INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH AND ANALYSIS

ISSN(print): 2643-9840, ISSN(online): 2643-9875

Volume 09 Issue 07 July 2026

DOI: 10.47191/ijmra/v9-i7-28, Impact Factor: 8.685

Page No. 4405-4421

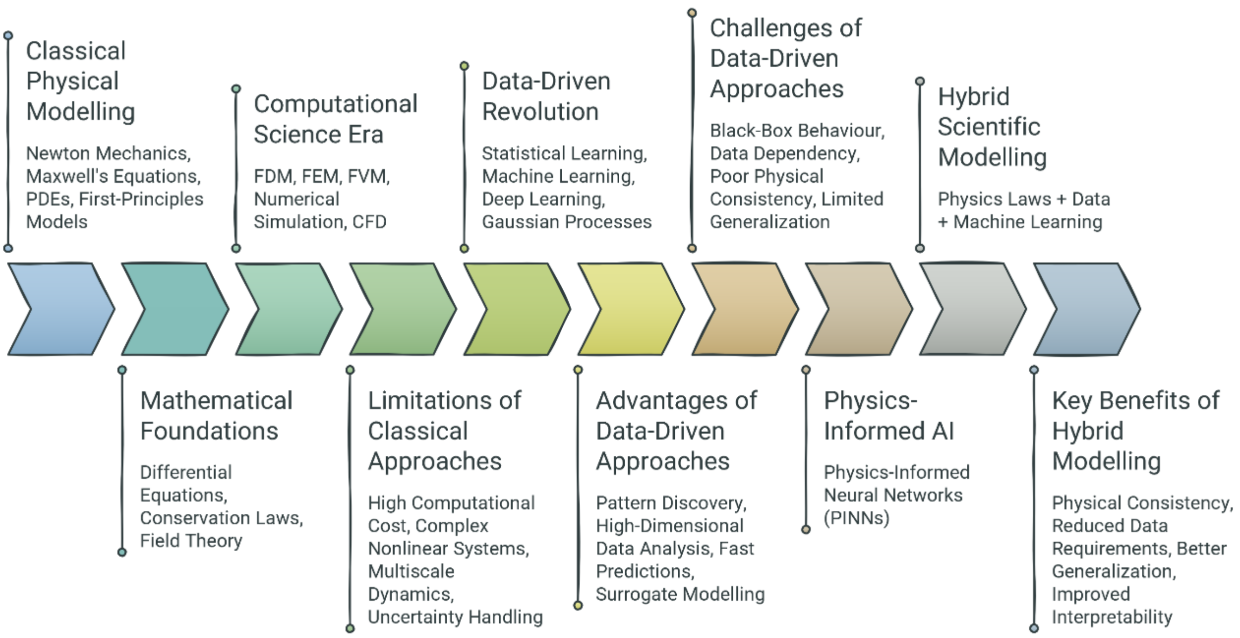
From Newton to Neural Networks: A Review of Data-Driven Physical Modelling and the Rise of Physics-Informed AI

Eshit Dhiman

Independent Researcher, Vega School, Gurugram, India

ABSTRACT: This review examines the evolution of physical modelling from classical first-principles approaches to contemporary data-driven and physics-informed learning frameworks. It begins with the historical foundations of scientific modelling, including classical mechanics, field equations, computational simulation, and statistical mechanics, and shows how these traditions established differential equations and numerical methods as core tools in science and engineering. The review then discusses the rise of data-driven modelling, covering statistical learning, machine learning, deep learning, and Gaussian processes, with attention to their strengths in handling complex nonlinear systems and their limitations in interpretability, data requirements, and physical consistency. A central focus is placed on Physics-Informed Neural Networks (PINNs) and related hybrid methods that integrate governing laws into learning algorithms to improve prediction, generalization, and physical plausibility. Applications across fluid dynamics, structural engineering, climate science, biomedicine, materials discovery, and energy systems are highlighted. Finally, the review identifies key challenges related to scalability, uncertainty quantification, robustness, and benchmarking, and outlines future directions such as active learning, operator learning, and scientific foundation models.

KEYWORDS: PINNs; Scientific AI; Machine Learning; Deep Learning; Physical Modelling; Data-Driven Modelling



1. INTRODUCTION

1.1 The Dual Heritage of Physical Modelling

The advancement of modern science has been intimately linked with the ability to mathematically and physically describe natural phenomena. Over the centuries, classical physics has provided fundamental laws governing the dynamics of objects, fluid motion, electromagnetism, and numerous other natural phenomena. Newton's laws of motion and Maxwell's equations are just two examples that show that complicated systems can often be modelled deterministically using mathematical relations (Newton, 1687)(Maxwell, 1865). These first principles approaches became the backbone of physical modelling and allowed scientists to predict how a system would behave under certain circumstances. Using equations and analytical approaches to model physical laws has made it possible to build interpretable models that have a clear physical background. However, there are many real-world systems where such an approach is insufficient due to their inherent complexity. Turbulent flows, biological systems, and climate dynamics exhibit strong nonlinearities, multiscale interactions, and uncertainties that are often difficult to represent using purely first-principles approaches. As scientific problems became increasingly complex, the role of experimental data and computational methods in solving them also expanded.

1.2 The Modern Convergence

The rapid growth of computational power, as well as the availability of large-scale scientific data, has brought about significant changes in the way we model physical processes. Over the past decade machine learnine (ML) algorithms mainly deep learning models, have demonstrated remarkable success in uncovering complex patterns from large-scale datasets (LeCun et al., 2015). Despite their predictive capabilities, however, the pure data-driven approaches usually lack both explainability and generalization to the domain outside the range of data used for training the algorithm. In many scientific areas, in addition to being able to make accurate predictions, it is crucial that a model be physically correct and consistent with existing scientific laws. In an effort to overcome such difficulties, physics-informed machine learning (ML) became particularly popular. Unlike other approaches where purely observational data serve as input information, the physics-informed methods combine physics with machine learning approaches, incorporating physical laws in the learning procedure. For instance, the PINN framework uses differential equations in the optimization process to build physically meaningful solutions (Raissi et al., 2019). Such a synergy between science and machine learning represents an interesting trend toward hybrid scientific models where theoretical and data-driven components are combined.

Although there is a rapid growth of scientific ML, but existing literature remains fragmented. Many review papers focus on specific methodologies such as Physics-Informed Neural Networks, operator learning or data-driven modelling frameworks. However, comparatively few studies provide a unified perspective that discuss the historical evolution of physical modelling from classical mechanics and field theory to contemporary hybrid approaches that combine physics and artificial intelligence (AI). So, there is a need for a comprehensive review that not only summarizes recent developments but also places them within the broader historical and scientific context of modelling. This review address that needs by studying the conceptual progression from first principles modelling to modern scientific ML framework.

1.3 Objectives and Scope of the Review

Physical modelling has undergone a remarkable transformation over the past few centuries, evolving from purely theoretical formulations based on physic laws to modern approaches that increasingly incorporate data driven learning techniques. Now a days scientific and engineering problems become more complex, so researchers are exploring new ways to combine traditional physics-based methods with advances in ML and AI. This shift has given rise to a new generation of modelling frameworks that seek to balance physical interpretability with predictive capability. The aim of this review is to study this evolution and provide a comprehensive overview of the major developments that have shaped modern scientific modelling. Mainly attention is given to the transition from classical first-principles approaches toward data-driven and hybrid methodologies that integrate physical knowledge with ML.

The review starts with historical perspectives of physical modelling with examples of classical mechanics, field equations, computational simulations, and statistical physics. The review moves on to classical modelling approaches rooted in physics by examining the first-principles modelling paradigm based on differential equations and its primary merits and disadvantages.

Next, this paper considers the rise of the data-driven methodology in scientific modelling by looking into statistical learning, machine learning, deep learning, and including probabilistic methods such as Gaussian processes. More specifically, this section focuses on the growing popularity of Physics-Informed Neural Networks (PINN). This includes the basic mathematical formulation, problem-solving capabilities in forward and inverse problems, benefits and limitations of PINN, and even new developments in this area.

