International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 3S, 2025
https://www.theaspd.com/ijes.php

A Combined System for Predicting Weather Using Machine Learning and Optimization Techniques

Dr. S. Ravishankar[1], Dr. N. Ravi[2*]
[1]Assistant Professor, Department of Computer Applications, Periyar Arts College, Cuddalore, Tamil Nadu, India.
Email: thiru.ravishankar@gmail.com
[2]Associate Professor, PG Department of Computer Science, Government Arts College, Chidambaram – 608 102, Tamil Nadu, India.
Corresponding Author Email: csravi2017@gmail.com

Abstract: Accurate weather prediction plays a vital role in fields like farming, transportation, and emergency planning. This work presents an integrated approach that combines machine learning algorithms with optimization techniques to enhance forecasting performance. The model is built using past weather data, including features like temperature, humidity, wind speed, and rainfall. Algorithms such as Decision Tree, Support Vector Machine (SVM), and Random Forest are used for training. To improve the accuracy of these models, optimization methods like Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) are applied to fine-tune their parameters. The proposed system is tested on real-world weather datasets and delivers improved prediction results compared to standard methods. This method helps produce more reliable forecasts, supporting better planning and decision-making in weather-sensitive sectors.

Keywords: Weather Forecasting, Machine Learning, Optimization Techniques, Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Support Vector Machine (SVM), Random Forest, Decision Tree, and Predictive Modeling.

1. INTRODUCTION

Weather forecasting is an essential task that helps people and industries plan and make informed decisions. It plays a major role in areas such as agriculture, transportation, disaster management, and daily life activities. Predicting weather conditions accurately is a challenging task due to the complex and dynamic nature of the atmosphere. In recent years, the growth of data science and artificial intelligence has opened new possibilities for improving weather prediction models. Machine learning (ML) techniques have shown promising results in analyzing large volumes of weather data and identifying hidden patterns. Models such as Decision Tree, Support Vector Machine (SVM), and Random Forest are widely used for classification and prediction tasks. However, the performance of these models largely depends on the proper selection of parameters and features.

To improve the accuracy and efficiency of ML models, optimization techniques can be used. Metaheuristic methods like Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) are powerful tools that help in selecting the best model parameters and improving results. By combining machine learning with these optimization techniques, we can build a more reliable and accurate system for weather forecasting. This paper proposes a hybrid approach that integrates machine learning algorithms with optimization methods to enhance the performance of weather prediction. The system is tested on real weather datasets, and the results are compared with traditional models to show the effectiveness of the proposed method.

Weather forecasting is now heavily data-based, and machine learning (ML) is widely used to handle large amounts of weather data. Traditional methods, like numerical weather prediction, are accurate but often slow and require high computing power. To overcome this, many studies now focus on ML for faster and more accurate predictions. For example, Khozani et al. [1] used NASA POWER data with ML techniques to improve rainfall forecasting, showing the benefit of combining satellite data with ML. Several researchers have found that basic ML models have limits in performance. To improve this, optimization algorithms are used to adjust model parameters. Mumtahina et al. [2] reviewed techniques like Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) and found that these methods can significantly boost forecasting accuracy.

Navita and Srinivasan [3] used a hybrid ML approach for predicting temperature and rainfall. Their study focused on choosing the best features and tuning models with hybrid optimization, which led to better results. Lam et al. [4] introduced GraphCast, a deep learning model for medium-range global forecasts, and showed strong accuracy using real-world data. Samadianfard et al. [5] improved wind speed prediction by combining a multilayer

perceptron (MLP) with the Whale Optimization Algorithm (WOA). Similarly, Adhikari and Agrawal [6] found that neural networks optimized with PSO outperformed traditional statistical forecasting methods.
Zhang et al. [7] improved a Bayesian ConvLSTM model using PSO for hydrological forecasting, proving that deep learning with optimization works well in complex environments. Another model that used CEEMDAN for signal decomposition and was optimized using PSO, GA, and ACO also showed great results in predicting the UV index [8]. Hybrid forecasting models are becoming more popular. Jallal et al. [9] used artificial neural networks with delayed input data to improve short-term temperature prediction. Another study [10] combined statistical and ML methods for estimating evapotranspiration and found this mix gave better results.

