Heliyon 10 (2024) e24292

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Combined measurement of serum zinc with PSA ameliorates prostate cancer screening efficiency via support vector machine algorithms

Muyu Wu [a,1], Yucan Zhang [b,1], Xiaoqun Zhang [b], Xiaozhu Lin [a], Qiaoqiao Ding [b], Peiyong Li [a,*]

[a] Department of Nuclear Medicine, Ruijin Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China
[b] School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai, China

ARTICLE INFO

Keywords:
Prostate cancer
Support vector machine
Zinc ion
Prostate-specific antigen
Age

ABSTRACT

Background: Early screening of prostate cancer (PCa) is pivotal but challenging in the clinical scenario due to the phenomena of false positivity or false negativity of some serological evaluations, e.g. PSA testing. Decline of serum Zn^{2+} levels in PCa patients reportedly plays a crucial role in early screening of PCa. Accordingly, we combined 4 indices comprising the serum levels of total PSA (tPSA), free PSA (fPSA), Zn^{2+} and demographic information (especially age) in order to ameliorate the efficacies of PCa screening with support vector machine (SVM) algorithms.

Methods: A total of 858 male patients with prostate disorders and 345 healthy male controls were enrolled. Patients' data included 4 variables and serum Zn^{2+} was quantified via a self-invented Zn^{2+} responsive AIE-based fluorescent probe as previously published. tPSA and fPSA were routinely determined by a chemiluminescent method. Mathematical simulations were conducted to establish a SVM model for the combined diagnostics with the four variables. Moreover, ROC and its characteristic AUC were also employed to evaluate the classification efficacy of the model. Sigmoid function was utilized to estimate corresponding probabilities of classifying the clinical subjects as per 5 grades, which were incorporated into our established prostate index (PI) stratification system.

Results: In SVM model, the mean AUC of the ROC with the quartet of variables was approximately 84% for PCa diagnosis, whereas the mean AUC of the ROCs with tPSA, fPSA, [Zn^{2+}] or age alone was 64%, 62%, 55% and 59%, respectively. We further established an integrated prostate index (PI) stratification system with 5 grades and a software package to support clinicians in predicting PCa, with the accuracy of our risk stratification system being 83.3%, 91.6% and 83.3% in predicting normal, benign and PCa cases in corresponding groups. Follow-up findings especially MRI results and PI-RADS scores supported the reliability of this stratification platform as well.

Conclusion: Findings from our present study demonstrated that index combination via SVM algorithms may well facilitate clinicians in early differential screening of PCa. Meanwhile, our established PI stratification system based on SVM model and Sigmoid function provided substantial accuracy in preclinical risk prediction of developing prostate cancer.

* Corresponding authorSchool of Medicine, Shanghai Jiao Tong University, Shanghai, China.
E-mail address: peiyli@vip.sina.com (P. Li).
[1] First authors, contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2024.e24292
Received 6 April 2023; Received in revised form 1 December 2023; Accepted 5 January 2024
Available online 8 January 2024

1. Introduction

Prostate cancer (PCa) is the most prevalent cancer in the male population worldwide, which has prevailed the prevalence of lung cancer as of 2020, with an increasing trend of development in youth and middle ages [1,2]. Accordingly, given the high morbidity and mortality of PCa, early screening and accurate intervention are essential, especially primary differentiation between PCa and other benign prostate disorders (BPD), e.g. benign prostate hyperplasia (BPH), acute and chronic prostatitis, benign prostatic nodules, and prostate calcification, etc. Notwithstanding, differentiation between PCa and BPD has always been pivotal but challenging, mainly due to the false positivity and negativity in prostate-specific antigen (PSA) determination in certain clinical scenarios, as exemplified by a gray zone of serum PSA concentrations from 4 to 10 ng/mL, which is the most ambiguous in diagnostics [3]. There were also reports on benign cases with serum PSA levels over 10 ng/mL and malignant cases with PSA below 4 ng/mL [4]. Hitherto, basic researchers and clinicians have been devising strategies, comprising determination of serum free PSA (fPSA), fPSA/tPSA rate, p2PSA and prostate health index (PHI), radiological evaluations with computed tomography (CT), ultrasonography (US), magnetic resonance imaging (MRI), prostate-specific membrane antigen (PSMA) positron emission tomography (PET), prostate puncture biopsy, etc. Despite their contribution to the differential diagnosis of PCa, most of these regimens were not cost-effective, thus restricting their early clinical and preclinical screening potentials. Meanwhile, the normal prostate has abundant zinc ion (Zn^{2+}) in contrast to decreased Zn^{2+} levels in PCa patients, with serum Zn^{2+} level also as an essential biomarker for PCa diagnosis [4–8]. In addition, the patients' age information (age) is a well-known influential factor in cancer incidence. Therefore, combined determination of several indices might be justifiable.

Machine learning (ML) is an interdisciplinary simulation and implementation of the learning activities of human beings, via integration of many realms e.g. probability theory, statistics, approximation theory, convex analysis, algorithm complexity theory, etc. Compared with conventional computation, ML created certain models from historical data training via certain algorithms to predict future events providing novel data is rendered. To date, diversities of ML algorithms comprising support vector machine (SVM), decision tree (DT), random forest (RF) and multilayer perceptron (MLP) etc. have been increasingly employed in the fields of natural sciences, especially clinical medicine. Notably, the SVM algorithm developed by Cortes and Vapnik in the 1990s is still applied in the classification and prediction of many disorders, such as PCa and cancer types of the stomach, breast, lungs, kidneys, etc. [9–12]. Furthermore, SVM with Gaussian kernel is very effective in multiple classification and problem prediction owing to its good generalization, superiority for small-sized data and high efficiency for multi-dimensionality data, etc. [13,14].

We combined the four features, i.e. levels of total PSA (tPSA), free PSA (fPSA), Zn^{2+} and patients' age (age) in order to ameliorate the efficacies of differential screening (classification) and further risk stratifications (prediction) of PCa with the assistance of several ML algorithms especially SVM. Receiver operating curve (ROC) and its characteristic area under the curve (AUC) were also employed to evaluate the diagnostic efficacies of corresponding ML models. Furthermore, risk stratifications and established prostate index (PI) stratification were performed via SVM model and Sigmoid function to predict the probabilities of PCa development.

2. Patients and methods

2.1. Patients

The study was approved by the independent ethics committee of Ruijin Hospital, School of Medicine, Shanghai Jiao Tong University, China. Between July 2021 and November 2022, 858 patients with clinically diagnosed prostate disorders including PCa (n = 150) and benign prostate disorders (n = 708) were recruited, the majority undergoing prostate biopsy at our hospital. The eligibility criteria were: 1. no prostate surgery history; 2. Confirmed diagnosis of prostate abnormalities. In parallel, we enrolled 345 healthy

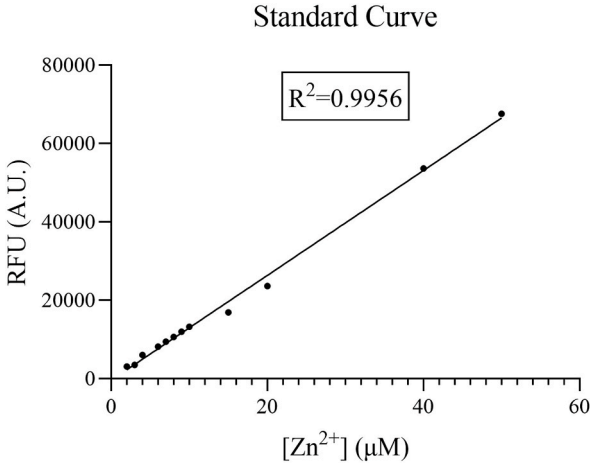

Fig. 1. Linear relationship and goodness-of-fit test of gradient Zn^{2+} concentrations with corresponding emitted fluorescence intensities, where RFU denotes relative fluorescence unit (A.U.) and [Zn^{2+}] denotes Zn^{2+} concentrations (μM).

individuals with regular physical examinations as the control group.

