REVIEW

Editorial Process: Submission:12/22/2024 Acceptance:05/10/2025

Identification of Factors Affecting Prostate Cancer Using Machine Learning Methods: A Systematic Review

Serveh Mohammadi[1], Behzad Imani[2]*, Soheila Saeedi[3], Mohammad Ali Amirzargar[4]

Abstract

Background: Prostate cancer is identified as the second cause of malignancy worldwide and the fifth cause of death among men. Considering the upward trend in cancer incidence and mortality rate due to this disease, the identification of risk factors can be of great help in prevention and conservative measures. Also, due to the significant growth in artificial intelligence and machine learning methods, many risk factors can be studied by identifying the most commonly used methods. **Methods:** The articles reviewed in this study were from 4 main databases: PubMed, Scopus, Web of Science, and IEEE Xplore. This systematic review was based on Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. Searching the databases was conducted from the beginning of 2015 to February 17, 2024 were included. Only the articles investigating factors affecting prostate cancer using machine learning are included in this systematic review. Non-English language studies, studies that did not involve human participants, review studies, meta-analyses, letters to editors, and commentary were excluded. **Results:** The findings showed that China had the most research in identifying prostate cancer risk factors with machine learning algorithms. Age, PSA level (prostate-specific antigen), tPSA (total PSA), fPSA (free PSA), and PSAD (PSA density) were identified as the most important risk factors in prostate cancer. R-software and Python were most employed in the data analysis. Random forest, support vector machine, and logistic regression were utilized more than other machine learning methods. Among data sources, MCC-Spain, SEER (surveillance, Epidemiology, and End Results), PLCO (National Cancer Institute's Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial), and NCBI (National Center for Biotechnology Information) were registries that were used in the studies. **Conclusion:** This research can help researchers use machine learning methods with better performance and registered data sources and identify the most influential risk factors for prostate cancer prevention and screening.

Keywords: Machine learning- PRISMA- PSA- systematic review- prostatic neoplasms

Asian Pac J Cancer Prev, **26 (5)**, 1519-1528

Introduction

Despite numerous advancements in the management of prostate cancer, there remains a lack of comprehensive understanding regarding the etiological elements linked to its onset and advancement [1]. Prostate cancer ranks as the second most frequently identified cancer and the fifth primary contributor to cancer-related mortality in the global male population [2]. In 2022, the World Health Organization (WHO) published statistics on the prevalence of cancer, which showed 1,467,854 cases of prostate cancer in the complete dataset, and the death rate from this cancer was 394,430, which was 7.3% of all cancers taken [3]. According to the National Institutes of Health (NIH), in 2024, there will be an estimated 299,010 cases of prostate cancer, which are expected to cause 34,250 deaths worldwide; people over 66 years old are responsible for 55% of deaths [4]. Also, by the year 2030, it is estimated that the incidence of prostate cancer will rise to 1.7 million cases globally as a result of the expansion and aging of the worldwide population [5].

The prostate gland, a reproductive organ of dimensions akin to a walnut, is positioned inferior to the bladder and superior to the penis [6]. Acting as an exocrine gland together with seminal vesicles it generates a fluid essential nourishing and aiding in the movement of sperm cells (manufactured in the testes) pre- and post-ejaculation also plays a crucial role in safeguarding the sperm cells within the acidic milieu of the vaginal environment [6, 7].

With aging, the prostate undergoes a natural

*[1]Student Research Committee, Hamadan University of Medical Sciences, Hamadan, Iran. [2]Department of Operating Room, School of Paramedicine, Hamadan University of Medical Sciences, Hamadan, Iran. [3]Department of Health Information Technology, School of Allied Medical Sciences, Hamadan University of Medical Sciences, Shahid Fahmideh Blvd, Hamadan, Iran. [4]Department of Urology, Faculty of Medicine, Hamadan University of Medical Sciences, Hamadan, Iran. *For Correspondence: behzadiman@yahoo.com*

enlargement process referred to as benign prostatic hyperplasia (BPH), which causes lower urinary tract symptoms (such as obstructive and/or irritative voiding, mainly due to bladder neck compression) in approximately one-third of males aged over 60 and around half of those over 80 [4]. There are known causes, such as age, ethnicity, and genetics, which are non-modifiable risk factors [8, 9]. Due to the impossibility of acting on modifiable risk factors (diet, physical activity, obesity, and smoking), there is still the opportunity to apply policies to reduce the effect of convertible risk factors [10]. Evidence for the impact of beta-carotene and obesity on prostate cancer carcinogens was revealed in the 2018 World Cancer Research Fund and American Institute for Cancer Research expert report on prostate cancer [10]. An in-depth examination of elements that can be changed or avoided is essential [11, 12]. Moreover, the risk factors for PC entail a history of prostatitis, which is characterized by the inflammation of the prostate gland, and the use of medications that obstruct five alpha-reductase, a treatment for benign prostatic hyperplasia [13]. The etiology of prostate cancer remains incompletely elucidated, and the task of pinpointing definitive risk factors has presented considerable challenges [14]. However, the implication that nongenetic, specifically environmental, influences play a significant role in the expression of this characteristic is robustly indicated by the observation that Asian immigrants residing in Western nations exhibit a greater prevalence of prostate cancer in contrast to their counterparts in their countries of birth [15].

In the healthcare field, digitizing and storing huge amounts of medical data has played a crucial role in facilitating the utilization of artificial intelligence-driven methodologies in diagnosing, treating, and predicting diseases. In the coming years, AI approaches will be applied in nearly every area of medicine because of the growing complexity and volume of data in this field [16].