Furthermore, hybrid methods that use physical knowledge and ML to optimize physical modelling for better computational efficiency and scalability are reviewed. Examples of applications of such hybrid approaches in fluid mechanics, structural engineering, quantum systems, climate science, biophysics, energy grid modelling, and material discovery are analysed. This paper also discusses the existing challenges and limitations of these methodologies, including computational scalability issues, data-driven uncertainty quantification, data quality requirements, explainability and robustness, and benchmarking in multiple physics scenarios. Some future directions in developing novel methods, for example, through active learning, operator-learning, and scientific foundation models, are also considered.

2. HISTORICAL FOUNDATIONS OF PHYSICAL MODELLING

2.1 Classical Mechanics and the Newtonian Paradigm (1687-1850)

The emergence of physical modelling is closely associated with the development of classical mechanics during the Scientific Revolution period. One of the most significant milestones in the history of physical modelling was the work of Sir Isaac Newton, who formulated mathematical laws governing motion. These laws help in quantitative prediction of motion based on applied forces and the properties of objects (mass and acceleration), which meant that one could describe motion quantitatively instead of only qualitatively as before. The introduction of Newtonian mechanical models, therefore represented the first major step toward deterministic physical modelling, since in principle, the further behaviour of any physical system can be predicted from the given laws of physics and its initial conditions.

Moreover, several important contributions were made in those times by prominent mathematicians such as Leonhard Euler, Gottfried Wilhelm Leibniz, and Joseph-Louis Lagrange to create mathematical descriptions of physical processes. The development of calculus provided a powerful framework for describing continuously varying quantities and their rates of change through differential equations. Besides, Euler made major contributions to analytical mechanics, differential equations, and fluid dynamics. Later, Lagrange introduced the concept of energy into classical mechanics.

Thus, another conceptual leap was achieved when scientists discovered that physical phenomena could be described using mathematical methods and that the mathematical expression of laws of physics could be used to predict the future evolution of physical processes. The deterministic nature of classical physics has had a major impact on the development of later science, particularly in developing the concept that any physical system can theoretically be predicted precisely, provided one knows its laws and starting conditions.

2.2 The Age of Field Equations (19^{th}-Early 20th Century)

In the nineteenth century, another significant revolution took place, namely the invention of the field concept. Unlike earlier approaches that primarily focused on individual particles and the forces acting between them, the field perspective described physical phenomena as continuous quantities distributed across space and time. The first notable example is, once again, the work done by Joseph Fourier, who presented a mathematical model of heat diffusion using partial differential equations (PDEs) (Fourier, 2009). This contribution showed that complex physical phenomena could be mathematically modelled using spatial and temporal variation of variables.

James Clerk Maxwell made another major contribution to the emergence of field theory in physics. The discovery that Maxwell equations can describe the unification of electricity and magnetism as a system of fields was an important milestone (Maxwell, 1865). Maxwell's work showed that physical interactions could be understood not only through the motion of particles but also through fields that extend throughout space.

During the same period, important advances were also made in fluid dynamics through the development of the Navier-Stokes equations by Claude-Louis Navier and George Gabriel Stokes. These equations describe the motion of fluids under different external forces, viscosity, and pressure. Nonetheless, it should be mentioned that many mathematical problems related to the Navier-Stokes equations are very challenging to solve and still require solutions, especially in terms of their existence and smoothness in three dimensions.

Field equations played a key role in promoting partial differential equations (PDEs) as primary modelling tools in physics. Namely, this happened because PDEs allowed describing how physical variables behave continuously in both space and time dimensions. As a result, differential equations became crucial tools for studying heat, electromagnetism, waves, and fluid flows. This era made an immense contribution to the mathematical scope of physical modelling.

2.3 The Birth of Computational Science (1940s-1970s)

Although many numerical techniques had been developed earlier, computational science emerged as a distinct discipline during the mid-twentieth century with the advent of electronic computers. The growing complexity of scientific and engineering problems created a need for computational approaches capable of solving mathematical models that were difficult or impossible to address analytically. With the increased sophistication of physical models, some of the governing equations especially nonlinear partial differential equations became unsolvable analytically. As a result, there was a need to develop numerical approaches that would allow obtaining approximate solutions to these equations using computers.

Some of the most significant approaches developed during that period include the finite difference method (FDM), finite element method (FEM), and finite volume method (FVM). In essence, these methods turned the continuous problem described with the differential equation into an approximate discrete form that could be solved numerically (Courant et al., 1967)(Zienkiewicz, 1971). The advent of computational models expanded the range of physical simulation greatly compared to previous analytic methods. It became possible to investigate analytically unsolvable problems such as fluid flow in turbulence, deformation of solids, heat transfer, and air flow around objects. As a result, it allowed the development of entirely new sciences such as CFD, computational mechanics, and numerical weather prediction.

At the same time, computational modelling posed certain restrictions. The first one included the requirement to have sufficient computational capacity, especially when simulating high-dimensional or multiscale systems. Additionally, the precision of the obtained results heavily depended on the numerical approach used and various boundary conditions assumed. Minor discrepancies in the simulation parameters could lead to significant differences in the obtained results. However, the emergence of computational science has dramatically changed the face of physics and engineering as it introduced simulation as the third pillar alongside theory and experiment.

2.4 Statistical Mechanics: When Physics Met Probability

Even though classical mechanics was mostly based on deterministic considerations, the investigation of physical systems composed of large amounts of interacting particles required probabilistic considerations and resulted in the birth of statistical mechanics. As an example, the work of scientists like Ludwig Boltzmann and Josiah Willard Gibbs, among others, has laid down the basis for probabilistic modelling in physics through the establishment of a connection between the dynamics of microscopic particles and the properties of temperature, pressure, and entropy (Boltzmann, 1970)(Gibbs, 2010).

Instead of analysing every single particle in the system, statistical mechanics focused on probabilistic properties of the entire distribution or ensemble. This represented an essential conceptual shift in physical modelling, whereby uncertainty and randomness ceased to be viewed merely as consequences of our inability to measure everything accurately and became recognized as intrinsic properties of the system under consideration. This probabilistic thinking introduced into physics with statistical mechanics would later influence modern computational and machine learning techniques dealing with uncertain information and stochastic processes. Indeed, many elements of contemporary probabilistic learning have been inspired by the results of statistical physics in the past.

Moreover, statistical mechanics has proven to be instrumental in thermodynamics, quantum mechanics, condensed matter physics, and materials science by demonstrating how complex phenomena and behaviours can arise from simple local interactions. Overall, along with classical mechanics, field theory, and computational science, the development of statistical mechanics has created the theoretical basis for modern scientific modelling. The next sections will focus on further progress in this area, specifically on data-driven and hybrid modelling frameworks.

3. CLASSICAL PHYSICAL MODELLING: METHODS AND LIMITATIONS

3.1 First-Principles Modelling

Classical physical modelling is fundamentally based on first-principles, or white-box, approaches in which the behaviour of a system is derived directly from established physical laws. These approaches rely heavily on governing equations such as ordinary differential equations (ODEs) and partial differential equations (PDEs), which describe how physical quantities evolve across time and space (Newton, 1687)(Fourier, 2009). Such an approach to modelling had dominated science for centuries, "providing, in principle, a framework for predicting the future behaviour of any system by solving the governing equations, provided the corresponding initial and boundary conditions. The conservation laws of mass, momentum, and energy became the bedrock on which descriptions of different physical processes, such as those described by mechanics, hydrodynamics, thermodynamics, and electromagnetism, were based.

The interpretability is perhaps one of the main advantages of this approach to modelling physical processes. Since first-principles approaches to modelling employ well-established physical laws, it is possible to understand the underlying physical mechanisms. It means that the researcher not only gets the opportunity to study physics but also to predict its further behaviour. However, there is a downside to the process. Often, constructing first-principal models requires profound knowledge of the domain and considerable effort in mathematics. Simplification and linearization of the resulting models are common practices for making them more tractable. As physical systems become more complicated, the creation of first-principles models tends to be quite difficult.

