Scientific Research Publishing

Journal of Intelligent Learning Systems and Applications, 2026, 18(2), 91-104
https://www.scirp.org/journal/jilsa
ISSN Online: 2150-8410
ISSN Print: 2150-8402

Explainable AI: Enhancing Clinician Trust and Accuracy in Prostate Cancer Diagnosis

Abdelrahman Elsharif Karrar

College of Computer Science and Engineering, Taibah University, Madinah, Saudi Arabia
Email: akarrar@taibahu.edu.sa

How to cite this paper: Karrar, A.E. (2026) Explainable AI: Enhancing Clinician Trust and Accuracy in Prostate Cancer Diagnosis. *Journal of Intelligent Learning Systems and Applications*, **18**, 91-104.
https://doi.org/10.4236/jilsa.2026.182006

Received: February 12, 2026
Accepted: March 9, 2026
Published: March 12, 2026

Abstract

Prostate cancer remains one of the most prevalent malignancies among men worldwide and achieving an accurate and timely diagnosis is essential for guiding appropriate treatment decisions and improving patient outcomes. In recent years, Artificial Intelligence (AI), particularly Machine Learning (ML) and Deep Learning (DL) techniques, has been increasingly applied to clinical, imaging, and histopathological data to enhance diagnostic accuracy. These approaches have shown strong potential in detecting complex patterns within medical datasets that may not be easily captured using traditional statistical methods. Nevertheless, despite the growing number of AI-based diagnostic models proposed in the literature, their adoption in routine clinical practice remains limited. This is largely due to concerns related to transparency, interpretability, reliability, and the extent to which clinicians can understand and trust automated decision-making systems. To address these limitations, Explainable Artificial Intelligence (XAI) has emerged as a promising direction aimed at improving the interpretability of predictive models without compromising performance. In this study, an XAI-based framework for prostate cancer diagnosis is developed by integrating ensemble learning with post-hoc explanation techniques. Multiple classifiers, including decision trees, support vector machines, and artificial neural networks, were combined within an ensemble strategy to leverage their complementary strengths and enhance overall predictive capability. To ensure model transparency, explanation methods such as SHAP (Shapley Additive Explanations) and feature importance analysis were employed to quantify and illustrate the contribution of each variable to the final prediction. The experimental results demonstrate that the proposed ensemble XAI framework achieves superior diagnostic accuracy compared with individual classifiers while also providing clear and clinically meaningful explanations of its decisions. These findings highlight the value of incorporating explainability into machine learning models to strengthen clinician trust, support responsible AI deployment, and enable more informed and reliable decision-making in prostate cancer management.

Keywords

Explainable Artificial Intelligence (XAI), Machine Learning (ML), Deep Learning (DL), Shapley Additive Explanations (SHAP), Prostate Cancer

1. Introduction

Prostate cancer continues to be one of the leading causes of morbidity and mortality related to cancer in men around the world. Accurate and timely detection facilitates the highest possible survival rate and optimized treatment for prostate cancer. The traditional methods of diagnosing prostate cancer involve the assessment of various biological markers and performing physical examinations of the prostate, including tests for PSA, imaging techniques, and biopsies [1] [2]. The current paper uses a simplified public dataset for the diagnosis of prostate cancer, which does not include the results from PSA tests or biopsies; instead, only the physical characteristics of the tumor and some factors associated with recurrence, as shown in **Table 1**, were made available. The development of an objective, automated diagnostic system for prostate cancer detection using data-driven artificial intelligence provides significant advantages over the limitations of current diagnostic methods. The ability to detect and analyze extensive amounts of imaging and clinical information through the use of Machine Learning (ML) algorithms can be a valuable source of information when looking for patterns in the data that may not be obvious to a trained clinical practitioner. However, because the majority of existing AI models are considered "black box" models, there is a significant barrier for clinical practitioners to feel confident in the AI recommendations provided by these types of models. Clinical practitioners need to be able to confidently grasp the rationale for the decisions that an AI makes in order to trust the AI in making treatment-related decisions, especially in high-urgency situations.

The primary goal of the Explainable AI (XAI) program is to empower users with a deeper understanding of AI's predictive capabilities, enabling them to grasp not only what the AI can predict, but also the underlying rationale and mechanisms that drive these predictions. Specifically, the XAI program aims to provide users with transparent and actionable insights into the AI's decision-making processes, shedding light on how the AI arrived at its predictions, and what specific factors or attributes contributed to its conclusions. By doing so, the XAI program seeks to foster a more nuanced and informed understanding of AI-driven predictions, allowing users to effectively evaluate and interpret the outputs of AI systems [3].