Artificial intelligence, commonly called AI, denotes a form of automated computational procedure integrated with pre-programmed intelligence, aiding decision-making within unfamiliar contexts. The machines are effectively conditioned to grasp concealed patterns or insights from a specified dataset through the progression in AI technologies, machine learning (ML) algorithms, and deep learning (DL) models built upon mathematical principles and statistical suppositions. These advanced algorithms have empowered AI-driven systems to enhance their predictive capabilities without explicit programming [17].

ML is a subset of AI in which the algorithm acquires knowledge from data without explicit programming [18]. ML is classified into supervised, semi-supervised, and unsupervised categories [19]. Today, machine learning methods are widely used to extract hidden knowledge from huge data sets in the health field. ML methods have also been used in many studies to diagnose prostate cancer or determine risk factors [20, 21]. Each of the conducted studies has introduced several risk factors affecting prostate cancer. but, based on the investigations carried out in different databases, no systematic review studies so far have been conducted to determine influential risk factors for prostate cancer using machine learning methods. This study aimed to introduce the influential factors on prostate cancer that have been identified with the help of ML methods by conducting a systematic review.

Materials and Methods

The guideline utilized for reporting the systematic review methodology and findings was the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [22]. This systematic review was conducted in the following steps:

Search strategy

A comprehensive search was conducted across four databases, including PubMed, Scopus, Web of Science, and IEEE Xplore, to investigate factors affecting prostate cancer using machine learning. Searching the databases was conducted from the beginning of 2015 to February 17, 2024. In the IEEE Xplore database, only documents that were in the two categories of “conferences” and “journals” were included in the review, but in Scopus, PubMed, and Web of Science, the limitation was only based on the Title of the articles. The keywords used for searching are listed in Table 1. These keywords comprised two main categories: keywords related to artificial intelligence and keywords related to prostate cancer. Keywords within each category were combined using the ‘OR’ operator, and then the ‘AND’ operator was used to combine the

Table 1. Keywords Related to Searching in the Database

Keywords used for Artificial Intelligence	Keywords used for disease
Machine learning” OR “Machine Learning” OR “Deep Learning” OR “Deep Learning” OR “Data Mining” OR “Neural Network” OR “Genetic Algorithms” OR “Supervised Machine Learning” OR “Unsupervised Machine Learning” OR “Supervised Machine Learning” OR “Unsupervised Machine Learning” OR “Random Forest” OR “Random Forest” OR “Random Forests” OR “Support Vector Machine” OR “Support Vector Machine” OR “Support Vector Network” OR “Support Vector Networks” OR “Bayes Theorem” OR “Bayesian Network” OR “Artificial Neural Network” OR “Clustering” OR “Decision Tree” OR “Decision Trees” OR “Regression” OR “Bayesian” OR “Naive Bayes” OR “K-Nearest Neighbors” OR “K-Nearest Neighbor” OR “Genetic Algorithms” OR “Gradient Boosting”	Prostatic Neoplasms” OR “Prostate Neoplasms” OR “Neoplasms, Prostate” OR “Neoplasm, Prostate” OR “Prostate Neoplasm” OR “Neoplasms, Prostatic” OR “Neoplasm, Prostatic” OR “Prostatic Neoplasm” OR “Prostate Cancer” OR “Cancer, Prostate” OR “Cancers, Prostate” OR “Prostate Cancers” OR “Cancer of the Prostate” OR “Prostatic Cancer” OR “Cancer, Prostatic” OR “Cancers, Prostatic” OR “Prostatic Cancers” OR “Cancer of Prostate”

two categories.

Study selection

Initially, a search was conducted across four databases using the mentioned keywords, and the retrieved articles from all databases were imported into EndNote reference management software. After searching and retrieving the articles, duplicates were removed. Following the removal of duplicates, the articles in EndNote were sorted by year. In the next step, the titles and abstracts of the articles were reviewed, and articles that appeared relevant were placed in a separate group for full-text review. Criteria for inclusion and exclusion were taken into account when evaluating the title and abstract of the articles, as detailed in Table 2.

Data Extraction

The full text of articles that seemed relevant was retrieved and analyzed, and the desired information was extracted from them. Also, the references of all articles were checked to find relevant articles. Any disagreement between the authors was resolved through discussion with BI. An Excel sheet was designed to extract the desired information from the articles. The following information was extracted from the articles and entered into the Excel spreadsheet: author name, year, country, type of publication (journal or conference paper), journal or conference name, machine learning methods, sample size, train and test size, risk factors, software, source of data, time of data gathering (prospective or retrospective) and limitations of studies. It is important to note that in this study, the country listed in the 'country' section corresponds to the first author's affiliation, as the country where the study was conducted was not specified in all articles.

Study analysis

The results of this study were reported descriptively, and due to the diverse findings, no meta-analysis was performed. EndNote software was used for data management and Excel software was used for data analysis.

Results

Results of the Literature Search

Four databases Scopus, PubMed, Web of Science, and IEEE Xplore were searched, and 3369 articles were retrieved from 2015 to February 17, 2024. From this number of retrieved articles, 1018 duplicate articles were removed, and 2351 articles remained for the title and abstract screen. After investigating the title and abstract of the remaining articles and matching them with the inclusion and exclusion criteria, 35 articles remained. The full text of 35 articles was retrieved and examined, and at this stage, 12 articles were removed for the reasons mentioned in Figure 1, finally, 23 articles were included in this systematic review. The desired information was extracted from these 23 articles.

General Characteristics of the Included Studies

Twenty-three articles were included in this systematic review (Supplementary Table 1). Among these studies, the oldest study was related to 2015 and the most recent was published in 2024. Most of the studies were published in 2019 (N=6, 26.09%) and 2018 (N=5, 21.74%), respectively.