2.2. Measurements of serum Zn^{2+} and PSA

All the patients underwent determination of serum levels of tPSA and fPSA with routine chemiluminescent kit (Abcam) in our hospital. It is worth mentioning that we invented a Zn^{2+}-responsive fluorescent quantification approach to the serum Zn^{2+} determination in the enrolled patients [15]. Briefly, our self-synthesized fluorescent probe HL (HL: 3,4,5-tribenzyl-Benzoic acid (2-hydroxy-benzylidene)-hydrazide) was dissolved and dispersed in the solvent ($DMF/H_2O = 2/3$, v/v). The detection system revealed serum Zn^{2+} concentrations in the patients exhibited a positive correlation with their corresponding fluorescence intensities, as illustrated by a close linear relationship in the standard curve (Fig. 1), facilitating the fluorescence quantification of serum Zn^{2+} concentrations.

2.3. Introduction to analytical tool SVM

Assuming that two types of training samples are linearly separable:

$$\{\boldsymbol{x}_1, y_1\}, \{\boldsymbol{x}_2, y_2\}, \ldots, \{\boldsymbol{x}_Q, y_Q\}, \boldsymbol{x}_i \in \mathbb{R}^n, y_i \in \{-1, 1\}, 1 \le i \le Q \tag{1}$$

If samples cannot be linearly separable, a nonlinear mapping can be adopted to map the nonlinearly separable classification to a higher-dimensional space to achieve linear separability with high probability.

As illustrated in Fig. 2, as per the SVM method, we selected the hyperplane with the largest margin $2d$ among all possible classification hyperplanes, thus attaining the optimal generalization.

In order for π_1 and π_2 to form a class divider, then the points xi satisfying wxi + b > 1 are all on the one side of π1, The points xj satisfying wxj + b < −1 are all on the other side of π2. The following constraints must be satisfied:

$$\begin{cases} \boldsymbol{w}^T\boldsymbol{x}_i + b \ge 1 \\ \boldsymbol{w}^T\boldsymbol{x}_i + b \le -1 \end{cases}, 1 \le i \le Q \tag{2}$$

The distance between plane π1 and π2 is the goal we want to maximize, so that we not only realize the classification, but also maximize the classification margin. The total margin between the hyperplane π_1 and π_2 is:

$$2d = 2\frac{|\boldsymbol{w}^T\boldsymbol{x}^* - \boldsymbol{w}^T\boldsymbol{x}_0|}{\|\boldsymbol{w}\|} = 2\frac{|\boldsymbol{w}^T\boldsymbol{x}^* + b|}{\|\boldsymbol{w}\|} = \frac{2}{\|\boldsymbol{w}\|} \tag{3}$$

Therefore, attainment of the optimal classification hyperplane can be translated into the following quadratic convex programming problem:

$$\begin{cases} \min\limits_{w,b} \dfrac{1}{2}\|\boldsymbol{w}\|^2 \\ subject\ to\ y_i\left(\boldsymbol{w}^T\boldsymbol{x}_i + b\right) \ge 1, 1 \le i \le Q \end{cases} \tag{4}$$

With solution of the above quadratic convex optimization problem, the optimal classification hyperplane can be achieved.

2.4. SVM simulation with python

The Scikit-Learn library in Python, with its wide application to ML and other fields, can better calculate the SVM classification

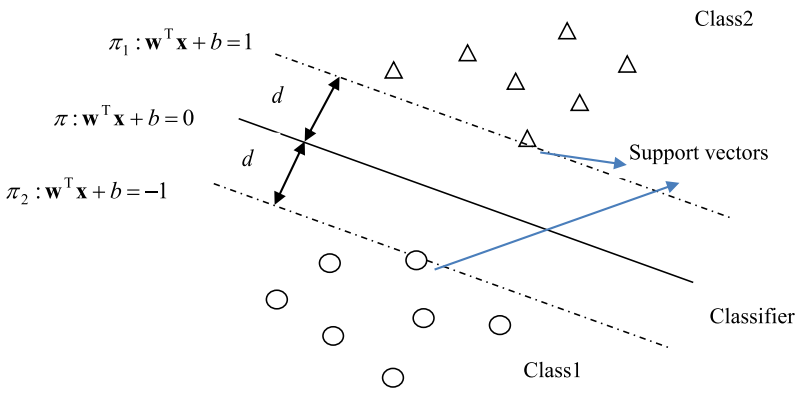

Fig. 2. Schema of the rationale for support vector machine model.

surface and facilitate SVM algorithm. For the utilization of Scikit-Learn library in SVM, the data set should be divided into the training set and the test set, and the fit ( ) method ought to be applied to the training set to solve the optimization problem in 2.4 to obtain the parameters of the classification hyperplane, and the predict ( ) method for the test set should be used to predict the probability. The accuracy of the classification and prediction of the test set is established by the following formula:

$$\text{Accuracy} = \frac{\#\text{ correctly predicted data}}{\#\text{ total testing data}} \times 100\% \tag{5}$$

Additionally, cross-validation methods were built into the Scikit-Learn library. As exemplified by 5-fold cross-validation (Fig. 3), this method can divide the sample data into five datasets, with one as the test data set in each validation and the other four as the training data set. Five head-to-head experiments were performed to preclude the computational imbalance of the classification surface due to the random division of the data set.

2.5. SVM model and sigmoid function in risk prediction

Since SVM model is a two-class classifier, only the algorithm of two-class classification can be achieved. Accordingly, to solve the three-classification problem, combinations of the obtained three categories (normal, benign and malignant) were adopted, involving three scenarios, i.e. "normal & benign", "normal & malignant", and "benign & malignant". Consequently, the three two-category sample data were utilized for the SVM model training and data classification.

The three-category SVM model was employed to conduct multiple cross-validations, with experimental data obtained. With attainment of the prespecified accuracy rate of experimental data, the classification surface with the highest accuracy rate was designated as the classification surface, with the accuracy rate no less than 80%. During the experiment, by five-fold cross-validation, the average accuracy rate was over 84.4%, with the variance of 0.059. The classification result with the highest accuracy rate was designated as the classification surface for the prostate index (PI) evaluation as follows.

$$\pi_1 : \text{-}0.05098266 * \text{Age-}0.99452169 * \text{tPSA-}0.8271752 * \text{fPSA} + 0.00320679 * \left[Zn^{2+}\right]$$

$$\pi_2 : \text{-}0.16478435 * \text{Age-}0.22402318 * \text{tPSA-}0.7961012 * \text{fPSA} + 0.18290099 * \left[Zn^{2+}\right]$$

$$\pi_3 : \text{-}0.05360206 * \text{Age} + 0.05138339 * \text{tPSA-}0.05095499 * \text{fPSA} + 0.16451015 * \left[Zn^{2+}\right]$$

Herein, π_1: distinguishing "normal" from "benign & malignant"; π_2: distinguishing "benign" from "normal & malignant"; π_3: distinguishing "malignant" from "normal & benign".

The detailed applications of the classification surface: x^* is set as a new sample, and x^* is input into the three SVMs (support vector machine models) with training completion. If there is $x^* \in$ *class i* for the *i*th SVM, then $x^* \in$ *class i* is discriminated.

