients showed the opposite trend of RT decision for similar accompanying oncological characteristics. Typically, high PPC had a positive impact on RT decision-making in T2 stage, but a negative impact in T1 stage patients (Supplementary Fig. 4).

Low-risk localized prostate cancer cohort. The same correlation was observed in the low-risk cohort as in the entire cohort, showing a positive correlation between regional factors (Fig. 3A). The decision-predicting machine learning model achieved fair discrimination, with an overall binary class AUC of 0.74 (Fig. 3B). The most important factor in the AS/WW decision was the year of diagnosis, followed by the PPC, state/county, age, SEER registry, and education (Fig. 4).

Interaction of features in the low-risk cohort model’s dependence plot. There were different characteristics between races and ethnicities. White patients tended to choose AS/WW when age was high or PPC was low, while patients of other race or ethnicity (black, Hispanic, Asian, and others) showed the opposite trend. For T1c stage, recently diagnosed patients or those with higher PSA levels chose AS/WW, whereas T2a stage patients preferred active treatments when recently diagnosed or with a lower PSA level (Supplementary Fig. 5).

Development of an online applicable prediction model using the SEER/WW platform. The model is open for patient access at <http://210.117.211.210:8501/> (Supplementary Fig. 6). Model information and details are listed in Supplementary Table 2.

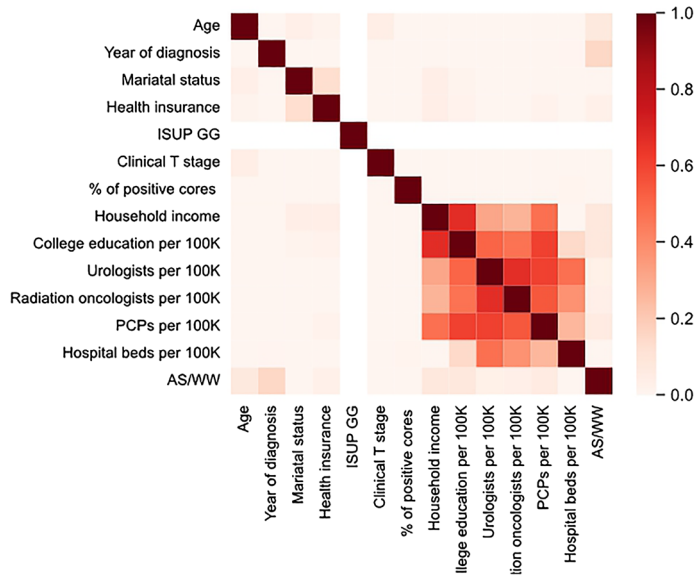
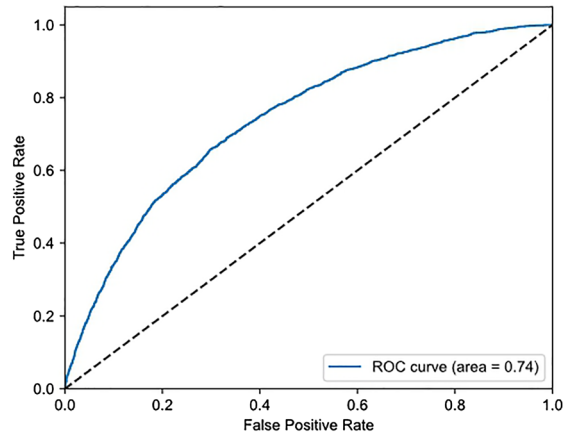
(A) Kendall Correlation**(B) ROC AUC of Low Risk Cohort**

Figure 3. (A) Correlation between categorical features by Kendall method in low-risk cohort. (B) Receiver operational characteristics curve and multiclass area under the curve (AUC) measure in a decision prediction model with a low-risk cohort.

