Global Research journal of Natural Science and Technology (Grjnst)
Volume: 03 - Issue 3 (2025)
p.ISSN: 2790-7643 e.ISSN: 2790-7651

Enhancing Weather Forecasting Accuracy: A Machine Learning Approach Using Genetic Algorithm and Random Forest

Fardad Ali Shah
BS-Computer Science, Department of Computer Science, University of Chitral, Pakistan.
fardadkhan6@gmail.com
Ayesha Bint E Meraj
BS-Computer Science, Department of Computer Science, University of Chitral, Pakistan
ayeshameraj337@gmail.com
Malak Roman
Lecturer, Department of Computer Science, University of Chitral, Pakistan
malak_5116@uoch.edu.pk
Masood Anwar
Lecturer, Department of Computer Science, University of Chitral, Pakistan
anwarmasood@uoch.edu.pk
Awrang Zaib
Lecturer, Department of Computer Science, University of Chitral, Pakistan
awrangzaib1717@gmail.com
Sana Shaiza Shams
BS-Computer Science, Department of Computer Science, University of Chitral, Pakistan.
sanashaizashams@gmail.com
Farzana Hussain
BS-Computer Science, Department of Computer Science, University of Chitral, Pakistan
farzanahussain809@gmail.com

Corresponding author: malak_5116@uoch.edu.pk (Malak Roman)

Abstract: *Weather prediction plays a vital role in numerous fields, including agriculture, transportation, and emergency response. However, the intricate and nonlinear nature of atmospheric progressions makes accurate forecasting a persistent challenge. This research explores the application of machine learning, particularly Random Forest (RF) algorithm to improve weather prediction using historical meteorological data. The study employs a dataset comprising 13,202 records with 11 weather-related variables. Two distinct models were developed: one using raw, unprocessed data and another incorporating Genetic Algorithm (GA)-based feature selection to identify optimal predictors. Both models were evaluated on a test of 3,960 instances. Results indicate that the GA-enhanced model outperformed the baseline, achieving an accuracy of 92.65% compared to 91.31%.*

Additional metrics, including kappa statistics, MAE, RMSE, precision, recall, and F-measure further validated the model's robustness. Notably, both models exhibited strong discriminative ability, with ROC and PRC areas exceeding 0.98, while the optimized model maintained high performance with reduced dimensionality. This study demonstrates that combining Random Forest with Genetic Algorithm-driven feature selection significantly enhances weather prediction accuracy. The proposed approach offers a reliable framework for developing efficient forecasting systems, with potential applications in real-time and long-term meteorological analysis.

Keywords *Weather Forecasting, Machine Learning, Random Forest Algorithm, Genetic Algorithm, Feature Selection.*

Introduction

Technology has seamlessly woven itself into the fabric of our everyday existence. It is no longer just a tool; it is an invisible assistant, a primary source of information, and a central hub for our social and professional interactions (Rahman, M. A. U., et al., 2021). In stream years, machine learning 'ML' has appeared as an effective tool in prediction and forecasting hidden and useful information. Data-Mining (DM) is the combination of two words data and mining, where we mine the useful data from a huge amount of data. Data mining is to excerpt knowledge and rules that are hidden and unknown (Wang, & Mujib, 2017). Data mining is the process of discovering patterns, relationships and insights from large sets of data. It is too difficult to check huge amounts of data it takes, our time so therefore we use data mining to reduce the checking time and effort. Data mining follows the following rules. "Data collection, data cleaning, data transformation and data mining". Data mining plays an imperative role in weather prediction by collecting insights and patterns from large datasets. Here are some ways data mining enhances predicting weather, data mining helps to predict weather conditions over a short period, data mining enhances climatic modeling by evaluating large datasets and identifying patterns (Sawant, S. S., et al., 2024).

Machine learning algorithms play the most important role in refining the accuracy and reliability of weather calculations/forecasting (Ullah, M. A., et al. 2025). Machine learning can analyze large volumes of data from different orchestras, including satellites, weather stations, radar and environmental sensors. The patterns to predict weather events like storms and rainfall. Accurate weather forecast can help inform decisions in many sectors like agriculture and transportation. Machine learning is used to augment computations by replacing some parts of the process. Learning can fully replace the whole model which can be less expensive. ML-enhanced model: machine learning algorithms can be combined into physical models to improve their performance.