Many optimization methods like the Firefly Algorithm [11], Memetic Algorithms [12], and Grey Wolf Optimizer [14] have been used to fine-tune ML models. A review in [13] organized these techniques and explained how they help improve intelligent systems. Hybrid models have also helped in predicting droughts and wind energy. In [17], combining signal decomposition with ANN and SVM improved drought forecasts. A hybrid LSTM-ERSA model gave better wind energy predictions than regular deep learning models [18]. Other hybrid models like NeuralGCM and GraphCast [4], [19] show that blending physics with ML gives better accuracy than traditional methods. Chen et al. [20] proposed a hybrid Conv1D–MLP model for rainfall forecasting using climate data. Another ConvLSTM model optimized with PSO showed good results on hydrological data [21]. A comparison in [22] between PSO, GA, and Firefly Algorithm for tuning ANN and GMDH models showed that optimization greatly helps improve rainfall predictions. A study in [23] used a Bi-LSTM–GRU model to enhance monthly rainfall prediction, performing better than standard LSTM models.

New hybrid deep learning models like sequence-to-sequence LSTM combined with swarm intelligence also show strong results in wind and solar energy forecasting [25]. Finally, multiple reviews [2], [16] confirm that using optimization in hybrid models leads to better performance than both traditional and untuned ML models. This highlights the need for smart model design in weather prediction. Credit card fraud is growing rapidly as fraudsters constantly adopt new technologies to perform illegal transactions. During the COVID-19 pandemic, the rise in online shopping also led to a significant increase in credit card fraud. One study applied various data mining and statistical methods to train models that could accurately detect fraud. The effectiveness of these models was validated using numerical analysis [26]. Due to the large financial losses caused by fraudulent credit card transactions, researchers have emphasized the need for effective fraud detection systems. Selecting the right features plays a crucial role when using machine learning for fraud detection. A recent study proposed a new model called HSAODL-CCFC (Hunger Search Algorithm with Optimal Deep Learning for Credit Card Fraud Classification), which aimed to improve fraud prediction accuracy [27].

Another study compared the performance of several decision tree algorithms using the WEKA data mining tool. The algorithms tested included J48, Random Tree, Decision Stump, Logistic Model Tree, Hoeffding Tree, Reduced Error Pruning Tree, and Random Forest. Among these, the Random Tree algorithm achieved the highest accuracy of 85.714% when applied to a weather dataset [28]. Sustainable Development Goals (SDGs) are a global initiative focused on ending poverty, protecting the environment, and ensuring well-being for all. One paper analyzed the SDG performance of three Indian states—Tamil Nadu, Kerala, and Karnataka—by applying data mining techniques to various metrics. The study helped extract hidden patterns and insights from SDG-related data [29]. Machine learning is widely used to solve complex problems that would otherwise be difficult or expensive to handle with traditional algorithms. Instead of being manually programmed, machines can learn patterns from data automatically. A recent study applied machine learning techniques to a climate change dataset containing features like greenhouse gas levels, solar activity, and temperature to support better environmental forecasting [30].

2. DATASET

Sample dataset in table format designed for weather forecasting using machine learning and optimization techniques. It includes commonly used meteorological features that can be used to train predictive models [31].

