Global Journal of Physics
Vol. 1 No.1 2026 Page 41-51
E-ISSN: XXXX-XXXX, DOI: 10.XXXXX/XXXX

Enhancing Data-Driven Physics Models through the Integration of Fundamental Theory and Physics-Informed Learning

Sari Fadillah[1]*, Taufik Pratama[2], Utami Wibowo[3]
[1,2] Department of Mathematics Education, Faculty of Teacher Training and Education, Universitas Pendidikan Indonesia, Bandung, Indonesia
[3] Department of Educational Technology, Faculty of Education, Universitas Negeri Malang, Malang, Indonesia.

Article Info

Received : 4 January 2026
Revised : 15 March 2026
Accepted : 29 March 2026

Keyword:

Data-driven physics models; Physics-informed learning; Scientific machine learning; Fundamental theory.

ABSTRACT

This study aims to analyze how the integration of fundamental physics theory and physics-informed learning enhances data-driven physics models. The study addresses the growing challenge of developing models that are not only accurate in prediction but also consistent with physical laws, especially when data are limited, noisy, or incomplete. A qualitative approach with an exploratory case study design was used. Data were collected through semi-structured interviews, limited observation, and documentation involving lecturers, researchers, postgraduate students, and scientific computing practitioners with experience in computational physics, machine learning, numerical simulation, or physics-informed learning. The data were analyzed using thematic analysis with inductive and deductive coding. The findings reveal five main themes: physics theory as a guide and constraint for model development, physics-informed learning as a response to data limitations, validation and interpretability challenges, the importance of documentation and transparency, and the need for interdisciplinary collaboration. These findings show that the quality of data-driven physics models depends not only on algorithmic sophistication, but also on physical validity, methodological reflection, and transparent research practice. The study contributes to scientific machine learning by positioning physics-informed learning as an epistemic bridge between theory and data. It also suggests the need for stronger scientific computing literacy, clearer model reporting standards, and further research across broader fields of applied physics and engineering.

***) Corresponding Author:**

Sari Fadillah,
Department of Mathematics Education, Faculty of Teacher Training and Education, Universitas Pendidikan Indonesia, Bandung, Indonesia
Email: sari.fadillah065@gmail.com

1. Introduction

The rapid development of machine learning has changed the way modern physics models are designed, trained, and evaluated. Data-driven models can process experimental data, simulation outputs, and sensor-based observations to predict complex physical phenomena. However, models that rely only on data often face serious limitations when the available data are scarce, noisy, incomplete, or unable to represent the full behavior of a physical system. Karniadakis et al. (2021) explain that physics-informed machine learning has emerged to address this limitation by embedding physical laws into the learning process. This approach is important because physics models should not only achieve predictive accuracy. They should also remain consistent with theory, differential equations, boundary conditions, conservation principles, and the causal structure of physical phenomena.

At the global level, the need for models that integrate theory and data has increased across several fields, including fluid mechanics, heat transfer, geoscience, manufacturing, energy systems, and nonlinear dynamics. Lu et al. (2021) show that Physics-Informed Neural Networks can solve forward and inverse problems by combining observational data with differential equations. Chen et al. (2021) also demonstrate that physics-informed learning can support the discovery of governing equations from scarce data. In fluid mechanics, Cai et al. (2021) explain that PINNs have strong potential because they can handle noisy data, complex domains, and inverse problems. These studies indicate that data-driven physics models should not only imitate patterns in data. They also need to preserve the scientific structure that explains physical phenomena.

Despite this progress, the implementation of physics-informed learning still faces several methodological challenges. Wang et al. (2022) found that PINNs may fail to train because of gradient imbalance, spectral bias, and different convergence rates across loss components. Yang et al. (2021) add that noisy data and parameter uncertainty can affect prediction quality. Therefore, Bayesian approaches are needed to estimate model uncertainty. Linka et al. (2022) also show that real-world nonlinear dynamical systems require models that are not only accurate but also capable of quantifying uncertainty and maintaining physical consistency. These findings show that the main issue in developing data-driven physics models is not only algorithm selection. It also concerns how researchers interpret the relationship among theory, data, assumptions, and scientific validation.

