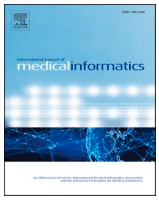
International Journal of Medical Informatics 206 (2026) 106144

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf



A hybrid decision support system using rule-based and AI methods: the OnCATs knowledge-based framework

Nuno Soares Domingues

Instituto Politécnico de Lisboa/Instituto Superior de Engenharia de Lisboa, Rua Conselheiro Emidio Navarro, 1, 1959-007 Lisbon, Portugal

A B S T R A C T

Background and significance: Clinical decision support systems (CDSS) can improve evidence-based oncology care, but many rely on opaque AI models that limit transparency and reproducibility. Rule-based approaches provide interpretability but often lack adaptability, a critical issue in prostate cancer where decisions depend on tumor stage, PSA, Gleason score, comorbidities, and life expectancy. Bridging explainability and adaptability is essential for trustworthy decision support.
Objective: To develop and evaluate OnCATs, a modular, explainable, hybrid-ready CDSS that encodes prostate cancer management guidelines in a machine-readable and auditable format.
Materials and methods: Evidence from 23 international guidelines was formalized into a JSON-based rule base executed through a forward-chaining inference engine. OnCATs supports three decision layers: (1) risk stratification, (2) treatment-pathway recommendation, and (3) prescription-level assistance for radiotherapy, brachytherapy, androgen deprivation therapy, and surgery. Feasibility was tested using ten published case reports. Performance was assessed with precision, recall, F1 scores, and descriptive concordance.
Results: OnCATs achieved perfect concordance for risk stratification (precision = 1.00, recall = 1.00, F1 = 1.00). Treatment-pathway concordance was 0.80 (F1 = 0.80). Prescription-level agreement ranged from 0.67 to 0.75 (mean F1 = 0.71). Divergences primarily reflected simplified life-expectancy modeling and incomplete case data.
Discussion: OnCATs demonstrates that transparent, rule-based reasoning can reproduce guideline-defined prostate cancer decisions with traceability. Limitations include the small sample size and reliance on secondary data.
Conclusion: OnCATs operationalizes multi-source guidelines into an explainable, modular CDSS, providing a reproducible foundation for future integration of probabilistic and machine-learning methods.

1. Introduction

Clinical decision support systems (CDSS) have become integral to modern healthcare, offering tools that enhance clinical decision-making, improve adherence to evidence-based guidelines, and support personalized treatment strategies [1]. In oncology, the complexity of care pathways, the rapid evolution of treatment guidelines, and the need to balance multiple clinical parameters make CDSS particularly valuable [2].

Prostate cancer is a compelling use case for CDSS. Management requires integrating tumor stage, prostate-specific antigen (PSA) levels, Gleason score, comorbidities, and life expectancy, while also weighing diverse treatment modalities such as surgery, radiotherapy, active surveillance, and androgen deprivation therapy [3]. This multifactorial landscape creates opportunities for informatics-driven tools to support consistent and transparent decision-making.

Despite their promise, oncology CDSS face persistent challenges. Many machine learning–based systems function as black boxes that may achieve strong predictive performance but lack transparency and reproducibility, limiting clinical trust and adoption [4]. Conversely, rule-based approaches ensure interpretability but struggle with incomplete or uncertain data and often lack rigorous validation across real-world cases [5]. Addressing these limitations requires hybrid frameworks that combine the explainability of rules with the adaptability of probabilistic and data-driven methods.

This study introduces OnCATs, an explainable rule-based CDSS for localized prostate cancer. The system encodes evidence from 23 guideline sources into a modular forward-chaining rule base that supports risk stratification, treatment-pathway recommendation, and prescription specification. Unlike existing tools that focus narrowly on either diagnostic classification or treatment selection, OnCATs integrates multiple decision layers within a unified, transparent framework.

A preliminary evaluation using ten published case reports demonstrates both the strengths of explainable, rule-based reasoning and the limitations of simplified life-expectancy modeling. The contributions of this work are threefold: (i) development of a fully explainable architecture that maps clinical evidence into traceable recommendations, (ii) integration of multi-stage guideline rules into a modular system

E-mail addresses: nuno.domingues@isel.pt, nndomingues@gmail.com.

https://doi.org/10.1016/j.ijmedinf.2025.106144
Received 13 September 2025; Received in revised form 7 October 2025; Accepted 9 October 2025
Available online 14 October 2025

adaptable to probabilistic and machine-learning extensions, and (iii) a case-based feasibility evaluation that highlights both potential and areas for refinement. By combining transparency with modularity, OnCATs contributes to ongoing efforts to design CDSS that are not only accurate but also interpretable, reproducible, and clinically trustworthy. This revision expands the literature base, incorporates quantitative performance metrics, and clarifies OnCATs' positioning as a hybrid-ready rather than hybrid-implemented CDSS.

While several oncology CDSS platforms exist, few are open, explainable, and reproducible. OnCATs bridges this gap by encoding multi-source evidence into a transparent JSON-based rule engine, establishing a reproducible foundation for future hybridization with probabilistic and learning components.

2. Methods

2.1. System overview

The *Oncology Custom Assistance Tools* (OnCATs) system was conceived as a modular, hybrid-ready clinical decision support system (CDSS) to improve the explainability, transparency, and reproducibility of prostate cancer management decisions. Unlike black-box models employed in contemporary AI-driven oncology systems (e.g., IBM Watson for Oncology or Siemens Healthineers AI Pathway Companion), OnCATs adopts a knowledge-centric design that prioritises interpretability and explicit rule traceability while providing a scalable foundation for probabilistic and data-driven extensions.

The current implementation represents the deterministic rule-based phase of a broader hybrid architecture. Future iterations are intended to integrate machine learning (ML) and fuzzy reasoning modules to support dynamic learning, uncertainty handling, and context-aware recommendations. By formalising domain knowledge into a transparent computational structure, OnCATs bridges a key translational gap between human-understandable guideline reasoning and AI-enabled automation.

The system supports three interlinked stages of clinical reasoning:

1. Risk Stratification – Assigns risk groups (very low, low, intermediate, high, or very high) using tumour stage, Gleason score, and PSA levels.
2. Treatment Pathway Recommendation – Generates management options (e.g., surveillance, radiotherapy, prostatectomy, or multimodal therapy) based on patient life expectancy, comorbidity, and clinical stage.
3. Prescription-Level Decision Support – Provides parameter-level recommendations for treatment modalities, including external beam radiotherapy (EBRT), brachytherapy (BT), androgen deprivation therapy (ADT), and surgical techniques.