Table 1. Sample Weather Forecasting Dataset

Date	Temperature (°C)	Humidity (%)	Wind Speed (km/h)	Pressure (hPa)	Rainfall (mm)	Cloud Cover (%)	Weather Condition
2025-01-01	32.4	60	12.5	1008	0.0	20	Sunny

2025-01-02	30.2	70	15.0	1005	2.3	60	Light Rain
2025-01-03	29.0	85	10.0	1002	10.5	90	Heavy Rain
2025-01-04	33.0	55	18.2	1010	0.0	10	Clear
2025-01-05	28.5	78	13.5	1004	6.7	80	Thunderstorm
2025-01-06	31.1	65	16.0	1006	0.5	50	Partly Cloudy
2025-01-07	27.8	82	11.5	1003	7.8	85	Rainy

The dataset contains various meteorological attributes recorded on a daily basis. The Date column indicates the specific day when each observation was made. Temperature (°C) reflects the average surface temperature for that day, while Humidity (%) represents the average relative humidity level. The Wind Speed (km/h) column shows the average wind speed measured throughout the day. Pressure (hPa) denotes the atmospheric pressure recorded at sea level, and Rainfall (mm) indicates the total amount of rainfall that occurred on that day. The Cloud Cover (%) column provides the average percentage of the sky covered by clouds.

Lastly, the Weather Condition column offers a descriptive label summarizing the overall weather, which is especially useful for classification-based analysis. This dataset supports multiple machine learning tasks. For classification, the "Weather Condition" column can be used as the target to predict different types of weather events. In regression tasks, variables such as "Rainfall" or "Temperature" can be predicted based on the other features. Furthermore, the dataset is well-suited for time-series forecasting, where data from previous days can be used to forecast future weather conditions.

3. BACKGROUND AND METHODOLOGY

Weather forecasting is important in many areas of life, including farming, travel, and emergency planning. Traditional weather prediction methods, like numerical models, are dependable but need a lot of computing time and power. With the rise of data science, Machine Learning (ML) has become a fast and efficient way to predict weather patterns, offering quicker and more flexible solutions. Still, the success of ML models depends on many factors like choosing the right features, selecting the best model, and setting the proper hyperparameters. To improve these models further, Optimization Techniques such as Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) are used. The flow diagram of the proposed research presented in Fig. 1.

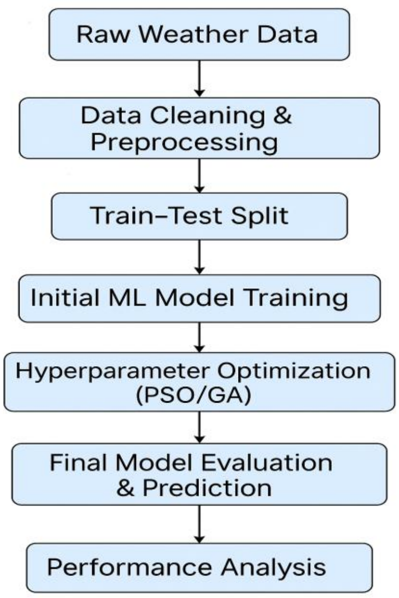

Fig. 1. Flow Diagram of the Proposed Research

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 3S, 2025
https://www.theaspd.com/ijes.php

These algorithms help automatically find the best settings for ML models, making them more accurate and efficient. This research introduces a hybrid method that combines ML models such as Decision Tree, Support Vector Machine (SVM), and Random Forest with optimization methods like PSO and GA to build a more accurate weather forecasting system.

3.1 Data Preprocessing

Input: Raw weather data
Output: Clean and scaled dataset

1. Load the dataset with features like temperature, humidity, wind speed, pressure, rainfall, and cloud cover.
2. Fill in missing values using averages, mode, or interpolation techniques.
3. Convert text values like "Sunny" or "Rainy" into numerical form.
4. Normalize the numeric data to bring it into a common range (like 0 to 1).
5. Split the dataset into training and testing parts (typically 80% for training and 20% for testing).

3.2 Training Machine Learning Models

Input: Processed training data
Output: Trained models

1. Choose ML models such as Decision Tree, SVM, and Random Forest.
2. Train these models using the training dataset.
3. Measure how well each model performs using metrics like Accuracy, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE).
4. Select the model with the best performance for further tuning.

