



Digital Twins for Radiation Oncology

P. James Jensen
Yale School of Medicine
james.jensen@yale.edu

Jun Deng
Yale School of Medicine
jun.deng@yale.edu

ABSTRACT

Digital twin technology has revolutionized the state-of-the-art practice in many industries, and digital twins have a natural application to modeling cancer patients. By simulating patients at a more fundamental level than conventional machine learning models, digital twins can provide unique insights by predicting each patient's outcome trajectory. This has numerous associated benefits, including patient-specific clinical decision-making support and the potential for large-scale virtual clinical trials. Historically, it has not been feasible to use digital twin technology to model cancer patients because of the large number of variables that impact each patient's outcome trajectory, including genotypic, phenotypic, social, and environmental factors. However, the path to digital twins in radiation oncology is becoming possible due to recent progress, such as multiscale modeling techniques that estimate patient-specific cellular, molecular, and histological distributions, and modern cryptographic techniques that enable secure and efficient centralization of patient data across multiple institutions. With these and other future scientific advances, digital twins for radiation oncology will likely become feasible. This work discusses the likely generalized architecture of patient-specific digital twins and digital twin networks, as well as the benefits, existing barriers, and potential gateways to the application of digital twin technology in radiation oncology.

CCS CONCEPTS

• :: • **Applied computing** → Life and medical sciences; Health care information systems; Life and medical sciences; Systems biology; Life and medical sciences; Computational biology; Imaging; Life and medical sciences; Computational biology; Biological networks;

KEYWORDS

Digital Twins, Radiation Oncology, Predictive Medicine, Artificial Intelligence

ACM Reference Format:

P. James Jensen and Jun Deng. 2023. Digital Twins for Radiation Oncology. In *Companion Proceedings of the ACM Web Conference 2023 (WWW '23 Companion)*, April 30–May 04, 2023, Austin, TX, USA. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3543873.3587688>

1 INTRODUCTION

Digital twin technology has recently innovated many industries, such as transportation engineering (aerospace [1], automobile [2],

and rail [3]), architectural design [4], and workplace safety [5]. Digital twins are aptly-named - they are virtual representations of a complicated system's components and environment. In each industry, digital twins are used to robustly predict the state and performance of the modeled system. For example, digital twins can predict the fuel efficiency of a jet design, or the hazardousness of a particular assembly line configuration in a warehouse.

Digital twin technology has a natural application to biomedical science. Human beings are immensely complex, and this complexity reduces the accuracy of statistical models that focus solely on a subset of this complexity, such as macroscopic/systemic data or histological data. As a result, it is difficult to predict the progression of certain types of diseases despite medicine's numerous scientific advances, and the outcomes of existing treatment options can be uncertain. Digital twins have the potential to provide superior models to existing statistical or machine learning models by more closely modeling the actual complexity of humans.

This is particularly true in radiation oncology. Cancer patients are a wide cohort with great interpatient variability, and there are numerous subtle differentiating factors between patients that strongly impact clinical outcomes. Current clinical decision-making does not completely account for this variability. As a result, many cancer patients experience unexpected responses to radiation, such as cancer recurrence and toxic side effects. To combat this challenging class of diseases, radiation oncology has become a field of precision medicine, where each individual patient's treatment course is determined by patient-specific features [6]. In radiation oncology, the oncologist's team uses patient-specific data to craft a custom-tailored radiation treatment plan designed to avoid critical organs-at-risk and maximally target the patient's cancer. The treatment planners use demographic data, histological biopsies, and three-dimensional tomographic imaging (such as CT, PET, and MR imaging) to create an optimal plan based on their experience. However, this data alone is not enough to precisely determine the results of treatment, so clinical decisions are more strongly influenced by the physician's personal experience, which introduces bias into the patient's treatment according to the physician's particular style. This ultimately contributes to the unpredictable treatment results in radiation oncology, decreasing cure rates and increasing toxic side effect rates.

Human digital twins could change all of this. Digital twins excel at modeling complex systems, so they are well-suited to model patients by incorporating all the subtle features that cannot be perfectly considered by another human. More importantly, human digital twins could enable a new paradigm of predictive medicine, where many clinical options are virtually explored to determine the optimal treatment course. In the paradigm of predictive medicine, human digital twins would go beyond the raw data by virtually simulating each potential treatment course's trajectory, producing a range of predicted treatment results that thoroughly represent the



This work is licensed under a Creative Commons Attribution International 4.0 License.

