Published in final edited form as:

Urol Oncol. 2022 April ; 40(4): 161.e1–161.e7. doi:10.1016/j.urolonc.2021.08.007.

A machine learning approach to predict progression on active surveillance for prostate cancer

Madhur Nayan, M.D.C.M., Ph.D.[a,*], **Keyan Salari, M.D., Ph.D.**[a,b], **Anthony Bozzo, M.D.C.M., M.Sc.**[c], **Wolfgang Ganglberger, M.Sc.**[d], **Gordan Lu, M.D.**[a], **Filipe Carvalho, M.D., Ph.D.**[a], **Andrew Gusev, M.D.**[a], **Adam Schneider, B.S.**[a], **Brandon M. Westover, M.D., Ph.D.**[d], **Adam S. Feldman, M.D., MPH**[a]

[a]Department of Urology, Massachusetts General Hospital, Boston, Massachusetts

[b]Broad Institute of Harvard and MIT, Cambridge, Massachusetts

[c]Division of Orthopaedic Surgery, Department of Surgery, McMaster University, Hamilton, Ontario, Canada

[d]Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts

Abstract

Purpose: Robust prediction of progression on active surveillance (AS) for prostate cancer can allow for risk-adapted protocols. To date, models predicting progression on AS have invariably used traditional statistical approaches. We sought to evaluate whether a machine learning (ML) approach could improve prediction of progression on AS.

Patients and Methods: We performed a retrospective cohort study of patients diagnosed with very-low or low-risk prostate cancer between 1997 and 2016 and managed with AS at our institution. In the training set, we trained a traditional logistic regression (T-LR) classifier, and alternate ML classifiers (support vector machine, random forest, a fully connected artificial neural network, and ML-LR) to predict grade-progression. We evaluated model performance in the test set. The primary performance metric was the F1 score.

Results: Our cohort included 790 patients. With a median follow-up of 6.29 years, 234 developed grade-progression. In descending order, the F1 scores were: support vector machine 0.586 (95% CI 0.579 – 0.591), ML-LR 0.522 (95% CI 0.513 – 0.526), artificial neural network 0.392 (95% CI 0.379 – 0.396), random forest 0.376 (95% CI 0.364 – 0.380), and T-LR 0.182 (95% CI 0.151 – 0.185). All alternate ML models had a significantly higher F1 score than the T-LR model (all $p < 0.001$).

Conclusion: In our study, ML methods significantly outperformed T-LR in predicting progression on AS for prostate cancer. While our specific models require further validation, we anticipate that a ML approach will help produce robust prediction models that will facilitate individualized risk-stratification in prostate cancer AS.

[*]Corresponding author. Tel.: 617-726-8078; fax: 617-643-8525, mnayan@mgh.harvard.edu (M. Nayan).

Supplementary materials

Supplementary material associated with this article can be found in the online version at https://doi.org/10.1016/j.urolonc.2021.08.007.

Keywords

Machine learning; Artificial intelligence; Prostatic neoplasm; Active surveillance; Progression

1. Introduction

The vast majority of patients undergoing screening for prostate cancer are diagnosed with very-low- or low-risk, localized disease [1,2]. In these disease risk categories, active surveillance (AS) has demonstrated excellent long-term outcomes [3,4] and is endorsed as the preferred initial management by several guidelines [5–7].

The objective of AS is timely identification of potentially aggressive disease to allow for a curative intervention, when necessary. Though variations exist in AS protocols, this objective is generally achieved through serial assessments involving physician visits, prostate specific antigen (PSA) testing, and repeat biopsies [8–10]. While the use of AS has reduced overtreatment for indolent disease, a rigorous AS protocol can be time-consuming for both the provider and patient and costly to the health-care system [11]. Furthermore, risks of prostate biopsy include bleeding, lower urinary tract symptoms, urinary retention, erectile dysfunction, and infection [12]. The infectious complications can be severe enough to warrant hospitalization and data indicate that the rates of infection may be increasing over time [13]. Finally, select men on AS may have anxiety related to the potential for deferred intervention of a diagnosed malignancy [14].

To date, studies that have developed models to predict progression on AS have invariably used traditional statistical approaches [8,15–17]. Recently, machine learning (ML) approaches have gained interest in medicine as they have demonstrated superior classification performance compared to traditional statistical approaches [18,19]. A robust model predicting progression on AS may allow for more selective use of serial assessments in those at lower risk of progression, and may reduce patient anxiety, potentially allowing for greater adoption. Therefore, the objective of this study was to evaluate whether a ML approach could improve prediction of disease progression on AS, compared to a traditional statistical approach.