Meteorology is the study of weather and weather forecasting. Machine-learning algorithms/techniques are used in weather to predict that tomorrow will be sunny or cloudy. Weather is a mix of events that have happened each day and hour in the atmosphere. The weather is totally different in different places, which changes in minutes, and days. Most of the weather occurs in the troposphere which is closest to the ground. Weather includes wind, storms, rain, sunny and snow. Meteorologists use different tools and techniques to measure or detect the weather, use thermometers to quantify the temperature, Barometers to notch air pressure and Hygrometers- percentage humidity, Anemometers are used to measure the wind speediness, weather balloons are used to carry the sensors that detect the temperature, pressure, wind speed, and wind direction at high altitudes. Supercomputers are used to process data from these techniques to predict forecasts using different mathematical models. Meteorologists also use the climate radar to monitor the weather and provide important evidence wherever rain or snow is falling.

The technique of anticipating the atmospheric conditions of a certain region over a specified time period is known as weather forecasting. It involves collecting data and using scientific models to predict meteorological conditions like temperature, humidity, wind speed, and air pressure. Meteorologists collect this data through various tools. Which monitor real-time changes in the atmospheric Numerical weather predictions models which rely on mathematical equations and algorithms play an important role in simulating future atmospheric states and have revolutionized modern forecasting methods (Bauer, P., et al. 2015). Improved the accuracy of short-term forecasting which predicts weather for a few hours to several days. Long-term forecasting focuses on seasonal but often less reliable due to the chaotic nature of the atmospheric system, a concept first explored in detail by “Lorenz, (1963)” (Kalnay, E., 1996).

Atmospheric conditions at a certain place and time, such as temperature, humidity, wind, air pressure, and precipitation, are referred to as the weather. It is a short-term phenomenon that can change rapidly and vary across different regions. Meteorologists use several tools to predict the weather, such as satellites, weather balloons, and radar. Satellites capture images of cloud patterns, temperature, and moisture in the atmosphere, providing data to track weather systems such as storms and hurricanes (Freeman, E., et al. 2019). Weather balloons, equipped with instruments like thermometers and barometers, are released into the atmosphere to measure changes in temperature, pressure, and humidity at various altitudes (National Research Council., 1998). Radar systems detect precipitation and help track the movement of storms by emitting radio waves that bounce off rain droplets and provide information about storm intensity and location (Joe, P. 1996).

Meteorologists combine all this data with sophisticated computer models to simulate weather patterns and forecast future conditions. Modern weather forecasting also utilizes machine learning techniques to enhance prediction accuracy by analyzing vast amounts of historical weather data to find patterns and trends (Iram, S., et al. 2023). Although weather predictions have become increasingly accurate over the years, certain extreme weather events like hurricanes still pose challenges due to their unpredictable nature (Stuart, L., et al. 2022). By using these tools and continuously updating models with new data, meteorologists aim to provide reliable forecasts that help people prepare for changing weather conditions (Zhang, H., et al. 2025).

Literature Review

Kadam, A., et al. (2023) proposed a research model to explore weather prediction using SVM, ANN, Decision Tree, Naïve Bayes, and Random Forest ML-algorithms. This study demonstrates the use and application of the weather forecasting technology that will be helpful in the near future to build prediction machinery. The study analyzed temperature, wind, and humidity parameters to improve prediction accuracy.

Khan, S., et al. (2022) the research article explores machine learning practices for accurate weather prediction. It compares the performance of algorithms like Naïve Bayes and Logistic regression using a Delhi weather dataset (1996-2016) with attributes like temperature, humidity, and wind direction. After data collection and pre-processing, the study finds that Naïve Bayes algorithm yields better accuracy than other methods.

Patkar, U., et al. (2021) researched evaluated deep learning algorithms for climate prediction using parameters like temperature, rainfall, and wind speed. The study compared Naïve Bayes, Logistic regression, KNN, and Naïve Bayes Gaussian algorithms. After data collection and pre-processing, the Naïve Bayes Bernoulli model outperformed the others, achieving 100% accuracy in both training and testing data.

Singh, S., et al. (2019) This research compared machine learning techniques for weather forecasting using a 12-year data set (2006-2018) from Indian airport weather stations. The study evaluated SVM, ANN, and Time Series RNN algorithms. Results showed Time series RNN outperformed the others, achieving the lowest RMSE (1.41) over an 8-week prediction window. RMSE values: Time series RNN: 1.41, SVM: 6.67, ANN: 3.1.