In the national context, this issue is relevant to the strengthening of computational research, physics education, engineering, energy, environmental studies, and industrial technology in Indonesia. Universities and research institutions have started to adopt simulation-based approaches, artificial intelligence, and scientific computing. However, the conceptual and methodological understanding of how to integrate physical laws into data-driven models still needs further development. Haghighat and Juanes (2021) show that tools such as SciANN can help researchers build neural network-based scientific models. Zobeiry and Humfeld (2021) demonstrate that physics-informed approaches can support the solution of heat transfer problems in manufacturing and engineering applications. Pachalieva et al. (2022) also show that the integration of full-physics models and machine learning can support subsurface reservoir pressure management. These studies confirm that the integration of theory and data has strong practical value. However, its success depends on researchers' ability to understand physical assumptions, model structures, and the limits of data use.

Previous studies on physics-informed learning have mostly focused on technical aspects, such as network architecture, loss functions, training stability, prediction accuracy, and computational efficiency. Cuomo et al. (2022) explain that the development of PINNs has been largely directed toward architecture optimization, training techniques, and applications to differential equations. Li et al. (2024) extend this approach through physics-informed neural operators for learning partial differential equations more flexibly. Jagtap et al. (2020) develop conservative physics-informed neural networks to preserve conservation laws in complex domains. Seyyedi et al. (2024) map the relationship between machine learning and physics as a growing interdisciplinary field. However, these studies have not fully explored how researchers qualitatively understand, negotiate, and apply fundamental theory during model development. This gap is important because methodological decisions in modeling are often shaped by experience, interpretation, data limitations, and researchers' conceptual understanding of physical laws.

Based on this background, this study aims to analyze qualitatively how the integration of fundamental theory and physics-informed learning can enhance data-driven physics models. The study focuses on the meaning of theory-data integration, the role of physical laws in the learning process, researchers' strategies for dealing with limited data, and the challenges of model validation and interpretability. Theoretically, this study contributes to the development of scientific machine learning by positioning physical theory as an epistemic foundation in model construction. Practically, this study can serve as a reference for researchers, lecturers, postgraduate students, and computational model developers in designing data-driven physics models that are not only accurate but also consistent, interpretable, and relevant to the physical phenomena under investigation.

2. Materials and Methods

This study uses a qualitative approach with an exploratory case study design. This design was selected because the study focuses on an in-depth understanding of researchers' processes, experiences, and methodological considerations in integrating fundamental physics theory with physics-informed learning. A case study is appropriate because the phenomenon under investigation occurs in a real context and involves scientific actors, computational practices, technical documents, and interrelated methodological decisions. Miller et al. (2023) explain that qualitative case studies help researchers understand a phenomenon in detail through multiple sources of data. Mtisi (2022) also states that case studies allow researchers to examine processes, actors, and contexts within a bounded system.

The study is conducted within academic and research communities that work in computational physics, scientific machine learning, numerical simulation, and Physics-Informed Neural Networks. The research sites include computational laboratories, physics study programs, artificial intelligence research centers, or interdisciplinary research groups at universities and research institutions. The study is designed to be conducted over three months, covering instrument preparation, data collection, data processing, thematic analysis, and validation of findings. The selection of these sites and this research period is based on the need to understand how researchers integrate data, physical theory, differential equations, boundary conditions, system parameters, and machine learning techniques in model development.

The participants consist of lecturers, researchers, postgraduate students, and scientific computing practitioners who have experience in developing or using data-driven physics models. Informants are selected through purposive sampling. The selection criteria include experience in computational physics modeling, machine learning, numerical simulation, or physics-informed learning. Informants must also have experience using experimental data, simulation data, or observational data to build predictive models. In addition, informants should understand the role of physical theory, governing equations, boundary conditions, system parameters, or model validation in the modeling process. The number of informants is flexible and depends on information sufficiency, with an initial range of 8 to 15 informants. If the initial data indicate the presence of other relevant informants, snowball sampling is used to reach additional participants with more specific expertise.

Data are collected through semi-structured interviews, limited observation, and documentation. Semi-structured interviews are used as the main technique because they allow the researcher to prepare guiding questions while still giving informants space to explain their experiences openly. Adeoye-Olatunde and Olenik (2021) explain that semi-structured interviews help researchers obtain rich data because core questions can be expanded through follow-up questions based on participants' responses. The interview topics include informants' understanding of data-driven physics models, reasons for embedding fundamental theory into models, experiences in using physics-informed learning, validation challenges, data limitations, model uncertainty, and considerations in interpreting prediction results. Limited observation is conducted during research discussions, code demonstrations, model presentation sessions, or model validation processes when access is available. Documentation includes scientific articles, research notes, open-source codes, experimental reports, model flow diagrams, presentation materials, and other relevant supporting documents.

