Clinical Research

Accuracy of machine learning models using ultrasound images in prostate cancer diagnosis: a systematic review

Retta Catherina Sihotang[1], Claudio Agustino[1], Ficky Huang[1], Dyandra Parikesit[2], Fakhri Rahman[1], Agus Rizal Ardy Hariandy Hamid[1]

pISSN: 0853-1773 • eISSN: 2252-8083
https://doi.org/10.13181/mji.oa.236765
Med J Indones. 2023;32:112–21

Received: January 31, 2023
Accepted: September 11, 2023

Authors' affiliations:
[1]Department of Urology, Faculty of Medicine, Universitas Indonesia, Cipto Mangunkusumo Hospital, Jakarta, Indonesia, [2]Urology Medical Staff Group, Universitas Indonesia, Universitas Indonesia Hospital, Depok, Indonesia

Corresponding author:
Agus Rizal Ardy Hariandy Hamid
Department of Urology, Faculty of Medicine, Universitas Indonesia, Cipto Mangunkusumo Hospital, Jalan Salemba Raya No. 6, Central Jakarta 10430, DKI Jakarta, Indonesia
Tel/Fax: +62-21-3912477
E-mail: rizalhamid.urology@gmail.com

ABSTRACT

BACKGROUND In prostate cancer (PCa) diagnosis, many developed machine learning (ML) models using ultrasound images show good accuracy. This study aimed to analyze the accuracy of neural network ML models in PCa diagnosis using ultrasound images.

METHODS The protocol was registered with PROSPERO registration number CRD42021277309. Three reviewers independently conducted a literature search in 5 online databases (PubMed, EBSCO, Proquest, ScienceDirect, and Scopus). We included all cohort, case-control, and cross-sectional studies in English, that used neural networks ML models for PCa diagnosis in humans. Conference/review articles and studies with combination examination with magnetic resonance imaging or had no diagnostic parameters were excluded.

RESULTS Of 391 titles and abstracts screened, 9 articles relevant to the study were included. Risk of bias analysis was conducted using the QUADAS-2 tool. Of the 9 articles, 5 used artificial neural networks, 1 used deep learning, 1 used recurrent neural networks, and 2 used convolutional neural networks. The included articles showed a varied area under the curve (AUC) of 0.76–0.98. Factors affecting the accuracy of artificial intelligence (AI) were the AI model, mode and type of transrectal sonography, Gleason grading, and prostate-specific antigen level.

CONCLUSIONS The accuracy of neural network ML models in PCa diagnosis using ultrasound images was relatively high, with an AUC value above 0.7. Thus, this modality is promising for PCa diagnosis that can provide instant information for further workup and help doctors decide whether to perform a prostate biopsy.

KEYWORDS artificial intelligence, machine learning, neural network model, prostate cancer, ultrasonography

Prostate cancer (PCa) is the third most common cancer globally and the second most common in men.[1] It significantly affects male health, and early detection facilitates curative treatment and reduces disease morbidity and mortality.[2,3]

Ultrasonography has a potential for PCa imaging because it is cost-effective, practical, and widely available.[4] However, standard transrectal ultrasound (TRUS) alone is not reliable due to its low sensitivity and specificity in detecting PCa.[5] The current gold standard for PCa detection is a prostate biopsy performed under TRUS guidance.[2,3,6,7] While ultrasonography is widely available, TRUS can be less comfortable for patients than the transabdominal approach. The best instruments currently available yield inaccurate results. More accurate diagnostic instruments are required to effectively detect disorders. Technological advancements, such as artificial intelligence (AI), may help overcome these challenges.[8,9]

AI is a revolutionary technology in the healthcare field that is gaining interest. Neural networks, such as artificial neural networks (ANNs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs), are machine learning (ML) models that mimic human biological neurons. For PCa, AI has been shown to aid in standardized pathological grading to guide cancer stratification and treatment. Nitta et al[10] and Djavan et al[11] applied ML models to predict PCa based on prostate-specific antigen (PSA) concentrations. ML tended to be superior to conventional methods, with a region-wise area under the receiver operating characteristic curve (ROC-AUC) value ranging from 0.63 to 0.91.

The accuracy of ML based on data from ultrasonography as the primary modality has been debated. Thus, this review aimed to analyze the accuracy of neural networks trained on ultrasound images for PCa diagnosis.

METHODS

Protocol registration

The protocol for this systematic review was registered with PROSPERO registration number CRD42021277309.

Search strategy

Three reviewers (RCS, CA, and FH) independently conducted a literature search of five online databases on January 13, 2023. The databases were PubMed, EBSCO, ProQuest, ScienceDirect, and Scopus. The following keywords with various combinations were used: "Prostate Cancer," "Machine Learning OR Neural Network," "Diagnosis," and "Ultrasonography" (Figure 1). The reference lists of the articles retrieved from the literature search were also reviewed to identify other relevant studies.

