



Article

Prediction System for Prostate Cancer Recurrence Using Machine Learning

Sun Jung Lee ^{1,2}, Sung Hye Yu ¹, Yejin Kim ³, Jae Kwon Kim ¹, Jun Hyuk Hong ⁴, Choung-Soo Kim ⁴, Seong Il Seo ⁵, Seok-Soo Byun ⁶, Chang Wook Jeong ⁷, Ji Youl Lee ⁸ and In Young Choi ^{1,2,*}

- Department of Medical Informatics, College of Medicine, Catholic University of Korea, Seoul 06591, Korea; lsjadd500@gmail.com (S.J.L.); sala5537@naver.com (S.H.Y.); jaekwonkorea@naver.com (J.K.K.)
- Department of Biomedicine & Health Sciences, College of Medicine, The Catholic University of Korea, Seoul 06591, Korea
- School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, TX 77030, USA; vejin.kim89@gmail.com
- Department of Urology, University of Ulsan College of Medicine, Seoul 05505, Korea; jhhong@amc.seoul.kr (J.H.H.); cskim@amc.seoul.kr (C.-S.K.)
- Department of Urology, Sungkyunkwan University School of Medicine, Seoul 06351, Korea; siseo@skku.edu
- Department of Urology, Seoul National University Bundang Hospital, Seongnam 13620, Korea; seoksoo.byeon@gmail.com
- Department of Urology, Seoul National University College of Medicine, Seoul 03080, Korea; drboss@gmail.com
- Department of Urology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul 06591, Korea; uroljy@catholic.ac.kr
- * Correspondence: iychoi@catholic.ac.kr; Tel.: +82-2-2258-7648

Received: 8 January 2020; Accepted: 13 February 2020; Published: 16 February 2020



Abstract: Prostate cancer is the fourth most common cancer affecting South Korean males, and the biochemical recurrence (BCR) of prostate cancer occurs in approximately 25% of patients five years after radical prostatectomy. The ability to predict BCR would help clinicians and patients to make better treatment decisions. Therefore, in this study, we have proposed a web-based clinical decision support system that predicts the BCR of prostate cancer in Korean patients. The data were obtained from the Korean Prostate Cancer Registry (KPCR) database, which contained information about 7394 patients with prostate cancer who were treated at one of the six major medical institutions in South Korea between May 2001 and December 2014. We tested 13 prediction models and selected the gradient boosting classifier because it demonstrated excellent prediction performance. Using this model, we were able to create a web application and once clinical data from patients were entered, the three- and five-year post-surgery BCR predictions could be extracted. We developed a clinical decision support system to provide a prostate cancer BCR predictive function to facilitate postoperative follow-up and clinical management. This system will help clinicians develop a strategic approach for prostate cancer treatment by predicting the likelihood of prostate cancer recurrence.

Keywords: prostate cancer; machine learning; prediction; clinical decision support system; gradient boost

1. Introduction

According to a 2017 Korean National Cancer Center survey, prostate cancer is the fourth most common cancer affecting South Korean males, with an incidence rate of 29 per 100,000 people in the country's population [1].

Appl. Sci. 2020, 10, 1333 2 of 9

Biochemical recurrence (BCR) of prostate cancer occurs in approximately 25% of patients within five years after radical prostatectomy [2]. BCR is accepted as evidence of cancer recurrence if measurable serum prostate-specific antigen (PSA) levels are confirmed after radical prostatectomy, or if PSA levels increase after radiation treatments [3–5]. BCR is also accepted as an indicator of postoperative progress and outcomes [6]. If we could predict prostate cancer BCR by analyzing data on the progress and recurrence patterns of prostate cancers, it would help clinicians to make better decisions with regard to future treatment options. In addition, depending on the predicted BCR probability, the incidence of recurrence could be reduced by providing patients with physical examinations or related medication.

Most medical institutions use computer-based systems to accurately and efficiently manage medical data [7]. In particular, the clinical decision support system (CDSS) is a valuable tool that helps healthcare providers to make decisions and solve complex problems [8,9]. There are few prostate cancer recurrence prediction systems currently in clinical use, so providing this function, through a facility's CDSS, will facilitate postoperative follow up and aid in the clinical management of patients with prostate cancer [10].