2. Materials and methods

2.1. Study design

Under the institutional review board–approved protocol, we utilized our previously described database [20] to perform a retrospective cohort study of consecutive patients diagnosed with very-low or low-risk (PSA <10 ng/ml, clinical stage up to T2a, and International Society of Urological Pathology (ISUP) grade group 1 on diagnostic biopsy [5–7]) localized prostate cancer between 1997 and 2016 and managed with AS at our institution. While some patients meeting intermediate-risk criteria were managed with AS during this time, we focused the model development in those in whom the model is most likely to be used, namely those with very-low and low-risk disease. We excluded those without a confirmatory biopsy.

Our AS protocol has been previously described [20]; briefly, after protocol creation in 2008, the standard follow-up schedule consisted of PSA and digital rectal exam every 4 months for the first year, every 6 months for 2 years, and then annually thereafter. A repeat 12-core biopsy was generally performed at 12 to 18 months after diagnosis, with subsequent biopsies performed at the discretion of the treating physician. Before 2008, the treating physician determined the follow-up schedule. All pathology was reviewed by Massachusetts General Hospital genitourinary pathologists at time of initial consultation.

2.2. Model features

Features considered for inclusion in the models were characteristics measured at initial prostate cancer diagnosis and comprised of age, year of diagnosis, family history of prostate cancer, clinical tumor stage, PSA level, prostate volume based on transrectal ultrasound, PSA density, diagnostic biopsy characteristics (total number of cores, number of positive cores, maximum percentage of core involvement, and perineural invasion), and use of a 5-alpha reductase inhibitor for ≥6 months at the time of diagnosis. Additional features considered were time between initial diagnosis and most recent biopsy, and number of biopsies while on AS.

In those with a history of 5-alpha reductase inhibitor use for ≥6 months at the time of diagnosis, we adjusted the PSA by multiplying by a factor of two [21], and the adjusted PSA was used to estimate PSA density in these patients.

Given that study enrollment began prior to 2005, we anticipated a shift in grade classification and therefore included an additional feature that dichotomized the year of diagnosis into before 2005 vs. 2005 and after, as has been done previously [8].

We acknowledge the potential for multicollinearity among the features included. However, the primary objective of this study was to optimize classification performance, rather than inferring the relationship between individual features and the outcome, and ML models can be robust despite the presence of multicollinearity [22].

2.3. Outcome

The outcome of interest was grade-progression, defined as the presence of any ISUP grade group 2 or greater disease, on any subsequent biopsy.

2.4. Analysis

For this classification problem, we compared a traditional statistical approach, specifically traditional logistic regression with an automated variable selection strategy (T-LR), as has been used by others to develop models to predict progression on AS [8,15,16], with several alternate ML approaches that have been frequently used in healthcare research. The alternate ML models included support-vector machine, random forest, a fully-connected artificial neural network, and a ML version of logistic regression (ML-LR). The details of these models have been described elsewhere [23].

The sample was initially split into a training and test set with a ratio of 80% vs. 20% using stratified random sampling. The models were then developed in the training set, leaving an independent test set to evaluate performance.

Although there are variations on the specific strategy for automated variable selection in traditional statistical approaches [8,15,16], we *a priori* selected variables for inclusion in the multivariable LR model using backward elimination with a cut-off value of $p < 0.10$. For the ML models, all features were included in model development.

The performance metrics of the final resulting models were compared in the test set. Given the clinical priority of correctly predicting those with grade progression and the imbalanced nature of the dataset, whereby a minority of patients were in the positive class (i.e. had grade progression), the primary performance metric was the F1 score [24]. The F1 score is the harmonic mean of the positive-predictive value, also known as precision, and sensitivity, also known as recall, the formulas of which has been provided previously [24]. The F1 score can range from 0 to 1, with a value of 1 indicating perfect precision and recall. Other performance metrics evaluated included sensitivity, specificity, positive predictive value, negative predictive value, and the c-statistic. We constructed confusion matrices.

