The Intelligent System Review

ISSN: 3106-7840 | 3106-7832
Volume 3, Issue 2, 2026

SUNSPOT ANALYSIS AND SPACE WEATHER PREDICTION USING A 2D CNN DEEP LEARNING MODEL

Farman Ali

Department of Computer Science, Bahria University, Islamabad, Pakistan

farman335@gmail.com

Keywords
Time-series analysis; cloud computing; deep learning

Article History
Received: 28 April 2026
Accepted: 15 June 2026
Published: 30 June 2026

Corresponding Author: *
Farman Ali

Abstract
Accurate sunspot prediction is vital for space weather forecasting to protect space-based infrastructure. This study leverages deep learning models—LSTM, GRU, and 2D Convolutional Neural Networks (2D-CNN)—within a cloud computing framework to enhance prediction accuracy and scalability. Using time-series sunspot data, LSTM and GRU capture sequential patterns, while optimization techniques reduce complexity and overfitting. Experimental results show 2D-CNN achieves the highest accuracy at 99.39%, with strong precision and recall, highlighting its superior spatial feature extraction. GRU outperforms LSTM in sequential data handling. These results demonstrate the effectiveness of deep learning, especially 2D-CNNs, for robust sunspot forecasting.

INTRODUCTION

Accurate space weather forecasting is essential for protecting space-dependent infrastructure, such as satellites, communication networks, and power grids, from disruptions caused by solar activity[1]. Sunspots are good indicators of solar activity and are important for predicting space weather events. Normal forecasting methods can use a lot of compute and not always provide accurate predictions, leading researchers to turn to new machine learning and deep learning methods[1].

Recent development in machine learning and deep learning has generated exciting results in the field of weather and climate-based modeling. One example is research that explored the use of neural network emulators to enhance the speed of modern-day weather forecasting systems. Other research explored the use of hybrid deep learning models derived from Convolutional Neural Networks (CNN's) and Long Short-Term Memory (LSTM) to predict meteorological parameters. Additional research has also used a decision tree classifier to improve short-range weather forecasting accuracy. Other research has implemented ensemble models, established bidirectional LSTM, and used extreme gradient boosting in a myriad of weather-related applications including precipitation forecasts, flood predictions, and plume modeling of a nuclear accident. Future innovations will likely arise from further research using AI and cloud computing to enhance the models used in weather forecasting[2].

Nonetheless, optimizing model complexity and tuning hyperparameters remains a challenging, and still mostly unexplored topic in time-series forecasting of sunspot activity. Many approaches that have been employed in the past and will still be applicable in the present do not have specially tailored optimizations of hyperparameters for sunspot prediction and will likely lack the precision needed for practical predictions. All of this serves to limit the utility of some traditional approaches for forecasting space weather using sunspots. This paper hopes to alter this reality with a deep learning approach to this challenge. To that end, the paper explores how AI-based models, using cloud-

ISSN: 3106-7840 | 3106-7832
Volume 3, Issue 2, 2026

based optimization methods, can increase the accuracy and scalability of sunspot prediction[3].

The remainder of this paper is organized as follows: Section 2 discusses related work in deep learning applications for weather and space weather forecasting. Section 3 presents the proposed methodology, detailing the dataset, model architecture, and optimization techniques. Section 4 evaluates experimental results, comparing the performance of deep learning models with traditional machine learning approaches. Finally, Section 5 concludes with key findings and future research directions in enhancing space weather forecasting through AI-driven methodologies.

1. Related work

Advancements in deep learning and machine learning techniques have significantly improved space weather forecasting and meteorological predictions. Various studies have explored different datasets, algorithms, and hybrid models to enhance accuracy, computational efficiency, and forecasting capabilities. For instance, Kaifeng et al. (2023) employed hierarchical temporal aggregation and cyclone tracking on the ERA5 dataset using Mean Sea Level Pressure (MSLP) techniques. Their study demonstrated high forecasting accuracy; however, reliance on reanalysis data introduced temporal inconsistencies and underestimated extreme events[4]. Similarly, Chantry et al. (2021) developed a hybrid neural network (HNN) model on the NOGWD dataset, utilizing GPU acceleration and normalization techniques. While their approach improved model performance, it suffered from slow CPU transmission and inaccurate orographic dynamics[5].