The distance to the classification surface was then calculated for each case, yielding an estimate of the probability of prostate cancer (PCa) development. For the input data, the distance from the point to the classification surface was calculated, and thereafter the distance underwent transformation via the Sigmoid function to yield a value between 0 and 1, or rather, prostate index (PI) of the estimated probability of PCa, which informs physicians as to preclinical the prostate state of the clinic visitors. In particular, the closer the points are to the classification surface, the higher the uncertainty.

2.6. Softwares and tools

Continuous variables were analyzed via one-way ANOVA with SAS (version 9.0). Model was established, performed and evaluated with the software Python (version 3.9.13 https://www.python.org/, with packages pandas 1.4.4 https://pandas.pydata.org/ and Scikit-Learn 1.0.2 https://scikit-learn.org/stable/), providing a measurement of the classification and prediction efficacy of our SVM model.

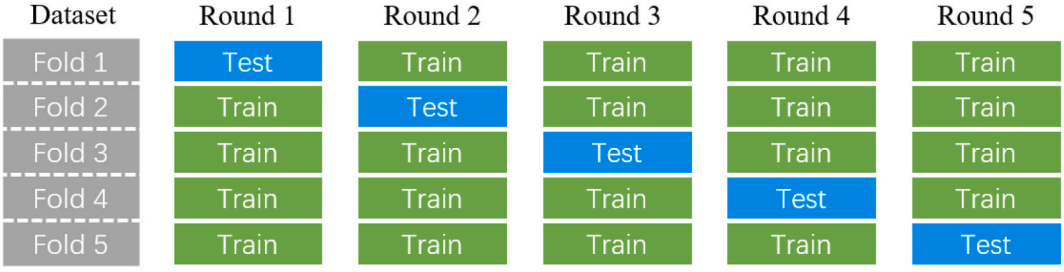

Fig. 3. Schema of the rationale for 5-fold cross-validation.

3. Results

3.1. Data distribution

858 patients with prostate disorders were enrolled, including cases of benign prostate hyperplasia ($n = 668$), prostatitis ($n = 38$), prostatic calcification ($n = 2$) and PCa ($n = 150$), with 345 healthy individuals as the normal control group. Table 1 illustrate the data distribution of age, serum Zn^{2+} concentration ([Zn^{2+}]), tPSA and fPSA levels in normal, benign and PCa group.

3.2. Efficacy evaluation of classification with SVM algorithm

Fig. 4 exhibits the ROC curves and their corresponding AUC values of classification efficacies of single index as age (Fig. 4a), PSA (Fig. 4b), fPSA (Fig. 4c) and [Zn^{2+}] (Fig. 4d) as well as their combination (Fig. 4e), respectively, with the assistance of SVM algorithms. As is depicted, the mean AUC value of the ROC curve with age alone is approximately 0.59 ± 0.02, whereas the mean AUC with individual index as tPSA, fPSA and [Zn^{2+}] was approximately 0.64 ± 0.13, 0.62 ± 0.11, and 0.55 ± 0.08, respectively. Notably, the mean AUC value of the ROC of index quartet was approximately 0.84 ± 0.05, which was much greater than the respective values of one index. Despite the unsatisfactory efficacy with the index [Zn^{2+}] alone in PCa screening, our combination strategy with SVM algorithms remedied the screening inefficacy.

3.3. Risk prediction with sigmoid function

Given the predicative performances and advantages of ML algorithms especially SVM, we further utilized Sigmoid function to stratify the probabilities of PCa development. Fig. 5(a–f) depicts the classification effect of the classification surface (hyperplane) on the test set, where the points represent the samples, and the area boundaries of different colors represent the classification surface. Since the model adopted 4 features as input values, the two-dimensional map was established by selecting two out of these four features as the axis, focusing on the malignant area. As shown in Table 2, to test the accuracy and veracity of our predictive model, we input all our cases in the model and checked their predictive and real values. Indeed, the classifier can better classify the test data with the prediction accuracy being 0.833 (normal group), 0.916 (benign group) and 0.833 (PCa group), respectively.

3.4. SVM model sensitivity evaluation

In order to verify the sensitivity of the model, a minor disturbance ε is given to the input value $\boldsymbol{x}$ of the model, and the variation of the output value $\boldsymbol{y}$ of the model will be evoked. In case of minor variation of model output value, the stability of the model is considered potent, and vice versa. We assumed a patient in a critical margin of PCa for this experiment, i.e., 'age' = 55 (y. o.), 'tPSA' = 5 (μmol/L), 'fPSA' = 4 (μg/mL), 'Zn' (i.e. [Zn^{2+}]) = 5 (μg/mL). Three out of four variables were fixed, with the output value of the probability of developing into PCa available. The relationship between the PCa probability and the value of the input variables (age, tPSA, fPSA and [Zn^{2+}]) is depicted (Fig. 6 a - d).

In our model, in case of the parameter modification as in age and tPSA, the probability of PCa development varies greatly, indicating that variables as age and tPSA are the pivotal contributors to the probability of PCa development. Despite the modification of variables fPSA and [Zn^{2+}], modest effect on the probability of PCa development was detected, implying the auxiliary roles of two parameters. The model output values did not exhibit marked fluctuation in response to modification of the input values, indicating the stability of our SVM model.

3.5. Establishment of prostate index (PI) for risk stratification

To evaluate by stratification of the prostate states of participants, we input the predicted probability of PCa development in each individual to establish a prostate index (PI) system (Fig. 7) via SVM model and Sigmoid function (Table 3).

Our PI scoring system can stratify the predicted probability of PCa development into 5 grades via SVM model. Consequently, the participant stratification was profiled as follows: grade one (PI score = 1), 304 normal participants and 56 benign prostatic patients; grade two (PI score = 2), 31 normal participants and 262 benign prostatic patients; grade three (PI = 3), 10 normal participant, 306 benign prostatic patients and 58 PCa patients. As for grades four (PI score = 4) and five (PI score = 5), only patients with either benign prostate diseases or PCa were involved. Patients with grades three to five should be recommended for radiological evaluations. The PI system and its component structures were integrated into a software program in both English and Chinese versions, which has been

Table 1
Data distribution (mean ± SD) of patient information.

Group	Sample size	Age (y. o.)	[Zn^{2+}] (μmol/L)	tPSA (μg/mL)	fPSA (μg/mL)
Normal	345	40.919 ± 12.800	4.513 ± 2.248	1.257 ± 1.168	0.379 ± 0.276
Benign	708	68.058 ± 10.701	8.339 ± 7.303	9.242 ± 6.887	1.790 ± 2.031
PCa	150	71.873 ± 8.626	3.267 ± 4.229	18.489 ± 28.609	2.660 ± 3.636
Total	1203	60.751 ± 16.830	6.609 ± 6.281	8.105 ± 12.569	1.494 ± 2.160