Discussion

The factors that influence the complex decision-making process in the treatment of localized PCa remain debatable¹⁰, suggesting that we still lack an understanding of these complex processes. Thus, clinicians collaboratively help patients make proper decisions through a shared decision-making process, such as informing evidence about the safety and value of each treatment method⁸ or informing the types of treatments offered to similar men with similar cancer characteristics to have more consistency in their decisions. In this regard, the Michigan Urological Surgery Improvement Collaborative (MUSIC) recently developed a new machine learning model to help men view predicted treatment decisions of similar patients using a prospective registry of 7543 men diagnosed with prostate cancer¹¹. However, this random forest model was unable to provide a distinct understanding of how a single decision was reached and it did not consider geographic and socioeconomic variations in the treatment decision¹².

To compensate for these shortcomings, we adopted an explainable machine learning model using the recently released SEER-WW dataset (2010–2016), which has a newly created variable clearly defined as “AS/WW”. Precise model training was enabled by a more accurate treatment group classification compared with the existing SEER databases. We trained machine learning models using a broad range of real-world features as input, including clinicopathologic characteristics as well as demographics, socio-economic information, and nationwide county level geographic data, including regional healthcare resources, which the MUSIC group (Michigan state-confined cohort) could not address. Usually, there is a trade-off between interpretability and accuracy for each machine learning model¹³; however, our model showed interpretability while having comparable model performance with the MUSIC group (AUC 0.77 vs. 0.81).

The overall characteristics of the machine learning model developed in this study included geographic features (state/county and SEER registry) that mostly affected RT treatment decisions, followed by AS/WW. This result agrees with a previous study by Wang et al. who showed that the availability of RT is the most important contributor to regional variation⁴. Meanwhile, a previous paper by Washington et al. demonstrating that geographic location is associated with AS/WW practice variation⁶, better explains the regional factors influencing AS/WW decisions. Contrary to our expectations, socio-economic features (marital status, educational status, household income, insurance, etc.) generally did not influence treatment decision-making, except that in RP patients, married men showed a tendency to undergo surgery compared to unmarried/single men. This likelihood is commonly observed in several other cancers (lung cancer, breast cancer, etc.)^{6,12,14}, and Schymura et al. demonstrated that unmarried men are more likely to choose conservative treatment or RT rather than RP¹⁵, which is consistent with our results.

Cancer characteristics (ISUP GG, PPC, clinical T stage, PSA, etc.) contributed significantly to AS/WW or RP treatment decisions, in contrast to RT treatment decisions. Age was the most important feature in active treatment decision-making (first in RP and second in RT), likely because of its correlation with accompanying comorbidities and life expectancy¹⁶. Consistent with previous data demonstrating the impact of ethnicity on PCa treatment decisions using logistic regression models¹², we observed a racial/ethnicity difference in the initial treatment. However, its impact on the overall model performance was relatively low, ranging from the seventh to tenth ranked feature. Finally, the year of diagnosis was the most important feature in deciding AS/WW in

