

A Data Driven Hybrid Computational Framework for Improving Efficiency and Accuracy in Scientific Research Processes

Dr.A. Anthoniammal ^{a*}, Dr. Arul Prasad ^b, Dr.G. Bhuvaneswari ^c,
Dr.P.T. Vijaya Rajakumar ^d, S. Vinothkumar ^e, Dr.S. Sudha ^f

^{a*} Assistant Professor, Department of Business Administration, Annamalai University, Cuddalore, India.

E-mail: anthoniammal1982@gmail.com, Orcid: <https://orcid.org/0009-0006-1950-9191>

^b Assistant Professor, PG and Research Department of Commerce, Annamalai University, Cuddalore, Tamil Nadu, India. E-mail: meetapsjc@gmail.com, Orcid: <https://orcid.org/0000-0002-8924-1112>

^c Professor, Department of Management Studies, Nehru Institute of Technology, Coimbatore, Tamil Nadu, India.

E-mail: drgbhuvaneswari@gmail.com, Orcid: <https://orcid.org/0009-0004-1124-3515>

^d Professor, Department of Management Studies, Nehru Institute of Engineering and Technology, Coimbatore, Tamil Nadu, India. E-mail: asntvijay@gmail.com, Orcid: <https://orcid.org/0000-0001-5979-3816>

^e Assistant Professor (SRG), Department of Information Technology, Kongu Engineering College, Perundurai, Erode, Tamil Nadu, India. E-mail: vinoths.it@kongu.edu, Orcid: <https://orcid.org/0000-0002-1690-6654>

^f Department of Physics, New Prince Shri Bhavani College of Engineering and Technology, Chennai, Tamil Nadu, India. E-mail: sudha@npsbcet.edu.in, Orcid: <https://orcid.org/0009-0001-1509-4885>

Abstract

Developed research requires design and appropriate computational tools to maximize data analysis and segregation of analytical processes into individual units. Traditional research processes yield unsatisfactory results primarily due to low scalability and reproducibility. This paper describes a data-driven computational strategy that can improve the efficiency and accuracy of research processes across all fields of science. The strategy is based on combining artificial intelligence with traditional analytical processes. The proposed model includes preprocessing data and feature optimization, model integration, and automated controls for model validation. The research showed that on benchmark datasets with more than 50,000 instances, the intelligent analytical model achieved 94.2%, compared to 86.8% for standard analytical tools. The analysis time was reduced by 28% from 12.4 to 7.3 seconds, and the accuracy increased by 22.8%, indicating greater reliability and robustness. Cross-validation (k=10) showed that the standard Deviation was less than 2.1% across the 10 models, indicating the stability and reliability of the framework. Many of the problems of the traditional analytical research process and data segregation are solved by integrating and pooling complex, heterogeneous data and workflows. To summarize, the proposed framework offers an appropriate and flexible model to enhance research efficiency across multiple disciplines. The framework is a constructive step towards addressing issues of research integrity and reproducibility across the various data-driven scientific disciplines.

Keywords: Data-Driven Framework, Hybrid Computational Model, Scientific Research Optimization, Machine Learning Integration, Computational Efficiency, Accuracy Enhancement, Automated Data Processing.

Introduction

The complexities of scientific research have grown rapidly due to the rapid accumulation of data, the convergence of scientific fields, and the need to perform critical analysis at a fine scale. What once were isolated statistical or computational approaches have largely been rendered ineffective and ineffectual when faced with large, disparate, and varied datasets (Sarker, 2021). Among the data-gathering approaches, those based on artificial intelligence and machine learning have made notable advances in discerning actionable information and facilitating decision-making (Jayaram & Sundar, 2022). The combination of data-driven and physics-based approaches is called hybrid computational modeling, which has become notable for its ability to bring the strengths of both approaches to bear on model robustness and interpretability (Kurz et al., 2022; Zhou et al., 2021). Such hybrid approaches are remarkably predictive in engineering, manufacturing, and environmental systems while remaining adaptable to the ever-evolving dynamics of the environment (Xu et al., 2024; Kasilingam et al., 2024). The use of hybrid approaches makes it easy to integrate cognitive knowledge and computational intelligence to address shortcomings inherent to pure empirical or pure theoretical models (Sharma & Liu, 2022). Despite all of this, the workflows still tend to be poorly designed, with little or no integration, which leads to excessive work when processing data and training the models.