For the alternate ML models, we scaled continuous features in the training data to have a mean of 0 and standard deviation of 1. We tuned hyperparameters to optimize the F1 score using randomized search with 10-fold cross-validation. The hyperparameters considered and selected are shown in the Supplemental Methods.

Confidence intervals (CI) were estimated using the jack-knife procedure [25]. We compared the F1 score between the alternate ML and T-LR models using a one-way ANOVA with Tukey's Test.

All analyses were completed using Python (version 3.7) with modules from Scikit-learn and Tensorflow 2.

3. Results

3.1. Cohort characteristics

Between 1997 and 2016, 1268 patients were managed with AS at our institution. Of these, we excluded 273 who did not undergo a confirmatory biopsy and an additional 97 did not meet very-low or low-risk criteria. Of the remaining 898 patients, we excluded a further 108 patients for missing data on prostate volume, total number of biopsy cores at diagnosis, or use of 5-alpha reductase inhibitor at the time of diagnosis. Our final cohort consisted of 790 patients, details of which are provided in (Table 1). With a median follow-up of 6.29 (interquartile range 4.30-8.92) years, 234 patients were found to have grade progression.

The cohort was split into a training (80%) and test (20%) set and their characteristics are shown in (Table 1).

3.2. Traditional statistical approach

After applying backward selection, 4 variables were included in the final multivariable T-LR model and these were age, PSA density, maximum percentage of core involvement, and time between initial diagnosis and most recent biopsy.

3.3. Model performance

The F1 scores measured in the test set in descending order were: support vector machine 0.586 (95% CI 0.579 – 0.591), ML-LR 0.522 (0.513 – 0.526), 2-layer artificial neural network 0.392 (95% CI 0.379 – 0.396), random forest 0.376 (95% CI 0.364 – 0.380), and T-LR 0.182 (95% CI 0.151 – 0.185). All ML models had a significantly higher F1 score than the T-LR model (all P <0.001, Fig. 1).

Other model performance metrics of sensitivity, specificity, positive predictive value, negative predictive value, and c-statistic are shown in (Table 2).

Compared to the ML models, the T-LR model had the lowest sensitivity and negative predictive value, but highest specificity and positive predictive value. The c-statistic of the T-LR model was lower than the support vector machine and ML-LR, but higher than the random forest and artificial neural network.

Confusion matrices are shown in (Fig. 2).

4. Discussion

Robust prediction of progression on AS for prostate cancer can allow for risk-directed protocols, thereby reducing the burden and risks of AS, as well as potentially decreasing patient anxiety. This is of increasing importance since AS is being used more frequently as the initial management for localized prostate cancer [26]. ML methods have gained interest in healthcare research as they have demonstrated superior classification performance to traditional statistical approaches [18,19], but have not yet been studied in prostate cancer AS. In this study, we developed and evaluated various models to predict progression on AS and found that ML approaches outperformed a traditional statistical approach. These findings emphasize the potential value of ML methods in developing robust AS prediction models.

Several performance metrics have been proposed in the ML literature and the ideal metric is based on the priority for accurately predicting either the positive or negative class and the characteristics of the dataset [27]. We chose the F1 score as the primary performance metric given the clinical priority for accurately predicting the positive class (disease progression) and the known class imbalance in our dataset. While the alternate ML models had a higher F1 score than the T-LR model, the T-LR model had the highest specificity and positive predictive value. This can be explained by the model development process and the characteristics of the AS population. The T-LR model was developed in a traditional statistical approach to maximize overall accuracy. In the present study, a minority of patients had disease progression, representing an unbalanced dataset, and therefore the T-LR model overpredicted the negative class while underpredicting the positive (disease progression)

class, as demonstrated in the confusion matrices. This resulted in the T-LR having relatively higher specificity but lower sensitivity. In contrast, the ML models were developed with clinical guidance to prioritize prediction of the positive class by tuning the hyperparameters to optimize the F1 score, resulting in improved sensitivity.

The c-statistic is another metric that has been used to describe model performance. However, the c-statistic places equal importance on the positive and negative classes [28] and therefore these measurements can be overly optimistic in unbalanced datasets. Indeed, the c-statistic for the T-LR model was superior to the random forest and artificial neural network, despite the low sensitivity of the T-LR model. The support vector machine was the overall best model by F1 score and c-statistic.