The Memorial Sloan Kettering Cancer Center (MSKCC) in the United States provides a function that can be used to predict the probability of 2-, 5-, 7-, 10-, and 15-year BCR-free survival after prostate cancer surgery. It applies statistical techniques such as linear regression, logistical regression, and survival progress models to provide a web-based cancer recurrence prediction system [11]. However, the incidence of prostate cancer varies with country and race [12–15], which means that the predictive system that was designed for MSKCC patients may not be as accurate at predicting prostate cancer recurrence in South Korea.

Therefore, in this study, we have proposed a web-based CDSS that can predict the BCR of Korean patients with prostate cancer to support radical prostatectomy postoperative care. Our study uses data from Korean patients with prostate cancer to predict the recurrence of prostate cancer in this population. We tested several statistical and machine learning techniques that could be used to predict prostate cancer recurrence and then adopted the model that exhibited the best prediction accuracy for use in our prostate cancer recurrence prediction system.

This study aimed to develop a machine learning model that could use the Korean Prostate Cancer Registry (KPCR) database to predict the three- and five-year BCR in Korean patients who previously had prostate cancer and had undergone radical prostatectomies.

2. Materials and Methods

2.1. BCR Prediction Data

The data used in this study was obtained from the KPCR database which included 7394 patients with prostate cancer. This database included Electronic Medical Records (EMRs) data collected from the six major medical institutions in South Korea between May 2001 and December 2014 (Figure 1). The study protocol was approved by the institutional review board of the Catholic University of Korea (IRB No. MC16RIMI0107).

Appl. Sci. 2020, 10, 1333 3 of 9

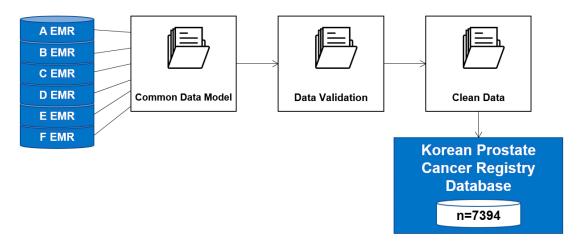


Figure 1. The Process of Building a Korean Prostate Cancer Registry (KPCR) Database Based on Electronic Medical Records (EMRs).

Data standardization and quality control were implemented to ensure data integrity, and exclusion criteria were created and applied to the database to refine the data used in the BCR prediction.

Our data subset was created from the main KPCR database using the following processes (Figure 2):

- There were 7394 potentially relevant records in the KPCR database
- 2280 records, including those of non-Koreans (51), those who did not have a full follow-up period of at least one year (633), those who did not have a neoclassical prostatectomy (80), those who performed pre-supplementary therapy (230), and those for whom critical data were missing (1286) were excluded
- We identified 5114 individual records as being statistically usable and relevant
- We classified the 5114 patient data items into BCR (1207) and Non-BCR (3907) groups to identify the characteristics of each group.

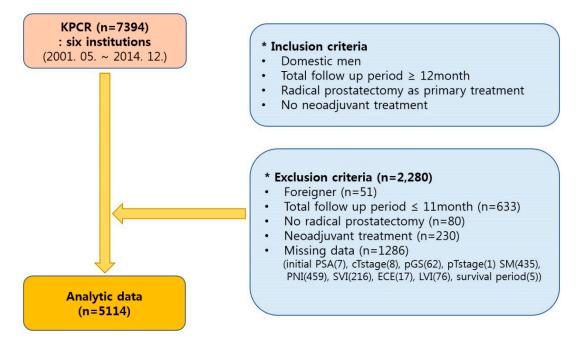


Figure 2. Patient Flow Chart.

Appl. Sci. 2020, 10, 1333 4 of 9

We found that follow-up periods ranged between 0 and 157 months, and out of the total of 7394 patients, 70% or more had a follow-up period of less than five years. Therefore, we decided to predict BCR likelihood at three and five years after radical prostate cancer surgery.

