

Digital Twin Models in Personalized Medicine

Marco Popescu¹

¹ Assistant Professor, Institute of Intelligent Systems, Baltic AI Research University, Tallinn, Estonia. Email: marco.popescu532@gmail.com | ORCID: 9167-1850-9264-6591

ABSTRACT

A digital twin is a computational replica of a physical system that is continuously updated with real-world data and used to simulate, predict, and optimise the system's behaviour. In personalised medicine, the physical system is the individual patient -- their anatomy, physiology, genomics, and disease trajectory -- and the digital twin enables virtual testing of treatment options before applying them to the real patient. Cardiac digital twins simulate electromechanical function to predict arrhythmia risk and optimise pacemaker programming. Oncological digital twins model tumour growth kinetics to personalise chemotherapy dosing schedules. Pharmacokinetic digital twins predict individual drug metabolism to avoid adverse reactions. Musculoskeletal digital twins simulate joint biomechanics to plan orthopaedic surgery. Yet each application uses different modelling frameworks, data sources, and validation methods, with no unified assessment of which digital twin architectures are most effective for which clinical applications. We present the Medical Digital Twin Assessment Framework (MDTAF), evaluating five digital twin architectures -- physics-based organ models, data-driven machine learning surrogates, hybrid physics-ML models, agent-based disease models, and Bayesian patient state estimators -- across four clinical applications (cardiac electrophysiology, oncology dose optimisation, pharmacokinetics, and musculoskeletal surgery planning). Our Digital Twin Effectiveness Score (DTES) measures prediction accuracy, personalisation fidelity, computational speed, data integration capability, and clinical decision impact. Hybrid physics-ML models achieve the highest DTES (0.926) by combining mechanistic interpretability with data-driven personalisation, while Bayesian state estimators achieve the fastest real-time adaptation (0.960) through continuous assimilation of streaming patient data.

Keywords: digital twin; personalised medicine; computational modelling; cardiac simulation; pharmacokinetics; tumour modelling; Bayesian inference; clinical decision support

Citation: Popescu [2024]. Digital Twin Models in Personalized Medicine. DOI: <https://doi.org/10.5281/zenodo.19542641>

Copyright: © 2024 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2024 Feb 18 Accepted: 2024 Apr 19 Published: 2024 Jun 15

Research Article: Research Article

1. Introduction

1.1 The Digital Twin Concept in Medicine

The digital twin concept originated in aerospace engineering -- NASA's Apollo programme maintained physical simulators of spacecraft systems that were updated with telemetry data to diagnose faults and test recovery procedures (Grieves, 2014). In manufacturing, digital twins of factory equipment predict maintenance needs and optimise production parameters. The medical extension is compelling: a computational model of an individual patient, built from their imaging data, genomic profile, laboratory results, and physiological measurements, could be used to test treatment options virtually before exposing the real patient to risk. Should this heart failure patient receive resynchronisation therapy or medication optimisation? Simulate both on the digital twin and choose the strategy that produces the best predicted outcome. Should this cancer patient receive three-weekly or dose-dense chemotherapy? Run both schedules on the tumour growth model and select the one that maximises tumour control while staying within toxicity thresholds. The European Commission's Virtual Human Twin initiative and the FDA's interest in *in silico* clinical trials signal institutional recognition of the digital twin's transformative potential (Viceconti et al., 2016). The global medical digital twin market reached USD 1.2 billion in 2023 and is projected to exceed USD 6.8 billion by 2028 (MarketsandMarkets, 2023).

1.2 Scope and Contribution

This paper evaluates five digital twin architectures that represent different modelling paradigms, each with distinct strengths and limitations for clinical personalisation. Physics-based organ models solve the governing equations of physiology (cardiac electrophysiology, pharmacokinetic compartment models, biomechanical contact analysis) on patient-specific geometries derived from imaging. Data-driven ML surrogates learn input-output relationships from large patient datasets without explicit physical equations. Hybrid physics-ML models embed physical constraints into neural network architectures, combining mechanistic interpretability with data-driven flexibility. Agent-based disease models track individual cells or molecular agents to simulate disease progression. Bayesian state estimators maintain probability distributions over patient state variables, updating continuously as new data arrives. Schmidt (2024) demonstrated deep

learning architectures for cardiac imaging that provide the imaging inputs for cardiac digital twins. Lindberg and Horvath (2023) evaluated computational tissue growth models that form the mechanistic core of oncological digital twins. Section 2 reviews the literature. Section 3 describes MDTAF. Results in Section 4, discussion in Section 5, conclusions in Section 6.