Study selection and data extraction

All articles that used ultrasound images to demonstrate the application of ML to the diagnosis of PCa were included. The literature search was limited to publications in English without regard to the publication date. A study was considered significant if it met the inclusion criteria, including using human participants, neural networks, ML models, and prostate biopsy as the criterion for diagnosis. Cohort, case-control, and cross-sectional studies were included. Conference or review articles and studies that involved a combined examination with magnetic resonance imaging (MRI) or had no diagnostic parameters were excluded. Three reviewers (RCS, CA, and FH) individually reviewed

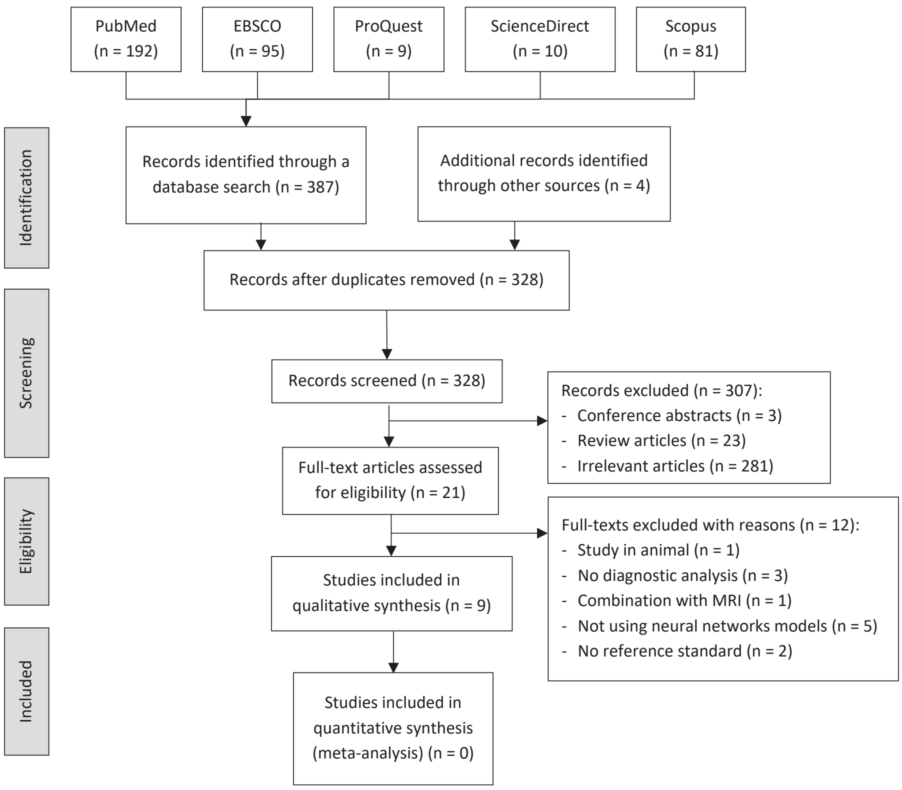

Figure 1. PRISMA flow diagram for the current study (a total of 391 articles obtained). MRI=magnetic resonance imaging; PRISMA=Preferred Reporting Items for Systematic Reviews and Meta-Analyses

the titles and abstracts of the selected studies. Disagreements were resolved through discussions with senior reviewers until a consensus was reached. All authors agreed with the final list of papers selected for extraction. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram was used to assist in selecting the articles.

The data extracted from the included articles were tabulated to summarize the outcomes. The data collection points included the number of samples and participants, ultrasound modes, ML methods, system specifications, software tools, programming languages, ML input data, ML outcomes, and diagnostic performance. The primary outcome was the accuracy of neural network ML models for PCa diagnosis. Additionally, the neural network models were compared with other ML models; we compared their available diagnostic performance data, including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and ROC-AUC. The receiver operating characteristic is a graph showing the performance of a classification model at all classification thresholds to determine its accuracy. The area under the curve (AUC) is the probability that a classifier ranks a randomly selected positive example more highly than a randomly selected negative example. Based on the test, an AUC of 0.5 indicates the inability to distinguish between patients with and without disease or condition, 0.7–0.8 is acceptable, 0.8–0.9 is considered excellent, and >0.9 is outstanding.

Risk of bias assessment

The methodological quality of the research was independently evaluated by three reviewers (RCS, CA, and FH) using the QUADAS-2 tool in the Review Manager software version 5.4 (Cochrane, United Kingdom) for Mac. The reviewers were not blinded to the identities of the authors of the articles, journals, and publishers. Based on the questions in the QUADAS-2 tool, the risks of bias were categorized as high, unclear, and low.

