Agentic Digital Twins for Precision Oncology: A Survey of Multimodal Foundation Models and Causal Reasoning for Personalized Cancer Care

Prasanna Vasudevan[1], *Pawan Kumar Chaurasia*[2]

[1] Postdoctoral Researcher, Lincoln University College, Malaysia

[2] Babasaheb Bhimrao Ambedkar Central University, Lucknow, Uttar Pradesh, India

Email ID [1]pdf.prasanna@lincoln.edu.my, [2]pkc.gkp@gmail.com

Abstract: Cancer treatment is increasingly guided by patient-specific molecular, radiological, and clinical evidence, yet most decision-support tools remain static, single-modality, and correlational. Three technological trends are converging to change this picture: (a) digital twins, which maintain continuously updated virtual replicas of a patient's tumor and physiology; (b) multimodal foundation models, which learn shared representations across imaging, pathology, genomics, and text; and (c) causal inference, which moves prediction beyond association toward estimating what would happen under an untried treatment. A fourth, more recent trend agentic artificial intelligence, in which large language model (LLM)-based agents plan, retrieve evidence, invoke tools, and act with limited autonomy offers a mechanism for orchestrating these components into a coherent clinical workflow. This paper surveys the state of the art in each of these four areas and synthesizes them into a conceptual framework for an agentic digital twin for precision oncology. We review recent literature on cancer digital twins, multimodal oncology foundation models, causal machine learning for treatment effect estimation, and agentic AI in clinical medicine, and we outline the architectural layers, applications, open challenges, and research directions relevant to building such systems responsibly.

Keywords: digital twin; genomics; oncology; modal; agentic AI

1. Introduction

Precision oncology aims to match each patient with the therapy most likely to benefit them, based on the molecular and clinical characteristics of their disease. Achieving this goal at scale requires integrating heterogeneous data radiology, digital pathology, next-generation sequencing, electronic health records (EHRs), and unstructured clinical notes and translating that integration into a decision that accounts for how the patient's tumor is likely to respond to a specific intervention. Historically, these tasks have been handled separately: predictive models are trained per modality, and treatment decisions are guided by population-level trial evidence rather than individualized counterfactual reasoning.

Digital twin technology, first popularized in aerospace and manufacturing, has been proposed as a way to maintain a living, patient-specific computational representation that is updated as new data arrive and can be used to simulate the effect of candidate interventions before they are given to the patient (Kemkar et al., 2024; Stefaniga et al., 2024). In parallel, multimodal foundation models large neural networks pretrained on broad, often self-supervised, biomedical data have demonstrated the ability to fuse imaging, omics, and text into shared representations that transfer across oncology tasks with limited labeled data (Huang et al., 2025; Kather et al., 2024). Causal inference methods provide the statistical machinery needed to ask counterfactual questions such as "would this patient benefit more from regimen A than regimen B," rather than merely predicting an outcome under the regimen the patient already received (Sanchez et al., 2022; Wang et al., 2024). Most recently, agentic AI LLM-based systems that plan multi-step tasks, retrieve external evidence, call specialized tools, and coordinate with other agents has begun to appear in clinical settings as a way of operating complex, multi-source reasoning (Collaco et al., 2026; Ferber et al., 2025).

This paper surveys these four strands of research and asks how they might be combined. We use the term agentic digital twin to describe a system in which an orchestrating agent (or set of agents) maintains and updates a patient-specific multimodal digital twin, uses foundation models to interpret incoming data, and uses causal inference to reason about candidate interventions, producing decision support that is continuously refreshed as new evidence becomes available.

2. Background and Related Work

2.1 Digital Twins in Oncology

A digital twin is a synchronized virtual representation of a physical entity that is updated as the physical entity changes and can be used to test interventions computationally before they are applied in the real world (National Academies of Sciences, Engineering, and Medicine, 2023). In oncology, this concept has been adapted to represent an individual tumor and patient physiology, integrating mechanistic tumor-growth models with patient-specific imaging, molecular, and clinical data (Kemkar et al., 2024). Scoping reviews report applications spanning treatment selection, radiotherapy planning, drug development, immuno-oncology modeling, surgical planning, and survivorship monitoring (Stefaniga et al., 2024). Karaman and Sebin (2025) draw an analogy to smart-city digital twins, arguing that tumors, like cities, are dynamic multiscale systems that benefit from continual data assimilation rather than a one-time model fit. Despite this promise, existing reviews consistently note that most published cancer digital twins rely on limited or synthetic data, are validated retrospectively rather than prospectively, and lack agreed-upon standards for verification, creating a gap between conceptual promise and clinical deployment (Stefaniga et al., 2024; Kemkar et al., 2024).