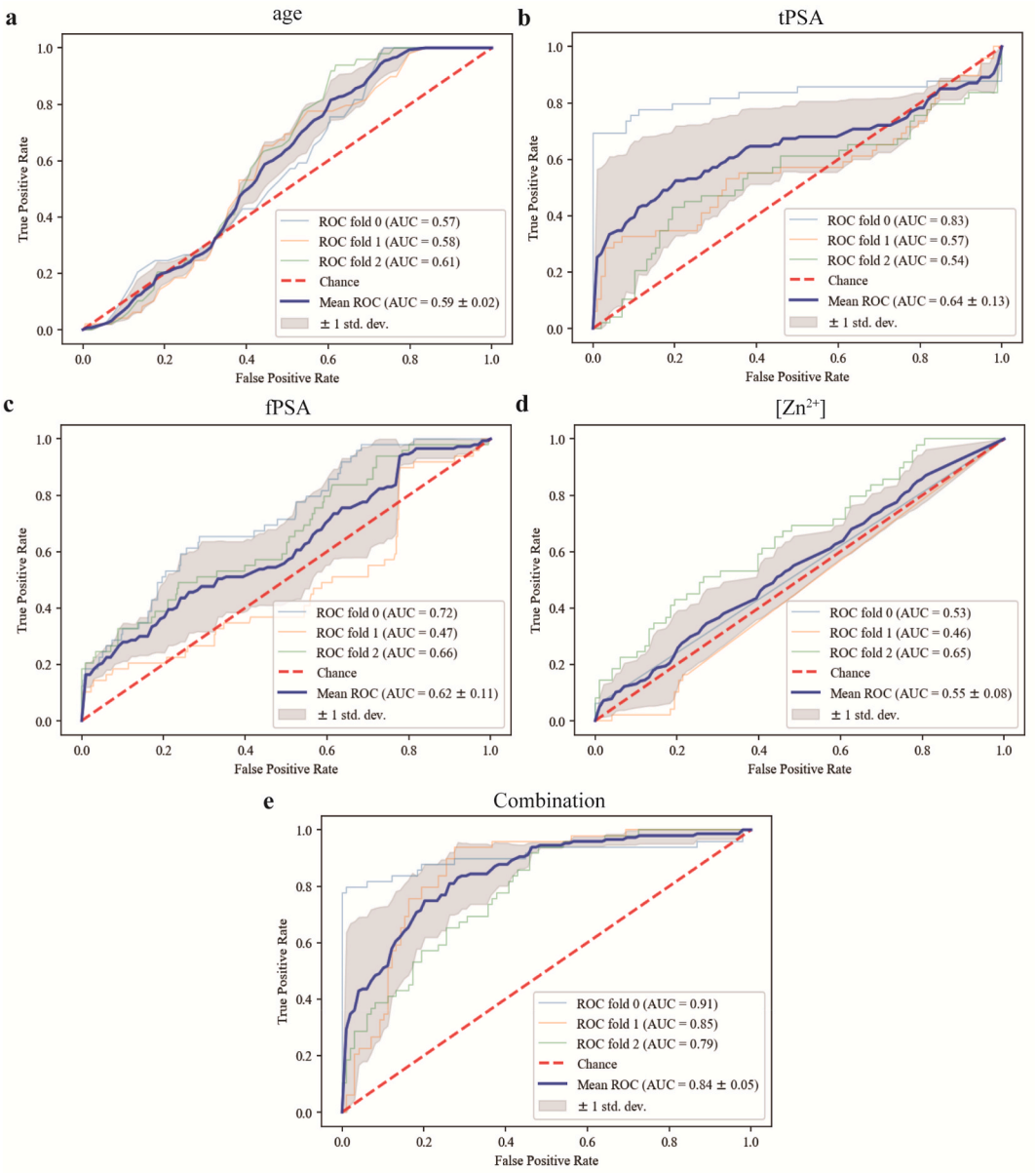

Fig. 4. ROC curves and their characteristic AUC values of sole-index diagnosis (**a** with sole index of Age, **b** with tPSA, **c** with fPSA, **d** with $[Zn^{2+}]$) and combined diagnosis (**e** 4 indices combined) with SVM algorithms.

applied clinically in several tertiary hospitals in Shanghai, China.

4. Discussion

As per many researchers in early differential screening of prostate cancer, diminished concentrations of free zinc ion (Zn^{2+}), whether in serum, in plasma, in urine, or in prostate tissue, are all considered as the marker of potential PCa development. A research group in Massachusetts General Hospital screened PCa via determination of endogenous zinc in samples of expressed prostatic secretion (EPS) [16]. In addition, Wakwe and colleagues conducted a descriptive cross-sectional study on the impact of low plasma zinc level on higher severity of Pca [17]. The association between low zinc level in prostate tissues and biochemical recurrence in PCa has been well documented [18]. These strategies are supportive of our measuring sensor for free zinc with self-invented fluorescent quantification regimen. Zhao group in China and Kuwahara group in Japan collaborated in the clinical applications of serum $[Zn^{2+}]$ measurement in PCa with the conclusions that $[Zn^{2+}]$ determination can well improve PCa detection efficiency in patients with PSA levels in the gray zone (i.e. 4-10 ng/mL) [19]. What is intriguing and worth mentioning is that their diagnostic efficacy with serum zinc

