



Multiscale Modeling Meets Machine Learning: What Can We Learn?

Grace C. Y. Peng¹ · Mark Alber² · Adrian Buganza Tepole³ · William R. Cannon⁴ · Suvranu De⁵ · Salvador Dura-Bernal⁶ · Krishna Garikipati⁷ · George Karniadakis⁸ · William W. Lytton⁶ · Paris Perdikaris⁹ · Linda Petzold¹⁰ · Ellen Kuhl¹¹ 

Received: 30 December 2019 / Accepted: 9 February 2020
© CIMNE, Barcelona, Spain 2020

Abstract

Machine learning is increasingly recognized as a promising technology in the biological, biomedical, and behavioral sciences. There can be no argument that this technique is incredibly successful in image recognition with immediate applications in diagnostics including electrophysiology, radiology, or pathology, where we have access to massive amounts of annotated data. However, machine learning often performs poorly in prognosis, especially when dealing with sparse data. This is a field where classical physics-based simulation seems to remain irreplaceable. In this review, we identify areas in the biomedical sciences where machine learning and multiscale modeling can mutually benefit from one another: Machine learning can integrate physics-based knowledge in the form of governing equations, boundary conditions, or constraints to manage ill-posed problems and robustly handle sparse and noisy data; multiscale modeling can integrate machine learning to create surrogate models, identify system dynamics and parameters, analyze sensitivities, and quantify uncertainty to bridge the scales and understand the emergence of function. With a view towards applications in the life sciences, we discuss the state of the art of combining machine learning and multiscale modeling, identify applications and opportunities, raise open questions, and address potential challenges and limitations. We anticipate that it will stimulate discussion within the community of computational mechanics and reach out to other disciplines including mathematics, statistics, computer science, artificial intelligence, biomedicine, systems biology, and precision medicine to join forces towards creating robust and efficient models for biological systems.

Keywords Machine learning · Multiscale modeling · Physics-based simulation · Biomedicine

✉ Ellen Kuhl
ekuhl@stanford.edu

Grace C. Y. Peng
penggr@mail.nih.gov

Mark Alber
malber@ucr.edu

Adrian Buganza Tepole
abuganza@purdue.edu

William R. Cannon
william.cannon@pnnl.gov

Suvranu De
des@rpi.edu

Salvador Dura-Bernal
salvador dura@gmail.com

Krishna Garikipati
krishna@umich.edu

George Karniadakis
georgekarniadakis@brown.edu

William W. Lytton
bill.lytton@downstate.edu

Paris Perdikaris
pgp@seas.upenn.edu

Linda Petzold
petzold@engineering.ucsb.edu

- ¹ National Institutes of Health, Bethesda, MD, USA
- ² University of California, Riverside, USA
- ³ Purdue University, Lafayette, IN, USA
- ⁴ Pacific Northwest National Laboratory, Richland, WA, USA
- ⁵ Rensselaer Polytechnic Institute, Troy, NY, USA
- ⁶ State University of New York, New York, USA
- ⁷ University of Michigan, Ann Arbor, MI, USA
- ⁸ Brown University, Providence, RI, USA
- ⁹ University of Pennsylvania, Philadelphia, PA, USA
- ¹⁰ University of California, Santa Barbara, CA, USA
- ¹¹ Stanford University, Stanford, CA, USA

1 Motivation

Machine learning is rapidly infiltrating the biological, biomedical, and behavioral sciences and seems to hold limitless potential to transform human health [123]. It already is widely considered to be one of the most significant breakthroughs in medical history [27]. But can this technology really live up to its promise? Machine learning is the scientific discipline that seeks to understand and improve how computers learn from data. As such, it combines elements from statistics, understanding relationships from data, with elements from computer science, developing algorithms to manage data. The success of machine learning relies heavily on our ability to collect and interpret big data. In many fields of medicine, we have successfully done this for multiple decades. So what's really new? The recent excitement around machine learning is generally attributed to the increase in computational resources, cloud storage, and data sharing, which we can witness in our own lives through smart watches, wearable electronics, or mobile devices [40]. For example, a recent success story of machine learning in medicine has shown that it is possible to classify skin cancer into malignant and benign subtypes using photographic images, for example from smartphones [31]. Unarguably, the two most compelling opportunities for machine learning in biomedicine are diagnosis and prognosis. Potential applications range from identifying bone fracture, brain hemorrhages, and head trauma to detecting lung nodules, liver masses, and pancreatic cancer [123]. But is machine learning powerful and accurate enough that we can simply ignore physics based simulations entirely?

Machine learning is exceptionally good at integrating multimodality multi-fidelity data with the goal to reveal correlations between different features. This makes the technology very powerful in fields like radiology and pathology where we seek to classify risk or stratify patients based on medical images [123] and the answer is either binary, as in the classification of skin cancer [31], or a discrete number, as in a recent classification of 12 types of arrhythmia [38]. In fact, arrhythmia classification is an excellent example, because more than 300 million electrocardiograms are acquired annually worldwide, and we have access to a vast amount of annotated data. Problems may arise, however, when dealing with sparse or biased data [98]. In these cases the naive use of machine learning can result in ill-posed problems and generate non-physical predictions. Naturally, this brings up the question that, provided we know the underlying physics, can we integrate our prior knowledge to constrain the the space of admissible solutions to a manageable size [96]?

Recent trends in computational physics suggest that exactly this approach [13]: to create data-efficient

physics-informed learning machines [92, 93]. Biomedicine has seen several application of these techniques in cardiovascular flows modeling [46] or in cardiac activation mapping [105], where we already have a reasonable physical understanding of the system and can constrain the design space using the known underlying wave propagation dynamics. Another example where machine learning can immediately benefit from multiscale modeling and physics-based simulation is the generation of synthetic data [102], for example, to supplement sparse training sets. This raises the obvious question—especially within the computational mechanics community—where can physics-based simulations benefit from machine learning?

Physics-based simulations are enormously successful at integrating multiscale, multiphysics data with the goal of uncovering mechanisms that explain the emergence of function [18]. In biomedicine, physics-based simulation and multiscale modeling have emerged as a promising technologies to build organ models by systematically integrating knowledge from the molecular, cellular, and tissue levels [25, 43] as evidenced by initiatives like the United States Federal Interagency Modeling and Analysis Group IMAG [82, 136]. Two immediate opportunities for machine learning in multiscale modeling include learning the underlying physics [100] and learning the parameters for a known physics-based problem. Recent examples of learning the underlying physics are the data-driven solution of problems in elasticity [22] and the data-driven discovery of partial differential equations for nonlinear dynamical systems [13, 91, 94]. This class of problems holds great promise, especially in combination with deep learning, but involves a thorough understanding and direct interaction with the underlying learning machines [96]. Are there also immediate opportunities for integrating machine learning and multiscale modeling, more from an end-user perspective, without having to modify the underlying tools and technologies at their very core?

This manuscript seeks to answer the question of how multiscale models can benefit from machine learning. As such, it is an extended version of a recent review article [3] and was inspired by a recent workshop on integrating machine learning with multiscale modeling. We have structured it around four methodological areas, ordinary and partial differential equations, and data and theory driven machine learning. For each area, we discuss the state of the art, identify applications and opportunities, raise open questions, and address potential challenges and limitations in view of specific examples from the life sciences. To make this work accessible to a broad audience, we summarize the most important terms and technologies associated with machine learning in boxes where they are first mentioned. We envision that this work will stimulate discussion and inspire scientists in the broader field of computational mechanics to explore the potential of machine learning towards creating reliable and

robust predictive tools for biological, biomedical, and behavioral systems to the benefit of human health.

2 Ordinary Differential Equations

Ordinary differential equations in time are ubiquitous in the biological, biomedical, and behavior sciences. At the molecular, cellular, organ, or population scales it is often easier to make observations and acquire data associated with ordinary differential equations than for partial differential equations, since the latter encode spatial variations, which are often more difficult to access. Ordinary differential equation based models can range from single equations to large systems of equations or stochastic ordinary differential equations. This implies that the number of parameters is typically large and can easily reach hundreds or more. Figure 1 illustrates an example of ordinary differential equations to explore the biophysical mechanisms of development [122].

2.1 State of the Art

Assuming we have acquired adequate data, the challenge begins with identifying the nonlinear, coupled driving terms. To analyze the data, we can apply formal methods of *system identification*. Common techniques include classical regression using L1 or LASSO and L2 or ridge regression, as well as stepwise regression with statistical tests [13, 130]. These approaches are essentially nonlinear optimization problems that learn the set of coefficients by multiplying combinations of algebraic and rate terms that result in the best fit to the observations. For adequate data, the *system identification* problem is usually relatively robust and can learn a parsimonious set of coefficients,

especially with stepwise *regression*. Clearly, parsimony is central to identifying the correct set of equations and the easiest strategy to satisfy this requirement is classical or stepwise *regression*.

System identification refers to a collection of statistical methods that identify the governing equations of a system from data. These methods can be applied to obtain either an equilibrium response or the dynamics of a system. Typical examples include inferring operators that form ordinary [68] or partial [130] differential equations.

Regression is a statistical process of estimating the relationship between a dependent variable and one or more independent variables. In the context of machine learning, regression is classified as a supervised learning approach in which the algorithm learns from a training set of correctly identified observations and then uses this learning to evaluate new observations. The underlying assumption is that the output variable is continuous over the input space. Examples in biomedicine include predicting an individual's life expectancy, identifying a tolerable dose of chemotherapy, or exploring the interplay between drug concentration and arrhythmogenic risk [102], among many others.

Any discussion of *system identification* from experimental data should address *uncertainty quantification* to account for both measurement errors and model errors. The Bayesian setting provides a formal framework for this purpose [44]. Prior probability distribution functions must be assumed for the errors. In the absence of deeper insights into the measurement techniques, a common choice is the Gaussian distribution. On this note, we observe that recent *system identification* techniques [13, 19, 68, 69, 88, 100, 130] start from a large space of candidate terms in the ordinary differential equations to systematically control and treat model errors. Machine learning provides a powerful approach to reduce the number of dynamical variables and parameters while maintaining the biological relevance of the model [13, 112].

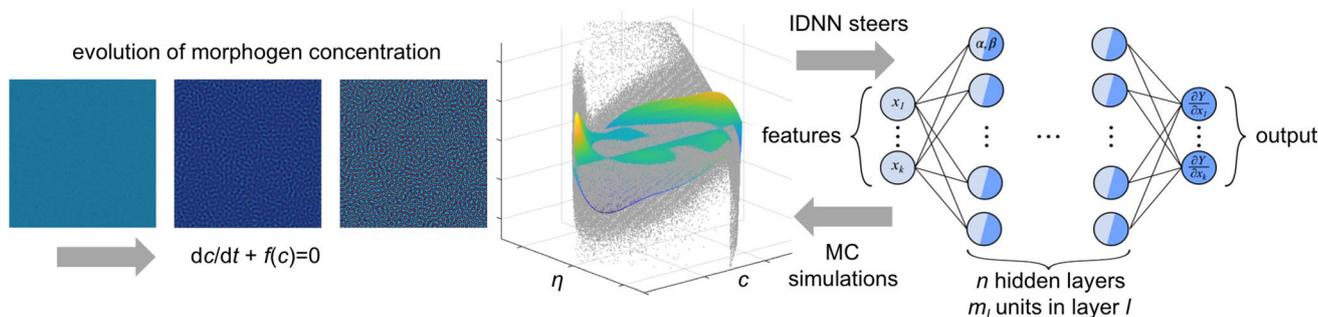


Fig. 1 Ordinary differential equations. Biophysical mechanisms of development can be discerned by identifying the nonlinear driving terms in ordinary differential equations that govern the evolution of morphogen concentrations, left. Metabolic processes evolve on a free energy landscape $g(c, \eta)$ that can be explored by Monte Carlo simula-

tions, thus generating large scale data, shown as a grey point cloud, that are used to train various classes of neural networks, right. The colored surface is an integrable deep neural network [122] representation of the metabolic data. (Color figure online)

Uncertainty quantification is the science of characterizing and reducing uncertainties. Its objective is to determine the likelihood of certain outcomes if some aspects of the system are not exactly known. Since standard variations in biomedical data are usually large and it is critical to know how small variations in the input data affect the output, uncertainty quantification is indispensable in medicine. The inherent nonlinearity of biological process also drives a need for uncertainty analysis, since noise in the inputs is nonlinearly propagated through the system. Sources of uncertainty can be experimental or computational related to the underlying governing equations, their parameters, and boundary conditions. Some examples include quantifying the effects of experimental uncertainties in heart failure [81], the effects of biomechanical stimuli in coronary artery bypass grafts [126], or the effects of material properties on stress profiles in reconstructive surgery [55].