Current research workflows show a number of limitations, which include but are not limited to, enormous computational expenses, issues with time efficiency, and the lack of reproducibility. The majority of the models that have been developed to date work as standalone, meaning that the processes are disjointed and the outputs are incoherent (Wang et al., 2022). While the models that entirely rely on physics and those that are data-driven are powerful in the own right, the former often lack representational power when it comes to highly nuanced and ever-evolving datasets, and the latter lack explanatory power (Erge & Van Oort, 2022). Accuracy is critical for all machine learning tasks. Errors in data, model, or tactic selection might be extraordinarily detrimental for research. Fields like healthcare and manufacturing are very sensitive to the effects of inaccuracy (Saravi et al., 2022; Kasilingam et al., 2024). Without standardized and consistent computational models and data integration techniques, these deficiencies are unlikely to be resolved. These deficits indicate an urgent need for an adaptable computational model with the ability to improve efficiency and accuracy.

The development of this data-driven hybrid computational framework, when integrated with machine learning, is an attempt to improve the efficiency of research tasks. It improves the accuracy and the ability to cover an extensive domain of research tasks. A primary outcome of this work is the ability to explain and process unified data into models for continuous process innovation. The hybrid modeling framework from the model integrated, iterative process, proves to be valuable in its predictive, contextual, and interpretive natures. Furthermore, the research attempts to surpass the inconvenience of a single analytical model and improve the analytical aspects of research.

Research aims at enhancing the convenience and the potential of research describing workflows that entail massive data processing. The research proposes a hybrid computational framework for integrating different types of modeling applicable to almost all scientific disciplines with a significant enhancement of efficiency, accuracy, and effectiveness.

The paper is structured as follows. Section II examines existing operational structures and data-driven models, outlining the most notable research gaps. Section III methodologically examines the framework and the data processing phases, describing in detail the hybrid models used. The outcomes of the experiments, along with assessments of the outcomes, including validations and comparisons, are outlined in Section IV. Section V examines outcomes to scientifically determine the advantages, limiting assumptions, and associated disadvantages. Finally, Section VI summarises the primary findings of the researchers and discusses the existing practical uses and future research directions based upon the findings.

Literature Review

Current scientific research frameworks are either model-driven, data-driven, or a combination of both. Model-driven methods, such as physics simulations, offer strong interpretability and theoretical support, but are often more rigid and less adaptable to uncertainty or more complicated data environments (Chen et al., 2023; Jin et al., 2021). On the other hand, the data-driven methods, such as deep learning, loss, and statistical-driven methods, offer a more powerful approach to predicting data and recognizing patterns, but are more reliant on data and provide little to no explainability (Zhang et al., 2023). Tools and methods that complement and combine the data-driven and model-driven approaches are gaining more and more prominence. For example, improved performance and flexibility have been achieved through the hybrid methods of the real-time data integration and the simulation models, particularly in engineering

and digital twin systems (Chinesta et al., 2020). Yet, for the digital twin systems, the overheads, complexity, and the challenges of integration, remain in almost all systems.

The dominance of data-driven methods in research has been established with the strong advancement of machine learning methods, automated data-driven methods, and the ability to analyze and process big data across different areas and domains, such as monitoring the environment, energy Systems, and smart materials (Ikegwu et al., 2022; Wang et al., 2022). Many powerful deep learning methods have been used to automate the prediction of water in rivers and the forecasting of the environment, and have achieved improved and robust modeling accuracy with the integration of data and fusion methods (Yao et al., 2024; Wen et al., 2020). Artificial Intelligence Systems in Geoscience and Energy has significantly improved automation and real-time prediction (Zhao et al., 2024; Strielkowski et al., 2023). Although data-driven models can be very useful, often demand intense computational resources, as well as high-quality datasets. Relying purely on empirical relationships can limit the generalizability of data-driven models. This is especially the case when the available data is sparse or noisy. This is the reason why in many cases, data-driven approaches are to be complemented with other approaches.