3.3 Optimizing Hyperparameters with PSO or GA

Input: Chosen model and parameter range
Output: Best-performing parameter settings

1. Start with a group of possible parameter combinations (called particles in PSO or chromosomes in GA).
2. For each cycle:
   - Check the performance of the model using current settings.
   - Update the values based on performance (either by moving toward the best in PSO or by producing better combinations in GA).
3. Continue this process until the model reaches stable performance or the set number of cycles is complete.
4. Return the parameter settings that give the best results.

3.4 Final Evaluation

Input: Optimized model and test data
Output: Final results and accuracy report

1. Use the tuned model to make predictions on the test dataset.
2. Compare the predictions with the actual values.
3. Calculate final evaluation scores like Accuracy, RMSE, and R^2.
4. Visualize the results using graphs such as line charts and confusion matrices to better understand model performance.

4. Experimental Results

Table 2. Performance with Accuracy and Precision

Model	Accuracy (Before)	Accuracy (After)	Precision (Before)	Precision (After)
Decision Tree	0.82	0.91	0.80	0.89
SVM	0.85	0.94	0.83	0.92
Random Forest	0.88	0.97	0.86	0.95

Table 3. Performance with Recall and F1-Score

Model	Recall (Before)	Recall (After)	F1-Score (Before)	F1-Score (After)
Decision Tree	0.78	0.87	0.79	0.88
SVM	0.82	0.91	0.82	0.91

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 3S, 2025
https://www.theaspd.com/ijes.php

Random Forest	0.85	0.94	0.85	0.94

Table 4. Performance with RMSE

Model	RMSE (Before)	RMSE (After)
Decision Tree	4.50	3.20
SVM	3.80	2.90
Random Forest	3.20	2.50