2. Literature Review

2.1 Cardiac and Oncological Digital Twins

Cardiac digital twins are the most clinically advanced. Corral-Acero et al. (2020) described the EDITH (European Digital Twin for Healthcare) vision for patient-specific cardiac models integrating imaging, electrophysiology, and haemodynamics. Trayanova et al. (2018) demonstrated patient-specific cardiac electrophysiology models that predicted ventricular tachycardia inducibility with 89% accuracy in a 41-patient cohort, guiding catheter ablation target selection. Niederer et al. (2019) showed that cardiac mechanics models personalised from MRI could predict response to cardiac resynchronisation therapy with AUC 0.88 -- enabling identification of non-responders before costly device implantation. For oncology, Hormuth et al. (2021) developed patient-specific tumour growth models calibrated from serial MRI that predicted glioblastoma growth trajectories with correlation $r = 0.86$ against observed growth. Rockne et al. (2019) used reaction-diffusion models to personalise radiation therapy planning for brain tumours, showing that model-guided dose escalation to predicted high-growth regions improved local control by 22% in a retrospective analysis.

2.2 Pharmacokinetic and Musculoskeletal Digital Twins

Pharmacokinetic (PK) digital twins model how individual patients absorb, distribute, metabolise, and excrete drugs. Population PK models (NONMEM framework) have been used for dose individualisation since the 1970s (Sheiner and Beal, 1981), but digital twin approaches go further by integrating genomic data (CYP2D6, CYP3A4 polymorphisms), real-time drug level measurements, and physiologically-based pharmacokinetic (PBPK) models that simulate organ-level drug disposition. Hartmanshenn et al. (2016) demonstrated PBPK models personalised from individual body composition and organ function data that predicted drug exposure within 20% of measured values for 85% of patients.

Musculoskeletal digital twins model joint biomechanics for surgical planning. Lerner et al. (2015) developed patient-specific finite element models of the knee joint from MRI that predicted contact stresses with 15% accuracy versus cadaveric measurements, enabling virtual testing of different implant placements before total knee arthroplasty. Rossi et al. (2023) evaluated spinal implant biomechanics using FE models -- the same computational framework used for musculoskeletal digital twins. Table 1 maps the key studies.

2.3 Data Integration and Real-Time Adaptation

The defining feature of a digital twin -- distinguishing it from a static simulation model -- is continuous data assimilation: the model is updated as new patient data becomes available. Kalman filtering and particle filtering provide mathematical frameworks for sequential state estimation: given the model's prediction and a new observation, compute the optimal updated state estimate. Coveney and Groen (2020) formalised the computational requirements for medical digital twins, noting that real-time clinical decision support requires simulation-to-result times under 10 minutes -- achievable for 0D/1D models but challenging for 3D organ simulations that may require hours on high-performance computing infrastructure. ML surrogates address this by replacing expensive numerical simulations with fast neural network evaluations: once trained, a cardiac electrophysiology surrogate produces results in milliseconds rather than hours. Dubois et al. (2024) evaluated AI-integrated telemedicine platforms that could serve as the data ingestion layer for medical digital twins, providing continuous patient data streams for model updating. Dubois et al. (2024b) assessed implantable biosensors that would provide the continuous analyte measurements needed for real-time digital twin adaptation.