Factors affecting BCR have been reported to include age, initial PSA, clinical T stage, pathology Gleason score sum, pathology T stage, surgical margin, perineural invasion, seminal vesicles, extracapsular extension, and lymphovascular invasion [16–19]. We extracted these factors from the 5114 analytical data items and generated a training data set for predicting three- and five-year BCR after radical prostate cancer surgery.

2.2. BCR Prediction Statistical Analysis

In this study, we tested 13 prediction models, including: decision tree [20]; random forest [21] with 20, 50, and 80 trees; a fully-connected feedforward neural network [22]; Cox proportional hazards model [23]; ridge regression [24]; lasso [25]; and gradient boosting classifier (GBC), and analyzed the results produced by each model. GBC iteratively combines weak prediction models to create a more powerful model by allowing optimization of the loss function [26–28].

To select the model with the best BCR prediction performance, we evaluated the area under the receiver operating characteristics curve (AUC), accuracy, sensitivity, specificity, and Matthews correlation coefficient (MCC) for each model. We then selected the analytical model that performed the best and applied it to the BCR prediction system.

2.3. BCR Prediction System

There was a need for a system that can analyze the database of patients with prostate cancer and review the results to make predictions about prostate cancer recurrence. This system can then be used to analyze data from a new patient and predict the probability of recurrence. Based on the results of the model, clinicians can select treatment options to reduce the incidence of BCR. This system was developed as a web-based application so clinicians could access it easily.

In this study, we proposed a system that used a prediction algorithm to analyze the KPCR database in order to develop a BCR prediction model using data collected from patients with prostate cancer.

3. Results

3.1. Patient Characteristics and Distribution

We analyzed the characteristics and distribution of patients by dividing them into groups with prostate cancer BCR and non-recurrence of prostate cancer (Non-BCR) (Table 1). The average age of the 5114 patients with prostate cancer was 66 years, and the average ages for BCR and Non-BCR were both 65 years, respectively. The oldest and youngest BCR patients were 84 and 38, respectively, while the oldest and youngest Non-BCR patients were 90 and 37, respectively. The average initial PSA value for all patients was 11.59 ng/mL, and the mean PSA values for BCR and Non-BCR patients were 17.58 ng/mL and 9.74 ng/mL, respectively.

Appl. Sci. 2020, 10, 1333 5 of 9

| | | | Non-BCR (3907 Patients) | | | |
|---------------------|---------|-------|-------------------------|--|--|--|
| Age | Average | 65 | 65 | | | |
| Initial PSA (ng/mL) | Average | 17.58 | 9.74 | | | |
| Gleason score (sum) | 2–4 | 0 | 5 | | | |
| | 5 | 3 | 19 | | | |
| | 6 | 111 | 994 | | | |
| | 7 | 719 | 2576 | | | |
| | 8–10 | 374 | 313 | | | |
| Clinical T stage | T1 | 399 | 1753 | | | |
| | T2 | 467 | 1416 | | | |
| | Т3 | 276 | 677 | | | |
| | T4 | 65 | 61 | | | |

Table 1. Patient Characteristics and Distribution.

For BCR patients, the initial PSA minimum value was 0.66, and the maximum was 261.77. For Non-BCR patients, the initial PSA minimum was 0.09, while the maximum was 305. Pathological Gleason scores were classified as grade 2–4, 5, 6, 7, and 8–10, and within each grade, 5 (0.09%), 22 (0.43%), 1105 (21%), 3295 (64%), and 687 (13.43%) patients were included, respectively. We classified clinical T stages into four of the existing eight grades [29]: 2152 for T1, 1883 for T2, 953 for T3, and 126 for T4. The BCR patients were widely distributed within T2, and the non-BCR patients were widely distributed within T1.

3.2. BCR Prediction Model Analysis

In order to analyze the statistical performance of the BCR predictions, we evaluated the AUC, accuracy, sensitivity, specificity, and MCC of each analysis method. The analytical data was divided randomly, with 80% of the data allocated to the training dataset and 20% allocated to the test dataset. We trained the model using the training data, and then we evaluated the performance of the model using the test data.