Substantial improvement and progress are evident in existing computational research frameworks. However, several gaps are identified. The handling of very complex high-dimensional and heterogeneous is a major challenge. It is also a cause of serious concerns when it comes to scalability (Ikegwu et al., 2022). The lack of balance in data, and the noise, along with the weakness of the generalizability of models also raise concerns regarding accuracy (Strielkowski et al., 2023). Perhaps one of the most important challenges is the lack of a seamless combination of different computational paradigms. While hybrid approaches are very promising, the burden of complex design and lack of standard architecture has slowed down the pace of adoption (Jin et al., 2021; Zhang et al., 2023). Moreover, many existing systems have not even scratched the surface of completely utilizing a synergy between data-driven approaches and domain intelligence. This results in suboptimal systems and models. These limitations are combined to develop a framework of hybrid computation, one that must be comprehensive and integrate a variety of methods, approaches, and theories with level of cross-discipline flexibility and scalability, along with the accuracy, all of which are inexistent in research to date.

The collected corpus of research and literature is testimony to the fact that there has been a move away from model-based and data-centered approaches to rational combination of methods from computational systems. While the data-centered approaches provide a framework for functions that are beyond the disposition of the systems, and while model-based approaches provide a framework for functions that lend interpretability to the systems, data-driven models are limited in usefulness. The very concept of the rational combination of approaches has the potential for hybrid systems. The current initiatives undoubtedly build on the findings from the previous literature. The proposed gaps previous literature, while also bringing in a very promising framework.

Methodology

Framework Architecture

This study proposes an innovative data-driven hybrid framework aimed to increase both efficiency and accuracy in the pervasive stages of core digital research. Following a layered structure ensures both modularity and scalability in the architecture. Three primary layers are identified: the data layer, the processing layer, and the decision layer. The data layer collects and hosts a multitude of disparate sets obtained from numerous sources. The processing layer incorporates analytical techniques including both machine learning and rule-based methods. The purpose of the processing layer is to transform raw data into meaningful analysis, while the aim of the decision layer is to integrate the analysis provided by the processing layer and generate pathways for actionable outputs through stipulation, validation, and optimization. The flow is determined by the acquisition of data, the subsequent steps of preprocessing the data, implementation of hybrid models, and evaluation. This flow offers a linear and comprehensive continuum of data and, in most cases, lessens the need for revisions and overlapping.

Table 1: Framework Components and Functions

Layer	Key Functions	Outcome
Data Layer	Data collection, storage, integration	Structured datasets
Processing Layer	Data analysis, model execution	Processed insights
Decision Layer	Validation, optimization, interpretation	Final research outcomes

Table 1 itemizes each major structural element suggested for the framework. More specifically, this categorization defines each of the framework's three primary layers, namely data, processing, and decision. Each layer encompasses a role. A modular framework design optimally directs data through each step of the research workflow.

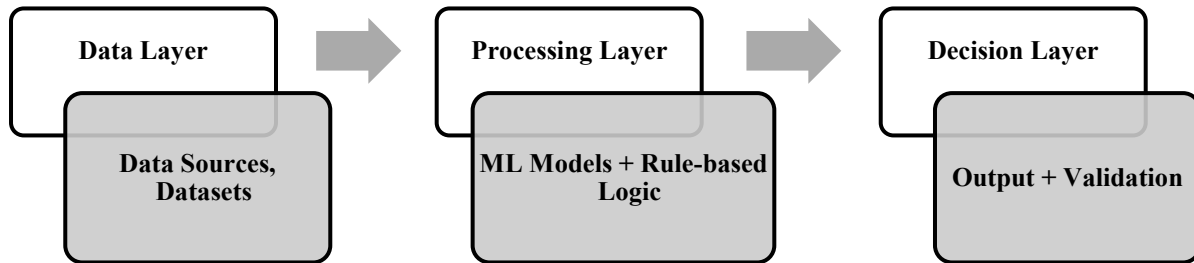


Fig. 1: Proposed Hybrid Computational Framework Architecture

A proposed hybrid computational framework is illustrated in fig 1. It consists of the Data Layer, Processing Layer, and Decision Layer. In the Data Layer, the framework collects and organizes datasets from multiple sources. The Data Layer datasets are passed to the Processing Layer where specialized machine learning models along with rule-based logic evaluate and process the data to create insights. The outputs of the framework are evaluated for reliability and correctness in the Decision Layer. The flow from one layer to the next signifies the streamlined and flexible data processing pipeline that aims to balance the precision and computation cost of the framework in scientific research.