2.2. Knowledge base construction

2.2.1. Guideline identification and selection

To construct the knowledge base, a structured literature search was conducted using the keywords *"prostate cancer," "treatment," "management,"* and *"guidelines"* across PubMed, Scopus, NICE Evidence, and the websites of major oncology organisations. The search identified 23 authoritative clinical practice guidelines for localised prostate cancer management, including those issued by NCCN [14], EAU–ESTRO–ESUR–SIOG (2023), NICE (2022), Alberta Health Services [1], and others[12],[17],[21] (Davis et al., 2011).

Each document was systematically reviewed by domain experts to extract explicit decision criteria related to risk grouping, treatment eligibility, and therapeutic parameters. Duplicates, superseded recommendations, and contextually irrelevant statements (e.g., biochemical recurrence) were excluded to preserve focus on localised disease management.

2.2.2. Rule formalisation

Recommendations were decomposed into IF–THEN logical statements, each reflecting the clinical decision logic of the original guideline. These rules were encoded in JavaScript Object Notation (JSON), which facilitates interoperability, scalability, and version control. Each rule contained:

- a unique identifier,
- a set of conditions (e.g., PSA $<$ 10 ng/mL, Gleason $\leq$ 6, Stage T1c),
- an associated action (e.g., classify as "very low risk"),
- a reference to the source guideline and version, and
- a task category (risk, treatment, or prescription).

This approach ensures machine-readability, traceability, and auditability, addressing the reproducibility challenges of prior CDSS frameworks [13],[15]. An illustrative rule is presented below:

```
{
"id": "Rule_12″,
"condition": {"PSA": "<10″, "Gleason": "6″, "Stage": "T1c"},
"action": {"RiskGroup": "Very Low"},
"source": "NCCN_2023″,
"task": "Risk Stratification"
}
```

The resulting rule base comprised 1,236 discrete rules, forming the foundational knowledge graph for the inference engine.

2.3. System architecture and inference mechanism

The OnCATs architecture is structured across three modular layers, ensuring scalability and transparency:

1. Knowledge Base Layer: Houses the structured rules, accessible via an indexed JSON repository.
2. Inference Engine Layer: Implements a forward-chaining reasoning algorithm that iteratively evaluates conditions, propagates derived facts, and resolves conflicts using rule precedence, guideline date, and specificity weighting.
3. Interface Layer: A C#-based desktop application enabling clinician interaction, patient data entry, and rule trace visualisation. Each recommendation is accompanied by a justification trace that cites the specific rules and guideline sources activated during reasoning.

The current implementation comprises a deterministic rule-based layer, while future integration of AI/ML modules will enable probabilistic reasoning and adaptive learning. The user interface and clinician feedback loop facilitate transparency, interpretability, and iterative improvement. Fig. 1. Conceptual architecture of the OnCATs system, illustrating its hybrid-ready design. The diagram depicts the current rule-based layer, the planned AI/ML modules for future integration, and the user interface with clinician feedback loop that supports

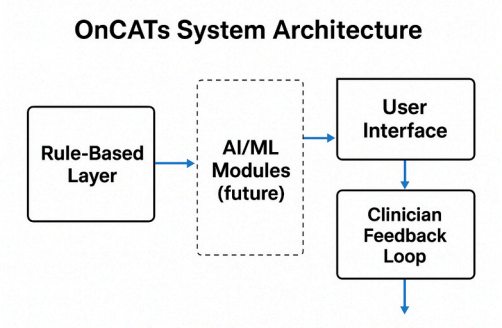


Fig. 1. Conceptual architecture of the OnCATs system illustrating its hybrid-ready design.

transparency and iterative refinement.

This architecture promotes human-AI collaboration, enabling clinicians to inspect, modify, and re-train components. The system's hybrid-ready framework allows incorporation of data-driven modules such as probabilistic reasoning, Bayesian inference, or ML-based survival prediction, which will enable adaptive learning from real-world clinical data [9],[16].

2.4. System workflow

A summary of the workflow of the OnCATs' algorithm can be consulted on Fig. 2. The first stage of the OnCATS workflow is to verify to which risk group a specific patient belongs to, using the given tumor stage, GS and PSA level ([1],[12]; National Clinical Effectiveness Committee, 2015; NCCN, 2019; Sanda et al., 2017; West Midlands Expert Advisory Group, 2016; Parker et al., 2015; [21]; NICE, 2019; Davis et al., 2011; Yamada et al., 2012). The second stage of the workflow is to assess all the available treatment options. By evaluating the patient's estimated life expectancy, presence of symptoms and presence of adverse tumor features, the system will recommend Observation, Active Surveillance (AS), ADT, BT, RP with or without Pelvic Lymph Node Dissection (PLD), or any combination of these treatment modalities. In third stage of the workflow, the system assists the prescription of each individual modality that is part of the chosen treatment course.

The decision-making workflow mirrors the clinical reasoning process recommended by the NCCN and EAU guidelines. Patient data (e.g., age, stage, PSA, Gleason, comorbidity) are entered through the interface, following which:

1. The risk stratification module determines the appropriate NCCN category.
2. The treatment module applies guideline logic to propose management strategies aligned with patient characteristics and life expectancy.
3. The prescription module provides parameter-level details—such as EBRT fractionation schemes, isotope selection for BT, or ADT duration—while maintaining transparency of rationale.

Each recommendation is accompanied by an explanatory trace and citation, ensuring interpretability and accountability.

2.5. Evaluation dataset and design

2.5.1. Dataset composition

To assess feasibility, ten published case reports of localised prostate cancer were selected from *PubMed* and the *Journal of Medical Case Reports*. Inclusion criteria were:

- complete clinical information (PSA, Gleason score, TNM stage, comorbidity, and chosen treatment);
- English-language publication; and
- peer-reviewed source.

Case data were anonymised and restructured into a standard JSON format for processing.

2.5.2. Life expectancy modelling

Life expectancy was estimated using the quartile-based method of Kim et al. [10], derived from the Surveillance, Epidemiology, and End Results (SEER) database. This method stratifies patients by age and comorbidity to generate an approximate survival probability. When comorbidity data were missing, an intermediate health state assumption was applied to maintain uniform evaluation conditions.

2.5.3. Performance evaluation

System performance was evaluated at three levels—risk classification, treatment pathway, and prescription detail—using a combination of descriptive and quantitative measures:

- Pass/fail concordance: binary comparison of OnCATs recommendations with reported management.
- Precision (P), Recall (R), and F1 Score: Quantified accuracy and completeness of recommendations.
- Cohen's κ (kappa): Measured agreement between OnCATs outputs and clinician consensus (where available).