2.2 Applications and Opportunities

There are numerous applications of ordinary differential equations that integrate machine learning and multiscale modeling for biological, biomedical, and behavioral systems.

Metabolic networks Machine learning has been applied to take advantage of large amounts of genomics and metabolomics data for the optimization of ordinary differential equation-based metabolic network models and their analysis [24]. For example, machine learning and genome-scale models were applied to determine the side effects of drugs [110]. Also, a recent study used a combination of machine learning and multiomics data, proteomics and metabolomics, to effectively predict pathway dynamics providing qualitative and quantitative predictions for guiding synthetic biology efforts [23]. **Supervised learning** methods are often used for finding the metabolic dynamics represented by coupled nonlinear ordinary differential equations to obtain the best fit with the provided time-series data.

Supervised learning defines the task of learning a function based on previous experience in the form of known input–output pairs or function evaluations. In many cases, this is a task that a trained person can do well, and the computer is trying to approximate human performance. When the input is high-dimensional, or the function is highly nonlinear, personal intuition may not be useful and supervised learning can overcome this limitation. Typical examples include classification and regression tasks. In biomedicine, a common classification problem is pattern recognition in an electrocardiogram to select from a limited set of diagnoses [38]; other examples include detecting cancer from medical images [31], estimating risk scores for coronary heart disease, guiding antithrombotic therapy in atrial fibrillation, and automating implantable defibrillators in hypertrophic cardiomyopathy [27].

Unsupervised learning defines the task of identifying naturally occurring patterns or groupings within datasets that consisting of input features without labeled output responses. The most common types of unsupervised learning techniques include clustering and density estimation used for exploratory data analysis to identify hidden patterns or groupings. In biomedicine, a promising example is precision medicine [27].

Microbiology, immunology, and cancer The coupled, nonlinear dynamics of intracellular and extracellular signaling are represented by cascades of tens of ordinary differential equations to model the onset of tuberculosis [97]. The same approach is applied to modeling the interaction of the immune system and drugs in the mathematical biology of cancer [74]. In this case, the major challenge is *system identification*. Another application is bridging scales in cancer progression in specific micro-environments by mapping genotype to phenotype using neural networks [34].

Neuroscience Machine learning is applied to *system identification* of the ordinary differential equations that govern the neural dynamics of circadian rhythms [11, 28, 73]. Principal component analysis and neural networks have been more widely applied to memory formation [79, 87], chaotic dynamics of epileptic seizures [1, 2], Alzheimers Disease, and aging.

Biomechanics The most prominent potential application of machine learning in biomechanics is in the determination of response functions including stress–strain relations or cell–scale laws in continuum theories of growth and remodeling [4]. These relations take the form of both ordinary differential equations, for which *system identification* is of relevance, and direct response functions, for which the framework of *deep neural networks* is applicable [101]. For example, a recent study integrated machine learning and multiscale modeling to characterize the dynamic growth and remodeling during heart failure across the scales, from the molecular via the cellular to the cardiac level [81].

Deep neural networks are a powerful form of machine learning strategies to approximate functions. The input features proceed through multiple hidden layers of connected neurons that progressively compose these features and ultimately produce an output. The key feature of deep learning is that the architecture of the network is not determined by humans, but rather by the data themselves. Deep neural networks have been successfully used in image and speech recognition [54]. The number of examples of deep learning in biomedicine is rapidly increasing and includes interpreting medical images to classify tuberculosis, identify bone fracture, detect lung nodules, liver masses, and pancreatic cancer, identify brain hemorrhages and head trauma, and analyze mammograms and electrocardiograms [123].

Public health The dynamics of disease spreading through a population, affected by environmental factors, has long been represented by cascades of ordinary differential equations. A major challenge in this application is determining the parameters of the ordinary differential equations by *system identification* [17]. Interestingly, the ordinary differential equations of disease spreading have recently been adopted to model the prion-like spreading of neurodegenerative diseases [133], where the parameters could potentially be identified from magnetic resonance images using machine learning.

2.3 Open Questions

Maximizing information gain An open question in modeling biological, biomedical, and behavior systems is how to best analyze and utilize sparse data. In such a setting, sparse identification techniques must be integrated with the experimental program. Optimal experimental design [41] methods allow the most efficient choice of experiments to maximize the information gain using criteria such as the Kullback–Leibler divergence or the Akaike Information Criterion. The information theoretic approach is particularly powerful to treat model form errors. In biological systems, where data may be obtained by resource-intensive wet lab experiments or multi-scale model simulations, the most efficient combination of such approaches could result in maximizing novel biological insight.

Optimizing efficiency The identification of the system of governing equations still leaves open the question of efficiency and time to solution, especially if equations are to be used in a sampling approach such as Monte Carlo. In this setting, the repeated generation of solutions inevitably suggests that we should circumvent the expense of time integration methods. Deep learning methods centered on neural networks offer a number of options. Particularly well-suited are *recurrent neural networks*, based on long short-term memory cells, which account for the inherent time dependence of ordinary differential equations and their solutions. Neural networks can efficiently encode the complex time dependence of ordinary differential equations. For example, extremely high-order time integration schemes such as Runge–Kutta algorithms of the order of one hundred have been replaced successfully by tailored *deep neural networks* [92]. The construction and use of such surrogate models is indispensable for sampling upward of tens of thousands of entire trajectories of dynamical systems such as reaction–diffusion of coupled ligand-morphogen pairs.

Recurrent neural networks are a class of neural networks that incorporate a notion of time by accounting not only for current data, but also for history with tunable extents of memory. A recent application is identifying unknown constitutive relations in ordinary differential equation systems [36]

Combining multi-fidelity neural networks Neural networks can be directly trained against labels for a quantity of interest such as the time-averaged solution or its frequency distribution. Principal component analysis can be applied to the dynamics to develop reduced order models. *Deep neural networks* can be combined into an approach of *multi-fidelity learning* [53] that integrates well with multiscale modeling methods. In this approach, multiple neural networks are trained. Coarse scale, but plentiful data, for example obtained from larger numbers of trajectories reported at fewer time instants, are used to train low-fidelity neural networks, which are typically shallow and narrow.

Progressively finer scale data at increasing numbers of time instants, but for fewer trajectories and expensive to obtain, are used to train higher fidelity *deep neural networks* to minimize the error between the labels and the output of the low-fidelity neural network. The low-fidelity neural network resolves the low frequency components of the response, with the progressively higher fidelity *deep neural networks* representing the higher frequencies. The underlying principle is that the knowledge base of the response is resolved by shallower and narrower neural networks, while the critical, high frequency response is left to the high-fidelity deep neural network. Developing these novel approaches is important to accurately resolve the dynamics of, for example, reaction–diffusion systems of ligands and morphogens that together control patterning in developmental biology.

2.4 Potential Challenges and Limitations

Dealing with inadequate resolution The most prominent challenge facing the application of machine learning and data-driven methods to obtaining a better understanding of biological systems is a paucity of data. In ordinary differential equation modeling, we often have to rely on classical data acquisition techniques, for example, microscopy or spectroscopy, which are known to have limited temporal resolution. Obtaining time series data at a sufficiently high temporal resolution to build and train mathematical models has always been and will likely remain a major challenge in the field.

Processing sparse data Sparse, incomplete, or heterogeneous data pose a natural challenge to modeling biological, biomedical, and behavioral systems. In principle, direct numerical simulations can fill this gap and generate missing data. However, the simulations themselves can be limited by poorly calibrated parameter values. There is, therefore, a pressing need to develop robust inverse methods that are capable of handling sparse data. One example is *robust system identification*, the creation of mathematical models of dynamical systems from measured experimental data. This naturally implies the optimal design of experiments to efficiently generate informative training data and iterative model refinement or progressive model reduction.

3 Partial Differential Equations

Partial differential equations describe the physics that govern the evolution of biological systems in time and space. The interaction between the different scales, both spatial and temporal, coupled with the various physical and biological processes in these systems, is complex with many unknown parameters. As a consequence, modeling biological systems in a multi-dimensional parametric space poses challenges of *uncertainty quantification*. Moreover, modeling these

systems depends crucially on the available data, and new multi-modality data fusion methods will play a key role in the effective use of partial differential equation modeling of multiscale biological systems. An additional challenge of modeling complex biological phenomena stems from the lack of knowledge of some of the processes that need to be modeled. Physics-informed machine learning is beginning to play a central role at this front, leveraging *multi-fidelity* or multi-modality data with any known physics. These data can then be exploited to discover the missing physics or unknown processes.

3.1 State of the Art

Modeling biological, biomedical, and behavioral systems crucially depends on both the amount of available data and the complexity of the system itself. The classical paradigm for which many numerical methods have been developed over the last 50 years is shown in the top of Fig. 2, where we assume that the only data available are the boundary and initial conditions, while the specific governing partial differential equations and associated parameters are precisely known. On the other extreme, in the bottom of Fig. 2, we may have a lot of data, for example in the form of time series, but we do not know the governing physical law, for example the underlying partial differential equation at the continuum level.

Many problems in social dynamics fall under this category, although work so far has focused on recovering known partial differential equations from data only. Perhaps the most interesting category for biological systems is sketched in the middle of Fig. 2, where we assume that we know the physics partially but not entirely. For example, we know the conservation law and not the constitutive relationship, but we have several scattered measurements in addition to

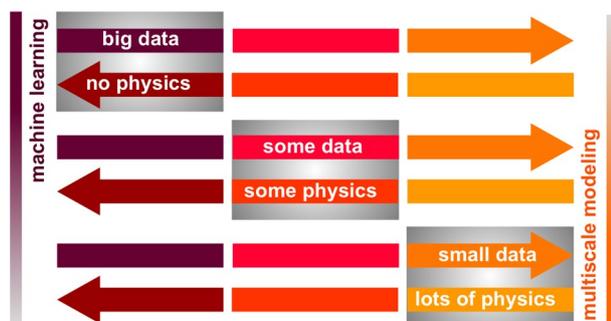


Fig. 2 Categories of modeling biomedical systems and associated available data and underlying physics. We use the term physics to imply the known physics for the target problem. Physics-informed neural networks can seamlessly integrate data and mathematical models, including models with missing biophysics, in a unified and compact way using automatic differentiation and partial differential equation-induced neural networks

the boundary and initial conditions that we can use to infer the missing functional terms and other parameters in the partial differential equation and simultaneously recover the solution. This middle category is the most general case. In fact, it is representative of the other two categories, if the measurements are too few or too many. This mixed case may lead to the significantly more complex scenarios, where the solution is a stochastic process due to stochastic excitation or an uncertain material property, for example the diffusivity in a tissue. Hence, we can employ stochastic partial differential equations to represent these stochastic solutions and other stochastic fields. Finally, there are many problems involving long-range spatio-temporal interactions, for example the viscoelasticity of arteries or the super-diffusion inside a cell, where fractional calculus and fractional partial differential equations, rather than the currently common partial differential equations with integer order derivatives, may be the proper mathematical model to adequately describe such phenomena. *Physics-informed machine learning* Prior physics-based information in the form of partial differential equations, boundary conditions and constraints regularize a machine learning approach in such a way that it can then learn from small and noisy data that can evolve in time. Very recently, the field has seen the leveraging of *Gaussian process regression* and *deep neural networks* into physics-informed machine learning [90–96]. For *Gaussian process regression*, the partial differential equation is encoded in an informative function prior; for *deep neural networks*, the partial differential equation induces a new neural network coupled to the standard uninformed data-driven neural network, see Fig. 3. We refer to this coupled data-partial differential equation deep neural network as a *physics-informed neural network*. New approaches, for example using generative adversarial networks, will be useful in the further development of *physics-informed neural networks*, for example, to solve stochastic partial differential equations, or fractional partial differential equations in systems with memory.

Gaussian process regression enables the creation of computationally inexpensive surrogates in a Bayesian approach. These surrogates do not assume a parametric form a priori and, instead, let the data speak for themselves. A significant advantage of Gaussian process surrogates is the ability to not only predict the response function in the parameter space, but also the associated epistemic uncertainty. Gaussian process regression has been used to creating surrogate models to characterize the effects of drugs on features of the electrocardiogram [102] and the effects of material properties on the stress profiles from reconstructive surgery [56].