To date, studies that have developed models to predict progression on AS have invariably used traditional statistical approaches. A study from Johns Hopkins developed a model to predict grade reclassification in patients managed with AS for very-low- and low-risk prostate cancer [8]. After applying a combination of forward and backward elimination, the c-statistic for the final model was 0.757. A more recent study developed a prediction model in those with grade group 1 disease from the Canary Prostate Active Surveillance Study and validated the model in a cohort from the University of California, San Francisco [16]. Variables in the final model were selected using forward selection and the c-statistic was 0.70 in both the development and validation cohort. Other models have been developed using similar traditional statistical approaches and resulting c-statistics [15,17].

Limitations of the models developed to date include the use of automated variable selection strategies and T-LR, as well as difficulty in interpreting classification performance based on the c-statistic. Disadvantages of automated variable selection strategies have been described elsewhere [29]; briefly, concerns include potentially selecting variables that may be unrelated to the outcome and inflating the association between each selected variable and the outcome. A concern of T-LR is that there are underlying assumptions of the relationship between the predictors and the outcome, which may not hold biologically [30].

More recently, ML methods are being used with increasing frequency in healthcare research [23]. Advantages of ML models include the ability to learn complex, non-linear relationships between predictors and the outcome, as well as being able to adjust the coefficients of the parameters to optimize model fit in a process known as regularization [30], with minimal human involvement. It should be noted that ML methods can be also applied to logistic regression, as was done in this study, to optimize the performance of this type of model. Despite the potential benefits of ML, traditional statistical methods may perform as well as ML methods in some contexts [31,32], and these scenarios cannot be predicted in advance.

While the performance of our ML models, as determined by the c-statistic, is in line with the previous models that have been developed using a traditional statistical approach, it should be noted that the c-statistic is dependent on the variables included in the model as well as the characteristics of the population evaluated [33]. Therefore, it is difficult to directly compare our c-statistic with those of others. Nonetheless, we did not achieve a robust prediction

model to predict progression on AS within our sample and available data. Rather, our study demonstrates the potential value of ML methods to improve predicting progression on AS, compared to traditional statistical approaches.

This study is limited by the lack of MRI data. Although MRI has been shown to be of potential value in selecting patients for AS, routine use is not yet endorsed and therefore a model utilizing this data would limit generalizability. Nonetheless, it is anticipated that combining MRI features through a convolutional neural network with clinical data would augment the performance of a model predicting AS progression. Furthermore, although model performance was evaluated on data that had not been used for model development, external validation of our results is needed as it remains conceivable that T-LR may be comparable to, or even outperform, ML methods in certain datasets, depending on the features and their distribution.

Despite these limitations, our study is the first to evaluate ML methods to predict progression on AS for prostate cancer. This study provides impetus to develop a robust prediction model for progression on AS using ML methods in a larger sample size with more features.

5. Conclusions

In this study, we demonstrate that a ML approach out-performs a traditional statistical approach to predict progression on AS for prostate cancer. We anticipate that a robust model for clinical use will be developed using ML methods in a large population with abundant informative data.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

[1]. Hugosson J, Godtman RA, Carlsson SV, et al. Eighteen-year follow-up of the göteborg randomized population-based prostate cancer screening trial: effect of sociodemographic variables on participation, prostate cancer incidence and mortality. Scand J Urol 2018;52:27. [PubMed: 29254399]

[2]. Pinsky PF, Miller E, Prorok P, et al. Extended follow-up for prostate cancer incidence and mortality among participants in the prostate, lung, colorectal and ovarian randomized cancer screening trial. BJU Int 2019;123:854. [PubMed: 30288918]

[3]. Tosoian JJ, Mamawala M, Epstein JI, et al. Intermediate and longer-term outcomes from a prospective active-surveillance program for favorable-risk prostate cancer. J Clin Oncol 2015;33:3379. [PubMed: 26324359]

[4]. Hamdy FC, Donovan JL, Lane JA, et al. 10-year outcomes after monitoring, surgery, or radiotherapy for localized prostate cancer. N Engl J Med 2016;375:1415. [PubMed: 27626136]

[5]. Mottet N, van den Bergh RCN, Briers E, et al. EAU-EANM-ESTRO-ESUR-SIOG guidelines on prostate cancer-2020 update. Part 1: screening, diagnosis, and local treatment with curative intent. Eur Urol 2021;79(2):243–62. [PubMed: 33172724]

[6]. Sanda MG, Cadeddu JA, Kirkby E, et al. Clinically localized prostate cancer: AUA/ASTRO/SUO guideline. Part I: risk stratification, shared decision making, and care options. J Urol 2018;199:683. [PubMed: 29203269]

[7]. Morash C, Tey R, Agbassi C, et al. Active surveillance for the management of localized prostate cancer: guideline recommendations. Can Urolo Assoc J 2015;9:171.