WWW '23 Companion, April 30–May 04, 2023, Austin, TX, USA
© 2023 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9419-2/23/04.
<https://doi.org/10.1145/3543873.3587688>

range of possible trade-offs in treatment planning. Equipped with this information, radiation oncologists can select the most optimal plan based on their experience, and they can also communicate with patients to determine the most appropriate trade-off between cure rate and toxicity rate based on patient preferences.

The potential to model human beings with digital twins has been suggested in broader categories, but no implementations have come to fruition yet. This is due to the large amount of groundwork that needs to be completed before generalized human digital twins can become a reality. However, human digital twins become much more feasible when restricted specifically to radiation oncology. We hypothesize that it is possible to create digital twins for cancer patients (DT-CPs) that can be implemented in current clinical practice. Currently, DT-CPs do not exist, and the field of medical physics is far off from the creation of DT-CPs. To address the current state of DT-CPs, this article presents a theoretical framework for DT-CPs, investigates the potential benefits of DT-CPs in clinical decision-making, discusses the existing barriers that prevent the implementation of DT-CPs, and proposes several solutions to these existing barriers.

2 NETWORK SYSTEMS OF DIGITAL TWINS

Although the details may vary, the eventual implementation of a digital twin network system is likely to resemble the network shown in Figure 1. In this network architecture, a central database is maintained, either by the consumer hospital system or by a central organization. Each time a new patient is to be added to the system, the central database enables the creation of that digital twin by providing models that are applied to that patient's data. Once created, the patient's digital twin predicts the expected patient health trajectories that result from a range of possible clinical decisions. For example, a physician could use the digital twin to explore what would happen if a lung cancer patient received only radiation therapy from a range of possible fractionations, or if the patient received radiation plus chemotherapy from a range of possible pharmaceuticals and doses, etc.

With this knowledge, the physician becomes much better equipped to make the best clinical decision for each patient. Later, the patient's progress is recorded and used to update the digital twin, which improves the accuracy of subsequent predictions for that patient. When the twin is updated, the universal database is also updated by the digital twin's data to improve the models of all digital twins in the system. This allows each digital twin in the database to learn from each other's baselines and progress. Figure 1 shows the cyclic flow of this network architecture, where each digital twin and the database are all constantly improving as more experience accrues. When there are many digital twins connected to the universal database, the database's models become much stronger. Therefore, it is important to facilitate inter-institutional data sharing to maximize the quality of all connected digital twins.

3 DIGITAL TWIN COMPONENTS

3.1 Patient Data

Patient data is the central component of a digital twin. The data needed to establish an accurate digital twin will at least include patient demographic and summarizing data, which can be used to

fill in any other data gaps with demographic-specific averages. Digital twins will also require three-dimensional tomographic imaging (CT, MR, etc.) to provide spatial data about the patient's tissues and anatomy. There will also be a need to transition from this physical spatial data (such as x-ray attenuation in CT images, magnetic spin density in MR images, etc.) to biological spatial data (such as tissue density, oxygenation values, histological profiling, etc.). Recent advances in multiscale modeling are currently exploring this possibility, and multiscale modeling is likely to advance in the future [7]. Digital twins may also incorporate genomic data, which can have a strong impact on treatment outcomes. In the future, genomic data may be used to augment the capabilities of multiscale modeling to provide better spatial tissue stratification.

3.2 Biophysical and Machine-Learned Modeling

Digital twin predictions will be enabled by a fusion of biophysical modeling and machine learning. This modeling approach can incorporate analytically derived models for microbiological processes, such as proliferation, cell death, cell repair, angiogenesis, mutation, and immune system responses, with hyperparameters for these models determined by modern deep learning methods. The potential to use this combination inference strategy has recently been demonstrated [8–11]. Once equipped with patient data, a digital twin can use these models to move forward in time to predict the results of treatment decisions. By directly simulating patient progress rather than predicting results indirectly, this strategy is likely to provide models with improved predictive performance. Importantly, digital twin simulations can enable Monte-Carlo-like probabilistic simulation, where many predictions are made to sample the probability distribution of possible treatment outcomes.