3.2 Key Equation Classes Reviewed

There are various types of equations that have served as mathematical models of classical physics-based modelling of nature within different areas of science. One of the notable examples is that of Newton's law of motion, which provided the basic principles of classical mechanics by linking force with mass and acceleration (Newton, 1687). With the advent of these equations, scientists were able to conduct quantitative analysis of motion. These are key principles underlying mechanics and engineering today. Another example is Fourier's heat equation, which describes thermal diffusion through spatial and temporal temperature variation (Fourier, 2009).

In addition, the field of fluid dynamics evolved with the help of the Navier-Stokes equations developed by Claude-Louis Navier and George Gabriel Stokes. They describe the flow of fluids subject to the influence of pressure, viscous forces, and external forces (Navier, 1823)(Stokes, 2007).

Likewise, the unification of electricity and magnetism was achieved through Maxwell's equations, leading to dramatic changes in our understanding of electromagnetism (Maxwell, 1865). Moreover, Schrödinger's equation became vital for the description of quantum phenomena (Schrödinger, 1926). As a result of all these developments, differential equations became the primary means of representing scientific models. However, even though most of the governing equations of nature are powerful tools of prediction, the equations are too complex and cannot be solved analytically.

3.3 Numerical Solvers and Simulation Methods

Due to the analytical complexity of many governing equations, numerical methods have been developed as critical aids to scientific computation. With the help of numerical methods, it becomes possible to approximate solutions for problems that involve nonlinearity, complex geometry, and difficult boundary conditions. Among the most widely used approaches are the Finite Difference Method (FDM), which approximates derivatives through discrete differences; the Finite Element Method (FEM), which divides a domain into smaller localized elements for approximation; and the Finite Volume Method (FVM), which ensures conservation laws across discrete control volumes.

Numerical methods have greatly expanded the scope of scientific modelling and allowed large-scale computations in areas such as Computational Fluid Dynamics (CFD), Structural mechanics, Heat transfer, and Electromagnetic field analysis (Courant et al., 1967)(Zienkiewicz, 1971). The numerical simulations have thus become an essential part of modern scientific research, in addition to theoretical and experimental studies. However, numerical methods often require significant computing power, fine discretization, and careful handling of boundary and initial conditions in order to achieve numerical stability and accuracy. For problems of extreme complexity and high dimensionality, the computational burden can be very heavy.

3.4 Strengths of Classical Approaches

Some key strengths of classical physical modelling include interpretability and a sound theoretical basis. The ability to interpret results in a physical way is one of the key advantages of classical modelling. Due to the direct derivation of the mathematical description from physical laws, all terms and parameters of the model are scientifically meaningful. Therefore, scientists are not only able to make predictions about the behaviour of the system under study, but can also explain the physics behind these predictions.

In addition, classical physical modelling benefits from a well-justified theoretical background. The differential equations used in the majority of classical models have been extensively verified through scientific research. Thus, classical models usually yield valid predictions in the context of their application. Another major advantage of classical modelling is its broad use in multiple fields of science, such as mechanics, thermodynamics, fluid dynamics, structural engineering, and electromagnetism.

3.5 Fundamental Limitations

Nevertheless, besides their advantages, some significant challenges have been posed by classical physical modelling techniques when working with highly complicated real-life systems. Most natural systems have inherently non-linear couplings, multiple time scales, stochastic behaviours, and a very large number of interacting variables. Considering all these elements in terms of equations may be difficult and might even prove to be infeasible. As a result, the assumptions made in most classic approaches may limit the accuracy of these approaches.

Furthermore, the use of simulations in combination with classical methods is extremely time-consuming, especially in cases where the problem involves turbulence, climate models, or molecular interactions. Deterministic modelling schemes often do not take into account factors like noise, uncertainty, and incomplete information typical of most real-life and experimental data sets. Adaptability to different conditions and environments also becomes problematic, as the governing equations of a particular model system are usually predefined. These difficulties gave rise to data-driven scientific modelling paradigms, which use the complementary power of learning-based techniques alongside physics-based models.

4. THE RISE OF DATA-DRIVEN METHODS

4.1 From First-Principles to Pattern Recognition

First principles methodologies have traditionally been used for the development of models through physical modelling techniques. The major advantage of this approach is that it gives the model high theoretical interpretability due to the presence of mathematical equations, which can be derived from scientific knowledge. Despite their effectiveness, first principles methodologies are often prone to difficulties, especially when modelling complex, nonlinear, and multidimensional systems whose behaviour might not be entirely defined by physical laws and equations.

The emergence of scientific data of greater magnitude and complexity has led to an evolution in modelling approaches where data plays a critical role in capturing the behaviour of a system rather than using physical laws alone. Data-driven modelling can be thought of as the move from deductive reasoning in science to induction, whereby the models learn the behaviour of a system from example (Brunton et al., 2016). This allows scientists to approximate the behaviour of complex systems even when the underlying physical laws are only partially known.

4.2 Classical Statistical Methods

Before the emergence of modern machine learning approaches, classical statistics was considered the primary method to conduct data analysis and prediction through models. Methods like Linear regression, Logistic regression, and Principal Component Analysis (PCA) were commonly used for exploring relationships between variables, predicting, and reducing the complexity of datasets through dimensionality reduction (Gorriz et al., 2026). Mathematically, linear regression provides a model describing the relation between dependent and independent variables, while logistic regression is used to apply linear regression concepts to classification problems where categorical outputs are expected. The main idea of PCA is finding the direction of maximum variance and allowing high-dimensional data to be represented in low dimensions.

These methods played a major role in early scientific data analysis because they were computationally efficient, mathematically interpretable, and relatively easy to implement. Furthermore, they provided important foundations for many later developments in machine learning and statistical learning theory. However, classical statistical approaches are generally limited in their ability to model strongly nonlinear relationships and highly complex interactions commonly found in real-world scientific systems. As scientific datasets grew in scale and complexity, more flexible learning-based methods became necessary.

4.3 Machine Learning in Scientific Modelling

With the rise of machine learning technologies, data-driven scientific modelling has gained much more power, allowing one to approximate extremely complicated relationships. Many types of algorithms, ranging from support vector machines, decision trees, ensembles, to various kinds of neural networks have proven to be powerful tools for scientific modelling in a variety of scientific and engineering areas. Of the above-mentioned machine learning algorithms, neural networks, have become especially popular because of their ability to approximate a wide range of nonlinear functions according to the universal approximation theorem (Cybenko, 1989). As a consequence, neural networks allow approximating the behaviour of rather complex systems that cannot be modelled analytically.

The development of deep learning has further extended this ability by providing the opportunity to extract hierarchical features from large high-dimensional datasets (LeCun et al., 2015). Therefore, nowadays machine learning algorithms are successfully used in scientific disciplines such as fluid dynamics, climate modelling, materials design, biomedicine, and others. Machine learning models have special significance in scientific research since they allow working with large, noisy, high-dimensional, and partially missing datasets. Such problems often occur in experimental and numerical sciences. Machine learning techniques allow discovering the relationship between different variables based on the principle of pattern recognition. However, at the same time, many machine learning systems often are criticized for being black box models because of the difficulty in interpretating of the inner workings of machine learning algorithms. This disadvantage is particularly problematic for scientific research.

4.4 Deep Learning Revolution

Deep learning represents a major breakthrough in the evolution of machine learning and relies on neural network architectures that are based on numerous layers of computation and can learn increasingly abstract representations of the data. In the last decade, deep learning has shown its great effectiveness in image classification, speech recognition, natural language understanding, and other areas of pattern recognition (LeCun et al., 2015). Within the field of scientific modelling, deep learning allows one to process large amounts of data and recognize highly intricate structures that are hidden within it. Various neural network architectures are tailored for different types of scientific data. Thus, Convolutional Neural Networks (CNNs) work best for spatial data, such as satellite images or medical images; Recurrent Neural Networks (RNNs) are extensively used for sequential and time-series data; and graph neural networks (GNNs) are utilized in cases when the studied system includes interconnections between its components.