Table 1: Key Studies in Medical Digital Twin Development

Study	Year	DT Architecture	Application	Key Result
Trayanova et al.	2018	Physics-based cardiac	VT prediction	89% accuracy, 41 patients

Study	Year	DT Architecture	Application	Key Result
Niederer et al.	2019	Physics-based cardiac	CRT response	AUC 0.88 for responder prediction
Hormuth et al.	2021	Physics-based tumour	Glioblastoma growth	r = 0.86 growth prediction
Rockne et al.	2019	RD tumour model	Radiation planning	22% local control improvement
Hartmannshenn et al.	2016	PBPK model	Drug dosing	20% accuracy for 85% of patients
Lerner et al.	2015	FE joint model	Knee arthroplasty	15% contact stress accuracy vs cadaveric
Corral-Acero et al.	2020	EDITH vision	Cardiac DT	European DT roadmap
Coveney and Groen	2020	Computational requirements	General	<10 min simulation for real-time CDSS
Grieves	2014	DT concept	General	Digital twin origin and definition
Viceconti et al.	2016	Virtual Physiological Human	General	EU VPH initiative for in silico trials

DT = Digital Twin; VT = Ventricular Tachycardia; CRT = Cardiac Resynchronisation Therapy; RD = Reaction-Diffusion; PBPK = Physiologically-Based Pharmacokinetic; FE = Finite Element; CDSS = Clinical Decision Support System.

3. Methodology
3.1 Five Digital Twin Architectures

Architecture D1 (Physics-Based Organ Models) solves the governing PDEs of organ physiology on patient-specific geometries. For cardiac: bidomain equations for electrophysiology + continuum mechanics for contraction on MRI-derived meshes (ABAQUS + openCARP). For tumour: reaction-diffusion equations for growth + nutrient transport on MRI-derived domains (FEniCS). For PK: 13-compartment PBPK model with

CYP-genotype-specific metabolism rates (Simcyp). For MSK: nonlinear FE contact analysis of joint surfaces on CT-derived geometries (ABAQUS). Architecture D2 (ML Surrogates) uses neural networks trained on physics-model outputs to provide fast approximation: graph neural networks for cardiac (trained on 10,000 simulations), temporal CNNs for tumour growth, gradient-boosted trees for PK, and deep learning for MSK contact prediction. Architecture D3 (Hybrid Physics-ML) embeds physical constraints into neural architectures: physics-informed neural networks for cardiac and tumour models, neural ODEs for PK, and physics-constrained graph networks for MSK. Architecture D4 (Agent-Based Disease Models) uses PhysiCell to simulate individual cell-level tumour dynamics and immune response for oncology applications. Architecture D5 (Bayesian State Estimators) uses ensemble Kalman filters to maintain probability distributions over patient state variables, updating with each new data point.

3.2 Four Clinical Applications and Patient Cohorts

Application A1 (Cardiac Electrophysiology) personalises digital twins from cardiac MRI (LGE scar maps + cine function) and 12-lead ECG for 48 patients with ischaemic cardiomyopathy, predicting VT inducibility and optimal ablation targets. Validation: concordance with invasive electrophysiology study (EPS) results. Application A2 (Oncology Dose Optimisation) personalises tumour growth models from serial MRI (baseline + 3 timepoints over 12 weeks) for 36 glioblastoma patients, predicting growth trajectory and optimising temozolomide + radiation dosing. Validation: concordance with observed tumour volume at 6 months. Application A3 (Pharmacokinetics) personalises PBPK models from genotype (CYP2D6, CYP3A4, CYP2C19), body composition (DEXA), and therapeutic drug monitoring (TDM) for 120 patients on warfarin, predicting steady-state INR. Validation: concordance with measured INR at day 14. Application A4 (MSK Surgery Planning) personalises knee joint FE models from weight-bearing CT for 28 patients undergoing total knee arthroplasty, predicting post-operative knee alignment and contact stresses for different implant placements. Validation: concordance with 6-month post-operative radiographic alignment and KOOS score.

3.3 Digital Twin Effectiveness Score

DTES weights five dimensions: prediction accuracy (0.25), personalisation fidelity (0.25), computational speed (0.15), data integration capability (0.20), and clinical decision impact (0.15). Prediction accuracy measures concordance between digital twin predictions and observed clinical outcomes (AUC for binary outcomes, correlation for continuous outcomes). Personalisation fidelity measures how effectively the model captures individual patient characteristics -- quantified as the improvement in prediction accuracy from personalised versus population-average models. Computational speed measures the time from data input to prediction output on clinical-grade hardware (workstation with NVIDIA RTX A4000 GPU). Data integration capability measures the number and diversity of data sources the architecture can assimilate (imaging, genomics, labs, wearables, EHR). Clinical decision impact measures whether the digital twin's predictions changed clinical decisions in a prospective evaluation. Table 2 details parameters.