Data validation is conducted through source triangulation, method triangulation, member checking, and audit trail. Source triangulation is carried out by comparing information from lecturers, researchers, postgraduate students, and scientific computing practitioners. Method triangulation is conducted by comparing interview data, observation notes, and supporting documents. Member checking is carried out by sending interview summaries or initial interpretations to informants to ensure that the meanings captured by the researcher align with participants' experiences. An audit trail is maintained by documenting research decisions, interview transcripts, analytic memos, coding changes, theme matrices, and supporting documents. Ahmed (2024) emphasizes that credibility, transferability, dependability, and confirmability are important principles in maintaining the trustworthiness of qualitative research. Johnson et al. (2020) also explain that qualitative research rigor should be maintained from the design stage to data collection, analysis, and reporting.

The data are analyzed using thematic analysis with a combination of inductive and deductive coding. The analysis begins with interview transcription, repeated reading, identification of meaning units, initial coding, code grouping, category formation, theme development, and theme review against the raw data. Inductive coding is used to capture experiences, terms, and meanings that emerge directly from informants. Deductive coding is used to connect the data with the initial concepts of the study, such as the integration of physical laws, model consistency, data limitations, interpretability, validation, and the role of theory in machine learning. Kiger and Varpio (2020) explain that thematic analysis helps researchers identify, organize, and interpret patterns of meaning in qualitative data systematically. Xu and Zammit (2020) also show that hybrid thematic analysis can combine inductive reading from the data with deductive reading based on a theoretical framework.

The analysis procedure is conducted in stages so that the findings can be traced and replicated in a limited way. In the first stage, the researcher reads all transcripts to understand the general context of informants' experiences. In the second stage, the researcher develops initial codes related to how informants understand the relationship among physical theory, data, and machine learning models. In the third stage, the codes are grouped into categories, such as the function of theory in constraining models, strategies for dealing with limited data, validation processes, computational barriers, and interpretability considerations. In the fourth stage, the categories are developed into main themes that explain the process of integrating fundamental theory and physics-informed learning in the development of data-driven physics models. In the final stage, the themes are compared again with transcripts, field notes, and supporting documents so that the interpretation remains grounded in empirical evidence. Ruslin et al. (2022) emphasize that qualitative instrument development should be conducted reflectively so that research questions, interview techniques, and interpretation processes remain aligned with the purpose of the study.

3. Results and Discussions

The results show that the integration of fundamental physics theory and physics-informed learning is understood as an important strategy for improving the quality of data-driven physics models. Based on semi-structured interviews, limited observation, and documentation, the participants did not view data-driven models only as computational tools for producing numerical predictions. They understood these models as scientific instruments that must remain connected to the physical laws underlying the observed phenomena. The thematic analysis produced five main themes: the role of physics theory as a guide and constraint for model development, the use of physics-informed learning to address data limitations, the challenges of validation and interpretability, the importance of documentation and transparency, and the need for interdisciplinary collaboration in developing data-driven physics models. These themes show that successful model development depends not only on algorithmic sophistication, but also on scientific reasoning, conceptual understanding, and careful methodological decisions.

The first theme concerns the role of physics theory as a guide and constraint for the model. Most participants stated that data-driven models have strong capacity to identify patterns, but they may produce predictions that do not reflect physical reality when they are not constrained by theory. In modeling practice, physics theory helps guide the model so that it does not merely follow statistical correlations in the data. Conservation laws, differential equations, boundary conditions, parameter relationships, and domain constraints are viewed as important components that keep the model within a physically plausible solution space. One participant stated, “If the model only learns from data, the result may look accurate, but it is not necessarily correct in terms of physics. Theory helps the model stay within the boundaries of the actual phenomenon.” This statement indicates that theory functions as a form of scientific control. It does not only provide a conceptual background. It also shapes the direction of the learning process.