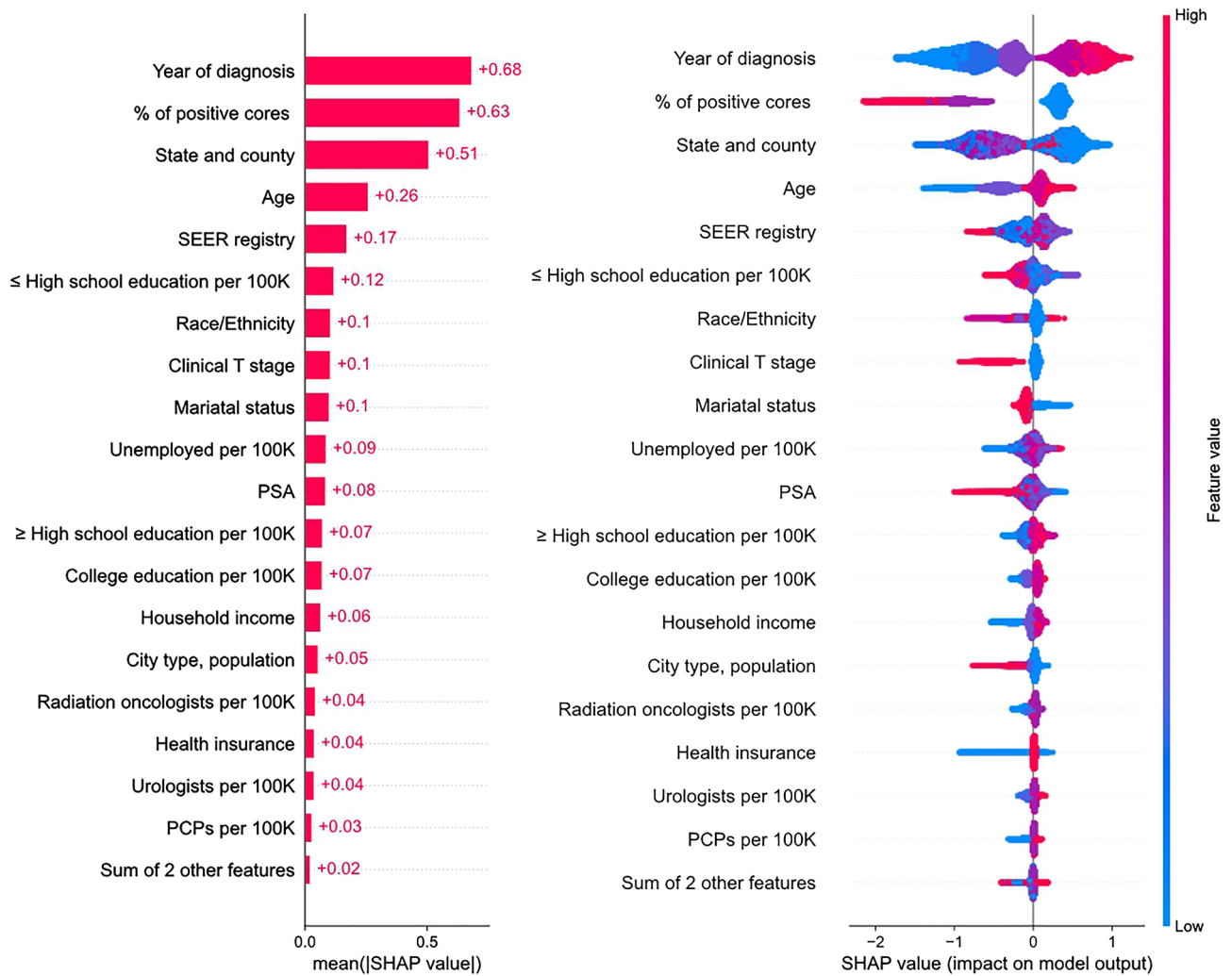


Figure 4. Global feature importance plot and Beeswarm plot for AS/WW decision in low-risk cohort using SHAP value.

the low-risk model, demonstrating the current time trend of an increasing proportion of AS/WW patients¹⁷. In other words, time trend reflects evolving evidence for AS/WW.

Several notable and solid relationship visualizations add further depth to this study. In the AS/WW treatment outcome model, the older (≥ 70 years) and younger age groups demonstrated different decision-making. The older group with higher oncological risk (high PSA or PPC or ISUP GG) chose AS/WW, while the younger group did not choose AS/WW in the same condition and underwent definitive treatment (Supplementary Fig. 2). This result is consistent with the overall feature contribution result that age was the most important feature in deciding active treatments (RP and RT). Furthermore, racial/ethnic differences were observed in the older age group. While older white men tended to pursue AS/WW, other races largely avoided it, which was more clearly identified in the low-risk cohort model (Supplementary Fig. 5). This may indicate that white men tend to emphasize treatment burden, while other races emphasize treatment efficacy/cure¹⁸, especially in the low-risk group. This cohort of patients was unlikely to progress and may not require radical treatment¹⁹. We also observed that patients who are presumed to have a large tumor burden (PPC $\geq 50\%$) avoided both RP and RT monotherapy treatment, and chose androgen-deprivation therapy or combinatory treatment when accompanied by other poor prognostic oncological features, such as high ISUP GG or high PSA.