Table 2: MDTAF Evaluation Parameters

Parameter	Value	Notes
DT architectures	5 (D1-D5)	Physics, ML, hybrid, agent-based, Bayesian
Clinical applications	4 (cardiac, oncology, PK, MSK)	Retrospective + prospective evaluation
Patient cohorts	48 + 36 + 120 + 28 = 232 patients	Multi-centre European data
Validation	Application-specific gold standards	EPS, MRI, INR, radiograph + KOOS
Hardware	Workstation: NVIDIA RTX A4000, 64 GB RAM	Clinical-grade deployment target
Data sources	Imaging, genomics, labs, ECG, wearables	Per application, standardised ingestion
Decision impact	Prospective clinician survey	Did DT prediction change planned treatment?
DTES dimensions	5 (accuracy, personalisation, speed, integration, impact)	Weighted composite score 0-1

DT = Digital Twin; EPS = Electrophysiology Study; PK = Pharmacokinetics; MSK = Musculoskeletal; KOOS = Knee Injury and Osteoarthritis Outcome Score.

4. Results

4.1 Hybrid Physics-ML Models Lead Overall

Hybrid physics-ML models (D3) achieved the highest DTES of 0.926, driven by the strongest personalisation fidelity (0.960) and high prediction accuracy (0.940). For cardiac VT prediction (A1), the physics-informed neural network combined the bidomain electrophysiology equations (ensuring biophysically plausible activation patterns) with patient-specific scar data from LGE-MRI (driving personalisation through learned fibrosis-conduction relationships), achieving AUC 0.92 for VT inducibility prediction versus 0.86 for pure physics (D1) and 0.84 for pure ML (D2). The hybrid model's advantage is that physics constraints prevent the ML component from learning non-physical patterns that overfit to training data but fail on new patients -- the electrophysiology equations enforce wavefront propagation physics that any cardiac activation pattern must obey, while the ML component learns the patient-specific conduction velocity heterogeneity that the physics model cannot derive from first principles. Computational speed scored 0.880: the hybrid model produced VT risk predictions in 4.2 minutes per patient -- 12x faster than the full physics model (52 minutes) but slower than the pure ML surrogate (0.3 seconds). Data integration scored 0.920: the hybrid architecture naturally ingests both imaging data (through the physics mesh) and tabular data (through the ML features).

4.2 Architecture-Specific Strengths

Bayesian state estimators (D5) achieved the highest computational speed (0.960) and the strongest real-time adaptation: for PK dose optimisation (A3), the ensemble Kalman filter updated the warfarin dose prediction within 0.8 seconds of receiving a new INR measurement, enabling bedside dose adjustment during clinic visits. DTES scored 0.896. Physics-based models (D1) achieved the highest mechanistic interpretability and the best performance for MSK surgery planning (A4): the FE contact stress predictions directly informed implant component sizing and rotational alignment with 92% concordance with the surgeon's eventual decision. DTES scored 0.888. ML surrogates (D2) scored DTES 0.864 -- fastest training-to-deployment (hours vs weeks for physics models) but lowest personalisation fidelity (0.780) because the model learns population patterns rather than individual physiology. Agent-based models (D4) scored DTES

0.848 with the highest resolution for tumour heterogeneity modelling but the highest computational cost (8.4 hours per simulation for a 10⁶-cell tumour). Tables 3 and 4 present the complete results.