The flexibility and abstraction capability of deep learning models have made them increasingly important in fields like climate prediction, molecular dynamics, fluid dynamics, material science, and computational biology. The technology has also enabled the creation of surrogate models that speed up simulations that are otherwise highly resource-intensive. However, while exhibiting great predictive capacity, deep learning models suffer from a number of significant drawbacks. Namely, their internal representations tend to be highly non-transparent and hard to interpret. Furthermore, deep learning models rely on vast amounts of training data and computational resources to function reliably.

4.5 Gaussian Processes

In addition to machine learning models based on determinism, probabilistic models have been increasingly adopted in science for scientific modelling because of their ability to incorporate uncertainty in their computations. One such probabilistic model is the Gaussian Process (GP), where a probability distribution over functions can be obtained (Rasmussen & Williams, 2005). Unlike many standard neural network models, Gaussian Processes naturally incorporate uncertainty into their predictions. This characteristic makes them particularly suitable for scientific applications where noise, incomplete measurements, and uncertainty are common (He et al., 2026).

Another important advantage of Gaussian Processes is their ability to incorporate prior knowledge through kernel functions. These kernels define assumptions regarding the structure, smoothness, and correlations present in the modelled system and allow physically meaningful behaviour to be incorporated into the learning process (Polke et al., 2026). Because GP models provide confidence intervals alongside predictions, they are often considered more interpretable than many deep learning approaches. This probabilistic interpretability is especially important in scientific domains where reliability and uncertainty quantification are critical. However, Gaussian Processes face some significant drawbacks. The computational cost increases exponentially with respect to the number of training samples, making the algorithm impractical for large-scale problems. Scalability has thus become one of the most critical issues in using Gaussian Processes for modelling. Despite these limitations, Gaussian Processes represent an important step toward integrating probabilistic reasoning, uncertainty estimation, and scientific learning within data-driven modelling frameworks.

4.6 Advantages and Concerns

Data-driven modelling techniques, especially those involving machine learning and deep learning, bring many benefits to scientific work. Machine learning models are capable of handling high dimensional data sets, learning complicated nonlinear relationships, and even adapting to the dynamic behaviour of the system that might have equations that are not completely known or difficult to solve. The ability to handle complicated tasks makes the use of data-driven techniques very important in many fields of science. Such methods are useful in speeding up simulations, creating approximations for expensive computations, prediction analysis, and discovering scientific patterns and behaviours.

However, there are several problems associated with data-driven modelling techniques that are worth noting. One major problem is the difficulty of interpreting the results from many machine learning algorithms due to their black box nature. In scientific studies, simply making predictions is usually not enough; understanding how the model arrives at its conclusions is equally important. Moreover, machine learning techniques depend on having huge amounts of good quality data available for training purposes. Gathering experimental or computational data in many fields of science is usually very expensive and requires a lot of effort and time. Additionally, there could be problems in generalizing the model predictions outside the training dataset distribution.

The last issue, related to machine learning models is their inability to satisfy certain physical laws like conservation laws or symmetry principles. In many scientific fields, it is crucial that machine learning models take into account such physical laws, which implies that predictions will be physically consistent. These limitations motivated the development of hybrid scientific modelling approaches that combine machine learning techniques with physical principles, leading to frameworks such as Physics-Informed Neural Networks and other forms of scientific machine learning.

5. PHYSICS-INFORMED NEURAL NETWORKS (PINNs)

5.1 Conceptual Foundation

A Physics-Informed Neural Network (PINN) is a machine learning approach in which physical laws are incorporated directly into the learning process. Different from the conventional data-driven approach in which a data-driven neural network is constructed based on the input data alone, a PINN utilizes governing physical laws such as PDEs in its training step (Raissi et al., 2019). As a traditional data-driven network does, a PINN aims at minimizing the discrepancy between output prediction and observed data. However, it further imposes constraints by adding the residuals of governing laws into the loss function, making sure that the solution learned not only approximates the observations well but also obeys the physical rules.

With this method, the integration of physical laws into machine learning models is possible without resorting to massive amounts of training data. It provides a powerful tool for the fusion of classical first-principles modelling with data-driven modelling.

5.2 Mathematical Framework of PINNs

PINNs essentially rely on the adjustment of the neural network's loss function to incorporate physics-based constraints (Khanra et al., 2026). The general differential equation can be represented in the following manner:

$$N[u(x,t)] = 0 \qquad (\mathbf{1})$$

In **equation (1)** N symbolizes the differential operator, whereas u(x,t) is the sought approximation of the function.

In a PINN framework, a neural network is used to approximate u(x,t). The loss function is then defined as a combination of multiple components:

- Data loss-the degree to which the predicted result deviates from the observed data;
- Physics loss-enforcing the governing law through the minimization of the residual of the equation;
- Boundary conditions and/or initial values are lost.

In practice, the physics-based constraints are enforced at a set of collocation points within the domain, where the governing equations are evaluated using the neural network approximation. The use of automatic differentiation allows the efficient calculation of necessary derivatives within the scope of PDE residual evaluation.

5.3 Solving Forward Problems with PINNs

In forward problems, we know the governing equations and parameters, and the goal is to determine the solution of the system. PINNs offer another approach to solving systems other than those of numerical solvers, where we can find an approximation of the solution using neural networks. For instance, PINNs have been used to solve differential equations such as the Burgers' equation and Navier-Stokes equations, which are encountered in fluid mechanics (Raissi et al., 2019). As opposed to classical methods that involve the discretization of the problem's domain, PINNs are able to deal with a continuous domain. This eliminates the process of mesh generation.

5.4 Solving Inverse Problems

An inverse problem refers to an approach whereby unknown parameters or properties associated with a system are estimated through observation data. These types of problems are frequently ill-posed and challenging to address through conventional means.

PINNs are particularly suitable for inverse problems because unknown system parameters can be integrated directly into the learning process. This approach leverages the use of both observational data and physics laws to enable inference of the unknowns from governing equations. Examples include parameter estimation, system identification, and model calibration, among others.

5.5 Strengths of PINNs

PINNs have several benefits compared to classical numerical methods and data-driven methods. First, they can work efficiently even with insufficient data or in situations where data may be corrupted, as physical laws incorporated in PINNs serve as regularization mechanisms during the process of learning. “PINNs also produce solutions that remain consistent with physical laws, which was one of the drawbacks of machine learning methods. The meshless nature of PINNs makes them efficient in solving tasks in complex geometries and higher dimensions than classical numerical techniques. These features make PINNs an effective instrument for scientific computing tasks.

5.6 Limitations and Open Challenges

PINNs, despite their benefits, have some limitations that include training difficulties, high computational cost, and remaining issues related to interpretability. The optimization procedure is vulnerable to initialization, which makes convergence challenging. Additionally, PINNs might encounter difficulties when working on highly intricate and stiff systems, where the loss function space is complicated to optimize. Moreover, although PINNs enhance physical coherence, they are unable to address concerns associated with interpretability and clarity of the model.

Thus, more investigation is required to enhance the stability and scalability of the training process and improve the architecture of PINNs. Additionally, alternative models can be considered, such as the operator learning approach. These limitations have motivated the development of more advanced hybrid and operator-learning approaches, which are explored in the following section.

5.7 Extensions and Variations of PINNs

Physics-Informed Neural Networks have come a long way since their introduction, with a variety of extensions emerging in an effort to further improve their performance and scalability (Farea et al., 2024). The first notable approach to achieve this goal is to use domain decomposition methods, where the domain is partitioned into small subdomains, and multiple neural networks are simultaneously trained to accelerate convergence and reduce computational costs. Second, loss weighting methods are employed in which the weights assigned to different loss terms such as data loss and physics loss are adaptively adjusted during the training phase. In this way, optimization difficulties can be mitigated.

Furthermore, variants of PINNs have been created to cater to more challenging problems with higher dimensions, such as multi fidelity PINNs and stochastic PINNs. These developments reflect ongoing efforts to improve the robustness, scalability, and practical applicability of PINNs for increasingly complex scientific problems.

