International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 12-No.1 (2026) pp. 1151-1162
http://www.ijcesen.com

ISSN: 2149-9144

Research Article

Digital Twin Architecture for Therapy and Imaging

Shrikant Chikhalkar*

Independent Researcher, USA
* **Corresponding Author Email:** shrikant.v.chikhalkar@gmail.com - **ORCID:** 0000-0002-5247-5550

Article Info:

DOI: 10.22399/ijcesen.4982
Received : 29 December 2025
Revised : 20 February 2026
Accepted : 22 February 2026

Keywords

Digital Twins,
Computational Modeling,
Verification and Validation,
Risk Management,
Medical Device Development

Abstract:

Digital twins—computational representations of devices, environments, and patient physiology—offer transformative potential for medical device development by enabling simulation-based evaluation of scenarios that are difficult, dangerous, or impossible to reproduce physically. However, deployment in regulated environments demands rigorous credibility frameworks addressing verification, validation, and uncertainty quantification. This article proposes a comprehensive digital twin architecture tailored to regulated medical devices spanning therapy delivery and imaging workflows, emphasizing credibility as the foundation for trustworthy simulation evidence. The architecture decomposes platforms into core components, including device models capturing software logic and actuation behaviors, physics and environment models representing electromagnetic coupling and imaging physics, physiology models encompassing anatomy and tissue response dynamics, and data assimilation modules enabling patient-specific parameter tuning. Central infrastructure elements, including twin orchestrators, registries, and evidence layers, provide coordination, version control, and comprehensive documentation supporting reproducibility and regulatory compliance. The credibility framework establishes systematic protocols for code verification, solver verification, numerical stability assessment, and stratified validation across device settings, patient anatomies, and workflow variants, with explicit uncertainty quantification and traceability from risk controls to simulation evidence aligned with regulatory standards. Risk-informed applications enable hazard-to-scenario mapping, boundary condition exploration, fault injection testing, and combined fault mode analysis that expand verification coverage beyond traditional testing limitations. Controlled personalization frameworks define permissible parameter spaces and safety envelopes, treat patient-specific parameter estimation as verified algorithms rather than unbounded learning processes, implement privacy preservation through data minimization and secure governance, and establish guardrails preventing unsafe adaptations. Deployment considerations address computational infrastructure requirements, reproducibility protocols through containerization and numerical tolerance management, tool qualification processes for commercial and custom simulation codes and lifecycle management, ensuring version control, change management, and evidence archival throughout product lifecycles. The article concludes by examining broader development implications, including prototype reduction and shorter development cycles, alongside future directions encompassing standardized credibility reporting, shared physiological model libraries, and hybrid physics-data approaches integrating machine learning within physics-based constraints to maintain explainability and bounded behavior essential for regulatory acceptance.

1. Introduction and Problem Definition

However, the medical device industry is under pressure to increase the pace of innovation without sacrificing safety and regulatory compliance. Customary verification and validation methods (bench tests, animal tests, and clinical trials) have intrinsic shortcomings in cost, time, and coverage. Bench, animal, and clinical testing are limited. Bench testing cannot reproduce the full range of variability seen in patients. Animal testing can only predict human physiology in qualitative terms, and clinical studies are limited by the ethical and logistical issues related to sample size. Regulatory

policy has been and remains that mathematical models and simulations are to be considered as supportive evidence provided the model is well validated and falls within a defined context of use. The bottleneck for regulatory acceptance is confidence in the accuracy of computational predictions [1].

Digital twins—computational representations of the operation of physical devices, their environment, and the physiology of a patient—could be a natural and attractive complement to existing verification techniques. Well-validated digital twins could be used for end-to-end testing for combinations of rare physiological states, sequences of unusual workflows, multiple fault modes, and boundary conditions at the extremes of design specifications that can be difficult, dangerous, or impossible to explore in the physical world. Guidance also includes recommendations that modeling considerations be made in the context of the totality of evidence, model credibility be established through systematic approaches, and considerations of context of use, questions of interest, and impact on regulatory decisions be made. The applicable standard of scrutiny should be commensurate with device risk and dependence on computational modeling versus physical testing data [1].

However, the use of digital twins in a regulated capacity for medical device creation gives rise to the question of credibility. In contrast to exploratory simulations, regulatory submissions, design verification, or patient-specific treatment planning applications require sufficiently accurate and reliable predictions for the task at hand. Credibility requires verification and validation, uncertainty quantification, and traceability. Verification ensures that the model has been implemented correctly. Validation shows the model is a reasonable approximation of a target phenomenon in contexts where it is applied. Uncertainty quantification bounds model predictions. Traceability maintains a clear path from evidence from simulation to decision-making. Credibility requires establishing before simulation the contexts of use, including questions models answer, anatomical region and conditions, target population, and device under investigation. More recently, it has been shown that physics-based modeling approaches that describe first principles of device operation and physiological response have higher explanatory power and generalizability than data-driven approaches, which are less likely to extrapolate well outside their training data domains [2].

In this article, It describes an end-to-end digital twin architecture for regulated medical devices integrating the therapy and imaging workflow. Technical and governance considerations for verification and validation planning, uncertainty management, configuration control, and integration engineering are explored. Specific use cases are exploration of design margins, validation of control logic against expected physiological changes, evaluation of the robustness of imaging pipelines to motion and variability of protocols, fault tolerance against failures of various components, and bounded personalization in which patient-specific parameters are obtained from within the safety envelopes. The FDA recognizes that computational modeling can be used to show, among other things, substantial equivalence, safety and effectiveness, and changes in design. Depending on whether computational modeling substituted for clinical data, was performed in conjunction with clinical data, or would support a mechanism, specific criteria were set for evidence used to support these claims [1].

