URSI AP-RASC, Sydney, Australia, 17 – 22 August 2025

Enhancing Ionospheric Prediction Accuracy Using Deep Learning: A Comparative Study

Stanislav Šafránek(1), Daniel Kouba(2), Pavel Čech(1) and Andrea Žváčková(1)

(1) University of Hradec Kralove, Faculty Informatics and Management, Rokitanského 62, 500 03 Hradec Králové, Czech Republic, e-mail: stanislav.safranek@uhk.cz, pavel.cech@uhk.cz, andrea.zvackova@uhk.cz
(2) Department of Ionosphere and Aeronomy, Institute of Atmospheric Physics, Czech Academy of Sciences, Boční II/1401, 14100 Prague, Czech Republic, e-mail: kouba@ufa.cas.cz

Abstract

This paper presents a comparative study aimed at evaluating the prediction accuracy and generalization capabilities of three algorithms: the Long Short-Term Memory (LSTM), the Convolutional Neural Networks-Long Short-Term Memory (CNN-LSTM) hybrid model, and the Variance Controlled Adaptive Sampling (VCAS). The study uses ionospheric data obtained from the Průhonice Observatory in the Czech Republic, which was divided into a training (80%) and a test (20%) dataset. Accuracy was assessed using an extensive set of metrics and graphical outputs

1. Introduction

The prediction of ionospheric parameters is a crucial aspect of space weather research. The ionosphere exhibits complex spatial and temporal variations influenced by solar activity, geomagnetic storms, and atmospheric dynamics. Traditional deterministic models, such as IRI or the NeQuick model, rely on empirical formulations derived from historical observations. While these models offer reasonable approximations, they struggle to capture transient and extreme ionospheric disturbances [1].
Deep learning models, including LSTM and CNN-LSTM, provide an alternative approach that incorporates sequential learning and feature extraction. LSTM networks are designed to handle long-term dependencies in sequential data by utilizing gated memory cells that regulate information flow. This makes them particularly effective for time series prediction, where past observations influence future outcomes [2]. The CNN-LSTM hybrid model extends this capability by integrating convolutional layers that capture spatial features in ionospheric maps. CNN layers process structured ionospheric datasets, identifying localized anomalies and feature correlations, while LSTM layers analyze temporal dependencies. This combination enhances prediction accuracy, particularly for modeling Total Electron Content (TEC) and F2 layer variations. Studies have demonstrated that CNN-LSTM models outperform standalone LSTM networks in TEC forecasting, with improved generalization across different geomagnetic conditions [3]. Additionally, Variance Controlled Adaptive Sampling (VCAS) are employed for anomaly detection and data reconstruction in ionospheric monitoring. These models leverage probabilistic latent space representations to encode ionospheric variability, improving the identification of abnormal TEC fluctuations and missing data interpolation. VCAS models have been particularly effective in reconstructing sparse GNSS-derived TEC datasets, contributing to enhanced predictive modeling [4].

2. Methodology

For this initial study we focused on a geomagnetically quiet period, which is why we selected the geomagnetically quiet year 2007. In our research, we used data collected by the Digisonde DPS-4 at the ionospheric station in Pruhonice. The critical frequency values of foF2 were taken from the automatic scaling of ionograms with a 15-minute cadence. We used this data to train and evaluate prediction models. For experimental verification of algorithm performance, simulation scenarios were created using the LSTM (Long Short-Term Memory) algorithm, a hybrid CNN-LSTM (Convolutional Neural Network - Long Short-Term Memory) model, and the VCAS (Variance Controlled Adaptive Sampling) method. At the beginning of each simulation, the initial configuration of the models was established, specifying parameters such as the number of epochs, learning rate, initial neural network weights, activation and loss functions, and the definition of training and validation datasets.
The datasets were divided into training (80% of samples) and testing (20% of samples) subsets to objectively assess the model's ability to generalize. The initial conditions were consistent across all models, allowing for a fair and reliable comparison of their predictive performance. After initialization, the actual training phase of the algorithms begins, during which the values of the error function and selected performance metrics are recorded at each iteration. Training continues until one of the predefined stopping criteria is met. These criteria include achieving the desired level of accuracy without significant deviations during the training process or reaching the maximum number of iterations (by default set to 100 epochs). The progress of training was continuously monitored to ensure stable and reliable results.