Multiscale modeling for biological systems Biological materials are known to have a complex hierarchy of structure, mechanical properties, and biological behavior across spatial and temporal scales. Throughout the past two decades, modeling these multiscale phenomena has been a point of attention, which has advanced detailed deterministic models

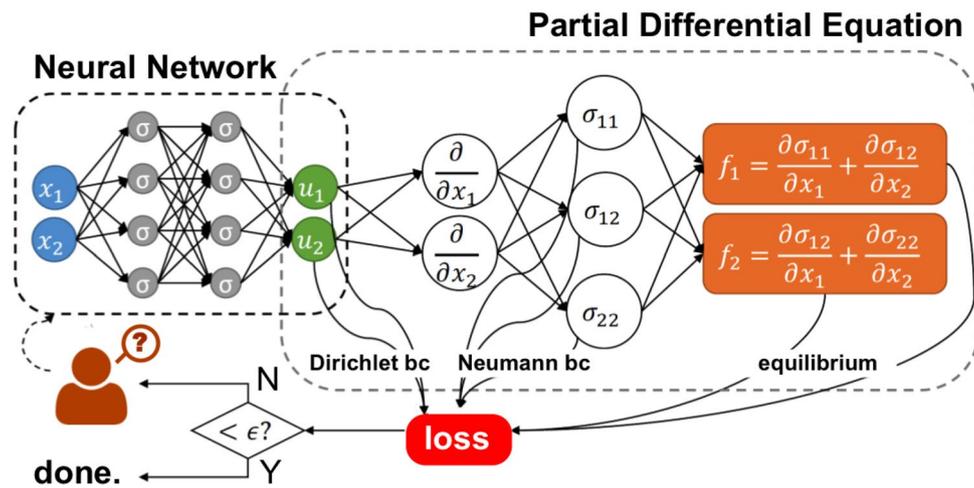


Fig. 3 Physics-informed neural networks. The left physics uninformed network represents the solution $u(x, t)$ of the partial differential equation; the right physics informed network describes the residual $f(x, t)$ of the partial differential equation. The example illustrates the nonlinear Schrödinger equation with unknown parameters λ_1 and λ_2 to be learned. In addition to unknown parameters, we can learn missing functional terms in the partial differential equation. Cur-

rently, this optimization is done empirically based on trial and error by a human-in-the-loop. Here, the u -architecture is a fully-connected neural network, while the f -architecture is dictated by the partial differential equation and is, in general, not possible to visualize explicitly. Its depth is proportional to the highest derivative in the partial differential equation times the depth of the uninformed u neural network

and their coupling across scales [25]. Strategies for multiscale modeling can be top down or bottom up, including: (i) models of representative volume elements for the microscopic scale coupled with the larger spatial scales through boundary conditions in terms of the first derivatives of the macroscale fields, first order coupling, or higher derivatives, higher order coupling [33, 49, 107]; (ii) micromechanics approaches [70, 89]; (iii) reduced order and simplified models for upscaling [108, 120]; (iv) scale-bridging or quasi-continuum methods [111, 129]. The explicit coupling of scales through representative volume elements is widely used because it can include many of the details and model complexities of the microscale, but it requires nested function evaluations such as nested finite element simulations or FE² that can easily become computationally intractable [20, 32, 48]. An additional complication in modeling biological systems comes from the inherent source of uncertainty in living matter, the high heterogeneity of the microscale, inter-subject variability, and stochastic nature of biological regulatory networks [15, 35, 63].

Machine learning for multiscale systems. Machine learning methods have recently permeated into composites research and materials design for example to enable the homogenization of representative volume elements with neural networks [51, 58, 60, 90] or the solution of high-dimensional partial differential equations with deep learning methods [37, 121, 122, 134, 135]. *Uncertainty quantification* in material properties is also gaining relevance, with examples of Bayesian model selection to calibrate of strain energy functions [66, 72] and uncertainty propagation with Gaussian processes of nonlinear mechanical systems [55, 56,

101]. These trends for non-biological systems point towards immediate opportunities for integrating machine learning and multiscale modeling in biological, biomedical, and behavioral systems and open new perspectives unique to the living nature of biological systems.

3.2 Applications and Opportunities

From localization to homogenization This application of machine learning in multiscale modeling is obvious, yet still unaccomplished. Leveraging data-driven approaches, the major objectives are localization and homogenization of information across the scales. The localization, the mapping of quantities from the macroscale, for example tissue stress or strain, to quantities at the microscale, for example cellular force or deformation, is crucial to understand the mechanobiology of the cell. The homogenization, the identification of the constitutive behavior at macroscale from the detailed behavior of representative units at the microscale, is critical to embed this knowledge into tissue or organ level simulations. In biological, biomedical, and behavioral systems, localization and homogenization present additional challenges and opportunities including the high-dimensional parameter space and inherent uncertainty, features that apply to both ordinary and partial differential equation based models, and the high degree of heterogeneity and microstructural complexity, features that mainly affect partial differential equation based models.

From single source to multi-modality and multi-fidelity modeling Figure 4, left, illustrates various sources of data

that that can potentially be combined with machine learning techniques. Biological, biomedical, and behavioral research crucially rely on experiments from different systems including *in vitro* culture, *in vivo* animal models and human data, and *in silico* experiments. The underlying assumption is that the data associated with each type of experiment or model are strongly correlated, even if they are not originating from the same system. Learning from one type of experiment or computer model can be used to improve the prediction at a higher level of fidelity for which information is scarce or difficult to obtain. Nonlinear multi-modality data fusion is a new way to combine information of various sources towards creating predictive models [83]. A typical example is integrating multi-omics data with biophysical models.

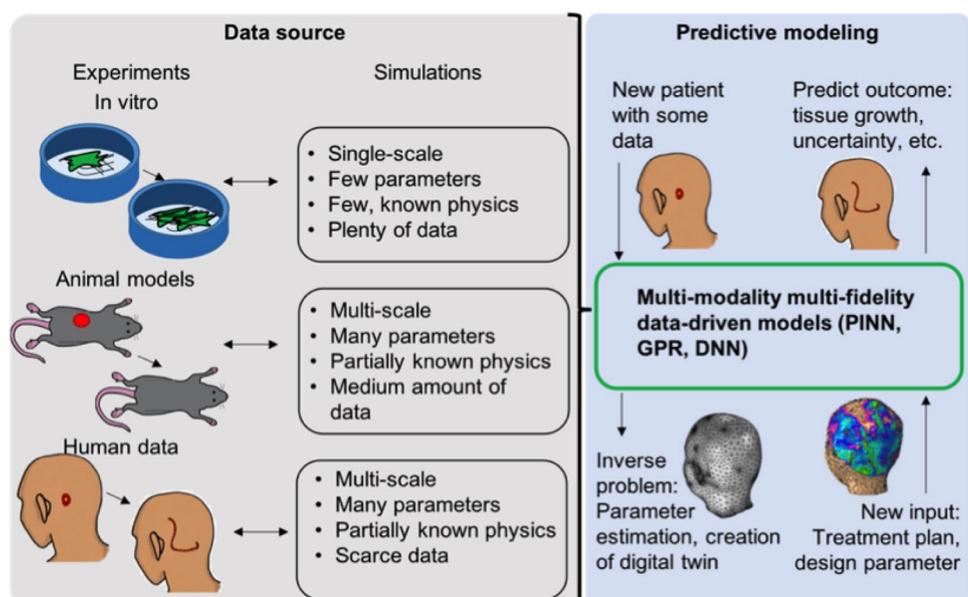
From parameter estimation to system identification to function discovery Figure 4, right, illustrates the combination of parameter estimation, *system identification*, and function discovery required to create a digital twin. The combination of multi-modality, *multi-fidelity*, data-driven techniques allows us to create a personalized computational model for an individual by combining previous observations from multi-scale simulations, experiments, and clinical data with continuously updated recordings of this individual. Using the digital twin, we can probe different treatment scenarios and screen the design parameter space to create personalized treatment plans.

From theoretical models to systems biology Living matter is characterized by its unique ability to respond and adapt to its environment. This can involve metabolic changes, inflammation, or mechanical changes such as growth and remodeling. Regulation of tissue activity is ultimately encoded in complex cell-signaling regulatory networks that operate at multiple spatial and temporal scales [114]. Modeling tissue

adaptation thus involves accounting for classical equilibrium principles, e.g., momentum and energy, as well as signaling network dynamics often described by stochastic reactive transport models with many variables and parameters. The complexity of the system often defies intuition [47]. Machine learning could enable discovery of reactions, e.g., coagulation cascades, the solution of inverse problems for parameter estimation, and the quantification of biological uncertainty in model predictions. Figure 5 shows a possible application of predicting growth and remodeling at the tissue level based on cell-level information.

From theoretical models to clinical applications Application of physics-based modeling in clinical practice is currently hindered by the difficulty to generate patient-specific predictions. Creating personalized models is time consuming and requires expert input and many different types of data, from multi-omics to biomechanical properties and measurements. On the opposite end, generic models are useful to understand mechanisms but are not suitable for planning individual interventions. Thus, there is an opportunity for machine learning and transfer learning in particular to generate individualized models or predictions in a fast and reliable manner without the need to create individualized models from scratch. Applications could include predicting the course of dissections in aortic dissections or quantifying wall shear stresses in aneurysms near the arterial wall [95]. Other applications include making predictions of whether thrombosis embolization will occur, or predictions of other cardiovascular diseases using multi-modality measurements integrated via *multi-fidelity* modeling to train neural networks instead of intricate models. Another possible application is optimization of surgery or prosthetic device design

Fig. 4 Multi-modality and multi-fidelity modeling of biomedical systems. Data from both experiments and computational models can be combined through machine learning to create predictive models. The underlying assumption is that, for a system of interest, data from different sources is correlated and can be fused. Parameter estimation, system identification, and function discovery result in inverse problems, for example, the creation of a digital twin, and forward problems, for example, treatment planning



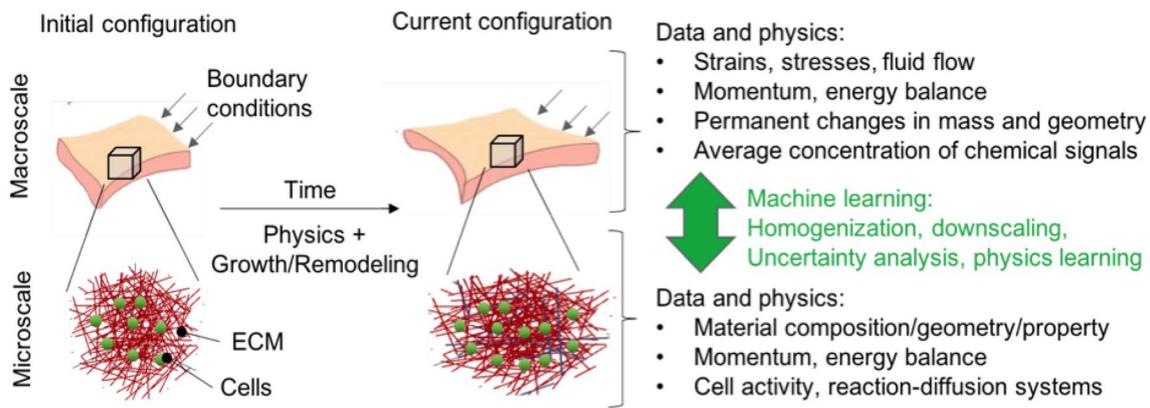


Fig. 5 Machine learning for multiscale modeling of biomedical systems. Tissues are characterized by hierarchical structure across spatial and temporal scales associated with inherent variability. At both the macroscale and microscales, biological systems satisfy physics-based partial differential equations for mass, momentum, and energy bal-

by combining off-line generic simulations with online data acquisition.

3.3 Open Questions

Modeling high-dimensional systems Can we model systems with high-dimensional input and hundreds of parameters? Biological systems are characterized by heterogeneous microstructures, spatial heterogeneity, many constituents, intricate and noisy cell-signaling networks, and inherent variability across subjects. Attempts to model these systems necessarily rely on a high-dimensional parametric input space. Despite the recent progress in *uncertainty quantification* methods and smart sampling techniques using sparse grids and polynomial chaos methods, handling data in high-dimensional spaces remains challenging. However, deep learning techniques can exploit the compositional structure of approximating functions and can, in principle, beat the curse of dimensionality [85]. Generative adversarial networks can also be useful for effective modeling of parameterized partial differential equations with thousands of uncertain parameters [140, 141].