Controlled personalization should also be considered a high-potential, high-risk area, as many modern medical devices are becoming adaptive, whereby the treatments or imaging protocols can be personalized for individual patients. In terms of personalization, digital twins can support patient-specific parameter estimation, prediction of patient-specific response, and personalization of device parameters. Unbounded personalization of digital twins could potentially lead to the device behaving inappropriately outside of the range of parameters to which it has been tested, overfitting to noise, and optimization to unsafe operating points. The proposed architecture is enabled by a notion of control-theoretic personalization, which works in validated regions of parameter space, is guaranteed to work as an algorithm with defined inputs/outputs, and has explicit safety guardrails that prevent excursions outside of the validated operating envelopes. Recent work has called for multi-scale modeling that captures device operation and tissue-level physiological responses. Furthermore, this representation needs to be validated against clinical data and needs to be computationally efficient for clinical use [2].

The degree of confidence in the digital twin and the level of proof required will depend on the application, including the consequences of a potential error. Digital twins used in early-stage design explorations may not be as validated. Twins that support regulatory claims should have high levels of evidence, and those computing patient-specific treatment parameters should have the highest levels of validation. The context of use can involve what decisions the twins are used for, what performance measures and thresholds are relevant, how to define a failure mode, and how to build

confidence that the twin is fit for purpose. Regulatory guidelines state that credibility evidence should address whether computational models have been verified as implemented and validated with data sets representative of the intended context of use while characterizing and assessing uncertainty. Documentation should also address model assumptions, limitations, verification, validation, sources of data, comparison to external data, studies supporting validation, and uncertainty quantification methods that support the use of modeling as evidence regarding regulatory decisions to enable independent assessment of its credibility [1].

2. Digital Twin System Architecture and Engineering Integration

For a digital twin platform of medical devices to be viable, it is important that it have a modular architecture, which separates concerns, allows for independent verification, and enables the independent evolution of the structure of any individual component without compromising the integrity of the platform. Device models can capture the control logic of the software that drives the device, the timing, the sensors, the actuation, and the communications involved in the system. They may be simple abstractions or include more detailed hardware in the loop simulations. Complete state machine representations may be used to answer control logic questions, while simpler functional models may be useful for answering system-level risk questions. Digital twins were first developed for use in industrial contexts. Typical architecture includes physical entities, virtual models, data connectivity between the physical and virtual entities, and two kinds of twin data stores: past states and simulation results. The ability to connect, observe, simulate, and optimize these models results in meaningful value in terms of shortening development cycles and improving product quality [3].

For therapy delivery devices, these consist of mechanical interactions such as contact forces and tissue deformation, electromagnetic interactions such as RF energy deposition and magnetic field interactions, fluid dynamics of blood flow and drug dispersion, and thermal interactions such as tissue heating and heat dissipation. For imaging systems they consist of radiation transport, signal formation (for example, detector responses), imaging reconstruction algorithms, and artifacts such as motion blur and metal artifacts. Aspects of the environment, such as the operating room, EMI, mechanical vibrations, and workflows (including the use of the patient positioning systems and the HIS), will need to be considered. A key future implementation issue will be balancing computational performance with model accuracy. These trade-offs are seen in industrial applications, where model fidelity and hierarchical modeling approaches to vary the resolution by task are used. [3]

These models include the geometry of the body organs and vasculature and various properties such as tissue mechanical stiffness, electrical and physiological properties, tissue responses and immunological responses, and dynamic responses, such as cardiac motion and respiratory motion. However, all these properties are not identical for every person due to biological variability. Thus, rather than constructing a universal model, a library of representative models can be created for relevant ranges of anatomical variation, tissue property distribution, and physiologic state. This provides a population approach that can be used to evaluate device performance over expected patient variability, to identify worst-case scenarios, and to quantify sensitivity. In the cardiovascular area, a digital twin may include multiple modeling scales, from cellular electrophysiology to tissue mechanics to organ-level hemodynamics, and each of these scales presents different challenges along the axes of parameter identification, validation data availability, and computational cost [4].

They also include algorithms for estimating parameters from imaging biomarkers, algorithms for combining model predictions and sensor measurements in real time, and algorithms for updating baseline population models. One difference in regulated settings is that algorithms must be operated as controlled and validated algorithms, rather than open-ended learners. Parameter estimation must be performed within a bounding physical range. State estimation should include estimation of uncertainty bounds and trigger an alert if limits are exceeded. Model updating must be reversible and auditable. Multi-modality imaging combined with physiological measurements has been shown to be stronger for parameter estimation than single-modality imaging data but requires consideration of time synchronization, data reconciliation, and propagation of measurement errors or biases [4].

These interfaces allow engineers to define simulations without any knowledge of model detail. Separation of scenario definition and model implementation is desirable for maintenance of verification and to avoid unexpected outcomes on underlying model implementations. The orchestrator translates the scenario specifications and controls the simulation execution and the gathering of the results and all the simulation

provenance metadata. In industrial deployments, it has been found that the orchestrator needs advanced workflow capabilities such as automated dependency resolution, clever scheduling, load balancing, and fault tolerance [3].