3. Results

The results of the experimental validation showed differences in the performance of the different algorithms (LSTM, hybrid CNN-LSTM and VCAS method). The performance was evaluated using the recorded values of the error function and the observed metrics during each training epoch.
The training progress of deep learning models applied to ionospheric prediction is crucial for evaluating their generalization capabilities. The provided training graphs illustrate the performance of VCAS, LSTM, and CNN-LSTM models, comparing their convergence speeds, validation errors, and robustness in forecasting ionospheric variations. VCAS models exhibit a rapid decrease in both training and validation loss, stabilizing within a few epochs. This behavior corresponds to their capability to compress and reconstruct ionospheric data efficiently. However, fluctuations in later epochs indicate sensitivity to missing data patterns, suggesting the need for enhanced regularization techniques [5,6]. The LSTM graph indicates slower convergence compared to CNN-based models, primarily due to the sequential nature of RNN architectures. Despite achieving reasonable RMSE values, the model's reliance on long-term dependencies may lead to vanishing gradient issues, necessitating careful hyperparameter tuning and gradient clipping techniques [7]. CNN-LSTM Hybrid Training Trends: CNN-LSTM models demonstrate improved stability in training loss, achieving a 30% reduction in RMSE compared to standalone LSTM models. This improvement indicates that spatial feature extraction through CNN layers enhances temporal learning by mitigating reliance on long-term dependencies.

3.1 Training Performance Overview

The training progress of deep learning models applied to ionospheric prediction is crucial for evaluating their generalization capabilities. The provided training graphs illustrate the performance of VCAS, LSTM, and CNN-LSTM models, comparing their convergence speeds, validation errors, and robustness in forecasting ionospheric variations.

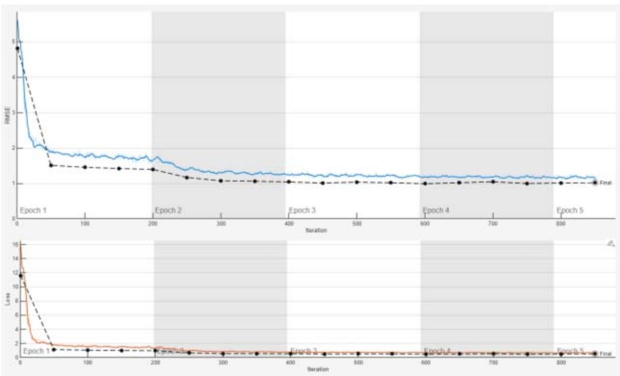

Figure 1. Training and Validation Performance of the LSTM Model

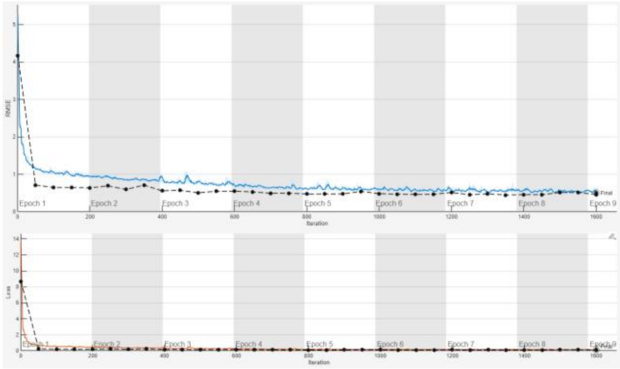

Figure 2. Training and Validation Performance of the CNN-LSTM Model

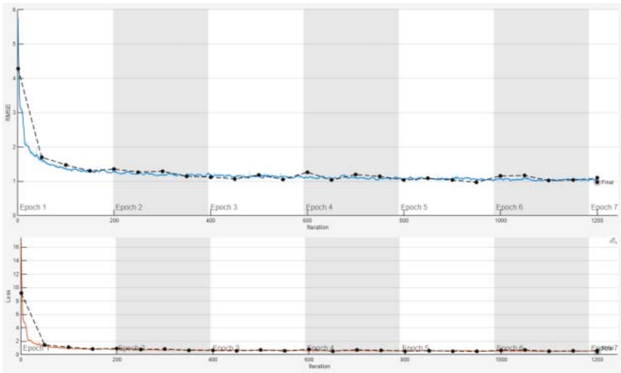

Figure 3. Training and Validation Performance of the VCAS Model

The training progress of deep learning models applied to ionospheric prediction is crucial for evaluating their generalization capabilities. The provided training graphs illustrate the performance of VCAS, LSTM, and CNN-LSTM models, comparing their convergence speeds, validation errors, and robustness in forecasting ionospheric variations.
The graphs illustrate the training progress of neural networks using three different methods: CNN+LSTM, LSTM, and VCAS. Each graph consists of two main parts—the upper section displays the evolution of the RMSE (Root Mean Squared Error) during training, while the lower section shows the loss progression.
In the upper part of the graph, the blue line represents the RMSE curve on the training dataset, with its decline indicating how the model gradually learns and improves its accuracy. Additionally, the light blue line captures individual training iterations with increased granularity, highlighting minor fluctuations in learning. The black dashed line with black dots represents the RMSE on the validation dataset—its stabilization or decrease suggests that the model generalizes well, while sudden fluctuations might indicate overfitting. The lower part of the graph illustrates the loss progression during training. The orange line represents the smoothed training loss, whereas the light orange line shows its unsmoothed version, providing more details about individual iterations. The black dashed line

with dots once again corresponds to the validation dataset, where the stabilization of this value suggests successful training.