[8]. Mamawala MM, Rao K, Landis P. Risk prediction tool for grade reclassification in men with favourable-risk prostate cancer on active surveillance. BJU Int 2017;120:25. [PubMed: 27469419]

[9]. Yamamoto T, Musunuru HB, Vesprini D. Metastatic prostate cancer in men initially treated with active surveillance. J Urol 2016;195:1409. [PubMed: 26707510]

[10]. Dall'Era MA, Konety BR, Cowan JE. Active surveillance for the management of prostate cancer in a contemporary cohort. J Am Cancer Society 2008;112:2664.

[11]. Dall'Era MA. The economics of active surveillance for prostate cancer. Curr Opin Urol 2013;23:278. [PubMed: 23449496]

[12]. Loeb S, Vellekoop A, Ahmed HU. Systematic review of complications of prostate biopsy. Eur Urol 2013;64:876. [PubMed: 23787356]

[13]. Halpern JA, Sedrakyan A, Dinerman B. Indications, utilization and complications following prostate biopsy: New York State analysis. J Urol 2017;197:1020. [PubMed: 27856226]

[14]. Marzouk K, Assel M, Ehdaie B. Long-term cancer specific anxiety in men undergoing active surveillance of prostate cancer: findings from a large prospective cohort. J Urol 2018;200:1250. [PubMed: 29886089]

[15]. Cooperberg MR, Brooks JD, Faino AV. Refined analysis of prostate-specific antigen kinetics to predict prostate cancer active surveillance outcomes. Eur Urol 2018;74:211. [PubMed: 29433975]

[16]. Cooperberg MR, Zheng Y, Faino AV. Tailoring intensity of active surveillance for low-risk prostate cancer based on individualized prediction of risk stability. JAMA Oncol 2020;6:e203187. [PubMed: 32852532]

[17]. Ankerst DP, Xia J, Thompson IM Jr. Precision medicine in active surveillance for prostate cancer: development of the canary-early detection research network active surveillance biopsy risk calculator. Eur Urol 2015;68:1083. [PubMed: 25819722]

[18]. Churpek MM, Yuen TC, Winslow C. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. Crit Care Med 2016;44:368. [PubMed: 26771782]

[19]. Ross EG, Shah NH, Dalman RL. The use of machine learning for the identification of peripheral artery disease and future mortality risk. J Vasc Surg 2016;64:1515. [PubMed: 27266594]

[20]. Preston MA, Feldman AS, Coen JJ, et al. Active surveillance for low-risk prostate cancer: need for intervention and survival at 10 years. Urol Oncol 2015;33:383.e9.

[21]. Guess HA, Gormley GJ, Stoner E. The effect of finasteride on prostate specific antigen: review of available data. J Urol 1996;155:3. [PubMed: 7490873]

[22]. Malhotra R Comparative analysis of statistical and machine learning methods for predicting faulty modules. ApplSoft Comput 2014;21:286.

[23]. Uddin S, Khan A, Hossain ME. Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak 2019;19:1. [PubMed: 30616584]

[24]. Jeni LA, Cohn JF, De La Torre F. Facing imbalanced data recommendations for the use of performance metrics. international. In: conference on affective computing and intelligent interaction and workshops: [proceedings]. ACII (Conference); 2013;2013:245–251.

[25]. Brennan RL, Harris DJ, Hanson BA. The bootstrap and other procedures for examining the variability of estimated variance components in testing contexts. IA: American College Testing Program Iowa City; 1987.

[26]. Butler SS, Mahal BA, Lamba N. Use and early mortality outcomes of active surveillance in patients with intermediate-risk prostate cancer. Cancer 2019;125:3164. [PubMed: 31150125]

[27]. Gu Q, Zhu L, Cai Z. Evaluation measures of the classification performance of imbalanced data sets. In: Presented at the Computational Intelligence and Intelligent Systems, Berlin, Heidelberg; 2009;51:461–471.

