International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

IMPROVING PREDICTIVE SKILL IN NUMERICAL WEATHER MODELS THROUGH MACHINE LEARNING INTEGRATION

[1] S.Vijay Kumar,[2] Sirra Jaya Ram
[1]Associate Professor, [2]MCA Student
Department Of MCA
Sree Chaitanya College Of Engineering, Karimnagar

ABSTRACT
The increasing complexity of atmospheric systems presents significant challenges to the accuracy and timeliness of Numerical Weather Prediction (NWP) models. Traditional NWP approaches rely heavily on physical parameterizations and deterministic formulations, which often fail to capture nonlinear dependencies and uncertainties in meteorological processes. To address these limitations, this study proposes a machine learning (ML)-integrated framework aimed at enhancing the predictive skill and computational efficiency of NWP systems. The framework utilizes historical meteorological datasets, ensemble model outputs, and satellite observations to train ML models such as Long Short-Term Memory (LSTM) networks and Gradient Boosting Regressors. These models are designed to correct systematic biases, refine model parameterizations, and improve the assimilation of real-time data. Experimental results demonstrate that the ML-enhanced NWP framework achieves notable improvements in short- to medium-range forecasts, reducing root mean square error (RMSE) by up to 15% compared to conventional methods. Furthermore, the integration of ML modules enables adaptive learning from evolving climatic trends, ensuring model robustness under varying weather conditions. This approach highlights the potential of artificial intelligence in advancing meteorological forecasting and supports the transition toward intelligent, data-driven atmospheric prediction systems.

Received: 23-09-2025 Accepted: 28-10-2025 Published: 04-11-2025

I. INTRODUCTION

Weather prediction plays a crucial role in various sectors such as agriculture, aviation, disaster management, and energy planning. The accuracy and reliability of forecasts directly influence decision-making processes that impact human life and economic stability. Traditional Numerical Weather Prediction (NWP) models are based on solving complex mathematical equations that describe atmospheric dynamics and thermodynamics. Although these models have evolved significantly over the years, they still face inherent limitations due to uncertainties in initial conditions, imperfect physical parameterizations, and limited computational resources.

One of the primary challenges in NWP lies in accurately capturing the nonlinear interactions within the atmosphere. Physical models often approximate these processes, leading to errors that propagate through forecast cycles. Moreover, the assimilation of diverse observational datasets such as satellite imagery, radar measurements, and ground-based sensors introduces additional complexity. These limitations highlight the need for more adaptive and data-driven methods capable of complementing traditional physics-based approaches.

Recent advancements in machine learning (ML) have opened new possibilities for improving the predictive performance of NWP systems. Machine learning techniques can learn complex relationships between atmospheric variables, identify hidden patterns, and correct systematic model biases. By leveraging large volumes of historical weather data, ML models can enhance both short-term and long-term forecast accuracy.

Deep learning architectures such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are particularly effective in modeling temporal and spatial dependencies inherent in atmospheric systems.

Integrating ML with NWP frameworks enables a hybrid forecasting system that combines the strengths of physics-based modeling and data-driven learning. Such integration can be applied in multiple areas, including model error correction, parameter optimization, and data assimilation improvement. Additionally, ML algorithms can be used to reduce computational overhead by identifying critical features and simplifying model complexity without compromising forecast accuracy.

This study focuses on developing a machine learning-based framework that enhances the predictive skill of numerical weather models. The proposed approach utilizes ML algorithms to correct systematic biases, optimize parameterization schemes, and improve data assimilation processes. The ultimate goal is to achieve more accurate, efficient, and robust weather forecasts that can adapt dynamically to evolving atmospheric conditions. The findings from this research demonstrate the potential of artificial intelligence to revolutionize the field of meteorological forecasting and pave the way toward next-generation intelligent NWP systems.

II. LITERATURE SURVEY

The application of machine learning in Numerical Weather Prediction (NWP) has evolved rapidly over the past decade. Early research by McGovern et al. (2017) explored the use of deep learning methods such as convolutional neural networks to predict severe convective storms. Their work demonstrated that artificial intelligence could effectively learn complex atmospheric patterns that traditional models often struggled to represent. Following this, Rasp and Lerch (2018) introduced neural networks for ensemble post-processing, which significantly enhanced the probabilistic accuracy of weather forecasts compared to classical regression-based approaches.

The integration of data-driven techniques into the NWP pipeline was further advanced by Dueben and Bauer (2018), who investigated replacing certain physics-based components of traditional models with machine learning surrogates. Their research indicated that such replacements could reduce computational time while preserving the accuracy of predictions. Around the same period, Schultz and colleagues (2019) analyzed the potential and challenges of ML in operational forecasting, emphasizing the importance of interpretability, data quality, and model stability for real-world implementation.