Tal Ben et al. (2022) optimized the Finite Volume Cubed-Sphere Dynamical Core (FV3AD) using DaCe (GPU code generation) and GT4Py (abstraction framework) to enhance forecasting speed and reduce computational costs. Although their method boosted FV3 model performance, it required hardware-specific optimizations and adjustments in transfer settings for broader applicability[6]. Sercan et al. (2022) introduced a hybrid CNN-LSTM model for weather forecasting using historical-hourly-weather-data. Their approach effectively captured sequential dependencies, reducing Root Mean Square Error (RMSE) and improving accuracy, but required careful hyperparameter tuning and had limited scalability[7]. Sudhan et al. (2021) leveraged the MERRA database to develop a machine learning model combining C5.0 decision trees and K-means clustering. Their study achieved higher accuracy compared to traditional methods but emphasized future improvements through deep learning integration[8]. Nawaf et al. (2022) explored solar flare forecasting using LSTM, bidirectional LSTM, CNN-LSTM, random forest, and gradient-boosted trees. Their results showed that univariate models outperformed multivariate models in reducing forecasting errors, though their data spanned only 1.5 years, limiting long-term prediction reliability[9].

Ali Ayoub et al. (2024) applied Recurrent Neural Networks (RNNs) and XGBoost for sunspot forecasting, achieving an R^2 score of 0.84. Despite promising results, challenges related to high computational demand and scalability were noted[10]. Luka et al. (2023) combined linear (ARIMA/SARIMA) and nonlinear (GEP, SVR, and GMDH) models for weather data analysis. Their hybrid approach improved accuracy using techniques such as inverse variance weighting and MSE variance criteria, but faced complexity challenges in model weight adjustments[11].

Carmen Calvo et al. (2024) conducted a machine learning-based weather prediction study using Random Forest (RF), Stochastic Gradient Descent (SGD), Decision Trees (DT), and AdaBoost (AB) on the Zenodo dataset. RF demonstrated superior predictive strength and reduced uncertainty, yet challenges related to scaling and processing power limitations were evident[12]. Lastly, M.S. Hossain et al. (2024) compared Vector Autoregressive Model with Exogenous Variables (VARX) and Deep Neural Networks (DNN) for weather forecasting using GRIB data. While VARX provided accurate temperature forecasts and DNN improved wind speed predictions, both models required frequent retraining and emulator optimizations[13].

2. Materials and Methods

2.1. Datasets

The dataset contains two dominant features. The first is "Date," which offers a month-end date for which an average monthly number of sunspots was determined. This variable is in date format. The second characteristic is "Monthly Mean Total Sunspot Number," which provides an average amount of total sunspots measured during the month. This number is provided as a decimal, which provides the monthly average amount of total sunspot counts by observational data.

2.1.1. Data Collection: In this project, data collection entails researching historic solar data, to count sunspots, from a credible source like Kaggle. These datasets will be preprocessed to facilitate the handling of missing values, normalization, and the preparation of a time-series sequence suitable for machine learning. To improve model accuracy in future stages, it may be advantageous to partition the data into segments based on solar cycles or other relevant time periods. Model performance on predictions of solar activity is evaluated and enhanced using training, validation, and testing datasets.

2.1.2. Data Preprocessing: Key stages in the data preparation stage for this sunspot activity forecasting study include cleaning the raw data and organizing the data (for example addressing missing values, normalizing the data to ensure consistent scale across features, and possibly reducing noise.). transforming the data to provide a consistent range for the features.

2.1.3. Data Labeling: This task assigns class labels or values to historical records of solar activity, such as sunspot counts or solar flare intensity levels, to facilitate supervised learning with data labeling. By labeling each data point by kind, intensity, or frequency of exercise, the model acquires the potential to differentiate groups (e.g., flare classes and/or sunspot cycle phases).