RESULTS

Of the 391 retrieved articles, only 9 met the inclusion criteria (Figure 1). The quality assessment of the included articles is shown in Table 1 using the QUADAS-2 tool. Several articles included in the analysis had an unclear or high risk of bias. Unclear risk of bias was common for the index test parameters due to the unclear threshold of the index test. Meanwhile, a high risk of bias was also common because the interpretation was limited to standard results in several articles.[12–14]

The characteristics of each study are presented in Table 2.[12–20] Five studies used an ANN, one used deep learning (DL), one used an RNN, and two used a CNN. Nine of the included studies had a cross-sectional design. All studies examined adult males with an unknown age range owing to unclear data. The sample sizes ranged from 48 to 1,151 patients; however, the studies by Ronco and Fernandez[12] and Akatsuka et al[13] only provided the number of cases. Five studies used TRUS data only for the input parameters, whereas the others used a combination of input data from clinical findings. All studies showed various accuracy analysis parameters, including

Table 1. Risk of bias assessment using the QUADAS-2 tool

First author, year	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Akatsuka,[13] 2022	Low	High	Low	Low	Low	Unclear	Low
Azizi,[17] 2018	Unclear	Unclear	Low	Low	Low	Unclear	Low
Hassan,[19] 2022	High	High	Low	Low	Unclear	Unclear	Low
Lee,[15] 2006	Low	Unclear	Low	Low	Low	Low	Low
Lee,[16] 2010	Low	Unclear	Low	Low	Low	Unclear	Low
Loch,[14] 1999	Unclear	Low	Low	Low	Low	Low	Low
Lorusso,[20] 2023	Unclear	Low	Low	Low	Unclear	Low	Low
Ronco and Fernandez,[12] 1999	Unclear	High	Low	Low	Unclear	High	Low
Wildeboer,[18] 2020	Low	Unclear	Low	Low	Low	Unclear	Low

Table 2. Characteristics and performance result of included studies

First author, year	Country	Samples	Imaging	ML method	System specifications, software tools, and programming language	Input data	Outcome	Performance results
Ronco and Fernandez,[12] 1999	Uruguay	442 cancer and benign cases	TRUS	ANN	1. System specifications: NA 2. Software tools: NeuroGenetic Optimizer version 2.5 for Windows 1995 3. Programming language: NA	1. Ultrasonographic variables (transverse axis, anteroposterior axis, longitudinal axis, prostatic volume, central zone, echoic level, volume of the pathological area, major diameter of the pathological area, minor diameter of the pathological area, presence/absence of calcifications, degree of bladder impression, PSA density [PSA/volume], and ultrasonographic diagnosis) 2. Non-ultrasonographic variables (age, previous clinical diagnosis, PSA level, and number of biopsies)	Accuracy for detecting PCa	1. PPV: 0.82 2. NPV: 0.97
Loch,[14] 1999	USA	553 specimens from 61 patients with confirmed PCa	TRUS	ANN	1. System specifications: NA 2. Software tools: Neuro-shell 2, Ward Systems Group, Inc., Frederick, MD 3. Programming language: NA	TRUS findings	Accuracy for detecting PCa	1. Benign pathology: 99% classified correctly 2. Cancer: 71% classified correctly
Lee,[15] 2006	Korea	684 patients who had undergone TRUS-guided prostate biopsy	TRUS and Doppler ultrasonography	ANN	1. System specifications: NA 2. Software tools: NeuroSolutions version 4.0, NeuroDimension Inc., Gainesville, FL 3. Programming language: NA	1. Model 1 (age, DRE findings, PSA level, PSA density, transitional zone volume, and PSA density in the transitional zone) 2. Model 2 (age, DRE findings, PSA level, PSA density, transitional zone volume, PSA density in the transitional zone, and TRUS findings [positive, suspicious, or negative])	Diagnostic performance of 2 ANN models	1. Model 2 showed better accuracy than Model 1. 2. Accuracy Model 1 (AUC PSA 0–4: 0.738, PSA 4–10: 0.753, and PSA>10: 0.774) 3. Accuracy Model 2 (AUC PSA 0–4: 0.859, PSA 4–10: 0.797, and PSA>10: 0.894)

Table continued on next page

Table 2. (continued)