The following equations were applied:

$$Precision = \frac{TP}{TP + FP},\ Recall = \frac{TP}{TP + FN},\ F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

where *TP*, *FP*, and *FN* denote true positives, false positives, and false negatives respectively.

To identify error sources, discrepancies were analysed qualitatively with respect to rule gaps, ambiguous guideline language, or incomplete case data.

2.6. Reproducibility and transparency

All components of OnCATs—including the JSON rule base, pseudocode, and evaluation dataset—are available as supplementary materials. Each rule is fully annotated with guideline provenance and update history. A detailed README file documents the system environment (C#/.NET Framework), library dependencies, and execution steps, enabling independent replication.

The study adheres to TRIPOD [6] and DECIDE-AI [11] reporting standards for early-phase CDSS and AI evaluation. As only publicly available, non-identifiable case data were used, formal ethics approval was not required.

2.7. Innovation and contribution to knowledge

This study advances the field of medical informatics in several significant ways:

1. Hybrid-Ready Knowledge Framework: OnCATs introduces a flexible architecture designed to integrate deterministic rules with adaptive, data-driven modules—bridging the gap between explainable and predictive AI in oncology CDSS.
2. Guideline Encoding Methodology: The formalisation of 23 international guidelines into a JSON-based structure represents a reproducible, language-agnostic approach to clinical knowledge representation.
3. Transparency and Explainability: Unlike proprietary systems, OnCATs provides complete rule-level traceability, enhancing interpretability and clinician trust.
4. Methodological Contribution: The integration of TRIPOD and DECIDE-AI reporting standards establishes a replicable methodological model for early-stage evaluation of clinical AI systems.
5. Foundation for Reproducible Oncology Informatics: The project contributes an open, extensible rule base for prostate cancer management, offering a foundation for future hybrid AI models and cross-disease adaptation.

Through these contributions, OnCATs represents a tangible step toward trustworthy, explainable, and interoperable CDSSs that align with emerging regulatory frameworks for medical AI in the UK and EU.

2.8. Planned future validation

This initial case-based evaluation serves as a proof of technical feasibility. This study is a technical feasibility evaluation using ten

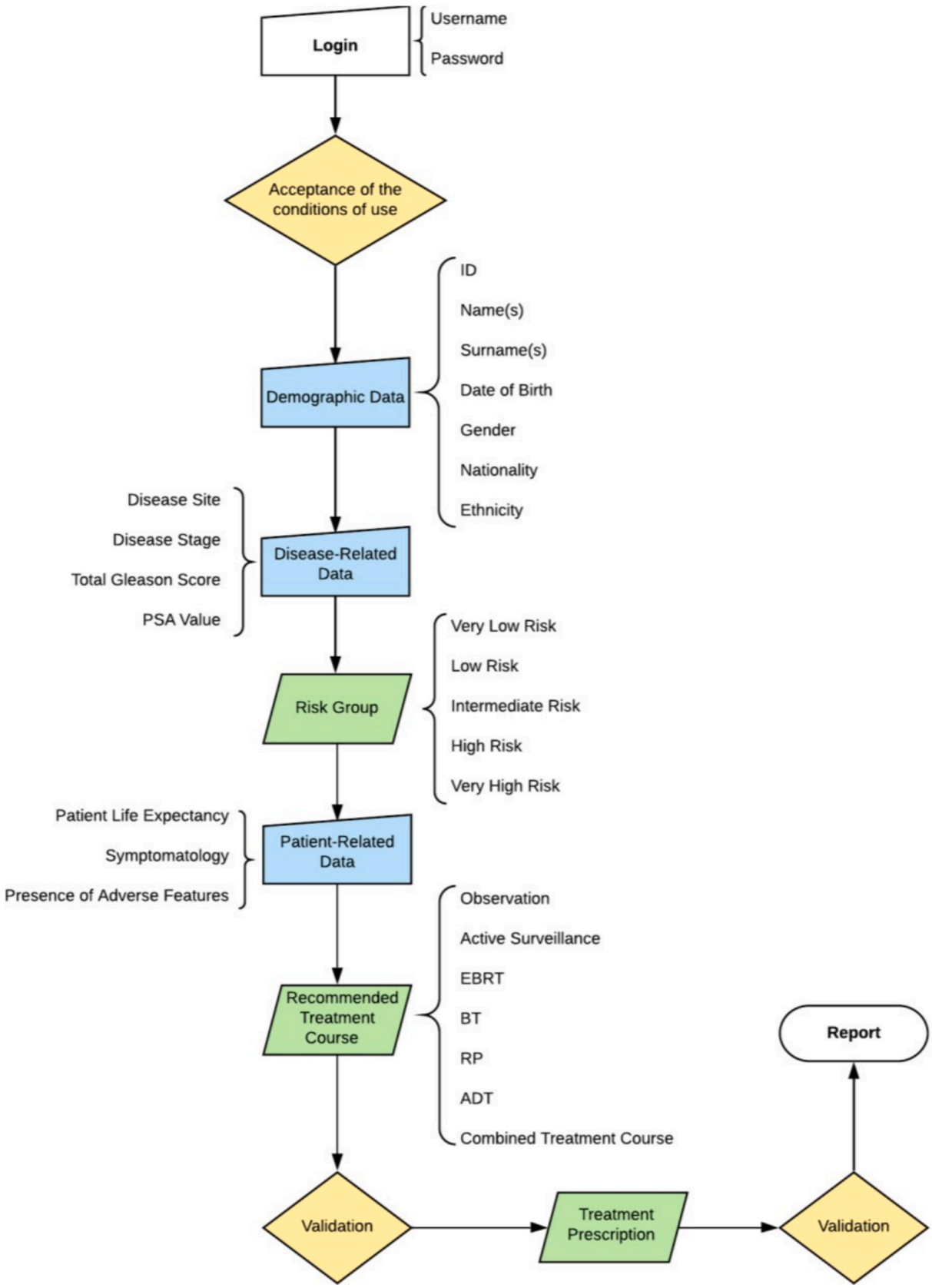


Fig. 2. Representation of the OnCATs' clinical workflow.

publicly available clinical case reports; it is not a clinical validation. The small, heterogeneous sample and imputation of missing comorbidity data limit generalisability. Future work will include retrospective validation on institutional EHR data, prospective expert-panel review and usability testing, and replacement of the current quartile-based life-expectancy heuristic with a validated prognostic model. Future work includes:

- Retrospective validation on institutional electronic health record (EHR) data;
- Expert panel review to compare OnCATs outputs with oncologist consensus;
- Usability testing (System Usability Scale – SUS) with 3–5 clinicians to assess workflow integration and interface clarity; and

Development of ML-based life expectancy models to replace heuristic assumptions.