3.3 High-Performance Computing

The computing power required to sustain a digital twin is likely to be very high. As explained above, digital twins for radiation oncology will require patient data with dense sampling across many patients, and the sheer volume of this data alone will be computationally intensive to process. Recent research has demonstrated the impact of this volume of data, as well as the need for high-performance computing to manage it [12]. This data will be even larger when accounting for the model-specific parameters that are generated, as well as the results created when propagating the patient data forward in time repeatedly during treatment optimization or virtual clinical trials (see sections 5.1 and 5.2 below). For the universal database (see section 3), this data is multiplied by the number of patients. Beyond the data processing requirements, model training and validation for both the multiscale models and machine learning involve intense algorithmic complexity. To address these requirements, high-performance computing will be needed. This computing power can be centralized along with the universal database; therefore, individual institutions will not be required to provide their own computing power.

4 THE BENEFITS OF DIGITAL TWINS

4.1 Greater Predictive Accuracy

The primary benefit of digital twins for cancer patients is the ability to make more precise predictions for the course of each patient.

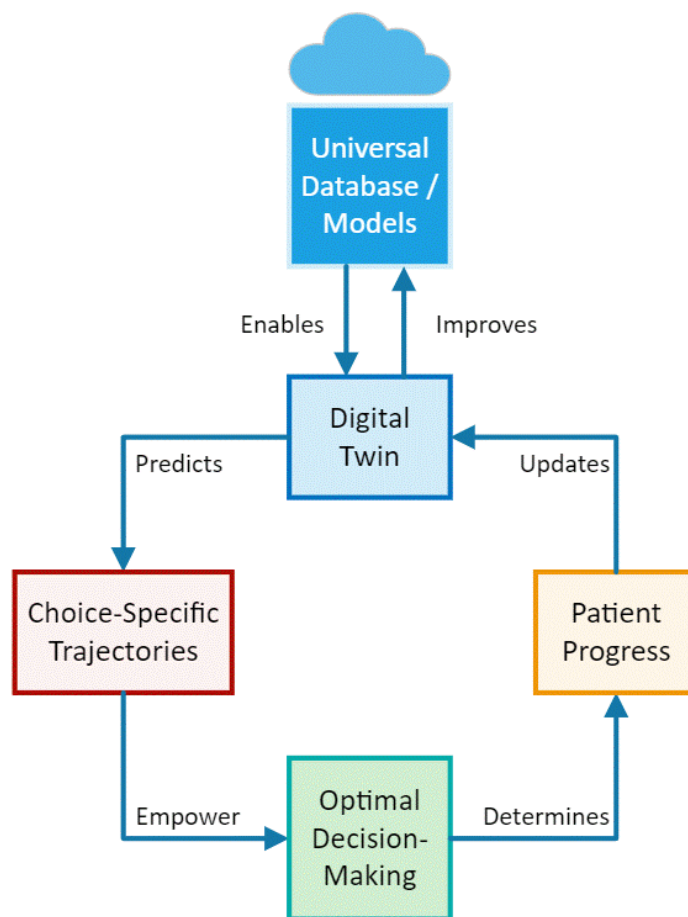


Figure 1: Network Architecture of Digital Twins

While a large amount of cancer research has focused on predicting the probabilities of patient outcomes and side effects using artificial intelligence or statistical models, much of this research attempts to predict the results directly from the patient's imaging and phenotyping. As a result, these models are "synthetic", i.e. they ignore all of the complex biological and physiological phenomena that fundamentally determine the patient's progress. Digital twins innovate the current scientific paradigm by reflecting these biological and physiological not just from AI or statistics but also through mechanistic modeling, genetic, environmental, and social factors. In doing so, digital twins progress away from synthetic inference and move closer to actual simulation of the patient. This makes digital twins much more likely to correctly predict the patient's outcome trajectory.

4.2 Treatment Plan Optimization

Digital twins can fundamentally change the paradigm of biological optimization for radiation therapy treatment planning. Historically, biological optimization relies solely on tumor volume and the delivered radiation dose distribution, discarding much of the information

that influences cancer radiobiology such as oxygenation and vasculature density. With digital twins, biological optimization can properly consider these factors, possibly greatly improving the resulting treatment plans. Similarly, digital twins can provide much more power to adaptive radiation therapy, a treatment paradigm in which the patient's radiation course is updated over the radiation course to account for the patient's anatomical changes. By using digital twins, adaptive radiation therapy can not only improve radiation dose homogeneity and conformality due to changes, but it can also modify the course based on the changing probabilities of cancer cure and toxic side effects. This can profoundly improve patient outcomes from radiation therapy in a previously impossible way.