First author, year	Country	Samples	Imaging	ML method	System specifications, software tools, and programming language	Input data	Outcome	Performance results
Lee,[16] 2010	Korea	1,077 patients who had undergone TRUS-guided prostate biopsy	TRUS and Doppler ultrasonography	MLRA, ANN, and SVM	1. System specifications: NA 2. Software tools and models: (MLRA: SPSS version 15, SVM: LIBSVM for multiclass classification, and ANN: detailed model of three-layer perceptron architecture, consisting of one input layer, one hidden layer, and one output layer) 3. Programming language: NA	Age, DRE findings, PSA level, PSA density, transitional zone volume, PSA density in the transitional zone, and TRUS findings (class I–V based on lesion location, outline, shape, and vascularity)	Accuracy of each model	1. ROC MLRA: 0.768 2. ROC ANN: 0.778 3. ROC SVM: 0.847
Azizi,[17] 2018	Canada	157 patients who had undergone prostate biopsy	TeUS	RNN comparing LSTM, GRU, vanilla RNN, and spectral	1. System specifications: GeForce GTX 980 Ti GPU with 6 GB of memory, hosted by a machine running Ubuntu 16.04 operating system on a 3.4 GHz Intel Core™ i7 CPU with 16 GB of memory 2. Software tools: NA 3. Programming language: Python 2.7	TeUS findings	Accuracy for detecting PCa	1. LTSM (specificity: 0.98, sensitivity: 0.76, accuracy: 0.93, and AUC: 0.96) 2. GRU (specificity: 0.95, sensitivity: 0.70, accuracy: 0.86, and AUC: 0.92) 3. Vanilla RNN (specificity: 0.72, sensitivity: 0.69, accuracy: 0.75, and AUC: 0.76) 4. Spectral (specificity: 0.73, sensitivity: 0.63, accuracy: 0.78, and AUC: 0.76)
Wildeboer,[18] 2020	Netherland	48 men with confirmed PCa	B-mode US, SWE, and DCE-US	DL	NA	TRUS findings	Accuracy of each model to localize PCa using ultrasound images	1. ROC-AUC for PCa: 0.75 2. ROC-AUC for Gleason >3+4: 0.90

Table continued on next page

Table 2. (continued)

First author, year	Country	Samples	Imaging	ML method	System specifications, software tools, and programming language	Input data	Outcome	Performance results
Hassan,[19] 2022	USA	61,119 images from 1,151 patients in open-source databases	TRUS	CNN compared to Nearest Neighbor, Gradient Boosting, SVM, and Random Forest	1. System specifications: PC with Core i7 11th generation intel processor, 16GB DDR RAM 2. Software tools: not mentioned (NN Models VGG-16) 3. Programming language: Python version 3.7 with Tensorflow 2.x and scikit learn 0.242 version	TRUS images	Accuracy of each model to efficiently classify PCa	Accuracy (VGG-16): (CNN: 0.99, Nearest Neighbor: 0.869, Gradient Boosting: 0.871, SVM: 0.872, and Random Forest: 0.875)
Akatsuka,[13] 2022	Japan	2,676 images from 691 cases	TRUS	SVM and CNN + SVM	1. System specifications: NA 2. Software tools and models (SVM: the e1071 package [version 1.7.0] and CNN: Grad-CAM) 3. Programming language: NA	1. Still ultrasound image data 2. Clinical data (age and PSA) 3. Integrated data (ultrasound image, total prostate volume, PSA density, age, and PSA)	Accuracy to detect high-grade PCa	1. ROC-AUC clinical data only (SVM): 0.691 2. ROC-AUC integrated data (CNN + SVM): 0.835
Lorusso,[20] 2023	German	64 patients with PCa	TRUS	ANN	NA	TRUS images	Diagnostic performances of the ANNA/C-TRUS system	1. Overall (sensitivity: 0.62, specificity: 0.81, NPV: 0.80, PPV: 0.64, and accuracy: 0.78) 2. PCa index lesion (sensitivity: 0.60, specificity: 0.87, NPV: 0.88, PPV: 0.58, and accuracy: 0.81) 3. ISUP grade >2 (sensitivity: 0.69, specificity: 0.77, NPV: 0.88, PPV: 0.50, and accuracy: 0.75) 4. Gleason 4 or 5 (sensitivity: 0.70, specificity: 0.74, NPV: 0.91, PPV: 0.41, and accuracy: 0.74)

ANN=artificial neural network; ANNA/C-TRUS=computerized artificial neural network analysis; AUC=area under the curve; CNN=convolutional neural network; DCE-US=dynamic contrast-enhanced ultrasound; DL=deep learning; DRE=digital rectal examination; Grad-CAM=gradient-weighted class activation mapping; GPU=graphic processing unit; GRU=gated recurrent units; LSTM=long short-term memory; ML=machine learning; MLRA=multiple logistic regression analysis; NA=not available; NPV=negative predictive value; PCa=prostate cancer; PPV=positive predictive value; PSA=prostate-specific antigen; RNN=recurrent neural network; ROC=receiver operating characteristic; ROC-AUC=a region-wise area under the receiver operating characteristics curve; SVM=support vector machine; SWE=shear-wave elastography; TeUS=temporal enhanced ultrasound; TRUS=transrectal ultrasonography

AUC, PPV, NPV, sensitivity, and specificity (Table 2). However, Loch et al[14] only used percentages. The performance results are presented in Table 2. Due to the varied parameters, a quantitative analysis could not be performed. Most of the studies used the AUC as an accuracy parameter. The AUC values of all the studies were greater than 0.7, ranging from 0.75 to 0.98.