In the healthcare sector, the implementation of XAI solutions is particularly vital, as it has the potential to address pressing concerns related to trust, ethics, and regulatory compliance. Clinicians, who are increasingly relying on AI-powered tools to inform their decisions, must be able to trust the accuracy and reliability of these systems. XAI can help alleviate concerns about the opacity of AI decision-making, providing clinicians with the necessary transparency and insights to confidently integrate AI-driven recommendations into their practice. Furthermore, XAI can facilitate compliance with regulatory requirements, such as those related to explainability and accountability, thereby facilitating the adoption of AI solutions in healthcare settings. By providing a clear understanding of AI-driven predictions and decisions, XAI can ultimately contribute to improved patient outcomes, enhanced patient safety, and more effective healthcare delivery [4].

Using a combination of approaches, an ensemble-based XAI framework for prostate cancer diagnosis is being developed to balance predictive accuracy and interpretability, thereby enhancing clinicians' confidence in their diagnoses. The objectives of this research include:

- Creating an ensemble-based XAI solution for prostate cancer diagnosis.
- Implementing a series of explainability techniques to provide users with a clear understanding of the predictions generated by the AI.
- Examining the effects of using an XAI framework to improve diagnostic accuracy and transparency.

The paper is organized as follows: Section 2 provides an overview of the related work, setting the context for the research. Section 3 delves into the methodology, explaining the approach and techniques employed in the study. The results are presented in Section 4, highlighting the key findings. These results are then discussed in Section 5, where the implications and significance of the outcomes are analyzed. Finally, Section 6 concludes the paper, summarizing the main contributions and takeaways.

2. Related Work

Studies continue to show that more clinician researchers are very interested in applying AI to assist with diagnosing, prognosing, and developing treatment plans for prostate cancer. The initial studies conducted by clinician-researchers primarily focused on applying standard AI (ML) techniques to structured clinical data, such as PSA, Gleason score, tumor stage, and demographics [5] [6].

The AI models discussed above generated significant interest in developing predictive models for cancer recurrence [7], tumor aggressiveness, and patient survival, and demonstrated superior predictive performance compared to common statistical techniques. However, limited sample sizes, a large number of features, and/or large class imbalances in clinical datasets significantly reduced the performance of these AI predictive models.

Ensemble learning techniques (that is, bagging, boosting, and random subspace) are increasingly popular ways of addressing some of the limitations in prostate cancer research. Each of these ensemble learning techniques can combine multiple base classifiers to improve classification generalization, reduce the variance of predicted results, and decrease model bias [8] [9].

Prior studies indicate that ensemble classifiers outperform their single-classifier counterparts in terms of improved classification accuracy and lower classification error rates [10], particularly when the ensemble is trained on prostate cancer datasets with significant imbalance. Boosting methods improve the sensitivity for identifying patients with high-risk or recurrent cancer, enabling early treatment or intervention planning.

A potential area of future investigation to address a deficit in the literature concerns the use of distinct model-building techniques, including dimensionality reduction and feature selection, as well as model interpretability. When using these methods correctly, clinicians will achieve improved model performance and reduce their workload by minimizing computational demands and extraneous, irrelevant, or noisy model inputs [11]. For example, the model's performance after feature selection on prostate cancer data can be demonstrated by characteristics such as malignant tissue scores, tumor size, lymph nodes involved, and other regional disease foci, which also provide insight into the comparability of machine learning model performance with clinical benchmarks.

Many AI studies of prostate cancer have made substantial progress in predictive performance but have typically produced complex model architectures that operate as "black boxes", severely limiting understanding of how the predictions were generated.

Due to the lack of interpretability, AI-assisted decisions may not be accepted for use in clinics, as caregivers must have a clear understanding of, believe in, and be able to rationalize the findings of such decisions [12]. This has guided healthcare researchers to focus their efforts on the development of "Explainable Artificial Intelligence" (XAI). In most cases, current XAI techniques for providing explanations or interpretation of a model's predictions are primarily based on post-hoc methods (such as SHAP, LIME, saliency maps, and feature attribution). These techniques allow users to gain insight into how their models arrived at a given prediction or decision.