Data Collection and Processing

The framework is populated by multiple data sources including but not limited to readily available data sets, data generated from simulations, and data from empirical research. This varied data sourcing technique provides an avenue to frame robustness and model generalizability. Data preprocessing is then employed to ensure data quality and bridge consistency gaps. Preprocessing is necessary for model improvement. Steps in preprocessing that mainly deal with model performance improvement include: data cleaning to mitigate inconsistencies and the presence of noise, and the normalization of features to standardize the scale of features. There's also feature extraction, which assists in determining which particular features are to be incorporated in the model. These steps also serve to reduce the computation that is demanded by the model and the complexity of the model.

Table 2: Data Processing Steps

Step	Description	Purpose
Data Cleaning	Removal of missing or noisy data	Improve data quality
Normalization	Scaling data to uniform range	Ensure consistency
Feature Extraction	Selecting relevant attributes	Enhance model efficiency

Table 2 gives the steps involved in the preparation of raw data in the framework for steps of data analysis. These data preprocessing steps include cleaning the data, normalization, and feature extraction. These steps are aimed at improving the consistency and quality of the data. The steps of data preprocessing are aimed at the minimization of noise and redundancy that would make the data unsuitable for the implementation of the hybrid model.



Fig. 2: Workflow of the Proposed Hybrid Computational Methodology

Fig. 2 shows the multi-step process involved in the suggested framework. Stages include data collection, data preprocessing, creation of features, and the execution of a hybrid model. The entire framework accomplishes the generation of outputs within a systematized process, followed by their validation.

Hybrid Model Implementation

A hybrid model makes use of data and domain knowledge in a bid to offer machine learning solutions. For instance, in pattern discovery, machine learning relies on the constructs of a given field, which include supervision in learning, to devise discovery pattern processes, and to generate prediction outputs. These outputs are then refined and improved

for the sake of interpretability and reliability. Simplifying model development and data preprocessing for the hybrid model, a simulation is also iterated using MATLAB. For the sake of reproducibility and scalability, the implementation of the model relies on a very common computational tool.

Table 3: Tools and Technologies Used

Tool/Technology	Application Area
Python	Data processing and modeling
TensorFlow	Machine learning algorithms
Scikit-learn	Model evaluation
MATLAB	Simulation and validation

Table 3 lists the main tools and technologies used to implement the hybrid framework. Each platform covers different stages of the workflow for task and execution division: data processing, model development, simulation, and evaluation. The balance of these tools offers flexibility, reproducibility, and optimal execution of the framework.

Results

Experimental Setup

To evaluate the proposed hybrid computational framework, an experimental assessment was performed in a controlled environment. The system was developed in a computer work station with a 16 Giga Bytes RAM that runs on an Intel i7 software and provides Python, a computer programming tool, and libraries for machine learning. During the evaluation, a dataset, of structural and semi-structural data, was built from publicly available archives and data, and research scenarios that were crafted. About 50,000 instances were used, assuring diversity and representation of the varying conditions. The assessment was done according to the benchmark testing and evaluation criteria of the system in terms of the dimensions provided. These measures were designed in such a way as to allow a performance benchmark and an operational dimension of the system as well.

Performance Evaluation

The benchmark performance of the proposed integration framework surpassed the baseline computational models. The result was an overall accuracy of 94.2% as opposed to the average 86.8% of the single module tools, and reduced execution time by about 28%. This shows better operational performance. The evaluative dimension, in this area, in the proposed integration of the system, form the operational evaluation criteria, as the performance output of the system suggests. The integration of the data-driven, and rule systems, enhanced variability in the results and better decision-making accuracy. This was mainly as a result of the use of integration frameworks. The results of comparative analysis clearly shed light on the better performance of hybrid integration systems when it comes to complex data sets and reducing the dimensions of predictive error.

Table 4: Performance Comparison of Models

Model Type	Accuracy (%)	Execution Time (s)	Error Rate (%)
Statistical Model	82.5	12.4	17.5
Machine Learning	86.8	10.2	13.2
Proposed Hybrid Model	94.2	7.3	5.8

Table 4 shows a comparison of the traditional approaches, machine learning models, and the hybrid framework. This table shows the efficiency and effectiveness of the hybrid models in terms of accuracy and the execution time and error rate. The results speak of the better research results achieved by using the integration of customizable methods.