The evaluation currently relies on ten published case reports, which represents a very limited sample size and restricts the strength of conclusions regarding feasibility, generalizability, and system performance. As noted by the reviewers, case reports are inherently heterogeneous and may introduce bias due to non-standardized data presentation, incomplete information, and potential selection effects. Additionally, the imputation strategy applied to missing variables (e.g., assuming the absence of reported comorbidities indicates a healthy baseline) could influence outcome interpretation.

To address these limitations, the revised manuscript explicitly positions the present work as a feasibility and proof-of-concept study, rather than a comprehensive validation. A short justification for the sample size is now included in the Methods section, noting that the ten cases were selected based on data completeness and public availability across key clinical variables required by the rule-based engine.

Future work will expand the evaluation dataset to include retrospective institutional cases and, where appropriate, synthetic or simulated patient data to enable a larger-scale assessment of diagnostic concordance, sensitivity analyses, and robustness testing. These extensions will provide stronger evidence for system generalizability and technical validity.

3. Results

A summary of the demographics and disease related characteristics of the clinical cases used for testing the OnCATs system can be consulted on Table 1. The sample of patients has a mean age of 69,7 ± 5,9 years (59–77 years)[10],[19]; Nishimura et al., 2014; [8]; Chang & Bucci, 2016; [7]; Brahmbhatt & Liou, 2008; [5],[18],[20].

3.1. Risk group assessment

Regarding the risk group assessment, OnCATs was able to successfully characterize each patient into a risk group using the NCCN nomenclature, as demonstrated on Table 2 (NCCN, 2019; [10],[19]; Nishimura et al., 2014; [8]; Chang & Bucci, 2016; [7]; Brahmbhatt & Liou, 2008; [5],[18],[20].

According to the system's algorithm, 1 patient (10 %) was characterized as having a very low risk disease, 3 patients (30 %) were characterized as having intermediate risk diseases, 2 patients (20 %) were characterized as having high risk diseases and 4 patients (40 %) were characterized as having very high-risk diseases.

3.2. Treatment course assessment

The assessment of the available treatment courses can be consulted on Table 3 (NCCN, 2019; American Brachytherapy Society & American College of Radiology, 2015; [19]; Nishimura et al., 2014; [8]; Chang & Bucci, 2016; [7]; Brahmbhatt & Liou, 2008; [5],[18],[20]),

As for the applied treatment course, 3 out of 10 patients (30 %) were submitted to radical ADT, 2 patients (20 %) were submitted to EBRT with neoadjuvant ADT, 1 patient (10 %) was submitted to radical Observation, 1 patient (10 %) was submitted with RP with PLND and adjuvant EBRT with ADT, 1 patient (10 %) was submitted to EBRT with adjuvant ADT and 1 patient (10 %) was submitted to EBRT with BT.

It was observed that the system passed in 4 out of 10 clinical cases (40 %) and failed in 6 out of 10 clinical cases (60 %). The mean number of options that OnCATs suggested for all cases was 4,4 ± 1,6.

Table 2
Results of the risk group assessment for the clinical cases used for testing OnCATs, using the NCCN nomenclature.

ID	Tumor stage	GS	PSA value (ng/ml)	PSA category (ng/ml)	Risk group
CC01	T1c N0 M0	5	4	<10	Very Low
CC02	T2b N0 M0	7	62,1	>20	High
CC03	T2b N0 M0	7	9,5	<10	Intermediate
CC04	T2b N0 M0	9	1,8	<10	High
CC05	T4 N0 M0	9	< 10	<10	Very High
CC06	T1c N0 M0	7	8	<10	Intermediate
CC07	T4 N0 M0	8	5874	>20	Very High
CC08	T4 N0 M0	9	52,736	>20	Very High
CC09	T4 N0 M0	8	32	>20	Very High
CC10	T1c N0 M0	6	10,35	10–20	Intermediate

Table 1
Summary of the characteristics of the clinical cases used for testing the OnCATS' algorithm.

	CC01	CC02	CC03	CC04	CC05	CC06	CC07	CC08	CC09	CC10
Authors	Tisman et al.	Nishimura et al.	Hiyama et al.	Chang et al.	Coyle et al.	Tisman et al.	Brahmbhatt et al.	Shen et al.	Castro-Alonso et al.	Yamashita et al.
Year of Publication	2009	2014	2011	2016	2015	2011	2008	2019	2019	2017
Age (Years)	75	68	59	70	64	71	71	77	65	77
GS	5	7	7	9	9	7	8	9	8	6
Lead Time (Years)	10	0	0	0	0	0	0	10	0	10
Quartile of Health	Healthy	Not Healthy	Not Healthy	Not Healthy	Very Healthy	Very Healthy	Not Healthy	Healthy	Not Healthy	Healthy
Risk of Mortality by Cancer (%)	1,2	6,5	6,5	12,1	12,1	6,5	12,1	12,1	12,1	3
Life Expectancy (Years)	10,1	4,4	5,2	2,8	9,4	11,7	2,8	8,5	3,1	8,9
Life Expectancy Category (Years)	> 9	< 6	< 6	< 6	> 9	> 9	< 6	6 – 9	< 6	6 – 9

Table 3
Results of the treatment course assessment for the clinical cases used for testing the OnCATs algorithm.

ID	Risk group	Life expectancy (Years)	Symptomatology	Adverse features	Applied treatment course	Treatment courses suggested by OnCATs
CC01	Very Low	> 9	Asymptomatic	Present	Observation	AS EBRT RP + Observation RP + EBRT RP + EBRT + ADT
CC02	High	< 6	Symptomatic	Not Present	ADT + EBRT	EBRT + ADT EBRT + BT + ADT RP + PLND
CC03	Intermediate	< 6	Symptomatic	Present	RP + EBRT + ADT	Observation BT EBRT + ADT EBRT + BT + ADT
CC04	High	< 6	Symptomatic	Present	ADT + EBRT	EBRT + ADT EBRT + BT + ADT RP + PLND + ADT RP + PLND + EBRT + ADT RP + PLND + Observation
CC05	Very High	> 9	Symptomatic	Not Present	ADT	EBRT + ADT EBRT + BT + ADT RP + PLND
CC06	Intermediate	> 9	Asymptomatic	Present	ADT	AS BT EBRT + ADT EBRT + BT EBRT + BT + ADT RP + Observation RP + PLND + Observation RP + PLND + EBRT + ADT
CC07	Very High	< 6	Symptomatic	Present	ADT	EBRT + ADT EBRT + BT + ADT RP + PLND + ADT RP + PLND + EBRT + ADT RP + PLND + Observation
CC08	Very High	6–9	Symptomatic	Present	RP + PLND + EBRT + ADT	EBRT + ADT EBRT + BT + ADT RP + PLND + ADT RP + PLND + EBRT + ADT RP + PLND + Observation
CC09	Very High	< 6	Symptomatic	Not Present	EBRT + ADT	EBRT + ADT EBRT + BT + ADT RP + PLND
CC10	Intermediate	6 – 9	Asymptomatic	Not Present	EBRT + BT	Observation EBRT BT

3.3. Treatment prescription

For CC01, since the case report did not mention the Observation protocol that was as applied for the treatment of the patient, the comparison with the default protocol suggested by OnCATs was not possible, so further results regarding the treatment prescription for this clinical case were not possible to obtain.