Research in different areas of medicine utilizes XAI. For example, researchers in radiology, dermatology, and diabetes related retinal diseases have extensive evidence indicating that effective use of XAI leads to an improvement in clinician confidence, facilitates the ability of clinicians to have shared decision-making with patients, and complies with ethical and regulatory standards [13] [14].

However, few studies have examined the use of XAI in prostate cancer AI systems. Furthermore, among the available studies, there are even fewer that have examined the integrating XAI with ensemble learning methods. Most studies conducted to date regarding prostate cancer have had a focus on the accuracy of the model, and they have seen interpretability as being an afterthought.

This research, therefore, fills an important gap in terms of creating prostate cancer AI models that enable high levels of diagnostic accuracy while providing clinicians with interpretability and thus trust regarding the prostate cancer diagnosis. By examining how to build a prostate cancer AI model using an ensemble learning method and combining it with XAI methods, A more accurate and reliable system is being developed to enhance clinicians' diagnostic accuracy and trust during the diagnostic process.

3. Methodology

3.1. Dataset

The UCI Prostate Cancer Diagnostic Dataset was obtained for this study and used in the context of machine learning applications in oncology. The UCI Prostate Cancer Diagnostic Dataset is an open-source public database that contains 100 patient records of patients who have been diagnosed with prostate cancer.

The dataset used in this study is presented in **Table 1** and only includes data regarding six variables: age, tumor size, node involvement, metastasis, degree of malignancy, and class (recurrence output). The six listed attributes comprise all data features contained in the dataset.

The dataset did not contain any information concerning PSA levels, Gleason Scores, or biopsy variables [15]. Therefore, no other features not previously identified were introduced in the preprocessing of the dataset, allowing for maximum consistency between the dataset description (**Table 1**) and the modelling processes employed in this study. The prostate cancer dataset used in this research is a small dataset from a publicly accessible database of 100 patients diagnosed with prostate cancer and six structured features. Due to the limited number of patients included in this sample, this study is presented as a methodological and proof-of-concept study and not as a definitive clinical prediction system.

The variables specified in **Table 1** represent both the complete and only the dataset factors that were utilized for all experimental purposes.

Table 1. Dataset attributes.

Attribute	Description	Type
Age	Patient age	Numeric
Tumor Size	Prostate tumor measurement	Numeric
Node	Lymph node involvement	Nominal
Metastasis	Presence of metastasis	Nominal
Degree of Malignancy	Cancer stage	Numeric
Class	Recurrence outcome	Nominal

3.2. Data Preprocessing

The researcher developed machine learning algorithms using advanced model development techniques in 4 stages. The first stage involved mean imputation for all numeric missing values. The second phase involved converting all categorical data to a computer-interpretable format for computer interpretation. The third stage was to normalize the numeric values before training the machine learning algorithms, thereby eliminating scale bias during training. Finally, in stage four, the researcher applied re-sampling techniques to balance the training classes of the model.

As a result of the four stages, the researcher generated models with reduced bias to predict the likelihood of future instances of previously analyzed cases.

3.3. Model Architecture

The ensemble learning framework comprises three distinct types of classifiers: Decision Tree (J48), Support Vector Machine (SVM), and Artificial Neural Network (ANN). The reason for employing three distinct types of classifiers is to enable each to identify unique patterns in the input dataset and, in turn, provide different advantages to the assembled model. All classifiers were built using a boosting algorithm, which allowed the ensemble to focus on instances of false positives generated by the individual classifiers and build more robust, accurate models. Once completed, all models were evaluated using cross-validation methods to eliminate bias from training on a single test dataset. Figure 1 illustrates the stages of the proposed diagnostic framework.