2.2 Multimodal Foundation Models in Cancer Care

Foundation models are large networks pretrained on broad data, typically using self-supervised objectives, and subsequently adapted to a range of downstream tasks with comparatively little labeled data (Huang et al., 2025). In oncology, foundation models have been developed for whole-slide pathology images, radiology, and molecular sequence data, and multimodal variants increasingly fuse two or more of these modalities together with structured EHR data and free-text clinical notes (Tripathi et al., 2024; Kather et al., 2024). Reported oncology applications include tumor segmentation and detection, prognosis and survival prediction, molecular subtype inference from histology, and prediction of response to specific drug classes, including checkpoint inhibitors, from combined radiology, pathology, and genomic inputs (Kather et al., 2024). Systematic reviews of multimodal foundation models for medical imaging note that most published systems are trained and evaluated within a single institution, are rarely validated prospectively, and are difficult to compare because of inconsistent reporting of fusion strategy and evaluation protocol (Huang et al., 2025).

2.3 Causal Inference for Personalized Treatment Decisions

Predictive accuracy alone does not answer the clinically relevant question of how a specific patient would respond to a treatment they did not receive. Causal inference formalizes this counterfactual question using frameworks such as the potential-outcomes model, structural causal models, propensity-score adjustment, and, increasingly, machine-learning-based estimators for heterogeneous treatment effects (Sanchez et al., 2022; Yao et al., 2021). Wang et al. (2024) survey methods for complex treatment settings multi-valued, continuous, or combination therapies that go beyond the classical binary-treatment case and are directly relevant to oncology, where regimens often involve multiple drugs, doses, and sequencing decisions. A dedicated survey of causal inference in the medical domain highlights applications to comparative effectiveness research from observational EHR data and cautions that unmeasured confounding, selection bias, and limited overlap between treatment groups remain persistent threats to validity (Wu et al., 2024). For oncology specifically, causal machine learning has been proposed as the mechanism by which a digital twin moves from correlational risk prediction to a genuine what-if simulation of alternative treatment plans (Sanchez et al., 2022).

2.4 Agentic AI in Clinical Medicine

Agentic AI refers to systems typically built around one or more LLMs that can decompose a goal into steps, retrieve external evidence, call specialized tools or models, maintain memory across steps, and in some architectures coordinate multiple specialized agents (Collaco et al., 2026). A recent scoping review identified deployed and prototype agentic systems spanning emergency medicine, radiology, rehabilitation, and oncology, reporting high task accuracy in narrow settings such as cancer diagnosis support and treatment-plan drafting, but noting that most systems remain exploratory, are evaluated on small or single-site cohorts, and rarely undergo prospective clinical validation (Collaco et al., 2026). A notable oncology-specific example is an autonomous agent developed and validated for clinical decision-making support in cancer care, which combined guideline retrieval with structured patient data to draft treatment recommendations for review by oncologists (Ferber et al., 2025). These systems illustrate how an orchestrating agent can serve as the connective layer between heterogeneous data sources, specialized foundation models, and the clinician, rather than as a monolithic diagnostic model in its own right.

3. Toward an Integrated Framework: The Agentic Digital Twin

Building on the four strands reviewed above, we outline a layered conceptual framework for an agentic digital twin in precision oncology. The framework is descriptive rather than a specific implemented system; it is intended to organize existing capabilities reported in the literature into a coherent architecture and to surface the integration challenges discussed in Section 5.

3.1 Data Layer

The data layer aggregates longitudinal, multimodal patient data, including radiology and digital pathology images, next-generation sequencing and other omics results, structured EHR fields, unstructured clinical notes, and, where available, wearable or patient-reported data. Consistent with observations in prior reviews, this layer must handle missingness, institutional heterogeneity in acquisition protocols, and asynchronous update frequency across modalities (imaging arrives episodically; vitals may stream continuously) (Stefaniga et al., 2024; Huang et al., 2025).

3.2 Foundation Model Layer

Modality-specific and multimodal foundation models transform raw data into clinically meaningful representations and predictions: tumor segmentation and radiomic features from imaging, molecular subtype and biomarker inference from pathology and sequencing, and summarized clinical state from notes and structured records (Tripathi et al., 2024; Kather et al., 2024). These representations populate and continuously refresh the state variables of the patient's digital twin, playing the role that sensor telemetry plays in industrial digital twins (Kemkar et al., 2024).

3.3 Causal Reasoning Layer

Given an updated twin state, the causal reasoning layer estimates the likely outcome of candidate interventions that have not yet been administered for example, comparing an additional line of chemotherapy against a targeted agent or immunotherapy using heterogeneous treatment-effect estimators and, where mechanistic tumor-growth models are available, hybrid simulation (Sanchez et al., 2022; Wang et al., 2024). This layer is what distinguishes an agentic digital twin from a conventional multimodal predictor: rather than only estimating risk under the observed treatment path, it is designed to answer counterfactual, decision-relevant queries.