DISCUSSION

Based on the included studies, the overall accuracy of ML showed promising results. The AUC values of nine studies were greater than 0.7, ranging from 0.75 to 0.98. Wildeboer et al[18] assessed a potential DL model based on TRUS B-mode US, shear-wave elastography (SWE), and dynamic contrast-enhanced ultrasound (DCE-US). The multiparametric classifier showed an AUC of 0.90 compared with 0.75 for the best-performing individual parameters for PCa and Gleason scores >3+4 significant PCa. This study revealed that combinations of the available modes were favored over a single mode. Lee et al[15] evaluated the accuracies of multiple logistic regression, ANN, and support vector machine (SVM) models in predicting the prostate biopsy outcomes of 684 patients (214 were confirmed to have PCa). The models were developed using the following input data: age, digital rectal examination (DRE) findings, PSA parameters, and TRUS findings. This study showed that image-based clinical decision support systems (ANN and SVM) were more accurate than multiple logistic

Table 3. Comparison of advantages and disadvantages of several ML models

ML models	Advantages	Disadvantages
ANN[26]	1. Stores data over an entire network 2. Capacity to operate with little information 3. Can overlook errors 4. Possesses a distributed memory system	1. Hardware reliant 2. Unexplained the network's behavior 3. Establishment of an appropriate network structure
CNN[26]	1. Extremely high accuracy when it comes to picture recognition challenges 2. Detects critical traits automatically and without human intervention 3. Weight distribution	1. Does not encode an object's location or orientation 2. Inability to be spatially invariant with respect to the supplied data 3. Requires numerous training data sets
RNN[27]	1. Retains all information over time and beneficial for time series prediction 2. Utilizes convolutional layers with RNNs to broaden the effective pixel neighborhood	1. Gradient difficulties of disappearing and exploding 2. Quite difficult to train 3. Incapable of processing extremely lengthy sequences
Linear regression[26]	1. Works exceptionally well in small data sets 2. Easy to build and comprehend 3. Analyzes model parameters in a statistical sense	1. Can only work in data sets that have linear relation 2. Overconfidence in the logic models 3. Can only classify dichotomous variables except multinomial linear regression
SVM[28,29]	1. Can handle several feature spaces with less risk of overfitting 2. Capable of classifying semi-structured and unstructured data well, such as texts or images	1. Results, weights, and impacts of variables are harder to comprehend and interpret. 2. Data's noise significantly impacts the classification results. 3. Expansive to build in a large data set environment
DT[28,29,30]	1. Results are simpler to comprehend and interpret. 2. Less time consuming data preparation 3. Can produce reliable classifiers that can be confirmed with statistical tests	1. Mutually exclusive classes 2. If any attribute or variable value for a non-leaf node is absent, the algorithm will not branch. 3. Less superior compared to ANN
RF[28,29]	1. A lower possibility of variance and overfitting of training data, compared to DT 2. Performs well in large data sets 3. Can calculate which variables or qualities are most significant in the categorization	1. Far more complex and expansive to build 2. When estimating variable significance, it favors variables or qualities that may take a large number of alternative values. 3. Commonly overfitting

ANN=artificial neural network; CNN=convulotional neural network; DT=decision tree; ML=machine learning; RF=Random Forest; RNN=recurrent neural network; SVM=support vector machine regression models. They evaluated the diagnostic performance of the ANN model with and without TRUS data. The ANN model used the primary input data of age, PSA levels, and DRE findings. However, with additional TRUS data, the ANN model showed better accuracy and a higher AUC value than without TRUS data. Azizi et al[17] proposed the temporal modeling of temporal enhanced ultrasound (TeUS) using an RNN to improve cancer detection accuracy. The TeUS data were acquired from 157 patients during fusion prostate biopsy. The model achieved an AUC value of 0.96. Hassan et al[19] demonstrated a higher accuracy (0.99) with a CNN (VGG-16) than with other algorithms (Gradient Boosting, SVM, and Random Forest). Akatsuka et al[13] reported an AUC of 0.835 for CNN combined with an SVM built on clinical data and TRUS images. This was higher than the AUC for the SVM based on only clinical data. A recent study by Lorusso et al[20] demonstrated increasing sensitivity and NPV of the ANN method using TRUS images for higher grades of PCa.