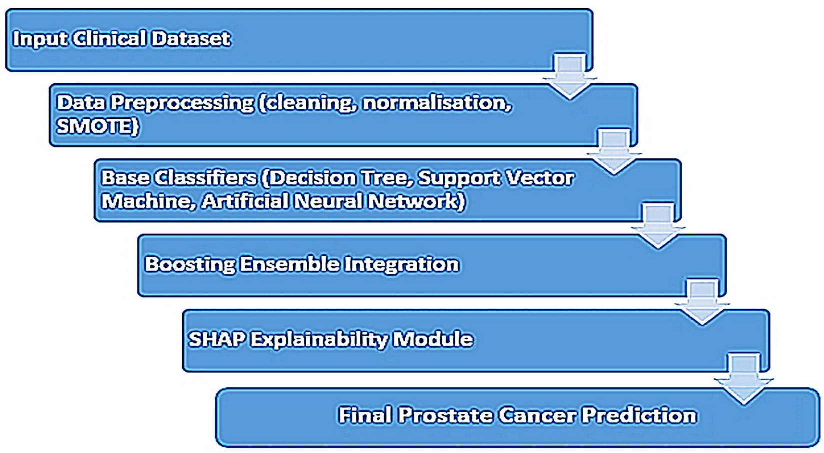

Figure 1. Proposed diagnostic framework.

3.4. Explainability Techniques

SHAP, which stands for Shapley Additive Explanations, through the stages depicted in Figure 2, was used as a post-hoc approach to improve model transparency and the interpretability of clinicians' predictions. Each SHAP value assigns each feature an important score relative to a patient's SHAP value, or the influence the feature has on a particular prediction, based on an average score [16]. The higher the SHAP value, the more important that feature was in determining the outcome of that particular prediction [17] [18]. By ranking features based on their total SHAP values, the model identifies clinically relevant or usual predictive features, enabling more intuitive explanations that align with clinically proven medical reasoning.

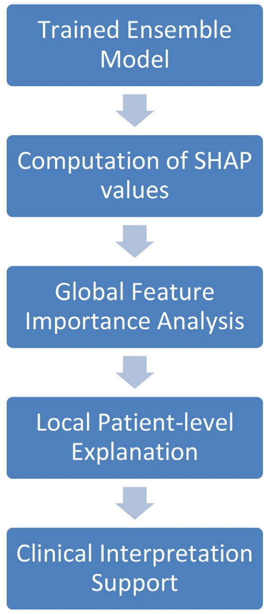

Figure 2. SHAP explainability workflow.

4. Results

4.1. Model Performance

This study aims to compare the new XAI ensemble classifier with its individual base classifiers to assess the overall classifier's performance using four classification metrics (accuracy, precision, recall, and F1), which were calculated using Equations (1) - (4) [19]. All four of these metrics are required for the development of an efficient medical diagnosis/decision-support system for prostate cancer. In addition, this analysis shows that the XAI ensemble classifier outperformed all four-classification metrics compared to the individual base classifiers, including the XAI ensemble classifier, demonstrating the great potential of XAI ensemble classifiers for identifying a broader range of complex heterogeneous prostate cancer data patterns.

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{All Instance}} \quad (1)$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (2)$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (3)$$

$$\text{F1 Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

All three models, namely decision trees, Support Vector Machines (SVMs), and Artificial Neural Networks (ANNs), demonstrated reasonable performance as single-classifier models, yielding acceptable predictive results. However, when examined individually, these models were unable to reveal any inconsistencies between the strengths of their respective components and the overall predictive performance, making it challenging to discern the intricacies of their decision-making processes. The decision tree model, exemplified by algorithms such as J48, generated interpretable prediction rules that provided valuable insights into the challenges associated with generalizing prediction accuracy across multiple folds of the validation data. In contrast, SVM produced effective separations between classes, but its performance was heavily reliant on the selection of optimal parameters to achieve peak predictive performance. The ANN, utilizing a nonlinear approach, yielded satisfactory results, albeit at the cost of requiring extensive tuning to attain optimal prediction performance. A notable limitation of the ANN, however, is its lack of transparency, as it does not provide insight into the processes underlying its predictions. In contrast, the ensemble model capitalizes on the strengths of individual classifiers, effectively combining their predictive capabilities to produce a model with enhanced robustness and generalization ability across validation folds.

The stability of the ensemble model has been further reinforced by the findings from cross-validation, which not only demonstrated improved performance metrics but also revealed diminished variance across folds. This suggests that the model's predictions are less dependent on specific subcomponents of the dataset, thereby enhancing its generalizability and robustness. Notably, the ensemble model exhibited a significant improvement in recall for recurrent cases, a finding of considerable clinical importance as it implies that the model is more effective in identifying patients at high risk, thereby reducing the likelihood of delayed intervention. The increased F1 score is also noteworthy, as it reflects a proper balance between precision and recall, rendering the model an apt choice for supporting patient diagnosis in a clinical setting. As illustrated in **Table 2** and **Figure 3**, the ensemble model's performance metrics underscore its potential as a reliable tool for clinicians.