With the aid of year of diagnosis variable annotation, we visually inspected the time trend of the associated variables. Although there was a trend of increasing AS/WW decision making, recently diagnosed patients chose not to undergo AS when high ISUP or PPC were present, which was in contrast to the previous period (Supplementary Fig. 2), possibly because of the recent publication of AS criteria²⁰. Regarding the active treatment groups, although both have been decreasing recently, patients with higher oncological risk (high ISUP GG, high PPC, or high PSA) more actively tended to decide on active treatment, which is consistent with previous data showing an increasing trend of RP in the intermediate-to-high-risk group¹⁷.

This study has several limitations. First, SEER database only covers 30% of US cancer patients, raising concerns about its generalizability. Furthermore, the SEER and Area Health Resource File (AHRF) dataset do not include the same patient groups, introducing potential issues during data integration. However, SEER database

is still the most well-established population-based epidemiologic cohort, and the AHRF data were linked to the county-level data where the patients reside within the SEER/WW dataset, utilizing a matching process based on the combined Federal Information Processing System (FIPS) codes for states and counties. To enhance the regional matching accuracy, year matching was also implemented. Second, the outcome of this predictive model represents the complicated treatment decision result; thus, it may be different from the decision prediction. Third, although data preprocessing was done to reduce the issue of feature imbalance, the potential clinical bias still remained. For example, as this cohort was developed in the active PSA screening era, the cohort distribution was heavily skewed towards earlier stage disease. Fourth, external validation was lacking. Fifth, there may be unidentified confounding factors in the analysis of the interaction dependence plot that hinder the ability to reach firm conclusions. Sixth, several important factors, such as life expectancy, preexisting urinary symptoms, genomic data (somatic and/or germline), family history, and types of health insurance could not be addressed due to the unavailability in SEER/WW dataset. Seventh, for patients classified into the observation group, WW or AS could not be discriminated, although they are completely different treatment approaches. To address this issue, we utilized explainable interaction plots to approximate the distinction between the two groups. And eighth, patient population were enrolled between 2010 and 2016, potentially preventing the reflection of the most recent trends. Also, there may be a drawback that the latest AS inclusion criteria were not applied to these patients. However, it is worth noting that in first decade of 2000, significant consensus was reached on the establishment of AS criteria and suggested as guideline, which is almost equivalent to the current standards²¹.

Despite these limitations, this study demonstrates important clinical implications. First, with the help of the explainable SHAP method and further descriptive interaction plots, we extracted a plausible description of each treatment decision in an orthogonal manner, leading to a comprehensive understanding and deeper insight. Second, our machine-learning-based model was trained on a large, contemporary, ethnically heterogeneous population using real-world data from a high-quality database²². Third, the potential power of this approach (offering decisions made by similar men) has been proven in a previous patient-led online community study by helping patients understand the decision-making process²³. Fourth, compared to a similar study performed by the MUSIC group, the predictive model in this study was fitted with additional non-oncological features, such as socioeconomic and geographical regional factors. Geographic features affected RT treatment decisions to the greatest degree.

Conclusions

Using a large population-based real-world database, we could have a deeper insight into the complex decision-making process and visualize nonlinear feature interactions in localized PCA.

Subjects and methods

Supplementary Fig. 7 provides a summary of the analysis. The experimental dataset utilized in this study consists of the clinical SEER/WW dataset and the regional County AHRF dataset. To improve the classification performance, preprocessing techniques were taken to address class imbalance, as well as to mitigate the presence of outliers or noisy data. The preprocessed dataset was then divided into training, validation, and test sets. The classification model was trained iteratively using the training set, aiming to identify the optimal machine learning algorithm and corresponding hyperparameters. The model's optimality was evaluated using the validation set. Finally, model explanation techniques were applied to gain insights into feature importance and interactions. The rest of this section includes more detailed information.