Singh, N., et al. (2019) researched aimed to predict weather using Delhi's 20-year weather data. After data collection and pre-processing, a new data set was derived by averaging temperature, humidity, pressure, and rain. The study used a Random Forest classification algorithm, training the model with 75% of the data and testing with 25%. The goal was to develop a low-cost, consistent, and effectual weather forecasting system using machine learning on a Raspberry Pi board.

Pavani et al., (2024) propose a multi-linear regression model to predict heavy rainfall, presented at the IEEE I3CEET conference. The study leverages historical meteorological data to identify key variables (e.g., temperature, humidity, wind speed) that influence extreme precipitation events. By applying statistical analysis, the model demonstrates improved accuracy in forecasting heavy rainfall compared to traditional methods. The authors highlight its potential for early warning systems and disaster preparedness, particularly in climatevulnerable regions. The paper includes validation results using real-world datasets and discusses practical implementation challenges.

Oshodi, et al. (2022) exploits a Seattle weather data collection (2012-2015) from Kaggle to teach robots to forecast weather conditions. The study contrasted algorithms such as Gaussian Naïve Bayes, Random Forest, Gradient Boosting, and Decision Tree. With an accuracy of 84.15%, the Gaussian Naïve Bayes model outperformed the others.

Methodology

Researchers obtained a data set for weather prediction from the Kaggle database repository. It contained 13,202 entries and 11 attributes. After data preprocessing, two data sets were created: one for training the machine and the second for model testing. A Genetic algorithm was implemented for best and optimal feature selection. The data set was loaded to train the model i.e., random forest (RE), and applied separately to verify and assess the model's correctness. The data was tested after the model had been trained. Accuracy was calculated using it, as Figure 1 illustrates.

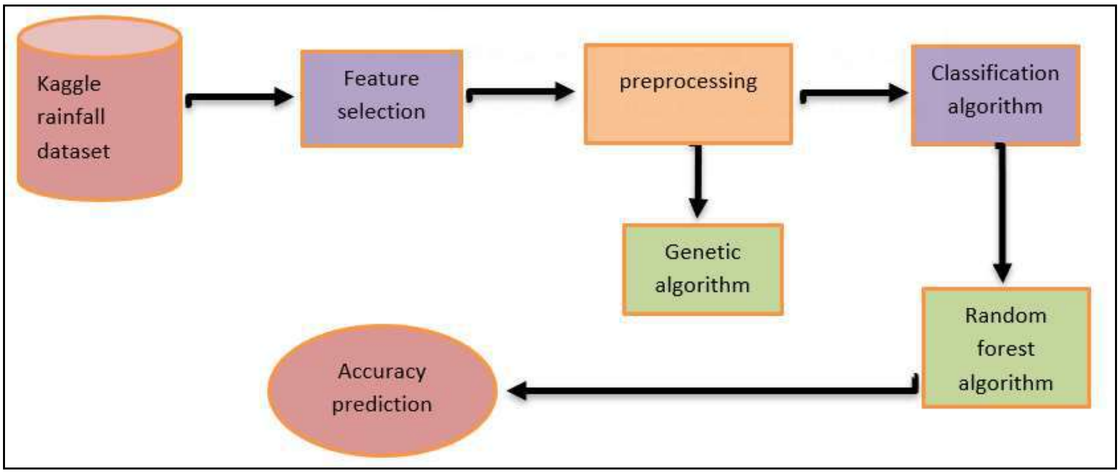

Fig 1. Research Flow Diagram

Data Collection and Preprocessing

The data set used in this research was sourced from Kaggle, comprising 13,202 entries with 11 attributes, which are shown in Table 1. It classifies the weather into four categories: rainy, sunny, cloudy, and snowy.

S.NO	NAME	TYPE	DESCRIPTION
1	Temperature	Numeric	The temperature is degrees Celsius. Ranging from extreme cold to extreme heat.
2	Humidity	Numeric	The humidity percentage, including values above 100% to introduce outers.
3	Wind speed	Numeric	The wind speed is kilometer per hour, with a range including unrealistically high values.
4	Precipitation (%)	Numeric	The precipitation percentage, including outer values.
5	Cloud cover	Categorical	The cloud cover description.
6	Atmospheric pressure	Numeric	The atmospheric pressure in upconverting a wide range.
7	UV index	Numeric	The UV index, indicating the strength of ultraviolet radiation.
8	Season	Categorical	The session during which the data was recorded.
9	Visibility (km)	Numerical	The visibility in kilometers, including very low or very high values.
10	Location	Categorical	The type of location where the data was recorded.
11	Weather	Categorical	The target variable for classification, including the weather type.