The findings within the first theme also show that participants distinguish between predictive accuracy and physical validity. Predictive accuracy refers to the ability of a model to produce values close to testing data. Physical validity refers to the ability of a model to follow the laws, structures, and principles that explain the phenomenon. Several participants argued that a model may have a low error value but still be problematic if it violates conservation laws, produces unrealistic parameters, or fails to explain system behavior under certain boundary conditions. During research discussions, the observed researchers checked model outputs by comparing them with theoretical expectations. They did not immediately accept a model result only because it showed good numerical performance. They examined whether the result matched the expected theoretical curve, energy distribution, direction of variable changes, and acceptable physical limits. This pattern shows that model evaluation took place through two layers, namely statistical evaluation and physical evaluation.

The second theme shows that physics-informed learning is used as a strategy to address data limitations. Participants explained that experimental data in physics are often difficult to obtain because measurement costs are high, instruments are limited, sensor errors may occur, or observations must be conducted under extreme conditions. In some cases, physical phenomena occur too fast, too slowly, at very small scales, at very large scales, or in conditions that are unsafe for direct measurement. Under these circumstances, purely data-driven models often require large amounts of data to learn effectively. Physics-informed learning allows models to use information from physics theory so that the learning process does not fully depend on empirical data. One participant explained, “In some experiments, our data are incomplete. However, if the physical equation is known, the model can still be directed to produce more reasonable predictions.” This statement shows that physics theory is understood as an additional source of information that helps the model learn from limited data.

Data limitations also appear in the form of noisy, imbalanced, or incomplete data. Participants noted that sensor data often contain technical disturbances, simulation data may contain numerical assumptions, and experimental data are sometimes available only at limited observation points. In these situations, physics-informed learning helps reduce the risk of models learning misleading patterns. By embedding physical equations into the loss function, the model is directed to consider the relationship between predictions and the governing laws of the system. However, participants also emphasized that this approach does not remove the need for high-quality data. Data are still needed to provide empirical information, while theory helps guide the learning process to become more stable and meaningful. This finding shows that the relationship between data and theory is complementary. Data provide empirical evidence, while theory provides explanatory structure.

The third theme relates to the challenges of model validation and interpretability. Participants stated that integrating physics theory into a model does not automatically make the model easy to validate. Validation still requires careful examination of data, equations, parameters, training procedures, and prediction results. One challenge frequently mentioned by participants is balancing data loss and physics loss. If the data loss weight is too large, the model may overfit the data and ignore physical laws. If the physics loss weight is too large, the model may follow an idealized equation too strongly and become less sensitive to empirical conditions. One participant stated, “The difficult part is not only inserting equations into the model. The difficult part is ensuring that the equation matches the data and the condition of the system being studied.” This statement shows that validation in physics-informed learning cannot rely only on prediction error. It must also examine the fit between theory, data, and the context of the phenomenon.

Interpretability also emerged as an important issue. Several participants stated that neural networks still contain black-box characteristics even when physical constraints are included. Researchers may know that a model produces good predictions, but they may not fully understand how the model reaches those predictions. In some cases, the model performs well on a certain dataset but performs poorly when applied to a different domain or boundary condition. This creates questions about model generalization. Participants argued that interpretability can be strengthened through result visualization, parameter sensitivity analysis, equation residual checking, comparison with numerical solutions, and documentation of training procedures. This finding shows that interpretability is not only related to model architecture. It is also related to the analytical practices that researchers conduct after the model has been trained.

The fourth theme shows the importance of documentation and transparency in model development. Observation and documentation data indicate that more systematic researchers tend to document data sources, equation forms, model assumptions, parameters, boundary conditions, training configurations, and evaluation results. Such documentation helps other researchers understand how the model was built and why specific decisions were made. One participant stated, “In this kind of model, documentation is very important. If it is not recorded, other people will not know what assumptions were used and which parts truly follow physics.” This statement indicates that transparency is an important element of model credibility. Without adequate documentation, a model is difficult to replicate, review, or develop further.

Documentation also helps maintain continuity between theoretical processes and computational processes. In physics-informed learning, researchers need to record how physical equations are selected, how variables are defined, how data are normalized, how the loss function is formulated, and how training parameters are determined. These decisions are often not visible in the final model output, but they strongly affect prediction quality and interpretation. The technical documents analyzed in this study show that model flow diagrams, experiment notes, and open-source codes help explain the relationship among data, theory, and training procedures. This indicates that documentation is not merely an administrative supplement. It is part of the scientific validation process.