6. HYBRID MODELLING FRAMEWORKS

6.1 Why Hybrid Approaches

Physics-based models and data-driven models both possess important advantages as well as limitations. Physics-based models rely on established scientific laws and therefore provide physical interpretability and consistency. However, they are often difficult to apply to highly complicated or poorly understood systems where deriving or solving governing equations becomes challenging. In contrast, data-driven models are capable of learning complicated patterns directly from data and can often achieve strong predictive performance. Nevertheless, such models may suffer from limited interpretability and reduced generalization outside the distribution of the training data.

Due to these complementary abilities and deficiencies, hybrid modelling techniques have been developed with the intent of merging the strengths of physical understanding with those of machine learning into one system. The aim is to achieve better performance in terms of prediction while maintaining physical integrity (Karniadakis et al., 2021).

6.2 Types of Hybrid Models

Hybrid approaches can be organized in several ways depending on how the physical and data-driven components interact with each other within the modelling framework. One common category is sequential models. In such approaches, a physics-based model first generates an approximate solution, after which a machine learning model is used to correct errors or account for missing physical effects (Wu et al., 2024). This method is often useful when the governing equations fail to fully describe the real system. Another category is parallel models, where both the physics-based and machine learning models operate simultaneously. Their outputs are then combined through weighting strategies or additional learning mechanisms. Such an approach attempts to benefit from both physical constraints and data-driven flexibility.

A third category includes embedded models, where physical laws are incorporated directly into the architecture or training procedure of the machine learning model itself. Physics-Informed Neural Networks are one example of this type of approach. Overall, these categories can be viewed as representing different levels of interaction between physical modelling and machine learning methods (Karniadakis et al., 2021).

6.3 Surrogate Modelling

A surrogate model is a computational tool for modelling the behaviour of expensive simulations using computationally cheap approximations. Rather than repeatedly solving complex differential equations for the governing dynamics of the simulator, surrogate modelling involves learning the mapping between input variables and output variables from already available data. Common surrogate modelling techniques include neural networks, Gaussian processes, regression-based approaches, and reduced-order models. These techniques have the potential to greatly decrease computational expense without sacrificing prediction accuracy.

Surrogate models find extensive use in fields like fluid dynamics, structural design, optimization problems, aerodynamic analysis, and energy systems modelling. In engineering, they assist in(Forrester & Keane, 2009) speeding up simulations that would otherwise demand substantial computational resources.

6.4 Transfer Learning in Physical Sciences

Transfer learning refers to the process of applying knowledge learned from one task or dataset to another related problem (Pan & Yang, 2010). This idea has become increasingly important in scientific machine learning, particularly in fields where collecting large datasets is expensive or difficult. In many cases, a neural network trained on one physical system can later be fine-tuned for another related system using a much smaller amount of data. This reduces both training time and computational cost. Transfer learning has been applied to various fields, including material science, medical imaging, climate simulation, and computational physics. It is highly applicable whenever the simulations or experiments are expensive to compute, or there is insufficient observational data.

6.5 Applications and Importance of Hybrid Modelling

Hybrid modelling approaches have become increasingly important in many areas of science and engineering because they allow physical knowledge and data-driven learning to complement one another. In such cases, machine learning techniques can be used to extract material behaviour from simulations or experiments, while physical constraints help maintain physically meaningful predictions. For advanced materials and complex systems, it may become hard to formulate exact constitutive relations using conventional techniques. Here, we can rely on machine learning techniques to extract material behaviour from simulations or experiments, and use physics as an additional constraint to guarantee physically meaningful results.

Hybrid models are also common in fields including fatigue analysis, fluid mechanics, weather prediction, energy systems, and material discovery. In many of these problems, purely physics-based simulations may be computationally expensive, whereas purely data-driven models may fail to satisfy important physical principles. Because of these advantages, hybrid modelling has become an increasingly important direction in modern computational science, combining the strengths of both physical modelling and machine learning approaches.

7. APPLICATIONS ACROSS SCIENTIFIC DOMAINS

7.1 Fluid Mechanics and Turbulence Modelling

Fluid mechanics is a tough area of physics to model because the equations that govern, it like the Navier-Stokes equations, are so complex. These equations help us understand how fluids move. They are very hard to solve, especially when we are dealing with turbulent flows. Turbulent flows are highly irregular and involve many different interacting scales. The traditional way to model mechanics is to use a lot of tiny spatial and temporal grids. This approach can give us results, but it takes a lot of computer power, especially for turbulent flows. It is just not possible to solve all the scales directly, which is called Direct Numerical Simulation for real-world problems. To reduce this cost, surrogate models have become widely used in aerodynamic design and related engineering applications. These models learn approximate mappings between input parameters (such as geometry, boundary conditions, or flow conditions) and output quantities (such as lift, drag, or pressure fields). Once these models are trained, they can make predictions quickly, which is really useful for designing and optimizing things.

Machine learning techniques are finding increasing use in turbulence modelling research. The field of fluid mechanics encounters its main obstacle when scientists try to create models that can simulate small-scale movements that are too tiny to detect through their existing simulation techniques. Data-driven approaches are being developed to improve closure models, which aim to approximate the effect of these unresolved scales on the larger flow dynamics, thereby improving the accuracy of reduced-order simulations (Brunton et al., 2016).

PINNs have also seen some successful applications in fluid dynamics problems by expressing the governing equations directly into the learning process. By enforcing the Navier-Stokes equations within the loss function, these models can produce physically consistent solutions even when training data is limited, making them particularly useful for inverse problems and scenarios with sparse observations (Raissi et al., 2019). Overall, these approaches surrogate modelling, machine learning-based turbulence modelling, and physics-informed learning are collectively advancing fluid mechanics by reducing computational cost while maintaining physical fidelity, enabling more efficient simulation and design workflows in aerodynamics and beyond.

7.2 Structural and Solid Mechanics

Structural and Solid Mechanics is concerned with predicting how materials will deform and react under loads, which is vital for safety during structural design processes. One of the biggest problems in Structural and Solid Mechanics is the formulation of constitutive laws for describing the relationship between stress and strain. For some advanced materials, the derivation of constitutive laws becomes too complicated to be carried out analytically using the traditional approach. The classical methodology depends on pre-designed constitutive laws and models, which often do not represent the real properties of the material. In order to address this issue, data-driven modelling of constitutive laws can be used to find the dependencies directly from input experimental data, without imposing restrictions on the form of the dependency beforehand.

Hybrid modelling methods utilize additional information about physics in Solid Mechanics to make data-driven models physically consistent and reliable. Besides that, data-driven methodologies are very common for fatigue and failure analysis, where they are used to analyze degradation processes and predict failures. These approaches provide significant improvements over purely classical modelling techniques.

7.3 Quantum Chemistry and Molecular Dynamics

Quantum chemistry and molecular dynamics involve the study of interactions at the atomistic and molecular levels, which require intensive computation using traditional techniques. Modern machine learning algorithms have made it possible to efficiently model quantum interactions and maintain accuracy through the development of Deep Potential Molecular Dynamics (DeePMD), and neural network-based models like SchNet that model interatomic potentials using deep learning (Schütt et al., 2017). Such methods have significantly boosted simulation speeds, allowing researchers to study larger-scale and more prolonged interactions than would be possible otherwise.

7.4 Climate Science and Geophysics

Both climate and geophysical sciences are highly non-linear in nature and involve complex interactions occurring over various space and time scales that are difficult to capture in any physics-based numerical models. Such models, as a rule, require large computational capabilities in order to generate accurate forecasts over extended time periods. Thus, there is a need for alternative approaches that will allow one to reduce computational effort without significantly compromising forecasting accuracy.