4.3 Innovative Research Tools

Beyond the immense clinical side effects, digital twins can also be powerful tools for cancer researchers. Because of their low-level simulations of patients, digital twins are much easier to generalize beyond their training data. This makes it easy for researchers to investigate numerous possible combinations and ratios of therapeutic agents and strategies. Even further, digital twins will be able

to extend their predictions to novel therapeutic strategies by simulating the patient's reaction to them. This goes beyond the typical artificial intelligence limitation of being unable to predict results that are significantly different from the training data. Because of their virtual nature, digital twins could enable large-scale virtual clinical trials that are more thorough and drastically faster than current clinical trials, accelerating the translation from basic science to clinical solutions.

5 THE BARRIERS TO DIGITAL TWINS

5.1 Centralized Data Commons

Historically, individual hospital systems have maintained their own IT departments and databases in the interest of preserving the security of their patient health information. Therefore, it is likely that the database powering the digital twins would be maintained by the individual hospital systems that are using the digital twins to improve the outcomes of their own patients. While this makes it easier for patient information to be secure, hospitals would be hampering the effectiveness of their digital twins by forcing their digital twins to learn from only the hospital's data. Artificial intelligence and machine learning models improve significantly when their training data increases, so we expect that digital twins should become much more accurate and powerful if they were trained from the data of many hospital systems rather than one. Moreover, this format would require each hospital system to have their own high-performance computers in order to train and maintain hospital-specific digital twin models. This increases the difficulty and cost of implementing digital twins on the hospital system's end.

However, with modern cryptography techniques and cloud computing readily available, it would be possible for hospital systems to anonymously share their patients' de-identified data with some central organization while still preserving the electronic key to that patient. This would allow the hospital system to periodically update the universal database with each patient's progress and clinical decisions, continuously improving the universal database by enabling online learning techniques [13]. The universal database would also be able to house the computers required to train the models without requiring high-end computing power on the side of the hospital systems. This would make it much easier and less costly for each hospital system to start using digital twins for their own patients, which incentivizes more hospital systems to join and even further improve the universal database.

5.2 Patient-Specific Data Assembly

Regardless of how the digital twins are maintained and trained, they will require an assembly of patient-specific data to make patient-specific inferences. Some of this data may be commonly acquired in typical patients, such as blood panels or MRI or CT imaging of the relevant part of the body. Depending on the final technical implementation of the digital twins, less-common data may need to be included, such as genetic sequencing of relevant genes. Even with the assistance of a central universal database, novel data commons and assembly methods need to be developed to minimize the number of additional tests needed to generate a digital twin. Moreover, each digital twin would operate better with more of its patient's

data, so steps should be taken to increase the density of pre-existing patient data. For example, a digital twin would be able to produce better predictions for a patient undergoing lung radiation therapy if that patient has previously undergone radiation therapy. In this example, it would be ideal for the hospital to access all the data associated with the prior radiation therapy, but this would be difficult if that data were secured in another hospital system's database. Naturally, this problem also vanishes in the presence of universal databases because the hospital would be able to provide the patient's anonymized key to the database and retrieve all the patient's historical data, maximizing the created digital twin's predictive performance.

5.3 Multiscale Modeling

Although artificial intelligence and digital twin technology have been applied broadly to other industries, they have not been extensively investigated in the low-level human domain generated by multiscale modeling. This is primarily because the relevant advancements in multiscale modeling are relatively new. Therefore, there has not been enough time or awareness of researchers to develop models in these domains yet. The domains also pose unique traits due to their representation of lower-level human data. These traits open the door to a new class of artificial intelligence models that exploit the well-known mechanical, chemical, and biological properties of materials. This will enable the path from high-level empirical models that ignore the fine details of the human body to fundamental, physically motivated models that perform realistic human simulation. These models are an active area of research, but because they are so new and underdeveloped, the coupling of modern AI techniques with mechanistic models poses a unique challenge to digital twin development. However, it is likely that significant progress will be made on these models in the near future.