The registry documents the versions of device models, physics solvers, libraries of physiology models, data assimilation algorithms, and parameter sets, and additionally records the validation and verification status, approved use cases, and known limitations of each component. It enforces compatibility rules to prevent them from being combined incorrectly and producing erroneous results; it also supports tracking the impact of changes to parts and complete traceability, which can be incorporated into audit trails (regulatory submissions) and compliance with quality systems. In practice, configuration management has proven vital for industrial deployments, where the number of model variants, parameters, and simulation scenarios can quickly grow without version control [3].

The evidence layer captures input specifications, assumptions and simplifications, solver parameters, intermediate results, final results, and metadata for each execution in order to support reproducibility, verification (by capturing a detailed record of execution), validation (by capturing an exact record of simulation conditions), and regulatory evidence (by providing complete provenance). For cardiovascular models, this extends beyond signal processing to multi-scale, multi-physics simulation where electrophysiological models couple to mechanical deformation and hemodynamic flow with different spatial and temporal resolution requirements [4].

The architecture should allow simulation to be embedded in the continuous integration pipeline, such that regression tests run, simulation campaigns are automatically triggered, parameter sweeps are executed to explore design margins each night or weekly, and risk coverage reports are automatically generated, linking scenarios to hazard analyses. Automation of scripting should be supported via APIs, and standard reporting formats should be defined. The industrial literature reports that PLM integration can reduce time to market and improve quality, at the cost of considerable upfront investment in automation infrastructure and organizational change management. A document control process is critical in maintaining the ability to reproduce simulations in the future if post-market surveillance or incident analysis is required. The development of cardiovascular digital twins has exposed challenges with the simulation workflow, including the expense of high-fidelity models, maintaining libraries of reduced-order models, and directional burst capacity through cloud computing [4].

3. Credibility Framework: Verification, Validation, and Uncertainty Quantification

Where digital twins are used in regulated domains, a corresponding process of verification, validation, and uncertainty quantification (VVUQ) is necessary to establish their credibility. Verification focuses on whether models are implemented correctly, whether the software implements the intended mathematical model, whether a numerical solver produces the correct solution, and whether the coupled systems behave correctly. Validation asserts that models predict the reality. Not to be confused with physical accuracy, verification applies software engineering best practices. Besides code review, unit testing with test cases and expected outputs, integration testing to see if the data flows correctly, and regression testing to find unexpected changes, formal verification for complex simulation codes extends customary testing to include static analysis, symbolic execution, and model checking. Formal verification can determine if the code for each component meets its specification, if the requirements trace back to the code, and if implementations pass verification tests. Current regulations state that simulated evidence as a partial replacement for physical testing can also be accepted as long as the proper levels of credibility are established [5].

Solver verification is a property of the mathematical correctness of numerical methods. Approximation errors come from discretization, iterative solution methods, and finite precision arithmetic and need to be controlled. Solver verification of a numerical method consists of showing that the approximation obtained by the method converges to the true solution as the size of the discretization or the tolerance in the solver decreases. One approach is the method of manufactured solutions, which constructs analytical solutions to modified problems. Grid refinement studies test whether calculations converge to a solution as the mesh or timestep size is decreased and yields quantitative acceptance criteria. Verification benchmarks check whether the solution produced by simulation codes matches a known analytical solution, a converged numerical solution, or a solution from a different code. In verification studies, isogeometric analysis methods are helpful since they can represent the geometric domain exactly without being affected by geometric approximation errors as in customary finite element methods [6].Meanwhile, numerical stability and sensitivity analysis take the response

of the solution to small changes to indicate reliability and robustness: for verified solvers, the output remains stable when the input is perturbed, except through physical sensitivity to fluctuations in the input. Large sensitivity to changes in parameters can indicate numerical instability, ill-posedness, or bifurcations; low sensitivity may indicate modeling errors, incorrect parameter values, or numerical damping. In automated sensitivity analysis, input parameters are perturbed and output values are computed based on derivative- or sampling-based methods. Output acceptance criteria, representing expected physical sensitivity, can indicate areas needing further scrutiny. The elimination of per-element geometry errors makes isogeometric analysis particularly attractive for thin structures, surface-dependent physics problems, and contact mechanics, where spatial errors in the geometry can jeopardize the accuracy and stability of the solution [6].

Validation deals with the physical correctness of the simulation results. A simulation result is validated by comparing it against measurements of the respective phenomenon. Validation is always context-dependent: a model that is validated for one purpose may not be valid for another purpose, and evidence is always presented with respect to the intended use. Validation planning generally begins with specification of validation metrics that describe the agreement of simulations and experiments, with mean absolute error and correlation coefficient often used for continuous observables. Confusion matrices and ROC curves may be useful for evaluating categorical outputs. Validation planning identifies data sources and acceptance criteria, defining what constitutes adequate validation for the intended application. The framework stresses that the amount of validation evidence should be proportional to the degree a model's results impact a regulatory action [5].

The validation data sets vary widely in nature: benchtop tissue phantoms provide tightly controlled conditions in which some modeling components can be studied but also reduce clinical realism. Animal models allow biological variation and anatomy to be investigated under experimental control, and the model species will determine the extent to which the results can be transferred to human anatomy. In clinical databases, the patient population is investigated under real conditions; however, experimental control and ground truth measurements are absent. This hierarchy of information encourages the use of multiple, complementary data sources, from controlled mechanistic validation to real-world outcome prediction, indicative of translational readiness [5].