Several factors influence the accuracies of models, including the AI model, TRUS modes, amount of input data, Gleason grading, and PSA concentrations. Based on the analysis of each AI model (Table 4), two included studies highlighted the superior diagnostic performance of the neural network model to those of other models.[13,20] ANN and CNN outperformed the other neural network models in terms of diagnostic performance.[14,15,19] TRUS modes are substantially related to the accuracy, with DCE-US/SWE/TeUS improving the visualization and distinction of prostate tissues over the B-mode. The amount of input data is also important for reliable predictions by ANN models. More complicated data will result in a more accurate diagnosis.[21,22] According to Lee et al,[16] Wildeboer et al,[18] and Akatsuka et al,[13] adding more complicated data increases the AUC, corresponding to better accuracy. Wildeboer et al[18] discovered a significant association between Gleason scores of >3+4 and accuracy of DL, but not in Gleason scores of 3+3 or 3+4. This could be due to a bias in patient selection; tumors with scores of 3+3 were disproportionately large for the doctors and were excluded from the study. According to Lee et al,[16] the AUC of ANN models was consistently higher for PSA concentrations greater than 10 ng/ml. This could be related to the serum PSA concentrations, corresponding to cancer extent and histological grade.[23] As a result, TRUS alone is insufficient for detecting PCa. However, TRUS data and its combinations with other pertinent input data can be used for ML. Despite its benefits, neural networks utilizing ultrasonic images have drawbacks that can be improved, such as the need for a large dataset for training.[24] Furthermore, the quality of scans, sample collection procedures, and human interpretation errors differ with datasets, making it impossible to create a gold standard.[24,25]

Reading ultrasound images requires several years of experience and training. ML has been introduced to medical imaging to address these constraints, speed up ultrasound picture analysis, and generate objective disease classification.[21] ML applications have advanced rapidly, thus reducing the time required to interpret a large amount of data and draw conclusions.[26] ML is an AI subfield in which computer algorithms learn connections between data instances for predictions.[22] As previously noted, ultrasound images are analyzed using various techniques such as classification, regression, registration, and segmentation. However, neural network techniques have been found to outperform other classifiers.[23] Neural networks function similarly to the human brain and can solve the limitations of regular ML. They can combine additional variables and produce outcomes for more complex scenarios.[23] A neural network can create input data from many variables to classify patients with PCa.

As shown in Table 3, the algorithms used to build ML have several advantages and disadvantages. Regardless of their differences, CNNs and ANNs are important in the ML field.[26,27] ANNs comprise multiple layers of interconnected artificial neurons activated by activation functions. Like traditional machine algorithms, the neural network learns specific values during training.[28] Other prominent ML models, such as SVM, work by adding a higher dimension to the input to differentiate the classes.[29] To assess whether the data meet the criteria, the decision tree (DT) employs several decision logics that act similarly to flowcharts. When numerous DTs are joined, a Random Forest method is used to reduce the overfitting tendency of the DT.[30]

The ML field is advancing rapidly, with corresponding hardware and software advancements. DL has advanced significantly in recent years, owing to data overflow and support from graphic processing unit hardware acceleration. Various DL libraries, including PyTorch, Keras, TensorFlow, Theano, and

Caffe, are currently available. Neural network fusion was recently developed to increase accuracy.[31] The utilization of ML with TRUS data could have a potential role as a diagnostic modality, especially when MRI is unavailable. Based on current guidelines, T2-weighted imaging remains the most useful method for local MRI.[32] However, a meta-analysis by de Rooij et al[33] showed that MRI had high specificity but poor sensitivity for local PCa staging. Its sensitivities and specificities for extracapsular extension, seminal vesicle invasion, and overall stage T3 detection were 0.57 (95% confidence interval [CI] = 0.49–0.64) and 0.91 (95% CI = 0.88–0.93), 0.58 (95% CI = 0.47–0.68) and 0.96 (95% CI = 0.95–0.97), and 0.61 (95% CI = 0.54–0.67) and 0.88 (95% CI = 0.85–0.91), respectively. Our findings showed that ML based on TRUS and other relevant data can improve diagnostic performance. Thus, it will become more affordable and easier to diagnose PCa without MRI. Furthermore, ML based on TRUS data can be implemented in combination with MRI for prostate biopsy and intraoperative mapping before robotic surgery. This will allow the surgeon to visualize suspected lesions on the instrument display during the procedure.