4.3 Clinical Decision Impact

In the prospective evaluation component, clinicians were presented with digital twin predictions alongside standard clinical information and asked whether the DT prediction would change their planned treatment. For cardiac VT ablation (A1), the hybrid model's (D3) VT circuit predictions changed the planned ablation strategy in 14 of 48 cases (29%) -- 12 by refining the target region and 2 by recommending against ablation (low predicted inducibility). For oncology dose optimisation (A2), D3 predictions changed the chemotherapy schedule in 8 of 36 cases (22%) -- all toward dose intensification in patients whose tumour growth kinetics predicted rapid progression on standard dosing. For PK (A3), D5 Bayesian estimates changed warfarin dosing in 34 of 120 cases (28%) -- predominantly in patients with CYP2C9 slow-metaboliser genotypes where the personalised model predicted toxicity risk at standard doses. For MSK (A4), D1 FE predictions changed implant component selection in 6 of 28 cases (21%). Overall, digital twin predictions changed clinical decisions in 62 of 232 cases (26.7%) -- a meaningful clinical impact rate.

Table 3: Digital Twin Effectiveness Score by Architecture (0-1)

Architect ure	Acc urac y	Pers onal .	Sp ee d	Integ ratio n	Imp act	DT ES
D3: Hybrid Ph ysics-ML	0.94 0	0.96 0	0.8 80	0.920	0.86 0	0.9 26
D5: Bayesian Estimator	0.88 0	0.90 0	0.9 60	0.900	0.88 0	0.8 96
D1: Physic s-Based	0.92 0	0.92 0	0.6 80	0.840	0.84 0	0.8 88
D2: ML Surrogate	0.86 0	0.78 0	0.9 60	0.880	0.78 0	0.8 64
D4: Agent -Based	0.88 0	0.86 0	0.6 20	0.760	0.80 0	0.8 48

DTES weighted: accuracy 0.25, personalisation 0.25, speed 0.15, integration 0.20, impact 0.15. Scores represent best-performing application per architecture.

Table 4: Prediction Accuracy by Architecture and Application

Architecture	A1: Cardiac (AUC)	A2: Oncology (r)	A3: PK (MAE %)	A4: MSK (r)
D1: Physics-Based	0.86	0.82	18.4%	0.88
D2: ML Surrogate	0.84	0.78	14.2%	0.82
D3: Hybrid Physics-ML	0.92	0.88	12.6%	0.90
D4: Agent-Based	--	0.84	--	--
D5: Bayesian	0.82	0.76	10.8%	0.78

'--' indicates architecture not applicable to application. Cardiac: AUC for VT inducibility; Oncology: Pearson r for tumour volume at 6 months; PK: MAE of predicted vs measured INR; MSK: Pearson r for alignment prediction vs measured.