3.4 Agentic Orchestration Layer

An orchestrating agent, or a small set of collaborating agents, coordinates the layers above: it decides when new data warrant an updated twin state, invokes the appropriate foundation model or causal estimator, retrieves relevant clinical guidelines or trial evidence, checks proposed recommendations against safety constraints, and prepares a structured summary for clinician review (Collaco et al., 2026; Ferber et al., 2025). Consistent with current practice in clinical agentic systems, this layer is best understood as decision support that keeps a human clinician in the loop rather than as an autonomous prescriber.

4. Applications Across the Cancer Care Continuum

Reviewed literature suggests plausible applications for agentic digital twins at several points in cancer care. In treatment selection, multimodal integration of radiology, pathology, and genomic data has been used to predict response to checkpoint-inhibitor immunotherapy, illustrating the kind of prediction a causal layer could extend into a comparative, counterfactual recommendation (Kather et al., 2024). In radiotherapy and surgical planning, twin-based simulation using patient-specific imaging has supported preoperative planning and dose optimization (Stefaniga et al., 2024). In monitoring and survivorship, continuously updated twins have been proposed to flag emerging resistance or relapse from longitudinal imaging and laboratory trends, a task naturally suited to an orchestrating agent that watches for clinically significant changes and triggers a re-evaluation of the twin (Karaman & Sebin, 2025). In clinical decision support more broadly, agentic systems have already been piloted for drafting oncology treatment recommendations for clinician review, foreshadowing the orchestration role described in Section 3.4 (Ferber et al., 2025; Collaco et al., 2026).

5. Open Challenges

Several challenges recur across the literature reviewed here. First, data scarcity and fragmentation: cancer digital twins and multimodal foundation models are frequently trained or validated on small, single-institution, or synthetic datasets, limiting generalizability (Stefaniga et al., 2024; Huang et al., 2025).

Second, causal identifiability: reliable treatment-effect estimation from observational oncology data requires strong, often untestable assumptions about unmeasured confounding, and combination or sequential treatments compound this difficulty (Wu et al., 2024; Wang et al., 2024). Third, validation and regulatory readiness: reviewers note that regulators such as the FDA and EMA have not yet established settled guidelines for evaluating digital twins or agentic clinical systems, and most reported systems lack prospective clinical trial evidence (Kemkar et al., 2024; Collaco et al., 2026). Fourth, autonomy and safety: as agentic systems gain the ability to call tools and take multi-step actions, ensuring that recommendations remain auditable, that failure modes such as hallucinated evidence are caught, and that a clinician retains final authority becomes a central design requirement rather than an afterthought (Collaco et al., 2026). Finally, equity: reviews of digital twins in healthcare caution that twinning solutions built on non-representative data risk encoding or amplifying existing disparities in cancer outcomes (Stefaniga et al., 2024).

6. Future Directions

Several directions follow from the gaps identified above. Prospective, multi-institutional validation studies are needed to move digital twins and agentic oncology assistants from retrospective demonstration to clinical evidence, mirroring the trajectory already argued for foundation models in medical imaging (Huang et al., 2025). Standardized benchmarks for causal treatment-effect estimation in oncology, with explicit reporting of confounding-control assumptions, would make comparisons across methods more meaningful (Wu et al., 2024). Tighter integration between mechanistic, biology-grounded tumor models and data-driven foundation models may improve extrapolation to treatments or patient subgroups that are underrepresented in historical data (Kemkar et al., 2024). Finally, governance frameworks specific to agentic clinical systems covering audit trails, human-in-the-loop checkpoints, and post-deployment monitoring will need to mature alongside the technology itself (Collaco et al., 2026).

7. Conclusion

Digital twins, multimodal foundation models, causal inference, and agentic AI have so far developed largely as separate research threads within oncology informatics. This survey has argued that they are complementary: foundation models supply the perceptual and representational backbone for a patient-specific twin, causal inference supplies the counterfactual reasoning needed for genuinely personalized treatment comparison, and agentic orchestration supplies the coordination layer that can keep the twin updated and keep a clinician informed. Realizing this integration safely will depend less on any single modeling breakthrough than on prospective validation, causal rigor, and careful attention to autonomy and equity as these systems move toward clinical use.

Summary Table

Table 1 summarizes the four technology pillars reviewed in this survey and how each maps onto the proposed agentic digital twin framework.

