e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 5, Issue 8, August 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.54

6381 907 438 6381 907 438 ijmrset@gmail.com www.ijmrset.com

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.54| Monthly, Peer Reviewed & Referred Journal|

| Volume 5, Issue 8, August 2022 |

| DOI:10.15680/IJMRSET.2022.0508010 |

Leveraging Machine Learning for Real-Time Short-Term Snowfall Forecasting Using Multi-Source Atmospheric and Terrain Data Integration

Vimal Raja Gopinathan

Principal Consultant, Oracle Financial Services Software Ltd., Bengaluru, India

ABSTRACT: This paper presents a machine learning-based framework for real-time short-term snowfall forecasting by integrating atmospheric and topographic data. The model uses real-time meteorological data such as temperature, humidity, and pressure, along with terrain data like elevation and land cover, to predict snowfall occurrence within a 12-hour forecast window. Random Forest (RF) and Support Vector Machine (SVM) models are employed to process these multi-source inputs, demonstrating a significant improvement in prediction accuracy over traditional methods. Experimental results show that the RF model achieved a root mean square error (RMSE) of 3.2 cm, while traditional regression models achieved an RMSE of 5.1 cm. The proposed framework provides a more reliable and accurate tool for real-time snowfall prediction, especially in mountainous regions, and can be deployed for operational weather forecasting applications.

KEYWORDS: Machine Learning, Snowfall Prediction, Atmospheric Data, Topographic Data, Real-Time Forecasting, Random Forest, Support Vector Machines

I. INTRODUCTION

Snowfall prediction is a vital aspect of weather forecasting, particularly in regions prone to heavy snowstorms. Accurate and timely predictions are essential for public safety, transportation, and disaster management. Traditional meteorological models rely heavily on numerical simulations and observational data, which often lack the precision necessary for short-term forecasts, particularly in complex terrains.

Machine learning (ML) has shown promise in improving forecasting accuracy by leveraging large datasets and identifying hidden patterns in complex systems. Recent studies have explored ML applications in weather prediction, but few have integrated both atmospheric and topographic data to address snowfall prediction in real-time, especially in mountainous areas.

This paper proposes a novel ML-based framework that integrates atmospheric conditions (temperature, humidity, wind speed) and topographic features (elevation, land cover, slope) to predict snowfall within a 12-hour window. The research leverages Random Forest (RF) and Support Vector Machine (SVM) models to analyze these multi-source inputs, providing enhanced prediction capabilities.

II. LITERATURE SURVEY

In recent years, machine learning (ML) has emerged as a powerful tool for weather forecasting, including the prediction of snowfall. Researchers have explored various machine learning techniques to improve the accuracy and timeliness of weather predictions, especially for short-term forecasts. Traditional forecasting methods, including numerical weather prediction (NWP) models, rely on mathematical simulations based on physical laws. However, these models often struggle to accurately predict snowstorms, especially in regions with complex topography. Machine learning methods, in contrast, are well-suited for capturing non-linear relationships in large, complex datasets, making them ideal for improving prediction accuracy.

Machine Learning in Snowfall Prediction

Several studies have employed machine learning techniques for snowfall prediction, often focusing on atmospheric variables such as temperature, humidity, and wind speed. A notable example is the work by Xie et al. (2018), who used artificial neural networks (ANNs) to predict snowfall in the Midwest U.S. Their model demonstrated that machine learning methods could significantly improve prediction accuracy when compared to traditional statistical methods.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.54| Monthly, Peer Reviewed & Referred Journal|

| Volume 5, Issue 8, August 2022 |

| DOI:10.15680/IJMRSET.2022.0508010 |

However, their approach relied solely on atmospheric data, ignoring the influence of terrain features like elevation, slope, and land cover, which can significantly affect snowfall patterns, particularly in mountainous areas.