Managing ill-posed problems Can we solve ill-posed inverse problems for parameter estimation or system identification? Many of the inverse problems for biological systems are ill posed, for example parameter estimation or *system identification*; they constitute boundary value problems with unknown boundary conditions. Classical mathematical approaches are not suitable in these cases. Methods for backward *uncertainty quantification* could potentially deal with the uncertainty involved in inverse problems, but these methods are difficult to scale to realistic settings. In view of the high dimensional input space and the inherent uncertainty of biological systems, posing

inverse problems is challenging. For instance, it is difficult to determine if there are multiple solutions or no solutions at all, or to quantify the confidence in the prediction of an inverse problem with high-dimensional input data. The inherent regularization in the loss function of neural networks allows us to deal with ill-posed inverse partial differential equations without boundary or initial conditions and to discover hidden states and biophysics not possible with classical methods. Moreover, advances in probabilistic programming offer a promising path for performing scalable statistical inference for large-scale inverse problems with a large number of uncertain parameters.

Discretizing space and time Can we remove or automate the tyranny of grid generation of conventional methods? Discretization of complex and moving three-dimensional domains remains challenging. It generally requires specific expertise and many hours of dedicated labor, and has to be re-done for each particular model. This is particularly important when creating personalized models with complex geometries and multiple spatial and temporal scales. While many efforts in machine learning are devoted to solving partial differential equations in a given domain, new opportunities include the use of machine learning to deal directly with the creation of the discrete problem. This includes automatic mesh generation, meshless interpolation, and parameterization of the domain itself as one of the inputs for the machine learning algorithms. Interestingly, some recent approaches with *physics-informed neural networks* entirely remove the notion of a mesh, and instead evaluate the conservation laws of mass, momentum, and energy at random points that are neither connected through a regular lattice nor through an unstructured grid.

Physics-informed neural networks are neural networks that solve supervised learning tasks while respecting physical constraints or encoding the partial differential equation in some way, for example, through the loss function. This technique is particularly powerful when dealing with sparse data from systems that obey known physical principles. Examples in biomedicine include diagnosing cardiovascular disorders non-invasively using four-dimensional magnetic resonance images of blood flow and arterial wall displacements [46], creating computationally efficient surrogates for velocity and pressure fields in intracranial aneurysms [95], and using nonlinear wave propagation dynamics in cardiac activation mapping [105].

Combining deterministic and stochastic models Can we couple conventional physics, mass and momentum balance, with stochastic reaction–diffusion over time to model the adaptation of living systems? While the laws that govern the physics of biological systems from the cellular scale to the tissue scale can be considered deterministic, the cell-signaling networks on the sub-cellular scales are inherently noisy [35, 109, 115]. We can model their multiphysics and multi-rate dynamics using neural networks by sharing the parameter space of coupled but separate neural networks, with each net representing a different multiscale process. This approach could alleviate some of the difficulties associated with stiff systems by exploiting the use of proper regularization terms in the loss function.

3.4 Potential Challenges and Limitations

Integrating multi-modality training data Even though nonlinear data fusion algorithms are currently being developed and improved, in the case of biological systems, as Fig. 3 left suggests, the data can come from vastly different sources. A potential challenge is to appropriately weight the data from these different modalities to improve predictions. For instance, it remains a challenge to quantify to what extent an in vitro model or an animal model is representative of the human response. It is possible that even the most powerful machine learning tools cannot substantially improve predictions needed in the clinical setting without high quality data of the human-specific response. Designing new composite neural networks that exploit correlations between multi-modality and *multi-fidelity* data is a top priority in the near future.

Multi-fidelity learning is a supervised learning approach used to fuse data from different sources to create a surrogate that can outperform predictions based on a single data source. Often times there is plenty of inexpensive, low fidelity data, for example from a simplified computational model or a simple experiment. At the same time, our confidence in the accuracy of this model or experiment is relatively low. In contrast, there is typically sparse expensive, high fidelity data from more complex computational models or experiments. Multi-fidelity learning exploits the correlations between the different fidelity levels to make predictions in regions of the input space for which no high fidelity data exist, but low fidelity measurements are easy to acquire. Recent examples include simulating the mixed convection flow past a cylinder [83], skin growth in tissue expansion [57], and cardiac electrophysiology [103].

Tuning physics inspired neural networks for stiff multiscale systems Multiscale systems lead to very complex landscapes of the loss function that current optimization solvers cannot deal with. Hence it could be impossible to train such systems. A possible approach is to use classical domain decomposition methods to deal with different regimes of scales and design corresponding distributed *physics-informed neural networks* that can be trained more easily and more accurately.

Increasing rigor and reproducibility The predictive power of models built with machine learning algorithms needs to be thoroughly tested. An important challenge is the creation of rigorous validation tests and guidelines for computational models. The use of open source codes and data sharing by the machine learning community is a positive step, but more benchmarks and guidelines are required for *physics-informed neural networks*. There is also an urgent need of sets of benchmark problems for biological systems. Reproducibility has to be quantified in terms of statistical metrics, as many optimization methods are stochastic in nature and may lead to different results.

Knowing limitations The main limitation of solving partial differential equations with neural networks and adversarial networks is the training time that results from solving non-convex optimization problems in high-dimensional spaces. This is a non-deterministic polynomial-time hard computational problem and one that will probably not be resolved in the near future. A possible solution is properly selecting the size of the system using domain decomposition techniques. Tuning the network parameters is a tedious and empirical job, but we can potentially adapt meta-learning methods that have been successfully used for *classification* problems to solve *regression* problems of partial differential equation modeling. Effective initialization, tailored to biological, biomedical, and behavioral systems, is a promising approach to alleviate some of these challenges and limitations.

Classification is a supervised learning approach in which the algorithm learns from a training set of correctly classified observations and uses this learning to classify new observations, where the output variable is discrete. Examples in biomedicine include classifying whether a tumor is benign or malignant [31], classifying the effects of individual single nucleotide polymorphisms on depression [7], the effects of ion channel blockage on arrhythmic risk in drug development [104], and the effects of chemotherapeutic agents in personalized cancer medicine [26].

4 Data-Driven Approaches

Often considered as an extension of statistics, machine learning is a method for identifying correlations in data. Machine learning techniques are function approximators and not predictors. This distinguishes them immediately from multiscale modeling techniques, which do provide predictions that can be based on parameter changes

suggested by particular pathological or pharmacological changes [42]. The advantage of machine learning over both manual statistics and multiscale modeling is its ability to directly utilize massive amounts of data through the use of iterative parameter changes. This ability to handle big data is important in view of recent developments of ultra-high resolution measurement techniques like cryo-EM, high-resolution imaging flow cytometry, or four-dimensional-flow magnetic resonance imaging. Machine learning also allows us to analyze massive amounts of health data from wearable devices and smartphone apps [40]. As a data-mining tool, machine learning can help us bring experiment and multiscale modeling closer together. Machine learning also allows us to leverage data to build artificial intelligence applications to solve biomedical problems [124]. Figure 6 illustrates a framework for integrating multiscale modeling and machine learning in data-driven approaches.

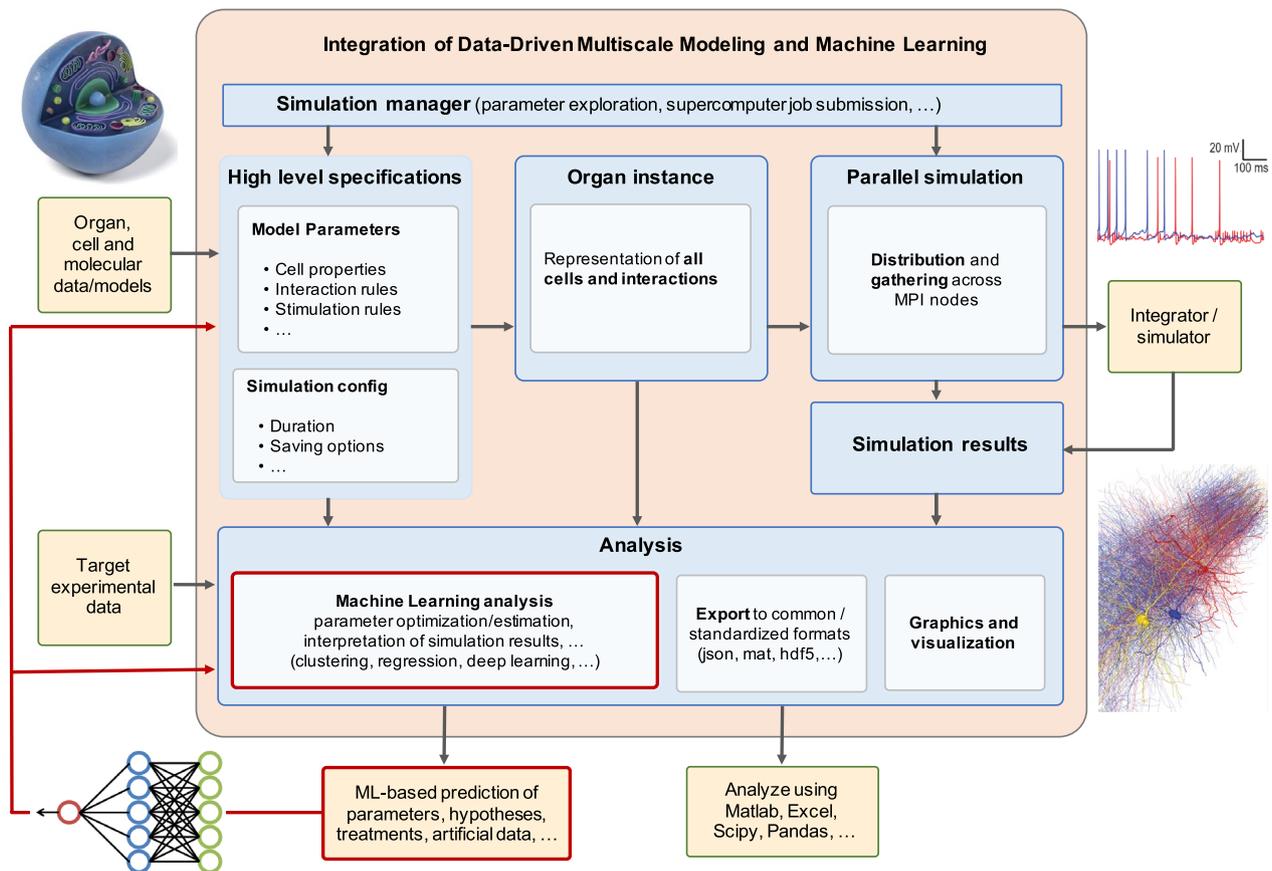


Fig. 6 Data-driven machine learning for multiscale modeling of biomedical systems. By performing organ, cellular, or molecular level simulations and systematically comparing the simulation results against experimental target data using machine learning analysis

including clustering, regression, dimensionality reduction, reinforcement learning, and deep learning we can identify model parameters and generate new hypotheses; adopted with permission from [3]

4.1 State of the Art

Most existing machine learning techniques identify correlations but are agnostic as to causality. In that sense, multiscale modeling complements machine learning: Where machine learning identifies a correlation, multiscale modeling can find causal mechanisms or a mechanistic chain [64]. Personalized medicine, where each patient's disease is considered a unique variant, can benefit from multiscale modeling to follow the particular parameters unique to that patient. Personalized models can then be based on individual differences measured by imaging [65, 86], by genomic or proteomic measures in the patient, or be based on the genomes of infectious agents or tumor cells. This will help in creating digital twins [67], models that incorporate both machine learning and multiscale modeling, for an organ system or a disease process in an individual patient. Using digital twins, we can identify promising therapies before trying them on the one patient [14]. As multiscale modeling attempts to leverage experimental data to gain understanding, machine learning provides a tool to preprocess these data, to automate the construction of models, and to analyze model output [128]. In the following, we focus primarily on applications of machine learning to multiscale modeling.

4.2 Applications and Opportunities

Since machine learning involves computer-based techniques for predicting outcomes or *classifications* based on training data, all data-driven approaches, including mechanistic multiscale modeling, can benefit from application of machine learning.

From simulation experiments to animal experiments In biomedicine, the original area of big-data research was the identification of the human genome, which employed machine learning to construct consistent frames out of genome fragments. Since then, extensions of genomic studies involve large numbers of patients and controls in genome-wide association studies, which compare single nucleotide polymorphisms in patients versus controls [7, 52]. Single nucleotide polymorphisms adjacent to coding sequences suggest which gene product *might* be involved in the disease. Once particular proteins are identified, multiscale modeling can track the effects up through the scales from molecular to cellular to intercellular to organ and organism. In addition to describing dynamics across scales, multiscale modeling can identify how multiple gene products may interact to produce the disease, given that most diseases are polygenic rather than being caused by a single mutation. Simulation experiments can also identify how other allele combinations would produce different disease manifestations and identify degrees of penetrance of a mutation. Simulation can then help us plan animal experiments to add further evidence.