6 CONCLUSION

Digital twins would have an undeniably strong positive effect on radiation therapy for cancer patients. While there are some barriers that currently prevent digital twins from being developed and implemented clinically, significant groundwork has been done to overcome them. The recently developed multiscale, multimodal models are primed to power and contribute to digital twins. In the future, digital twins will significantly improve patient outcomes and enable a new class of large-scale virtual clinical trials, bringing us closer to a world where cancer has changed from being a deadly disease to a mild inconvenience.

REFERENCES

- [1] L. Li, S. Aslam, A. Wileman and S. Perinpanayagam, "Digital Twin in Aerospace Industry: A Gentle Introduction," in *IEEE Access*, vol. 10, pp. 9543-9562, 2022, doi: 10.1109/ACCESS.2021.3136458.
- [2] Piromalis D, Kantaros A. Digital Twins in the Automotive Industry: The Road toward Physical-Digital Convergence. *Applied System Innovation*. 2022; 5(4):65. <https://doi.org/10.3390/asi5040065>
- [3] Dirnfeld, Ruth. (2022). Digital Twins in Railways. 10.13140/RG.2.2.32690.68804.
- [4] Chelsea Finn. 2018. Learning to Learn with Gradients. PhD Thesis, EECS Department, University of Berkeley.
- [4] Al-Sehrawy, R., Kumar, B. (2021). Digital Twins in Architecture, Engineering, Construction and Operations. A Brief Review and Analysis. In: Toledo Santos, E., Scheer, S. (eds) *Proceedings of the 18th International Conference on Computing in Civil and Building Engineering. ICCCBE 2020. Lecture Notes in Civil Engineering*, vol 98. Springer, Cham. https://doi.org/10.1007/978-3-030-51295-8_64

- [5] Hou L, Wu S, Zhang G, Tan Y, Wang X. Literature Review of Digital Twins Applications in Construction Workforce Safety. *Applied Sciences*. 2021; 11(1):339. <https://doi.org/10.3390/app11010339>
- [6] Hall WA, Bergom C, Thompson RF, Baschnagel AM, Vijayakumar S, Willers H, Li XA, Schultz CJ, Wilson GD, West CML, Capala J, Coleman CN, Torres-Roca JF, Weidhaas J, Feng FY. Precision Oncology and Genomically Guided Radiation Therapy: A Report From the American Society for Radiation Oncology/American Association of Physicists in Medicine/National Cancer Institute Precision Medicine Conference. *Int J Radiat Oncol Biol Phys*. 2018 Jun 1;101(2):274–284. doi: 10.1016/j.ijrobp.2017.05.044. Epub 2017 Jun 9. PMID: 28964588.
- [7] Peng, G.C.Y., Alber, M., Buganza Tepole, A. *et al*. Multiscale Modeling Meets Machine Learning: What Can We Learn?. *Arch Computat Methods Eng* 28, 1017–1037 (2021). <https://doi.org/10.1007/s11831-020-09405-5>
- [8] Azuaje, F. Artificial intelligence for precision oncology: beyond patient stratification. *npj Precision Onc* 3, 6 (2019). <https://doi.org/10.1038/s41698-019-0078-1>
- [9] Wainberg, M., Merico, D., Delong, A., & Frey, B. J. (2018). Deep learning in biomedicine. *Nature biotechnology*, 36(9), 829–838.
- [10] Hormuth, D. A., Jarrett, A. M., & Yankeelov, T. E. (2020). Forecasting tumor and vasculature response dynamics to radiation therapy via image based mathematical modeling. *Radiation Oncology*, 15, 1–14.
- [11] Gaw, N., Hawkins-Daarud, A., Hu, L. S., Yoon, H., Wang, L., Xu, Y., ... & Li, J. (2019). Integration of machine learning and mechanistic models accurately predicts variation in cell density of glioblastoma using multiparametric MRI. *Scientific reports*, 9(1), 10063.
- [12] Bhattacharya, T., Brettin, T., Doroshov, J. H., Evrard, Y. A., Greenspan, E. J., Gryshuk, A. L., ... & Zaki, G. (2019). AI meets exascale computing: Advancing cancer research with large-scale high performance computing. *Frontiers in oncology*, 9, 984.
- [13] Bottou, Léon (1998). "Online Algorithms and Stochastic Approximations". *Online Learning and Neural Networks*. Cambridge University Press. ISBN 978-0-521-65263-6.