[28]. Vakili M, Ghamsari M, Rezaei M. Performance analysis and comparison of machine and deep learning algorithms for IoT data classification. 2020

[29]. Austin PC, Tu JV. Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality. J Clin Epidemiol 2004;57:1138. [PubMed: 15567629]

[30]. James G, Witten D, Hastie T. An introduction to statistical learning. 1st ed. New York: Springer; 2013.

[31]. Kate RJ, Perez RM, Mazumdar D. Prediction and detection models for acute kidney injury in hospitalized older adults. BMC Med Inform Decis Mak 2016;16:39. [PubMed: 27025458]

[32]. Thottakkara P, Ozrazgat-Baslanti T, Hupf BB. Application of machine learning techniques to high-dimensional clinical data to forecast postoperative complications. PLoS One 2016;11:e0155705. [PubMed: 27232332]

[33]. Austin PC, Reeves MJ. The relationship between the C-statistic of a risk-adjustment model and the accuracy of hospital report cards: a monte carlo study. Med Care 2013;51:275. [PubMed: 23295579]

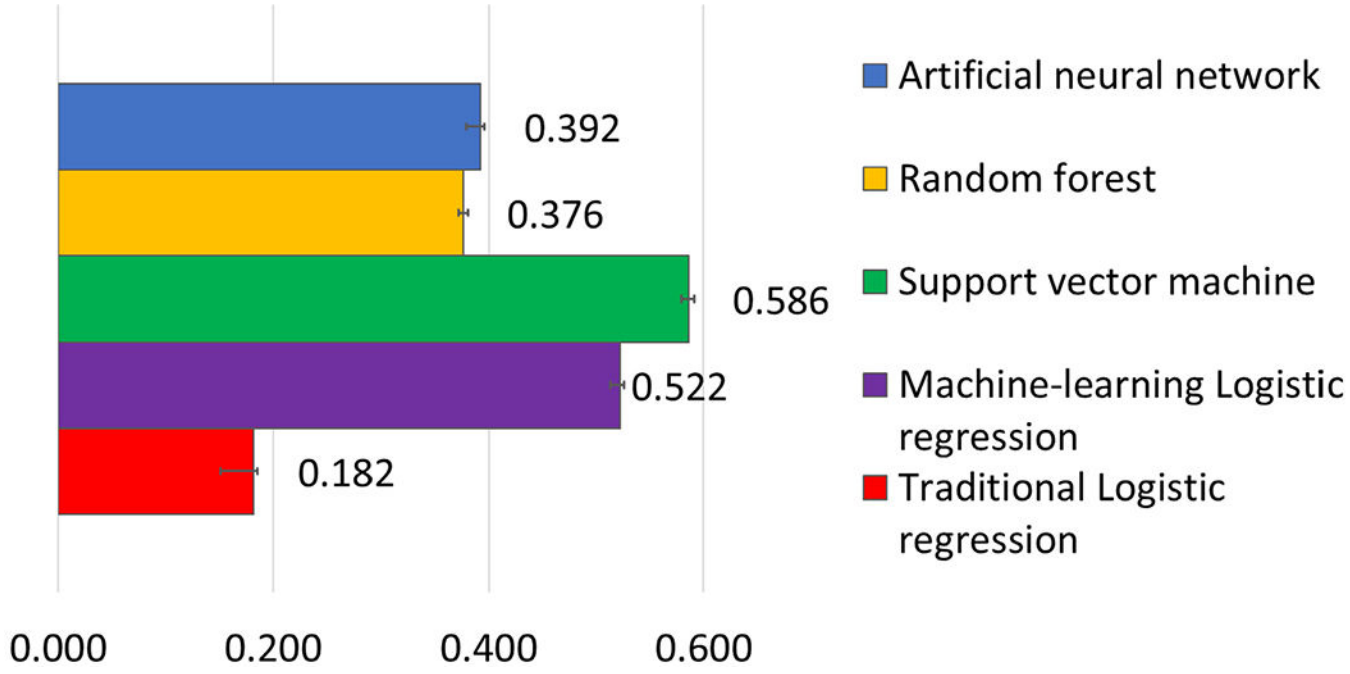

Fig. 1.

F1 scores of machine learning and traditional statistical models predicting progression on active surveillance for prostate cancer.