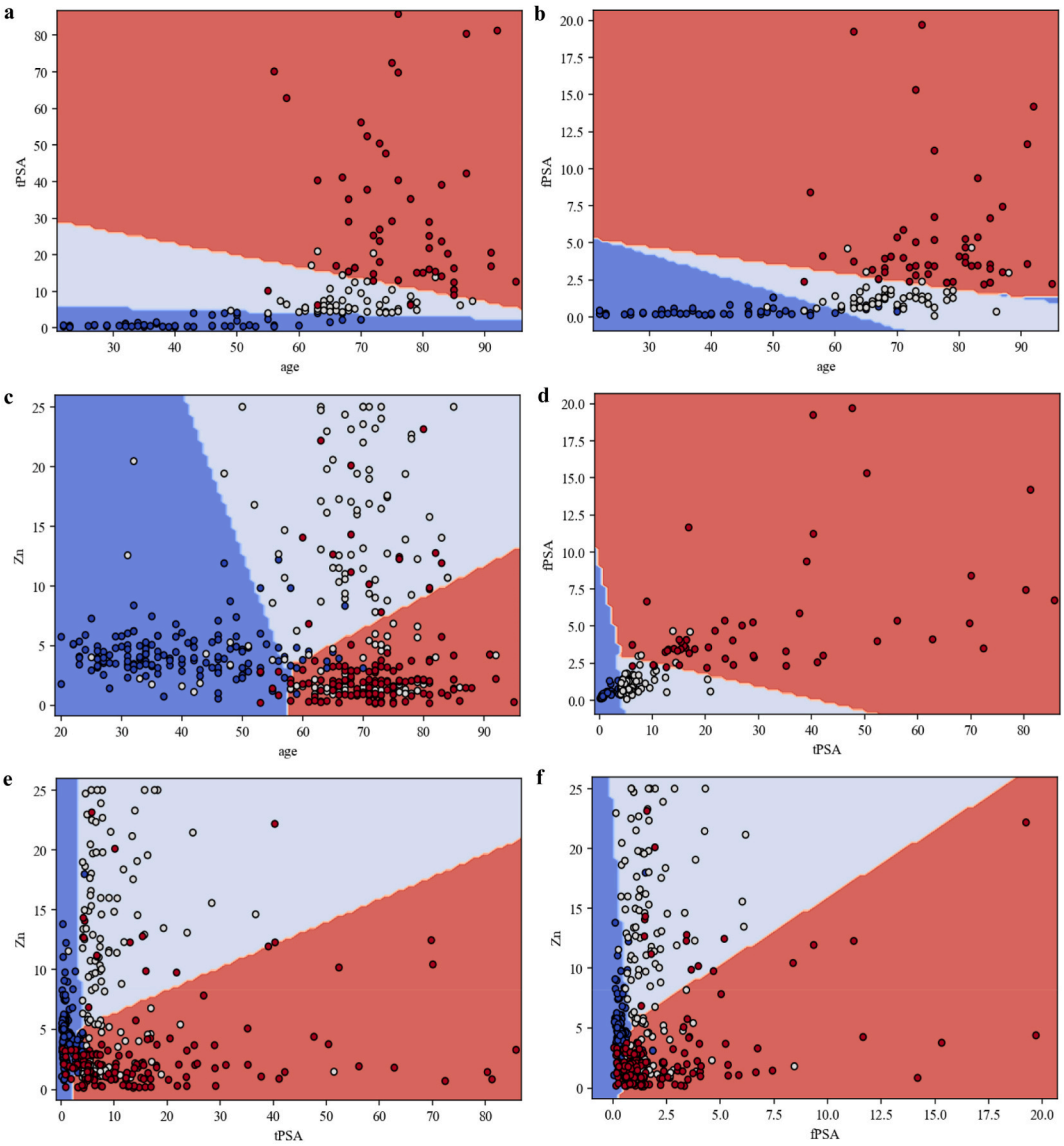

Fig. 5. Classification effect of the hyperplane (classification surface **an** of [Zn^{2+}]-tPSA; classification surface b of [Zn2+]-fPSA; classification surface c of [Zn2+]-age; classification surface d of tPSA-age; classification surface **e** of fPSA-age; classification surface **e** of fPSA-tPSA) on the test set, where the points denote the corresponding samples; the area boundaries of different colors denote the corresponding classification surfaces; the blue, white, red regions denote the "normal", "benign", "malignant" prediction results, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 2
Accuracy evaluation of the prediction performances with SVM model.

Group	Training Cases	Testing Cases	Accuracy
Normal	276	69	0.833
Benign	566	142	0.916
PCa	118	29	0.833

alone was marvelously higher than our objectively and strictly collected data and their AUC value with serum zinc alone was as high as 0.7. In our results, however, serum [Zn^{2+}] alone was merely eligible for an adjuvant (mean AUC of 0.55) in screening PCa. Nonetheless, our combined diagnostic efficacy with four features can augment the mean AUC value to approximately 0.84, which is superior to the other diagnostic regimens irrespective of indexes alone or in combination.

Notwithstanding, utilization of the reduced [Zn^{2+}] alone could not suffice early screening of PCa. Accordingly, combined

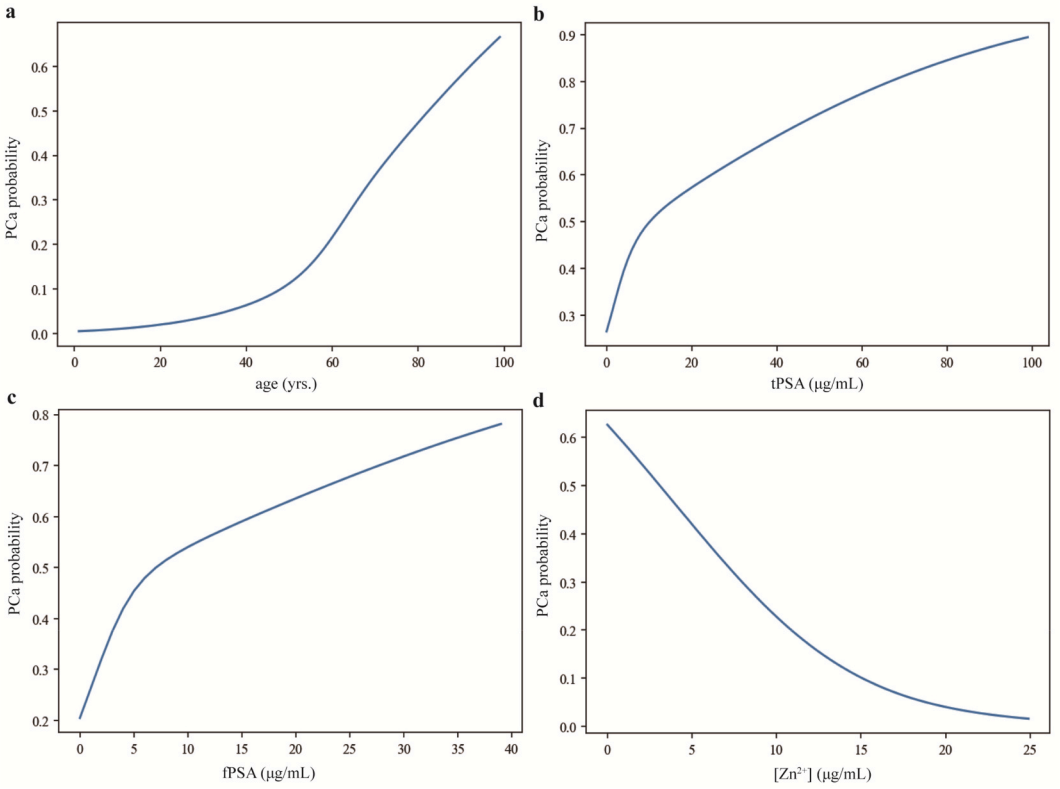

Fig. 6. The relationship between PCa probability and values of input variables is represented, where the input variable is age (yrs, **a**), tPSA (μg/mL, **b**), fPSA (μg/mL, **c**) and [Zn^{2+}] (μmol/L **d**), respectively.

diagnostic strategies flourish nowadays. Of note, Temiz and co-workers employed zinc to PSA ratio as a marker in prostate biopsy outcome prediction, despite the suboptimal mean AUC value of 0.67 [20]. Encouragingly, binary combination of [Zn^{2+}] with tPSA or fPSA in our SVM model yielded the mean AUC value of 0.78 and 0.75, respectively.

Machine learning (ML) has evoked revolution of modern research paradigms in many fields, especially in biomedicine. The classification and prediction of ML algorithms are already applied in PCa worldwide. Satoshi Nitta and colleagues verified that ML methods could predict more efficiently than some advanced PSA-based serologic screenings for PCa like PSA density and PSA velocity (21). Nevertheless, such approaches are limited due to the monofactorial application in scenarios with multifactors and there is a paucity of application of ML in quartet of [Zn^{2+}], tPSA, fPSA and Age, which justified our further comparisons of four different ML algorithms (including Decision Tree, Random Forest, SVM and Multilayer Perceptron) (Supplementary information). On SVM complexity, the more classical one is Lloyd N. Trefethen, David Bau III, Numerical Linear Algebra, SIAM, 1997. P81 Equation (11.22), which is an O (n^3) algorithm in the training process and for the testing process is O(n).

With respect to the data distribution, the maximum (23.131 μM) of serum [Zn^{2+}] in PCa group was paradoxically higher than in normal or benign group (Table 1), but the minimum (0.146 μM) and the average value (3.267 μM) of [Zn^{2+}] in PCa cases were much lower than in normal or benign group (Table 2). Given the circumstances of serum [Zn^{2+}] elevation in PCa patients, such as high-zinc diets, complication or comorbidity of other disorders like chronic gastritis, atherosclerosis, hypertension, etc., the discrepancy may be attributable to the presence of complications or comorbidities and the exclusion of such factors would substantially affect the AUC values. Hence, we excluded 37 PCa patients with complications or comorbidities from analysis. The consequent ROC profiles and its corresponding AUC values were significantly ameliorated. The combined classification efficacies of 4 indices via SVM algorithms yielded a mean AUC value of 0.94 (Fig. 8 d), which was superior to that with inclusion of complication or comorbidity (0.84) (Fig. 8 a - d). Despite the superiority of the results in this context, the robustness and stability of the model required more verifications due to the decline of PCa sample size. Thenafter, the initial model (the AUC of which was 0.84) was recommended in practice.