Study cohort. For the experimental analysis, we used the following two datasets: (1) the SEER/WW dataset (2010–2016) which had undergone multiple imputations for missing data handling²⁴, and (2) the AHRF, which includes data on variable health care-related features, such as healthcare providers by specialty, health facilities, population demographic characteristics, income, and hospital utilization⁶. We merged these two datasets in one final experimental dataset ($n = 255,837$). Supplementary Table 1 shows the labels encoding the correspondence between categorical variables, including nominal variables. Initial treatment variable was divided into four groups: AS/WW, RT, RP, and other/unknown treatments. We included men with concurrent RP and RT treatment into the other treatment group as a small number of RP + RT class could skew the whole class distributions. The low-risk cohort was refined to include patients with clinical T stage T1c and T2a, Gleason grade group 1, and PSA ≤ 10 ng/mL. Men aged > 80 years were excluded from the study, consistent with an appropriate patient population for AS⁶. The initial treatment features in the low-risk group were relabeled into two classes: AS/WW (AS + WW) and other treatments (Supplementary Fig. 1).

The Seoul National University Hospital institutional review board deemed this study exempt from review and informed consent because patient information in these databases was completely de-identified and publicly available.

Data preprocessing. Oversampling and undersampling techniques were adopted to circumvent the class imbalance problem. MSMOTE (Modified SMOTE)²⁵ as an oversampling method, was primarily applied to consider the distribution of minority class instances and remove noisy instances. For the undersampling method, we used the edited nearest neighbor (ENN)²⁶, which is based on a nearest-neighbor algorithm to remove samples whose class differs from the majority class of their neighborhood in a broad sense.

Model development. We split the dataset into training (70%), validation (15%), and test (15%) sets. The validation set was allocated to search for appropriate hyperparameters of the model. We trained three different gradient boosting decision tree algorithm-based ensemble method models, eXtreme Gradient Boosting

(XGBoost)²⁷, LightGBM²⁸, and CatBoost²⁹, and found that the XGBoost model had the highest explanatory power in our experimental setting.

Model validation. After fitting the XGBoost model using the training set, the discrimination of the model was evaluated in the test set using a multiclass area under the curve (AUC) measure. Model calibration was evaluated using a calibration plot that compared the predicted classes for each outcome with the observed classes.

Model explanation. SHapley additive exPlanations (SHAP)³⁰ serves two functions. (1) To demonstrate the contribution of each feature to the model's overall prediction results, visualizing with the global feature importance and Beeswarm plots. (2) To unravel the nonlinear complicated relationship between two principal features using a two-way dependence scatter plot and make it possible to gain deeper insight into the complex decision process.

Development of the web platform. We developed the Proca (Prostate Cancer treatment Advisor; Slogan: Friendly advice from patients like you) web platform based on Streamlit, an open-source Python library for building data apps. The web platform shows the initial treatment prediction for a patient using the given inputs. The inputs consisted of demographic and clinical information that appeared during the experiment. Furthermore, the web platform presents several important factors that influence the prediction. This tool is predictive of patient choice in a given scenario, so would suggest what other people have chosen. This is not about what each patient should do for their localized PCa.

Software. R version 3.6 to perform multiple imputation and statistical analysis for the completion of the experimental dataset. Python 3.8.10, and XGBoost 1.4.2 were used for the model algorithm. Train/test split, shuffle, and other data preprocessing procedures were conducted using Scikit-learn 0.24.2. Smote-variants 0.4.0 were used to address the class imbalance problem. Finally, the shap 0.39.0 library was employed for model interpretation.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 21 December 2022; Accepted: 4 July 2023

