International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 6, 2025
https://theaspd.com/index.php

Weather Predictive System By Using Machine Learning

Katta Trinadha Ravi Kumar[1], Dr. P Suresh Varma[2], Dr. M V Rama Sundari[3]
[1]Research Scholar, Dept. Of CSE, Adikavi Nannaya University, Rajamahendravaram &Assoc. Professor, Dept. of Computer Science, SVKP & Dr.K.S.Raju Arts & Science College(A),Penugonda-534320,
[2]Professor, Department of CSE, Adikavi Nannaya University, Rajamahendravaram, India.
[3]Professor, Dept.of AIML, Gokaraju Ranga Raju Institute of Engineering and Technology, Hyderabad, India, trinadhaknu9@gmail.com

Abstract
Forecasting the weather has a significant impact on both life and production. Technological advancements have led to the emergence of many weather forecasting techniques, including numerical weather forecasting, quantitative forecasting, and weather map forecasting. These conventional techniques for analysing data, however, have drawbacks, including inadequate objectivity, limited analysis, and an inability to make accurate weather predictions. The purpose of this predictive system is to forecast the weather and alert individuals to changes in air quality or other weather conditions that might have an impact on their general health. It is still difficult to provide a prediction model for climate forecasting that is both highly accurate and effective. Thus, by combining machine learning with this prediction method, precise results may be achieved. This prediction allows humans to get warnings in advance of natural calamities such as floods. This will lessen property damage and assist preserve the lives of residents in low-lying regions. This technique primarily aids farmers in preventing serious crop damage. Therefore, when compared to Naive Bayes (NB), this suggested model Support Vector Machine (SVM) does better when it comes to F1 Score, Accuracy, and Recall.

Keywords: *Support Vector Machine (SVM), Naive Bayes (NB), Weather Forecasting, Prediction*

1. INTRODUCTION

Weather forecasting has always been one of the world's most difficult scientific and technological issues. The development of technology has made it easier to produce forecasts through the use of intricate mathematical models. Better results have been obtained with the development of machine learning-based learning models, such as neural networks, genetic algorithms, and neuro-fuzzy logic. [1]. Modern weather forecasting is characterised by a highly quantitative and complicated nature. Statistical methodologies, numerical methods, and synoptic weather forecasting are a few of the several techniques utilized in weather forecasting [2].

For the purpose of planning and organising our daily schedules, weather forecasting is essential. Weather forecasts help us make wise decisions on a given day. The predicting of the weather might affect our everyday routines. With fast and accurate weather forecasts, a strong forecasting model may assist reduce damages and losses. For the majority of applications, including those in agriculture, the military, aviation, etc., weather attribute estimate or prediction is essential. A system that enhances forecasting outcomes is required for many applications such as weather warnings and advisories, cloud behaviour prediction for air travel, agricultural development, military uses, etc. Techniques for forecasting weather attributes include sky observation, radar usage, ground-based and/or satellite image use, machine learning approaches, etc. [3].

Machine learning models employ techniques like neural networks, fuzzy inference, genetic algorithms, and multiple regressions, whereas statistical models use techniques like exponential smoothing and multiple regressions. The primary distinction between the two is that machine learning methods are often connected with nonlinear data, whereas statistical approaches are typically linked with linear data. Genetic algorithms have been used in research for a variety of weather prediction applications [4]. Meteorologists analyse and forecast future weather patterns in addition to the present weather trend. Since the camera points upward, it is easy to capture low-lying clouds with whole sky imagers, which are increasingly being used by researchers for a variety of purposes and fields where the final photos produced have a greater resolution than what can be obtained from satellites.

In contrast to geographic data, meteorological conditions are always changing, even at the same location. As a result, fluctuations in weather over time give richer data and have an influence on a larger range of sectors [5]. While the components of weather predictions differ from location to place and from season to season, in an ideal world, they would include pressure, depressions, wind speed, temperature, relative humidity, and cloud cover in the sky.

Humans and weather conditions are tightly intertwined since the weather has a direct impact on every action that humans take. Additionally, a variety of industries including agriculture, transportation, tourism, and others greatly depend on this climatic data. The variables influencing the weather are temperature, humidity, air pressure, wind speed, cloud cover, and sun radiation. One of the numerous detrimental repercussions of anomalous weather events is flooding, which may seriously impair public infrastructure, transportation, and business ventures in local communities. Since research has forecast the presence of rain, it is important to have more specific knowledge on rainfall.

Numerous techniques and models have been used to determine weather conditions, such as micro rain radar data to identify clouds that produce high daily rainfall and rainfall characteristics [6].The weather in the time series data is then predicted using the Adaptive Nero Fuzzy Inference System (ANFIS). Weather forecasting can be done by data mining, and rainfall during the dry season can be predicted using Support Vector Regression Modelling based on SOI and NINO. A fuzzy inference method can also be used to estimate rainfall in northern Surabaya based on the neighbour's proximity.

The ability to forecast future events and make arrangements for them is a must for living in the modern world. A sound forecast is not one that is based on conjecture without supporting evidence, instead, one that is predicated on the way symptoms behave or on recurring patterns,. Numerous tools and techniques are available for prediction; they include time series analysis and Support Vector Machine (SVM), an artificial intelligence-based technique [7]. Support Vector Machines may be used to solve classification problems, where the objective is to use functions derived from the given data to distinguish between two or more classes of examples.

The following are some examples of research cases that make use of the Support Vector Classification method: an analysis of how Support Vector Machines (SVM) affect the classification of Microarray data for cancer detection, the method's implementation, and the SVM-based diabetic retinopathy detection system [8]. Identification of Cabbage Plant Leaf Disease Using Support Vector Machine Technique Loss function regression, or support vector regression (SVR), is another way that Support Vector may be used to solve regression problems. SVR loss functions include quadratic, Huber, Laplace, and ɻ-insensitive functions [9]. SVR has been used to prediction scenarios in earlier research, such as the forecasting of crude palm oil utilising a SVR Radial kernel base and the prediction of the rupiah's exchange rate vs the US dollar using SVR experts. Support vector regression is used to anticipate the supply and demand of pulpwood and to model rainfall prediction in the dry season using the Southern Oscillation Index and NINO [10].