Incorporating Terrain Data

The incorporation of topographic data has been explored by several researchers to improve weather prediction models. For example, Smith et al. (2019) proposed a machine learning approach to integrate both atmospheric and terrain data to improve weather forecasting accuracy. Their model used digital elevation models (DEMs) and land cover data alongside traditional meteorological features to enhance the prediction of temperature and precipitation patterns. While their work demonstrated the benefits of incorporating terrain data, their focus was more general weather forecasting, rather than specifically addressing snowfall prediction. Additionally, the study did not focus on real-time forecasting or short-term prediction windows, which is a critical challenge for operational applications.

Real-Time Snowfall Prediction

Real-time snowfall prediction has been a major challenge, particularly in regions with rapidly changing weather conditions. Previous studies have primarily focused on offline prediction models that require substantial amounts of historical data to train. However, real-time prediction is necessary for timely snowstorm warnings and snow management operations. Recent advancements in real-time data processing, coupled with machine learning, have enabled more accurate and timely snowfall predictions. For example, recent work by Zhang et al. (2020) applied Random Forest (RF) and Gradient Boosting Machines (GBMs) for real-time snowfall predictions in the Rocky Mountains. Their models incorporated both atmospheric data and real-time snowfall measurements, demonstrating improvements over traditional weather models. However, their model focused primarily on atmospheric data and lacked integration with detailed topographic features, which are crucial in mountainous regions.

Multi-Source Data Integration for Snowfall Forecasting

There have been limited studies on the integration of multi-source data (atmospheric and topographic) for snowfall prediction, particularly using machine learning methods. One of the closest approaches was presented by Johnson et al. (2021), who explored the fusion of satellite-derived snow cover data with atmospheric observations to predict snowfall events in the Alps. While this study provided insights into data fusion techniques, it did not utilize machine learning models for prediction, and its focus was primarily on post-event snow analysis rather than forecasting.

This paper builds upon these studies by proposing a framework that integrates atmospheric and topographic data using advanced machine learning techniques, such as Random Forest and Support Vector Machine (SVM), for short-term, real-time snowfall prediction. By combining these diverse datasets, our approach aims to address the gap in the literature by providing a model that not only accounts for meteorological conditions but also integrates the influence of terrain on snowfall, especially in mountainous regions.

III. METHODOLOGY

3.1 Data Collection

The dataset consists of meteorological and terrain data from multiple sources:

- **Atmospheric Data**: Temperature, pressure, humidity, wind speed, and wind direction were collected from regional weather stations and global meteorological databases.
- **Topographic Data**: Digital Elevation Models (DEMs) provided elevation, slope, and land cover data.

The dataset spans 3 winter seasons (2015-2019), covering mountainous and flat terrains in the Rocky Mountains. Data was collected hourly, with features including:

- Temperature (°C)
- Humidity (%)
- Wind Speed (m/s)
- Elevation (meters)
- Slope (degrees)
- Land Cover (categorical: forest, urban, etc.)

The data was preprocessed to fill missing values using linear interpolation, normalize the data to a range of [0,1], and split into training (70%), validation (15%), and test (15%) sets.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.54| Monthly, Peer Reviewed & Referred Journal|

| Volume 5, Issue 8, August 2022 |

| DOI:10.15680/IJMRSET.2022.0508010 |

3.2 Machine Learning Models

We chose Random Forest (RF) and Support Vector Machine (SVM) models due to their proven success in handling high-dimensional data and their ability to capture complex non-linear relationships.

- **Random Forest (RF)**: An ensemble learning method that combines multiple decision trees to improve prediction accuracy. RF is effective in reducing overfitting and provides insights into feature importance.
- **Support Vector Machine (SVM)**: A supervised learning algorithm that classifies data by finding a hyperplane that best separates different classes. SVM is well-suited for high-dimensional feature spaces and small datasets.

Both models were trained on the atmospheric and topographic features to predict the snowfall occurrence (binary classification: snowfall/no snowfall) and the snowfall intensity (measured in centimeters).

3.3 Model Training and Validation

We employed 5-fold cross-validation for model training, optimizing hyperparameters such as the number of trees (RF) and the kernel type (SVM). The models were evaluated based on two metrics:

- **Root Mean Square Error (RMSE)**: Measures the model's accuracy in predicting snowfall intensity.
- **F1 Score**: Evaluates the classification performance, considering both precision and recall for snowfall predictions.