According to the result of analysis (Table 2), the highest AUC value for the three year predictions was 0.8419, which was associated with the GBC model. For the five year predictions, the highest AUC value (0.8071) was associated with the Ridged Regression model. The highest MCC values for the three year and five year predictions were associated with the Random Forest (ntrees = 80) model (0.4621) and the GBC model (0.4836), respectively. The Cox proportional hazard model showed good results in terms of accuracy and specificity but performed poorly in terms of sensitivity, and so was not used as a predictive model. In terms of overall performance across the five analysis criteria (AUC, accuracy, sensitivity, specificity, MCC), the GBC model had the highest average value and showed excellent performance. Therefore, we applied this model to develop the BCR prediction system. To develop prediction models for three and five years after prostate cancer surgery using the GBC model, we developed the BCR prediction system using Python.

Appl. Sci. 2020, 10, 1333 6 of 9

| Analysis | Method | AUC | | Accuracy | | Sensitivity | | Specificity | | MCC | |
|------------------------|--|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|---------------------------------------|
| | Wichiou | 3-year | 5-year |
| Trees | Decision Tree Random Forest (ntrees = 20) ¹ Random Forest (ntrees = 50) Random Forest (ntrees = 80) | 0.8186 0.8320 0.8349 0.8362 | 0.7561 0.7956 0.8047 0.8050 | 0.7817 0.7817 0.7832 0.7876 | 0.7033 0.7126 0.7266 0.7196 | 0.4135 0.4135 0.4279 0.4327 | 0.7009 0.7308 0.7393 0.7308 | 0.9447 0.9447 0.9404 0.9447 | 0.7062 0.6907 0.7113 0.7062 | 0.4448 0.4448 0.4495 0.4621 | 0.4055 0.4210 0.4310 0.4360 |
| Neural Networks | 1hidden, dropout = 0.3 at input 1hidden dropout = 0.1 at input 2hidden, dropout = 0.3 at input 2hidden dropout = 0.1 at input | 0.7939 0.8027 0.7978 0.7988 | 0.7895 0.7977 0.7941 0.7967 | 0.7611 0.7699 0.6932 0.7611 | 0.7056 0.6916 0.7056 0.7056 | 0.3606 0.5144 0.0000 0.3702 | 0.7821 0.7393 0.7051 0.8162 | 0.9383 0.8830 1.0000 0.9340 | 0.6134 0.6340 0.7062 0.5722 | 0.4385 0.4089 0.3393 0.3963 | 0.4001 -0.0367 0.3986 0.4128 |
| Survival Regression | Cox PH | 0.7944 | 0.7816 | 0.8025 | 0.7830 | 0.2160 | 0.3333 | 0.9568 | 0.9300 | 0.2894 | 0.3159 |
| Logistic Regression | Ridged Regression (L2) Lasso (L1) GBC | 0.8288 0.8319 0.8419 | 0.8071 0.7993 0.8031 | 0.7802 0.7861 0.7891 | 0.7290 0.7243 0.7407 | 0.4279 0.4519 0.4567 | 0.7265 0.7179 0.7436 | 0.9362 0.9340 0.9362 | 0.7320 0.7320 0.7371 | 0.4413 0.4590 0.4348 | 0.4210 0.4481 0.4836 |

Table 2. Statistical Method Analysis.

3.3. Development of BCR Prediction System

The BCR prediction system used data from Korean patients with prostate cancer to implement a web-based application. The structure of the prediction system was classified into two parts: predictive modeling and web application development (Figure 3). The web application used the predictive model to make predictions.

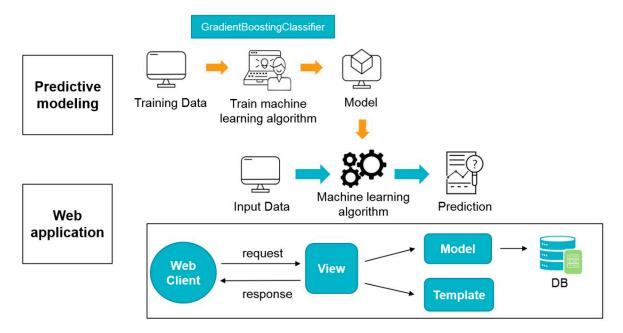


Figure 3. System Flow.