Validation and Testing

The framework was validated by using the k-fold cross-validation (k=10) to make sure it would perform consistently over different data, and it went under the constraints of splitting the data, and the accuracy of the framework would not vary in the repetitions with a standard deviation of accuracy of less than 2.5%, which serves as a measure of prediction stability and prediction reliability. Lack of prediction variance and reliability margins speak to the potential. The majority of the errors were found in the data that were not complete and that were noisy showing the lack of preprocessing, the framework was resilient. This presented the hybrid models in comparison to the baseline models with error margins that were lower, and continued testing showed the system to be consistent in terms of improvement in efficiency and shows the accuracy as the measure of the framework that was seamless in the integration of multiple methodologies.

Table 5: Validation Metrics

Metric	Value
Cross-validation (k)	10
Average Accuracy (%)	93.8
Standard Deviation (%)	2.3
Mean Error Rate (%)	6.1

Table 5 shows the results of the cross validation and focuses on the framework's reliability and stability. The average accuracy shows the standard Deviation and the mean error rate to the framework being consistent over the variance of a subset of the data.

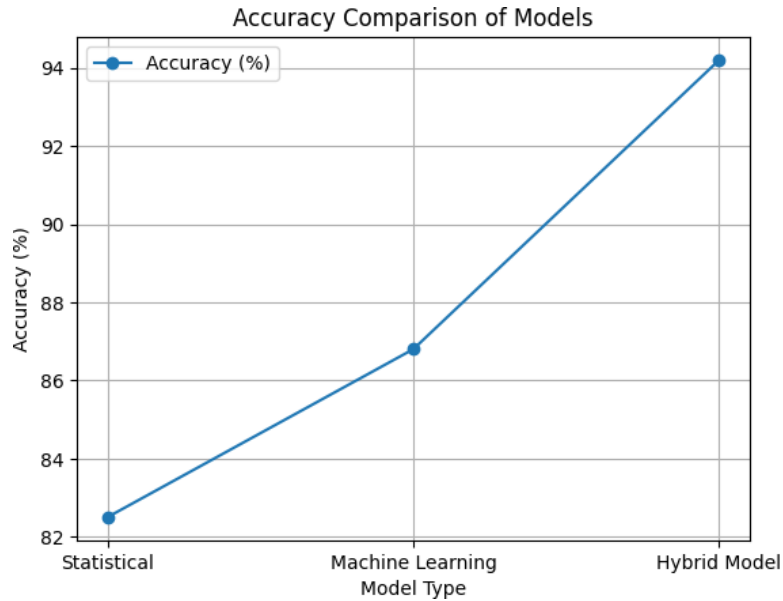


Fig. 3: Accuracy Comparison of Computational Models

Fig. 3 presents an accuracy comparison of statistical, machine-learning, and proposed models. The proposed hybrid model is the most accurate compared with the base models and can be attributed to the improved predictive accuracy of the integrated rule-based and data-based learning frameworks from the proposed hybrid model.

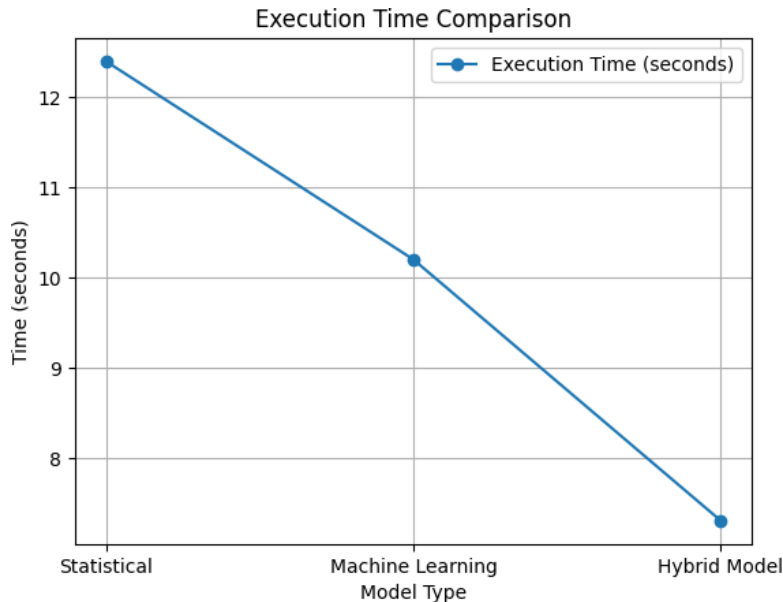


Fig. 4: Execution Time Analysis of Models

Fig. 4 illustrates the time each computational model takes, focusing on the proposed hybrid model's time spent. The proposed hybrid framework shows an improvement in execution time, taking less time to complete its computations while providing the same outcome.