The first authors of the studies included in this systematic review were from Australia, China, Germany, Iran, Israel, Italy, Serbia, Singapore, Spain, Taiwan, the UK, the United States, and Vietnam. The countries of China with 7 studies, Spain with 3 studies, Serbia, and Taiwan with 2 studies accounted for the most studies. Other countries had conducted one study each.

Out of 23 included studies, 21 studies were presented in journals and 2 studies were also presented in conferences. *"The Journal of Urology"* has published two articles in this field, and other journals have also published one study each.

Machine learning methods, software, and performance

In most studies, more than one machine-learning method was used for data analysis. Logistic regression, support vector machine, and random forest methods were used more than other machine learning methods so each of these methods was used in 6 studies. Also, the artificial neural network was used in 4 studies. Five studies did not mention the software used for data analysis.

R software was employed in 6 studies, Python in 5, and MATLAB in 3 for data analysis. SPSS, WinBUGS, SAS, and Kaluza software were among the other mentioned software for data analysis (Figure 2).

Source of data

The data collection method was prospective in 2 studies (8.7%) and retrospective in 21 (91.3%). The sample size used in different studies to train and test machine learning methods differed greatly. The lowest number of samples was 71, and the highest was 514,878. The median number of samples was 941. In 10 studies, how much data was used for training and how much was used for testing machine learning algorithms was not

Table 2. Inclusion and Exclusion Criteria for Selecting Articles

Inclusion Criteria	Exclusion Criteria
(1) Studies involving human participants diagnosed with prostate cancer	(1) Non-peer-reviewed articles, systematic reviews, meta-analyses, letters to editors, commentary
(2) Studies that investigate potential risk factors for prostate cancer used machine learning	(2) Studies involving animals or in vitro experiments
(3) Studies published in English	(3) Studies not reporting risk factors for prostate cancer.

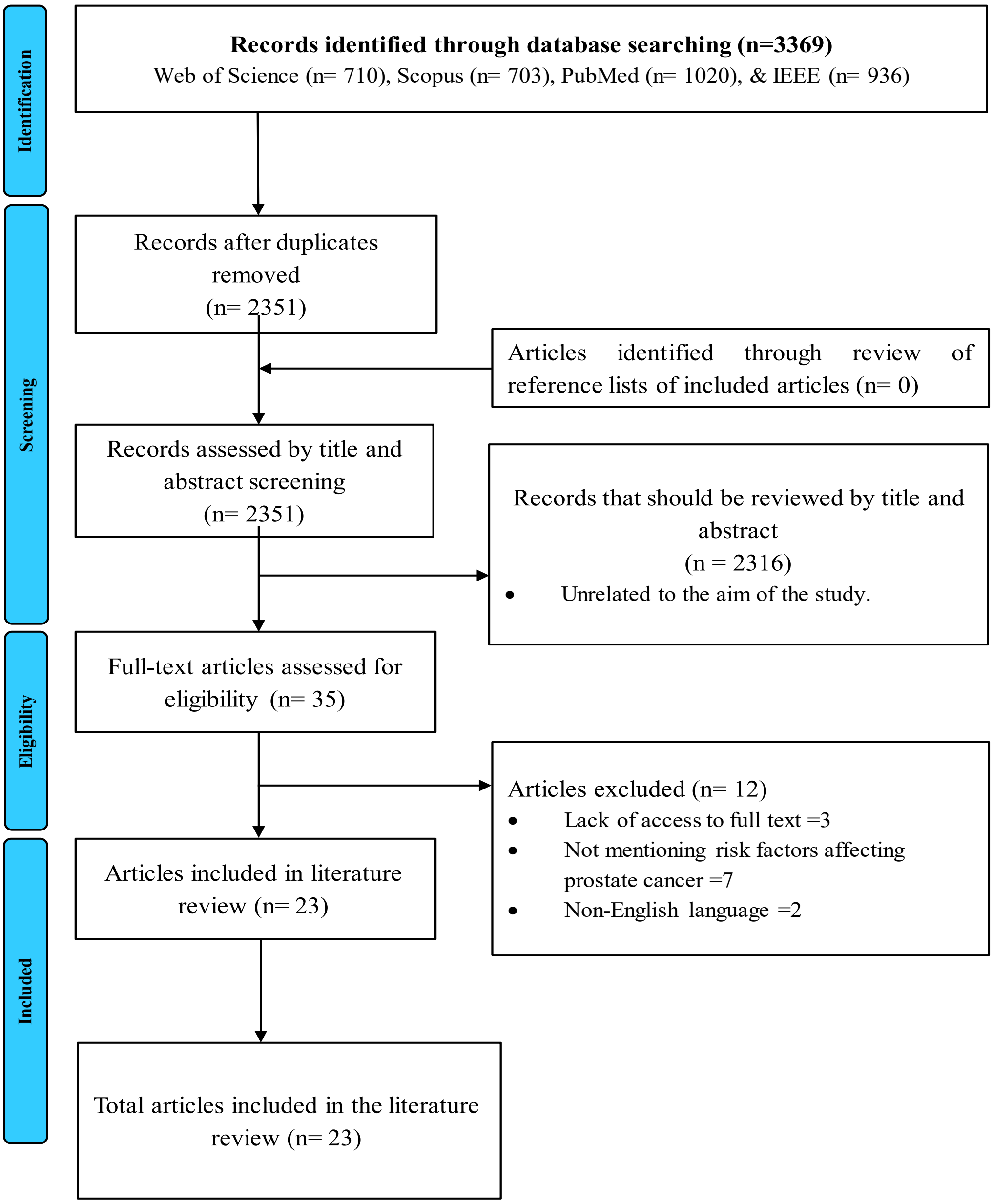

Figure 1. Flow Diagram of the Literature Search and Study Selection

mentioned. In the studies where the percentage of data for training and testing was mentioned, 70-80% of the data was used for training and 20-30% of the data was used for validation or testing. The source of data utilized