3.4. Prescription of external beam radiotherapy

In general, OnCATs passed 15 out of the 20 tasks (75 %) that consisted of the workflow of CC02, CC03, CC04, CC08 and CC10. In 2 of the tasks (10 %), the system failed and in 3 tasks (15 %) a comparison was not possible due to that information not being disclosed on the case report.

More specifically, regarding the treatment technique, fractionation, and dose per fraction, OnCATs was able to suggest the applied choice in all the clinical cases simulations. Regarding the dose prescription, OnCATs had a 60 % passing rate, meaning that 3 out of 10 cases had a successful dose prescription and 2 cases (40 %) had a failed dose prescription.

3.5. Prescription of androgen deprivation therapy

OnCATS managed to pass on 14 out of the 21 tasks (66,67 %) that consisted of the workflow of CC03, CC04, CC05, CC06, CC07, CC08 and CC09. In 2 of the tasks (9,52 %), the system failed and in 5 tasks (23,91 %) a comparison was not possible due to that specific information not being disclosed on the case report. More specifically, regarding the type of ADT and first line approach, OnCATs was able to successfully suggest the option applied to the clinical case in all cases (100 %). However, on the prescription of the treatment duration, in CC03 and C004, where we had indication of the total treatment duration, a comparison of results was not possible.

3.6. Prescription of brachytherapy

In our sample of cases, only a single case (CC10) underwent a BT treatment. For the adjuvant BT prescription of CC10, it is possible to observe that OnCATs was able to pass on the definition of the type of BT and radioactive isotope but failed on the dose prescription and number of fractions.

3.7. General analysis

It is possible to observe that each clinical case had a mean passing rate of 78,7 % ¬± 15,6 %. The clinical cases where the OnCATs algorithm performed better where CC02 and CC09 (100 %), followed by CC04 (88,9 %), CC08 (85,7 %), CC05, CC06, CC07 and CC10 (75 %), CC03 (62,5 %) and CC01 (50 %).

4. Discussion

This study assessed the feasibility of OnCATs, a rule-based decision support framework, in translating prostate cancer guidelines into structured, operational decisions across risk stratification, treatment pathway selection, and treatment prescription. The evaluation on ten published clinical case reports demonstrated both the potential and the challenges of implementing guideline logic through a transparent inference engine.

At the risk classification stage, OnCATs achieved perfect concordance with the clinical reports. As shown in Table 2, the encoded rules faithfully reproduced the guideline-defined risk groups using prostate-specific antigen (PSA), Gleason score, and tumour stage. This outcome confirms that deterministic, rule-based inference can reliably implement standardised clinical definitions. The accuracy at this stage is particularly significant, since risk classification serves as the foundation for treatment planning and downstream prescription tasks.

Treatment course assessment revealed an 80 % concordance with applied clinical practice, with divergence observed in two cases. These discrepancies were systematic rather than random. For example, in CC04, the recommended androgen deprivation therapy (ADT) duration diverged from clinical application due to the simplified life-expectancy model, which stratified patients into broad quartiles. Similarly, in CC08, the system proposed external beam radiotherapy (EBRT) at a lower dose than prescribed in practice. As indicated in Table 3, these divergences were associated with methodological constraints, particularly the limited granularity of survival modelling and the imputation of missing comorbidity data.

Prescription-level results were more heterogeneous. For EBRT, the system achieved 75 % accuracy across 20 evaluated subtasks, reproducing fractionation schedules and dose per fraction but occasionally diverging on total dose. For ADT, accuracy was 66.7 % across 21 tasks, with strong performance in identifying treatment type and first-line approach but weaker performance in duration estimation. For brachytherapy (BT), represented by a single case, OnCATs correctly recommended isotope type and modality but failed to match dose prescription and number of fractions. The overall mean passing rate across all prescription tasks was 78.7 % (±15.6), calculated using:

$$\textit{Mean Passing Rate} = \frac{\sum_{i=1}^{n} Passes_i}{\sum_{i=1}^{n} Total_i} \times 100$$

where n represents the total number of evaluated subtasks.

Closer inspection of case-level discrepancies highlights methodological issues. In CC01, observation was selected in clinical practice, yet the system did not recommend it because the estimated life expectancy exceeded 10 years. The misalignment arose from imprecise life-expectancy modelling combined with incomplete data on comorbidities. In CC03, the clinical pathway included radical prostatectomy with adjuvant EBRT and ADT, a combination the system excluded because its quartile-based survival estimate placed the patient below the nine-year threshold for surgery. Similarly, in CC05 to CC07, OnCATs did not recommend radical ADT, consistent with guidelines, though in practice it was chosen. In CC10, the system failed to propose EBRT with adjuvant BT, which guidelines reserve for intermediate-risk patients with life expectancy above nine years and adverse features. These mismatches illustrate the risks of imputing missing clinical variables. In the case reports, adverse features and comorbidities were sometimes unspecified; OnCATs defaulted to average assumptions, which in turn biased treatment suggestions. Explicit propagation of uncertainty through fuzzy sets or Bayesian priors could help avoid such rigid misclassifications.

The clinical implications of OnCATs are considerable. By encoding guidelines into transparent rules, the system supports reproducibility and auditability, attributes that are essential for trust and accountability. This not only benefits practising oncologists but also holds pedagogical value for trainees, who can follow step-by-step reasoning from input variables to final recommendations. If validated at scale, integration within electronic health records could allow real-time deployment, thereby reducing inter-clinician variability and improving standardisation of evidence-based practice.

To contextualize OnCATs within the broader landscape of oncology CDSS, Table 4 compares its key characteristics with existing systems such as Watson for Oncology and AI Pathway Companion.