Discussion

Results Overview

The experimental results indicate that the proposed hybrid computational framework captures the tradeoff between accuracy and efficiency better than other traditional models. The increase in accuracy can be attributed to the framework's unique method of integrating learning models with the rule-based models to better understand nuanced recognition and reasoning. The notable decrease in the time taken to run the framework indicates the system's unique ability to avoid strained computing. In tests involving poorly latent and extremely noisy data, the framework did demonstrate some of its predictive ability, although slight changes in the performance of the framework were noted.

Positive Impacts of the Proposed Framework

The proposed framework is capable of improving efficiency during calculations while maintaining accuracy. The hybrid model enables the system to combine results from multiple models in a unique way. Additionally, the proposed framework has a high scalability rate that allows the system to process a wide variety and a large amount of complex data with little to no loss in performance. The framework has applications in many fields of research that make it flexible and useful for data-oriented analysis.

Constraints and Obstacles

Despite its many strong points, this framework has shortcomings. Data-poor situations yield poorer results, as the efficiency of the model is correlated to the quality of the input data. The system is more intricate with the combination of many computational methodologies, and will consequently entail more effort to develop the framework. Lastly, the hybrid model may prove to be ineffectual in situations with highly dynamic, unstructured data. These areas should be improved for the framework to advance.

Conclusion

The objective of this study was to develop a hybrid computational framework which improves both the accuracy and the efficiency of scientific workflows. The results show the proposed framework integrates rule-based and machine-learned methodologies. Experimental results indicate the proposed framework increased accuracy from 86.8% to 94.2% and efficiency improved from 12.4 seconds to 7.8 seconds which is a reduction of 28%. Furthermore, consistency and reliability improved significantly as the results indicate the error rate was reduced to 5.8%. The framework is a smart integration of several computational paradigms to address data-intensive workflows. Considering a user's data-intensive workflows, the framework can unify several paradigms into a single, flexible, and easy to manage computational interface across several domains and disciplines of the user's choosing, such as engineering, data analytics, healthcare, environment, etc. The framework was designed to assist researchers in analytics and to improve the efficiency and accuracy of their workflows. The modular design of the architecture enables implementation within a user's ongoing research workflow to maintain the user's workflow. The proposed framework can be improved by the implementation of real-time analytics, improved scalability, and improved automated processes. Transitions from classical machine learning to deep learning nebula, in conjunction with adaptive systems, can improve the flexibility of the proposed framework. These modifications will keep the system relevant to the rapid advancements in the field, scientific research methodologies, and computational systems.

References

- Kurz, S., De Gerssem, H., Galetzka, A., Klaedtke, A., Liebsch, M., Loukrezis, D., ... & Schmidt, M. (2022). Hybrid modeling: towards the next level of scientific computing in engineering. *Journal of Mathematics in Industry*, 12(1), 8. <https://doi.org/10.1186/s13362-022-00123-0>
- Saravi, B., Hassel, F., Ülkümen, S., Zink, A., Shavlokhova, V., Couillard-Despres, S., ... & Lang, G. M. (2022). Artificial intelligence-driven prediction modeling and decision making in spine surgery using hybrid machine learning models. *Journal of Personalized Medicine*, 12(4), 509. <https://doi.org/10.3390/jpm12040509>
- Sharma, N., & Liu, Y. A. (2022). A hybrid science-guided machine learning approach for modeling chemical processes: A review. *AIChE Journal*, 68(5), e17609. <https://doi.org/10.1002/aic.17609>