From smaller scales to larger scales A major problem for multiscale modeling is the identification of appropriate model parameters. Ideally, multiscale modeling parameters are all based on consistent experimental measurements. Realistically, biological parameters may be measured in various species, in various cell types, at various temperatures, at various ages, and in different *in vivo* and *in vitro* preparations. Medical multiscale modeling, and medical conclusions, are applied to humans, but are often based on animal models. Additionally, many enzymes have large numbers of isoforms and different phosphorylation states that make generalization problematic. For all these reasons, it is typically necessary to identify the parameters on a multiscale model to ensure realistic dynamics at the higher scales of organization. Machine learning techniques have been used extensively to tune the parameters to replicate these higher-level dynamics. An example of this is the use of genetic algorithms and *evolutionary algorithms* in neural models [16, 29, 76]. Going beyond inference of parameters, recurrent neural networks have been used to identify unknown constitutive relations in ordinary differential equation systems [36].

Evolutionary algorithms are generic population-based optimization algorithms that adopt mechanisms inspired by biological evolution to generate new sampling points for further function evaluation. Strategies include reproduction, mutation, recombination, and selection. Evolutionary algorithms have been used successfully for automatic parameter tuning in multiscale brain modeling [29].

From multiscale modeling to machine learning Multiscale models can provide insight into a biological, biomedical, or behavioral system at a high level of resolution and precision. They can systematically probe different conditions and treatments, faster, more cost-effective, and often beyond what is possible experimentally. This parameter screening naturally produces massive output datasets, which are ideally suited to machine learning analysis. To no surprise, machine learning methods are progressively becoming part of the tool suite to analyze the output of multiscale models [30, 142]. A recent example is the study of *clustering* to study the effects of potential simulated treatments [59, 76].

Clustering is an unsupervised learning method that organizes members of a dataset into groups that share common properties. Typical examples in biomedicine include clustering the effects of simulated treatments [59, 76].

4.3 Open Questions

Classifying simulation and experiment To what extent does the multiscale modeling differ from experiment? Machine learning tools allow us to *cluster* and *classify* the predictions of the multiscale model and systematically compare simulated and experimental datasets. Where simulation and experiment differ, machine learning can identify potential

high-order features and suggest iterative refinements to improve the multiscale model.

Identifying missing information Do the chosen parameters provide a basis set that allows production of the needed higher-scale model dynamics? Multiscale simulations and generative networks can be set up to work in parallel, alongside the experiment, to provide an independent confirmation of parameter sensitivity. For example, the circadian rhythm generators provide relatively simple dynamics but have very complex dependence on numerous underlying parameters, which multiscale modeling can reveal. We could then use *generative models* to identify both the underlying low dimensionality of the dynamics and the high dimensionality associated with parameter variation. Inadequate multiscale models could be identified with failure of generative model predictions.

Generative models are statistical models that capture the joint distribution between a set of observed or latent random variables. A recent study used deep generative models for chemical space exploration and matter engineering [106].

Creating surrogates How can we use generative adversarial networks to create new data sets for testing multiscale models? Conversely, how can we create training or test instances using multiscale modeling to use with deep learning models? A deep learning network could be deployed more widely and provide answers more quickly than a multiscale modeling dynamic simulation, permitting, for example, prediction of pharmaceutical efficacy for patients with particular genetic inheritance in personalized medicine. This would be particularly important for on-body digital twins which need to function rapidly with limited computational resources

Identifying relevant processes and interactions How can we use machine learning to bridge scales? For example, machine learning could be used to explore responses of both immune and tumor cells in cancer based on single-cell data. A multiscale model could then be built on the families of solutions to codify the evolution of the tumor at organ- or metastasis-scale.

Supplementing training data Supervised learning, as used in deep networks, is a powerful technique but requires large amounts of training data. Recent studies have shown that, in the area of object detection in image analysis, simulation augmented by *domain randomization* can be used successfully as a supplement to existing training data [127]. In areas where multiscale models are well-developed, simulation across many parameter has been used as a supplement to existing training data for nonlinear diffusion models to provide physics-informed machine learning [96, 118, 119]. Similarly, multiscale models can be used in biological, biomedical, and behavioral systems to augment insufficient experimental or clinical data sets. Machine learning can provide tools to verify the validity of the simulation results. Multiscale models can then expand the datasets towards

developing machine learning and artificial intelligence applications.

Domain randomization is a technique for randomizing the field of an image so that the true image is also recognized as a realization of this space. Domain randomization has been used successfully to supplement training data [127].

4.4 Potential Challenges and Limitations

Developing new architectures and algorithms inspired by biological learning Chess and go are two games, difficult for humans, which have been solved successfully by artificial intelligence using deep learning. Deep learning has also shown success in image recognition, a function that utilizes large amounts of brain real estate [125]. By contrast, activities that real brains networks are very good at remain elusive. For example, the control systems of a mosquito engaged in evasion and targeting are remarkable considering the small neuronal network involved. This limitation provides opportunities for more detailed brain models to assist in developing new architectures and new learning algorithms. Incorporating spiking [45] or oscillatory dynamics at the mesoscopic or macroscopic levels could inspire novel low-energy architectures and algorithms. Deep learning and *reinforcement learning* were both motivated by brain mechanisms. Understanding biological learning has the potential to inspire novel and improved machine learning architectures and algorithms [39].

Reinforcement learning is a technique that circumvents the notions of supervised learning and unsupervised learning by exploring and combining decisions and actions in dynamic environments to maximize some notion of cumulative reward. Of broad relevance is understanding common learning modes in biological, cognitive, and artificial systems through the lens of reinforcement learning [12, 75].

Identifying disease progression biomarkers and mechanisms There are abundant challenges for data-driven approaches for integrating machine learning and multiscale modeling towards understanding and diagnosing specific disease states. If machine learning could identify predictive disease progression biomarkers, multiscale modeling could follow up to identify mechanisms at each stage of the disease with the ultimate goal to propose interventions that delay, prevent, or revert disease progression.

5 Theory-Driven Approaches

Theory-driven approaches aspire to answer the following questions: How can we leverage structured physical laws and mechanistic models as informative prior information in a machine learning pipeline towards advancing modeling capabilities and expediting the simulation of multiscale

systems? Given imperfect and irregularly sampled data, how do we identify the form and parameters of a governing law, for example an ordinary or partial differential equation, and use it for forecasting? Fig. 7 illustrates a closed-loop integration of theory-driven machine learning and multiscale modeling to accelerate model- and data-driven discovery.

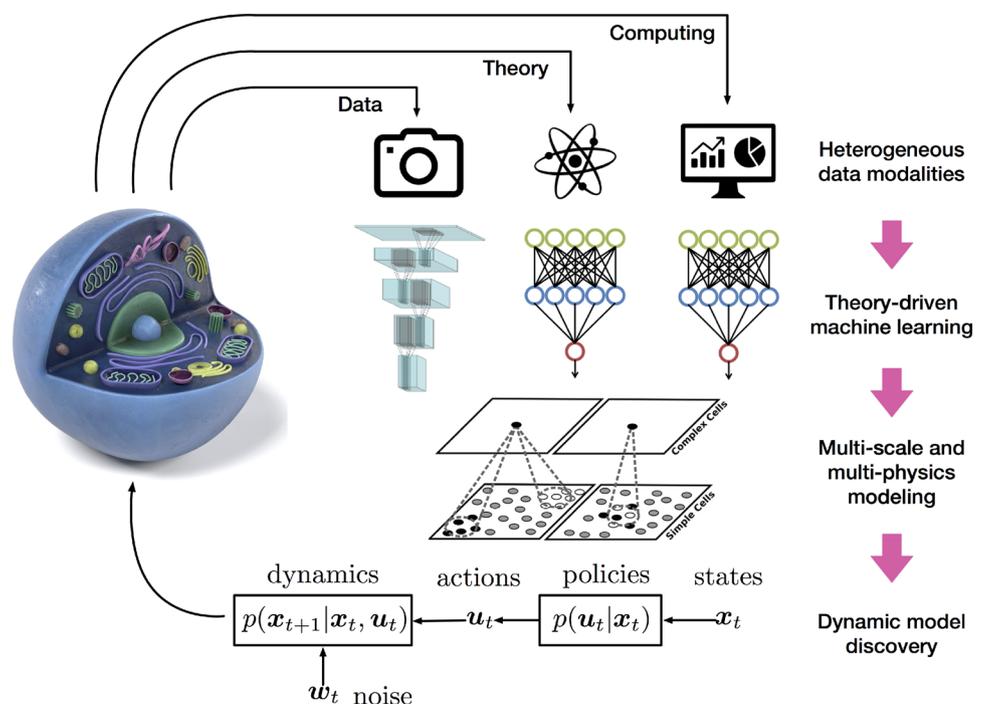
5.1 State of the Art

Theory-driven machine learning is both a decently mature and quickly evolving area of research. It is mature in that methods for learning parameters for a model such as dynamic programming and variational methods have been known and applied for a long time. Although these methods are generally not considered to be tools of machine learning, the difference between them and current machine learning techniques may be as simple as the difference between a deterministic and a stochastic search [138]. Dynamic programming and variational methods are very powerful when we know the form of the model and need to constrain the parameters within a specified range to reproduce experimental observations. Machine learning methods, however, can be very powerful when the model is completely unknown or when there is uncertainty about its form. Mixed cases can exist as well. For example, when modeling the dynamics of a cell, we may know the rate laws that need to be solved and obtain the rate parameters using an optimization algorithm; yet, which reactions are regulated under what conditions may be a mystery for which there are no adequate models.

Theory-driven machine learning can enable the seamless synthesis of physics-based models at multiple temporal and spatial scales. For example, *multi-fidelity* techniques can combine coarse measurements and reduced order models to significantly accelerate the prediction of expensive experiments and large-scale computations [83]. In drug development, for example, we can leverage theory-driven machine learning techniques to integrate information across ten orders of magnitude in space and time towards developing interpretable classifiers that enable us to characterize the potency of pro-arrhythmic drugs [102]. Based on *Gaussian process regression*, these approaches can effectively explore the interplay between drug concentration and drug toxicity by probing the effect of different drugs on ion-channel blockage, cellular action potentials, and electrocardiograms using coarse and low-cost models, anchored by a few, judiciously selected, high-resolution simulations of cardiac electrophysiology [103].

Leveraging probabilistic formulations, theory-driven machine learning techniques can also inform the judicious acquisition of new data and actively expedite tasks involving the exploration of large parameter spaces or the calibration of complex models. For example, we could devise an effective data acquisition policy for choosing the most informative meso-scale simulations that need to be performed to recover detailed constitutive laws as appropriate closures for macroscopic models of complex fluids [143]. Building on recent advances in *automatic differentiation* [9], techniques such as *neural differential equations* [21] are also expanding our capabilities in calibrating complex dynamic models using noisy and irregularly sampled data.

Fig. 7 Theory-driven machine learning for multiscale modeling of biological systems. Theory-driven machine learning can yield data-efficient workflows for predictive modeling by synthesizing prior knowledge and multi-modal data across the scales. Probabilistic formulations enable uncertainty quantification and can guide the judicious acquisition of new data towards dynamic model-refinement; adopted with permission from [3]



Automatic differentiation is a family of techniques to efficiently and accurately evaluate derivatives of numeric functions expressed as computer programs. Not to be confused with symbolic or numerical differentiation, automatic differentiation is an exact procedure for differentiating computer code by redefining the semantics of the operators, composing a complex program to propagate derivatives per the chain rule of differential calculus [9]. Along with the use of graphics processing units, automatic differentiation has been one of the key backbones of the current machine learning revolution. It significantly expedites the prototyping and deployment of predictive data-driven models.