4.2. Explainability Analysis

Explainability analysis was instrumental in evaluating the behavior of the models, providing crucial insights into their decision-making processes [20]. While performance predictions were crucial in assessing model quality, understanding the underlying mechanisms driving their conclusions was equally important. To

Table 2. Model performance comparison.

Model	Accuracy	Precision	Recall	F1-score
Decision Tree (J48)	0.86	0.83	0.81	0.82
SVM	0.88	0.85	0.84	0.84
ANN	0.89	0.87	0.85	0.86
Ensemble XAI Model	**0.92**	**0.90**	**0.89**	**0.89**

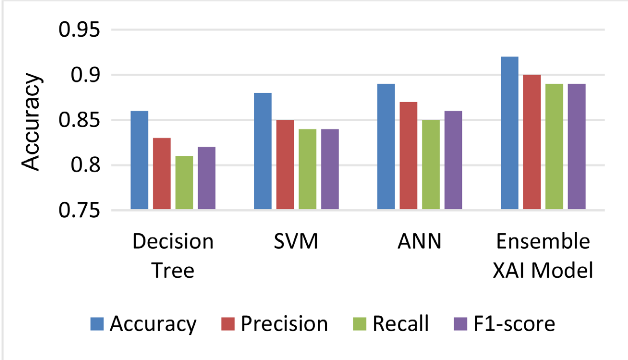

Figure 3. Comparison of algorithms across evaluation metrics.

achieve this, SHAP-based approaches, specifically Shapley values-based methods, were employed to analyze feature importance. This enabled the identification of the most influential factors contributing to the models' predictions. Notably, analysis revealed that the degree of malignancy was the strongest predictor for patients, surpassing tumor size and metastatic status. Additionally, age and lymph node involvement were found to have a smaller, yet still significant, impact on the models' decisions. These findings corroborate established clinical knowledge regarding the interplay between these variables, lending credibility to the models and underscoring the value of explainability analysis in understanding complex predictive models.

To help clinicians interpret the results of this study, a simple bar chart was created showing feature importance rankings for each clinical variable contributing to diagnostic predictions at a broader level. The use of this bar chart provided a quick overview of the relative importance of the various clinical variables for diagnostic predictions. In addition to the feature importance rankings presented in the bar chart, patient-level explanation outputs allow evaluation of each prediction to determine which specific feature values contributed to either an increase or decrease in recurrence risk. Overall, the findings of this research provide evidence that the proposed "explainable ensemble" model produces accurate diagnostic predictions and provides insight into the important variables used to develop those predictions, thereby offering an opportunity to serve as a reliable clinical decision support tool.

5. Discussion

The results of the current research indicate that applying ensemble learning with explainable artificial intelligence offers the opportunity to improve both the accuracy of diagnosing prostate cancer and the level of trust clinicians have in AI-based decision support systems. The ensemble model's enhanced accuracy provides additional support for previous findings on the positive effect of combining multiple classifiers to mitigate challenges in medical datasets, such as class imbalance, feature variability, and small sample sizes [21]-[23].

The incorporation of diverse algorithms, including decision trees, support vector machines, and artificial neural networks, enabled the creation of ensemble models that exhibited significantly enhanced predictive stability and reliability compared to any single model type in isolation. This amalgamation of different algorithmic capabilities yielded models that were more robust and dependable in their predictive performance. The overall stability of an ensemble model is a crucial factor to consider when developing and implementing AI-based clinical decision support systems, as it has a direct impact on the trust and confidence of clinicians in the system's outputs. If an AI-based clinical decision support system generates inconsistent predictive outputs, such as predictions made at different time points (e.g., one to three weeks apart), clinicians may become uncertain and lose faith in the system's ability to provide reliable guidance, ultimately undermining its utility in clinical practice.

In addition to focusing on predictive performance, this study places significant emphasis on explainability as a deliberate design principle, rather than treating it as an afterthought. This is a notable advancement, as previous research has highlighted that a lack of trust among clinicians is a substantial obstacle to the adoption of AI in healthcare settings. Typically, AI-based predictive models are developed and utilized in a black-box manner, depriving clinicians of insight into the decision-making process behind specific predictions. By integrating SHAP-based explainability, the authors provide clinicians with transparent information elucidating how and why the model arrived at a particular prediction, thereby addressing a major concern related to clinician trust. The system not only furnishes clinicians with a binary outcome but also equips them with evidence to substantiate their clinical reasoning and accountability, ultimately facilitating more informed decision-making.