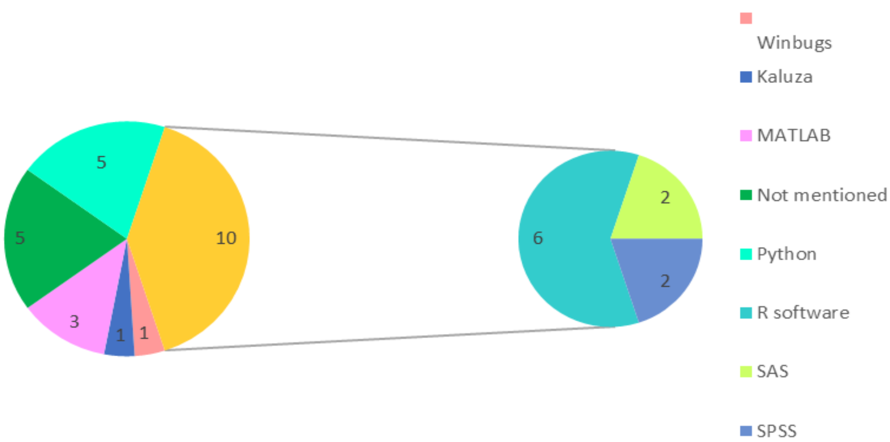

Figure 2. Distribution of Studies based on Software Used for Data Analysis

Table 3. Factors Affecting Prostate Cancer

#	Factors	References
1	Age	[23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33]
2	PSA levels	[24], [34], [35], [36], [28], [29], [31]
3	fPSA	[23], [25], [29], [30], [32], [33]
4	tPSA	[25], [26], [29], [30], [32], [33]
5	PSAD [PSA density]	[23], [37], [38], [39], [32]
6	Prostate volume	[23], [25], [26], [31]
7	Single nucleotide polymorphism	[40], [41]
8	Body Mass Index	[27], [42]
9	f/t PSA	[29], [32]
10	Boundary between internal and external glands	[24]
11	DRE results	[26]
12	Multiple genetic variants	[34]
13	Gleason score	[35]
14	Height	[27]
15	Some foods and drinks including intake of artificial sweeteners, sodas, bread, sugar, glucose, junk food, dairy desserts, tomatoes, tinned fish, red and processed meat, proteins, soy milk,	[27]
16	Vitamin D	[27]
17	MicroRNAs [miRs]	[36]
18	Heavy metals [Zn, AS, Mn, Sb]	[43]
19	Overall, ROC [rate of change]	[28]
20	Recent ROC	[28]
21	Metformin	[44]
22	Brain natriuretic peptide precursor	[30]
23	Free calcium	[30]
24	Apolipoprotein E ratio	[30]
25	Apolipoprotein A1	[30]
26	Creatine Protein T	[30]
27	Chloride	[30]
28	Phenotypic features	[45]
29	fPSA/PSA ratio	[31]
30	IGPSAD	[32]
31	EGPSAD	[32]
32	2D-US score	[32]
33	CEUS score	[32]
34	Elasticity score	[32]
35	Serum zinc concentration	[33]

PSA, Prostate Specific Antigen; PASD, Prostate Specific Antigen Density; fPSA, free Prostate-Specific Antigen; tPSA, total Prostate-Specific Antigen; DRE results, Digital Rectal Exam

Table 4. The Main Limitations Mentioned in the Studies

#	Limitations	References
1	Not taking into account all factors that may be useful for prostate cancer diagnosis	[24], [38], [39], [35], [43]
2	Small sample size	[32], [31], [45], [39], [38]
3	Single center	[38], [39], [31], [32],
4	Retrospective design	[38], [39], [29], [31]
5	Restricted generalization	[24], [39], [38]

in the studies was explained in all 23 articles. The most sources of data (12 studies) were hospitals. The MCC-Spain study database was also used in 3 studies. Also, in several studies, existing datasets or registries such as SEER, PLCO, and NCBI were used.

Factors affecting prostate cancer

Examining the articles included in this systematic review indicated that machine learning methods have revealed that many factors can affect prostate cancer. More than 30 factors were mentioned in these studies (Table 3). Among the mentioned factors, age was mentioned as an effective factor in 11 studies. Also, PSA levels, fPSA, tPSA, PSAD (PSA density), prostate volume, single nucleotide polymorphism, body mass index, and f/t PSA were other factors that were introduced as influential factors in more than one study.

Limitations of studies included in systematic review

Table 4 shows the main limitations mentioned in the studies. Among the constraints mentioned in the studies, two challenges of not considering all the effective factors and also the small sample size are more noticeable than other limitations.

Discussion

The present systematic review study investigated the application of machine learning methods to identify the most influential risk factors of prostate cancer. 23 articles were included in this study, and age was cited as an influential factor in 11 studies. Also, PSA levels, fPSA, tPSA, PSAD) (PSA density) prostate volume, single nucleotide polymorphism, body mass index, and f/t PSA were other factors that were introduced as influential factors in more than one study. This study showed that most studies used R and Python software for data analysis. The source of data utilized in the studies was explained in all 23 articles. Most of the data (12 studies) were from hospitals. The MCC-Spain study database was also used in 3 studies. Also, in several studies, existing datasets or registries such as SEER, PLCO, and NCBI were used. The most used machine learning methods were logistic regression, support vector machine, and random forest. Also, among the reviewed studies, 30% of the studies were conducted in China.

The results of this study showed that R software and Python were used more than other tools for data analysis. R software has several packages and libraries to assist with the development of artificial intelligence, including OneR, Ranger, iml, Tm, XGBoost, and partial dependence plots (PDP). Python software also has some libraries that can be used for data analysis including Scikit-learnt, TensorFlow, Pandas, NumPy, Matplotlib, SciPy, PyTorch, Keras, and Theano. Since this software has many capabilities, it can be used for machine learning studies.