Table 1: Attributes of data set.

Preprocessing

Transforming the collected data into a comprehensible format, eliminating duplicate or null values, and eliminating undesirable features are all part of data preparation (Oshodi, I. 2022). Data preprocessing required better accuracy in rainfall prediction. For preprocessing we use genetic algorithms.

Genetic Algorithm (GA)

In 1975, American researcher Holland proposed the Genetic Algorithm (GA). It is a type of effective search algorithm and optimisation that mimics the survival of the fittest principle of genetics and natural selection (Zhang, J. 2018). Its broad adaptability and powerful problem-solving skills have allowed it to successfully enter a number of technical and research domains in recent years. One type of random global search algorithm that does a random search of the target space is the Genetics algorithm (Haldurai, L., et al. 2016). Genetic algorithms, which draw inspiration from the biological development of living things, abstract the problem space as a population of individuals and iteratively produce generations to find the fittest individual. GA matures into a population of excellent people, each of whom is a solution to the issue at hand. A fitness function, which quantifies how well a rule adapts to a certain environment, is used to evaluate each rule's quality. An initial population of randomly chosen people is used to begin the process (Roman. M., at al. 2022).

To create the next generation, the following stages are used.

a) Selection stage: In this stage, an individual is selected for a further process called parent, to incipient a new generation.

b) Crossover stage: The individuals as parents are intersected to reproduce the next offspring.

c) Mutation stage: The superlative offspring offspring is completely transmuted from the prior population (Roman. M., at al. 2022).

S.NO	**NAME**	**S.NO**	**NAME**
1	Temperature	2	Precipitation (%)
3	Cloud cover	4	Atmospheric pressure
5	UV index	6	Season
7	Visibility (km)	8	Location

Table 2: Best Feature Selected by Genetic Algorithm

In the preprocessing stage, GA selects the eight best attributes i.e. temperature, perception%, cloud cover, atmospheric pressure, UV index, season, visibility km, and location as shown in table 2. The genetic algorithm searching steps with pseudocode is discussed in figure 2.

$C, T, fbest \leftarrow \varnothing \rightarrow$ Initialize.

$Po \leftarrow$ Gauss random distribution with $\sigma = 0.15$ *and* $\mu = 0.3$

Each genes is converted to binary discrete.

Such that $\begin{matrix}\{0, & pi < 0.5\} \\ \{1, & otherwise\}\end{matrix}$.

while $t \leq T$ **do.**

$t \leftarrow t + 1$

GA (Pt).

if $argmaxt\ (Pt) \geq then\ fbest \leftarrow argmaxt(Pt).$

end while

return *best p* $\rightarrow$ Return the best individual.

Fig 2: Genetic Algorithm Pseudocode adapted from (Roman. M., at al. 2024)

Random Forest Algorithm (Rf)

In 2001, Breiman introduced the concept of Random Forests, which solves the overfitting issue and outperforms existing classifiers such as Support Vector Machines, Neural Networks, and Discriminant Analysis (Ali, J., et al. 2012). An ensemble of decision trees is produced using Random Forest. Breiman chose the randomization approach, which is effective with bagging or random subspace approaches, to provide variety among the basis decision trees (Kulkarni, & Sinha, 2013). Random forests can be applied to a continuous response known as regression or to a categorical response variable like classification. In a similar vein, expectancy variables may be continuous or categorical (Salman, H.A., et al. 2024). They can accommodate a huge number of input variables without overfitting, generate incredibly accurate predictions, and are quick and simple to construct. In fact, they are one of the most accurate general-purpose learning techniques available (Ren, Q., et al. 2017).