The analysis of model explainability shows that the most influential features in the model align closely with established medical knowledge. The greatest contributions to the model's ability to predict recurrence were observed in variables such as degree of malignancy, tumor size, and metastatic status; this finding aligns well with current knowledge of these variables' roles in assessing progression and prognosis of prostate cancer. Therefore, there must be agreement between how the model operates and clinicians' expectations for its functioning, so that they can be confident that the AI system is not finding random correlations [22]. Additionally, the primary and secondary features identified, such as age and lymph node involvement, are similar to the types of information a physician considers when making clinical decisions.

One of the primary benefits of the proposed approach is its ability to offer clinicians a dual-layered understanding of predictive outcomes, providing explanations at both the global and individual levels. This multifaceted capability enables clinicians to gain a more comprehensive appreciation of the factors driving diagnostic decisions. At the global level, feature importance rankings furnish clinicians with valuable insights into the most influential factors contributing to a particular diagnosis for a specific group of patients. This information can help identify trends and patterns, informing treatment strategies and policy decisions. Conversely, patient-level explanations empower clinicians to drill down into the specifics of individual cases, allowing them to precisely determine how actual feature values contributed to the predicted outcome for a particular patient. This granular understanding can facilitate personalized treatment planning, enhance patient engagement, and foster more effective clinician-patient communication. By bridging the gap between population-level insights and individual patient needs, this approach has the potential to revolutionize the way clinicians interact with and act upon predictive models [24]. This dual perspective enables clinicians to assess the appropriateness of using AI recommendations in specific cases and to communicate the reasons for their decisions to patients.

Despite the numerous strengths of this study, as outlined above, it is essential to acknowledge certain limitations that may impact the generalizability and interpretability of the findings. A primary concern is that all evidence presented in this study is derived from a single, structured data set, which may not be representative of broader or more diverse populations. As a result, the applicability of the results to different demographic or clinical contexts may be limited, underscoring the need for further research using more varied and extensive data sets. Furthermore, while the post-hoc explainability techniques employed in this study, such as SHAP, provide valuable insights into the nature of feature relationships and their contributions to the model's predictions, they are not without their limitations. Specifically, these techniques are primarily designed to elucidate the relationships between features and predictions, rather than identifying the underlying causal mechanisms driving the observed outcomes. Consequently, they do not provide definitive answers regarding the actual causes of the outcomes, highlighting the need for additional research aimed at disentangling the complex interplay of factors influencing patient outcomes [25]. Additionally, an explanation may oversimplify the true complexity of the interactions within a given model.

Future research needs to integrate diverse and multimodal data sources, including medical imaging, histopathology, and genomic testing metrics, to create a comprehensive representation of the disease process's characteristics.

6. Conclusions

This investigation developed an innovative explanatory Artificial Intelligence (AI) tool to enhance healthcare professionals' trust in their diagnoses of prostate cancer, using ensemble techniques and SHAP-based explanations. Findings demonstrate not only predictive ability but also model transparency and interpretability. Using multiple classifiers produced more robust, generalizable outcomes than using a single classifier alone. A further contribution of this study was the identification of clinically useful explanations generated by top-performing AI technologies, demonstrating a high degree of correlation between the features identified in the explanation evaluation and those used throughout the professional evaluation process for prostate cancer. Demonstrating the validity and double-checking of AI insights will increase clinicians' confidence in AI-generated insights, enhance ethical accountability in healthcare decision-making, and improve medical decision-making by increasing providers' perception of accountability.