More recently, efforts have been made to directly bake theory into machine learning practice [95]. This enables the construction of predictive models that adhere to the underlying physical principles, including conservation, symmetry, or invariance, while remaining robust even when the observed data are very limited. For example, a recent model only utilized conservation laws of reaction to model the metabolism of a cell. While the exact functional forms of the rate laws was unknown, the equations were solved using machine learning [23]. An intriguing implication of such theory-driven machine learning approaches is related to their ability to leverage auxiliary observations to infer quantities of interest that are difficult to measure in practice [96]. An example includes the use of *physics-informed neural networks* to infer the arterial blood pressure directly and non-invasively from four-dimensional magnetic resonance images of blood velocities and arterial wall displacements by leveraging the known dynamic correlations induced by first principles in fluid and solid mechanics [46]. A common thread between these approaches is that they rely on conventional neural network architectures, including fully connected or convolutional, and constrain them using physical laws as penalty terms in the loss function that drives the learning process. An alternative, yet more laborious route is to design new neural architectures that implicitly remain invariant with respect to the symmetry groups that characterize the dynamics of a given system. A representative example of this school of thought can be found in covariant molecular neural networks; a rotationally covariant neural network architecture for learning the behavior and properties of complex many-body physical systems [6]

Neural differential equations are machine learning models that aim to identify latent dynamic processes from noisy and irregularly sampled time-series data [21]. Leveraging recent advances in automatic differentiation [9], they can efficiently back-propagate through ordinary or partial differential equation solvers to calibrate complex dynamic models and perform forecasting with quantified uncertainty. Examples in biomedicine include predicting in-hospital mortality from irregularly sampled time-series containing measurements from the first 48 h of a different patient's admission to the intensive care unit [99].

5.2 Applications and Opportunities

From models and patient data to personalized medicine. Theory-driven and computational methods have long aspired to provide predictive tools for patient monitoring, diagnostics, and surgical planning. However, high-fidelity predictive models typically incur a large computational cost and rely on tedious calibration procedures that render them impractical for clinical use. Theory-driven machine learning, for example in the form of *multi-fidelity* or *physics-informed* approaches, has the potential to bridge the gap between modeling, predictions, and clinical decision making by enabling the seamless and cost-effective integration of computational models and disparate data modalities, for example from medical imaging, laboratory tests, and patient records. In the age of the Digital Twin, this integration enables new capabilities for assimilating data from medical devices into predictive models that enable the assessment of health risks and inform preventative care and therapeutic strategies on a personalized basis.

From protein biology to physics of the cell In contrast to the clinical case, theory-driven computational methods have a long history of providing high-fidelity, predictive models for many fields in the basic sciences with protein biology as one of the most prominent examples [71, 131]. However, once we move to the scale where non-equilibrium phenomena occur, the rate parameters necessary to solve either the full mass action rate law or even approximations such as the Michealis–Menten equation have been difficult and expensive to obtain, as they depend on careful in vitro kinetic studies. The Costello–Garcia–Martin study mentioned above is a significant step, but the functional form of the rate laws, which describe the underlying physics remain unknown. Accordingly, a grand challenge application that is ripe for further development is combining theory-driven machine learning with multi-scale modeling to understand the physics of the cell, especially with a view towards the emergence of function. We could, for example, generate high-fidelity data from simulations based on maximum entropy assumptions, and then use these data to learn feasible solution spaces.

From interpolation to extrapolation When data can be generated from theory-driven models, machine learning techniques have enjoyed immense success. Such tasks are usually based on interpolation in that the input domain is well specified, and we have sufficient data to construct models that can interpolate between the dots. This is the regime where discriminative, black box methods such as deep learning perform best. When extrapolation is needed instead, the introduction of prior knowledge and appropriate inductive biases through theory-driven methods can effectively steer a machine learning algorithm towards physically consistent solutions. Theory-driven machine learning approaches present a unique opportunity for leveraging the

domain knowledge and mechanistic insight brought by the multi-scale modeling community to develop novel learning algorithms with enhanced robustness, data-efficiency, and generalization performance in data-limited regimes. We anticipate such developments to be crucial for leveraging the full potential of machine learning in advancing multi-scale modeling for biological, biomedical, and behavioral systems.

5.3 Open Questions

Elucidating mechanisms Can theory-driven machine learning approaches enable the discovery of interpretable models that can not only explain data, but also elucidate mechanisms, distill causality, and help us probe interventions and counterfactuals in complex multi-scale systems? For instance, causal inference generally uses various statistical measures such as partial correlation to infer causal influence. If instead, the appropriate statistical measure were known from the physics such as a statistical odds ratio from thermodynamics, would the causal inference be more accurate or interpretable as a mechanism?

Understanding the emergence of function Understanding the emergence of function is of critical importance in biology and medicine, environmental studies, biotechnology and other biological sciences. The study of emergence necessitates the ability to model collective action on a lower scale to predict how the phenomena on the higher scale emerge from the collective action. Can theory-driven machine learning, combined with sparse and indirect measurements of the phenomena, produce a mechanistic understanding of how biological phenomena emerge?

Exploring massive design spaces Can theory-driven machine learning approaches uncover meaningful and compact representations for complex inter-connected processes, and, subsequently, enable the cost-effective exploration of vast combinatorial spaces? A typical example is the design of bio-molecules with target properties in drug development.

Predicting uncertainty *Uncertainty quantification* is the back-bone of decision making. Can theory-driven machine learning approaches enable the reliable characterization of predictive uncertainty and pinpoint its sources? The quantification of uncertainty has many practical applications such as decision making in the clinic, the robust design of synthetic biology pathways, drug target identification and drug risk assessment, as well as guiding the informed, targeted acquisition of new data.

Selecting the appropriate tools Is deep learning necessary in theory-driven learning? In principle, the more domain knowledge is incorporated into the model the less needs to be learned and the easier the computing task will become. More knowledge will enable researchers to take on even greater challenges, which, in turn, may require more learning. It is likely that the applications will utilize a range

of techniques, from dynamic programming to variational methods to standard machine learning to deep learning. Is high performance computing required when theory-driven models are employed? The answer here probably depends on the application and the depth of the model used in learning, with the larger the multi-scale model, the more computing necessary.

5.4 Potential Challenges and Limitations

A major challenge in theory-driven approaches towards understanding biological, biomedical, and behavioral systems, is obtaining sufficient data to answer the driving question of interest.

Combining low- and high-resolution data Can theory-driven machine learning be utilized to bridge the gap between qualitative 'omics data and the quantitative data needed for prediction? For example, RNASeq data has become fairly quantitative, but the amplification of transcripts using polymerase chain reaction can add uncertainty to the final measures. Proteomics and metabolomics assays can be quite quantitative when using nuclear magnetic resonance or multiple reaction monitoring [50], but nuclear magnetic resonance has a relatively narrow dynamic range for quantification and multiple reaction monitoring mass spectrometry is not high throughput. Similarly, isotope labeling studies such as metabolic flux analysis [137] and absolute quantitation by mass spectroscopy [10, 62, 80] provide highly valuable information, but are low-throughput and relatively costly. High throughput methods such as shotgun proteomics [139] or global metabolomics determine whether a protein, peptide, or metabolite is observed, but not whether it is actually present since different species have different detectability characteristics. For these reasons, the use of high-throughput biological data in machine learning remains a challenge, but combining theory-driven approaches with *multi-fidelity* data would help reduce the uncertainty in the analysis.

Minimizing data bias Can arrhythmia patients trust a neural net controller embedded in a pacemaker that was trained under different environmental conditions than the ones during their own use? Training data come at various scales and different levels of fidelity. These data are typically generated by existing models, experimental assays, historical data, and other surveys, all of which come with their own biases. As machine learning algorithms can only be as good as the data they have seen, proper care needs to be taken to safe-guard against biased models and biased data-sets. Theory-driven approaches could provide a rigorous foundation to estimate the range of validity, quantify the uncertainty, and characterize the level of confidence of machine learning based approaches.

Knowing the risk of non-physical predictions Can new data fill the gap when the multi-scale model lacks a clean separation between the fast and slow temporal scales or between the small and large spatial scales? From a conceptual point of view, this is a problem of supplementing the set of physics-based equations with constitutive equations, an approach, which has long been used in traditional engineering disciplines. While data-driven methods can provide solutions that are not constrained by preconceived notions or models, their predictions should not violate the fundamental laws of physics. Sometimes it is difficult to determine whether the model predictions obey these fundamental laws. This is especially the case when the functional form of the model cannot be determined explicitly, for instance in deep learning. This makes it difficult to know whether the analysis predicts the correct answer for the right reasons. There are well-known examples of deep learning neural networks that appear to be highly accurate, but make highly inaccurate predictions when faced with data outside their training regime [78], and others that make highly inaccurate predictions based on seemingly minor changes to the target data [116]. Integrating machine learning and multiscale models that a priori satisfy the fundamental laws of physics would help address this limitation.

6 Conclusion

Many exciting new applications emerge at the interface of machine learning and multiscale modeling. Immediate applications in multiscale modeling include system identification, parameter identification, sensitivity analysis, and uncertainty quantification and applications in machine learning are physics-informed neural networks. Integrating machine learning and multiscale modeling can have a massive impact in the biological, biomedical, and behavioral sciences, both in diagnostics and prognostics, and this review has only touched the surface. Undeniably, applications become more and more sophisticated and have to be increasingly aware of the inherent limitations of overfitting and data bias. A major challenge to make progress in this field will be to increase transparency, rigor, and reproducibility. We hope that this review will stimulate discussion within the community of computational mechanics and reach out to other disciplines including mathematics, statistics, computer science, artificial intelligence, and precision medicine to join forces towards personalized predictive modeling in biomedicine.

Acknowledgements The authors acknowledge support of the National Institutes of Health Grants U01 HL116330 (Alber), R01 AR074525 (Buganza Tepole), U01 EB022546 (Cannon), R01 CA197491 (De), U24 EB028998 (Dura-Bernal), U01 HL116323 and U01 HL142518 (Karniadakis), U01 EB017695 (Lytton), R01 EB014877 (Petzold) and U01 HL119578 (Kuhl), as well as DARPA Grant HR0011199002 and

Toyota Research Institute Grant 849910, (both Garikipati). This work was inspired by the 2019 Symposium on Integrating Machine Learning with Multiscale Modeling for Biological, Biomedical, and Behavioral Systems (ML-MSM) as part of the Interagency Modeling and Analysis Group (IMAG), and is endorsed by the Multiscale Modeling (MSM) Consortium, by the U.S. Association for Computational Mechanics (USACM) Technical Trust Area Biological Systems, and by the U.S. National Committee on Biomechanics (USNCB). The authors acknowledge the active discussions within these communities.

Compliance with ethical standards

Conflict of Interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- Ahmed OJ, Sudhakar SK (2019) High frequency activity during stereotyped low frequency events might help to identify the seizure onset zone. *Epilepsy Curr* 19(3):184–186
- Ahmed OJ, John TT (2019) A straw can break a neural network's back and lead to seizures but only when delivered at the right time. *Epilepsy Currents* 19(2):115–116
- Alber M, Buganza Tepole A, Cannon W, De S, Dura-Bernal S, Garikipati K, Karniadakis G, Lytton WW, Perdikaris P, Petzold L, Kuhl E (2019) Integrating machine learning and multiscale modeling: perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. *NPJ Digit Med* 2:115
- Ambrosi D, Ateshian GA, Arruda EM, Cowin SC, Dumais J, Goriely A, Holzapfel GA, Humphrey JD, Kemkemer R, Kuhl E, Olberding JE, Taber LA, Garikipati K (2011) Perspectives on biological growth and remodeling. *J Mech Phys Solids* 59:863–883
- Ambrosi D, BenAmar M, Cyron CJ, DeSimone A, Goriely A, Humphrey JD, Kuhl E (2019) Growth and remodelling of living tissues: perspectives, challenges, and opportunities. *J R Soc Interface* 16:20190233
- Anderson B, Hy TS, Kondor R (2019) ArXiv preprint [arXiv:1906.04015](https://arxiv.org/abs/1906.04015)
- Athreya AP, Neavin D, Carrillo-Roa T, Skime M, Biernacka J, Frye MA, Rush AJ, Wang L, Binder EB, Iyer RK, Weinsilboum RM, Bobo WV (2019) Pharmacogenomics-driven prediction of antidepressant treatment outcomes: a machine learning approach with multi-trial replication. *Clin Pharmacol Thera* 106:855–865. <https://doi.org/10.1002/cpt.1482>
- Baillargeon B, Rebelo N, Fox DD, Taylor RL, Kuhl E (2014) The living heart project: a robust and integrative simulator for human heart function. *Eur J Mech A/Solids* 48:38–47
- Baydin AG, Pearlmutter BA, Radul AA, Siskind JM (2018) Automatic differentiation in machine learning: a survey. *J Mach Learn Res* 18:153
- Bennett BD, Kimball EH, Gao M, Osterhout R, Van Dien SJ, Rabinowitz JD (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat Chem Biol* 5(8):593–599
- Booth V, Xique IJ, Diniz Behn CG (2017) One-dimensional map for the circadian modulation of sleep in a sleep-wake regulatory network model for human sleep. *SIAM J Appl Dyn Syst* 16:1089–1112
- Botvinick M, Ritter S, Wang JX, Kurth-Nelson Z, Blundell C, Hassabis D (2019) Reinforcement learning, fast and slow. *Trends Cognit Sci* 23:408–422