Validation should be stratified across clinically relevant ranges of the device settings, the patient characteristics, and the environmental and workflow characteristics so as to cover the intended use space. For example, for therapy delivery devices, validation could be stratified across ranges of power settings, tissue type, patient anatomy, and procedure workflow. In imaging systems, stratification factors can include scan protocols, patient motion, patient size, and reconstruction algorithms. Risk analysis, sensitivity analysis, and practical considerations should guide the choice of stratification factors to use in a validation plan. Statistical power analysis also decides the number of validation cases to be used per stratum. When dealing with the validation of techniques involved with complex patient-specific geometries, isogeometric analysis techniques are preferable, as reconstructed CAD models from imaging data do not suffer from geometric approximation errors [6].

Explicit uncertainty quantification recognizes that irreducible uncertainties are always present in a computational model, which can only be characterized but cannot be reduced to a single value. Irreducible uncertainties can include parameter uncertainty, model-form uncertainty, numerical uncertainty, input uncertainty, and scenario uncertainty. For credible digital twins, uncertainties need to be characterized explicitly and propagated through the digital twin, and decisions need to be made on the effect of uncertainties. These uncertainty quantification methods can be sensitivity analysis, interval analysis, or probabilistic uncertainty propagation based on Monte Carlo simulation. The choice also depends on the information available about input uncertainties, the computational cost, and the requirements of the decision framework. Uncertainty quantification is identified as important credibility evidence in regulatory guidance [5].

Risk controls, whether the verification or, less frequently, the validation kind, can be linked to simulation evidence through traceability. For example, this can be achieved by linking hazards to risk controls and linking those in turn to verification evidence and the validation evidence required to meet regulatory and quality system requirements. Digital twin platforms should support this through the hazard analysis table, simulation scenarios, quantification of residual risk, and validation studies that show prediction performance. It is important that adequate documentation supports traceability both ways, forward from the hazard analysis table to the evidence and back from the result to the risks. Considering the standards to which credibility activities conform helps to ensure regulatory

compliance and conformity to the industry's best practice. There are further theoretical advantages of isogeometric analysis approaches that could improve the credibility of a computational model by reducing the contribution of approximation error and uncertainty [6].

4. Risk-Informed Applications and Controlled Personalization

Digital twins are also being linked with risk management to systematically improve verification coverage and safety arguments. The International Organization for Standardization standard requires systematic hazard identification, risk estimation, risk controls, and verification that the risk controls are effective. Digital twins allow scenarios to be explored that may not be safe, possible, or cost-effective to apply to the real system. Digital twin simulation eases assessment of residual risk after control application, representing the entire range of operational modes, patient variations, and environmental scenarios, while establishing a risk-based scenario library that provides traceability between identified hazards and scenarios for hazard evaluation and control verification. For each hazard, the scenario library captures the physical mechanism for the hazard, the scenarios of test case runs where the hazard is modeled in simulation, the metrics, and the acceptance criteria identifying when the hazard is realized. All hazards should have been analyzed with simulation, physical tests, or documented justification in the library. Risk management is an iterative process throughout the device's life cycle [8].

Boundary condition exploration utilizing a digital twin can systematically explore the performance of a device at the corners of its operating envelope where validation is relatively sparse. Medical devices specify ranges of adjustable parameters, patient-related characteristics, environmental conditions, and use scenarios where the medical device claims to be safe and effective. These nominal conditions are well validated, but corner cases of multiple factors reaching boundary conditions simultaneously have not been systematically validated. Digital twins can support the design of experiments, parameter sweeps, corner case analysis, and Latin hypercube sampling. The analysis finds the worst-case scenario and identifies safety margins. This technique verifies whether the chosen operating ranges are sufficient or overestimate the risk, and it informs the distributed positioning of test points. The risk management standard states risk estimation should include the full range of conditions of intended use and foreseeable misuse scenarios [8].

Degradation scenarios entail studying how a system behaves when its components' performance is below full due to physical reasons such as wear, aging, or partial functional failures. Digital twins can be deployed to simulate degradation scenarios by modeling components with less than full performance capability. Particularly, graceful degradation paths, cliff-edge failures, compensating failures, failure combinations, component specifications, and system-level indicators, including whether safety interlocks for degraded components work as intended. The risk management framework used for such analysis also examines device aging, wear, and the effect of environmental exposure on safety over the intended design lifetime [8].

Combined fault mode analysis considers the interactions between multiple faults. Combinations of faults may not be precluded by any single redundancy or fault detection scheme and may therefore bypass protections against single faults. Because combined fault scenarios are too complex for physical testing, digital twins have been used for systematic enumeration (with different fault combinations sharing the same common cause mechanism, random sampling, and building up scenarios in detail). As well as these, combined fault analysis may be applied to situations in which cybersecurity vulnerabilities are simultaneously exploited, in conjunction with component failures, to create dangerous system states [7].

Safety interlock verification tests ensure that they will detect and avoid the unsafe state. Physical testing can be an implicit hazard and is often impractical. For digital twins, tests can be run in simulation across the full range of conditions to be avoided and those not to cause interlocks, to test threshold values, timing, noise immunity, ability to evade, and whether the expected protective reaction is triggered. This is required, as interlocks are last resorts: a positive verification is that the interlock fires, a negative verification is that it does not, and a fault tolerance verification is that it functions with a degraded part [8].