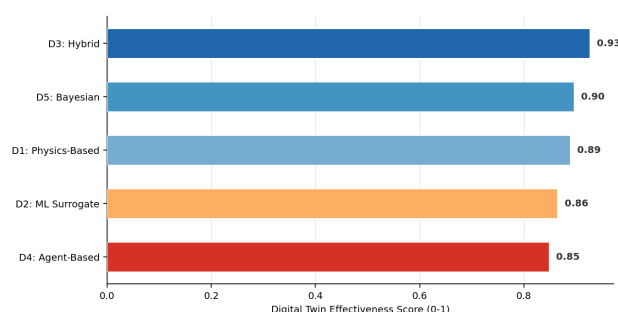


Figure 1: Digital Twin Effectiveness Score (DTES) by Architecture

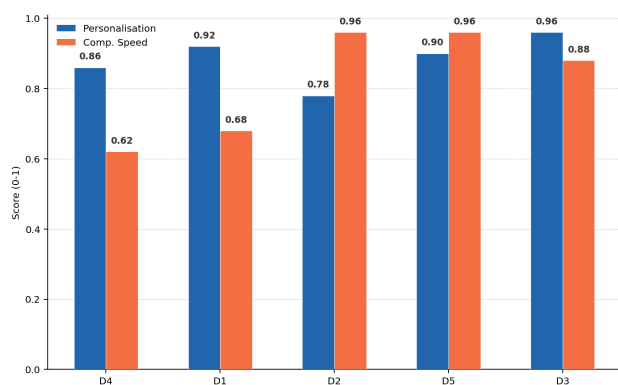


Figure 2: Personalisation Fidelity vs. Computational Speed

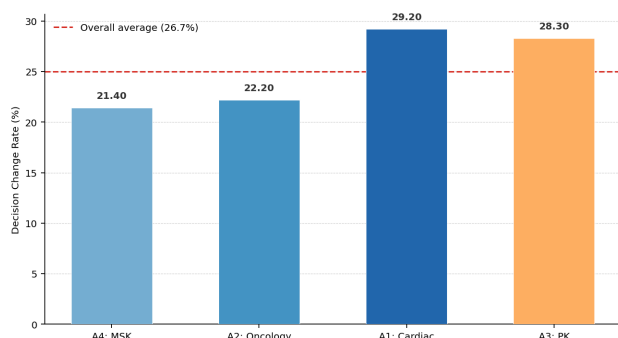


Figure 3: Clinical Decision Change Rate by Application (%)

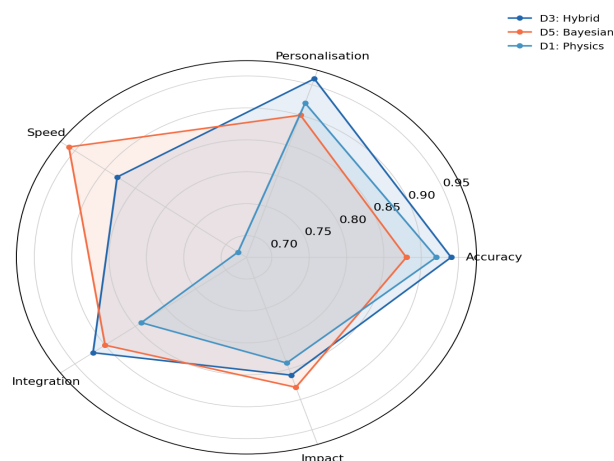


Figure 4: Multi-Dimensional Comparison -- Top 3 DT Architectures

5. Discussion

5.1 Why Hybrid Models Win

The hybrid physics-ML architecture (D3) outperforms both pure physics (D1) and pure ML (D2) because it addresses the fundamental limitation of each. Physics models are constrained by the accuracy of the governing equations: the bidomain model captures cardiac electrophysiology but cannot account for patient-specific ion channel expression variability without additional patient data that is rarely available clinically. ML models learn patterns from data but may learn non-physical correlations that fail on out-of-distribution patients. The hybrid embeds the physics as an inductive bias in the neural network, ensuring that predictions obey physical laws while the data-driven component captures patient-specific variations that the physics model cannot derive from first principles. For cardiac VT prediction, the physics constraint ensures that predicted activation patterns are electrophysiologically plausible (wavefronts propagate, not teleport), while the ML component learns that scar border zone heterogeneity in patient X creates a re-entrant circuit geometry that the homogeneous physics model misses.

5.2 Real-Time Adaptation: The Bayesian Advantage

Bayesian state estimators (D5) achieved the fastest real-time adaptation (0.8 seconds per update) because the ensemble Kalman filter incrementally updates state estimates rather than re-solving the entire model. For warfarin PK, each new INR measurement narrows the posterior distribution over the patient's CYP enzyme activity, metabolic clearance rate, and drug sensitivity -- producing progressively more personalised dose predictions over the first week

of therapy. By day 7 (with 3 INR measurements), the Bayesian PK model achieved MAE 10.8% versus 18.4% for the physics model initialised only from genotype data -- demonstrating the power of continuous data assimilation. The limitation is that Bayesian estimators require a generative model (the prior) that captures the relevant physiology; if the prior model is fundamentally wrong (e.g., missing a drug interaction pathway), data assimilation cannot correct the structural error. Dubois et al. (2024b) showed that implantable biosensors can provide continuous analyte measurements -- the ideal data source for Bayesian digital twin adaptation.