1.1 Novelty: A new, multi-perspective method of forecasting is introduced by the use of many machine learning algorithms, including Naive Bayes (NB), Logistic Regression (LR), and Support Vector Machine (SVM), in weather prediction systems. When used in tandem, the distinct advantages of each method improve the model's accuracy, resilience, and flexibility across a range of weather scenarios. Naïve Bayes is perfect for preliminary predictions and probability estimation because of its simplicity and speed, which enable it to handle big datasets and noisy features. Conversely, logistic regression offers interpretability and unambiguous probabilistic results, as well as a strong statistical basis for binary and multi-class classification tasks. Support Vector Machines are essential for capturing intricate weather patterns because of their strong classification capabilities and non-linear modeling with kernel functions.

Making use of these models' complementing qualities is what's innovative. SVM tackles the complex and non-linear elements of atmospheric dynamics, whereas NB and LR are effective for linear patterns and broad trends. More accurate and balanced forecasts are made possible by combining them using ensemble learning or voting-based methods. The overall performance of the system is also improved by using hybrid or comparative modeling approaches, in which the advantages of each algorithm are assessed across several meteorological parameters (such as temperature, precipitation, and wind speed). Prediction accuracy is increased by such a multi-model framework's improved ability to adjust to regional and seasonal differences.

Additionally, this method promotes interpretability and explainability of the model, which is particularly crucial when making decisions on weather in urban planning, agriculture, and disaster management. This prediction system's integration of NB, LR, and SVM guarantees increased accuracy and flexibility while also introducing a novel, layered learning approach that can outperform individual models. An important development in data-driven meteorological forecasting is this multi-algorithm framework.

1.2 Data Set Description: Selecting the right dataset is essential for developing a successful weather prediction system with machine learning techniques like Naive Bayes (NB), Logistic Regression (LR), and Support Vector Machine (SVM). The perfect dataset should cover a large range of meteorological variables

over a sizable time period, be thorough, and be well-structured. In order for the models to forecast variables like rainfall, temperature, humidity, and weather conditions (such as sunny, cloudy, rainy, and stormy), the dataset should enable both classification and regression tasks. The Weather Dataset from NASA's POWER Project, NOAA's Global Historical Climatology Network (GHCN), or the Indian Meteorological Department (IMD) would all be excellent choices. Long-term historical data, frequently spanning decades, is included in these databases, which is crucial for developing reliable predictive models. Essential Elements of the Perfect Dataset

Temporal Coverage: The dataset should include daily or hourly granularity and at least 10–30 years of historical weather data. Capturing seasonal, monthly, and annual fluctuations requires this.

Geographical Scope: To enable regional forecasts, location-specific information (latitude, longitude, elevation) should be provided. Spatial analysis and cross-regional model generalization are supported by multi-location datasets.

Essential Meteorological Parameters:

- Temperature: daily average, maximum and minimum temperatures.
- Precipitation/Rainfall: The amount of rainfall each day (mm).
- Humidity: The relative humidity during the day.
- Wind Direction and Speed. The pressure of the atmosphere.
- Solar radiation and cloud cover (if available).
- Evaporation, Dew Point, and Visibility (optional but helpful).

Target Labels: The target variable may vary depending on whether the prediction objective is regression or classification.

- Binary, like "Yes" or "No" for rain.
- Multi-class (for example, meteorological conditions: rainy, cloudy and sunny).
- Continuous (e.g., rainfall in millimeters or temperature in degrees Celsius).

Data Quality and Consistency:

- The dataset needs to allow for imputation or have a small number of missing values.
- Standardized time formats and units. Model performance depends on clean, de-duplicated and normalized values.

Data Format: The best format is a structured time-series or CSV file. Clearly defined timestamps, feature columns, and target labels are essential.

Why It’s Suitable for NB, LR, and SVM:

- Naive Bayes: Effectively manages probabilistic predictions from tabular data and performs well with categorical characteristics such as season or weather type. For binary or multi-class weather occurrences, logistic regression works best when there are linear or semi- linear connections between features and outputs.
- SVM: Excels that contain non-linear and high-dimensional data. Because of the dataset's extensive feature set and depth of history, complex weather behavior can be modelled using kernel methods. Implementing a multi-model forecasting system with NB, LR, and SVM requires a rich, clean, and comprehensive meteorological dataset with both continuous and categorical variables, a large time span, and a variety of parameters. It facilitates scalability across locations and meteorological events in addition to improving generality and accuracy.

Thus, this study presents the application of machine learning algorithms for weather prediction. The remaining material is grouped in the following way: Section II provides a description of the literature review. Section III illustrates the machine learning approach used by the weather interactive forecast system. Section IV discusses the outcome analysis of the recommended method. Section V represents the end of the work.

2. LITERATURE SURVEY

Convolutional Recurrent Neural Networks (CNN) was suggested as a method for estimating weather temperature forecast by W.-T. Chu, K.-C. Ho, and A. Borji et al. [11]. Two scenarios were used to estimate the temperature: one used a single outside photograph, while the other used an image series. Only picture data is used to estimate the temperature. When estimating the temperature of a single picture, CNN is utilised, and when estimating the temperature of the final scene from a series of photos, RNN (Recurrent Neural Network) is employed.

Wei-Ta Chu, et al. [12] claimed to be able to "estimate weather information from single images by constructing computational models using random forest classifiers." Creating a vast image collection with different weather conditions, image metadata, and elevation from different platforms was the major objective. Subsequently, data were collected to illustrate the connections between photo-taking practices and meteorological characteristics. Lastly, a random forest approach-based weather type classifier based on visual features and photo-taking time was constructed

The authors emphasise the necessity for specifically created features or for comments and tags to be added to photos in order to create more sophisticated computer models, in order to advance this field of study.