The BCR prediction system was set up to perform several functions on one screen, including reading the data list of patients with prostate cancer, inputting the input values, and extracting the prediction results (Figure 4). After the input fields, including age at diagnosis, initial PSA, clinical T stage, pathological Gleason score, and pathological T stage values have been entered and status information such as surgical margin, perineural invasion, seminal vesicle invasion, extracapsular extension, and lymphovascular invasion has been selected, the three and five year post-surgery BCR predictions could be extracted. The prediction results appear as an output item at the bottom of the screen.

¹ "ntrees" means the number of trees in the forest.

Appl. Sci. 2020, 10, 1333

Patient Show Search: Age: 67 Initial psa: 4.1 entries Clinical t stage ID Initial_PSA Age 1a. 1b. 1c ○ 2a. 2b ○ 3a. 3b ○ 4 67 7 43 Pathology gleason score sum: 6 Pathology t stage restaging: 2 4.1 Surgical margin: 3.242 o yes o no 62 3.11 Perineural invasion: 71 4 61 Seminal vesicle: 72 10.45 o ves o no 63 2.58 Extracapsular extension: 4.5 ves no 60 5.61 Lymphovascular invasion: ⊚ yes ⊛ no 10 66 3.18 calculate clear Showing 1 to 10 of 100 entries Previous 2 3 4 5 ... 10 Next 3year BCR : 25.0 % 5year BCR: 47.0 %

Figure 4. BCR Prediction System.

4. Discussion

BCR Prediction

4.1. Follow-Up Period for Predicting BCR

Long-term follow-up is needed to identify prostate cancer recurrence. The current data includes follow-up periods of up to 13 years, although the average is just four years, which is short. As a result, our work was limited to predicting prostate cancer recurrence three and five years after radical prostate cancer surgery. If Korean prostate cancer data could be accumulated over a longer time period, the BCR prediction system developed here could be applied to predicting the likelihood of cancer recurrence more than 10 years after surgery.

4.2. Prostate Cancer Patients Data and Data Provider Characteristics

The data on patients with prostate cancer used in this study were collected from six Korean agencies. Although this is not a large number of institutions, they are large hospitals that would have been visited by most cancer patients in South Korea. These institutions are therefore considered to have provided representative datasets, and it was assumed that institutional dataset characteristics were unlikely to exhibit significant differences.

The analytical classifier used in this study is specific to the Korean prostate cancer dataset. There are racial differences in the recurrence of prostate cancer, and since we currently only have a dataset that represents Korean patients, we can only develop a specific predictive model that applies to that population. We would, therefore, have difficulty predicting recurrence in other races. If we can collect data from other racial populations in the future, we will be able to develop more general models that can predict recurrence in various groups of people.

4.3. The Potential of Misdiagnosis

The CDSS developed in this study can help clinicians determine treatment options by predicting the probability of prostate cancer recurrence. However, there was an error rate of about 20 percent in

Appl. Sci. 2020, 10, 1333 8 of 9

this study, which means that there is a potential for misdiagnosis. The CDSS should be used as an ancillary tool, and, in the future, measures to improve data quality and the accuracy of predictive models should be investigated to.

5. Conclusions

In this study, we designed and developed a CDSS that could predict the likelihood of BCR in Korean patients with prostate cancer. Performance-stable GBC predictive models were used, and the KPCR, a database of Korean patients with prostate cancer, was used for training to develop BCR prediction models that were specialized for use with Korean data. This prediction model was incorporated into a web-based system to allow easy access to data from patients with prostate cancer and facilitate its use for BCR predictions. The BCR prediction could facilitate postoperative follow-up and clinical management of patients with prostate cancer. This system will be able to help clinicians develop a strategic approach by predicting the likelihood that prostate cancer will recur in a particular patient.