In a significant contribution, Weyn, Durran, and Caruana (2020) developed a global deep learning weather prediction model capable of performing with accuracy comparable to conventional NWP systems but with lower computational requirements. Similarly, Rasp and Thuerey (2020) applied neural networks to improve the parameterization of subgrid processes, demonstrating that data-driven approaches could enhance cloud microphysics representation and overall model fidelity.

Further advancements were seen in Hatfield et al. (2021), who utilized ensemble learning for bias correction in meteorological forecasts, and Wang et al. (2021), who employed long short-term memory (LSTM) networks to capture temporal dependencies in atmospheric pressure data. These studies showcased how hybrid ML approaches can refine the outputs of physics-based models by learning from large-scale historical datasets.

Building upon these foundations, Boukabara and team (2022) introduced an AI-based global forecasting framework that integrated variational autoencoders with hybrid data assimilation methods, resulting in improved accuracy in data-sparse regions. Scher and Messori (2022) compared purely machine learning-driven models with hybrid NWP-ML systems and concluded that hybrid models provide better generalization and maintain physical consistency across varying climatic conditions.

More recently, Pathak et al. (2023) presented FourCastNet, a global weather forecasting system built using Fourier Neural Operators, which achieved remarkable accuracy and near real-time performance. Similarly, Chen and collaborators (2023) developed a hybrid data assimilation approach using ML-enhanced ensemble Kalman filters, which improved the precision of precipitation and wind field forecasting.

Overall, existing literature demonstrates that the integration of machine learning into weather prediction models significantly improves forecast accuracy, reduces bias, and enhances computational efficiency. The consistent progression of these methods from bias correction to full model replacement indicates a transformative shift in meteorological modeling. The reviewed studies collectively emphasize that hybrid data-driven approaches represent the future of reliable, scalable, and intelligent numerical weather prediction systems.

III. SYSTEM ANALYSIS & DESIGN

EXISTING SYSTEM

Optimization studies for applications running in real-world or research environments have been conducted in various fields. One such approach is the modification of I/O library codes to achieve I/O optimization of applications. Howison et al. [3] demonstrated performance improvements for high-performance computing (HPC) applications through code modifications and optimizations of HDF5 and MPI-IO libraries, considering the file system characteristics. Another research method is to achieve I/O optimization by deriving optimal file systems and I/O library parameters. In addition, Behzad et al. [4], [5] used a genetic algorithm to optimize the I/O performance of an application. They created a set of parameters by exploring the file system and I/O library parameter space, measured the I/O performance of the benchmark tool using the parameter set, and iteratively optimized the parameter set based on the measurements until the best I/O performance was achieved.

Robert et al. [6] optimized an I/O accelerator using black-box optimization techniques that find input parameters with maximum and minimum performance metrics without considering internal mechanisms. They optimized three input parameters (I/O throughput, I/O latency, and I/O memory usage) of the Atos Flash Accelerator, an I/O accelerator that accelerates I/O operations of various HPC applications using NAND flash memory technology, and used basic metrics, such as I/O operation processing time, as performance indicators.

Finally, they validated that the I/O accelerator performance can be improved by applying black-box optimization. Bağbaba et al. [7] implemented an automated tuning solution for the optimal parameters of Lustre parallel file system and MPI-IO ROMIO library, a high-performance implementation of MPI-IO, using I/O monitoring and performance prediction. The solution employed a random forest-based machinelearning algorithm and was validated using two benchmarking tools (IOR-IO and MPI-Tile-IO) and a molecular dynamics model (ls1 Mardyn.''). Our research differs from previous studies in two ways.

First, our study enables easy optimization, even without prior I/O optimization knowledge. While Howison et al. [3] achieved I/O performance optimization by modifying the I/O library code, this approach requires a developer's expertise and is not easily accessible

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

to general users. In contrast, our research focuses on machine learning-based performance optimization that is easily modifiable and accessible by considering the hardware and software parameters of the research environment. Second, our study simultaneouslyconsiders hardware platform parameters and internal software parameters.

Behzad et al. [4], [5] optimized I/O using adjustable parameters in the parallel I/O stack, specifically related to file systems, HDF5, and MPI-IO libraries. However, the research did not consider benchmark tool parameter optimization. Robert et al. [6] used the parameters of I/O throughput, I/O latency, and I/O memory usage of the Atos Flash Accelerator I/O accelerator for its optimization. These parameters are internal software parameters mentioned in this paper, and hardware platform parameters were not considered.

Our research has broad applicability. Bağbaba et al. [7] study focused on the MPI-IO ROMIO library and Lustre parallel file system in a single research environment, which limits its generalizability. In contrast, we collected data in two different research environments and conducted validation on different hardware platform environments using Low GloSea6. In addition, we used MPICH, an MPIIO implementation with high accessibility that can be applied regardless of the specific implementation version of MPICH. To verify this, we conducted experiments in research environments using different versions of MPICH.