One way to cope with this problem is to use machine learning models that act as emulators of physical models' behaviour. Instead of solving differential equations at each time step, these models learn the mapping from input parameters (initial conditions, forcings, etc.) to output ones (fields of temperature, pressure, etc.). After appropriate training, this mapping provides for fast approximation of complex physical processes without significant loss of prediction accuracy. Another promising approach based on deep learning is the application of physics-constrained neural networks in geophysics. The idea is to integrate the known physical laws and governing equations into the learning process, making the model's predictions compliant with basic principles of mass and energy conservation, for instance. This approach is especially useful when dealing with sparse observational data since it reduces possible errors resulting from incorrect physical interpretation (Reichstein et al., 2019).

7.5 Biomedical and Biological Systems

Hybrid modelling has also found important applications in biomedical sciences because biological systems are often too complex to be described accurately using only physics-based or purely data-driven approaches. The idea of the application of hybrid modelling in thermodynamics is the creation of a model of blood flow in the cardiovascular system. Physics-based differential equations can describe general properties of the blood flow; however, actual human physiology changes from person to person, which means that it is impossible to make accurate predictions only based on such equations. The combination of physics and machine learning will allow creating an adaptive approach capable of reflecting patient-specific features.

Medical imaging involves an inverse problem, meaning the problem of reconstruction of certain internal body parts based on measurements obtained from outside. Such reconstruction is complicated by the limited amount of noisy data; therefore, physics-based modeling is important due to knowledge of physical signal forming processes, while machine learning helps fill in gaps in the data (Karniadakis et al., 2021).

7.6 Energy Systems and Power Grids

The energy systems consist of power grids and electrical networks, where there is an interplay between the physical hardware and the fluctuating load requirements. The systems are dynamic in nature, and therefore, the need for accurate models for proper functioning is necessary. Machine learning is commonly applied in the sector in applications such as load forecasting, fault identification, and system optimization. The problem with the data-based models is that they do not respect the physical constraints of the electrical systems; thus, they generate results that are not practical or even realistic. In order to solve the problem, there is a combination of machine learning models with physics laws that dictate how the system operates. There are applications such as the optimal power flow, stability analysis, and energy management, which greatly benefit from the approach.

7.7 Material Discovery and Design

The field of materials science has seen great developments owing to the employment of machine learning techniques for the development and discovery of new materials with desirable attributes. Generative models can be employed to investigate the chemical and crystalline structures to aid the creation of novel material compositions. Graph Neural Networks (GNNs), on the other hand, can be utilized to capture the relationships between atoms at the atomic level, making it easier to predict properties of materials based on structural information. Such an approach greatly enhances the screening process compared to the traditional methods.

Combining expert knowledge with the principles of learning has enabled researchers to achieve greater success in predictive modelling of material properties. In spite of the great successes in scientific disciplines due to the employment of such techniques, there exist certain shortcomings that need addressing. This will be discussed in the next section.

8. CHALLENGES AND LIMITATIONS OF CURRENT APPROACHES

8.1 Computational Scalability

One of the major challenges associated with modern data-driven and physics-informed methods is computational scalability. Many scientific and engineering problems are highly dimensional, involving a lot of variables, a wide range of spatial locations, and time scales. As the dimensions of a problem increase, the computational cost needed to solve or simulate increases rapidly, a phenomenon known as the "curse of dimensionality." Classical numerical methods already require a lot of computational resources to simulate large systems such as turbulent flows, climate simulations, and molecular dynamics. The introduction of machine learning and hybrid simulation methods further increases computational demands because of the need to train large neural networks and optimization procedures.

For example, Physics-Informed Neural Networks (PINNs) must compute the solution and its derivative with respect to the system's equation at many collocation points within the system's domain. This is very costly when dealing with complex or high-dimensional problems. Training of deep neural networks can also be expensive because of the high amount of memory space and computational time needed. Although progress in hardware acceleration has improved computational efficiency, scalability is still an important limitation.

8.2 Data Quality and Availability

The performance of data-driven and hybrid models highly depends on the quantity and quality of data used for modelling purposes. Often, large amounts of high-quality structured data are hard to acquire and expensive, if not entirely unavailable. Data acquired through experiments is usually noisy, uncertain, and contains missing observations due to imperfect sensors and interference with the environment. Additionally, the data used in science is often unstructured, which means that there might be gaps in the data, as observation does not always occur consistently through time or space. This lack of regularity can lead to problems during the learning stage.

Data-based machine learning models are extremely vulnerable to inaccurate data since their functioning is based on statistics obtained from real-life examples. The presence of any bias in data will result in an unreliable model. Despite physics-informed techniques' ability to compensate for some data-related challenges, having high-quality data is still important.

8.3 Selection of Physical Models

The use of hybrid models and physics-based models relies heavily on including the right physics within the framework of the models being used. It is not always easy to choose the appropriate physical laws and physical constraints to be applied. Many real-world phenomena rely on physics that is not fully known, certain or complete. In such cases, researchers have to make simplifying assumptions about the physics at hand. Using inappropriate physical laws in the training procedure might lead to inaccurate results even when the required constraints have been met.

The performance of physics-based and hybrid models can vary significantly depending on the assumptions made regarding the underlying physical processes. It is therefore very important to include the correct physics in the learning process.

8.4 Explainability and Ethical Challenges

Even though machine learning and deep learning models can attain good predictive performance, a large number of models currently used are black box systems for which there is no transparent interpretation of their inner workings. However, this limitation in terms of interpretability constitutes an alarming issue in scientific and engineering applications, where understanding why a prediction was made often turns out to be equally important as making the prediction itself. There must be ways of verifying whether the results obtained make physical sense. Another important problem arises when machine learning models are being used for applications in healthcare, structural engineering, climatic forecasting, and energetics. Errors in prediction in these fields may lead to serious consequences, making it important to understand how model decisions are produced.

Apart from interpretability problems, ethical concerns emerge due to increased reliance on automated decision-making systems. Issues related to transparency, accountability, and reliability remain unresolved as machine learning algorithms continue to be increasingly embedded within scientific and industrial processes.

8.5 Robustness and Generalization

Another crucial limitation of modern machine learning algorithms is the fragility of such systems when encountering scenarios that are significantly different from those encountered during the training process. The performance of many algorithms is confined to a specific set of situations covered by their training datasets, and thus, outside the training domain, the algorithms' predictive capacity may degrade substantially. It becomes especially critical in the context of scientific modelling since it often takes place in complex and unpredictable conditions.

As an example, consider a model that is highly accurate when applied to one specific flow in fluid dynamics; however, its performance may become poor when used to describe another physical situation. Likewise, rare events in climate or biomedicine are unlikely to be found in any dataset and, therefore, can present challenges for prediction. Maintaining the stability of the model under different physical situations is a key issue.

8.6 Lack of Benchmarks

With the fast progress in the realm of physics-informed and hybrid modelling techniques, various architecture types, data sets, and assessment criteria have emerged. Still, there is no agreed upon benchmark problem or criteria. In many papers, different data sets, governing equations, criteria, and conditions are used, which does not allow a straightforward comparison of different methodologies. The assessment of actual efficiency and applicability of novel approaches is hindered by that.

Another issue that persists in the area of machine learning applications is reproducibility in many cases. The lack of consistent benchmarks makes further scientific work difficult, as it impedes validating novel approaches. The development of standardized benchmarking frameworks and publicly available datasets will be crucial for future progress in the field.

9. RESEARCH GAPS AND FUTURE DIRECTIONS

9.1 3D Turbulence at Engineering Scales

Even with advancements made in physics-based and data-driven modelling, simulating the turbulent flow in 3D in real-world settings is still an unsolved problem that remains difficult. Turbulent systems show interactions at varying non-linear scales, making their simulation computationally expensive for both classical numerical solvers and modern machine learning techniques. While PINNs have shown promising results for simpler fluid dynamics cases, their utilization for fully-developed 3D turbulence still remains relatively underexplored. Issues related to training stability, expensive computations, and challenges associated with capturing small scales make it difficult for PINNs to scale effectively.