One of the challenges that researchers usually face in machine learning studies is data collection. The results of this study showed that the data used in the field of prostate cancer were either extracted from patient's medical records or data stored in existing databases or registries were used. The MCC-Spain study database, SEER, PLCO, and NCBI were the databases in the prostate cancer domain that were used in the studies. MCC-Spain is a multicase-control study on cancer performed in Spain, with fieldwork conducted between 2008 and 2013 [27]. Ten thousand one hundred six participants, aged 20 to 85, were assessed in 23 different hospitals and primary care offices [46]. SEER (Surveillance, Epidemiology, and End Results), the only comprehensive population-based cancer database in the United States, makes studying racial and ethnic disparities in PCa mortality possible [35]. The Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial was a large randomized controlled trial designed and sponsored by the National Cancer Institute (NCI) [47] to determine the effects of screening on cancer-related mortality and secondary endpoints in men and women aged 55 to 74 [47]. NCBI (National Center for Biotechnology Information) acquires its data through three main methods: direct submissions from researchers, collaborations or agreements with national and international research groups and data providers, and its internal curation efforts [48]. Researchers are suggested to use registries or databases in future studies to access a large amount of data quickly.

Seven of the reviewed articles were conducted in China, and this shows the high statistics of studies about the identification of prostate cancer risk factors using machine learning methods; according to the Global Burden of Disease research, it is predicted that by 2050, the incidence of prostate cancer will account for an average of 3.03% of the total number of patients annually and 2.4% of deaths per year, which indicates the rising trend of the disease in this country [49]. Also, based on the studies, machine learning technology and artificial intelligence have grown significantly in this country [50]. Since prostate cancer is influenced by geographical and geospatial factors, the results of this study can be more applicable to Western Asian countries.

The results of this systematic review study showed that age is one of the influencing factors in prostate cancer. In the study conducted by Dite et al., in the age range of 60-69 years, the prevalence of prostate cancer was higher than in other age groups [51]. They also concluded that age is one of the influencing factors in prostate cancer; based on their research, older people are more likely to get the disease. In a study, Omankwu et al. concluded that men aged 55-69 are a key demographic information for prostate cancer screening, and age is an essential factor in clinical assessments and machine learning model accuracy [52]. Barlow et al.'s study highlights the age range of 55 to 69 years as a significant factor for prostate cancer risk assessment and screening, while the recommended routine screening age is over 70 years [28]. In Wang et al.'s study, they concluded that age is one of the most critical risk factors in the prostate cancer predictive model [53]. The result of this study is consistent with the findings of previous studies that identified age as a crucial determinant of prostate cancer risk.

The results of this study revealed that PSA levels, fPSA, tPSA, and PSAD were risk factors in prostate cancer. Harvey et al. [54] conducted a systematic review regarding the diagnostic accuracy of PSA in prostate cancer. Their study, which was performed on the European population, concluded that the accuracy of PSA for all types of prostate cancer can be generally considered. The systematic review study conducted by Merriel et al. [55] concluded that PSA is very sensitive for diagnosing prostate cancer in symptomatic patients. Ilic et al. concluded that PSA screening increases prostate cancer detection at early stages and reduces mortality but does not affect overall mortality [56]. The results of Nhung's study showed that tPSA is a crucial factor in differentiating prostate cancer and benign prostatic hyperplasia, with a tPSA range of less than four ng/ml. fPSA demonstrated 100% sensitivity and 81.2% specificity, demonstrating its outstanding performance as a PCa diagnosis technique. The results of their study showed the importance of tPSA, fPSA, and the f/tPSA ratio in accurately classifying prostate cancer patients, and each factor plays a vital role in prostate cancer diagnosis [25]. In both the US and Europe, PCa screening has become commonplace. For instance, widespread and strict prostate cancer screening programs have contributed to a recent drop in the death rate from PCa in the United States3. Naturally, the percentage of early PCa is rising as more and more advanced PCa is discovered and treated, and there might be a slight overdiagnosis and overtreatment. As a result, population-based prostate cancer screening is currently the subject of intense debate in both Europe and the US, and some policy recommendations would have the opposite effect. However, there should be a sizable number of extremely aggressive or advanced PCa cases in the population of China due to the lack of widespread PCa screening. Thus, a PSA-based PCa screening approach might be wise at this point in China and Western Asian countries [57].

The limitations observed in the reviewed studies included not considering all factors relevant to prostate cancer diagnosis, small sample sizes, single-center designs, retrospective designs, and limited generalizability. To address these limitations in future research, it is recommended to conduct studies with larger sample sizes, prospective designs, and multi-center collaborations. Additionally, considering all factors relevant to prostate cancer diagnosis, including genetic, environmental, and clinical factors, can enhance diagnostic accuracy. Furthermore, future studies should focus on the practical implications of the findings for clinical applications. For example, evaluating the effectiveness of various artificial intelligence-based methods, which this study highlighted for their high accuracy, in the early diagnosis of prostate cancer and determining risk factors influencing disease progression can improve prostate cancer management.

Limitations

Some limitations need to be addressed. Initially, only four databases were searched, and the search was restricted to English language studies. Additionally, the full text of certain eligible studies was unavailable. Furthermore, this study did not conduct a meta-analysis, making it challenging to assess the quality of the results. Despite these limitations, this systematic review provided valuable insights into the significance of identifying prostate cancer risk factors using machine learning methods. Data Collection Methods: Retrospective and prospective studies were reviewed. A total of 91.3% of the included studies employed retrospective data collection, which might increase the risk of bias. Sample Size: The sample size of the studies varied significantly, ranging from 71 to 514,878 participants, reflecting substantial diversity in the statistical power of the studies. Studies with small sample sizes (<500 participants) faced challenges in generalizing their findings. Transparency in Data Analysis: In 10 studies, data division into training and testing sets was not explicitly reported, indicating a lack of clarity in data handling. Adequacy of Study Design: Single-center studies encountered limitations in the generalizability of their findings due to the confined data sources.