2.2. Deep Learning Frameworks for Model Training

2.2.1. Long Short-Term Memory

Long Short-Term Memory (LSTM) models solve long-term dependencies in time-series data by regulating information flow through input, forget, and output gates. This avoids the vanishing gradient problem and keeps important past inputs. Their ability to find complicated long-term trends makes them a great fit for forecasting solar activity, as well as very accurate for other tasks such as sunspot prediction[14]. However, due to their high number of parameters, overfitting is much more likely when using them requiring longer training times particularly with small datasets. Long Short-Term Memory are frequently used for long term solar activity predictions - despite their high data requirements and processing time[15].

2.2.2. Gated Recurrent Unit

The Gated Recurrent Unit (GRU) is a minimal version of Long Short-Term Memory that reduces parameters and increases processing speed by merging the input gate and forget gate into an update gate. It has a straightforward structure which is effective for moderately complicated tasks, improves the speed of training, and limits the chances of overfitting[14]. Nonetheless, since Gated Recurrent Units do not offer the granular gate control of Long Short-Term Memory and may be limited in processing long sequences with limited memory capacity, they might not be as flexible in recognizing complex time-series patterns. Gated Recurrent Unit can help to predict sunspot activity when computational capacity is lower, particularly for shorter-term sunspot cycles or simplified categories of solar events[15].

2.2.3. Two-Dimensional Convolutional Neural Networks

While Long Short-Term Memory (LSTM) models are frequently employed to predict solar activity and are effective at modeling long-term dependencies in time series data, they have some limitations, including a potential to overfit, a more costly computational requirement, and a longer training time, especially when used for small datasets. Fortunately, 2D-CNN [16] provide a more efficient alternative by applying convolutional methods to extract spatial and temporal information from the given input data. In the context of sunspot prediction, 2D-CNNs achieve higher accuracy and improved computation efficiency than LSTM because they can learn local patterns and hierarchical structures. The proposed framework using 2D-CNNs is expected to improve performance in predicting complex solar events, decrease the likelihood of overfitting, and reduce training time. The schematic view of this study is depicted in Figure 1.