Published online: 17 July 2023

References

- Siegel, D. A., O'Neil, M. E., Richards, T. B., Dowling, N. F. & Weir, H. K. Prostate cancer incidence and survival, by stage and race/ethnicity—United States, 2001–2017. *MMWR Morb. Mortal. Wkly. Rep.* **69**, 1473–1480. <https://doi.org/10.15585/mmwr.mm6941a1> (2020).
- Fraser, M. *et al.* Genomic hallmarks of localized, non-indolent prostate cancer. *Nature* **541**, 359–364. <https://doi.org/10.1038/nature20788> (2017).
- El-Haouly, A., Dragomir, A., El-Rami, H., Liandier, F. & Lacasse, A. Treatment decision-making in men with localized prostate cancer living in a remote area: A cross-sectional, observational study. *Can. Urol. Assoc. J.* **15**, E160–E168. <https://doi.org/10.5489/cuaj.6521> (2021).
- Wang, E. H. *et al.* Shared decision making and use of decision AIDS for localized prostate cancer: Perceptions from radiation oncologists and urologists. *JAMA Intern. Med.* **175**, 792–799. <https://doi.org/10.1001/jamainternmed.2015.63> (2015).
- Hoffman, K. E. *et al.* Physician variation in management of low-risk prostate cancer: A population-based cohort study. *JAMA Intern. Med.* **174**, 1450–1459. <https://doi.org/10.1001/jamainternmed.2014.3021> (2014).
- Washington, S. L. 3rd. *et al.* Regional variation in active surveillance for low-risk prostate cancer in the US. *JAMA Netw. Open* **3**, e2031349. <https://doi.org/10.1001/jamanetworkopen.2020.31349> (2020).
- Vromans, R. D. *et al.* Communicative aspects of decision aids for localized prostate cancer treatment - A systematic review. *Urol. Oncol.* **37**, 409–429. <https://doi.org/10.1016/j.urolonc.2019.04.005> (2019).
- Riikonen, J. M. *et al.* Decision aids for prostate cancer screening choice: A systematic review and meta-analysis. *JAMA Intern. Med.* **179**, 1072–1082. <https://doi.org/10.1001/jamainternmed.2019.0763> (2019).
- Linardatos, P., Papastefanopoulos, V. & Kotsiantis, S. Explainable AI: A review of machine learning interpretability methods. *Entropy* **23**, 200. <https://doi.org/10.3390/e23010018> (2020).
- Cox, J. & Amling, C. L. Current decision-making in prostate cancer therapy. *Curr. Opin. Urol.* **18**, 275–278. <https://doi.org/10.1097/MOU.0b013e3282fba5f2> (2008).
- Auffenberg, G. B. *et al.* askMUSIC: Leveraging a clinical registry to develop a new machine learning model to inform patients of prostate cancer treatments chosen by similar men. *Eur. Urol.* **75**, 901–907. <https://doi.org/10.1016/j.eururo.2018.09.050> (2019).
- Krupski, T. L., Kwan, L., Afifi, A. A. & Litwin, M. S. Geographic and socioeconomic variation in the treatment of prostate cancer. *J. Clin. Oncol.* **23**, 7881–7888. <https://doi.org/10.1200/JCO.2005.08.755> (2005).
- Johansson, U., Sonstrod, C., Norinder, U. & Boström, H. Trade-off between accuracy and interpretability for predictive in silico modeling. *Future Med. Chem.* **3**, 647–663. <https://doi.org/10.4155/fmc.11.23> (2011).
- Harlan, L. C. *et al.* Factors associated with initial therapy for clinically localized prostate cancer: Prostate cancer outcomes study. *J. Natl. Cancer Inst.* **93**, 1864–1871. <https://doi.org/10.1093/jnci/93.24.1864> (2001).
- Schymura, M. J. *et al.* Factors associated with initial treatment and survival for clinically localized prostate cancer: Results from the CDC-NPCR Patterns of Care Study (PoC1). *BMC Cancer* **10**, 152. <https://doi.org/10.1186/1471-2407-10-152> (2010).
- Jeldres, C. *et al.* Survival after radical prostatectomy and radiotherapy for prostate cancer: A population-based study. *Can. Urol. Assoc. J.* **3**, 13–21 (2009).
- Gray, P. J., Lin, C. C., Cooperberg, M. R., Jemal, A. & Efstathiou, J. A. Temporal trends and the impact of race, insurance, and socioeconomic status in the management of localized prostate cancer. *Eur. Urol.* **71**, 729–737. <https://doi.org/10.1016/j.eururo.2016.08.047> (2017).