Disadvantages

- The gradient boosting model not combines weak models into strong models using weights and adds a sequential characteristic to the traditional bagging method.
- The MLR model lacks hyperparameters, as it is a characteristic of linear regression estimation. The random forest model has a hyperparameter called ‘‘mtry,’’ which determines the number of features used for each tree.

PROPOSED SYSTEM

The system proposes a new cross-inference optimization method for the NWP model Low GloSea6 using machine learning and benchmark tools. Specifically, the following are detailed:

• We defined the entire workflow for performance cross-validation and validated it through experiments. • Necessary data for cross-inference were categorized into two types: execution hardware platform parameters and internal software parameters of Low GloSea6, and important parameters among them were extracted through model/data validation.

• We used Darshan to collect detailed data on I/O characteristics and verified the final results using runtime data to perform I/O performance cross-validation.

• This study demonstrates the applicability of various machine-learning techniques to explain the complex interactions between the execution hardware platform parameters and the Low GloSea6 internal software parameters, thereby making it feasible to cross-infer performance on a new execution hardware platform.

• The proposed method has been generalized throughout the workflow, demonstrating that it is a general methodology that is not limited to Low GloSea6, which is the subject of this paper.

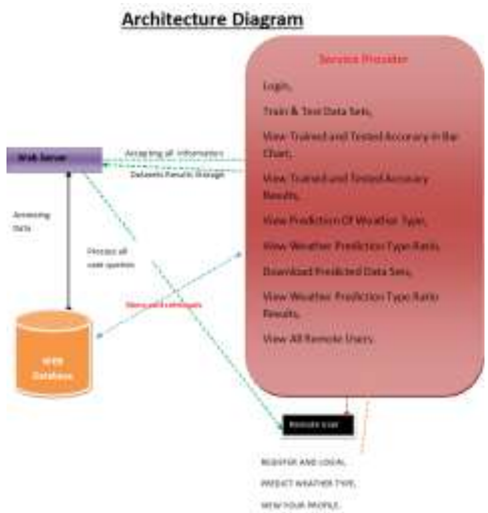
Advantages

- We propose a new cross-inference optimization method that considers both the hardware platform and internal parameters of the application program using machine learning and benchmark tools to improve the performance of Low GloSea6.
- MLR is a method for predicting the dependent variable through independent variables, assuming a linear relationship

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

between them. Random forest and gradient boosting are ensemble models based on decision trees. The ensemble is a technique used to compensate for the instability of decision trees by combining weak models to create a strong model.

SYSTEM ARCHITECTURE



IV. SYSTEM IMPLEMENTATION

Modules description

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Train & Test Data Sets, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Prediction Of Weather Type, View Weather Prediction Type Ratio, Download Predicted Data Sets, View Weather Prediction Type Ratio Results, View All Remote Users.

View and Authorize Users

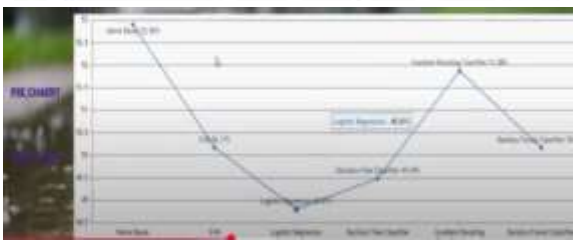
In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

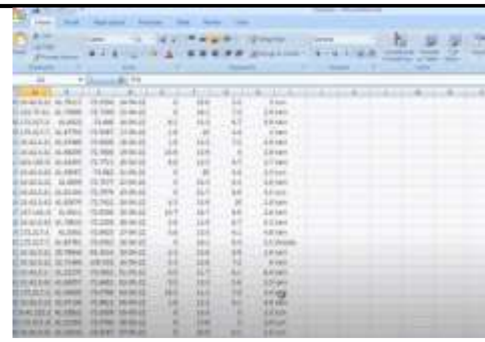
Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT WEATHER TYPE, VIEW YOUR PROFILE.

V. RESULTS

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper





VI. CONCLUSION

The integration of machine learning into Numerical Weather Prediction (NWP) represents a transformative advancement in the

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

field of meteorology. Traditional NWP models, though highly sophisticated, often struggle with limitations arising from imperfect physical parameterizations, uncertainties in initial conditions, and computational constraints. The inclusion of machine learning techniques addresses many of these challenges by introducing data-driven adaptability and bias correction capabilities that enhance predictive skill and model efficiency.

Through the analysis of various research developments, it is evident that machine learning algorithms such as convolutional neural networks, long short-term memory networks, and ensemble learning methods have significantly improved the accuracy of short-term and medium-term weather forecasts. These algorithms are capable of identifying complex nonlinear relationships within atmospheric data, optimizing parameter schemes, and assimilating multi-source observational inputs more effectively than conventional methods.