The fifth theme concerns the need for interdisciplinary collaboration. Participants argued that data-driven physics models cannot be developed optimally by one discipline alone. Physicists are needed to ensure that the model remains consistent with the phenomenon under investigation. Applied mathematicians are needed to help formulate equations, boundary conditions, solution stability, and numerical structures. Computational experts are needed to build network architecture, manage data, and optimize training. Domain practitioners are needed to assess whether model outputs are relevant to real-world needs. One participant explained, “Physicists understand the phenomenon, but computational implementation requires other expertise. Without collaboration, the model may be strong in theory but weak in implementation, or the opposite.” This finding shows that theory-data integration is a collaborative practice that requires communication across disciplines.

Interdisciplinary collaboration also influences how researchers interpret model results. During research discussions, different perspectives appeared when participants evaluated predictions. Physicists tended to question the consistency of results with physical laws and physical intuition.

Computational experts focused on training stability, optimization, and data quality. Applied mathematicians paid attention to equation forms, scale, and solution properties. These differences did not necessarily become barriers. Instead, they enriched the model evaluation process. Cross-disciplinary discussion helped identify weaknesses that might not be visible if the model were evaluated from only one perspective. Therefore, collaboration becomes an important mechanism for strengthening the scientific and technical quality of the model.

Overall, the findings show that the integration of fundamental theory and physics-informed learning improves the quality of data-driven physics models through several mechanisms. First, physics theory constrains the solution space so that the model does not produce predictions that conflict with physical principles. Second, theory helps the model learn when data are limited, noisy, or incomplete. Third, theory encourages a more critical validation process because the model must be evaluated based on both numerical accuracy and physical consistency. Fourth, documentation strengthens transparency and limited replication. Fifth, interdisciplinary collaboration improves researchers' ability to understand the model from conceptual, mathematical, computational, and applied perspectives. However, the findings also show several challenges, including the selection of appropriate equations, the alignment of assumptions with real data, the weighting of loss functions, uncertainty measurement, model interpretation, and effective communication across disciplines.

Discussion

The findings support the view that physics-informed machine learning is an important approach in the development of modern scientific models. Karniadakis et al. (2021) argue that integrating physical laws into machine learning can help overcome the limitations of purely data-driven models. The present study shows that participants view physics theory not as an additional component added after model construction, but as a central part of the modeling process. Physics theory functions as an epistemic foundation that determines how data are read, how models are guided, and how predictions are evaluated. Therefore, model quality should not be assessed only through numerical error. It should also be assessed through the model's ability to maintain consistency with the scientific structure of the physical phenomenon.

The finding that theory functions as a model constraint is also consistent with Jagtap et al. (2020), who emphasize the importance of conservation laws in conservative physics-informed neural networks. In this study, participants showed that constraining the model through physics theory helps reduce the risk of unrealistic predictions. This is important because machine learning models are highly flexible in adapting to data. This flexibility can be useful when the data are representative, but it can become a weakness when the data are limited or noisy. Physics theory helps reduce this weakness by limiting possible solutions. In this sense, theory serves as a form of scientific discipline that prevents the model from learning patterns without physical meaning.

The findings on data limitations support Lu et al. (2021), who show that PINNs can solve forward and inverse problems by combining data with differential equations. They also align with Chen et al. (2021), who explain that physics-informed learning can help discover governing equations from scarce data. In this study, data limitations are understood not only as a problem of data quantity, but also as a problem of quality, distribution, and coverage. Participants stated that limited data can make it difficult for models to recognize stable patterns. However, theory enriches the information available to the model. This perspective extends previous discussions by showing that theory not only improves technical performance but also supports researchers in making methodological decisions when data are not ideal.

The use of theory in modeling does not remove uncertainty. Yang et al. (2021) explain that Bayesian approaches are needed in PINNs to handle noisy data and uncertainty in PDE problems. Linka et al. (2022) also emphasize the importance of uncertainty estimation in real-world nonlinear dynamical systems. The present findings align with these studies. Participants explained that physics-informed models still need uncertainty measurement, especially when data contain noise, parameters are uncertain, or equations are simplified representations of real phenomena. Therefore, physics-informed learning should not be treated as an automatic guarantee of model truth. It still requires critical evaluation of data sources, theoretical assumptions, and the limits of model application.