The current study has yielded valuable insights, but it is essential to acknowledge that further investigation is required to validate and expand upon these findings. Specifically, additional research utilizing larger, more diverse, and more varied datasets is necessary to substantiate the claims made in this study. The results of this research indicate that explainable ensemble models have the potential to play a significant role in providing clinical decision support, and as such, they warrant further exploration. Future studies should focus on evaluating the deployment of these models in real-world clinical practice, examining how they can be seamlessly integrated into existing workflows, and investigating how clinicians interact with XAI systems. By doing so, researchers can gain a deeper understanding of the practical applications and implications of explainable AI in healthcare. This research contributes to the growing body of evidence demonstrating that the use of explainable AI can have numerous advantages in improving the safety and efficacy of prostate cancer diagnosis, and it highlights the need for continued investigation in this area.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Khan, T., Altamimi, M.A., Hussain, A., Ramzan, M., Ashique, S., Alhuzani, M.R., *et al.* (2022) Understanding of PSA Biology, Factors Affecting PSA Detection, Challenges, Various Biomarkers, Methods, and Future Perspective of Prostate Cancer Detection and Diagnosis. *Advances in Cancer Biology—Metastasis*, **5**, Article ID: 100059. https://doi.org/10.1016/j.adcanc.2022.100059

[2] Kania, E., Janica, M., Nesterowicz, M., Modzelewski, W., Cybulski, M. and Janica, J. (2025) Advances and Challenges in Prostate Cancer Diagnosis: A Comprehensive Review. *Cancers*, **17**, Article 2137. https://doi.org/10.3390/cancers17132137

[3] Knapič, S., Malhi, A., Saluja, R. and Främling, K. (2021) Explainable Artificial Intelligence for Human Decision Support System in the Medical Domain. *Machine Learning and Knowledge Extraction*, **3**, 740-770. https://doi.org/10.3390/make3030037

[4] Niu, S., Yin, Q., Ma, J., Song, Y., Xu, Y., Bai, L., *et al.* (2024) Enhancing Healthcare Decision Support through Explainable AI Models for Risk Prediction. *Decision Support Systems*, **181**, Article ID: 114228. https://doi.org/10.1016/j.dss.2024.114228

[5] Dhivya, V., Yadalam, P.K., Sri Shivasankari, T., Ayyachamy, S. and Anegundi, R.V. (2025) PSA-Based Machine Learning Model in the Detection of Prostate Cancer. In: Agarwal, S., Lakshmi, D. and Singh, L., Eds., *Navigating Innovations and Challenges in Travel Medicine and Digital Health*, IGI Global, 219-236. https://doi.org/10.4018/979-8-3693-8774-0.ch011

[6] Bhise, S., Bepari, S., Gadekar, S., Kale, D., Singh Gaur, A. and Aswale, S. (2021) Breast Cancer Detection Using Machine Learning Techniques. *International Journal of Engineering Research and Technology*, **10**, 98-103. https://www.researchgate.net/publication/353285629

[7] Karrar, A.E. and Elarabi, R. (2023) An Intelligent Machine Learning-Based System for Predicting Heart Disease Using Mixed Feature Creation Technique. In: Saeed, F., Mohammed, F., Mohammed, E., Al-Hadhrami, T. and Al-Sarem, M., Eds., *Advances on Intelligent Computing and Data Science*, Springer, 353-367. https://doi.org/10.1007/978-3-031-36258-3_31

[8] Mienye, I.D. and Sun, Y. (2022) A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. *IEEE Access*, **10**, 99129-99149. https://doi.org/10.1109/access.2022.3207287

[9] Karrar, A.E. (2022) Investigate the Ensemble Model by Intelligence Analysis to Improve the Accuracy of the Classification Data in the Diagnostic and Treatment Interventions for Prostate Cancer. *International Journal of Advanced Computer Science and Applications*, **13**, 181-188. https://doi.org/10.14569/ijacsa.2022.0130122

[10] Umair, M., Majeed, F., Shoaib, M., Qaiser Saleem, M., S. Adrees, M., Elsharif Karrar, A., *et al.* (2022) Main Path Analysis to Filter Unbiased Literature. *Intelligent Automation & Soft Computing*, **32**, 1179-1194. https://doi.org/10.32604/iasc.2022.018952

[11] Shukla, A.K., Pippal, S.K., Gupta, S., Ramachandra Reddy, B. and Tripathi, D. (2020) Knowledge Discovery in Medical and Biological Datasets by Integration of Relief-F and Correlation Feature Selection Techniques. *Journal of Intelligent & Fuzzy Systems*, **38**, 6637-6648. https://doi.org/10.3233/jifs-179743

[12] Agrawal, S. and Vagha, S. (2024) A Comprehensive Review of Artificial Intelligence in Prostate Cancer Care: State-of-the-Art Diagnostic Tools and Future Outlook. *Cureus*, **16**, e66225. https://doi.org/10.7759/cureus.66225