A collaborative learning strategy for dividing the weather into two categories (sunny and overcast) utilising data-driven CNN features with weather-specific features was developed by C. Lu, D. et al. [13]. The Sun, Labelme, and Flickr datasets were used. Larger datasets that can be used to classify additional meteorological situations can be used to make further breakthroughs.

" created a system using modular neural networks and multi-feature texture analysis that will enable the automated process of classifying different forms of clouds via satellite image interpretation (The following: stratocumulus, stratus, cirrus cirrostratus, cumulonimbus, and cumulus - stratocumulus) which, according to C. I. Christodoulou, et al., can be utilized for weather analysis. [14]. The neural network Self Organising Feature Map (SOFM) classifier and the statistical K-nearest Neighbour (KNN) classifier were used to classify the cloud photos. The developed system effectively classified cloud pictures for each of the six classes, achieving a success rate of 64% with the SOFM classifier and 65% with the KNN classifier.

A method for "temperature prediction using CNN with VGG–16 architecture" has been described by A. Volokitin,. Van Gool, and others [15]. was presented. Compared to the widely used fully linked layers, temperature forecasts made using the pooling layers produce superior characteristics. The temperature prediction was mostly accurate. The author suggests that because the temperature is not a direct predictor of the time of year, but rather the product of many interactions between the temperature, sunshine, and scene objects, the prediction's accuracy is low at the day level. A weather estimation model based on clustering learning was presented by Jiwan Lee, et al. [16] to assist reduce road risks caused by varying weather conditions. Any setting where many CCTVs are placed and real-time picture collection and storage is possible can use this technique. It is said that using many ROIs (Region of Interest) makes it easier to identify even the smallest variations in the weather. Super-Pixel Segmentation (SPS) approach was proposed by S. Liu, L. Zhang, Z. Zhang, C. Wang, and B. Xiao et al. [17] as an automated cloud identification method. The Kiel [URL] and IapCAS [URL] datasets were utilised to conduct the experiment. Compared to fixed threshold, global threshold, and local threshold interpolation, this method outperforms the others in cloud identification.

"METOSTAT image using Contextual Spatio-Temporal labelling approach" was proposed by C. Papin, P. Bouthemy, and G. Rochard et al. [18]. They made use of local motion-based data in this study, as well as local contextual information and temperature parameters measured over blocks and intensity photographs. The proposed method yields segmentation maps with temporal coherency along the photo series. When the cloud classification problem was formulated to minimize the global energy function, a novel minimization scheme based on the ICM (Iterated Conditional Mode) deterministic iterative relaxation algorithm embedded in a spatially "progressive" scheme was created. The primary goal is to spread knowledge from trustworthy sources to regions that create uncertainty.

M. Mahrooghy, N. H. Younan, V. G. Anantharaj, J. Aanstoos, and S. Yarahmadian et al. [19] proposed a link-based cluster ensemble (LCE) approach to improve precipitation predictions and cloud categorization. Satellite Precipitation Estimation (SPE), which uses Precipitation Estimation from Remotely Sensed Imagery using an Artificial Neural Network Cloud Classification algorithm, is altered by using LCE, which comprises segmenting infrared cloud images into patches, extracting cloud patch features, clustering cloud patches using LCE, and dynamically applying brightness temperature (Tb) and rain-rate relationships derived from GOES-12 satellite images.

S. Dev, Y. H. Lee, S. Winkler, et al. [20] mention "A supervised segmentation framework for ground-based sky/cloud images based on systematic analysis of different colour spaces and components using Partial Least-Squares regression (PLS)". One recommended method for segmenting sky/cloud photographs taken from the ground was a probabilistic approach based on PLS regression. The Hybrid Thresholding Algorithm (HYTA) database and the Singapore Whole Sky Imaging Segmentation (SWIMSEG) database are used to obtain the experimental results. The authors suggest using this approach for high dynamic range photos and anticipate using it to estimate cloud movement and height

as well as categorise clouds into distinct categories. Using polar-orbiting satellite remote sensing observations (MWHS on Chinese meteorological satellite FY-3B, GMI on GPM satellite), geostationary orbit simulations (designed for FY-4M), conventional datasets, and forecasting using the mesoscale Weather Research and Forecasting (WRF) model at temporal and spatial resolution of 15 km and 5 minutes, J. He, H. Chen, S. Zhang, N. Li, et al. [21] examined the evolution process of Hurricane Sandy (October 20–29, 2012). Hurricane Sandy's track and intensity are predicted using the WRF Data Assimilation (WRFDA) model with the radiance previously described. In order to validate the track and intensity of previous forecasts and analysis results, the forecast results are cross-checked with the best track. This suggests a form of fresh observations to improve the track and intensity for future hurricanes. The role of SPARK in weather forecasting is examined by R. Dhoot, S. Agrawal, M. Shushil Kumar, et al. [22] by comparing weather forecasting with and without the usage of Spark cluster for the ARIMA model and Kalman Filter. To categorise the weather, the values predicted by the aforementioned models are sent to the XGBoost Classifier. The last 20 years' worth of data was selected from Kaggle. Three characteristics are first selected by pre-processing: temperature, dew point, and humidity, together with a timestamp. The models mentioned above are used for weather forecasting and XGBoost is used based on these three characteristics. To assess the model quality, the actual values in the dataset have been compared with the forecasted values from weather forecasting. Second order differencing is used into the Kalman filter for forecasting in order to ensure the precision of the model's real-time predictions. The computational time and model quality have been examined using graphical analysis.