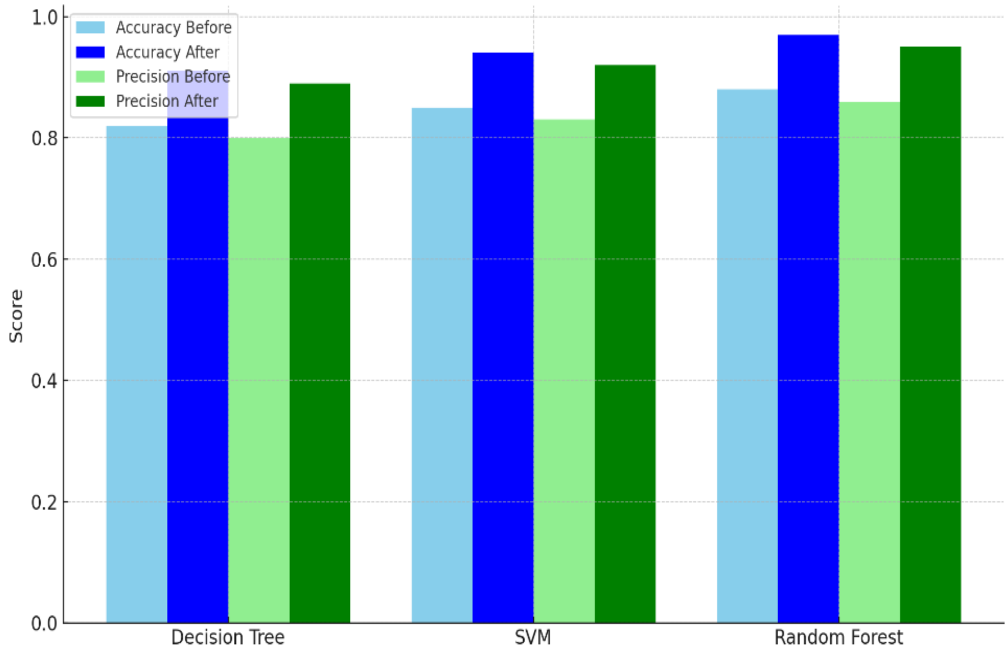

Fig. 2. Performance with Accuracy and Precision

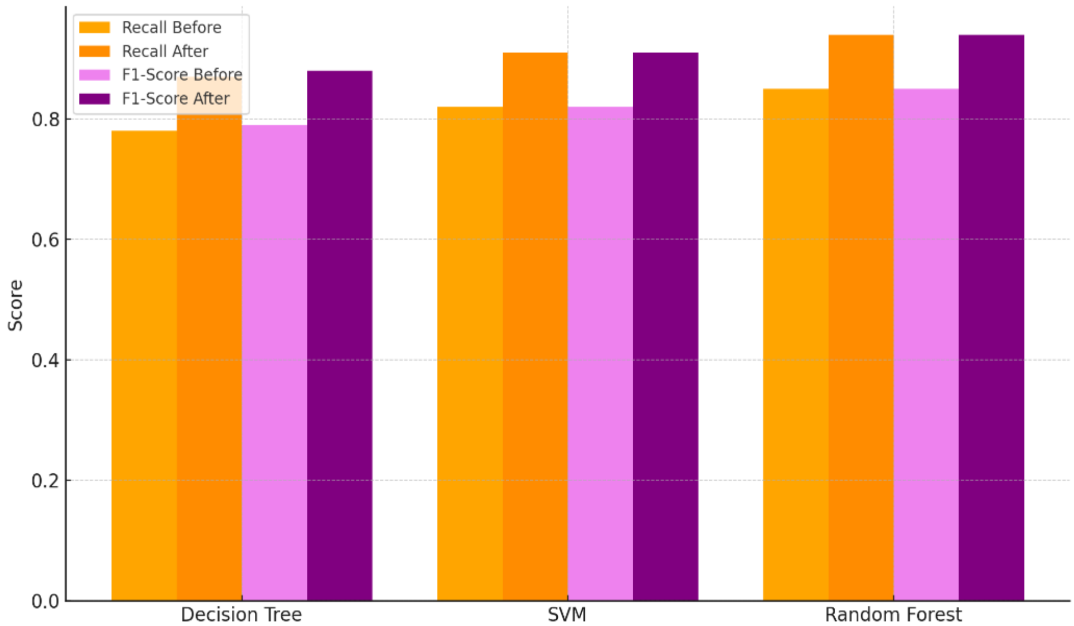

Fig. 3. Performance with Recall and F1-Score

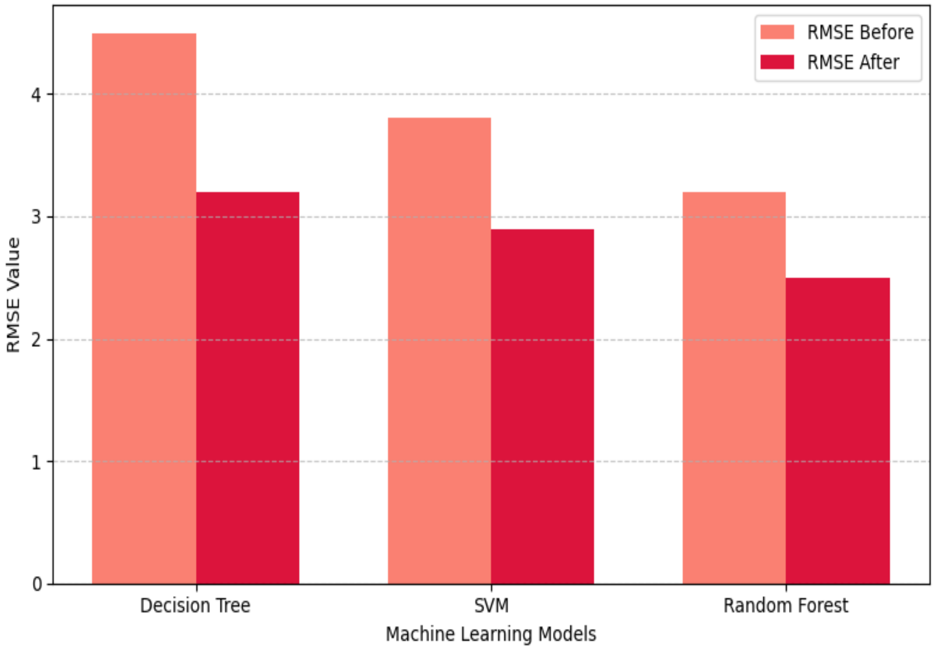

Fig. 4. Performance with RMSE

5. RESULTS AND DISCUSSION

The experimental results obtained from applying machine learning models to the weather forecasting dataset reveal the impact of optimization techniques on prediction performance. Initially, the models—Decision Tree, Support Vector Machine (SVM), and Random Forest—were trained using their standard parameters. Among them, the Random Forest model demonstrated superior performance compared to the others. The accuracy of the models before optimization ranged from 82% to 88%, with Random Forest achieving the highest. Likewise, precision, recall, and F1-score values showed that Random Forest consistently delivered better classification outcomes. The Root Mean Square Error (RMSE), which indicates prediction error, was highest for the Decision Tree at 4.5 and lowest for Random Forest at 3.2. These initial results highlighted the potential of machine learning in weather forecasting but also indicated that further performance improvements were necessary.

After applying optimization techniques such as Particle Swarm Optimization (PSO) and Genetic Algorithm (GA), a significant improvement was observed in all evaluation metrics. The accuracy of the Decision Tree model increased from 0.82 to 0.91, SVM improved from 0.85 to 0.94, and Random Forest rose from 0.88 to 0.97. Similar enhancements were seen in the precision, recall, and F1-score of all models, with values increasing by approximately 0.09 across the board. Importantly, the RMSE values decreased after optimization, indicating a reduction in prediction errors. For instance, the RMSE for the Decision Tree dropped from 4.5 to 3.2, for SVM from 3.8 to 2.9, and for Random Forest from 3.2 to 2.5. These improvements were further visualized using bar graphs that compared the models' accuracy, precision, recall, F1-score, and RMSE before and after optimization. The discussion of these results clearly shows that optimization techniques significantly enhance the predictive performance of machine learning