[13] Shoughi, A., Dowlatshahi, M.B. and Amiri, A. (2025) New Challenges and Opportunities to Explainable Artificial Intelligence (XAI) in Smart Healthcare. In: Saba, T., Azar, A.T. and Kadry, S., Eds., *Explainable AI in Healthcare Imaging for Medical Diagnoses*, Elsevier, 503-527. https://doi.org/10.1016/b978-0-443-23979-3.00020-8

[14] Adeniran, A.A., Onebunne, A.P. and William, P. (2024) Explainable AI (XAI) in Healthcare: Enhancing Trust and Transparency in Critical Decision-Making. *World Journal of Advanced Research and Reviews*, **23**, 2447-2658. https://doi.org/10.30574/wjarr.2024.23.3.2936

[15] Lichman, M. (2025) Prognostic Dataset for Prostate Cancer. UCI Machine Learning Repository, University of California. http://archive.ics.uci.edu/ml

[16] Atmakuru, A. (2025) Enhancing the Performance and Transparency of Machine Learning (ML) Using MRI-Derived Data: Alternative Approaches to ML Interpretability-Explainability. Ph.D. Thesis, University of Reading.

[17] S Band, S., Yarahmadi, A., Hsu, C., Biyari, M., Sookhak, M., Ameri, R., *et al.* (2023) Application of Explainable Artificial Intelligence in Medical Health: A Systematic Review of Interpretability Methods. *Informatics in Medicine Unlocked*, **40**, Article ID: 101286. https://doi.org/10.1016/j.imu.2023.101286

[18] Gerlings, J., Jensen, M.S. and Shollo, A. (2021) Explainable AI, but Explainable to Whom? An Exploratory Case Study of Xai in Healthcare. In: Lim, C.P., Chen, Y.W., Vaidya, A., Mahorkar, C. and Jain, L.C., Eds., *Handbook of Artificial Intelligence in Healthcare*, Springer, 169-198. https://doi.org/10.1007/978-3-030-83620-7_7

[19] Masum, A.K.M., Khan, M.F.I., Hassan, M.M., Bitto, A.K., Farid, D.M. and Rahman, T. (2025) Enhancing Dengue Fever Diagnosis: A Machine Learning Framework with Stacking Ensemble and SHAP Explainability. 2025 *International Conference on Quantum Photonics, Artificial Intelligence, and Networking* (*QPAIN*), Rangpur, 31 July-2 August 2025, 1-6. https://doi.org/10.1109/qpain66474.2025.11171651

[20] Radja, H., Djouadi, Y. and Tabia, K. (2025) Model-Agnostic Explainability for Multi-

Label Classification via Formal Concept Analysis. *Informatica*, **49**, 307-336. https://doi.org/10.31449/inf.v49i19.9947

[21] Salmi, M., Atif, D., Oliva, D., Abraham, A. and Ventura, S. (2024) Handling Imbalanced Medical Datasets: Review of a Decade of Research. *Artificial Intelligence Review*, **57**, Article No. 273. https://doi.org/10.1007/s10462-024-10884-2

[22] Tuan, D.A. (2024) Bridging the Gap Between Black Box AI and Clinical Practice: Advancing Explainable AI for Trust, Ethics, and Personalized Healthcare Diagnostics. https://doi.org/10.20944/preprints202409.1974.v2

[23] Eltayeb, R.Y., Karrar, A.E., Osman, W.I. and Ali, M.M. (2023) Handling Imbalanced Data through Re-Sampling: Systematic Review. *Indonesian Journal of Electrical Engineering and Informatics* (*IJEEI*), **11**, 503-514. https://doi.org/10.52549/.v11i2.4471

[24] Naheed, N., Shaheen, M., Ali Khan, S., Alawairdhi, M. and Attique Khan, M. (2020) Importance of Features Selection, Attributes Selection, Challenges and Future Directions for Medical Imaging Data: A Review. *Computer Modeling in Engineering & Sciences*, **125**, 315-344. https://doi.org/10.32604/cmes.2020.011380

[25] Bhati, D., Amiruzzaman, M., Zhao, Y., Guercio, A. and Le, T. (2025) A Survey of Post-Hoc XAI Methods from a Visualization Perspective: Challenges and Opportunities. *IEEE Access*, **13**, 120785-120806. https://doi.org/10.1109/access.2025.3581136