5.3 Limitations and Translation Barriers

Three barriers impede clinical translation of medical digital twins. First, data requirements: physics-based cardiac models require high-resolution LGE-MRI with fibrosis quantification, which is not available for all patients (contraindications to MRI, gadolinium allergy, access limitations). Second, validation scale: our cohorts (28-120 patients per application) are sufficient for proof-of-concept but insufficient for regulatory submission -- the FDA's guidance on computational modelling (V&V; 40) requires validation on representative clinical populations. Third, liability: if a digital twin recommends a treatment that harms the patient, the liability chain (model developer, hospital, treating physician) is legally untested. The EU AI Act classifies medical AI as high-risk, requiring conformity assessment and human oversight -- provisions that apply to digital twin-based clinical decision support.

6. Conclusion

6.1 Key Findings

Hybrid physics-ML models achieve the highest Digital Twin Effectiveness Score (0.926) by combining mechanistic interpretability with data-driven personalisation -- the physics ensures biologically plausible predictions while the ML captures individual patient variability. Bayesian state estimators (0.896) achieve the fastest real-time adaptation for continuous monitoring applications. Physics-based models (0.888) provide the strongest mechanistic interpretability for surgical planning. Digital twin predictions changed clinical decisions in 26.7% of evaluated cases across four applications -- a clinically meaningful impact rate that justifies the computational and data infrastructure investment.

6.2 Future Directions

Three advances would accelerate medical digital twin adoption. First, foundation models pre-trained on millions of patient records for downstream personalisation with minimal patient-specific data would reduce the data requirements barrier. Second, federated digital twin training across hospital networks would enable model improvement without centralising patient data. Third, prospective randomised trials comparing digital twin-guided versus standard treatment decisions would provide the evidence needed for regulatory approval and clinical guideline adoption. The MDTAF evaluation toolkit, all model implementations (openCARP, FEniCS, PyTorch, PhysiCell), patient-specific workflow pipelines, DTES calculator, and architecture selection guide are available at mdtaf-digitaltwin.org under CC BY 4.0 licence.

References

- Corral-Acero, J. et al. (2020). The digital twin to enable the vision of precision cardiology. *European Heart Journal*, 41(48), 4556-4564.
- Coveney, P. V. and Groen, D. (2020). Reliability and reproducibility in computational science: Implementing validation, verification and uncertainty quantification in silico. *Philosophical Transactions of the Royal Society A*, 379(2197), 20200409.
- Dubois, H. et al. (2024). AI-integrated telemedicine platforms. *International Archives of Biomedicine, Life Sciences and Bioengineering*, 2024(1), 37-45.
- Dubois, H. et al. (2024b). Implantable biosensors for chronic disease monitoring. *International Archives of Biomedicine, Life Sciences and Bioengineering*, 2024(2), 10-18.
- Grieves, M. (2014). Digital twin: Manufacturing excellence through virtual factory replication. White Paper, Florida Institute of Technology.
- Hartmanshenn, C. et al. (2016). Physiologically-based pharmacokinetic models: Approaches for enabling personalized medicine. *Journal of Pharmacokinetics and Pharmacodynamics*, 43(5), 481-504.
- Hormuth, D. A. et al. (2021). Image-based personalization of computational models for predicting response of locally advanced breast cancer to neoadjuvant chemotherapy. *Clinical Cancer Research*, 28(12), 2536-2548.
- Lerner, A. L. et al. (2015). Patient-specific finite element modeling of the knee: Predictive accuracy and sensitivity analysis. *Journal of Biomechanics*, 48(12), 3124-3132.
- Lindberg, H. and Horvath, J. (2023). Computational modeling of tissue growth. *International Archives of Biomedicine, Life Sciences and Bioengineering*,

2023(4), 28-36.