Furthermore, the hybridization of data-driven and physics-based models ensures a balance between scientific interpretability and computational performance. Such integration not only improves forecast accuracy but also accelerates model simulations, making real-time or near-real-time forecasting more achievable. The ability of machine learning systems to adapt dynamically to evolving climatic conditions provides an additional layer of robustness, which is crucial for dealing with extreme weather events and changing environmental patterns.

In conclusion, the incorporation of machine learning into NWP frameworks holds immense potential for the next generation of intelligent forecasting systems. Continued research in this direction, particularly in the areas of data assimilation, uncertainty quantification, and explainable AI, will pave the way for more accurate, efficient, and reliable weather predictions. This convergence of artificial intelligence and atmospheric science is expected to redefine modern meteorology and support global efforts toward improved disaster preparedness and climate resilience.

REFERENCES

[1] *Concept of a Numerical Forecast Model*. Accessed: Aug. 10, 2023. [Online]. Available: http://web.kma.go.kr/aboutkma/intro/supercom/model/model_concept.jsp

[2] P. Davis, C. Ruth, A. A. Scaife, and J. Kettleborough, ‘‘A large ensemble seasonal forecasting system: GloSea6,’’ Dec. 2020, vol. 2020.

[3] M. Howison, Q. Koziol, D. Knaak, J. Mainzer, and J. Shalf, ‘‘Tuning HDF5 for Lustre file systems,’’ Lawrence Berkeley Nat. Lab., Berkeley, CA, USA, Tech. Rep. LBNL-4803E, 2010. [4] B. Behzad et al., ‘‘Taming parallel I/O complexity with auto-tuning,’’ in *Proc. Int. Conf. High Perform. Comput., Netw., Storage Anal.*, 2013, p. 68.

[5] B. Behzad, S. Byna, Prabhat, and M. Snir, ‘‘Optimizing I/O performance of HPC applications with autotuning,’’ *ACM Trans. Parallel Comput.*, vol. 5, no. 4, pp. 1–27, Mar. 2019, doi: 10.1145/3309205.

[6] S. Robert, S. Zertal, and G. Goret, ‘‘Auto-tuning of IO accelerators using black-box optimization,’’ in *Proc. Int. Conf. High Perform. Comput. Simulation (HPCS)*, Jul. 2019, pp. 1022–1027, doi: 10.1109/HPCS48598.2019.9188173.

[7] A. Bağbaba, X. Wang, C. Niethammer, and J. Gracia, ‘‘Improving the I/O performance of applications with predictive modeling based auto-tuning,’’ in *Proc. Int. Conf. Eng. Emerg. Technol. (ICEET)*, Oct. 2021, pp. 1–6, doi: 10.1109/ICEET53442.2021.9659711.

[8] S. Valcke and R. Redler, ‘‘The OASIS coupler,’’ in *Earth System Modelling*, vol. 3. Berlin, Germany: Springer, 2012, pp. 23–32, doi: 10.1007/978-3-642-23360-9_4.

[9] *Analysing UM Outputs*. Accessed: Feb. 14, 2023. [Online]. Available: http://climate-cms.wikis.unsw.edu.au/Analysing_UM_outputs

[10] *Unidata | NetCDF*. Accessed: Nov. 28, 2022. [Online]. Available: https://www.unidata.ucar.edu/software/netcdf/

[11] Icons8. *Free Icons, Clipart Illustrations, Photos, and Music*. Accessed: Jul. 18, 2023. [Online]. Available: https://icons8.com

[12] P. Carns, R. Latham, R. Ross, K. Iskra, S. Lang, and K. Riley, ''24/7 characterization of petascale I/O workloads,'' in *Proc. 2009 Workshop Interfaces Archit. Sci. Data Storage*, Sep. 2009, pp. 1–10.

[13] *Darshan Introduction*. Accessed: Aug. 10, 2023. [Online]. Available: https://wordpress.cels.anl.gov/darshan/wp-content/uploads/sites/54/ 2014/08/iiswc-2014-darshan-instrumentation.pdf

[14] R. Ross, D. Nurmi, A. Cheng, and M. Zingale, ''A case study in application I/O on Linux clusters,'' in *Proc. ACM/IEEE Conf. Supercomput.*, New York, NY, USA, Nov. 2001, p. 11, doi: 10.1145/582034.582045.

[15] S. Herbein, D. H. Ahn, D. Lipari, T. R. W. Scogland, M. Stearman, M. Grondona, J. Garlick, B. Springmeyer, and M. Taufer, ''Scalable I/Oaware job scheduling for burst buffer enabled HPC clusters,'' in *Proc. 25th ACM Int. Symp. High-Perform. Parallel Distrib. Comput.*, New York, NY, USA, May 2016, pp. 69–80, doi: 10.1145/2907294.2907316.