The goal of research by E. S. Barus, M. Zarlis, Z. Nasution, Sutarman, et al. [23] is to forecast plant esters blossom using neural network approach. By observing a number of factors, including stalk length, flower quantity, flower diameter, bloom appearance time, leaf count, and sprout count, this study aims to construct a model of plant growth. Every variable is compared to the development of organic fertilizer-fed plants. IoT devices were positioned at every place in the landscape and used to make observations. By observation, it was discovered that after 30 days, plants were found to develop as predicted by the neural network approach in 100 iterations, reaching an iteration point of 9 stable plant models at point 3. This indicated that the optimal circumstances for the plant models

Advanced weather condition monitoring is provided by V. Kadrolli, et al. [24]. We are assembling a weather station with a range of sensors in this project. The EduArm board with Wi-Fi module is the platform in use. The sensor will be able to communicate thanks to this. A wireless hardware module that senses air temperature, wind speed, and temperature. The interface technique each sensor uses and the inputs that are accessible on the EduArm determine which sensors are used for this project. The LPC1768 based micro controller that collects the data from the sensor and uses a serial connection to transfer it to the PC is what powers the EduArm. Creating a protocol, transferring data from the device to the server, and send data over Wi-Fi to server. An application interface that runs on a PC linked to an EduArm and wifi module collects this data. Monitoring the air samples, predicting the weather, and keeping an eye on the efforts involved in alerting the public to impending disasters. To optimise distribution and consumption, H. Duong-Ngoc, et al. [25] propose a DNN-based electrical load forecasting system. In my paper, I used a range of activation functions, such as rectifier linear unit (ReLU), hyperbolic tangent (tanh), and sigmoid. The proposed scheme is administered in accordance with Ho Chi Minh City's necessary offline load requirements. Weather attributes analysis load predictions with DNN. It makes it possible to record the variables influencing the model of power usage. The proposed approach is formalised using three performance metrics: correlation coefficients, normalised root mean square (NRMSE), and absolute percentage error (MAPE). The ultimate outcomes have confirmed the authenticity of FF-DNN and R-DNN.

2.1 Research Gaps: A popular machine learning approach for handling classification and regression issues, such as weather prediction, is the Support Vector Machine (SVM). Although SVM has demonstrated encouraging outcomes in the analysis of meteorological data, a number of research gaps prevent it from reaching its full potential in practical weather forecasting applications. First, the scarcity of high-quality, real-time meteorological datasets is a significant obstacle. For SVM models to train well, the data must be labelled and well-structured. But meteorological information is frequently lacking, erratic, or impacted by noise. To clean and normalize such data before supplying it to the SVM model, more reliable pre-processing methods are required. Second, dimensionality reduction and feature selection continue to be crucial problems.

Numerous factors, including temperature, humidity, wind speed, and atmospheric pressure, are commonly included in weather data. Finding the most important characteristics for prediction is still a difficult research problem. Ineffective feature selection may cause the model to over fit and lose its capacity to generalize. The dynamic and non-linear character of weather systems represents another gap. The temporal relationships and quick variations present in atmospheric conditions may be difficult for kernel-based SVMs to capture, despite their ability to predict non-linear patterns. This suggests that hybrid models that combine SVM with time-series models or deep learning methods like LSTM may be advantageous.

Furthermore, SVMs' computational cost and scalability when used on extensive meteorological datasets are worrisome. Training SVMs on massive datasets can be time-consuming and resource-intensive. Finally, there is a discrepancy between context-aware and localized predictions. The majorities of models are overly generic and fail to sufficiently take seasonal or geographic variances into consideration. Studying region- specific, adaptive SVM models may improve prediction accuracy. To create weather prediction systems with SVM that are more precise, effective, and dependable, these deficiencies must be filled.

The Support Vector Machine (SVM) is a potent algorithm for weather prediction jobs because of its many benefits. Its capacity to efficiently handle high-dimensional data is one of its main advantages; this is important because meteorological datasets frequently contain several features, including temperature, humidity, wind speed, and atmospheric pressure. In order to capture intricate patterns in meteorological data, SVM is especially well-suited for modeling non-linear interactions utilizing kernel functions such as polynomial kernels or the Radial Basis Function (RBF).SVM's resilience to overfitting is another significant benefit, particularly in situations with sparse or noisy data. For real-world weather datasets that could contain errors, this makes it extremely dependable. Additionally, SVM is well-known for its powerful generalization capabilities, which enable it to function effectively on unseen data and improve the precision of upcoming weather forecasts.

SVM can predict a wide range of weather conditions, including sunny, wet, cloudy, and stormy, because it supports both binary and multiclass classification. Unlike deep learning models, which usually require vast amounts of data, it performs well with small to medium-sized datasets. All things considered, SVM's accuracy, adaptability, and flexibility make it a useful tool for creating precise and effective weather forecasting systems across many climatic zones.

2.2 Scientific Merit: Since it affects public safety, transportation, agriculture, and disaster management, weather forecasting has always been a crucial field of study. Conventional numerical weather prediction (NWP) models simulate atmospheric conditions using high-performance computing and equations based on physics. These models, however, can be sluggish, computationally demanding, and initial condition sensitive. Machine Learning (ML) has become a potent substitute or addition to these models in recent years because of its capacity to handle big datasets, identify intricate patterns, and generate precise forecasts quickly. An ML-based weather prediction system is a noteworthy scientific breakthrough with broad applicability.

The foundation of this strategy is the use of historical meteorological data to train forecast models, including temperature, humidity, pressure, wind speed, and satellite images. Weather data contains temporal dependencies and nonlinear interactions that can be modelled by machine learning (ML) algorithms like Random Forests, Support Vector Machines, Gradient Boosting, and particularly deep learning models like Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks. These models can accurately predict future situations by learning from historical patterns. Efficiency is one of the main benefits of science. In contrast to NWP models, which frequently need hours of simulation time, ML models, once trained, can produce forecasts that are almost instantaneous with little computational overhead. This capacity to make predictions in real time is especially helpful when weather conditions like heat waves, cyclones, and thunderstorms are changing quickly. Second, precision and flexibility are important benefits. ML models can learn from current weather trends and anomalies because they can be updated with new data on a regular basis. ML systems can use alternate data sources, such remote sensing or crowd-sourced data, to deliver localized and adaptive forecasts in areas with inadequate meteorological infrastructure.