13. Brunton SL, Proctor JL, Kutz JN (2016) Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc Natl Acad Sci* 113:3932–3937
14. Bruynseels K, Santoni de Sio F, van den Hoven J (2018) Digital Twins in health care: ethical implications of an emerging engineering paradigm. *Front Genet* 9:31
15. Buehler MJ (2006) Atomistic and continuum modeling of mechanical properties of collagen: elasticity, fracture, and self-assembly. *J Mater Res* 21:19471961
16. Carlson KD, Nageswaran JM, Dutt N, Krichmar JL (2014) An efficient automated parameter tuning framework for spiking neural networks. *Front Neurosci* 8(10):00010
17. Cao YH, Eisenberg MC (2018) Practical unidentifiability of a simple vector-borne disease model: implications for parameter estimation and intervention assessment. *Epidemics* 25:89–100
18. Chabiniok R, Wang V, Hadjicharalambous M, Asner L, Lee J, Sermesant M, Kuhl E, Young A, Moireau P, Nash M, Chapelle D, Nordsletten DA (2016) Multiphysics and multiscale modeling, data-model fusion and integration of organ physiology in the clinic: ventricular cardiac mechanics. *Interface Focus* 6:20150083
19. Champion KP, Brunton SL, Kutz JN (2019) Discovery of nonlinear multiscale systems: sampling strategies and embeddings. *SIAM J Appl Dyn Syst* 18:312–333
20. Chandran PL, Barocas VH (2007) Deterministic material-based averaging theory model of collagen gel micromechanics. *J Biomech Eng* 129:137147
21. Chen TQ, Rubanova Y, Bettencourt J, Duvenaud DK (2018) Neural ordinary differential equations. In: *Advances in neural information processing systems*, pp 6571–6583
22. Conti S, Müller S, Ortiz M (2018) Data-driven problems in elasticity. *Arch Ration Mech Anal* 229:79–123
23. Costello Z, Martin HG (2018) A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data. *NPJ Syst Biol Appl* 4:19
24. Cuperlovic-Culf M (2018) Machine learning methods for analysis of metabolic data and metabolic pathway modeling. *Metabolites* 8:4
25. De S, Wongmuk H, Kuhl E (eds) (2014) *Multiscale modeling in biomechanics and mechanobiology*. Springer, Berlin
26. Deist TM, Patti A, Wang Z, Krane D, Sorenson T, Craft D (2019) Simulation assisted machine learning. *Bioinformatics* 35:4072–4080. <https://doi.org/10.1093/bioinformatics/btz199>
27. Deo RC (2015) Machine learning in medicine. *Circulation* 132:1920–1930
28. DeWoskin D, Myung J, Belle MD, Piggins HD, Takumi T, Forger DB (2015) Distinct roles for GABA across multiple timescales in mammalian circadian timekeeping. *Proc Natl Acad Sci* 112:E2911
29. Dura-Bernal S, Neymotin SA, Kerr CC, Sivagnanam S, Majumdar A, Francis JT, Lytton WW (2017) Evolutionary algorithm optimization of biological learning parameters in a biomimetic neuroprosthesis. *IBM J Res Dev* 61(6):114
30. Dura-Bernal S, Suter BA, Gleeson P, Cantarelli M, Quintana A, Rodriguez F, Lytton WW (2019) NetPyNE, a tool for data-driven multiscale modeling of brain circuits. *Elife* 8:e44494. <https://doi.org/10.7554/eLife.44494.001>
31. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S (2017) Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542:115–118
32. Fritzen F, Hodapp M (2016) The finite element square reduced (FE2R) method with GPU acceleration: towards three-dimensional two-scale simulations. *Int J Numer Methods Eng* 107:853881
33. Geers MGD, Kouznetsova VG, Brekelmans WAM (2010) Multi-scale computational homogenization: trends and challenges. *J Comput Appl Math* 234:21752182
34. Gerlee P, Kim E, Anderson ARA (2015) Bridging scales in cancer progression: mapping genotype to phenotype using neural networks. *Semin Cancer Biol* 30:3041
35. Gillespie DT (2007) Stochastic simulation of chemical kinetics. *Ann Rev Phys Chem* 58:3555
36. Hagge T, Stinis P, Yeung E, Tartakovsky AM (2017) Solving differential equations with unknown constitutive relations as recurrent neural networks. [arXiv:1710.02242](https://arxiv.org/abs/1710.02242)
37. Han J, Jentzen A, Weinan E (2018) Solving high-dimensional partial differential equations using deep learning. *Proc Natl Acad Sci* 115(34):8505–8510
38. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, Ng AY (2019) Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med* 25:65–69
39. Hassabis D, Kumaran D, Summerfield C, Botvinick M (2017) Neuroscience-inspired artificial intelligence. *Neuron* 95(2):245258
40. Hicks JL, Althoff T, Sosic R, Kuhar P, Bostjancic B, King AC, Leskovec J, Delp SL (2019) Best practices for analyzing large-scale health data from wearables and smartphone apps. *NPJ Digit Med* 2:45
41. Huan X, Marzouk YM (2013) Simulation-based optimal experimental design for nonlinear systems. *J Comput Phys* 232:288–317
42. Hunt CA, Erdemir A, Lytton WW, Mac Gabhann F, Sander EA, Transtrum MK, Mulugeta L (2018) The spectrum of mechanism-oriented models and methods for explanations of biological phenomena. *Processes* 6(5):56
43. Hunter PJ, Borg TK (2003) Integration from proteins to organs: the physiome project. *Processes* 4:237–243
44. Kennedy M, O'Hagan A (2001) Bayesian calibration of computer models (with discussion). *J Roy Stat Soc B* 63:425–464
45. Kim R, Li Y, Sejnowski TJ (2019) Simple framework for constructing functional spiking recurrent neural networks. *Proc Natl Acad Sci* 116:22811–22820. <https://doi.org/10.1101/579706>
46. Kissas G, Yang Y, Hwuang E, Witschey WR, Detre JA, Perdikaris P (2019) Machine learning in cardiovascular flows modeling: predicting pulse wave propagation from non-invasive clinical measurements using physics-informed deep learning. *ArXiv preprint arXiv:1905.04817*
47. Kitano H (2002) Systems biology: a brief overview. *Science* 295:16621664
48. Kouznetsova V, Brekelmans WAM, Baaijens FPT (2001) Approach to micro-macro modeling of heterogeneous materials. *Comput Mech* 27:3748
49. Kouznetsova VG, Geers MGD, Brekelmans WAM (2004) Multi-scale second-order computational homogenization of multi-phase materials: a nested finite element solution strategy. *Comput Methods Appl Mech Eng* 193:55255550
50. Lange V, Picotti P, Domon B, Aebersold R (2008) Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol Syst Biol* 4:222
51. Le BA, Yvonne J, He QC (2015) Computational homogenization of nonlinear elastic materials using neural networks. *Int J Numer Methods Eng* 104:10611084
52. Leal LG, David A, Jarvelin M-R, Sebert S, Ruddock M, Karhunen V, Sternberg MJE (2019) Identification of disease-associated loci using machine learning for genotype and network data integration. *Bioinformatics* 35:5182–5190. <https://doi.org/10.1093/bioinformatics/btz310>

53. Leary SJ, Bhaskar A, Keane AJ (2003) A knowledge-based approach to response surface modelling in multifidelity optimization. *J Glob Optim* 26(3):297319
54. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444
55. Lee T, Turin SY, Gosain AK, Bilionis I, Buganza Tepole A (2018) Propagation of material behavior uncertainty in a nonlinear finite element model of reconstructive surgery. *Biomech Model Mechanobiol* 17(6):1857–18731
56. Lee T, Gosain AK, Bilionis I, Buganza Tepole A (2019) Predicting the effect of aging and defect size on the stress profiles of skin from advancement, rotation and transposition flap surgeries. *J Mech Phys Solids* 125:572590
57. Lee T, Bilionis I, Buganza Tepole A (2020) Propagation of uncertainty in the mechanical and biological response of growing tissues using multi-fidelity Gaussian process regression. *Comput Methods Appl Mech Eng* 359:112724
58. Liang G, Chandrashekhara K (2008) Neural network based constitutive model for elastomeric foams. *Eng Struct* 30:20022011
59. Lin C-L, Choi S, Haghighi B, Choi J, Hoffman EA (2018) Cluster-guided multiscale lung modeling via machine learning. In: *Handbook of materials modeling*, vol 120. https://doi.org/10.1007/978-3-319-50257-1_98-1
60. Liu Y, Zhang L, Yang Y, Zhou L, Ren L, Liu R, Pang Z, Deen MJ (2019) A novel cloud-based framework for the elderly healthcare services using Digital Twin. *IEEE Access* 7:49088–49101
61. Liu Z, Wu CT, Koishi M (2019) A deep material network for multiscale topology learning and accelerated nonlinear modeling of heterogeneous materials. *Comput Methods Appl Mech Eng* 345:11381168
62. Lu W, Su X, Klein MS, Lewis IA, Fiehn O, Rabinowitz JD (2017) Metabolite measurement: pitfalls to avoid and practices to follow. *Ann Rev Biochem* 86:277–304
63. Lueberding S, Krueger N, Kerscher M (2014) Mechanical properties of human skin in vivo: a comparative evaluation in 300 men and women. *Skin Res Technol* 20:127135
64. Lytton WW, Arle J, Bobashev G, Ji S, Klassen TL, Marmarelis VZ, Sanger TD (2017) Multiscale modeling in the clinic: diseases of the brain and nervous system. *Brain Inform* 4(4):219230
65. Lytton WW (2017) Computers, causality and cure in epilepsy. *Brain* 140(3):516–526
66. Madireddy S, Sista B, Vemaganti K (2015) A Bayesian approach to selecting hyperelastic constitutive models of soft tissue. *Comput Methods Appl Mech Eng* 291:102122
67. Madni AM, Madni CC, Lucerno SD (2019) Leveraging Digital Twin technology in model-based systems engineering. *Systems* 7:1–13
68. Mangan NM, Brunton SL, Proctor JL, Kutz JN (2016) Inferring biological networks by sparse identification of nonlinear dynamics. *IEEE Trans Mol Biol Multi-Scale Commun* 2:52–63
69. Mangan NM, Askham T, Brunton SL, Kutz NN, Proctor JL (2019) Model selection for hybrid dynamical systems via sparse regression. *Proceedings R Soc A Math Phys Eng Sci* 475:20180534
70. Marino M, Vairo G (2012) Stress and strain localization in stretched collagenous tissues via a multiscale modelling approach. *Comput Methods Biomech Biomed Eng* 17:1130
71. McCammon JA, Gelin BR, Karplus M (1977) Dynamics of folded proteins. *Nature* 267:585–590
72. Mihai LA, Woolley TE, Gorieli A (2018) Stochastic isotropic hyperelastic materials: constitutive calibration and model selection. *Proc R Soc A/Math Phys Eng Sci* 474:0858
73. Myung J, Hong S, DeWoskin D, De Schutter E, Forger DB, Takumi T (2015) GABA-mediated repulsive coupling between circadian clock neurons encodes seasonal time. *Proc Natl Acad Sci* 112:E2920
74. Nazari F, Pearson AT, Nor JE, Jackson TL (2018) A mathematical model for IL-6-mediated, stem cell driven tumor growth and targeted treatment. *PLOS Comput Biol* 14:e1005920. <https://doi.org/10.1371/journal.pcbi.1005920>
75. Neftci EO, Averbeck BB (2019) Reinforcement learning in artificial and biological systems. *Nat Mach Intell* 1:133–143
76. Neymotin SA, Dura-Bernal S, Moreno H, Lytton WW (2016) Computer modeling for pharmacological treatments for dystonia. *Drug Discov Today Disease Models* 19:5157
77. Neymotin SA, Suter BA, Dura-Bernal S, Shepherd GMG, Migliore M, Lytton WW (2016) Optimizing computer models of corticospinal neurons to replicate in vitro dynamics. *J Neurophysiol* 117:148–162
78. Nguyen A, Yosinski J, Clune J (2015) Deep neural networks are easily fooled: high confidence predictions for unrecognizable images. In: 2015 IEEE conference on computer vision and pattern recognition
79. Ognjanovski N, Broussard C, Zochowski M, Aton SJ (2018) Hippocampal network oscillations drive memory consolidation in the absence of sleep. *Cereb Cortex* 28(10):1–13
80. Park JO, Rubin SA, Amador-Noguz D, Fan J, Shlomi T, Rabinowitz JD (2016) Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. *Nat Chem Biol* 12(7):482–489
81. Peirlinck M, Sahli Costabal F, Sack KL, Choy JS, Kassab GS, Guccione JM, De Beule M, Segers P, Kuhl E (2019) Using machine learning to characterize heart failure across the scales. *Biomech Model Mechanobiol* 18:1987–2001. <https://doi.org/10.1007/s10237-019-01190-w>
82. Peng GCY (2016) Moving toward model reproducibility and reusability. *IEEE Trans Biomed Eng* 63:1997–1998
83. Perdikaris P, Karniadakis GE (2016) Model inversion via multi-fidelity Bayesian optimization: a new paradigm for parameter estimation in haemodynamics, and beyond. *J R Soc Interface* 13(118):20151107
84. Perdikaris P, Raissi M, Damianou A, Lawrence ND, Karniadakis GE (2016) Nonlinear information fusion algorithms for robust multi-fidelity modeling. *Proc R Soc A/Math Phys Eng Sci* 473:0751
85. Poggio T, Mhaskar H, Rosasco L, Miranda B, Liao Q (2017) Why and when can deep-but not shallow-networks avoid the curse of dimensionality: a review. *Int J Autom Comput* 14:503–519
86. Proix T, Bartolomei F, Guye M, Jirsa VK (2017) Individual brain structure and modeling predict seizure propagation. *Brain* 140:651–654
87. Puentes-Mestri C, Roach J, Niethard N, Zochowski M, Aton SJ (2019) How rhythms of the sleeping brain tune memory and synaptic plasticity. *Sleep* zsz42:095. <https://doi.org/10.1093/sleep/zsz095>
88. Quade M, Abel M, Kutz JN, Brunton SL (2018) Sparse identification of nonlinear dynamics for rapid model recovery. *Chaos* 28:063116
89. Raina A, Linder C (2014) A homogenization approach for non-woven materials based on fiber undulations and reorientation. *J Mech Phys Solids* 65:1234
90. Raissi M, Perdikaris P, Karniadakis GE (2017) Inferring solutions of differential equations using noisy multi-fidelity data. *J Comput Phys* 335:736746
91. Raissi M, Perdikaris P, Karniadakis GE (2017) Machine learning of linear differential equations using Gaussian processes. *J Comput Phys* 348:683–693
92. Raissi M, Perdikaris P, Karniadakis GE (2017) Physics informed deep learning (Part I): data-driven solutions of nonlinear partial differential equations. *ArXiv preprint arXiv:1711.10561*
93. Raissi M, Perdikaris P, Karniadakis GE (2017) Physics informed deep learning (Part II): data-driven discovery of nonlinear partial differential equations. *ArXiv preprint arXiv:1711.10566*