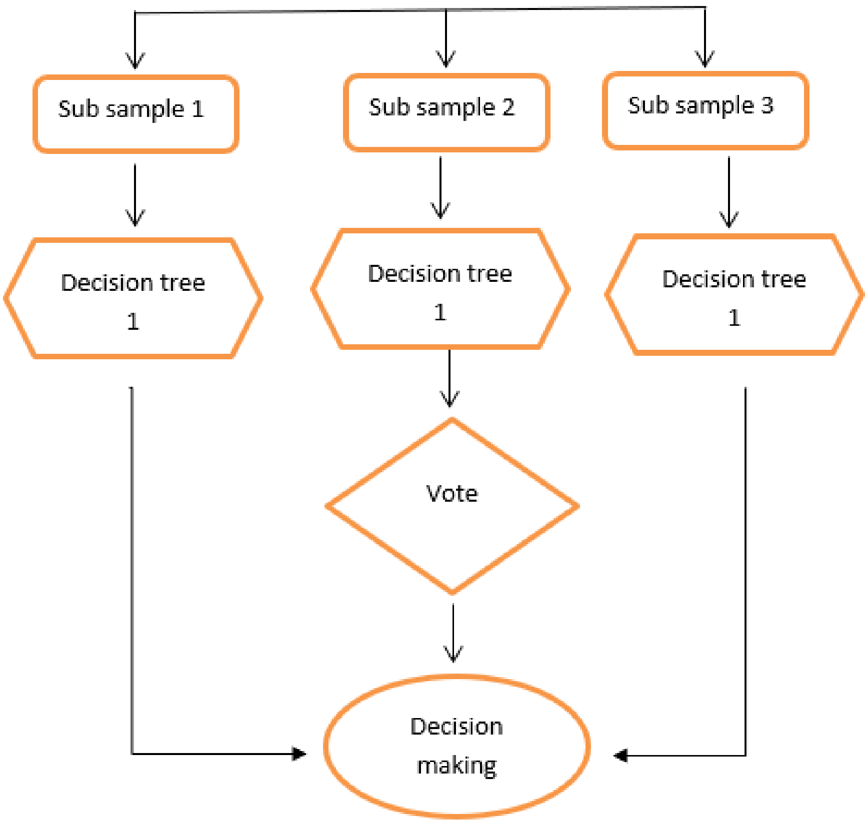

Fig 3: Working Flow chart of Random Forest Algorithm adopted from (Ren, Q., et al. 2017)

Results And Discussion

A machine learning approach using the Random Forest (RF) algorithm was applied to predict weather conditions on a dataset consisting of 13,202 entries and 11 attributes sourced from Kaggle. After data preprocessing using the Genetic Algorithm (GA), the eight most relevant features were selected to improve prediction accuracy. The model was trained and tested on a distinct test and training datasets. The classifiers were assessed using standard evaluation metrics including accuracy, precision, recall (sensitivity), false positive rate (FPR), and false negative rate (FNR) (Anwar, M., et al. 2025). These metrics were computed using the following formulas:

$$Accuracy = \frac{No\ of\ correct\ predictions}{Total\ no\ of\ predictions} \qquad (equ. 2)$$

$$Precision = \frac{True\ Positive}{True\ positive + False\ Positive} \qquad (equ.3)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \qquad (equ.4)$$

$$F1_{Score} = 2 \times \frac{Precision\ \times Recall}{Precision + Recall} \qquad (equ.5)$$

The Random Forest classifier achieved a high classification accuracy of 92.65%, correctly classifying 3,669 instances and misclassifying only 291. The Kappa statistics were recorded as 0.902, indicating strong agreement between predicted and actual classes beyond chance. Error metrics such as Mean Absolute Error (0.0444) and Root Mean Squared Error (0.1489) were low, affirming the precision of the model used by (Roman, M., et al. 2025). In class-wise performance, snowy had the highest precision (0.948) and an F-measure of (0.941). Sunny had a precision of (0.943) and recall of (0.924) as demonstrated in fig 4 and table 3. Cloudy achieved balanced precision (0.914) and recall (0.908). Rainy showed the highest recall (0.941) and a precision of 0.902. The ROC Area and PRC Area for all classes were greater than 0.98, which demonstrates that the model is highly effective in distinguishing between different weather types.

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.933	0.017	0.948	0.933	0.941	Snowy
0.924	0.019	0.943	0.924	0.933	Sunny
0.908	0.027	0.914	0.908	0.911	Cloudy
0.941	0.035	0.902	0.941	0.921	Rainy

Table 3: Comparison in Term of Classes (snowy, sunny, cloudy, and rainy)