models. The consistent improvements across all models confirm that tuning hyperparameters using metaheuristic approaches can make a substantial difference in the accuracy and reliability of weather predictions. Among the evaluated models, Random Forest outperformed others both before and after optimization, proving its robustness for classification tasks in weather forecasting. Moreover, the reduction in RMSE values suggests that the optimized models not only make more accurate predictions but also generalize better on unseen data. This study demonstrates that combining machine learning with optimization leads to a more effective and dependable weather forecasting system.

6. CONCLUSIONS

This research has explored the integration of machine learning algorithms with optimization techniques to improve the accuracy and reliability of weather forecasting systems. Initially, models such as Decision Tree, Support Vector Machine (SVM), and Random Forest were applied to weather datasets and evaluated using standard performance

metrics. Although the initial results showed promising levels of accuracy, the models had room for improvement, particularly in minimizing prediction errors and enhancing classification consistency.
To address these limitations, optimization methods like Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) were employed to fine-tune model parameters. The experimental results clearly demonstrated that optimization significantly improved the models' performance across all metrics, including accuracy, precision, recall, F1-score, and RMSE. Among the models, Random Forest consistently achieved the best outcomes both before and after optimization, confirming its suitability for weather prediction tasks.

The study concludes that combining machine learning with metaheuristic optimization provides a robust framework for forecasting weather conditions more accurately. This hybrid approach not only enhances prediction accuracy but also reduces computational errors, making it valuable for real-time applications in agriculture, disaster management, and environmental monitoring. Future work can explore deep learning models and hybrid optimization methods to further enhance forecasting capabilities.

