Baku State University: Journal of Physics & Space Sciences, **2026**, v3 (1), p. 49-56

journal homepage: http://bsuj.bsu.edu.az/en

Machine learning in physics and semiconductor materials: approaches to modeling and prediction

Vusala J. Mammadova*

Department of Optics and Molecular Physics, Baku State University, Baku, Azerbaijan
R e c e i v e d 29-Dec-2025; A c c e p t e d 12-Feb-2026
DOI: https://doi.org/10.30546/209501.101.2026.3.01.316

Abstract

Machine learning (ML) has recently emerged as a powerful complement to traditional theoretical and computational methods in physics and materials science. It enables faster and more efficient predictions of physical properties compared to resource-intensive approaches such as Density Functional Theory (DFT). In this work, ML applications in physical systems are systematically analyzed within a unified conceptual framework encompassing surrogate models, physics-informed approaches, and hybrid methods. Attention is given to semiconductor materials, where ML significantly accelerates the prediction of energy gaps and optical properties based on composition and structural parameters. Overall, the integration of ML with physics enhances modeling flexibility and provides new opportunities for data-driven materials design.

Keywords: Machine Learning, Semiconductor Materials, Density Functional Theory, Surrogate Modeling, Physics-Informed Neural Networks, Band Gap Prediction, Optical Properties
PACS Numbers: 07.05. Mh; 07.05. Kf; 72.80. Ey; 89.20. Ff.

1. Introduction

Physics is a fundamental field of science aimed at revealing the laws governing natural phenomena. From classical mechanics to quantum mechanics, all major theories are based on the mathematical description of observed processes. However, the accurate modeling of complex systems, particularly many-body quantum systems and real materials, often requires significant computational resources.

*Corresponding author – Tel.: (+994) 77 616 90 16
e-mail address: vusala.mammadova@bsu.edu.az; *ORCID ID*: 0009-0008-2256-2203.

In this context, machine learning (ML) methods have emerged as a promising approach. These methods enable the discovery of underlying patterns in data and allow for the prediction of physical properties, complementing traditional models and making the modeling process more flexible and efficient.

In recent years, ML has moved beyond data analysis and has begun to play an active role in the modeling of physical systems. For instance, neural networks have been successfully applied to represent quantum wavefunctions, providing new computational possibilities.

The role of ML is particularly significant in semiconductor systems. Determining properties such as band gap energy, refractive index, and other optoelectronic characteristics using conventional methods require considerable time and computational effort. In contrast, ML models can learn the relationships between compositional and structural parameters and these properties, significantly accelerating the prediction process.

This paper reviews the application of ML methods in physics and semiconductor systems. The main novelty of this work lies in the systematic organization of these approaches within a unified conceptual framework.

2. Theoretical Foundations of Machine Learning

The primary goal of machine learning methods is to identify patterns present in data and express them in the form of a generalized model. In this sense, the ML approach has a conceptual similarity to classical physics: in both cases, the aim is to determine the functional relationships underlying observed phenomena [1].

Within the ML framework, this problem is formulated as a function approximation task. The function that maps input variables $\boldsymbol{x}$to the corresponding output variable $\boldsymbol{y}$can be written as follows [1]:

$$y = f(x)$$

However, in modern approaches, this relationship is no longer considered purely deterministic but rather probabilistic, and can be expressed as [1]:

$$p(y \mid x)$$

This perspective is particularly important for physical systems, since both experimental measurements and theoretical calculations are often accompanied by uncertainties. Therefore, representing results in terms of probability distributions provides a more realistic description.

The model is constructed through learning from data. In supervised learning, the model is trained using input–output pairs and gains the ability to make predictions for new data. This approach is widely used in predicting physical parameters.

The effectiveness of a model depends on its complexity and its ability to genera-

lize.

An overly simple model fails to capture the underlying patterns in the data, while an overly complex model tends to memorize the training data and performs poorly on new cases. This balance is known as the bias–variance trade-off and is one of the central challenges in building ML models [1, 9]. This issue is particularly relevant in the study of material systems, where data are often limited and the model's generalization capability becomes critical [4].

In probabilistic approaches, model parameters are estimated using the likelihood function. The likelihood measures how well a model explains the observed data, and in Maximum Likelihood Estimation (MLE), the parameters are chosen to maximize this function [1, 9].

A more general approach is the Bayesian method. In this framework, prior knowledge (prior) is combined with observed data (likelihood) to obtain the posterior distribution:

$$p(\theta \mid D) \propto p(D \mid \theta)\, p(\theta)$$

This approach allows physical constraints to be incorporated into ML models and improves the interpretability of the results [1]. Bayesian methods play an important role in estimating uncertainties and assessing the reliability of predictions [5,6].

Thus, machine learning is not limited to identifying functional relationships but represents a comprehensive modeling framework that integrates probability theory, optimization, and statistical methods. This makes ML highly effective for studying physical systems and complex material structures.

3. Integration of Machine Learning and Physics

Initially, the relationship between machine learning and physics was largely limited to data analysis. However, in recent years, this interaction has become deeper and more mutually influential. ML methods are no longer used only for processing experimental or computational results but also play an active role in modeling physical systems.