One of the strengths of this study was that we checked the references of the articles, and the possibility that an unrelated article was included is low; also, in our study, all articles with any risk factor that they had investigated were included.

In conclusion, the results of this study showed that the most important risk factors for prostate cancer included age, PSA levels, fPSA, tPSA, and PSAD. Logistic regression, support vector machine, and random forest were the most employed machine learning methods. Although machine learning has identified these factors with high performance, the studies had reported some limitations, including not examining all risk factors, small sample size, single-center, retrospective data collection, and limited generalizability. It is suggested that future studies consider these limitations.

Author Contribution Statement

SM, BI, SS, and MA, conceptualized the study design and methodology and conducted the literature review. SM and SS proceeded with the data extraction and formal analysis of data. BI and SS collaborated with the study coordination (supervision and project administration). SM and SS prepared the first draft, and BI improved it. All authors read and approved the final manuscript.

Acknowledgements

We extend our heartfelt gratitude to all those who have helped in the study and preparation of the manuscript.

Scientific Approval

This systematic review is a component of Serveh Mohammadi MSc thesis

Funding

The study was funded by the Vice-chancellor for Research and Technology, Hamadan University of Medical Sciences (14030204783).

Statement of Ethics

This study was reviewed and approved by the Ethics Committee of Hamadan University of Medical Sciences with the approval number: IR.UMSHA.REC.1403.042.

This study does not require informed consent.

Data Availability

All data generated or analyzed during this study are included in the article.

Conflicts of 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.

References

1. Mandair D, Rossi RE, Pericleous M, Whyand T, Caplin ME. Prostate cancer and the influence of dietary factors and supplements: A systematic review. Nutr Metab. 2014;11:30. https://doi.org/10.1186/1743-7075-11-30.
2. Wang L, Lu B, He M, Wang Y, Wang Z, Du L. Prostate cancer incidence and mortality: Global status and temporal trends in 89 countries from 2000 to 2019. Front Public Health. 2022;10:811044. https://doi.org/10.3389/fpubh.2022.811044.
3. Ferlay J, Ervik M, Lam F, Laversanne M, Colombet M, Mery L, et al. Global cancer observatory: cancer today (version 1.1). Lyon, France: International Agency for Research on Cancer. 2024 Sep 23. Available from: https://gco.iarc.who.int/today.
4. Langan RC. Benign prostatic hyperplasia. Primary care. 2019;46(2):223-32. https://doi.org/10.1016/j.pop.2019.02.003.
5. Jain S, Saxena S, Kumar A. Epidemiology of prostate cancer in india. Meta gene. 2014;2:596-605. https://doi.org/10.1016/j.mgene.2014.07.007.
6. Barsouk A, Padala SA, Vakiti A, Mohammed A, Saginala K, Thandra KC, et al. Epidemiology, staging and management of prostate cancer. Med Sci (Basel). 2020;8(3). https://doi.org/10.3390/medsci8030028.
7. Hayward SW, Cunha GR. The prostate: Development and physiology. Radiol Clin North Am. 2000;38(1):1-14. https://doi.org/10.1016/s0033-8389(05)70146-9.
8. Eeles RA, Olama AA, Benlloch S, Saunders EJ, Leongamornlert DA, Tymrakiewicz M, et al. Identification of 23 new prostate cancer susceptibility loci using the icogs custom genotyping array. Nat Genet. 2013;45(4):385-91, 91e1-2. https://doi.org/10.1038/ng.2560.
9. Tan SH, Petrovics G, Srivastava S. Prostate cancer genomics: Recent advances and the prevailing underrepresentation from racial and ethnic minorities. Int J Mol Sci. 2018;19(4). https://doi.org/10.3390/ijms19041255.
10. Ziglioli F, Patera A, Isgrò G, Campobasso D, Guarino G, Maestroni U. Impact of modifiable lifestyle risk factors for prostate cancer prevention: A review of the literature. Front Oncol. 2023;13:1203791. https://doi.org/10.3389/fonc.2023.1203791.
11. Clinton SK, Giovannucci EL, Hursting SD. The world cancer research fund/american institute for cancer research third expert report on diet, nutrition, physical activity, and cancer: Impact and future directions. J Nutr. 2020;150(4):663-71. https://doi.org/10.1093/jn/nxz268.
12. World Cancer Research Fund/American Institute for Cancer Research. Diet, Nutrition PAaCAGPC, Project U. Available from: http://dietandcancerreport.org.
13. Lesourd M, Roumiguié M, Prudhomme T, Assaf E, Gamé X, Soulié M, et al. Impact of preoperative 5-alpha reductase inhibitor treatment in holmium laser enucleation of the prostate. Prog Urol. 2023;33(2):73-8. https://doi.org/10.1016/j.purol.2022.09.020.
14. Goldenberg SL, Nir G, Salcudean SE. A new era: Artificial intelligence and machine learning in prostate cancer. Nat Rev Urol. 2019;16(7):391-403. https://doi.org/10.1038/s41585-019-0193-3.
15. Kimura T. East meets west: Ethnic differences in prostate cancer epidemiology between east asians and caucasians. Chin J Cancer. 2012;31(9):421-9. https://doi.org/10.5732/cjc.011.10324.
16. Rezayi S, Ghazisaeedi M, Kalhori SRN, Saeedi S. Artificial intelligence approaches on x-ray-oriented images process for early detection of covid-19. J Med Signals sens. 2022;12(3):233-53. https://doi.org/10.4103/jmss.jmss_111_21.
17. Koteluk O, Wartecki A, Mazurek S, Kołodziejczak I, Mackiewicz A. How do machines learn? Artificial intelligence as a new era in medicine. J Pers Med. 2021;11(1):32. https://doi.org/10.3390/jpm11010032.
18. Panch T, Szolovits P, Atun R. Artificial intelligence, machine learning and health systems. J glob health. 2018;8(2):020303. https://doi.org/10.7189/jogh.08.020303.
19. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak. 2019;19(1):281. https://doi.org/10.1186/s12911-019-1004-8.
20. Saeedi S, Maghooli K, Amirazodi S, Rezayi S. Towards a better diagnosis of prostate cancer: Application of machine learning algorithms. Front Health Inform. 2022;11(1):119. https://doi.org/10.30699/fhi.v11i1.382.
21. Zhang W, Zhang K. Quantifying the contributions of environmental factors to prostate cancer and detecting risk-related diet metrics and racial disparities. Cancer Inform. 2023;22:11769351231168006. https://doi.org/10.1177/11769351231168006.
22. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The prisma statement. PLoS Med. 2009;6(7):e1000097. https://doi.org/10.1371/journal.pmed.1000097.
23. Ge P, Gao F, Chen G, editors. Predictive models for prostate cancer based on logistic regression and artificial neural network. 2015 IEEE International Conference on Mechatronics and Automation (ICMA); 2015: IEEE.
24. Xiao LH, Chen PR, Gou ZP, Li YZ, Li M, Xiang LC, et al. Prostate cancer prediction using the random forest algorithm that takes into account transrectal ultrasound findings, age, and serum levels of prostate-specific antigen. Asian J Androl. 2017;19(5):586-90. https://doi.org/10.4103/1008-682X.186884.
25. Nhung NTH, Khuong VTM, Huy VQ. Classifying prostate cancer patients based on total prostate-specific antigen and free prostate-specific antigen features by support vector machine. J Cancer Res Ther. 2016;12(2):818-25. https://doi.org/10.4103/0973-1482.172133.
26. Allameh F, Qashqai H, Salavati A. A dynamic model for predicting prostate cancer in iranian men based on a perceptron neural network. Int J Cancer Manag. 2017;10(3).
27. Lasheras JES, Lasheras FS, Donquiles CG, Tardón A, Castano-Vinyals G, Palazuelos C, et al. Hybrid algorithm for the classification of prostate cancer patients of the mcc-spain study based on support vector machines and genetic algorithms. Neurocomputing. 2021;452:386-94. https://doi.org/10.1016/j.neucom.2019.08.113.
28. Barlow H, Mao S, Khushi M. Predicting high-risk prostate cancer using machine learning methods. Data. 2019;4(3):129.
29. Perera M, Mirchandani R, Papa N, Breemer G, Effeindzourou