Despite the absence of specificity of serum PSA-related biomarkers (i.e. tPSA and fPSA) and [Zn^{2+}] in determination, the relatively high sensitivity of serological indices is a superiority in PCa screening, which is further supplemented by the mathematical advantages of the prediction functionality of SVM and other ML algorithms. We focused on risk stratifications during the screening procedure of PCa, with reference of the establishment of the renowned prostate imaging reporting and data system (PI-RADS) of MRI and the prostate health index (PHI) scoring system, etc. Either of the scoring platforms were not concerned with any ML/AI algorithm and cannot provide recommendation in radiology. In contrast, our SVM-based prostate index (PI) stratification system is capable of risk prediction of PCa and further corresponding recommendations for radiological evaluations. To our knowledge, our risk stratification

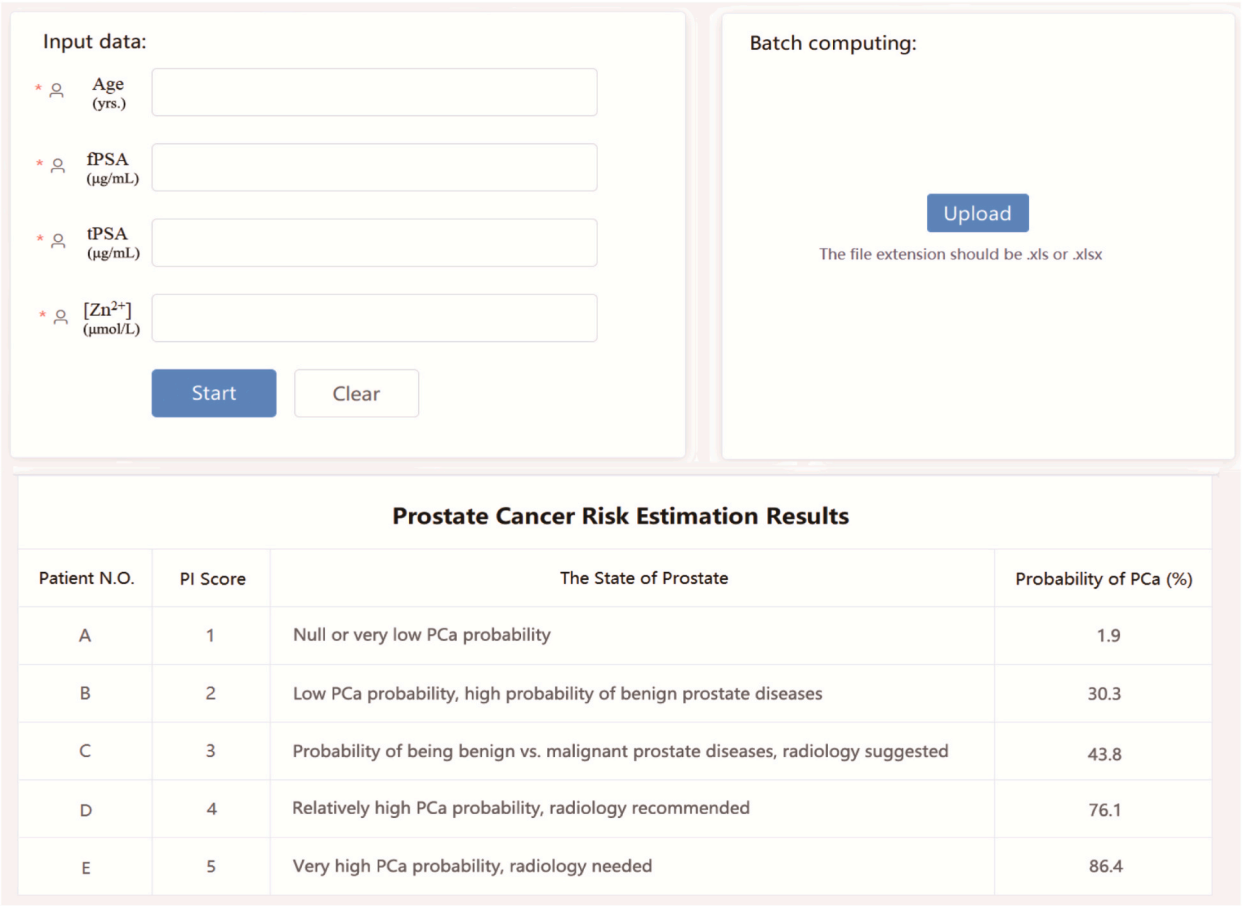

Fig. 7. The program platform "Prostate Index stratification system" with input parameters: age, free prostate-specific antigen, total prostate-specific antigen, and serum zinc ion concentration.

Table 3
Prostate index (PI) establishment via risk stratification results.

PI	Normal Group	Benign Group	PCa Group	Prostate State	Probability of PCa (%)
1	304	56	0	null or very low PCa probability	<15%
2	31	262	0	low PCa probability, high probability of benign prostate diseases	15%–35%
3	10	306	58	probability of being benign *vs.* malignant prostate diseases, radiology suggested	35%–65%
4	0	77	74	Relatively high PCa probability, radiology recommended	65%–85%
5	0	7	15	Very high PCa probability, radiology needed	>85%

system is superior to other predictors available in literature, with the accuracy of 83.3%, 91.6% and 83.3% in predicting the normal, benign and PCa cases, respectively.

In order to investigate and further verify the advantages of our established PI stratification system and its syncretism with MRI, we conducted a retrospective analysis of the clinical data of five patients with different PI scores, including 4 indices, PI scores, predicative probabilities of PCa development, clinically evidenced prostate states and follow-up profiles (Table 4). Patient C, in particular, had a PI score of 3 and the predicative PCa probability of approximately 45.6%. According to our PI system, patient C should have been recommended for further radiological evaluations especially MRI. Meanwhile, the patient was diagnosed as having prostate adenocarcinoma by post-operative pathology, which further underlined the values of our PI system. In addition, our PI system was supported by retrospective data analysis. The PI-RADS scores of patients D and E were 3 and 5, respectively in MRI reports. Both patients were confirmed with having prostate adenocarcinoma by post-surgery pathology. As per our PI system, patient D was given a PI score of 4 and predicative PCa probability of 67.2%, and patient E had a PI score of 5 and predicative PCa probability of 86.3%, respectively, which was in line with the follow-up findings but with higher accuracy. The followed up MR images and corresponding PI-RADS scores supported our PI scoring system as well (Fig. 9 a1 – a3, b1 – b3, c1 – c3).

With the requirement of personalized regimen for diagnostics and therapeutics, the PI stratification system for PCa risk prediction