The validation challenges identified in this study also support Wang et al. (2022), who found that PINNs may fail to train because of gradient imbalance and spectral bias. Participants described a similar challenge through their experience in determining the balance between data loss and physics loss. This finding is important because it shows that applying theory to machine learning is not a mechanical process. Researchers must understand how loss functions operate, how physical equations are translated into computational components, and how training results are evaluated. Without careful evaluation, the model may appear to follow physical theory but still produce unstable or difficult-to-explain predictions.

From the perspective of interpretability, the findings show that integrating physics can strengthen model explanation, but it does not fully remove the complexity of neural networks. Cuomo et al. (2022) explain that PINNs still face challenges in optimization, stability, and application expansion. This study adds that these challenges also have an interpretive dimension. Researchers need to know not only whether a model works, but also why it works, when it fails, and which part of the learning process is influenced by physical theory. Therefore, interpretability is not only a technical matter. It is also related to the researcher's ability to explain the relationship among the model, data, theory, and phenomenon.

The importance of documentation found in this study strengthens methodological standards in scientific model development. Documentation helps make the modeling process traceable. In scientific machine learning, documentation should not only cover data and code. It should also include the reasons for selecting equations, physical assumptions, normalization procedures, training configurations, validation strategies, and model limitations. This is relevant to scientific validity because poorly documented models are difficult for other researchers to verify. This finding provides a practical perspective that data-driven physics model development requires clearer reporting protocols. Such protocols can improve transparency, limited replication, and trust in research results.

The finding on interdisciplinary collaboration shows that theory-data integration does not only occur at the algorithmic level. It also occurs at the social and organizational levels of research. Model development involves interactions among physics knowledge, mathematics, computation, and domain experience. Seyyedi et al. (2024) describe the relationship between machine learning and physics as a rapidly growing interdisciplinary field. The present findings support this view by showing that cross-disciplinary collaboration is necessary to ensure that models are not only technically strong but also scientifically relevant. Without collaboration, a model may be computationally accurate but physically weak, or theoretically strong but difficult to implement.

In the Indonesian context, these findings are relevant for strengthening computational research and physics education. Universities and research institutions need capacity development that does not only focus on software use or algorithm implementation. They also need stronger conceptual understanding of the relationship between theory and data. Machine learning education in physics should train students to formulate problems, select appropriate equations, understand boundary conditions, evaluate data quality, and interpret model outputs. Therefore, scientific computing education should integrate theoretical physics, numerical methods, programming, and scientific reporting ethics. The findings also indicate the need to strengthen collaborative research culture so that students and researchers can work more effectively across disciplines.

Theoretically, this study offers the view that physics-informed learning can be understood as an epistemic integration between theory and data. In this practice, theory is not only a conceptual reference, and data are not only training materials. Both interact and shape each other in the modeling process. Data help test and adjust the model, while theory provides structure, boundaries, and meaning to the learning results. This perspective is important because much of the previous literature focuses more on computational performance. This study shows that model success is also influenced by how researchers understand theory, interpret data, and make methodological decisions. Therefore, the development of data-driven physics models should be seen as a scientific practice that combines computation, theory, interpretation, and validation.

Practically, the findings produce several implications. First, researchers need to prepare clear modeling guidelines before building a model, including the definition of the phenomenon, variables, governing equations, data sources, and validation strategies. Second, researchers need to conduct dual evaluation by assessing both statistical performance and physical consistency. Third, researchers need to document all important decisions so that the modeling process can be reviewed. Fourth, research institutions need to encourage collaboration among physicists, computational experts, applied mathematicians, and domain practitioners. Fifth, higher education institutions need to strengthen scientific machine learning literacy so that students do not only use algorithms but also understand the physical basis and methodological limits of the models they develop.

This study has several limitations. Since it uses a qualitative approach, the findings are not intended for statistical generalization. They are intended to provide an in-depth understanding of researchers' experiences and reasoning processes in integrating physics theory with physics-informed learning. Another limitation concerns possible differences across physics subfields. Researchers in fluid mechanics, geophysics, materials science, energy, or biomedical engineering may face different challenges when using data-driven models. In addition, this study depends on participants' openness in explaining their modeling experiences and practices. Therefore, documentation and observation were used to strengthen the validity of the findings.