Controlled personalization permits helpful device personalization for an individual patient given preconditions and verification, motivated by patient performance and characteristics. Digital twins could be used for personalization; however, unrestricted personalization can create unstable behavior outside of the parameter range, overfitting with respect to the noise present in the medical data, optimization towards unsuitable operating points, and neglect of systematic validation. Reducing these risks can be achieved through

ensuring that personalization is performed within safety envelopes (clearly delineated parameters), that the algorithm itself can be verified with defined inputs and outputs, and that the entire product has been made fully transparent. Cybersecurity is often raised as a concern for AI/ML-powered devices, as they can be an attack vector [7].

Permissible parameter spaces are the multidimensional spaces in which patient-specific device parameters can vary while ensuring that the device is safe and effective. These boundaries may be defined by validation space, safety analysis, physical constraints, and regulatory spaces. Boundaries can be set either analytically, empirically, or as hybrids, and conservatively defined zones of the parameter space should be specified on each side of the boundary. Documentation should indicate the basis of the evidence for each boundary, integrated margin, and update procedure. Patient-specific parameter estimation algorithms should not select parameters outside their defined spaces with hard constraints. Thus, to include this requirement in the risk management framework, the risk controls must be implemented by means of software mechanisms that are impossible to bypass [8].

Patient-specific parameter estimation should be implemented, verified, and documented as a deterministic algorithm (with defined inputs, outputs, and failure modes). Steps should provide clinical input(s), define computational procedures to generate patient-specific model parameters, and outline acceptance criteria. Verification also includes testing the algorithm over input data, assessing its response to input quality and the acceptability of quality descriptors, testing derived parameters against experimental observations, and documenting its limitations. The algorithm should be able to check the quality of input and parameter confidence automatically. For devices including personalization functions to increase the security, the algorithm can implement security measures for input validation, bounds checking, and parameter estimation logging [7].

Privacy protection is critical as data-driven models are developed. Principles of data minimization are applied, whereby no more patient data than needed is used. Technical details include removing anatomy not related to the model, extracting high-level statistical parameters, and decoupling identifiable demographics from the parameters of the model. Security mechanisms ensure that data are only stored and transmitted securely (e.g., encryption of data at rest and in transit, access rights management, and physical security), while governance mechanisms specify the rules and regulations of what patient data can be gathered and recovered by whom, for how long it is kept, and how and when it is deleted. The technical architecture should support the separation of functions in distinct modules for patient data handling. De-identification may be applied in situations when identification of persons is not necessary for a function. Other cybersecurity controls include limiting access to protected information, protection against malware, applying software patches, and checking for anomalous data access patterns [7].

Guardrails preventing unbounded adaptation consist of manifold layers of protection ensuring that adaptation does not exceed safe, validated boundaries. Hard constraints are ultimate parameter value boundaries that are enforced with certainty. Soft constraints flag parameter values relevant to the boundary, which may require confirmation by the clinician. Patient-specific models are removed from consideration if parameter uncertainty exceeds clinician-acceptable boundaries. Override mechanisms enabling clinicians to use population-average models or manually selected parameters when estimation fails should, as a minimum, be stress tested using fault injection studies to show that they are invoked to prevent use of an unsafe configuration. Where risk controls depend on the software to perform certain functionalities, the risk management standard specifies that the software lifecycle processes are used to develop and verify the functionalities. Cybersecurity controls shall not allow guardrails to be bypassed via a vulnerability, unauthorized change to configuration, or a malicious software update [8].

5. Deployment Considerations, Lifecycle Management, and Future Directions

Successful adoption of digital twin platforms in medical device companies requires consideration of practical aspects of digital twin implementation, including computing infrastructure, reproducibility, tool qualification, and lifecycle management. Computing infrastructure requirements depend heavily on the complexity and frequency of the simulations. Device control logic verification simulations with simple physics may run in seconds on an office workstation, while high-fidelity multiphysics simulations of complex therapy delivery systems and advanced imaging devices may run for hours or days on high-performance computing clusters with hundreds or thousands of processor cores. The compute infrastructure may be on-premise for maximum control over the computing environment and data privacy, but at the cost of large capital investment; cloud computing for elastic scaling and specialized compute

hardware, but at the cost of data residency; or hybrid infrastructure with on-premise compute resources for routine simulation and cloud bursting for parameter sweeps or high-priority in-depth investigations. Infrastructure-related aspects are computational resources, data storage for inputs and outputs of the simulations that may grow over the life of the device, networks to manage large data transfers, visualization systems for multidimensional data, and backup to avoid loss of simulation products. Standards for software validation also state that the software's credibility and the rigor of the software validation process should be consistent with the level of concern for the software and in line with the confidence the software is used with. For example, software supporting critical device functions or influencing important regulatory decisions will require more validation than software supporting ancillary functions of low safety importance.

The reproducibility requirements for simulations are stricter than for exploratory studies because simulation results are used in regulatory submissions or when the claim is about the design or safety of the product developed from the simulation. If a simulation result is challenged long after the original research is complete, independent reviewers should be able to reproduce the result or assert the claim is true. This can be achieved through controlled computational environments where all software, including operating systems, compilers, numerical libraries, codes for simulation, and settings for solvers, are captured and 'frozen' at specific software versions. Container technologies like Docker and Singularity can export images that contain complete computational environments, which run consistently on different machines and over time. Numerical tolerance management addresses differences in calculations between computer architectures and compilers. Techniques include specifying a difference for numerical results to be treated as equal, documenting what results are expected to be reproducible, executing tests to assure that repeated decisions within the tolerance band are stable, and regression testing by comparing values across configurations against a reference. All software lifecycle processes of medical device software (including software maintenance) must identify the software items, control changes, and maintain the software version data, which is necessary for reconstructing a released configuration [10].