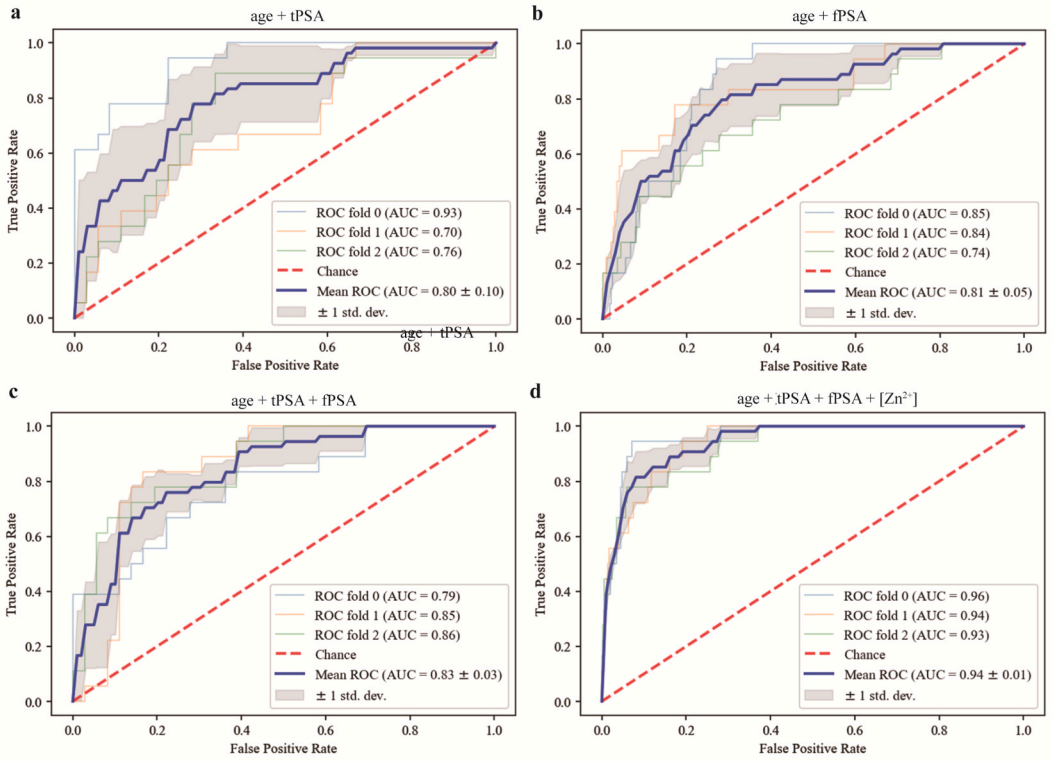

Fig. 8. ROC curves and their characteristic AUC values of Age combined with tPSA (**a**) or [Zn^{2+}] (**b**) or tPSA & fPSA (**c**) diagnosis and 4 indices-combined diagnosis (**d**) with SVM algorithms.

Table 4
Prostate index (PI) evaluations and follow-up profiles of certain patients.

Patient NO.	Age (y.o.)	tPSA (μg/mL)	fPSA (μg/mL)	Serum [Zn^{2+}] (μmol/L)	PI Score	The State of Prostate	Probability of PCa (%)	Follow-up profiling (esp. with MRI)
A	46	0.393	0.216	2.909	1	Prostate is nearly healthy, null or very low PCa probability	4.3%	N/A (no nodule identified in ultrasonography)
B	63	2.364	0.688	1.879	2	low PCa probability, high probability of benign prostate diseases	31.7%	N/A (only prostate hyperplasia but no nodule identified in ultrasonography)
C	66	6.349	0.843	1.243	3	Near probability of benign *vs.* malignant prostate diseases, radiology suggested	45.6%	N/A (diagnosed prostate adenocarcinoma in post-surgery pathology)
D	85	16.291	3.532	4.1706	4	Relatively high PCa probability, radiology recommended	67.2%	PI RADS score = 3 (nodules were seen in the left peripheral zone of prostate in MRI and prostate adenocarcinoma were diagnosed in post-surgery pathology)
E	83	23.637	5.365	1.044	5	Very high PCa probability, radiology needed	86.3%	PI RADS score = 5 (nodules were seen in the transitional & right peripheral zone of prostate in MRI and prostate adenocarcinoma were diagnosed in post-surgery pathology)

was established with the use of SVM model and Sigmoid function. Meanwhile, further radiological evaluation modes such as MRI and ^{68}Ga-PSMA nuclear imaging might be recommended in the case of PI score higher than a certain threshold.

5. Conclusion

We established an SVM classification model with the combination of four parameters ([Zn^{2+}], tPSA, fPSA and age) via SVM

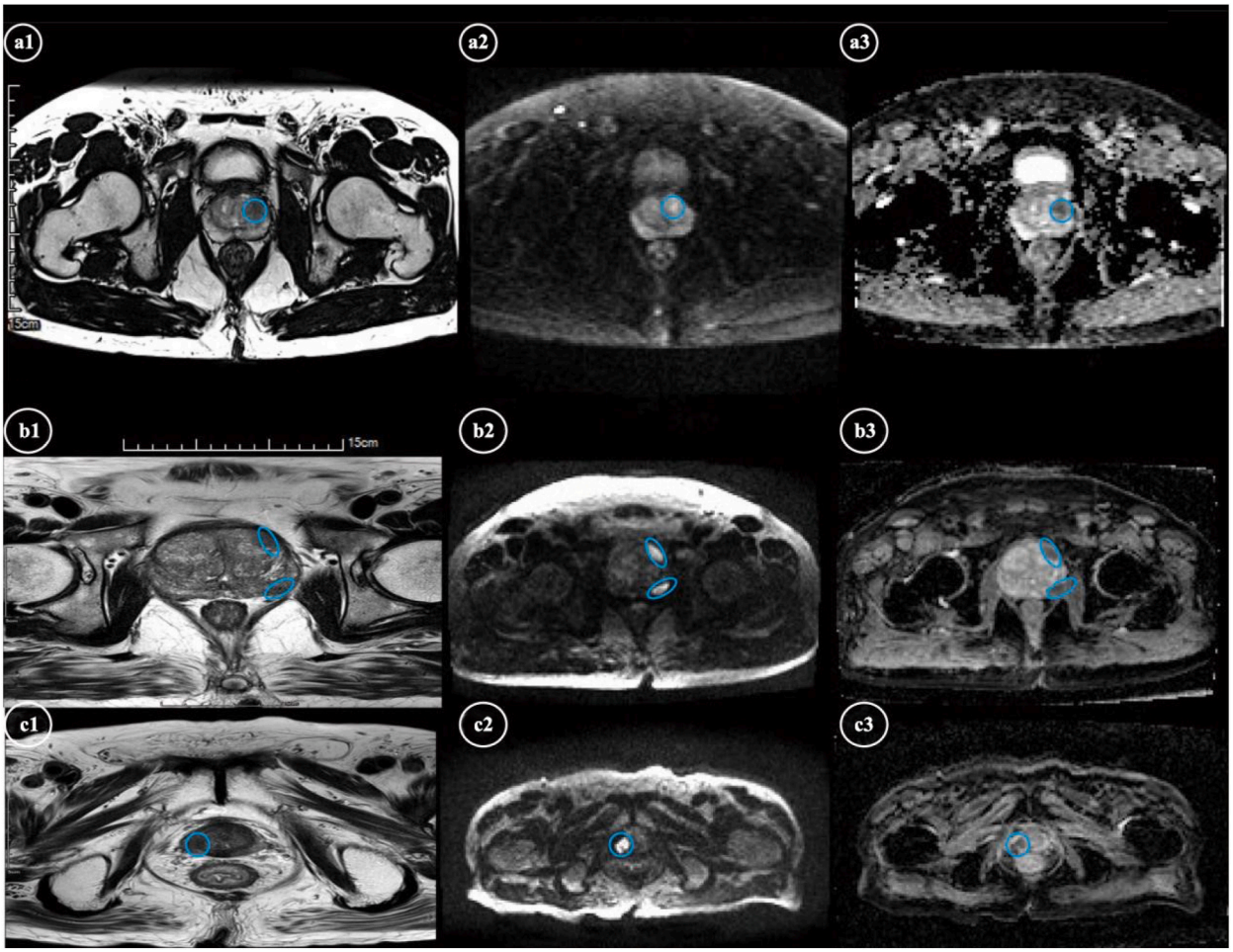

Fig. 9. MRI follow-up profiles of three patients. This figure illustrates the T2-WI (**a1**), DWI (**a2**), and ADC imaging (**a3**) of a patient whose PI score = 3 and PI-RADS = 4, the T2-WI (**b1**), DWI (**b2**), and ADC imaging (**b3**) of a patient whose PI score = 4 and PI-RADS = 5 and the T2-WI (**c1**), DWI (**c2**), and ADC imaging (**c3**) of a patient whose PI score = 5 and PI-RADS = 4. *T2-WI denotes T2-weighted imaging; DWI denotes diffusion-weighted imaging; ADC denotes apparent diffusion coefficient.

algorithms in early differential screening of PCa. Our PI stratification system based on SVM model and Sigmoid function presents optimal accuracy in preclinical prediction of prostate state in individuals. The combination of zinc ion concentration ($[Zn^{2+}]$) and integration of support vector machine (SVM) algorithm meliorate prostate cancer (PCa) differential screening efficacy of our model, performing better than most of other screening regimens. Our research provided a successful attempt that combination of several features integrated with SVM algorithm can help a lot in both classification (differential screening) and prediction (risk stratification) of prostate cancer. SVM and other ML (machine learning)/DL (deep learning) algorithms can be applied and integrated in more clinical scenarios, including not only differential screening/diagnosis or prognosis prediction of diseases.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable requests.