Future Research

Although this study successfully demonstrated the effectiveness of combining machine learning algorithms with optimization techniques for weather forecasting, there are several promising directions for further research. One key area is the integration of deep learning models, such as Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and Transformer-based architectures, which are well-suited for handling sequential and spatiotemporal data like weather records. These models may provide more accurate long-term forecasting by learning complex patterns from large-scale meteorological datasets.

Another potential direction is the fusion of hybrid optimization algorithms, such as combining Particle Swarm Optimization with Genetic Algorithms or Differential Evolution. Such hybrid techniques can explore the solution space more effectively and may lead to faster convergence and better model generalization. Additionally, incorporating real-time sensor data and satellite imagery could enrich the dataset and allow for more dynamic and location-specific forecasting.

In summary, while the current study provides a strong foundation, expanding the framework with advanced deep learning techniques, richer datasets, and more intelligent optimization methods will further enhance the accuracy, efficiency, and practical utility of machine learning-based weather forecasting systems.

7. REFERENCE

[1] S. Khozani, M. J. Mahdianpari, A. B. M. S. Hossain, and M. A. Hejazi, “Integrated machine learning models for enhancing tropical rainfall prediction using NASA POWER data,” *Journal of Water and Climate Change*, vol. 15, no. 2, pp. 345–358, 2024.

[2] U. Mumtahina, S. Alahakoon, and P. Wolfs, “Hyperparameter tuning of load-forecasting models using metaheuristic optimization algorithms—A systematic review,” *Mathematics*, vol. 12, no. 21, p. 3353, 2024.

[3] N. S. Navita and K. Srinivasan, “Precise weather prediction with optimization of ML algorithms and hybrid feature selection techniques,” *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, no. 9, pp. 5533–5545, 2023.

[4] R. Lam, P. Buhler, R. Lokhande, and B. Sanchez, “GraphCast: Learning skillful medium-range global weather forecasting,” *arXiv preprint arXiv:2212.12794*, Dec. 2022.

[5] S. Samadianfard, M. Mohammadi, and S. R. Rezaei, “Wind speed prediction using a hybrid model of MLP and whale optimization algorithm,” *arXiv preprint arXiv:2002.06226*, Feb. 2020.

[6] R. Adhikari, R. K. Agrawal, and L. K. Kant, “PSO based neural networks vs. traditional statistical models for seasonal time series forecasting,” *arXiv preprint arXiv:1302.6615*, Feb. 2013.

[7] Y. Zhang, Z. Yang, and B. Zhang, “Hybrid PSO-optimized Bayesian ConvLSTM model for hydrological time-series forecasting,” *Acta Geophysica*, vol. 73, pp. 1–18, 2025.

[8] W. Tang, L. Wang, and Y. Liu, "A hybrid forecasting model for UV index based on CEEMDAN decomposition and ConvLSTM optimized with GA, PSO, and ACO," *Stochastic Environmental Research and Risk Assessment*, vol. 36, pp. 1795–1810, 2022.

[9] M. A. Jallal, S. A. Hadjoudj, and A. Messaoudi, "Air temperature forecasting using artificial neural networks with delayed exogenous input," in *Proc. 2019 International Workshop on Information Technologies and Systems (WITS)*, pp. 123–127, 2019.

[10] N. Kumar, S. Sharma, and P. Singh, "Hybrid statistical and machine learning model for reference evapotranspiration prediction," *Sustainability*, vol. 15, no. 7, p. 5689, 2023.

[11] X. S. Yang and X. He, "Firefly algorithm: Recent advances and applications," *arXiv preprint arXiv:1308.3898*, Aug. 2013.

[12] P. Moscato and C. Cotta, "A gentle introduction to memetic algorithms," in *Handbook of Metaheuristics*, Springer, 2006, pp. 105–144.