A, Smith L, et al. Psa-based machine learning model improves prostate cancer risk stratification in a screening population. World J Urol. 2021;39(6):1897-902. https://doi.org/10.1007/s00345-020-03392-9.
30. Li M, Fu X, Meng C. The Risk Prediction of Prostate Cancer Based on A Improved Hybrid Algorithm. In2020 IEEE International Conference on Smart Cloud (SmartCloud) 2020 Nov 6 (pp. 94-99). IEEE.
31. Zhang X, Feng C, Bai X, Peng X, Guo Q, Chen L, et al. Predicting positive repeat prostate biopsy outcomes: Comparison of machine learning approaches to identify key parameters and optimal algorithms. Arch Esp Urol. 2023;76(7):494-503. https://doi.org/10.56434/j.arch.esp.urol.20237607.61.
32. Zhang M, Liu Y, Yao J, Wang K, Tu J, Hu Z, et al. Value of machine learning-based transrectal multimodal ultrasound combined with psa-related indicators in the diagnosis of clinically significant prostate cancer. Front Endocrinol (Lausanne). 2023;14:1137322. https://doi.org/10.3389/fendo.2023.1137322.
33. Wu M, Zhang Y, Zhang X, Lin X, Ding Q, Li P. Combined measurement of serum zinc with psa ameliorates prostate cancer screening efficiency via support vector machine algorithms. Heliyon. 2024;10(2):e24292. https://doi.org/10.1016/j.heliyon.2024.e24292.
34. Li-Sheng Chen S, Ching-Yuan Fann J, Sipeky C, Yang TK, Yueh-Hsia Chiu S, Ming-Fang Yen A, et al. Risk prediction of prostate cancer with single nucleotide polymorphisms and prostate specific antigen. J urol. 2019;201(3):486-95. https://doi.org/10.1016/j.juro.2018.10.015.
35. Hanson HA, Martin C, O'Neil B, Leiser CL, Mayer EN, Smith KR, et al. The relative importance of race compared to health care and social factors in predicting prostate cancer mortality: A random forest approach. J urol. 2019;202(6):1209-16. https://doi.org/10.1097/JU.0000000000000416.
36. Mello-Grand M, Gregnanin I, Sacchetto L, Ostano P, Zitella A, Bottoni G, et al. Circulating micrornas combined with psa for accurate and non-invasive prostate cancer detection. Carcinogenesis. 2019;40(2):246-53. https://doi.org/10.1093/carcin/bgy167.
37. Zhang Y-Y, Li Q, Xin Y, Lv W-Q. Differentiating prostate cancer from benign prostatic hyperplasia using psad based on machine learning: Single-center retrospective study in china. IEEE/ACM Trans Comput Biol Bioinform. 2018;16(3):936-41. https://doi.org/10.1109/TCBB.2018.2822675.
38. Pantic DN, Stojadinovic MM, Stojadinovic MM. Decision tree analysis for prostate cancer prediction in patients with serum psa 10 ng/ml or less. Exp appl biomed res. 2020;21(1):43-50. https://doi.org/10.2478/sjecr-2018-0039.
39. Stojadinovic M, Stojadinovic M, Pantic D. Decision tree analysis for prostate cancer prediction. Srp Arh Celok Lek. 2019;147(00):39. https://doi.org/10.2298/SARH181127039S.
40. Sánchez Lasheras JE, González Donquiles C, García Nieto PJ, Jiménez Moleon JJ, Salas D, Suárez Gómez SL, et al. A methodology for detecting relevant single nucleotide polymorphism in prostate cancer with multivariate adaptive regression splines and backpropagation artificial neural networks. Neural Comput Appl. 2020;32:1231-8. https://doi.org/10.1007/s00521-018-3503-4.
41. Sánchez Lasheras JE, Suárez Gómez SL, Santos JD, Castaño-Vinyals G, Pérez-Gómez B, Tardón A. A multivariate regression approach for identification of snps importance in prostate cancer. J Exp Theor Artif Intell. 2019;31(6):817-28. https://doi.org/10.1080/0952813X.2018.1552319.
42. Yeh S-J, Chung Y-C, Chen B-S. Investigating the role of obesity in prostate cancer and identifying biomarkers for drug discovery: Systems biology and deep learning approaches. Molecules. 2022;27(3):900. https://doi.org/10.3390/molecules27030900.