Tool qualification determines whether the tool has technical data sufficient to claim it has been properly verified and validated for the intended use in regulatory submission. The steps include reviewing vendor verification and validation, delimiting tool capability covered by verification and validation, confirming that the intended use is covered, and validating the tool where necessary. Qualification often involves verifying code, validating it against experimental data, documenting the software development processes employed, and often independent code review. Simulations with code-based tools may also be qualified against the level of reliance on the simulations' results and the severity of the consequences that would be associated with software faults. The highest qualification level is for simulations replacing clinical tests or for pivotal safety claims. The lowest qualification level is for discovery, screening, or preclinical tools, including recommendations for appropriate application and limitations of use. Typical software validation activities include requirements definition, test design, test execution under controlled conditions, and the evaluation of test outcomes to ensure that all requirements have been met [9].

As medical device designs or patient populations change, new evidence is generated, or new computational methods are developed, the version of a DT platform needs to be version controlled and must maintain thorough traceability from any simulation result back to the specific versions of all models, solvers, and parameters that were used in that simulation. Version control policies specify when to create a new version and when modifications should occur. Compatibility management is the prevention of combining particular components with particular versions outside of fully tested configurations defined in compatibility matrices. Change control governs the updating of validated components. An impact analysis is performed before the change to determine what existing validation evidence is still valid and what new validation evidence needs to be created. Regression testing ensures that previously validated scenarios still produce expected results. Evidence archival policies ensure that results from a simulation platform used as evidence for a design decision or regulatory claim are accessible with all their detail even as the platform evolves. Policies for retaining such evidence should specify a retention period for types of simulation evidence. This is typically decades after a product has been decommissioned. Corrective, adaptive, and perfective maintenance activities should be carried out in accordance with documented procedures, and the identification, approval, implementation, verification, and documentation of changes to the maintained software are treated with the same rigor as for development. [10]

Further potential project-level benefits of effective digital twin platforms include reduced time to

market through timely identification and resolution of design issues, reduced need for prototyping through greater simulation-based verification, improved risk management through simulation of scenarios impractical to assess physically, and better-regarded regulatory submissions with a richer body of evidence. Delivery of such platforms requires sustained commitment and investment in software and hardware infrastructure. personnel and simulation skills by engineering teams, and effective governance across the different uses of their digital twins to build credibility and foster appropriate use. The software validation guidance is clear that it is a process with planned, systematic, and documented activities during all phases of the software life cycle, not a one-time event [9].

Future digital twin technologies may also benefit from standardized reporting of credibility, which could enable regulators to easily assess validation and limitations of a digital twin and enable assessment and comparison of models built with different simulation methods. Standardization of libraries of validated physiological models could reduce duplication and ease evidence sharing, although intellectual property, data privacy, technical interoperability, and governance issues would need to be addressed first. Hybrid digital twins using physics-based sub-models and constrained data-driven models represent a promising approach that supports the use of machine learning with the interpretability of physics. Physics-based sub-models provide explainability and generalization; constrained data sub-models can represent complex non-linearity and emergent behavior without full mechanistic understanding. Hybrid models include physics-informed neural networks where the governing equations are incorporated in the loss function, reduced-order models where the accurate physics simulation is used to train a fast surrogate model across regimes, and ensemble methods where physics is combined with a data-driven model corrector. A major barrier to regulatory acceptance is that data-driven models should not compromise explainability or induce pathological behavior. Novel validation methods are required to guarantee robustness, stability, and bounded error on the use case [10].

Ultimately, it has been shown how digital twins can be integrated into the medical device development lifecycle as validated, trustworthy tools to support safety if the digital twin architecture is specified to include context-of-use definition, scenario coverage traceability, verification and validation, uncertainty management, and engineering integration. The digital twin architecture proposed here covers technical modeling capability as well as verification methodology governance, validation planning, uncertainty quantification, configuration management, and lifecycle management. Implementation of design verification and appropriately limited personalization has already been made possible where patient safety was the dominant consideration in developing, deploying, and maintaining a platform. To realize the full potential of digital twins, device manufacturers, regulators, SDOs, and researchers will need to jointly develop the methodology, evidentiary expectations, and organizational competencies needed for using digital twins to accelerate the delivery of safer and more innovative medical devices without compromising safety and effectiveness for the benefit of patients.