Third, multi-scale prediction is made possible by the scientific adaptability of ML-based weather systems. In addition to being region-specific, these models can be tailored for medium-, long-, or short-term (daily or hourly) projections. Deep learning models, for instance, have proven effective at detecting cyclones in coastal areas and predicting rainfall in areas that depend on the monsoon.

Furthermore, ML makes it possible to combine many data types—numerical, categorical, spatiotemporal, and visual into a single framework. This is especially helpful for integrating radar data, satellite photos, and inputs from IoT-based sensors. The predictive system becomes more extensive and robust as a result of this integration. Crucially, the creation of such systems advances environmental monitoring, catastrophe preparedness, and climate modeling, all of which benefit the larger scientific community. It backs the Sustainable Development Goals (SDGs), particularly those pertaining to sustainable cities and climate action.

In summary, a machine learning-based weather prediction system signifies a revolution in meteorological research. It uses computational intelligence to improve and supplement conventional forecasting techniques, providing a quicker, more scalable, and maybe more accurate substitute. Its multidisciplinary approach, which combines data science, artificial intelligence and atmospheric science to tackle one of humanity's oldest and most important problems weather prediction is what gives it scientific merit.

2.3 Substantial Technical Details: A machine learning (ML)-based weather prediction system uses both historical and current data to forecast atmospheric conditions by utilizing sophisticated computational algorithms. ML-based systems use statistical and data-driven techniques to find patterns and provide predictions, in contrast to conventional numerical weather prediction (NWP) models, which rely on resolving intricate physical equations. In order to create a reliable machine learning (ML)-driven weather prediction system, this technical exposition describes the system architecture, data pre-processing techniques, model selection strategies, evaluation metrics, and implementation pipeline.

2.3.1. System Architecture

The following are the main parts of an ML-based weather prediction system:

- Data Collection Layer: Compiles meteorological data from a variety of sources, including sensors, weather stations, satellites and APIs like ECMWF, NOAA and Open Weather Map. The raw data is cleaned, normalized, and transformed into a structured format that is appropriate for modeling by the data pre-processing layer.
- Feature Engineering Layer: Generates extra characteristics to improve model performance by extracting Significant features from raw data.
- Modeling Layer: Using historical weather data, ML algorithms are trained to identify trends.
- Evaluation & Deployment Layer: Uses test data to validate the models and incorporates the top-performing model for real-time predictions into an application or service.

2.3.2. Data Sources and Pre-processing

A weather prediction system's input data usually consists of:

- Meteorological factors: precipitation, barometric pressure, wind direction and speed, temperature and humidity.
- Time-series data: Captured over a period of days, months, or years in order to detect cyclical or seasonal patterns.
- Radar and satellite data: Offer inputs for storm tracking, thermal pictures, and cloud cover.
- Geographical information: A region's topography, latitude and altitude.

Among the pre-processing actions are:

- Using imputation or interpolation techniques to deal with missing values.
- Noise filtering: To lessen volatility, smoothing filters such as rolling averages are applied.
- Normalization and scaling: To put all features on a comparable scale, use Z-score standardization or Min-Max scaling.

Aligning several data sources to a standard time interval (such as hourly or daily) is known as time alignment.

2.3.3. Feature Engineering

In order to increase model accuracy, feature engineering is essential. Typical engineered features consist of:

- Lag variables: Historical data on humidity, temperature, and other variables that are used to forecast Future conditions.
- Statistical summaries: Moving averages over time frames, mean, or standard deviation.
- Temporal features: To capture cyclic patterns, use the time of day, day of the week, month, or season.
- Weather indices: These are based on fundamental variables and include the heat index, wind chill, and dew point.

2.3.4. Model Selection and Algorithms

- Various machine learning methods are utilized based on the kind of prediction (classification or regression):

A baseline model for forecasting continuous quantities, such as temperature, is called linear regression.

- Random Forest Regressor/Classifier: An ensemble technique that works well with feature significance and nonlinear relationships. Support Vector Machines (SVM) is used for regression as well as classification (e.g., rain/no rain).
- Gradient Boosting Models: Effective and potent models for structured weather data, such as XGBoost and Light GBM.
- NNs or neural networks:
- Forward Neural Networks (FNNs): Good for making static forecasts. Because they can recall previous inputs, recurrent neural networks (RNNs) and long short-term memory (LSTM) are perfect for sequential time-series forecasting.
- Convolutional Neural Networks (CNNs): Used to extract spatial features from radar maps or satellite photos.
- Transformers: New time-series forecasting models that are becoming popular because of their attention mechanisms and parallel processing

2.3.5. Training and Validation

During training, the dataset is divided into test, validation, and training sets (usually a 70-15-15 split). To make sure the model performs effectively when applied to new data, cross-validation more specifically, time series cross validation is employed. For temporal datasets, the walk-forward validation method is also frequently used. Model parameters like learning rate, tree depth (for boosting models), or number of neurons/layers (for neural networks) are refined through hyper parameter tuning utilizing grid search or Bayesian optimization.

2.3.6. Model Evaluation Metrics

Common metrics for regression tasks (such forecasting temperature and humidity) include:

- Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE)
- R2 Score (Determination Coefficient)

Regarding categorization tasks (such as storm detection and rain/no rain):

- Precision
- F1-Score, Precision and Recall
- The Confusion Matrix
- AUC-ROC o are under the ROC curve

To clarify model predictions and boost confidence, model interpretability methods like SHAP (Shapley Additive explanations) and LIME (Local Interpretable Model-Agnostic Explanations) can be employed.

2.3.7. Deployment and Real-Time Prediction

After being verified, the model is serialized (for example, with Pickle, ONNX, or Tensor Flow Saved Model) and introduced into a mobile or web application. Frameworks like Flask or Fast API can be used to create a REST full API that serves predictions. Message queues (Kafka, MQTT) facilitate real-time data ingestion from APIs or IoT devices, while cloud platforms such as AWS or Azure offer scalable infrastructure for ongoing forecasting.