43. Lim JT, Tan YQ, Valeri L, Lee J, Geok PP, Chia SE, et al. Association between serum heavy metals and prostate cancer risk–a multiple metal analysis. Environ int. 2019;132:105109. https://doi.org/10.1016/j.envint.2019.105109.
44. Dankner R, Agay N, Olmer L, Murad H, Keinan Boker L, Balicer RD, et al. Metformin treatment and cancer risk: Cox regression analysis, with time-dependent covariates, of 320,000 persons with incident diabetes mellitus. Am J Epidemiol. 2019;188(10):1794-800. https://doi.org/10.1093/aje/kwz157.
45. Hood SP, Cosma G, Foulds GA, Johnson C, Reeder S, McArdle SE, et al. Identifying prostate cancer and its clinical risk in asymptomatic men using machine learning of high dimensional peripheral blood flow cytometric natural killer cell subset phenotyping data. Elife. 2020;9:e50936. https://doi.org/10.7554/eLife.50936.
46. Castaño-Vinyals G, Aragonés N, Pérez-Gómez B, Martín V, Llorca J, Moreno V, et al. Population-based multicase-control study in common tumors in spain (mcc-spain): Rationale and study design. Gac sanit. 2015;29(4):308-15. https://doi.org/10.1016/j.gaceta.2014.12.003.
47. Black A, Huang WY, Wright P, Riley T, Mabie J, Mathew S, et al. Plco: Evolution of an epidemiologic resource and opportunities for future studies. Rev Recent Clin Trials. 2015;10(3):238-45. https://doi.org/10.2174/157488711003150928130654.
48. Sayers EW, Cavanaugh M, Clark K, Pruitt KD, Sherry ST, Yankie L, et al. Genbank 2023 update. Nucleic Acids Res. 2023;51(D1):D141-d4. https://doi.org/10.1093/nar/gkac1012.
49. Visualization. IfHMaEIGF. Seattle, WA: IHME, University of Washington, 2024.
50. Wu F, Lu C, Zhu M, Chen H, Zhu J, Yu K, et al. Towards a new generation of artificial intelligence in china. Nat Mach Intell. 2020;2(6):312-6. https://doi.org/10.1038/s42256-020-0183-4.
51. Dite GS, Spaeth E, Murphy NM, Allman R. Development and validation of a simple prostate cancer risk prediction model based on age, family history, and polygenic risk. Prostate. 2023;83(10):962-9. https://doi.org/10.1002/pros.24537.
52. Omankwu OC, Osodeke EC, Kanu C. Machine learning algorithms for predicting high-risk of prostate cancer using prostate-specific antigen (psa), age and body mass index (bmi). NIPES J Sci Technol Res. 2023;5(1):133-49.
53. Wang W, Jin X. Prostate cancer prediction model: A retrospective analysis based on machine learning using the mimic-iv database. Intelligent Pharmacy. 2023;1(4):268-73. https://doi.org/10.1016/j.ipha.2023.04.010.
54. Harvey P, Basuita A, Endersby D, Curtis B, Iacovidou A, Walker M. A systematic review of the diagnostic accuracy of prostate specific antigen. BMC urol. 2009;9:1-9. https://doi.org/10.1186/1471-2490-9-14.
55. Merriel SW, Pocock L, Gilbert E, Creavin S, Walter FM, Spencer A, et al. Systematic review and meta-analysis of the diagnostic accuracy of prostate-specific antigen (psa) for the detection of prostate cancer in symptomatic patients. BMC Med. 2022;20(1):54. https://doi.org/10.1186/s12916-021-02230-y.
56. Ilic D, Djulbegovic M, Jung JH, Hwang EC, Zhou Q, Cleves A, et al. Prostate cancer screening with prostate-specific antigen (psa) test: A systematic review and meta-analysis. BMJ. 2018;362:k3519. https://doi.org/10.1136/bmj.k3519.
57. Chen J, He L, Ni Y, Yu F, Zhang A, Wang X, et al. Prevalence

and associated risk factors of prostate cancer among a large chinese population. Sci Rep. 2024;14(1):26338. https://doi.org/10.1038/s41598-024-77863-z.