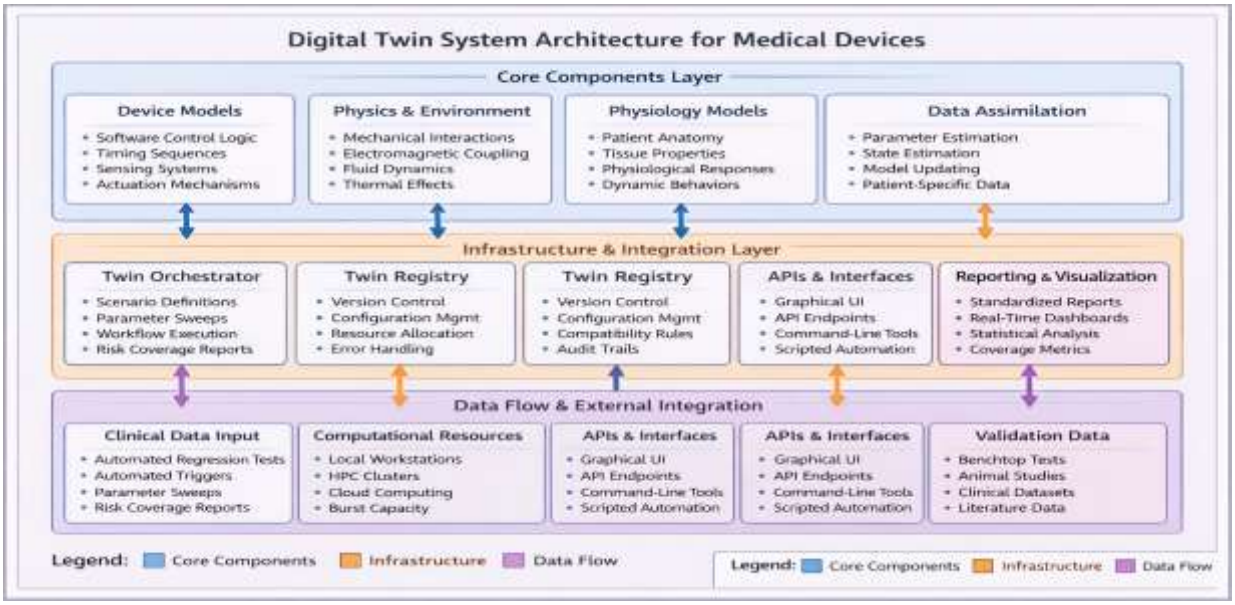

Figure 1: *Digital Twin System Architecture for Medical Device Development and Verification. [3, 4]*

Table 1: *Verification and Validation Hierarchy for Digital Twin Credibility. [5, 6]*

Credibility Component	Methods and Techniques	Acceptance Criteria and Outputs
Code and Solver Verification	Structured code reviews, unit testing with known outputs, integration testing, regression testing, method of manufactured solutions, grid convergence studies, verification benchmarks against analytical solutions	Documented evidence of specification compliance, convergence at predicted rates (first-order: linear, second-order: quadratic), quantitative error metrics, traceability from requirements to verification tests
Numerical Stability and Sensitivity Analysis	Automated sensitivity analysis through derivative-based methods (adjoint sensitivity, finite-difference) or sampling-based methods (Morris screening, Sobol indices); perturbation studies examining solution response to input variations	Stable results with output changes proportional to physical sensitivity, detection of numerical instability or ill-posed problems, acceptance criteria based on expected physical sensitivities, investigation triggers for significant deviations
Validation Data Sources	Benchtop experiments (tissue phantoms, controlled conditions), animal studies (biological variability, realistic anatomy), retrospective clinical datasets (real-world evidence, actual patient populations)	Context-specific validation metrics (mean absolute error, correlation coefficients for continuous quantities; confusion matrices, ROC curves for categorical outcomes), hierarchical evidence spanning mechanistic validation to clinical outcome prediction

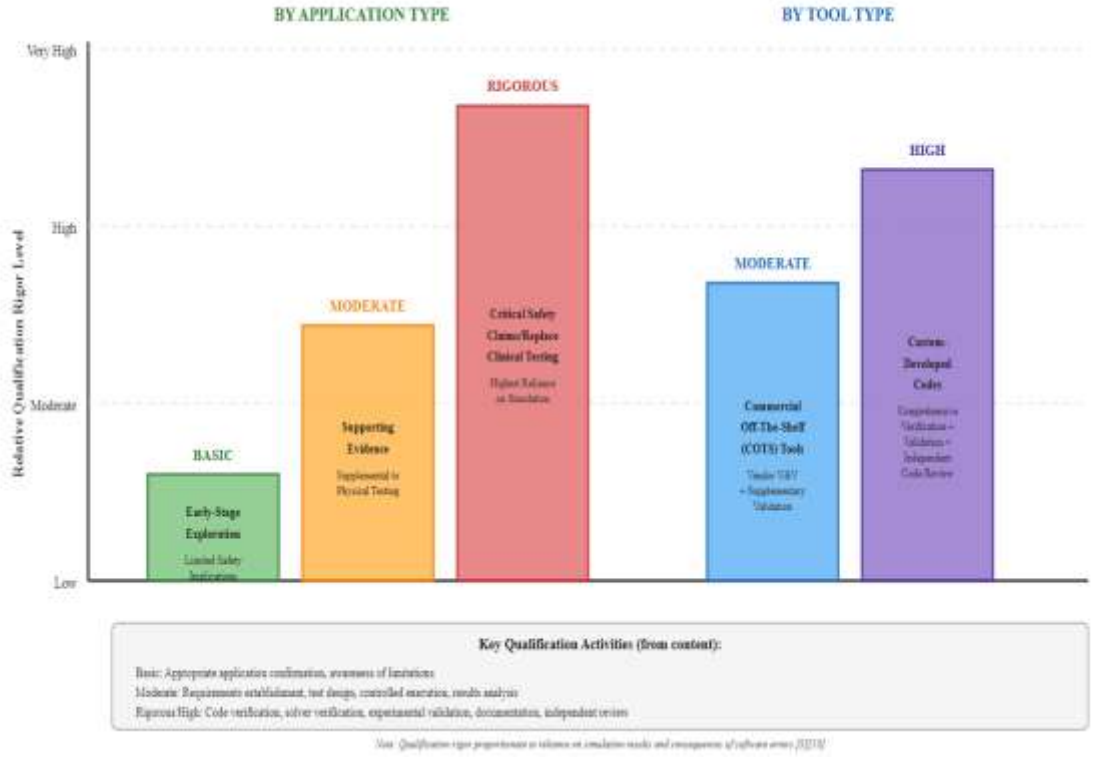

Figure 2: *Risk-Based Tool Qualification Framework for Digital Twin Platforms Qualification Rigor Levels by Application Type and Tool Category*