Recent works involving operator-learning approaches such as Fourier Neural Operators (FNOs) are promising replacements. Instead of learning specific solutions, these networks seek to learn operators that map between function spaces. This approach helps with faster simulation in high dimensional scenarios but may require additional study to ensure physical validity. The growing interest in operator learning methods reflects a broader shift within scientific ML. Rather than approximating individual solutions, these approaches attempt to learn mappings between entire classes of functions, potentially offering superior computational efficiency and generalization capabilities. Although the field is still evolving, operator-learning frameworks are increasingly viewed as one of the most promising directions for addressing challenges associated with large-scale scientific simulations.

9.2 Uncertainty Quantification

Most modern machine learning algorithms focus on generating point estimates while ignoring the inherent uncertainty associated with their outputs. In science and engineering domains, uncertainty estimation may be necessary since measurements, simulations, and natural phenomena are subject to some form of variability or noise. Without uncertainty estimation, it becomes difficult to assess how reliable a prediction actually is, especially in important fields such as climate modelling, health diagnostics, and stress analysis of structures.

Methods that incorporate probabilistic reasoning, like Bayesian neural networks and Gaussian processes, can provide reliable frameworks for estimating uncertainty, although they are often resource-intensive and hard to implement in complex systems.

9.3 Sparse and Noisy Observation Regimes

Many practical problems in the field of science often have insufficient, incomplete, or noisy observation data. Experimental data are usually affected by inaccurate sensors, environmental factors, and irregular sampling times. This problem poses special challenges to machine learning algorithms, as many models rely on large amounts of data. With insufficient data, it may be difficult for models to correctly characterize the physics of the system.

Physics-informed learning methods can help alleviate this problem to some extent by taking advantage of physics equations. However, it still faces considerable difficulties when dealing with highly-noisy data and data-limited situations.

9.4 Multiscale and Multiphysics Coupling

Most of the practical cases contain interactions occurring at more than one scale and multiple physical phenomena. Examples include climate phenomena, which have interactions between fluid dynamics in the atmosphere, dynamics of the ocean, radiation, and chemistry. The same is true for biological phenomena, which have interactions involving fluid mechanics, deformations, and chemical phenomena. It is very difficult to model such interactions, as there can be differences in scales at which each phenomenon occurs. Artificial intelligence models currently struggle to incorporate the interactions that happen between these various physical phenomena.

While hybrid approaches appear to be viable methods, no frameworks currently exists to solve the problem effectively. Further work is still needed before such systems can be modelled efficiently.

9.5 Community Benchmarks and Standardization

The absence of agreed upon benchmarks in datasets and their evaluation continues to hinder development in physics-informed and hybrid machine learning techniques. Multiple papers often utilize various datasets, simulation parameters, error functions, and computational assumptions. Hence, it becomes impossible to evaluate if the new method is actually superior to the previous ones. In order to increase transparency and advance science, the development of open data sets, reproducible evaluation methodologies, and benchmarking standards will be absolutely necessary.

9.6 Explain ability of Physics-AI Systems

The combination of physical rules within ML increases physical validity; however, numerous hybrids still fail to meet expectations in the domain of their interpretability. The ability to understand how models generate predictions is a necessity for science, particularly in important applications that require reliability and verifiability. Scientists are now actively investigating the possibilities of explainable artificial intelligence techniques that will be capable of revealing which physical properties, variables, and principles affect the model's activity the most. Increased interpretability not only contributes to increased reliability of models but can also help scientists discover new physical relations in data. Because of this, improving interpretability has become an important research direction.

9.7 Active Learning and Experimental Design

Active learning is another emerging research direction that enables machine learning models to actively select new data that would maximize their improvement. In science, experiments and simulations can be both expensive and resource-intensive. Instead of generating a vast amount of unnecessary data, active learning aims to uncover which data would provide the most valuable information in terms of maximizing knowledge and saving computation/experiments. There are important applications for this method within experimental physics, materials science, fluid dynamics, and biomedical sciences, where data generation is very important. The combination of active learning with physics-based approaches could have a major impact on science in the coming years.

9.8 Foundation Models for Physical Sciences

Capitalizing on advances made by foundation models for natural language processing and computer vision, research efforts have more recently been directed toward the study of foundation models that can be used for scientific purposes. The development of such models involves training on very large scientific datasets and simulations that help learn representations of physical systems. Instead of training an independent model for each specific task, foundation models will hopefully allow cross-domain adaptation with minimal effort and data in many scientific domains.

It is likely that the domain-specific pre-training will lead to much greater efficiency in fields including fluid dynamics, climate science, material discovery, and molecular modelling. However, issues of physical consistency, efficiency, explainability, and data availability are far from being solved. **Table 1** summarizes the major challenges identified in the literature and outlines promising directions for future research. The development of efficient, scalable, explainable, and physically consistent foundation models should contribute greatly to scientific machine learning in the years ahead.

Table 1: Challenges and future developments in physics informed and hybrid scientific modelling

Research Area	Current Limitation	Potential Future Direction
3D Turbulence Modelling	PINNs struggle with scalability and fine-scale turbulence resolution	Development of scalable operator-learning methods such as FNOs
Uncertainty Quantification	Most models provide deterministic point predictions	Integration of probabilistic and Bayesian learning frameworks
Sparse and Noisy Data	Scientific observations are often limited and corrupted by noise	Improved physics-guided learning under low-data regimes
Multiscale and Multiphysics Systems	Existing models struggle to couple multiple physical processes	Unified hybrid frameworks for multiscale scientific systems
Benchmarking and Standardization	Lack of common datasets and evaluation protocols	Open benchmarks and reproducible evaluation pipelines
Explainability	Hybrid AI systems remain partially interpretable	Development of explainable physics-AI methodologies
Active Learning	Experimental data acquisition is expensive	Adaptive sampling and information-efficient learning
Foundation Models for Science	Scientific foundation models remain computationally expensive and underdeveloped	Domain-specific pretrained scientific models

These future directions indicate that the integration of physics and artificial intelligence is still in its early stages, with substantial opportunities remaining for the development of more reliable, scalable, and interpretable scientific learning systems.

10. CONCLUSION AND OUTLOOK

10.1 From Classical Physics to Scientific Machine Learning

Scientific modelling throughout history is characterized by attempts to increase understanding, prediction and simulation of nature. Since the emergence of deterministic laws in classical mechanics, the concept of physical phenomena modelling based on governing equations of physics has been formed. Later, the area was broadened significantly as a result of new field theories, statistical mechanics, and computational science. However, when problems became even more nonlinear, high-dimensional, and complex for computational analysis, pure approaches based on first principles have become less efficient due to their inability to model turbulence, biochemistry, climate and other complicated phenomena.

The rapid growth of machine learning and artificial intelligence led to a major shift and the introduction of modelling approaches based on pattern recognition from data rather than physical laws. These approaches enabled the approximation of nonlinear systems, acceleration of simulations, and discovery of hidden relationships within scientific datasets. However, pure data-driven models had low physical interpretability, robustness and consistency with scientific laws.

Thus, a need has emerged for approaches that combine physical knowledge with machine learning. Hybrid modelling or scientific machine learning combines physical and learning approaches in one computational system. It has seen several successful developments, among which physics-informed neural networks (PINN), operator learning, surrogate models, and physics-based learning are to be highlighted. All of the mentioned methods unite the advantages of both approaches by combining machine learning predictive power with the reliability and physical interpretability of scientific laws. Therefore, scientific modelling has started to move away from purely equation-based computational techniques to adaptive hybrid models combining theory, simulation, data and learning in one framework.

10.2 The Emerging Scientific Paradigm

The integration of artificial intelligence and physical modelling is not only an improvement of technology but also a more profound transformation of scientific research techniques in general. Historically, science mainly relied on two major approaches – theoretical modelling and experimentation. Later, simulation became the third technique that was widely used for science purposes. Today, machine learning is increasingly emerging as a fourth pillar of scientific investigation alongside theory, experimentation, and computational simulation. Machine learning does not replace physics in its function, but rather serves as a computational tool that can speed up simulations, approximate unknown relations, deal with large datasets, and help conduct scientific reasoning.