Author Contributions: Conceptualization, S.J.L., S.H.Y., Y.K., and I.Y.C.; Methodology, S.J.L., Y.K., and I.Y.C.; Software, S.J.L. Formal Analysis, S.J.L.; Investigation, S.J.L, S.H.Y., Y.K., J.H.H., C.-S.K., S.I.S., C.W.J., S.-S.B., J.Y.L., and I.Y.C.; Data Curation, S.H.Y.; Writing—Original Draft Preparation, S.J.L.; Writing—Review & Editing, Y.K., J.K.K., and I.Y.C.; Visualization, S.J.L.; Supervision, I.Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2019R1A5A2027588 and No. 2016R1A2B4015922).

Acknowledgments: We thank Hyunmin Kim for reviewing the abstract and providing valuable comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. National Cancer Information Center. 2017 Cancer Statistics; Cancer Incidence Rates. Available online: https://www.cancer.go.kr/lay1/S1T639C641/contents.do (accessed on 6 January 2020).
- 2. Lee, J.W.; Cho, K.S.; Han, K.S.; Kim, E.K.; Joung, J.Y.; Seo, H.K.; Chung, J.; Park, W.S.; Lee, K.H. Epidermal growth factor receptor as predicting factor on biochemical recurrence after radical prostatectomy: A prospective study. *Korean J. Urol.* 2008, 49, 974–980. [CrossRef]
- 3. Song, S.H.; Kwak, C. Definition of BCR and further evaluation. Korean J. Urol. Oncol. 2011, 9, 104–111.
- 4. Cox, J.D.; Grignon, D.J.; Kaplan, R.S.; Parsons, J.T.; Schellhammer, P.F. Consensus statement: Guidelines for PSA following radiation therapy. *Int. J. Radiat. Oncol. Biol. Phys.* **1997**, *37*, 1035–1041.
- 5. Lange, P.H.; Ercole, C.J.; Lightner, D.J.; Fraley, E.E.; Vessella, R. The value of serum prostate specific antigen determinations before and after radical prostatectomy. *J. Urol.* **1989**, 141, 873–879. [CrossRef]
- 6. Hong, J.H.; Lee, H.M.; Choi, H.Y. The predictors of biochemical recurrence and metastasis following radical perineal prostatectomy in clinically localized prostate cancer. *Korean J. Urol.* **2005**, *46*, 1161.
- 7. Sa, S. Intelligent heart disease prediction system using data mining techniques. *Int. J. Healthc. Biomed. Res.* **2013**, *1*, 94–101.
- 8. Paydar, K.; Kalhori, S.R.N.; Akbarian, M.; Sheikhtaheri, A. A clinical decision support system for prediction of pregnancy outcome in pregnant women with systemic lupus erythematosus. *Int. J. Med Inform.* **2017**, 97, 239–246. [CrossRef]
- Berner, E.S. (Ed.) Clinical Decision Support Systems; Springer Science + Business Media, LLC: New York, NY, USA, 2007.
- Foster, D.; McGregor, C.; El-Masri, S. A survey of agent-based intelligent decision support systems to support clinical management and research. In Proceedings of the Second International Workshop on Multi-Agent Systems for Medicine, Computational Biology, and Bioinformatics, Utrecht, The Netherlands, 25 July 2005; pp. 16–34.
- 11. Memorial Sloan Kettering Cancer Center. Prostate Cancer Nomograms. Available online: https://www.mskcc.org/nomograms/prostate (accessed on 6 January 2020).
- 12. National Cancer Information Center. 2017 Cancer Statistics; Global Cancer Incidence. Available online: https://www.cancer.go.kr/lay1/S1T639C644/contents.do (accessed on 6 January 2020).