Table 2: *Risk-Informed Digital Twin Applications and Verification Methods. [7, 8]*

Application Type	Digital Twin Capabilities	Verification Outputs
Boundary Condition Exploration	Design of experiments methods, parameter sweeps, corner case analyses, Latin hypercube sampling for systematic evaluation at operating	Worst-case scenario identification, safety margin characterization, labeled operating range verification, intelligent physical test point placement

	envelope edges	
Component Degradation Scenarios	Virtual reliability testing through modified component models representing reduced performance characteristics (sensor noise, actuator wear, processor throttling)	Graceful degradation paths, cliff-edge failure identification, compensatory mechanism discovery, component specification refinement, system-level monitor requirements
Combined Fault Mode Analysis	Safe computational exploration through systematic enumeration, common-cause prioritization, random sampling, scenario building for multiple concurrent faults	Safety architecture effectiveness verification, redundancy scheme validation, fault detection algorithm testing, cybersecurity vulnerability assessment

6. Conclusions

Digital twins can become practical, safety-enhancing components of regulated medical device development when engineered for credibility through explicit context-of-use specification defining what decisions twins will support, traceable scenario coverage mapping from risk analysis through simulation scenarios to evidence, rigorous verification and validation processes establishing computational correctness and physical accuracy, transparent uncertainty handling quantifying confidence in predictions, and controlled integration with engineering workflows ensuring appropriate use. The architecture proposed in this article provides a comprehensive framework addressing both technical modeling capabilities, including device, physics, and physiology models with appropriate fidelity for intended applications, and essential governance elements, including verification protocols, validation planning, uncertainty quantification, configuration management, and lifecycle controls that collectively enable trustworthy simulation-based evidence generation. Both design verification, enabling exploration of design margins, fault scenarios, and boundary conditions beyond the practical scope of physical testing, and carefully bounded personalization, adapting device operation to individual patients while maintaining safety through defined parameter spaces and verified algorithms, become achievable when supported by well-architected platforms maintaining patient safety as the primary objective throughout development, deployment, and lifecycle management. Realizing the full potential of digital twins requires continued collaboration among device manufacturers developing and deploying twin platforms, regulators establishing evidence standards and review processes, standards organizations developing consensus frameworks and best practices, and researchers advancing modeling methods, validation approaches, and uncertainty quantification techniques, collectively building the methodological foundations, evidentiary standards, and organizational capabilities enabling digital twins to deliver their promise of safer, more innovative medical devices developed more efficiently while maintaining the rigorous safety standards essential for patient protection and regulatory compliance in an evolving healthcare technology landscape.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.
- **Use of AI Tools:** The author(s) declare that no generative AI or AI-assisted technologies were used in the writing process of this manuscript.

References

[1] U.S. Department of Health and Human Services Food and Drug Administration Center for Devices and Radiological Health, "Assessing the Credibility of Computational Modeling and Simulation in Medical Device Submissions," 2021. [Online].

Available: https://www.fda.gov/media/154985/download

[2] Alessandra Aldieri et al., "Credibility assessment of computational models according to ASME V&V40: Application to the Bologna Biomechanical Computed Tomography solution," Science Direct, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169260723003930

[3] Fei Tao et al., "Digital Twin in Industry: State-of-the-Art," ResearchGate, 2019. [Online]. Available: https://www.researchgate.net/publication/345078627_Digital_Twin_in_Industry_State-of-the-Art

[4] Jorge Corral-Acero et al., "The 'Digital Twin' to enable the vision of precision cardiology," Eur Heart J, 2020. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/32128588/

[5] Tina M. Morrison et al., "Assessing Computational Model Credibility Using a Risk-Based Framework: Application to Hemolysis in Centrifugal Blood Pumps," ASAIO J. 2019. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC6493688/

[6] T.J.R. Hughes et al., "Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement," ScienceDirect, 2005. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0045782504005171

[7] U.S. Food and Drug Administration, "Cybersecurity in Medical Devices: Quality Management System Considerations and Content of Premarket Submissions," 2026 [Online]. Available: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/cybersecurity-medical-devices-quality-management-system-considerations-and-content-premarket

[8] International Organization for Standardization, "Medical devices—Application of risk management to medical devices," 2019. [Online]. Available: https://www.kmedhealth.com/wp-content/uploads/2024/03/EN-ISO-14971-2019-Application-of-risk-management.pdf

[9] U.S. Food and Drug Administration, "General Principles of Software Validation; Final Guidance for Industry and FDA Staff," 2002. [Online]. Available: https://www.fda.gov/media/73141/download

[10] Małgorzata Kruszynska, "IEC 62304:2006 – software life cycle processes explained," Spyrosoft, 2021. [Online]. Available: https://spyro-soft.com/blog/healthcare/iec-623042006-software-life-cycle-processes-explained