3.METHODOLOGY

The machine learning-based weather interactive prediction system utilized in this part is depicted in block diagram form in Fig. 1. The ground-based picture datasets will be the ones taken into consideration in this case. the picture databases derived from the ground, like Sky Finder. Pre-processing is a term used to describe actions performed on pictures at the lowest possible abstraction level. Intensity pictures are used for both input and output pre-processing. The original information acquired by the image sensor is comparable to the kind of these iconic pictures; A matrix of image function values is frequently used to depict an intensity image. (brightness). Pre-processing seeks to enhance certain visual characteristics or suppress undesired distortions in order to improve the image data. It's possible that noise was introduced into the photos from the input data. Prior to extracting any features, denoising these pictures becomes essential. The terms absolute and relative humidity are sometimes used interchangeably by meteorologists when discussing humidity. Absolute humidity is defined as the ratio of water vapour to dry air in a specific

volume of air at a specific temperature. Temperature affects how much water vapour the air can hold. The proportion between the current and greatest absolute humidity levels, which is dependent on the air temperature at that moment, is known as relative humidity. The most common word used by weather forecasters is relative humidity.

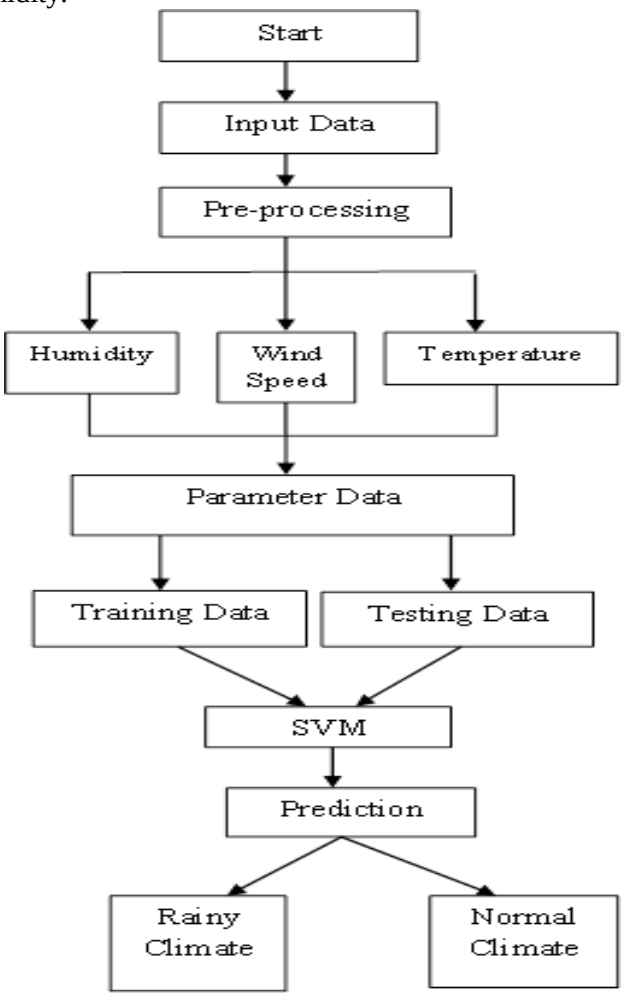

Fig.1: Block Diagram of Weather Interactive Predictive System by Using Machine Learning

Our everyday weather is a result of changes in broad-scale wind circulation patterns. For the purpose of assessing the condition of the atmosphere at certain periods and locations on Earth, wind speed and direction observations are crucial, as are observations of other components like temperature and moisture. The data for the prediction parameter is acquired. Rainfall predictions are made using the Support Vector Machine (SVM) algorithm. The data is first normalized before being divided into training and test sets. Eventually, until the model is optimized for rainfall prediction, the parameters for the training and testing data are initialized. Separate the outcome into test and training sets. SVM classifiers are used to forecast the presence of rainy and normal climates.

4.RESULT ANALYSIS

This section shows the results of a machine learning-based weather interactive prediction system analysis.

Table 4.1 Measurement analysis

Parameters	SVM	NB	LR
Accuracy	84.19	81.3	83.8
Recall	88.2	82.3	42.6
F1 Score	91.9	85.7	53.2

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 6, 2025
https://theaspd.com/index.php

A comparison graph of accuracy between SVM, NB, and logistic regression is shown in Figure 2. for weather prediction.

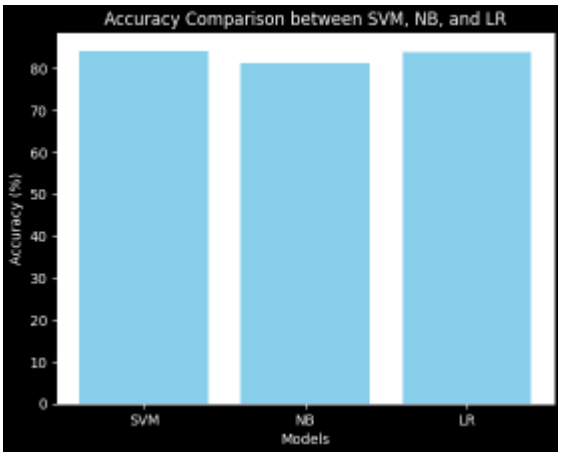

Fig.2: Accuracy Comparison Graph

A recall comparison graph comparing SVM, NB, and logistic regression can be seen in Figure.3 SVM's recall value is greater.