Through such an approach, scientists get a possibility to explore the systems which before were too computationally intensive and/or analytically complex to examine. Hybrid methods allow combining observational information with laws of physics, which helps to make modelling of complex systems faster. Therefore, the association between AI and physics is becoming more collaborative, leading to a new era of adaptive physical models for science.

10.3 Remaining Challenges

Despite advances made in the field of scientific machine learning, several problems have yet to be solved. Scalability remains an issue in implementing many hybrid learning systems because they rely on high-dimensional and multiscale computations. Physics-informed learning approaches often require significant computation resources when optimizing solutions.

Scientific datasets usually face issues such as sparsity, noise, incompleteness, and expense of acquisition. While physics-informed machine learning approaches could help address the issue, predictions will still be highly dependent on the availability and accuracy of the data. Reliability and generalizability remain major concerns since machine learning models tend to function well only in circumstances similar to the conditions used for training. However, scientific applications often demand models to extrapolate accurately into new physical situations.

In addition, interpretability and explainability become significant concerns since scientific knowledge and safety should not be sacrificed in favour of prediction performance. In fields such as engineering, medicine, climate science, and energy systems, it is crucial to understand the underlying mechanisms that make the predictions accurate. Finally, a set of benchmarking protocols has yet to be defined to test different approaches to hybrid scientific machine learning techniques. Solving these challenges will ensure that scientific machine learning systems are both reliable and robust.

10.4 Long-Term Outlook

The future course of scientific modelling may well see a rise in intelligent and self-sufficient computational systems that can combine physical reasoning, observations, and learning. One such promising direction is digital twins, virtual replicas of physical systems that can be continuously updated using live data streams. Digital twins might be able to provide real-time monitoring, optimization, and predictive analytics in areas ranging from industrial engineering, climate science, and medicine.

Active learning and experiment assistance by artificial intelligence may also emerge as an important research direction, in which machine learning algorithms are used to help researchers determine which experiments or simulations will give the most useful results. It would help make scientific discovery faster and cheaper. Scientific foundation models, similar to foundation models currently used for natural language processing and computer vision, may also grow in popularity in the future. Scientific foundation models might be pretrained on huge simulation and experimental datasets and then transferred to multiple scientific problems and tasks, similarly to how large natural language processing and computer vision models are applied today.

In addition, operator learning and advanced hybrid architectures may dramatically improve large-scale simulations in fluid dynamics, molecular dynamics, weather forecasting, and materials science. Overall, these trends suggest that future scientific modelling systems might not only simulate physical systems but also help scientists formulate hypotheses and discover novel physical laws from data.

10.5 Final Remarks

Physical modelling can be traced back to human beings' persistent drive to understand the principles of nature. The introduction of classical physics led to the development of a framework on which mathematical modelling of physical events could be done with a high degree of accuracy. Computational science contributed to these achievements in that it made use of numerical simulations to study physical processes that could not be modelled analytically. Today, the advancement in machine learning and artificial intelligence is bringing yet another wave of change to this process. Instead of competing with classical physics, the emerging field of scientific machine learning attempts to unite physical laws, computations, and machine learning in single models. Despite facing many unresolved issues related to scalability, uncertainty, interpretability, and robustness, the integration of physics and artificial intelligence represents one of the most important innovations in scientific computing today. The future of scientific modelling will likely depend not on choosing between physics and artificial intelligence, but on developing systems capable of combining physical reasoning, computational simulation, and adaptive learning into unified frameworks for scientific discovery.

REFERENCES

1) Boltzmann, L. (1970). 2. Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen. In *Einführung und Originaltexte* (pp. 115–225). De Gruyter.
https://doi.org/doi:10.1515/9783112596760-011

2) Brunton, S. L., Proctor, J. L., & Kutz, J. N. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, *113*(15), 3932–3937.
https://doi.org/10.1073/pnas.1517384113

3) Courant, R., Friedrichs, K., & Lewy, H. (1967). On the Partial Difference Equations of Mathematical Physics. *IBM Journal of Research and Development*, *11*(2), 215–234.
https://doi.org/10.1147/rd.112.0215

4) Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, *2*(4), 303–314.
https://doi.org/10.1007/BF02551274

5) Farea, A., Yli-Harja, O., & Emmert-Streib, F. (2024). Understanding physics-informed neural networks: Techniques, applications, trends, and challenges. *Ai*, *5*(3), 1534–1557.

6) Forrester, A. I. J., & Keane, A. J. (2009). Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences*, *45*(1), 50–79.
https://doi.org/https://doi.org/10.1016/j.paerosci.2008.11.001

7) Fourier, J. B. J. (2009). Théorie Analytique de la Chaleur. In *Cambridge Library Collection - Mathematics*. Cambridge University Press.
https://doi.org/DOI: 10.1017/CBO9780511693229

8) Gibbs, J. W. (2010). Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundation of Thermodynamics. In *Cambridge Library Collection - Mathematics*. Cambridge University Press.
https://doi.org/DOI: 10.1017/CBO9780511686948

9) Gorriz, J. M., Ramirez, J., Segovia, F., Jimenez-Mesa, C., Martinez-Murcia, F. J., & Suckling, J. (2026). Statistical agnostic regression: A machine learning method to validate regression models. *Journal of Advanced Research*, *80*, 503–533. https://doi.org/https://doi.org/10.1016/j.jare.2025.04.026
10) He, W., Jiang, Z., Xiao, T., Xu, Z., & Li, Y. (2026). A survey on uncertainty quantification methods for deep learning. *ACM Computing Surveys*, *58*(7), 1–35.
11) Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, *3*(6), 422–440. https://doi.org/10.1038/s42254-021-00314-5
12) Khanra, S., Kukreja, V. K., & Bala, I. (2026). Physics-informed neural networks for differential equation solutions: A comprehensive review. *Neurocomputing*, *680*, 133317. https://doi.org/https://doi.org/10.1016/j.neucom.2026.133317
13) LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539
14) Maxwell, J. C. (1865). VIII. A dynamical theory of the electromagnetic field. *Philosophical Transactions of the Royal Society of London*, (155), 459–512. https://doi.org/10.1098/rstl.1865.0008
15) Navier, C. (1823). Mémoire sur les lois du mouvement des fluides. *Mémoires de l'Académie Royale Des Sciences de l'Institut de France*, *6*(1823), 389–440.
16) Newton, I. (1687). *Philosophiae naturalis principia mathematica*. Jussu Societatis Regiae ac Typis Josephi Streater. Prostat apud plures bibliopolas. https://doi.org/10.5479/sil.52126.39088015628399
17) Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191
18) Polke, D., Ahle, E., & Söffker, D. (2026). Adaptive Learning with Gaussian Process Regression: A Comprehensive Review of Methods and Applications. *Machine Learning and Knowledge Extraction*, *8*(4), 101.
19) Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, *378*, 686–707. https://doi.org/https://doi.org/10.1016/j.jcp.2018.10.045
20) Rasmussen, C. E., & Williams, C. K. I. (2005). *Gaussian Processes for Machine Learning*. The MIT Press. https://doi.org/10.7551/mitpress/3206.001.0001
21) Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1
22) Schrödinger, E. (1926). Quantisierung als Eigenwertproblem. *Annalen Der Physik*, *384*(6), 489–527. https://doi.org/https://doi.org/10.1002/andp.19263840602
23) Schütt, K., Kindermans, P.-J., Sauceda Felix, H. E., Chmiela, S., Tkatchenko, A., & Müller, K.-R. (2017). Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. *Advances in Neural Information Processing Systems*, *30*.
24) Stokes, G. G. (2007). *On the theories of the internal friction of fluids in motion, and of the equilibrium and motion of elastic solids*.
25) Wu, Y., Sicard, B., & Gadsden, S. A. (2024). Physics-informed machine learning: A comprehensive review on applications in anomaly detection and condition monitoring. *Expert Systems with Applications*, *255*, 124678. https://doi.org/https://doi.org/10.1016/j.eswa.2024.124678
26) Zienkiewicz, O. C. (1971). *The finite element methods in engineering science*. McGraw-Hill.