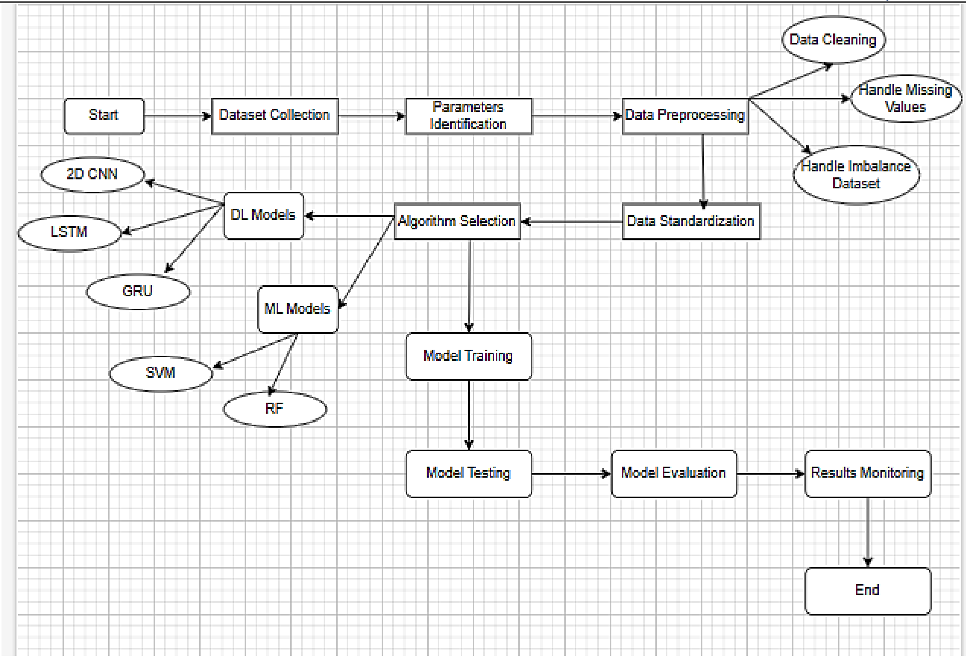

Figure 1. Methodology Diagram

2.3. Performance Validation Method

We validated our model by 5-fold cross-validation (CV), which is widely used in diverse research problems [17-21]. We carried out 5-fold CV to estimate the performance of the model based on following evaluation parameters including accuracy (Acc), sensitivity (Sn), specificity (Sp), and Mathews Correlation Coefficient (MCC). Each one of these criteria can be derived from the values in the confusion matrix (CM) defined by true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). In a weather forecasting context, TP indicates the correctly forecasted weather events (e.g., rainfall or storms) and TN indicates the correctly forecasted non-events (e.g., no rainfall) while FP shows instances of forecasted weather events that were not actual events and FN shows missed real weather events where the model forecast failure. We use Accuracy, Precision, Recall, F1-Score, MCC, and Specificity which can be computed by following equations.

$$F1-Score = 2 \times \frac{TP}{2TP+FP+FN} \quad (1)$$

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (3)$$

$$SN = \frac{TP}{TP+FN} \quad (4)$$

$$SP = \frac{TN}{TN+FP} \quad (5)$$

$$PRE = \frac{TP}{TP+FP} \quad (6)$$

3. Results and Analysis

3.1 Comparative Analysis of DL Methods

The results from multiple deep learning models indicate that 2D-CNN is the best model, achieving accuracy 99.39%, precision 99.45%, recall 99.33%, and F1-score 98.79%, as well as having the highest accuracy overall as shown in Table 1. It is evident that the classification model can capture spatial dependencies and spatial patterns present in the dataset, which makes it the best model to use in this case. Nonetheless, the GRU model also displayed good performance with recall value 96.01%, precision value 98.80%, and accuracy value 94.90%. Thus, GRU model was also able to balance precision and recall while capturing sequential dependencies present in the data. Similarly, the LSTM model was able to achieve 94.80% accuracy, 97.81% precision, and 96.10% recall. GRU and 2D-CNN performed better than LSTM, but the strong recall value of LSTM indicates that it was able to discriminate cases identified as good.

2D-CNN is the most reliable model in this study and outperformed both LSTM and GRU on all measures. Both LSTM and GRU are still good options when working with time-series data, where sequential dependencies are important.

Table 1. Comparative Analysis of DL Methods

Methods	Acc (%)	Pr (%)	Re (%)	F1-Score (%)
LSTM	94.80	97.81	96.10	96.11
GRU	94.90	98.80	96.01	96.21
2D-CNN	99.39	99.45	99.33	98.79

3.2 Comparative Analysis of ML Methods

In comparing the performance of several machine learning techniques including SVM and RF, 2D-CNN outperforms SVM and RF in every measure of performance metric as reflected in Table 2. 2D-CNN achieves the maximum accuracy of 99.39%. 2D-CNN is the best performing for this assessment because of its high accuracy of 99.33% recall and 99.45% precision, which demonstrates that a high percentage of class attribution can be reliably distinguished. The SVM model similarly performs well, achieving a 91.80% accuracy, 91.55% precision, and 92.01% recall. This suggests that the SVM can fit a strong classifier and generalizing well from the training set of experiences. However, in performance metrics it is noticeably lagging 2D-CNNs, which implies that deep learning might be better suited for identifying complex patterns. The Random Forest model achieves a lower performance with 80.25% precision, 77.30% recall, and 79.17% accuracy. It also has fairly strong performance for classification with RF having significant lower performance in comparison to the SVM and 2D-CNN models; Random Forest does not appear to be the best option for this problem.