This integration can be broadly categorized into three main directions: surrogate models, data-driven approaches, and hybrid methods enriched with physical constraints.

3.1 Surrogate Models

Many computational methods in physics, particularly those based on quantum mechanics such as Density Functional Theory (DFT), provide high accuracy but require significant computational time. For this reason, ML models are not used as direct replacements for these methods, but rather as their fast approximations –

so-called surrogate models [3].

In this approach, high-accuracy calculations are performed only for a limited number of systems, and the results obtained are then used to train an ML model. Once trained, the model can predict the same properties for new systems in a much shorter time.

The general principle of this process is illustrated in Figure 1 (prepared by the author). As shown, the initial data obtained from high-accuracy methods are used to construct the ML model, which then enables rapid evaluation of material properties.

It has been shown that this approach is effective for the rapid evaluation and optimization of material properties [3].

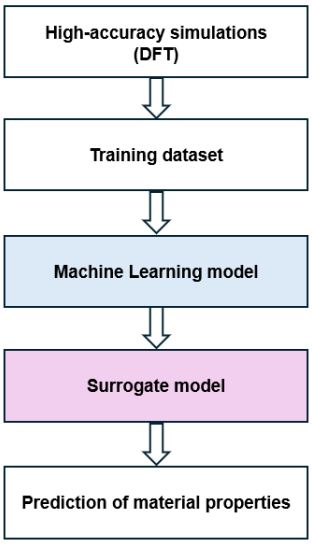

Fig. 1. Schematic representation of the surrogate model approach: the ML model trained on data obtained from DFT calculations enables rapid prediction of the properties of new materials (prepared by the author).

3.2. Machine Learning as a Physical Model

One of the recent developments is the use of machine learning models for the direct representation of physical systems. In particular, the work of Giuseppe Carleo and collaborators has demonstrated that neural networks can be applied to model the wavefunctions of quantum systems [8].

The general principle of this approach is illustrated in Figure 2. As shown, the representation of the physical system (such as the wavefunction or other forms of representation) is provided as input to the machine learning model, which learns

the system's behavior and reproduces its underlying physical relationships. As a result, the ML model is not limited to data approximation but acts as an effective model of the physical system, enabling the prediction of its properties.

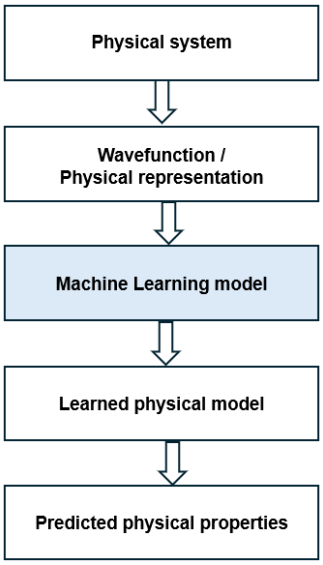

Fig. 2. Schematic representation of machine learning as a physical model: the ML model, constructed based on the representation of a physical system, learns its behavior and enables the prediction of physical properties (prepared by the author).

3.3. Physics-Informed Neural Networks (PINNs)

One of the main challenges of machine learning models is that they may sometimes produce results that are inconsistent with physical laws. To address this limitation, approaches have been developed in which physical constraints are directly incorporated into the model.

One of the most important approaches in this direction is the Physics-Informed Neural Networks (PINNs) method [5]. In this framework, the neural network is not trained solely on data but also incorporates physical laws–such as differential equations and boundary conditions–into the loss function.

The general principle of this approach is illustrated in Figure 3. As shown, the machine learning model operates together with physical constraints, ensuring both consistency with the data and adherence to physical laws. As a result, the model functions not only as an approximation tool but also as a more reliable representation of the physical system.

These characteristics make the PINNs approach particularly effective in situations where experimental data are limited.

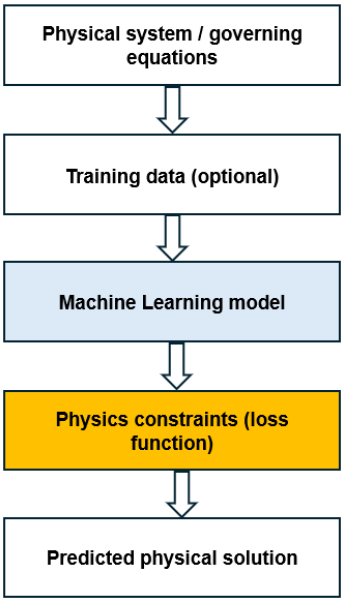

Fig. 3. Schematic representation of the Physics-Informed Neural Networks (PINNs) approach: the machine learning model, integrated with physical laws (constraints and loss function), enables more accurate and reliable modeling of physical systems (prepared by the author).

3.4. Hybrid Approaches and Future Perspectives

One of the most promising directions in modern research is the development of hybrid approaches that integrate machine learning (ML) with physical models. In this approach, knowledge obtained from physical models is combined with machine learning methods to form a more powerful modeling framework.