Based on these limitations, future research can be directed toward several agendas. First, future studies can include participants from more applied fields, such as energy, materials science, geoscience, biomedical engineering, and industrial engineering. Second, future research can compare the experiences of novice and expert researchers in using physics-informed learning. Third, mixed-methods studies can examine the relationship among physics theory understanding, validation strategy, and model performance quality. Fourth, future studies can explore how computational infrastructure, research policy, and collaborative culture influence the adoption of scientific machine learning in universities and research institutions. Fifth, methodological studies can develop clearer reporting guidelines for data-driven physics models to support transparency and limited replication.

In conclusion, the results and discussion show that the integration of fundamental theory and physics-informed learning makes an important contribution to improving the quality of data-driven physics models. Physics theory helps keep models consistent with scientific laws. Data help adjust models to empirical evidence. Physics-informed learning connects both elements in the modeling process. However, the success of this approach still depends on researchers' conceptual understanding, careful validation, transparent documentation, and interdisciplinary collaboration. These findings contribute to qualitative studies in scientific machine learning by positioning researchers' experiences, reasoning, and methodological decisions as important elements in the quality of data-driven physics models.

4. Conclusions

This study concludes that the integration of fundamental physics theory and physics-informed learning strengthens data-driven physics models by linking predictive performance with physical consistency. The findings show that physics theory functions as a guide, constraint, and validation basis in the modeling process. It helps models remain aligned with governing equations, conservation principles, boundary conditions, and realistic system behavior. The study also shows that physics-informed learning is valuable when empirical data are limited, noisy, or incomplete, because theory can provide an additional structure that supports model learning. However, the findings confirm that this approach still requires careful validation, transparent documentation, uncertainty awareness, and interdisciplinary collaboration. Thus, this study contributes to the literature by showing that the quality of data-driven physics models depends not only on algorithmic performance, but also on researchers' conceptual reasoning and methodological decisions.

Theoretically, this study extends the discussion of scientific machine learning by positioning physics-informed learning as an epistemic integration between theory and data. Practically, the findings suggest that researchers need to evaluate models through both statistical accuracy and physical validity, document all modeling assumptions, and collaborate across physics, computation, mathematics, and applied domains. From a policy perspective, universities and research institutions need to strengthen scientific computing literacy, provide better computational infrastructure, and support interdisciplinary research ecosystems. Future studies should involve broader research fields, compare novice and expert researchers, and apply mixed-method designs to examine how theoretical understanding, validation strategy, and computational practice influence model quality.

References

Adeoye-Olatunde, O. A., & Olenik, N. L. (2021). Research and scholarly methods: Semi-structured interviews. *JACCP: Journal of the American College of Clinical Pharmacy, 4*(10), 1358–1367. https://doi.org/10.1002/jac5.1441

Ahmed, S. K. (2024). The pillars of trustworthiness in qualitative research. *Journal of Medicine, Surgery, and Public Health, 2*, Article 100051. https://doi.org/10.1016/j.glmedi.2024.100051

Cai, S., Mao, Z., Wang, Z., Yin, M., & Karniadakis, G. E. (2021). Physics-informed neural networks (PINNs) for fluid mechanics: A review. *Acta Mechanica Sinica, 37*(12), 1727–1738. https://doi.org/10.1007/s10409-021-01148-1

Chen, Z., Liu, Y., & Sun, H. (2021). Physics-informed learning of governing equations from scarce data. *Nature Communications, 12*, Article 6136. https://doi.org/10.1038/s41467-021-26434-1

Cuomo, S., Schiano di Cola, V., Giampaolo, F., Rozza, G., Raissi, M., & Piccialli, F. (2022). Scientific machine learning through physics-informed neural networks: Where we are and what's next. *Journal of Scientific Computing, 92*, Article 88. https://doi.org/10.1007/s10915-022-01939-z

Haghighat, E., & Juanes, R. (2021). SciANN: A Keras/TensorFlow wrapper for scientific computations and physics-informed deep learning using artificial neural networks. *Computer Methods in Applied Mechanics and Engineering, 373*, Article 113552. https://doi.org/10.1016/j.cma.2020.113552