IV. EXPERIMENTAL RESULTS

4.1 Performance Evaluation

The models were evaluated on the test dataset, which consisted of 1,000 hourly snowfall observations. The experimental results are summarized in Table 1.

Model	RMSE (cm)	F1 Score (Snowfall)
Random Forest (RF)	3.2	0.91
Support Vector Machine (SVM)	4.0	0.87
Traditional Regression	5.1	0.82

From the results, it is clear that the Random Forest model outperforms both the Support Vector Machine and traditional regression methods in terms of both RMSE and F1 Score.

4.2 Model Comparison

The RF model's RMSE of 3.2 cm indicates a high degree of accuracy in predicting snowfall intensity. In comparison, traditional regression models, which rely only on atmospheric data, show a significantly higher RMSE of 5.1 cm. The integration of terrain data in the RF model allows for better handling of complex topographic variations, which is crucial for snowfall prediction in mountainous areas.

4.3 Real-Time Prediction Evaluation

The real-time prediction accuracy was tested using a rolling window approach. The model successfully predicted snowfall events in real-time, with minimal latency, making it suitable for operational forecasting. The results showed that the RF model provided predictions with an average lead time of 12 hours, with an F1 score of 0.88 for real-time predictions.

V. CONCLUSION

The integration of atmospheric and topographic data through machine learning has shown significant promise in improving short-term snowfall predictions. The Random Forest model, in particular, demonstrated a strong capability in capturing the complex relationships between the variables and accurately predicting snowfall intensity. The real-time prediction capability further enhances the model's applicability in operational forecasting, providing valuable insights for decision-making in snow management and public safety.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.54| Monthly, Peer Reviewed & Referred Journal|

| Volume 5, Issue 8, August 2022 |

| DOI:10.15680/IJMRSET.2022.0508010 |

However, there are several areas for future work. First, incorporating more granular data, such as real-time satellite imagery or radar-based snow accumulation measurements, could improve prediction accuracy. Additionally, deep learning models, such as convolutional neural networks (CNNs), could be explored to handle the spatial data from terrain more effectively.

This paper presents a machine learning-based framework that integrates atmospheric and topographic data for short-term snowfall prediction. The results demonstrate that the RF model outperforms traditional methods, achieving higher accuracy and real-time prediction capabilities. The proposed framework holds potential for deployment in operational weather forecasting systems, particularly in mountainous regions. Future work will focus on incorporating more data sources, enhancing real-time data assimilation, and exploring advanced deep learning techniques to improve prediction accuracy and robustness.

REFERENCES

[1] Xie, H., Zhang, X., and Wang, Y., "Application of Artificial Neural Networks for Snowfall Prediction in the Midwest U.S.," *Journal of Atmospheric Sciences*, vol. 79, no. 4, pp. 1345-1360, Apr. 2018.

[2] Smith, J., Brown, T., and Lee, K., "Integrating Terrain Data for Weather Forecasting Using Machine Learning," *Proceedings of the International Conference on Machine Learning*, pp. 88-92, Jun. 2019.

[3] Zhang, L., Chen, Q., and Zhao, R., "Real-Time Snowfall Prediction Using Random Forest and Gradient Boosting Machines in the Rocky Mountains," *Journal of Hydrometeorology*, vol. 21, no. 10, pp. 2678-2694, Oct. 2020.

[4] Johnson, A., Lee, S., and Carter, P., "Fusion of Satellite Snow Cover Data with Atmospheric Observations for Snowfall Event Prediction," *Journal of Remote Sensing and Atmospheric Research*, vol. 45, no. 2, pp. 125-139, Feb. 2021.

[5] Xie, Q., and Zhao, F., "Application of Support Vector Machines for Snowfall Prediction in Complex Terrain," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 11, pp. 5821-5829, Nov. 2020.

[6] Jones, R., et al., "Digital Elevation Models and Their Application in Snowfall Prediction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 10, pp. 5000-5012, Oct. 2020.

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com