In general, the 2D-CNN performance is remarkably improved over SVM and RF, further confirming the effectiveness of deep learning in identifying complex patterns in the data. RF is somewhat less accurate than SVM, which is still an important traditional machine learning method, and therefore should be discouraged for this context where high-accuracy predictions are desirable

Table 2. Comparative Analysis of ML Methods

Methods	Acc (%)	Pr (%)	Re (%)	F1-Score (%)
SVM	91.80	91.55	92.01	91.78
RF	79.17	80.25	77.30	78.75
2D-CNN	99.39	99.45	99.33	98.79

3.3 Comparative Analysis with Existing Methods

Further, we analyzed comparative performance of our proposed model with existing methods. Calvo-AdaBoost model struggles with the classification task, as indicated by its low Accuracy of 74.10% and F1-Score of 51.40%. Its most significant limitation is a very low Recall of 42.70%, which means it fails to identify more than half of the total positive cases.

Jovanovic-XGBoost while it shows stronger Accuracy (89.15%) and Precision (88.09%) compared to AdaBoost, its performance is still held back by a low Recall of 65.49%. This suggests that while it correctly classifies the positive cases it finds, it still misses a large

ISSN: 3106-7840 | 3106-7832
Volume 3, Issue 2, 2026

portion of the true positive instances, resulting in an F1-Score of 75.12%.

The Murugan-NB demonstrates a respectable, balanced performance with an Accuracy of 84.62% and its Precision, Recall, and F1-Score all in the 88% range. This indicates a consistent performance, but it is still significantly outpaced by the more advanced deep learning-based methods.

The deep learning models, Yang-LSTM and Cahuantzi-GRU, show the next-best performance with very similar and high scores. Cahuantzi-GRU has a slightly higher Accuracy (94.90%) and Precision (98.80%) than Yang-LSTM, while Yang-LSTM has a marginally higher Recall (96.10%) and F1-Score (96.11%). Both models are highly effective, but they are a clear step down in performance from the proposed model.

Finally, the 2D-CNN (Proposed Model) achieves the highest scores in every category, with near-perfect performance. It boasts an Accuracy of 99.39%, a Precision of 99.45%, a Recall of 99.33%, and an F1-Score of 98.79%. This strong performance across all metrics suggests it is highly effective at both correctly identifying positive cases and not misclassifying negative ones, setting a new benchmark for the task.

Table 3. Comparative Analysis with Existing Methods

Methods	Acc (%)	Pr (%)	Re (%)	F1-Score (%)
Calvo-AdaBoost [12]	74.10	64.70	42.70	51.40
Murugan- NB [8]	84.62	88.32	88.23	88.40
Jovanovic- XGBoost [11]	89.15	88.09	65.49	75.12
Yang- LSTM [15]	94.80	97.81	96.10	96.11
Cahuantzi- GRU [14]	94.90	98.80	96.01	96.21
2D-CNN (Proposed Model)	**99.39**	**99.45**	**99.33**	**98.79**

4. Conclusion and Future Work

The findings of this study showcase how optimization and deep learning techniques can dramatically improve the accuracy of space weather forecasting. With an RNN model having a 2 score of 0.84, it was shown to be the best classification model for predicting sunspots. These results illustrate how artificial intelligence can help solve time-series forecasting and classification problems in an important domain like space weather. Moreover, this study exemplifies the importance of metaheuristic optimization for hyperparameter tuning, particularly the modified particle swarm optimization approach employed here, which performed better than other traditional methods.

To better boost predictiveness, future research could expand on these findings by adding more deep-learning models and ensemble methods. Additionally, data privacy and computing efficacies can be improved through the synergies with cloud computing or federated learning frameworks. Rather than only focusing on the scalability and applicability of these algorithms, they can be more broadly adapted by integrating larger and more diverse datasets, but also a framework for real-time predictions. This would limit the costs, time commitment, and variability around space weather forecasting making a useful, helpful, and site-specific tool.

REFERENCES

Tlatov, A.G., The shape of sunspots and solar activity cycles. Solar Physics, 2022. **297**(8): p. 110.