Traditional logistic regression

	Predicted: Yes	Predicted: No	
Actual: Yes	TP=5	FN=42	47
Actual: No	FP=3	TN=108	111
	8	150	

Support vector machine

	Predicted: Yes	Predicted: No	
Actual: Yes	TP=34	FN=13	47
Actual: No	FP=35	TN=76	111
	69	89	

Random forest

	Predicted: Yes	Predicted: No	
Actual: Yes	TP=19	FN=28	47
Actual: No	FP=35	TN=76	111
	54	104	

Artificial neural network

	Predicted: Yes	Predicted: No	
Actual: Yes	TP=19	FN=28	47
Actual: No	FP=31	TN=80	111
	50	108	

Machine learning logistic regression

	Predicted: Yes	Predicted: No	
Actual: Yes	TP=30	FN=17	47
Actual: No	FP=38	TN=73	111
	68	90	

Fig. 2.

Confusion matrices for models predicting progression on active surveillance for prostate cancer. Abbreviations: FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.

Table 1.

Characteristics of the cohort used to develop a classification model predicting progression on active surveillance for very-low and low-risk prostate cancer

Features	**Full cohort (*n* = 790)**	**Training set (*n* = 632)**	**Test set (*n* = 158)**
Age at diagnosis, years (median (IQR))	64.5 (59.1 – 69.5)	64.6 (59.3 – 69.4)	64.1 (58.1 – 69.6)
Family history of prostate cancer (*n* (%))			
Yes	275 (34.8)	212 (33.5)	63 (39.9)
No	496 (62.8)	405 (64.1)	91 (57.6)
Unknown	19 (2.4)	15 (2.4)	4 (2.5)
PSA at diagnosis, ng/dL (median (IQR))	4.8 (3.8 – 6.1)	4.9 (3.9 – 6.3)	4.7 (3.4 – 5.6)
Clinical tumor stage at diagnosis (*n* (%))			
1	733 (93)	588 (93)	145 (92)
2a	57 (7)	44 (7)	13 (8)
Number of cores (median (IQR))	12 (12 – 12)	12 (12 – 12)	12 (12 – 12)
Number of positive cores (median (IQR))	1 (1 – 2)	1 (1 – 2)	1 (1 – 2)
Maximum core involvement (median, (IQR))	8 (5 – 15)	5 (5 – 15)	10 (5 – 15)
Perineural invasion present (*n*, (%))	30 (4)	23 (4)	7 (4)
Number of surveillance biopsies (median (IQR))	2 (1 – 2)	2 (1 – 2)	2 (1 – 2)
Year of diagnosis (median (IQR))	2010 (2008 – 2013)	2011 (2008 – 2013)	2011 (2008 – 2013)
Estimated prostate volume, mL (median (IQR))	41 (31 –55)	41 (31 –57)	41 (31 – 56)
PSA density, ng/ml^2 (median (IQR))	0.11 (0.07 – 0.15)	0.11 (0.07 – 0.14)	0.10 (0.06 – 0.14)
Grade progression (*n* (%))	234 (30)	187 (30)	47 (30)

Abbreviations: IQR= interquartile range

Table 2.

Comparison of performance metrics between classification models predicting progression on active surveillance for prostate cancer

Model	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Negative predictive value (%)	c-statistic
T-LR	10.6 (8.7 – 10.9)	97.3 (97.3 – 97.3)	62.5 (57.1 – 62.5)	72.0 (71.8 – 72.5)	0.686 (0.681 – 0.699)
ML-LR	63.8 (63.0 – 65.2)	65.8 (65.5 – 66.4)	44.1 (43.3 – 44.8)	81.1 (80.9 – 82.0)	0.701 (0.696 – 0.708)
Support-vector machine	72.3 (71.7 – 73.9)	68.5 (68.2 – 69.1)	49.3 (48.5 – 50.0)	85.4 (85.2 – 86.4)	0.701 (0.696 – 0.709)
Random forests	40.4 (39.1 – 41.3)	68.4 (68.2 – 69.1)	35.2 (34.0 – 35.8)	73.1 (72.8 – 73.8)	0.603 (0.595 – 0.611)
Artificial neural network	40.4 (39.1 – 41.3)	72.1 (71.8 – 72.7)	38.0 (36.7 – 38.8)	74.1 (73.8 – 74.8)	0.545 (0.536 – 0.556)

Abbreviations: T-LR= traditional logistic regression; ML-LR= machine learning logistic regression

95% confidence interval shown in parentheses.