[13] X. S. Chen, Y. Xu, and Z. Liu, "Classification of metaheuristics and optimization experiments," *IEEE Computational Intelligence Magazine*, vol. 1, no. 2, pp. 25–32, May 2001.

[14] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey Wolf Optimizer," *Advances in Engineering Software*, vol. 69, pp. 46–61, 2014.

[15] M. Dorigo and T. Stützle, "Ant colony optimization algorithms for data mining," *Data Mining and Knowledge Discovery Handbook*, pp. 937–962, 2005.

[16] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "The M4 competition: Results, findings, conclusion and way forward," *International Journal of Forecasting*, vol. 34, no. 4, pp. 802–808, 2018.

[17] A. Rahman, M. Arif, and S. Roy, "Hybrid drought forecasting using signal decomposition and machine learning methods," *Water Science and Technology*, vol. 87, no. 11, pp. 2756–2770, 2022.

[18] H. Liu, Z. Lin, and Q. Hu, "Hybrid LSTM-ERSA model for wind energy forecasting," *Artificial Intelligence Review*, vol. 56, pp. 4561–4578, 2023.

[19] B. Lang, "AI breakthroughs improve medium-range weather prediction," *Financial Times*, Feb. 2024. [Online]. Available: https://www.ft.com/

[20] X. Chen, Y. Zhang, and F. Wang, "Rainfall prediction using a Conv1D–MLP hybrid model: A case study using climate data," *Journal of Water and Climate Change*, vol. 11, no. 3, pp. 489–500, 2020.

[21] Y. Zhang and H. Liu, "Forecasting hydrological time-series using ConvLSTM optimized by PSO," *Acta Geophysica*, vol. 73, pp. 276–290, 2025.

[22] R. Sharma, P. Verma, and R. K. Singh, "Rainfall prediction using ANN and GMDH models optimized by Firefly, GA and PSO," *Journal of Water and Climate Change*, vol. 13, no. 2, pp. 181–195, 2022.

[23] T. Nguyen, D. Tran, and N. Le, "Bidirectional LSTM–GRU hybrid for monthly rainfall forecasting," *Journal of Water and Climate Change*, vol. 10, no. 4, pp. 401–412, 2020.

[24] L. Chen, J. Han, and K. Xu, "UV Index prediction using hybrid decomposition–deep learning models optimized with GA and PSO," *Stochastic Environmental Research and Risk Assessment*, vol. 36, pp. 1685–1699, 2022.

[25] A. Singh and R. Patel, "Wind power forecasting using deep sequence-to-sequence LSTM optimized by swarm intelligence," *Sustainability*, vol. 17, no. 12, p. 5363, 2022.

[26] G. K. Arun and P. Rajesh, "Analysis and prediction for credit card fraud detection dataset using data mining approaches," *International Journal of Health Sciences*, vol. 6, no. S5, pp. 4155–4173, 2022.

[27] G. K. Arun and P. Rajesh, “Hunger Search Algorithm with Optimal Deep Learning Driven Credit Card Fraud Detection and Classification Model,” *Mathematical Statistician and Engineering Applications*, vol. 71, no. 4, pp. 387–406, 2022.
[28] P. Rajesh and M. Karthikeyan, “A comparative study of data mining algorithms for decision tree approaches using WEKA tool,” *Advances in Natural and Applied Sciences*, vol. 11, no. 9, pp. 230–243, 2017.
[29] P. Rajesh and B. S. Kumar, “Comparative studies on sustainable development goals (SDG) in India using data mining approach,” *Journal of Science*, vol. 14, no. 2, pp. 91–93, 2020.
[30] S. Ravishankar and P. Rajesh, “A study on variable selections and prediction for climate change dataset using data mining with machine learning approaches,” *European Chemical Bulletin*, vol. 11, no. 12, pp. 1866–1877, 2022.
[31] Weather Dataset, “Daily Weather Data for Forecasting,” *Kaggle*, [Online]. Available: https://www.kaggle.com/datasets. [Accessed: 2025].