18. Xu, J., Janisse, J., Ruterbusch, J., Ager, J. & Schwartz, K. L. Racial differences in treatment decision-making for men with clinically localized prostate cancer: A population-based study. *J. Racial Ethn. Health Disparities* **3**, 35–45. <https://doi.org/10.1007/s40615-015-0109-8> (2016).
19. Bryant, R. J. *et al.* The ProtecT trial: Analysis of the patient cohort, baseline risk stratification and disease progression. *BJU Int.* **125**, 506–514. <https://doi.org/10.1111/bju.14987> (2020).
20. Liu, J. L., Patel, H. D., Haney, N. M., Epstein, J. I. & Partin, A. W. Advances in the selection of patients with prostate cancer for active surveillance. *Nat. Rev. Urol.* **18**, 197–208. <https://doi.org/10.1038/s41585-021-00432-w> (2021).
21. Han, C. S., Parihar, J. S. & Kim, I. Y. Active surveillance in men with low-risk prostate cancer: Current and future challenges. *Am. J. Clin. Exp. Urol.* **1**, 72–82 (2013).
22. Lee, C. *et al.* Application of a novel machine learning framework for predicting non-metastatic prostate cancer-specific mortality in men using the Surveillance, Epidemiology, and End Results (SEER) database. *Lancet Digit. Health* **3**, e158–e165. [https://doi.org/10.1016/S2589-7500\(20\)30314-9](https://doi.org/10.1016/S2589-7500(20)30314-9) (2021).
23. Wicks, P. *et al.* Sharing health data for better outcomes on PatientsLikeMe. *J. Med. Internet Res.* **12**, e19. <https://doi.org/10.2196/jmir.1549> (2010).
24. Jeong, C. W. *et al.* The new surveillance, epidemiology, and end results prostate with watchful waiting database: Opportunities and limitations. *Eur. Urol.* **78**, 335–344. <https://doi.org/10.1016/j.eururo.2020.01.009> (2020).
25. Hu, S., Liang, Y., Ma, L. & He, Y. In *2009 Second International Workshop on Computer Science and Engineering* 13–17 (2009).
26. Wilson, D. L. Asymptotic properties of nearest neighbor rules using edited data. *IEEE Trans. Syst. Man Cybern.* **3**, 408–421 (1972).
27. Friedman, J. H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **29**, 1189–1232 (2001).
28. Ke, G. *et al.* Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **30**, 3146–3154 (2017).
29. Drogush, A. V., Ershov, V. & Gulin, A. CatBoost: gradient boosting with categorical features support. arXiv preprint [arXiv:1810.11363](https://arxiv.org/abs/1810.11363) (2018).
30. Lundberg, S. M. & Lee, S.-I. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* 4768–4777.

Author contributions

C.W.J. had full access to all data in the study and takes responsibility for the integrity of the data and accuracy of the data analysis. Concept and design: C.W.J., M.R.C. Acquisition of data, Analysis and Interpretation of data: All authors Initial drafting of the manuscript: J.H.H., S.L., B.L., O.-k.B., C.W.J. Critical revision of the manuscript for important intellectual content: All authors. Statistical analysis: C.W.J., S.L.W., M.R.C. Obtained funding: C.W.J. Administrative, technical, and material support: P.R.C., M.R.C. Supervision: C.W.J., P.R.C., M.R.C.

Funding

This study was supported by a grant from the National R&D Program for Cancer Control, Ministry of Health and Welfare, Republic of Korea (HA17C0039), the Patient-Centered Clinical Research Coordinating Center funded by the Ministry of Health & Welfare, Republic of Korea (HI19C0481, HC19C0164), an Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government (22ZS1100, Core Technology Research for Self-Improving Integrated Artificial Intelligence System, and the New Faculty Startup Fund of Seoul National University (800-20200267). Neither sponsor had access to the data or any influence on or access to the analysis plan, results, or manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-38162-1>.

Correspondence and requests for materials should be addressed to C.W.J.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023