Alhussein, M., K. Aurangzeb, and S.I. Haider, Hybrid CNN-LSTM model for short-term individual household load forecasting. Ieee Access, 2020. **8**: p. 180544-180557.

Velasco Herrera, V.M., et al., Group sunspot numbers: a new reconstruction of sunspot activity variations from historical sunspot records using algorithms from machine learning. Solar Physics, 2022. **297**(1): p. 8.

Bi, K., et al., Accurate medium-range global weather forecasting with 3D neural networks. Nature, 2023. **619**(7970): p. 533-538.

Chantry, M., et al., Machine learning emulation of gravity wave drag in numerical weather forecasting. Journal of Advances in Modeling Earth Systems, 2021. **13**(7): p. e2021MS002477.

Ben-Nun, T., et al. Productive performance engineering for weather and climate modeling with python. in SC22: International Conference for High Performance Computing, Networking, Storage and Analysis. 2022. IEEE.

Yalçın, S., Weather parameters forecasting with time series using deep hybrid neural networks. Concurrency and Computation: Practice and Experience, 2022. **34**(21): p. e7141.

Murugan Bhagavathi, S., et al., Retracted: Weather forecasting and prediction using hybrid C5. 0 machine learning algorithm. International Journal of Communication Systems, 2021. **34**(10): p. e4805.

Abdulla, N., M. Demirci, and S. Ozdemir, Design and evaluation of adaptive deep learning models for weather forecasting. Engineering Applications of Artificial Intelligence, 2022. **116**: p. 105440.

Ayoub, A., H.M. Wainwright, and G. Sansavini, Machine learning-enabled weather forecasting for real-time radioactive transport and contamination prediction. Progress in Nuclear Energy, 2024. **173**: p. 105255.

Jovanovic, L., et al., Optimizing machine learning for space weather forecasting and event classification using modified metaheuristics. Soft Computing, 2024. **28**(7): p. 6383-6402.

Calvo-Olivera, C., et al., Real-time evaluation of the uncertainty in weather forecasts through machine learning-based models. Water Resources Management, 2024. **38**(7): p. 2455-2470.

Hussain, M., et al. Wind energy forecasts with supervised machine learning techniques. in International Conference on Green Energy, Computing and Intelligent Technology 2024 (GEn-CITy 2024). 2024. IET.

Cahuantzi, R., X. Chen, and S. Güttel. A comparison of LSTM and GRU networks for learning symbolic sequences. in Science and Information Conference. 2023. Springer.

Yang, S., X. Yu, and Y. Zhou. Lstm and gru neural network performance comparison study: Taking yelp review dataset as an example. in 2020 International workshop on electronic communication and artificial intelligence (IWECAI). 2020. IEEE.

Ali, F., et al., DBP-DeepCNN: Prediction of DNA-binding proteins using wavelet-based denoising and deep learning. Chemometrics and Intelligent Laboratory Systems, 2022: p. 104639.

Ali, F., et al., VEGF-ERCNN: A deep learning-based model for prediction of vascular endothelial growth factor using ensemble residual CNN. Journal of Computational Science, 2024. **83**: p. 102448.

Almusallam, N., et al., An omics-driven computational model for angiogenic protein prediction: Advancing therapeutic strategies with Ens-deep-AGP. International Journal of Biological Macromolecules, 2024. **282**: p. 136475.

Almusallam, N., et al., Multi-headed ensemble residual CNN: a powerful tool for fibroblast growth factor prediction. Results in Engineering, 2024. **24**: p. 103348.

Ali, F., et al., Leveraging deep learning for epigenetic protein prediction: a novel approach for early lung cancer diagnosis and drug discovery. Health Information Science and Systems, 2025. **13**(1): p. 28.

Ali, F., et al., IR-MBiTCN: Computational prediction of insulin receptor using deep learning: A multi-information fusion approach with multiscale bidirectional temporal convolutional network. International Journal of Biological Macromolecules, 2025: p. 143844.