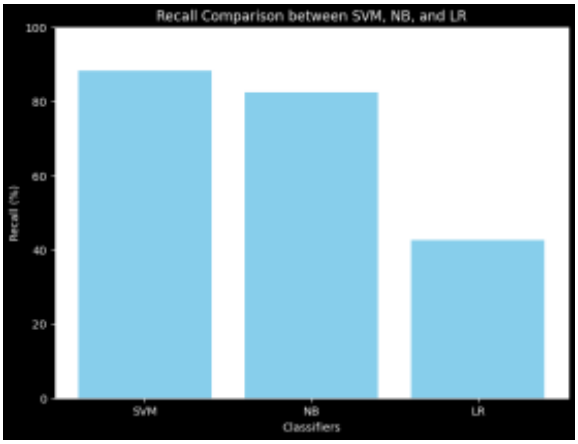

Fig.3: Recall Comparison Graph

For weather prediction an F1-Score comparison graph comparing SVM, NB, and logistic regression is shown in Figure. 4

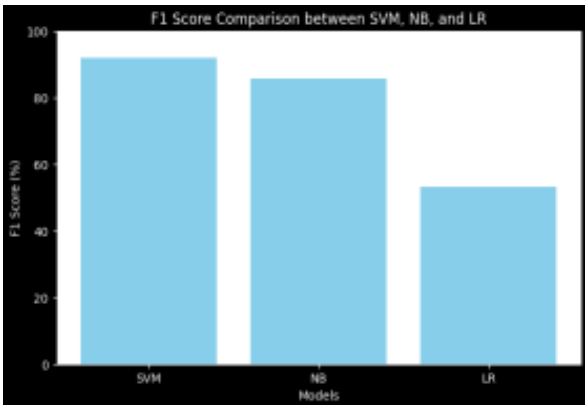

Fig.4: F1-Score Comparison Graph

4.1 Limitations: Although incorporating Naive Bayes (NB), Logistic Regression (LR), and Support Vector Machine (SVM) into weather prediction systems provides a potent method for data-driven forecasting, a number of restrictions impact the models' accuracy, performance, and practicality. Researchers and

developers hoping to improve the accuracy of machine learning-based weather prediction must be aware of these constraints.

1. Dependency on Data Availability and Quality: For all three algorithms—NB, LR, and SVM—to function well, clean, organized, and comprehensive datasets are necessary. However, due to sensor failures or restricted geographic coverage, databases frequently have missing values, measurement errors, and inconsistencies in real-world meteorological applications. SVM performance deteriorates with noisy or irrelevant features, but Naive Bayes is very sensitive to data imbalance. If the dataset is not appropriately pre-processed, logistic regression, which assumes linear relationships, may yield inaccurate findings.

2. The incapacity to record time-dependent relationships. By its very nature, weather is a time-dependent process that is impacted by successive variations in atmospheric variables. However, because they are not time-series models, NB, LR, and SVM find it difficult to identify long-term patterns and temporal dependencies in data. Their capacity to forecast abrupt weather changes or uncommon occurrences like storms or cyclones is limited since they evaluate each input separately and do not learn from previous sequences.

3. Premises and Simplified Concepts Every algorithm is predicated on the following:

- In meteorological data, when temperature, humidity, and pressure are heavily connected, naive Bayes Assumes feature independence which is rarely the case.
- SVM can model non-linear data with kernels, but selecting the appropriate kernel and adjusting hyper parameters can be computationally demanding and dataset specific.
- Logistic regression assumes a linear relationship between features and the log odds of the output, which frequently oversimplifies complex atmospheric interactions.

4. Scalability and Computational Complexity SVM can become resource-intensive when used on large-scale datasets, particularly when non-linear kernels are used, whereas NB and LR are computationally efficient. It can be time-consuming and memory-intensive to train SVM models on high-dimensional meteorological data with thousands of records. This restricts the system's real-time usefulness, particularly in areas where conditions change quickly.

5. Overfitting and Generalization Models for predicting the weather must be able to generalize effectively over many seasons, years, and geographic locations. But overfitting is a frequent problem:

- When characteristics are highly linked, NB may over fit.
- If non-linearity is ignored, LR might be under fit.
- In high-dimensional spaces, SVM may over fit if regularization is not managed appropriately.

6. Inability to Interpret in Complicated Situations While LR's linear coefficients make it interpretable, non-technical users may find SVM (particularly with complex kernels) and NB (with probabilistic outputs) less comprehensible. Adoption in vital fields like catastrophe management, where comprehending the reasoning behind forecasts is essential, may be hampered by this lack of openness. When employed separately for weather prediction, NB, LR, and SVM have significant drawbacks despite their advantages. The necessity for hybrid models, better data processing, and the integration of sophisticated time-series or deep learning approaches are highlighted by their shortcomings in managing temporal data, feature dependencies, and computational needs.

5. CONCLUSION

This section has covered the interactive prediction system that uses machine learning to anticipate weather in detail. Numerous weather predicting techniques, such as quantitative forecasting, weather map forecasting, and numerical weather forecasting, have developed as technology advances. In order to protect public health, this predictive technology is essential in forecasting weather patterns and sending out alerts about changes in air quality. This prediction method effectively uses machine learning to meet the difficulty of creating a highly accurate and efficient climate forecasting model. In addition to providing accurate forecasts, the deployment of this kind of technology helps prevent property damage and perhaps save lives by providing timely alerts for natural catastrophes like floods. As can be seen, the suggested Support Vector Machine (SVM) model performs better than Naïve Bayes (NB) in terms of Accuracy, Recall, and F1 Score, demonstrating its superiority. Similar to how it does in this predictive system, Logistic Regression (LR) is equally essential, offering insightful information that enhances the overall efficacy of weather forecasting.

Future Work: A solid basis for categorizing and predicting weather patterns is demonstrated by the present weather prediction system, which uses Naive Bayes, Logistic Regression and Support Vector

Machine models. Nonetheless, there is a great deal of room for improvement and growth in further research. First, by using more sophisticated machine learning methods like ensemble models (Random Forest, Gradient Boosting) and deep learning architectures (e.g., LSTM and CNN for time-series data), the system's accuracy and resilience can be increased. Compared to conventional techniques, these models are better able to identify intricate patterns and connections in big datasets. Second, by utilizing live weather data streams via APIs from sources such as NOAA or OpenWeatherMap, the system can be expanded to provide real-time forecasts. The system would become more useful and dynamic for real-world applications as a result. Incorporating geospatial data, satellite imaging, and climatic factors like air pressure, humidity, wind speed, and solar radiation can also improve feature engineering. Forecast accuracy could be greatly increased by using these factors. Incorporating explainable AI (XAI) approaches is another crucial topic for future research. Users and meteorologists can have a better understanding of the forecast's reasoning by making model forecasts easier to interpret, which will boost the system's dependability and credibility. Lastly, the system's adaptability would be enhanced by extending its coverage to several geographic areas and modifying the models to account for regional climates. Either region-specific model training or transfer learning can do this.