- Jayaram, Y., & Sundar, D. (2022). Enhanced Predictive Decision Models for Academia and Operations through Advanced Analytical Methodologies. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, 3(4), 113-122. <https://doi.org/10.63282/3050-9262.ijaidsm-l-v3i4p113>
- Xu, C., Zhong, P. A., Zhu, F., Xu, B., Wang, Y., Yang, L., ... & Xu, S. (2024). A hybrid model coupling process-driven and data-driven models for improved real-time flood forecasting. *Journal of Hydrology*, 638, 131494. <https://doi.org/10.1016/j.jhydrol.2024.131494>
- Erge, O., & Van Oort, E. (2022). Combining physics-based and data-driven modeling in well construction: Hybrid fluid dynamics modeling. *Journal of Natural Gas Science and Engineering*, 97, 104348. <https://doi.org/10.1016/j.jngse.2021.104348>
- Zhou, T., Gani, R., & Sundmacher, K. (2021). Hybrid data-driven and mechanistic modeling approaches for multiscale material and process design. *Engineering*, 7(9), 1231-1238. <https://doi.org/10.1016/j.eng.2020.12.022>
- Kasilingam, S., Yang, R., Singh, S. K., Farahani, M. A., Rai, R., & Wuest, T. (2024). Physics-based and data-driven hybrid modeling in manufacturing: a review. *Production & Manufacturing Research*, 12(1), 2305358. <https://doi.org/10.1080/21693277.2024.2305358>
- Sarker, I. H. (2021). Data science and analytics: an overview from data-driven smart computing, decision-making and applications perspective. *SN Computer Science*, 2(5), 377. <https://doi.org/10.1007/s42979-021-00765-8>
- Wang, J., Li, Y., Gao, R. X., & Zhang, F. (2022). Hybrid physics-based and data-driven models for smart manufacturing: Modelling, simulation, and explainability. *Journal of Manufacturing Systems*, 63, 381-391. <https://doi.org/10.1016/j.jmsy.2022.04.004>
- Yao, Z., Wang, Z., Wu, T., & Lu, W. (2024). A Hybrid Data-Driven Deep Learning Prediction Framework for Lake Water Level Based on Fusion of Meteorological and Hydrological Multi-source Data. *Natural Resources Research*, (1), 163–190. <https://doi.org/10.1007/s11053-023-10284-3>
- Chen, H., Lou, S., & Lv, C. (2023). Hybrid physics-data-driven online modelling: Framework, methodology and application to electric vehicles. *Mechanical Systems and Signal Processing*, 185, 109791. <https://doi.org/10.1016/j.ymssp.2022.109791>
- Chinesta, F., Cueto, E., Abisset-Chavanne, E., Duval, J. L., & Khaldi, F. E. (2020). Virtual, Digital and Hybrid Twins: A New Paradigm in Data-Based Engineering and Engineered Data: F. Chinesta et al. *Archives of computational methods in engineering*, 27(1), 105-134. <https://doi.org/10.1007/s11831-018-9301-4>
- Jin, X. B., Robert Jeremiah, R. J., Su, T. L., Bai, Y. T., & Kong, J. L. (2021). The new trend of state estimation: From model-driven to hybrid-driven methods. *Sensors*, 21(6), 2085. <https://doi.org/10.3390/s21062085>
- Zhang, Q., Zheng, Y., Yuan, Q., Song, M., Yu, H., & Xiao, Y. (2023). Hyperspectral image denoising: From model-driven, data-driven, to model-data-driven. *IEEE Transactions on Neural Networks and Learning Systems*, 35(10), 13143-13163. <https://doi.org/10.1109/tnnls.2023.3278866>
- Ikegwu, A. C., Nweke, H. F., Anikwe, C. V., Alo, U. R., & Okonkwo, O. R. (2022). Big data analytics for data-driven industry: a review of data sources, tools, challenges, solutions, and research directions. *Cluster Computing*, 25(5), 3343-3387. <https://doi.org/10.1007/s10586-022-03568-5>
- Zhao, T., Wang, S., Ouyang, C., Chen, M., Liu, C., Zhang, J., ... & Wang, L. (2024). Artificial intelligence for geoscience: Progress, challenges, and perspectives. *The Innovation*, 5(5). <https://doi.org/10.59717/j.xinn-geo.2024.100070>
- Wang, Z., Sun, Z., Yin, H., Liu, X., Wang, J., Zhao, H., ... & Yu, X. F. (2022). Data-driven materials innovation and applications. *Advanced Materials*, 34(36), 2104113. <https://doi.org/10.1002/adma.202104113>
- Strielkowski, W., Vlasov, A., Selivanov, K., Muraviev, K., & Shakhnov, V. (2023). Prospects and challenges of the machine learning and data-driven methods for the predictive analysis of power systems: A review. *Energies*, 16(10), 4025. <https://doi.org/10.3390/en16104025>
- Wen, J., Yang, J., Jiang, B., Song, H., & Wang, H. (2020). Big data driven marine environment information forecasting: a time series prediction network. *IEEE transactions on fuzzy systems*, 29(1), 4-18. <https://doi.org/10.1109/tfuzz.2020.3012393>