To date, no study has analyzed the cost-effectiveness of ML for PCa diagnosis. For severe cases of PCa, AI is used to reduce the processing time and facilitate early detection, resulting in a superior prognosis. Additionally, reducing the quantity of human labor enables the service to be provided at a reduced price compared with multiparametric MRI.[34] A systematic review by Khanna et al[35] reported that AI models demonstrated significant cost savings for medical diagnosis and treatment, and this is applicable to PCa diagnosis.

The present study had some limitations. The major limitations were the low to moderate quality of the included studies and the small sample of articles. The literature search was restricted to studies written in English, and some articles in other languages might have been missed. None of the studies used the same output parameters to generate a quantitative analysis. Additionally, most studies did not blind the diagnosis when testing the ML models, which might have resulted in bias. The approximate AUC and sensitivity values of the ML models in this study were not high and might have led to missed PCa cases among the patients. Further advancements in ML will continue to improve diagnostic accuracy.

In conclusion, the accuracy of the neural network models for PCa diagnosis using ultrasound images was relatively high, with AUCs greater than 0.7. Neural network models are promising for PCa diagnosis and can provide instant information for further workup with relatively high accuracy. Image-based ML models can help doctors decide on proceeding with or deferring a prostate biopsy. Further development of AI will be beneficial for diagnosis, treatment evaluation, and predicting patient prognosis. Future studies should investigate and compare the diagnostic performance of neural networks based on ultrasound images and MRI for PCa.

A preprint of this manuscript has previously been published (https://www.medrxiv.org/content/10.1101/2022.02.03.22270377v1).

Conflict of Interest

Agus Rizal Ardy Hariandy Hamid is the editor-in-chief of this journal but was not involved in the review or decision making process of the article.

Acknowledgment

Technical assistance and critical advice are provided by the staff of the Department of Urology, Cipto Mangunkusumo Hospital.

Funding Sources

None.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistic 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209–49.
2. Hayes JH, Barry MJ. Screening for prostate cancer with the prostate-specific antigen test: a review of current evidence. JAMA. 2014;311(11):1143–9.
3. Naji L, Randhawa H, Sohani Z, Dennis B, Lautenbach D, Kavanagh O, et al. Digital rectal examination for prostate cancer screening in primary care: a systematic review and meta-analysis. Ann Fam Med. 2018;16(2):149–54.
4. Ganie FA, Wanie MS, Ganie SA, Lone H, Gani M, Mir MF, et al. Correlation of transrectal ultrasonographic findings with histopathology in prostatic cancer. J Educ Health Promot. 2014;3:38.
5. Harvey CJ, Pilcher J, Richenberg J, Patel U, Frauscher F. Applications of transrectal ultrasound in prostate cancer. The British Journal of Radiology. 2012;85 Spec No 1(Spec Iss 1):S3–17.
6. Kretschmer A, Tilki D. Biomarkers in prostate cancer - Current clinical utility and future perspectives. Crit Rev Oncol Hematol. 2017;120:180–93.
7. Bratan F, Niaf E, Melodelima C, Chesnais AL, Souchon R, Mège-Lechevallier F, et al. Influence of imaging and histological factors on prostate cancer detection and localization on multiparametric MRI: a prospective study. Eur Radiol. 2013;23(7):2019–29.
8. Loeb S, Vellekoop A, Ahmed HU, Catto J, Emberton M, Nam R, et al. Systematic review of complications of prostate biopsy. Eur Urol. 2013;64:876–92.
9. Ukimura O, Coleman JA, de la Taille A, Emberton M, Epstein JI, Freedland SJ, et al. Contemporary role of systematic prostate biopsies: indications, techniques, and implications for patient care. Eur Urol. 2013;63(2):214–30.
10. Nitta S, Tsutsumi M, Sakka S, Endo T, Hashimoto K, Hasegawa