94. Raissi M, Karniadakis GE (2018) Hidden physics models: machine learning of nonlinear partial differential equations. *J Comput Phys* 357:125–141
95. Raissi M, Yazdani A, Karniadakis GE (2018) Hidden fluid mechanics: a Navier-Stokes informed deep learning framework for assimilating flow visualization data. ArXiv preprint [arXiv:1808.04327](https://arxiv.org/abs/1808.04327)
96. Raissi M, Perdikaris P, Karniadakis GE (2019) Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J Comput Phys* 378:686707
97. Rhodes SJ, Knight GM, Kirschner DE, White RG, Evans TG (2019) Dose finding for new vaccines: the role for immunostimulation/immunodynamic modelling. *J Theor Biol* 465:51–55
98. Riley P (2019) Three pitfalls to avoid in machine learning. *Nature* 572:27–28
99. Rubanova Y, Chen RTQ, Duvenaud D (2019) Latent odes for irregularly-sampled time series. ArXiv preprint [arXiv:1907.03907](https://arxiv.org/abs/1907.03907)
100. Rudy SH, Brunton SL, Proctor JL, Kutz JN (2017) Data-driven discovery of partial differential equations. *Sci Adv* 3(4):e1602614
101. Sahli Costabal F, Choy JS, Sack KL, Guccione JM, Kassab GS, Kuhl E (2019) Multiscale characterization of heart failure. *Acta Biomater* 86:66–76
102. Sahli Costabal F, Matsuno K, Yao J, Perdikaris P, Kuhl E (2019) Machine learning in drug development: characterizing the effect of 30 drugs on the QT interval using Gaussian process regression, sensitivity analysis, and uncertainty quantification. *Comput Methods Appl Mech Eng* 348:313–333
103. Sahli Costabal F, Perdikaris P, Kuhl E, Hurtado DE (2019) Multi-fidelity classification using Gaussian processes: accelerating the prediction of large-scale computational models. *Comput Methods Appl Mech Eng* 357:112602
104. Sahli Costabal F, Seo K, Ashley E, Kuhl E (2020) Classifying drugs by their arrhythmogenic risk using machine learning. *Biophys J* 118:1–12
105. Sahli Costabal F, Yang Y, Perdikaris P, Hurtado DE, Kuhl E (2020) Physics-informed neural networks for cardiac activation mapping. *Front Phys*. <https://doi.org/10.3389/fphy.2020.00042>
106. Sanchez-Lengeling B, Aspuru-Guzik A (2018) Inverse molecular design using machine learning: generative models for matter engineering. *Science* 361:360–365
107. Sander EA, Stylianopoulos T, Tranquillo RT, Barocas VH (2009) Image-based multiscale modeling predicts tissue-level and network-level fiber reorganization in stretched cell-compacted collagen gels. *Proc Natl Acad Sci* 106:1767517680
108. Sankaran S, Moghadam ME, Kahn AM, Tseng EE, Guccione JM, Marsden AL (2012) Patient-specific multiscale modeling of blood flow for coronary artery bypass graft surgery. *Ann Biomed Eng* 40(10):2228–2242
109. Schoeberl B, Eichler-Jonsson C, Gilles ED, Miller G (2002) Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat Biotechnol* 20:370375
110. Shaked I, Oberhardt MA, Atias N, Sharan R, Ruppin E (2018) Metabolic network prediction of drug side effects. *Cell Systems* 2:209213
111. Shenoy VB, Miller RE, Tadmor EB, Rodney D, Phillips R, Ortiz M (1999) An adaptive finite element approach to atomic scale mechanics-the quasicontinuum method. *J Mech Phys Solids* 47:611642
112. Snowden TJ, van der Graaf PH, Tindall MJ (2017) Methods of model reduction for large-scale biological systems: a survey of current methods and trends. *Bull Math Biol* 79(7):14491486
113. Song D, Hugenberg N, Oberai AA (2019) Three-dimensional traction microscopy with a fiber-based constitutive model. *Comput Methods Appl Mech Eng* 357:112579
114. Southern J, Pitt-Francis J, Whiteley J, Stokeley D, Kobashi H, Nobes R, Kadooka Y, Gavaghan D (2008) Multi-scale computational modelling in biology and physiology. *Prog Biophys Mol Biol* 96:6089
115. Stelling J, Gilles ED (2004) Mathematical modeling of complex regulatory networks. *NanoBioscience, IEEE Transactions* 3:172179
116. Szegedy C, Zaremba W, Sutskever I, Bruna J, Erhan D, Goodfellow I, Fergus R (2013) Intriguing properties of neural networks. ArXiv preprint [arXiv:1312.6199](https://arxiv.org/abs/1312.6199)
117. Tank A, Covert I, Foti N, Shojaie A, Fox E (2018) Neural Granger causality for nonlinear time series. [arXiv:1802.05842](https://arxiv.org/abs/1802.05842)
118. Tartakovsky AM, Marrero CO, Perdikaris P, Tartakovsky GD, Barajas-Solano D (2018) Learning Parameters and constitutive relationships with physics informed deep neural networks. [arXiv:1808.03398](https://arxiv.org/abs/1808.03398)
119. Tartakovsky G, Tartakovsky AM, Perdikaris P (2018) Physics informed deep neural networks for learning parameters with non-Gaussian non-stationary statistics. [arXiv:2018agufm.h21j1791t](https://arxiv.org/abs/2018agufm.h21j1791t)
120. Taylor CA, Figueroa CA (2009) Patient-specific modeling of cardiovascular mechanics. *Annu Rev Biomed Eng* 11:109134
121. Teichert G, Garikipati K (2019) Machine learning materials physics: surrogate optimization and multi-fidelity algorithms predict precipitate morphology in an alternative to phase field dynamics. *Comput Methods Appl Mech Eng* 344:666–693
122. Teichert GH, Natarajan AR, Van der Ven A, Garikipati K (2019) Machine learning materials physics: integrable deep neural networks enable scale bridging by learning free energy functions. *Comput Methods Appl Mech Eng* 353:201–216
123. Topol EJ (2019) Deep medicine: how artificial intelligence can make healthcare human again. Hachette Book Group, New York
124. Topol EJ (2019) High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 25:44–56
125. Topol EJ (2019) Deep learning detects impending organ injury. *Nature* 572:36–37
126. Tran JS, Schiavazzi DE, Kahn AM, Marsden AL (2019) Uncertainty quantification of simulated biomechanical simuli in coronary artery bypass grafts. *Comput Methods Appl Mech Eng* 345:402–428
127. Tremblay J, Prakash A, Acuna D, Brophy M, Jampani V, Anil C, Birchfield S (2018) Training deep networks with synthetic data: bridging the reality gap by domain randomization. [arXiv:1804.06516](https://arxiv.org/abs/1804.06516)
128. Vu MAT, Adali T, Ba D, Buzsaki G, Carlson D, Heller K, Dzirasa K (2018) A shared vision for machine learning in neuroscience. *J Neurosci* 38(7):16011607
129. Wagner GJ, Liu WK (2003) Coupling of atomistic and continuum simulations using a bridging scale decomposition. *J Comput Phys* 190:249274
130. Wang Z, Huan X, Garikipati K (2019) Variational system identification of the partial differential equations governing the physics of pattern-formation: inference under varying fidelity and noise. *Comput Methods Appl Mech Eng* 356:44–74
131. Warschel A, Levitt M (1976) Theoretical studies of enzymic reactions—dielectric, electrostatic and steric stabilization of carbonium-ion in reaction of lysozyme. *J Mol Biol* 103:227–249
132. Weickenmeier J, Kuhl E, Goriely A (2018) The multiphysics of prion-like diseases: progression and atrophy. *Phys Rev Lett* 121:158101
133. Weickenmeier J, Jucker M, Goriely A, Kuhl E (2019) A physics-based model explains the prion-like features of neurodegeneration in Alzheimers disease, Parkinsons disease, and amyotrophic lateral sclerosis. *J Mech Phys Solids* 124:264–281

134. Weinan E, Han J, Jentzen A (2017) Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. *Commun Math Stat* 5(4):349–380
135. Weinan E, Yu B (2018) The deep Ritz method: a deep learning-based numerical algorithm for solving variational problems. *Commun Math Stat* 6(1):1–12
136. White R, Peng G, Demir S (2009) Multiscale modeling of biomedical, biological, and behavioral systems. *IEEE Eng Med* 28:12–13
137. Wiechert W (2001) 13C metabolic flux analysis. *Metab Eng* 2:195–206
138. Wiering M, van Otterlo M (2013) Reinforcement learning and Markov decision processes. In: *Reinforcement learning*, pp 3–39
139. Wolters DA, Washburn MP, Yates JR III (2001) An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem* 73(23):5683–5090
140. Yang L, Zhang D, Karniadakis GE (2018) Physics-informed generative adversarial networks for. [arXiv:1811.02033](https://arxiv.org/abs/1811.02033) [StatML]
141. Yang Y, Perdikaris P (2019) Adversarial uncertainty quantification in physics-informed neural networks. *J Comput Phys* 394:136–152
142. Zangooei MH, Habibi J (2017) Hybrid multiscale modeling and prediction of cancer cell behavior. *PLoS ONE* 12(8):e0183810
143. Zhao L, Li Z, Caswell B, Ouyang J, Karniadakis GE (2018) Active learning of constitutive relation from mesoscopic dynamics for macroscopic modeling of non-Newtonian flows. *J Comput Phys* 363:116–127

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.