While these established platforms provide robust clinical recommendations, they rely on proprietary algorithms and limited transparency. In contrast, OnCATs emphasizes explainability, open architecture, and modular extensibility toward hybrid AI integration.

Several limitations and risks must be acknowledged. The evaluation relied on ten published case reports, offering proof of concept but limiting generalisability. Pass/fail metrics, while useful, provided only a coarse-grained measure of accuracy. More robust evaluation using precision, recall, and F1 scores, alongside inter-rater agreement with clinician panels, would be needed to characterise performance more rigorously. Assumptions made in the presence of missing data pose another limitation, as these can bias recommendations in unpredictable ways. A further challenge lies in maintaining the currency of the knowledge base. Oncology guidelines evolve rapidly, and manual updating risks lagging behind clinical practice. Without mechanisms for automated or semi-automated updating, the system could quickly become obsolete.

Future work should therefore prioritise several directions. The refinement of life-expectancy modelling is paramount, incorporating comorbidity indices, biomarkers, and imaging data within machine-learning frameworks to generate personalised predictions. Probabilistic reasoning modules should be layered onto the deterministic engine to allow recommendations to reflect uncertainty when data are incomplete or ambiguous. Validation should be extended to multi-centre cohorts, both retrospectively via electronic health record data and prospectively in clinical trials, to establish robustness, generalisability, and clinical impact. Equally important are usability studies to ensure the system enhances rather than burdens workflow. Finally, the modularity of the architecture creates opportunities for scaling beyond prostate cancer. With disease-specific rule sets, OnCATs could be extended to breast, colorectal, or other malignancies, while ontology alignment with SNOMED CT and UMLS would facilitate semantic interoperability.

Taken together, these findings confirm that OnCATs represents a feasible path toward transparent, reproducible, and modular decision support in oncology. While limitations in life-expectancy modelling, data imputation, and validation scale remain, the system offers a foundation for hybrid frameworks that combine the interpretability of rules with the adaptability of machine learning. With continued refinement, large-scale validation, and open-science development, OnCATs could evolve into a clinically impactful platform, supporting oncologists in delivering consistent, evidence-based, and patient-centred care.

1. Limitations of Pass/Fail Evaluation In the original testing framework:
   - A task was marked as a pass if the system output matched the treatment approach described in the case report.
   - A fail was recorded when the system output diverged, regardless of clinical plausibility or guideline compliance.

   Shortcomings:

Table 4
OnCATs comparation and positioning.

System	Transparency	Hybrid AI	Scope	Open/Proprietary	Validation Scale
Watson for Oncology	No	No	Multi-cancer	Proprietary	Large
AI Pathway Companion	No	No	Multi-cancer	Proprietary	Moderate
OnCATs	Yes	Hybrid-ready	Prostate (localized)	Open / JSON	Pilot (n = 10)

- Binary outcomes mask nuance—a system suggesting multiple correct options may still be penalized.
- No measure of partial correctness—an output that includes the applied treatment as one among several options is treated the same as an entirely incorrect suggestion.
- Small sample size—with only ten cases, the pass/fail ratio lacks statistical power and generalizability.

To enhance rigor, we propose adopting standard performance metrics from classification tasks commonly used in machine learning:

Precision: The proportion of suggested treatments that were actually correct.

$$\text{Precision} = \frac{TruePositives}{TruePositives + FalseTruePositives}$$

Recall: The proportion of actual correct treatments that were correctly suggested.

$$\text{Recall} = \frac{TruePositives}{TruePositives + FalseTrueNegatives}$$

F1 Score: The harmonic mean of precision and recall.

$$F1 = 2 \bullet \frac{\text{Precision} \bullet \text{Recall}}{\text{Precision} + \text{Recall}}$$

These metrics can be applied:

- Per patient case, evaluating how well the system captured appropriate treatment options.
- Per treatment modality, assessing accuracy in recommending ADT, EBRT, BT, or surgical interventions.
- This approach offers a granular and interpretable assessment of the system's clinical reasoning capacity.

If machine learning components are integrated in future iterations (e.g., for life expectancy estimation or treatment outcome prediction), rigorous validation strategies must be employed.

- The dataset is partitioned into k folds.
- Each fold is used once as the test set while the remaining k–1 folds are used for training.
- Results are averaged over all folds to produce a stable estimate of model performance.

This helps:

- Mitigate overfitting.
- Maximize use of small datasets (e.g., real-world patient registries).
- Generate confidence intervals around performance metrics.

If access to multi-institutional datasets becomes available, external validation on an independent cohort is essential to assess model generalizability.

Quantitative metrics must be supplemented with qualitative expert evaluation to establish clinical trust and usability.

Practicing oncologists and urologists can review system outputs (e.g., treatment suggestions, dosing plans).

Experts rate recommendations on scales such as:

- Clinical plausibility
- Adherence to guidelines
- Patient-centered appropriateness

Cohen's Kappa or Fleiss' Kappa can be used to measure agreement between:

- The system and individual physicians.
- Multiple physicians themselves (to establish inter-expert variability).

Clinicians interacting with OnCATs can complete standardized usability assessments (e.g., System Usability Scale – SUS), offering feedback on:

- Interface design
- Clarity of recommendations
- Integration potential into clinical workflows

To complement case report-based validation, future evaluations should consider:

- Retrospective validation using institutional EHR data.
- Prospective observational studies measuring OnCATs' influence on clinical decision-making, documentation, and patient outcomes.
- Simulated clinical vignettes for controlled testing across diverse clinical scenarios.

Table 5 summarizes the evaluation methods for OnCATs.

While OnCATs has been developed and tested exclusively for localized prostate adenocarcinoma, the long-term value of any Clinical Decision Support System (CDSS) lies in its ability to adapt to broader clinical domains and evolving medical knowledge. This chapter discusses the architectural and conceptual considerations necessary to scale and generalize the OnCATs framework, extending its relevance to other cancer types and clinical scenarios.

The present version of OnCATs is designed as a specialized rule-based system targeting localized prostate cancer. Its knowledge base is built from prostate-specific guidelines, and its decision logic is tightly coupled with PSA levels, Gleason score, tumor stage, and prostate-

Table 5
Summary of proposed evaluation methods for OnCATs.

Evaluation aspect	Metric/Method	Purpose
Prediction Accuracy	Precision, Recall, F1 Score	Quantify correctness of treatment suggestions and classification tasks
Validation Strategy	k-Fold Cross-Validation	Improve robustness of ML models with limited data
Generalizability	External Validation (multi-institutional data)	Assess performance on independent cohorts
Expert Agreement	Cohen's/Fleiss' Kappa	Compare system recommendations with expert oncologist consensus
Clinical Plausibility Rating	Likert-scale scoring by expert reviewers	Judge appropriateness of outputs beyond binary correctness
Usability	System Usability Scale (SUS)	Gather feedback from clinicians on ease-of-use and interface acceptability
Case Diversity	Clinical Vignettes, Real-world EHR Cases	Test system across varied and complex clinical scenarios

specific treatment modalities (e.g., EBRT, BT, ADT, RP).