Component	Role in the Framework	Key Capabilities	Representative Sources	Main Limitations
Digital Twins	Maintains a continuously updated, patient-specific virtual replica of the tumor and physiology	Simulates candidate interventions before real-world use; integrates mechanistic tumor models with patient data	Kemkar et al. (2024); Stefaniga et al. (2024); Karaman & Sebin (2025)	Small/synthetic datasets; retrospective (not prospective) validation; no agreed verification standards
Multimodal Foundation Models	Perceptual/representational layer interprets imaging, pathology, genomics, EHR, and text	Fuses modalities into shared representations; transfers across tasks with limited labeled data	Huang et al. (2025); Tripathi et al. (2024); Kather et al. (2024)	Single-institution training; inconsistent fusion/evaluation reporting; limited prospective testing
Causal Inference	Reasoning layer estimates outcomes under treatments not yet given (counterfactual comparison)	Heterogeneous treatment-effect estimation; handles multi-valued/combination therapies	Sanchez et al. (2022); Wang et al. (2024); Wu et al. (2024); Yao et al. (2021)	Unmeasured confounding; weak identifiability for sequential/combination regimens
Agentic AI	Orchestration layer coordinates data, models, and evidence retrieval into clinician-facing recommendations	Multi-step planning; tool use; evidence retrieval; multi-agent collaboration	Collaco et al. (2026); Ferber et al. (2025)	Mostly exploratory/small-cohort studies; limited prospective clinical validation; autonomy/safety oversight still immature

References

1. Collaco, B. G., Haider, S. A., Prabha, S., Gomez-Cabello, C. A., Genovese, A., Wood, N. G., Bagaria, S. P., Gopala, N., Tao, C., & Forte, A. J. (2026). The role of agentic artificial intelligence in healthcare: A scoping review. npj Digital Medicine, 9, Article 345. https://doi.org/10.1038/s41746-026-02517-5
2. Ferber, D., et al. (2025). Development and validation of an autonomous artificial intelligence agent for clinical decision-making in oncology. Nature Cancer, 6(8), 1337–1349. https://doi.org/10.1038/s43018-025-00991-6
3. Huang, S.-C., Jensen, M., Yeung-Levy, S., Lungren, M. P., Poon, H., & Chaudhari, A. S. (2025). A systematic review and implementation guidelines of multimodal foundation models in medical imaging. Research Square. https://doi.org/10.21203/rs.3.rs-5537908/v1
4. Karaman, I., & Sebin, B. (2025). From data-driven cities to data-driven tumors: Dynamic digital twins for adaptive oncology. Frontiers in Artificial Intelligence, 8, Article 1624877. https://doi.org/10.3389/frai.2025.1624877
5. Kather, J. N., et al. (2024). Large language models and multimodal foundation models for precision oncology. npj Precision Oncology, 8, Article 33. https://doi.org/10.1038/s41698-024-00573-2
6. Kemkar, S., Tao, M., Ghosh, A., Stamatakos, G., Graf, N., Poorey, K., Balakrishnan, U., Trask, N., & Radhakrishnan, R. (2024). Towards verifiable cancer digital twins: Tissue level modeling protocol for precision medicine. Frontiers in Physiology, 15, Article 1473125. https://doi.org/10.3389/fphys.2024.1473125
7. National Academies of Sciences, Engineering, and Medicine. (2023). Foundational research gaps and future directions for digital twins. The National Academies Press. https://doi.org/10.17226/26894
8. Sanchez, P., Voisey, J. P., Xia, T., Watson, H. I., O'Neil, A. Q., & Tsaftaris, S. A. (2022). Causal machine learning for healthcare and precision medicine. Royal Society Open Science, 9(8), Article 220638. https://doi.org/10.1098/rsos.220638

9. Stefaniga, S. A., Cordoș, A. A., Ivascu, T., Feier, C. V. I., Muntean, C., Stupinean, C. V., Călinici, T., Aluaș, M., & Bolboacă, S. D. (2024). Advancing precision oncology with digital and virtual twins: A scoping review. Cancers, 16(22), Article 3817. https://doi.org/10.3390/cancers16223817
10. Tripathi, S., et al. (2024). Multimodal data integration for oncology in the era of deep neural networks: A review. Frontiers in Artificial Intelligence, 7, Article 1408843. https://doi.org/10.3389/frai.2024.1408843
11. Wang, Y., Li, H., Zhu, M., Wu, A., Xiong, R., Wu, F., & Kuang, K. (2024). Causal inference with complex treatments: A survey (arXiv:2407.14022). arXiv. https://doi.org/10.48550/arXiv.2407.14022
12. Wu, X., et al. (2024). Causal inference in the medical domain: A survey. Applied Intelligence, 54, 4911–4934.
13. Yao, L., Chu, Z., Li, S., Li, Y., Gao, J., & Zhang, A. (2021). A survey on causal inference. ACM Transactions on Knowledge Discovery from Data, 15(5), Article 74. https://doi.org/10.1145/3444944