Appl. Sci. 2020, 10, 1333 9 of 9

13. Hoffman, R.M.; Gilliland, F.D.; Eley, J.W.; Harlan, L.C.; Stephenson, R.A.; Stanford, J.L.; Albertson, P.C.; Hamilton, A.S.; Hunt, W.C.; Potosky, A.L. Racial and ethnic differences in advanced-stage prostate cancer: The prostate cancer outcomes study. *J. Natl. Cancer Inst.* **2001**, *93*, 388–395. [CrossRef]

- 14. Hankey, B.F.; Feuer, E.J.; Clegg, L.X.; Hayes, R.B.; Legler, J.M.; Prorok, P.C.; Ries, L.A.; Merrill, R.M.; Kaplan, R.S. Cancer surveillance series: Interpreting trends in prostate cancer—Part I: Evidence of the effects of screening in recent prostate cancer incidence, mortality, and survival rates. *J. Natl. Cancer Inst.* 1999, 91, 1017–1024. [CrossRef]
- 15. Stanford, J.L.; Stephenson, R.A.; Coyle, L.M.; Cerhan, J.; Correa, R.; Eley, J.W.; Gilliland, F.; Hankey, B.; Kolonel, L.N.; Ross, R. *Prostate Cancer Trends* 1973–1995; NIH Pub. No. 99-4543; Surveillance Research Program, National Cancer Institute: Bethesda, MD, USA, 1999.
- 16. Ahn, K.I.; Chang, S.G.; Choi, B.K.; Jeon, S.H. The study of biochemical and clinical failure after radical prostatectomy or radiation therapy in localized prostate cancer. *Korean J. Urol. Oncol.* **2004**, *2*, 171–177.
- 17. De la Tallle, A.; Rubin, A.A.; Bagiella, E.; Olsson, C.A.; Buttyan, R.; Burchardt, T.; Knight, C.; O'Toole, K.M.; Katz, A.E. Can perineural invasion on prostate needle biopsy predict prostate specific antigen recurrence after radical prostatectomy? *J. Urol.* 1999, 162, 103–106. [CrossRef] [PubMed]
- 18. Blute, M.L.; Bergstralh, E.J.; Iocca, A.; Scherer, B.; Zincke, H. Use of Gleason score, prostate specific antigen, seminal vesicle and margin status to predict biochemical failure after radical prostatectomy. *J. Urol.* **2001**, 165, 119–125. [CrossRef] [PubMed]
- 19. Han, M.; Partin, A.W.; Zahurak, M.; Piantadosi, S.; Epstein, J.I.; Walsh, P.C. Biochemical (prostate specific antigen) recurrence probability following radical prostatectomy for clinically localized prostate cancer. *J. Urol.* **2003**, *169*, 517–523. [CrossRef]
- 20. Pal, M.; Mather, P.M. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sens. Environ.* **2003**, *86*, 554–565. [CrossRef]
- 21. Liaw, A.; Wiener, M. Classification and regression by randomForest. R News 2002, 2, 18–22.
- 22. Acion, L.; Kelmansky, D.; van der Laan, M.; Sahker, E.; Jones, D.; Arndt, S. Use of a machine learning framework to predict substance use disorder treatment success. *PLoS ONE* **2017**, *12*, e0175383. [CrossRef]
- 23. Van Houwelingen, H.; Putter, H. *Dynamic Prediction in Clinical Survival Analysis*, 1st ed.; CRC Press: Boca Raton, FL, USA, 2011.
- Saunders, C.; Gammerman, A.; Vovk, V. Ridge regression learning algorithm in dual variables. In Proceedings
 of the Fifteenth International Conference on Machine Learning (ICML 1998), Madison, WI, USA, 24–27 July
 1998.
- 25. Meier, L.; Van De Geer, S.; Bühlmann, P. The group lasso for logistic regression. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2008**, 70, 53–71. [CrossRef]
- 26. Wang, J.; Li, P.; Ran, R.; Che, Y.; Zhou, Y. A short-term photovoltaic power prediction model based on the gradient boost decision tree. *Appl. Sci.* **2018**, *8*, 689. [CrossRef]
- 27. Friedman, J.H. Stochastic gradient boosting. Comput. Stat. Data Anal. 2002, 38, 367–378. [CrossRef]
- 28. Krauss, C.; Do, X.A.; Huck, N. Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *Eur. J. Oper. Res.* **2017**, 259, 689–702.
- 29. Ohori, M.; Wheeler, T.M.; Scardino, P.T. The new American joint committee on cancer and international union against cancer TNM classification of prostate cancer. *Cancer* **1994**, *74*, 104–114. [CrossRef]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).