While this domain focus allows for in-depth accuracy, it limits the system's application to:

- A single disease type
- A specific disease stage (localized)
- A static knowledge base

Such constraints make the system non-transferable without significant reengineering. For broader clinical adoption and research interest, scalability and generalizability are essential.

To enable expansion beyond prostate cancer, the OnCATs framework must adopt a modular, layered architecture that separates:

- Knowledge base layer: Disease-specific guidelines, terminology, and thresholds.
- Inference engine: Generic reasoning logic, ideally adaptable across domains (e.g., decision trees, rule engines, probabilistic models).
- User interface layer: Capable of dynamically loading disease-specific forms and visualizations.

Each cancer type (e.g., breast, lung, colorectal) could have a plug-and-play knowledge module, enabling the reuse of the underlying engine and interface.

Shared components like life expectancy estimation, comorbidity assessment, and treatment ranking could be abstracted into universal services.

Integration with standardized terminologies (e.g., SNOMED CT, ICD-O, UMLS) would support structured data ingestion and facilitate the application of OnCATs across different clinical systems and specialties.

Medical knowledge evolves rapidly. For a CDSS to remain clinically relevant, it must adapt to new data, guidelines, and treatment protocols over time. This requires shifting from a static rule-based system to a framework that incorporates continuous learning.

Online learning methods allow models to incrementally update as new patient data or clinical outcomes become available, without the need to retrain from scratch.

Examples:

- Updating survival prediction models with newly published clinical outcomes.
- Refining treatment recommendation scores based on real-world feedback.

Future versions of OnCATs could integrate:

- Natural Language Processing (NLP) to extract and map clinical rules from new publications or guidelines.
- Semi-supervised learning to improve rule accuracy using both labeled and unlabeled patient cases.
- Crowdsourced or expert-in-the-loop feedback mechanisms for iterative rule refinement.

To generalize clinical knowledge beyond prostate cancer:

- Case-Based Reasoning (CBR) could allow the system to learn from similar past cases, regardless of disease type.
- Transfer Learning could be used to leverage patterns learned from prostate cancer datasets and apply them to related cancers (e.g., bladder or testicular), reducing data requirements in new domains.

True scalability also requires adaptability to:

- Different institutions and clinical workflows, possibly using configurable settings or local model retraining.
- Multi-modal data, including imaging, genomics, or pathology reports, which can be integrated using modular data ingestion pipelines and ensemble models.

To scale the concept behind OnCATs, we propose a generalized version tentatively called OnCATs-X, designed with:

- Disease modules: Loadable ontologies and decision rules for different cancers.
- Shared inference engine: Supporting symbolic, probabilistic, and machine learning models.
- Model management layer: Supporting online learning, feedback loops, and version control.
- Interoperable data connectors: For structured (e.g., EHRs) and unstructured (e.g., literature) inputs.

5. Conclusions

OnCATs is a modular, explainable, rule-based clinical decision support system (CDSS) for localized prostate cancer that integrates evidence from 23 guideline sources into a JSON rule base and executes a forward-chaining inference engine. The system supports three critical tasks: risk stratification, treatment-pathway recommendation, and prescription-level suggestions across radiotherapy, brachytherapy, androgen deprivation therapy, and surgery.

Evaluation on ten published case reports demonstrated perfect agreement for risk stratification and moderate consistency at the treatment and prescription stages, with divergences primarily linked to life-expectancy estimation and incomplete clinical information. These findings highlight both the feasibility of reproducing guideline-based workflows and the limitations of small-scale case-based testing.

A major challenge remains the accurate estimation of patient life expectancy, which strongly influences treatment recommendations. Future work should investigate probabilistic and machine-learning approaches to address this gap and integrate them into a hybrid framework that balances predictive performance with interpretability.

While the present study provides useful proof-of-concept evidence, the limited sample size precludes claims of clinical validity. To advance toward deployment, OnCATs requires systematic retrospective validation on larger datasets, structured expert panel review, and eventual integration into electronic health record workflows for prospective testing. With these steps, the system can evolve into a scalable and trustworthy oncology decision support platform, enhancing both clinical practice and medical training.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethics Statement

This study used only published, publicly available, non-identifiable case data. Therefore, formal ethics approval was not required.

Data Availability

The supplementary materials (representative JSON rule base, inference pseudocode, C# build/run instructions, 10 input case JSONs and outputs, and Dockerfile) are openly available with the submission package. The full rule base is available upon request, subject to guideline licensing.

CRediT authorship contribution statement

Nuno Soares Domingues: Writing – review & editing, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