As illustrated in Figure 4, this process involves constructing an ML model based on both physical models and available data, followed by the integration of these two approaches into a hybrid model. As a result, the model can simultaneously learn from data and adhere to physical laws.

Studies by Butler et al. and Himanen et al. have shown that this approach can significantly accelerate the discovery and optimization of materials [2,7]. At the same time, the development of interpretable ML models allows for a deeper understanding of the physical meaning of the obtained results.

4.Machine Learning Approaches in Semiconductor Systems

In recent years, the application of machine learning (ML) methods in semiconductor systems has gained significant attention. The main objective of this approach is to identify the complex relationships between the chemical composition of a material and its physical and optical properties. Traditionally, such relationships have been analyzed using quantum mechanical methods, particularly Density Functional Theory (DFT). However, due to their high computational cost and time re-

quirements, ML models have emerged as an effective alternative for accelerating this process.

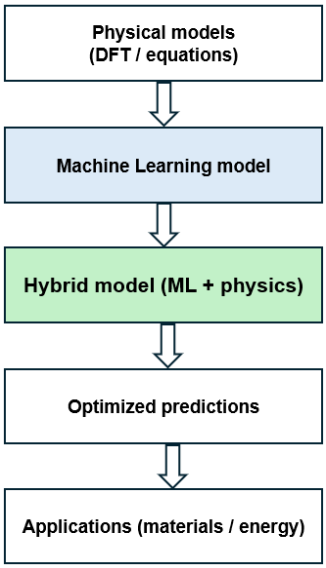

Fig. 4. Schematic representation of hybrid approaches: the integration of physical models and machine learning methods enables more accurate and optimized predictions, with broad applications in materials and energy systems (prepared by the author).

One of the key parameters of semiconductor materials is the band gap energy (Eg), which determines their optoelectronic properties and plays a crucial role in devices such as solar cells and photodetectors. ML models are widely used to predict Eg for various material systems. These models are trained using chemical composition, atomic radius, crystal structure parameters, and other characteristic features, enabling fast and reasonably accurate predictions for new materials [2, 10].

In addition, the prediction of optical properties such as the refractive index can also be effectively performed using ML approaches. While classical empirical and semi-empirical models rely on certain simplifications, ML models can capture more complex and nonlinear relationships.

These approaches become particularly clear when applied to specific material systems. For example, in solid solution systems such as $Cd_{1-x}Zn_xO$, changes in composition lead to variations in band gap energy and optical properties. In such systems, ML models can be used to determine composition–property relationships, identify optimal material compositions, compare experimental and theoretical results, and predict new materials.

Thus, ML approaches in semiconductor research are not limited to explaining existing properties but also enable a more systematic and efficient material design process. In general, these methods significantly accelerate computational workflows and allow for a deeper understanding of complex physical relationships.

5. Conclusion

This study shows that machine learning methods, when used alongside traditional approaches, provide more effective solutions for modeling physical systems. Surrogate models and data-driven approaches enable faster solutions to complex problems, while models integrated with physical constraints ensure more reliable and robust results.

In semiconductor systems, these approaches facilitate the rapid identification of composition–property relationships and support material optimization. This demonstrates that ML is not only a tool for analysis but also a powerful approach for material discovery and design.

Overall, the integration of machine learning with physics represents a new stage in the development of modeling approaches and opens opportunities for constructing more reliable and interpretable models in future research.

References

[1] Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, **2006**.

[2] Butler, K.T.; Davies, D.W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559*, 547–555.

[3] Schmidt, J.; Marques, M.R.G.; Botti, S.; Marques, M.A.L. Recent advances and applications of machine learning in solid-state materials science. *npj Comput. Mater.* **2019**, *5*, 83.

[4] Morgan, D.; Jacobs, R. Opportunities and challenges for machine learning in materials science. *Annu. Rev. Mater. Res.* **2020**, *50*, 71–103.

[5] Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems. *J. Comput. Phys.* **2019**, *378*, 686–707.

[6] Carleo, G.; Cirac, I.; Cranmer, K.; Canziani, A.; Feuerstein, G.; Fuentes, J.S.; Lauk, N.; Lanzieri, M.; Peotta, S.; Provasoli, L.; et al. Machine learning and the physical sciences. *Rev. Mod. Phys.* **2019**, *91*, 045002.

[7] Himanen, L.; Geurts, A.; Foster, A.S.; Rinke, P. Data-driven materials science: status, challenges, and perspectives. *Adv. Sci.* **2019**, *6*, 1900808.

[8] Carleo, G.; Troyer, M. Solving the quantum many-body problem with artificial neural networks. *Science* **2017**, *355*, 602–606.

[9] Wang, A.Y.-T.; Murdock, R.J.; Kauwe, S.K.; Oliynyk, A.O.; Gurlo, A.; Brgoch, J.; Persson, K.A.; Sparks, T.D. Machine learning for materials scientists: An introductory guide toward best practices. *Chem. Mater.* **2020**, *32*, 4954–4965.

[10] Zhou, Q.; Tang, P.; Liu, S.; Pan, J.; Yan, Q.; Zhang, S.-C. Learning atoms for materials discovery. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E6411–E6417.