Jagtap, A. D., Kharazmi, E., & Karniadakis, G. E. (2020). Conservative physics-informed neural networks on discrete domains for conservation laws: Applications to forward and inverse problems. *Computer Methods in Applied Mechanics and Engineering, 365*, Article 113028. https://doi.org/10.1016/j.cma.2020.113028

Johnson, J. L., Adkins, D., & Chauvin, S. (2020). A review of the quality indicators of rigor in qualitative research. *American Journal of Pharmaceutical Education, 84*(1), Article 7120. https://doi.org/10.5688/ajpe7120

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

Kiger, M. E., & Varpio, L. (2020). Thematic analysis of qualitative data: AMEE Guide No. 131. *Medical Teacher, 42*(8), 846–854. https://doi.org/10.1080/0142159X.2020.1755030

Li, Z., Kovachki, N. B., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A. M., & Anandkumar, A. (2024). Physics-informed neural operator for learning partial differential equations. *ACM Transactions on Mathematical Software, 50*(2), Article 21. https://doi.org/10.1145/3648506

Linka, K., Schäfer, A., Meng, X., Zou, Z., Karniadakis, G. E., & Kuhl, E. (2022). Bayesian physics informed neural networks for real-world nonlinear dynamical systems. *Computer Methods in Applied Mechanics and Engineering, 402*, Article 115346. https://doi.org/10.1016/j.cma.2022.115346

Lu, L., Meng, X., Mao, Z., & Karniadakis, G. E. (2021). DeepXDE: A deep learning library for solving differential equations. *SIAM Review, 63*(1), 208–228. https://doi.org/10.1137/19M1274067

Miller, E. M., Porter, J. E., & Barbagallo, M. S. (2023). Simplifying qualitative case study research methodology: A step-by-step guide using a palliative care example. *The Qualitative Report, 28*(8), 2363–2379. https://doi.org/10.46743/2160-3715/2023.6478

Mtisi, S. (2022). The qualitative case study research strategy as applied on a rural enterprise development doctoral research project. *International Journal of Qualitative Methods, 21*, 1–13. https://doi.org/10.1177/16094069221145849

Pachalieva, A., O'Malley, D., Harp, D. R., & Viswanathan, H. (2022). Physics-informed machine learning with differentiable programming for heterogeneous underground reservoir pressure management. *Scientific Reports, 12*, Article 18734. https://doi.org/10.1038/s41598-022-22832-7

Ruslin, R., Mashuri, S., Rasak, M. S. A., Alhabsyi, F., & Syam, H. (2022). Semi-structured interview: A methodological reflection on the development of a qualitative research instrument in educational studies. *IOSR Journal of Research & Method in Education, 12*(1), 22–29. https://doi.org/10.9790/7388-1201052229

Seyyedi, A., Bohlouli, M., & Nedaaee Oskoee, S. (2024). Machine learning and physics: A survey of integrated models. *ACM Computing Surveys, 56*(5), Article 115. https://doi.org/10.1145/3611383

Wang, N., Chen, Y., & Zhang, D. (2025). A comprehensive review of physics-informed deep learning and its applications in geoenergy development. *The Innovation Energy, 2*(2), Article 100087. https://doi.org/10.59717/j.xinn-energy.2025.100087

Wang, S., Yu, X., & Perdikaris, P. (2022). When and why PINNs fail to train: A neural tangent kernel perspective. *Journal of Computational Physics, 449*, Article 110768. https://doi.org/10.1016/j.jcp.2021.110768

Xu, W., & Zammit, K. (2020). Applying thematic analysis to education: A hybrid approach to interpreting data in practitioner research. *International Journal of Qualitative Methods, 19*, 1–9. https://doi.org/10.1177/1609406920918810

Yang, L., Meng, X., & Karniadakis, G. E. (2021). B-PINNs: Bayesian physics-informed neural networks for forward and inverse PDE problems with noisy data. *Journal of Computational Physics, 425*, Article 109913. https://doi.org/10.1016/j.jcp.2020.109913

Zobeiry, N., & Humfeld, K. D. (2021). A physics-informed machine learning approach for solving heat transfer equation in advanced manufacturing and engineering applications. *Engineering Applications of Artificial Intelligence, 101*, Article 104232. https://doi.org/10.1016/j.engappai.2021.104232