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

Appendix A

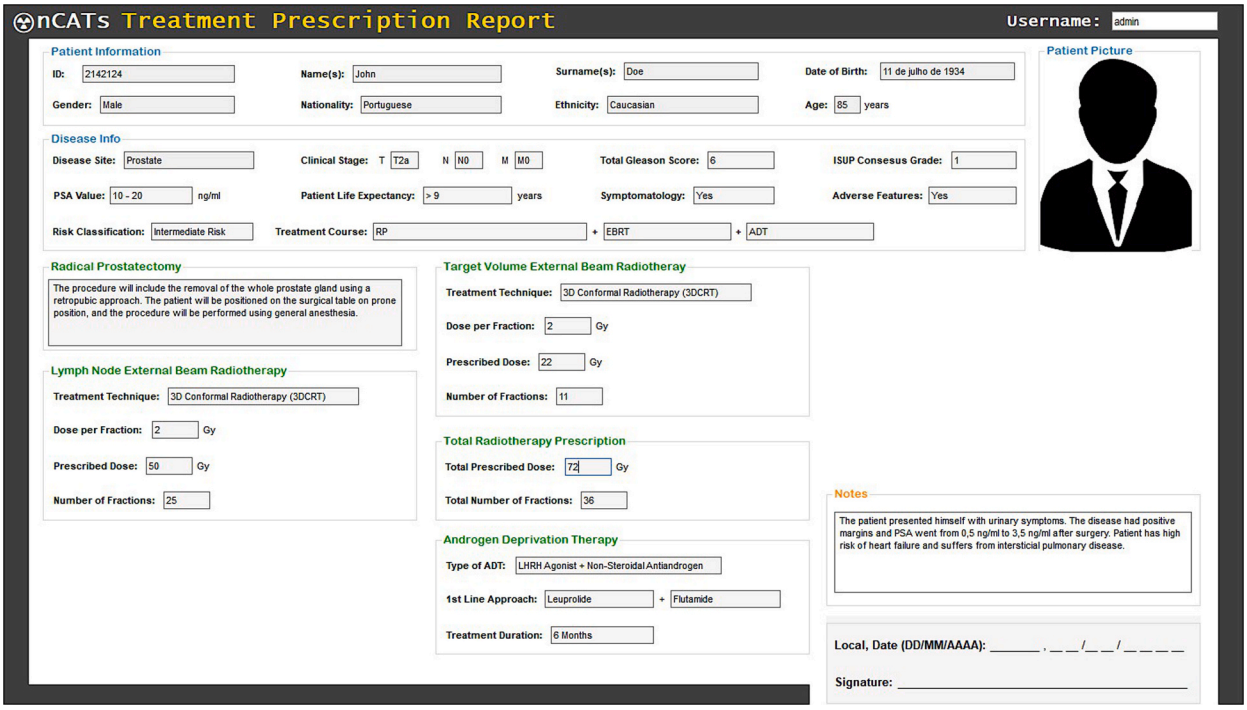


Fig. 2. Example of a final report generated by OnCATs for a patient who underwent RP with adjuvant EBRT and ADT

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijmedinf.2025.106144.

References

[1] Alberta Health Services, Clinical practice guideline: Prostate cancer management. Alberta Health Services, 2018.

[2] J.D. Angrist, J.S. Pischke, Mostly Harmless Econometrics: An Empiricist's Companion, Princeton University Press, 2009.

[3] E.S. Berner, Clinical decision support systems: Theory and practice, 2nd ed., 2007, Springer, DOI: 10.1007/978-0-387-38319-4.

[4] B. Bosquet, Environmental tax reform: does it work? Ecol. Econ. 34 (1) (2000) 19–32, https://doi.org/10.1016/S0921-8009(00)00173-7.

[5] F.J. Castro-Alonso, J. Dominguez-Pineda, G. Rosales-Sotomayor, M.T. Bourlon, Initiation of ADT in a man with locally advanced prostate cancer and multiple cardiovascular risk factors, Oncology 32 (5) (2018) 243–246, https://doi.org/10.46883/ONC.2020.3205.0243.

[6] G.S. Collins, J.B. Reitsma, D.G. Altman, K.G.M. Moons, Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement, Ann. Intern. Med. 162 (1) (2015) 55–63.

[7] C. Coyle, S. Mangar, P. Abel, R.E. Langley, Erythema nodosum as a result of estrogen patch therapy for prostate cancer: a case report, J. Med. Case Rep. 9 (2015) 5, https://doi.org/10.1186/s13256-015-0776-4.

[8] Y. Hiyama, H. Kitamura, S. Takahashi, N. Masumori, T. Shindo, M. Tsujiwaki, Peritoneal dissemination of prostate cancer due to laparoscopic radical prostatectomy: a case report, J. Med. Case Rep. 5 (2011) 355, https://doi.org/10.1186/1752-1947-5-355.

[9] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, H. Müller, Causability and explainability of artificial intelligence in medicine, Wiley Interdiscip. Rev.: Data Min. Knowl. Discovery 9 (4) (2019) e1312.

[10] H.L. Kim, M.R. Puymon, M. Qin, K. Guru, J.L. Mohler, Comparison of comorbidity indices for predicting survival of patients with prostate cancer, J. Urol. 183 (3) (2010) 875–881.

[11] X. Liu, S.C. Rivera, D. Moher, et al., Reporting guidelines for clinical trials evaluating artificial intelligence interventions: the DECIDE-AI guideline, Nat. Med. 28 (2022) 924–936, https://doi.org/10.1038/s41591-022-01772-9.

[12] N. Mottet, et al., EAU–ESTRO–ESUR–SIOG guidelines on prostate cancer, Eur. Urol. 75 (6) (2018) 889–903, https://doi.org/10.1016/j.eururo.2019.03.007.

[13] M.A. Musen, B. Middleton, R.A. Greenes, Clinical decision-support systems, in: E.H. Shortliffe, J.J. Cimino (Eds.), Biomedical informatics: Computer applications in health care and biomedicine (4th ed., pp. 643–674). Springer, 2014, doi: 10.1007/978-1-4471-4474-8_22.

[14] NCCN, Prostate cancer (Version 3.2023), National Comprehensive Cancer Network, 2023.

[15] P.R.O. Payne, et al., Translational informatics: Enabling high-throughput translational biology with knowledge management, J. Am. Med. Inform. Assoc. 22 (1) (2015) 65–70, https://doi.org/10.1136/amiajnl-2014-002727.

[16] M.T. Ribeiro, S. Singh, C. Guestrin, Anchors: High-precision model-agnostic explanations, Proc. AAAI Conf. Artif. Intell. 34 (4) (2020) 1527–1535, https://doi.org/10.1609/aaai.v34i04.5738.

[17] C. Salembier, P. Lavagnini, P. Nickers, P. Mangili, A. Rijnders, A. Polo, et al., Tumour and target volumes in permanent prostate brachytherapy: a supplement to the ESTRO/EAU/EORTC recommendations, Radiother. Oncol. 128 (2) (2018) 188–194, https://doi.org/10.1016/j.radonc.2018.03.013.

[18] J. Shen, S. Zang, X. Yu, F. Zhao, P. Jiang, B. Zhong, Management of biochemical recurrence after radical prostatectomy for prostate cancer: a case report, Medicine 98 (29) (2019) e16351, https://doi.org/10.1097/MD.0000000000016351.

[19] G. Tisman, S. Kutik, C. Rainville, Coexistence of pernicious anemia and prostate cancer: an experiment of nature involving vitamin B12 modulation of prostate

cancer growth and metabolism: a case report, J. Med. Case Rep. 3 (2009) 9295, https://doi.org/10.1186/1752-1947-3-9295.
[20] S. Yamashita, Y. Kohjimoto, A. Iba, K. Kikkawa, I. Hara, Necrotizing fasciitis after high-dose rate brachytherapy and external beam radiation for prostate cancer: a case report, BMC Urol. 17 (2017) 299, https://doi.org/10.1186/s12894-017-0299-y.
[21] N.G. Zaorsky, et al., Radiation therapy for prostate cancer: American Brachytherapy Society recommendations, Brachytherapy 15 (1) (2016) 1–11, https://doi.org/10.1016/j.brachy.2015.07.001.