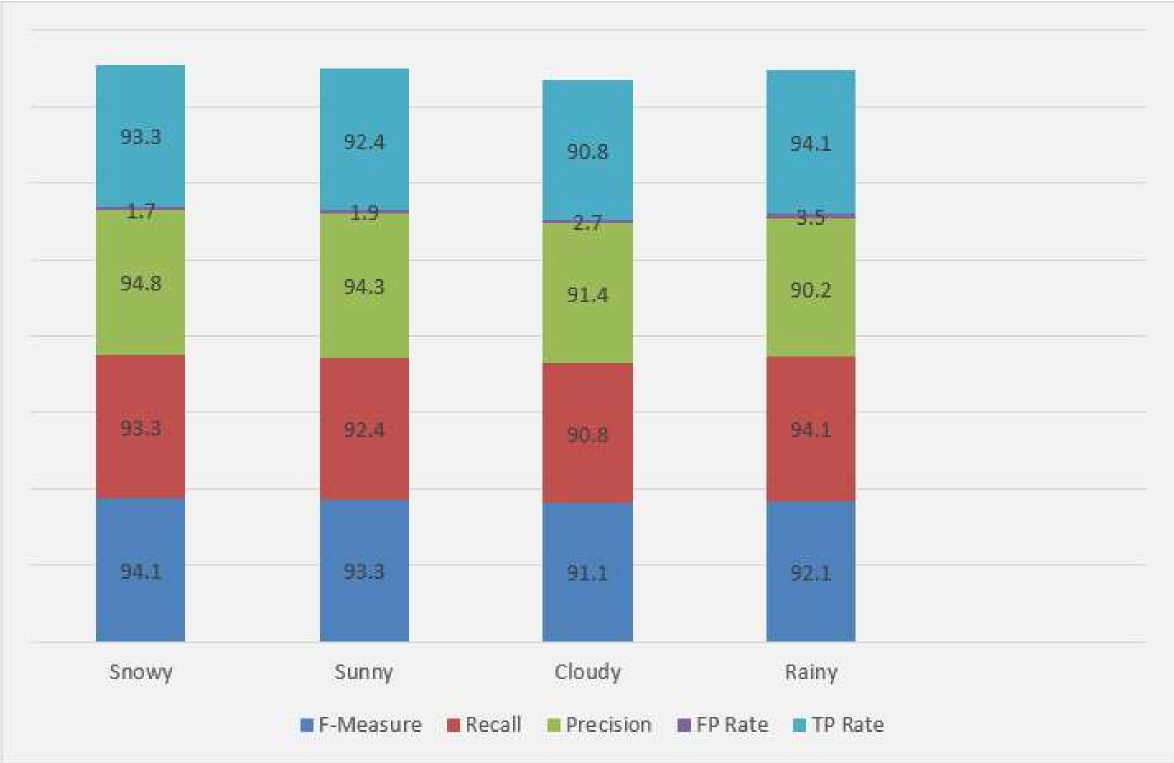

Fig 4: Comparison in Term of Classes (snowy, sunny, cloudy, and rainy)

In the subsequent experimental phase, the Random Forest algorithm was trained and tested on the original dataset using all 11 attributes, without applying any preprocessing or feature selection techniques. Performance metrics are correctly classified instances 3,616, incorrectly classified instances 344, the accuracy is 90.31%, kappa statistics 0.8841, Mean absolute error (MAE) 0.0496, Root mean squared error (RMSE) 0.1568. From table 4 and figure 5, it is evident that preprocessing using Genetic Algorithm significantly improved the model's accuracy and reliability. The refined dataset helped reduce overfitting, improved learning efficiency, and eliminated noisy or redundant features that may have confused the classifier.

Model	**Description**	**Accuracy (%)**
Model 1	Random Forest with Genetic Algorithm	92.65
Model 2	Random Forest without Genetic Algorithm	90.31