M, et al. Machine learning methods can more efficiently predict prostate cancer compared with prostate-specific antigen density and prostate-specific antigen velocity. Prostate Int. 2019;7(3):114–8.
11. Djavan B, Remzi M, Zlotta A, Seitz C, Snow P, Marberger M. Novel artificial neural network for early detection of prostate cancer. J Clin Oncol. 2002;20(4):921–9.
12. Ronco AL, Fernandez R. Improving ultrasonographic diagnosis of prostate cancer with neural networks. Ultrasound Med Biol. 1999;25(5):729–33.
13. Akatsuka J, Numata Y, Morikawa H, Sekine T, Kayama S, Mikami H, et al. A data-driven ultrasound approach discriminates pathological high grade prostate cancer. Sci Rep. 2022;12(860).
14. Loch T, Leuschner I, Genberg C, Weichert-Jacobsen K, Küppers F, Yfantis E, et al. Artificial neural network analysis (ANNA) of prostatic transrectal ultrasound. Prostate. 1999;39(3):198–204.
15. Lee HJ, Kim KG, Lee SE, Byun SS, Hwang SI, Jung SI, et al. Role of transrectal ultrasonography in the prediction of prostate cancer: artificial neural network analysis. J Ultrasound Med. 2006;25(7):815–21.
16. Lee HJ, Hwang SI, Han SM, Park SH, Kim SH, Cho JY, et al. Image-based clinical decision support for transrectal ultrasound in the diagnosis of prostate cancer: comparison of multiple logistic regression, artificial neural network, and support vector machine. Eur Radiol. 2010;20(6):1476–84.
17. Azizi S, Bayat S, Yan P, Tahmasebi A, Kwak JT, Xu S, et al. Deep recurrent neural networks for prostate cancer detection: analysis of temporal enhanced ultrasound. IEEE Trans Med Imaging. 2018;37(12):2695–703.
18. Wildeboer RR, Mannaerts CK, van Sloun RJG, Budäus L, Tilki D, Wijkstra H, et al. Automated multiparametric localization of prostate cancer based on B-mode, shear-wave elastography, and contrast-enhanced ultrasound radiomics. Eur Radiol. 2020;30(2):806–15.
19. Hassan R, Islam F, Uddin Z, Ghoshal G, Hassan MM, Huda S, et al. Prostate cancer classification from ultrasound and MRI images using deep learning based explainable artificial intelligence. Future Gener Comput Syst. 2022;127:462–72.
20. Lorusso V, Kabre B, Pignot G, Branger N, Pacchetti A, Thomassin-Piana J, et al. External validation of the computerized analysis of TRUS of the prostate with the ANNA/C-TRUS system: a potential role of artificial intelligence for improving prostate cancer detection. World J Urol. 2023;41(3):619–25.
21. Noorbakhsh-Sabet N, Zand R, Zhang Y, Abedi V. Artificial Intelligence Transforms the Future Healthcare. Am J Med. 2019;132(7):795–801.
22. Alaloul WS, Qureshi AH. Data processing using artificial neural networks. Dynamic data assimilation - beating the uncertainties. IntechOpen; 2020.
23. Carter HB. Differentiation of lethal and non-lethal prostate cancer: PSA and PSA isoforms and kinetics. Asian J Androl. 2012;14(3):355–60.
24. Pai RK, Van Booven DJ, Parmar M, Lokeshwar SD, Shah K, Ramasamy R, et al. A review of current advancements and limitations of artificial intelligence in genitourinary cancers. Am J Clin Exp Urol. 2020;8(5):152–62.
25. Shahid N, Rappon T, Berta W. Application of artificial neural networks in health care organizational decision-making: a scoping review. PLos One. 2019;14(2):e0212356.
26. Brattain LJ, Telfer BA, Dhyani M, Grajo JR, Samir AE. Machine learning for medical ultrasound: status, method, and future opportunities. Abdom Radiol (NY). 2018;43(4):786–99.
27. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data. 2021;8(1):53.
28. Li B, He Y. An attention mechanism oriented hybrid CNN-RNN deep learning architecture of container terminal liner handling conditions prediction. Comput Intell Neurosci. 2021;2021: 3846078.
29. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak. 2019;281(19):281.
30. Juarez-Orozco LE, Martinez-Manzanera O, Nesterov SV, Kajander S, Knuut J. The machine learning horizon in cardiac hybrid imaging. European J Hybrid Imaging. 2018;15(2):1–15.
31. Dhawale CA, Dhawale K. Current trends in deep learning frameworks with opportunities and future prospectus. Adv Electr Comput Eng. 2020;63–77.
32. Mottet N, van den Bergh RCN, Briers E, Van den Broeck T, Cumberbatch MG, De Santis M, 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.
33. de Rooij M, Hamoen EH, Witjes JA, Barentsz JO, Rovers MM. Accuracy of magnetic resonance imaging for local staging of prostate cancer: a diagnostic meta-analysis. Eur Urol. 2016;70(2):233–45.
34. Rabaan AA, Bakhrebah M A, AlSaihati H, Alhumaid S, Alsubki RA, Turkistani SA, et al. Artificial intelligence for clinical diagnosis and treatment of prostate cancer. Cancers. 2022;14(22):5595.
35. Khanna NN, Maindarkar MA, Viswanathan V, Fernandes JFE, Paul S, Bhagawati M, et al. Economics of artificial intelligence in healthcare: diagnosis vs. treatment. Healthcare (Basel). 2022;10(12):2493.