Consent statement

Written informed consent statement was waived by the Institutional Review Board.

Ethical approval

Institutional Review Board was obtained.

CRediT authorship contribution statement

Muyu Wu: Writing – review & editing, Writing – original draft, Software, Project administration, Methodology, Data curation, Conceptualization. **Yucan Zhang:** Visualization, Software, Methodology. **Xiaoqun Zhang:** Supervision. **Xiaozhu Lin:** Supervision, Resources. **Qiaoqiao Ding:** Validation, Software. **Peiyong Li:** Validation, Supervision, Resources, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was financially supported by Shanghai Municipal Science and Technology Commission 2019 “Science and Technology Innovation Action Plan” Funded Project (19441905900). Especially, the manuscript of our research is through thorough proof-reading and language polishment by Prof. Li and G. T. Mao.

References

[1] R.L. Siegel, K.D. Miller, Cancer Statistics, 2021 71 (1) (2021) 7–33, https://doi.org/10.3322/caac.21654.
[2] H. Sung, J. Ferlay, R.L. Siegel, Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries 71 (3) (2021) 209–249, https://doi.org/10.3322/caac.21660.
[3] C.H. Bangma, R.H. van Schaik, B.G. Blijenberg, M.J. Roobol, H. Lilja, U.H. Stenman, On the use of prostate-specific antigen for screening of prostate cancer in European Randomised Study for Screening of Prostate Cancer, European journal of cancer (Oxford, England : 1990 46 (17) (2010) 3109–3119, https://doi.org/10.1016/j.ejca.2010.09.031.
[4] I.M. Thompson, D.K. Pauler, P.J. Goodman, C.M. Tangen, M.S. Lucia, H.L. Parnes, et al., Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter, N. Engl. J. Med. 350 (22) (2004) 2239–2246, https://doi.org/10.1056/NEJMoa031918.
[5] M. Lekili, A. Ergen, I. Celebi, Zinc plasma levels in prostatic carcinoma and BPH, Int. Urol. Nephrol. 23 (2) (1991) 151–154, https://doi.org/10.1007/bf02549712.
[6] P. Whelan, B.E. Walker, J. Kelleher, Zinc, vitamin A and prostatic cancer, Br. J. Urol. 55 (5) (1983) 525–528, https://doi.org/10.1111/j.1464-410x.1983.tb03362.x.
[7] J.O. Ogunlewe, D.N. Osegbe, Zinc and cadmium concentrations in indigenous blacks with normal, hypertrophic, and malignant prostate, Cancer 63 (7) (1989) 1388–1392, https://doi.org/10.1002/1097-0142(19890401)63:7<1388::aid-cncr2820630725>3.0.co;2-m.
[8] V. Zaichick, T.V. Sviridova, S.V. Zaichick, Zinc in the human prostate gland: normal, hyperplastic and cancerous, Int. Urol. Nephrol. 29 (5) (1997) 565–574, https://doi.org/10.1007/bf02552202.
[9] Y. Zhang, Y. Wu, Z.Y. Gong, H.D. Ye, X.K. Zhao, J.Y. Li, et al., Distinguishing rectal cancer from colon cancer based on the support vector machine method and RNA-sequencing data, Current medical science 41 (2) (2021) 368–374, https://doi.org/10.1007/s11596-021-2356-8.
[10] M.A. Aswathy, M. Jagannath, An SVM approach towards breast cancer classification from H&E-stained histopathology images based on integrated features, Med. Biol. Eng. Comput. 59 (9) (2021) 1773–1783, https://doi.org/10.1007/s11517-021-02403-0.
[11] Y. Li, X. Xie, X. Yang, L. Guo, Z. Liu, X. Zhao, et al., Diagnosis of early gastric cancer based on fluorescence hyperspectral imaging technology combined with partial-least-square discriminant analysis and support vector machine, J. Biophot. 12 (5) (2019) e201800324, https://doi.org/10.1002/jbio.201800324.
[12] C.M. Lynch, B. Abdollahi, J.D. Fuqua, A.R. de Carlo, J.A. Bartholomai, R.N. Balgemann, et al., Prediction of lung cancer patient survival via supervised machine learning classification techniques, Int. J. Med. Inf. 108 (2017) 1–8, https://doi.org/10.1016/j.ijmedinf.2017.09.013.
[13] K. Kourou, T.P. Exarchos, K.P. Exarchos, M.V. Karamouzis, D.I. Fotiadis, Machine learning applications in cancer prognosis and prediction, Comput. Struct. Biotechnol. J. 13 (2015) 8–17, https://doi.org/10.1016/j.csbj.2014.11.005.
[14] S. Albaradei, M. Thafar, A. Alsaedi, C. Van Neste, T. Gojobori, M. Essack, et al., Machine learning and deep learning methods that use omics data for metastasis prediction, Comput. Struct. Biotechnol. J. 19 (2021) 5008–5018, https://doi.org/10.1016/j.csbj.2021.09.001.
[15] X.M. Lu, M.Y. Wu, S.W. Wang, J.C. Qin, P.Y. Li, An AIE/PET-based fluorescent probe for Zn2+/Al3+ detection and its application in fluorescence-assisted diagnosis for prostate cancer, Dyes Pigments (2022) 203, https://doi.org/10.1016/j.dyepig.2022.110372.
[16] Z. Medarova, S.K. Ghosh, M. Vangel, R. Drake, A. Moore, Risk stratification of prostate cancer patients based on EPS-urine zinc content, Am. J. Cancer Res. 4 (4) (2014) 385–393.
[17] V.C. Wakwe, E.P. Odum, C. Amadi, The impact of plasma zinc status on the severity of prostate cancer disease 60 (3) (2019) 162–168, https://doi.org/10.4111/icu.2019.60.3.162.
[18] A.G. Sarafanov, T.I. Todorov, J.A. Centeno, V. Macias, W. Gao, W.M. Liang, et al., Prostate cancer outcome and tissue levels of metal ions, Prostate 71 (11) (2011) 1231–1238, https://doi.org/10.1002/pros.21339.
[19] X.M. Li, L. Zhang, J. Li, Y. Li, H.L. Wang, G.Y. Ji, et al., Measurement of serum zinc improves prostate cancer detection efficiency in patients with PSA levels between 4 ng/mL and 10 ng/mL, Asian J. Androl. 7 (3) (2005) 323–328, https://doi.org/10.1111/j.1745-7262.2005.00044.x.
[20] M.Z. Temiz, O.O. Cakir, S. Aykan, S.H. Kucuk, N.O. Tiryakioglu, G. Bilek, et al., The use of serum zinc to prostate-specific antigen ratio as a biomarker in the prediction of prostate biopsy outcomes, Biomarkers Med. 13 (2) (2019) 59–68, https://doi.org/10.2217/bmm-2018-0260.