Table 4: Models' Performance Comparison in Terms of Accuracy

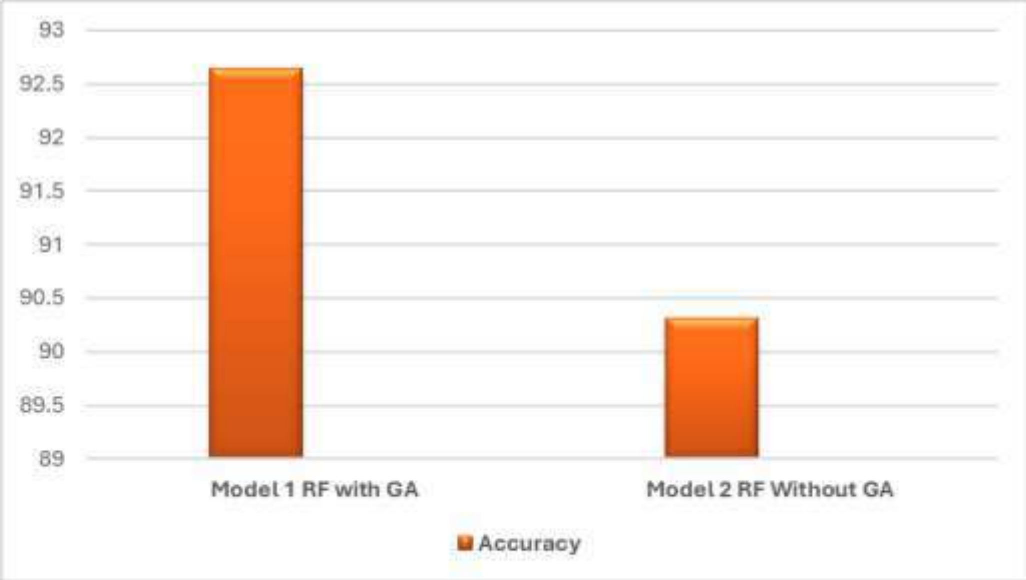

Fig 5: Models' Performance Comparison in Terms of Accuracy

Conclusion

This study demonstrates how Random Forest Algorithms, combined with Genetic Algorithm driven feature selection, significantly enhance weather forecasting accuracy. By filtering out irrelevant data, the model efficiently identifies the most critical weather attributes, improving both performance and computational efficiency. The system achieved high accuracy and consistency across all weather conditions (rain, snow, sun, and clouds), proving its reliability for real-world applications. These results highlight the potential of machine learning to deliver scalable solutions for industries like agriculture, transportation, and emergency response. To further refine prediction, future work could explore: integration of real time sensor data dynamic updates, adoption of deep learning model e.g. LSTMs for temporal patterns, CNNs for spatial data Ensemble techniques to better predict are or extreme weather events.

References

Rahman, M. A. U., Roman, M., Ahmad, S., Jan, M. S., & Nawab, H. U. (2021). Optimizing Collaboration: Insights into Inter-Team Coordination and Self-Management in Distributed Agile Software Development. *Webology (ISSN: 1735-188X)*, *18*(5).

Wang, Z., & Mujib, A. M. (2017). The Weather Forecast Using Data Mining Research Based on Cloud Computing. In *Journal of Physics: Conference Series* (Vol. 910, No. 1, p. 012020). IOP Publishing.

Sawant, S. S., Karandikar, A. A., Dhatrak, K. M., Kapse, N. Y., Gedam, K. R., & Gosavi, K. D. (2024). A Comparative Study of Machine Learning Algorithms for Weather Prediction in India. In *2024 7th International Conference on Circuit Power and Computing Technologies (ICCPCT)* (Vol. 1, pp. 1489-1493). IEEE.

Ullah, M. A., Ullah, A., Roman, M., Anwar, M., Siddiqi, M. U. M., Jaffar, H., & Ali, J. (2025). A Comparative Study of Machine Learning Algorithms for Cardiovascular Risk Prediction: Support Vector Machine, Gradient Boosting, And Rotation Forest. *Spectrum of Engineering Sciences*, *3*(8), 481-491.

Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, *525*(7567), 47-55.

Kalnay, E. (1996). The NCEP/NCAR 40-yr reanalysis project. *Bull. Am. Meteorol. Soc.*, *77*, 431-477.

Freeman, E., Kent, E. C., Brohan, P., Cram, T., Gates, L., Huang, B., .& Zhang, H. M. (2019). The international comprehensive ocean-atmosphere data set–meeting users needs and future priorities. *Frontiers in Marine Science*, *6*, 435.

National Research Council, Division on Engineering, Physical Sciences, Commission on Engineering, Technical Systems, & National Weather Service Modernization Committee. (1998). *Future of the National Weather Service Cooperative Observer Network*. National Academies Press.

Joe, P. (1996). Precipitation at the ground: Radar techniques. In *Radiation and Water in the Climate System: Remote Measurements* (pp. 277-321). Berlin, Heidelberg: Springer Berlin Heidelberg.

Iram, S., Al-Aqrabi, H., Shakeel, H. M., Farid, H. M. A., Riaz, M., Hill, R., & Alsboui, T. (2023). An innovative machine learning technique for the prediction of weather based smart home energy consumption. *IEEE Access*, *11*, 76300-76320.