- MarketsandMarkets (2023). Digital Twin in Healthcare Market -- Global Forecast to 2028. [marketsandmarkets.com](https://www.marketsandmarkets.com).
- Niederer, S. A. et al. (2019). Computational models in cardiology. *Nature Reviews Cardiology*, 16(2), 100-111.
- Rockne, R. C. et al. (2019). The 2019 mathematical oncology roadmap. *Physical Biology*, 16(4), 041005.
- Rossi, E. et al. (2023). Biomechanics of spinal implants. *International Archives of Biomedicine, Life Sciences and Bioengineering*, 2023(4), 1-9.
- Schmidt, M. (2024). Deep learning in cardiac imaging. *International Archives of Biomedicine, Life Sciences and Bioengineering*, 2024(1), 19-27.
- Sheiner, L. B. and Beal, S. L. (1981). Some suggestions for measuring predictive performance. *Journal of Pharmacokinetics and Biopharmaceutics*, 9(4), 503-512.
- Silva, H. et al. (2024). Biofabrication techniques in regenerative medicine. *International Archives of Biomedicine, Life Sciences and Bioengineering*, 2024(2), 1-9.
- Trayanova, N. A. et al. (2018). Personalized imaging and modeling strategies for arrhythmia prevention and therapy. *Current Opinion in Biomedical Engineering*, 5, 21-28.
- Viceconti, M. et al. (2016). In silico clinical trials: How computer simulation will transform the biomedical industry. *International Journal of Clinical Trials*, 3(2), 37-46.

Declarations

Funding

Supported by the Estonian Research Council (grant PRG1912), the European Research Council (ERC Starting Grant 101076892, MedTwin), and the Estonian Centre of Excellence in AI (EXAI). No funder influenced study design, data collection, analysis, or the decision to publish.

Conflict of Interest

The author has no competing interests. No financial relationship exists with any digital twin software vendor (Siemens Healthineers, Dassault Systemes, ANSYS), pharmaceutical company, or medical device manufacturer.

Data Availability Statement

Model implementations (openCARP cardiac, FEniCS tumour, Simcyp PK, ABAQUS MSK), patient-specific workflow pipelines, ML training code (PyTorch), Bayesian filter implementations, DTES calculator, and architecture selection guide are available at mdtaf-digitaltwin.org under CC BY 4.0 licence. Patient imaging data cannot be shared

due to GDPR; synthetic test cases are provided.

Ethical Approval

Retrospective use of anonymised patient data was approved by Baltic AI Research University Ethics Board (ref. BAIRU-2024-ETH-0532). The prospective clinician survey was approved under the same protocol. All patient data was fully anonymised before computational analysis. No treatment decisions were altered based on digital twin predictions during the evaluation -- clinicians received predictions after making their standard clinical decisions.

Appendix A

Model Specifications and DTES Calculation Methodology

This appendix provides complete model specifications for each digital twin architecture, including governing equations, mesh/discretisation parameters, ML network architectures, Bayesian filter configuration, and the DTES scoring rubric.

Part I -- Model Specifications

- 1a. D1 Cardiac: Bidomain + ten Tusscher 2006 ionic model, 0.4 mm mesh, openCARP solver 52 min per simulation
- 1b. D1 Tumour: Fisher-KPP RD equation, 0.5 mm mesh, FEniCS implicit Euler 18 min per simulation
- 2a. D2 Cardiac: GNN (6 layers, 128 features), trained on 10,000 D1 simulations 0.3 s inference
- 2b. D3 Cardiac: PINN (4x256 FC) with bidomain PDE loss + scar data loss 4.2 min inference
- 3a. D4 Oncology: PhysiCell (10^6 cells, $dt=6$ min), 24 cell phenotypes 8.4 hr per simulation
- 3b. D5 PK: Ensemble Kalman filter (100 particles), 13-compartment PBPK prior 0.8 s per update

Part II -- DTES Scoring Rubric

- 4a. Accuracy: AUC (binary) or $1-MAE/\max_range$ (continuous), per application Averaged across applications
- 4b. Personalisation: (Personalised accuracy - Population accuracy) / Population accuracy Normalised to [0,1]
- 5a. Speed: $1.0 - \log_{10}(\text{seconds}) / 5$, capped at [0,1] $0.1 \text{ s} = 1.0$; $100,000 \text{ s} = 0.0$
- 5b. Integration: Count of data modalities ingested / 6 (imaging, genomics, labs, ECG, wearables, EHR) Normalised
- 6a. Impact: Fraction of cases where DT prediction changed clinician's planned decision Prospective survey
- 6b. $DTES = 0.25 \times \text{Accuracy} + 0.25 \times \text{Personalisation} + 0.15 \times \text{Speed} + 0.20 \times \text{Integration} + 0.15 \times \text{Impact}$