The integration of more complex algorithms, real-time data, new characteristics, and explainability will be crucial paths for future study and practical deployment, even if the existing system offers a foundation for weather prediction employing NB, LR, and SVM.

REFERENCES

[1] Jain G, Mallick B. A review on weather forecasting techniques. International Journal of Advanced Research in Computer and Communication Engineering. 2016 Dec;5(12):177-80.

[2] Saima H, Jaafar J, Belhaouari S, Jillani TA. Intelligent methods for weather forecasting: A review. In2011 national postgraduate conference 2011 Sep 19 (pp. 1-6). IEEE.

[3] Didal VK, Brijbhooshan AT, Choudhary K. Weather Forecasting in India: A Review. Int. J. Curr. Microbiol. App. Sci. 2017 Nov;6(11):577-90.

[4] Chu WT, Ho KC, Borji A. Visual weather temperature prediction. In2018 IEEE Winter Conference on Applications of Computer Vision (WACV) 2018 Mar 12 (pp. 234-241). IEEE.

[5] Chu WT, Zheng XY, Ding DS. Camera as weather sensor: Estimating weather information from single images. Journal of Visual Communication and Image Representation. 2017 Jul 1;46:233-49.

[6] Lu C, Lin D, Jia J, Tang CK. Two-class weather classification. InProceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2014 (pp. 3718-3725).

[7] Christodoulou CI, Michaelides SC, Pattichis CS. Multifeature texture analysis for the classification of clouds in satellite imagery. IEEE transactions on geoscience and remote sensing. 2003 Nov 17;41(11):2662-8.

[8] Bautu A, Bautu E. Meteorological data analysis and prediction by means of genetic programming. InProceedings of the 5th Workshop on Mathematical Modeling of Environmental and Life Sciences Problems Constanta, Romania 2006 Sep (pp. 35-42).

[9] Esfandeh S, Sedighizadeh M. Meteorological data study and forecasting using particle swarm optimization algorithm. World Academy of Science, Engineering and Technology. 2011 Mar 1;59:2117-9.

[10] Mohapatra SK, Upadhyay A, Gola C. Rainfall prediction based on 100 years of meteorological data. In2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN) 2017 Oct 12 (pp. 162-166). IEEE.

[11] Chu WT, Ho KC, Borji A. Visual weather temperature prediction. In2018 IEEE Winter Conference on Applications of Computer Vision (WACV) 2018 Mar 12 (pp. 234-241). IEEE..

[12] Chu WT, Zheng XY, Ding DS. Camera as weather sensor: Estimating weather information from single images. Journal of Visual Communication and Image Representation. 2017 Jul 1;46:233-49.

[13] Lu C, Lin D, Jia J, Tang CK. Two-class weather classification. InProceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2014 (pp. 3718-3725).

[14] Christodoulou CI, Michaelides SC, Pattichis CS. Multifeature texture analysis for the classification of clouds in satellite imagery. IEEE transactions on geoscience and remote sensing. 2003 Nov 17;41(11):2662-8.

[15] Volokitin A, Timofte R, Van Gool L. Deep features or not: Temperature and time prediction in outdoor scenes. InProceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops 2016 (pp. 63-71)..

[16] Lee J, Hong B, Jung S, Chang V. Clustering learning model of CCTV image pattern for producing road hazard meteorological information. Future Generation Computer Systems. 2018 Sep 1;86:1338-50.

[17] Liu S, Zhang L, Zhang Z, Wang C, Xiao B. Automatic cloud detection for all-sky images using superpixel segmentation. IEEE Geoscience and Remote Sensing Letters. 2014 Aug 8;12(2):354-8.

[18] Papin C, Bouthemy P, Rochard G. Unsupervised segmentation of low clouds from infrared METEOSAT images based on a contextual spatio-temporal labeling approach. IEEE Transactions on Geoscience and Remote Sensing. 2002 Jan;40(1):104-14.

[19]Mahrooghy M, Younan NH, Anantharaj VG, Aanstoos J, Yarahmadian S. On the use of a cluster ensemble cloud classification technique in satellite precipitation estimation. IEEE journal of selected topics in applied earth observations and remote sensing. 2012 Jul 10;5(5):1356-63.

[20] Dev S, Lee YH, Winkler S. Color-based segmentation of sky/cloud images from ground-based cameras. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2016 May 18;10(1):231-42..

[21] He J, Chen H, Zhang S, Li N. Observations and Forcasting Analysis of Hurricane Sandy Using Satellite Microwave Remote Sensing. InIGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium 2019 Jul 28 (pp. 7552-7555). IEEE..
[22] Dhoot R, Agrawal S, Kumar MS. Implementation and analysis of arima model and kalman filter for weather forcasting in spark computing environment. In2019 3rd international conference on computing and communications technologies (ICCCT) 2019 Feb 21 (pp. 105-112). IEEE.
[23] Barus ES, Zarlis M, Nasution Z. Forcasting plant growth using neural network time series. In2019 International Conference of Computer Science and Information Technology (ICoSNIKOM) 2019 Nov 28 (pp. 1-4). IEEE..
[24] Kadrolli V, Melge K, Kamane Y, Gaikwad P, Sangle R. Portable Weather Station Using GUI. In2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) 2019 Mar 27 (pp. 806-811). IEEE..
[25] Duong-Ngoc, Hung, Hoan Nguyen-Thanh, and Tam Nguyen-Minh. "Short term load forcast using deep learning." 2019 Innovations in Power and Advanced Computing Technologies (i-PACT). Vol. 1. IEEE, 2019.