Stuart, L., Lüterbacher, J., Paterson, L., Deviller, R., & Castonguay, S. (2022). United In Science 2022: a multi-organization high-level compilation of the most recent science related to climate change, impacts and responses.

Zhang, H., Liu, Y., Zhang, C., & Li, N. (2025). Machine learning methods for weather forecasting: a survey. *Atmosphere*, *16*(1), 82.

Kadam, A., Sathe, R., & Idhate, S. (2023). *Weather Prediction Using Machine Learning*. pp. 99-102

Khan, S., Mishra, M. P., Khatoon, R., & Singh, R. (2022). Weather Prediction using Machine Learning. *IJARIIE-ISSN (O)*.

Patkar, U., Masked, S., Ahmad, S., Mangoade, R., & Sadawarti, G. (2021). Weather Prediction using Machine Learning. *GIS Science Journal*.

Singh, S., Kaushik, M., Gupta, A., & Malviya, A. K. (2019). Weather forecasting using machine learning techniques. In *Proceedings of 2nd international conference on advanced computing and software engineering (ICACSE)*.

Singh, N., Chaturvedi, S., & Akhter, S. (2019). Weather forecasting using machine learning algorithm. In *2019 International Conference on Signal Processing and Communication (ICSC)* (pp. 171-174). IEEE.

Pavani, G., Ganesh, D., Ramana, P. V., Ubbala, V., & Reena, T. (2024). Application of Multi Linear Regression Model for Predicting Heavy Rainfall. In *2024 International Conference on Communication, Computing and Energy Efficient Technologies (I3CEET)* (pp. 273-278). IEEE.

Oshodi, I. (2022). Machine learning-based algorithms for weather forecasting. *Int. J. Artif. Intell. Mach. Learn*, *2*(2), 12-20.

Hussain, S., & Khan, M. Q. (2023). Student-performulator: Predicting students' academic performance at secondary and intermediate level using machine learning. *Annals of data science*, *10*(3), 637-655.

Roman, M., Naz, I., Luqman, M. A., Ali, J., Jan, M. S., & Nawab, H. U. (2024). Stroke Disease Prediction Using K-Nearest Neighbor and Decision Tree Algorithms with Machine Learning Pre-Processing Techniques. *Migration Letters*, *21*(S4), 2015-2027.

Roman, M., Nawab, H. U., Ahmad, S., & Khan, I. A. (2022). K-Nearest Neighbor and Fuzzy K-Nearest Neighbor Algorithm Performance Analysis for Heart Disease Classification. *Webology (ISSN: 1735-188X)*, *19*(1).

Zhang, J. (2018). The application research of genetic algorithm. In *2018 3rd International Workshop on Materials Engineering and Computer Sciences (IWMECS 2018)* (pp. 138-141). Atlantis Press.

Haldurai, L., Madhubala, T., & Rajalakshmi, R. (2016). A study on genetic algorithm and its applications. *Int. J. Comput. Sci. Eng*, *4*(10), 139-143.

Ali, J., Khan, R., Ahmad, N., & Maqsood, I. (2012). Random forests and decision trees. *International Journal of Computer Science Issues (IJCSI)*, *9*(5), 272.

Kulkarni, V. Y., & Sinha, P. K. (2013). Random forest classifiers: a survey and future research directions. *Int. J. Adv. Comput*, *36*(1), 1144-1153.

Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random forest algorithm overview. *Babylonian Journal of Machine Learning*, *2024*, 69-79.

Biau, G. (2012). Analysis of a random forests model. *The Journal of Machine Learning Research*, *13*(1), 1063-1095.

Ren, Q., Cheng, H., & Han, H. (2017). Research on machine learning framework based on random forest algorithm. In *AIP conference proceedings* (Vol. 1820, No. 1, p. 080020). AIP Publishing LLC.

Anwar, M., Rahman, T., & Roman, M. (2025). Voice-Activated Smart Environments: Deep Learning Approach for Pashto Speech Command Processing. *Spectrum of Engineering Sciences*, *3*(5), 541-550.

Roman, M., Ullah, A., Ullah, M. A., Hussain, F., Shams, S. S., Bint-e-Meraj, A., & Ali, S. (2025). Predicting Academic Success: A Machine Learning Approach Using Decision Tables and Random Forests Algorithms. *Spectrum of Engineering Sciences*, *3*(5), 205-213.