A comparison of the individual graphs reveals that CNN+LSTM achieved the best results, as training stopped at epoch 9 with the lowest validation RMSE (~0.50044). The model quickly learned optimal parameters and stabilized. The LSTM method stabilized earlier, around epoch 5, but had a higher RMSE, indicating a weaker generalization capability. The VCAS model performed the worst, with a validation RMSE of 0.96816, and training concluded at epoch 7.

Overall, the CNN+LSTM model delivered the best performance, achieving the lowest RMSE with stable training. The LSTM model was slightly worse but still relatively stable. The VCAS method exhibited the highest RMSE, suggesting that it was the least effective approach for this particular task.

3.2 Detailed Model Comparison

To evaluate the accuracy of the selected algorithms, various performance metrics were applied. Specifically, Root Mean Squared Error (RMSE) was used to assess the average magnitude of prediction errors, emphasizing larger deviations due to its sensitivity toward substantial errors [8]. The Mean Absolute Error (MAE) was employed to measure the average absolute difference between predicted and actual values, providing equal weight to all errors and thus reducing sensitivity to outliers. Additionally, Mean Squared Error (MSE) was calculated to penalize larger prediction errors by squaring deviations, also forming the basis for deriving RMSE. The correlation coefficient quantified the strength and direction of the linear relationship between predicted and observed values, with values close to 1 or -1 indicating strong correlations. The standard error of estimate (STD_Error) was analyzed to assess the precision of predictions; lower values indicate higher reliability [9]. The Mean Bias Error (MBE) quantified systematic prediction biases, distinguishing between overestimation (positive values) and underestimation (negative values) [8]. Median Absolute Error (MedAE) provided robust evaluation against extreme prediction errors, and Symmetric Mean Absolute Percentage Error (sMAPE) was utilized as a symmetric relative accuracy measure, especially suitable for datasets containing values near zero [10]. The combination of these metrics ensured comprehensive and objective evaluation of accuracy and reliability of the tested algorithms.

Table 2. Comparison of LSTM, CNN-LSTM and VCAS accuracy metrics

	LSTM	CNN + LSTM	VCAS
RMSE	1.519	0.713	1.503
MAE	1.219	0.541	1.218
MSE	2.309	0.508	2.259
Correlation	0.792	0.962	0.818
STD_Error	1.503	0.701	1.486
MBE	-0.228	0.135	0.226
MedAE	1.056	0.439	1.109
sMAPE	26.714	12.871	26.701

The results of the experiments are summarized in Table 1, presenting a comprehensive comparison of the selected accuracy metrics across the three evaluated algorithms: LSTM, CNN-LSTM hybrid, and VCAS. Based on the evaluated criteria, the CNN-LSTM algorithm clearly demonstrated superior predictive accuracy and generalization capability compared to the other two methods. Specifically, CNN-LSTM achieved the lowest values for RMSE (0.713), MAE (0.541), MSE (0.508), and STD_Error (0.701), indicating that this model delivered the most precise and stable predictions. Additionally, the CNN-LSTM hybrid displayed the highest correlation coefficient (0.962), signifying a strong linear relationship between predicted and observed values. The sMAPE of 12.871 for CNN-LSTM was substantially lower compared to LSTM (26.714) and VCAS (26.701, further emphasizing the superior relative accuracy of the CNN-LSTM approach.

In contrast, the LSTM and VCAS methods produced very similar results, both showing higher values of error metrics. LSTM achieved an RMSE value of 1.519, while VCAS reached 1.503, and their corresponding MAE values were 1.219 and 1.218, respectively. Both models exhibited lower correlation coefficients (0.792 for LSTM and 0.818 for VCAS), reflecting weaker predictive relationships. The Mean Bias Error (MBE) revealed that the CNN-LSTM had the lowest systematic deviation (0.135), whereas LSTM systematically underestimated (-0.228), and VCAS showed a tendency toward slight overestimation (0.226). These observations underscore the effectiveness of the CNN-LSTM hybrid model, making it the recommended approach due to its ability to accurately model the data while maintaining stability during the training process.

4. Discussion and conclusion

The experiments conducted confirmed that the hybrid CNN-LSTM algorithm achieved superior prediction accuracy and generalization capabilities compared to traditional LSTM and the adaptive VCAS method. Specifically, the CNN-LSTM model demonstrated the lowest prediction errors (RMSE, MAE, and MSE), the highest correlation coefficient (0.962), and minimal systematic bias (MBE = 0.135). Interestingly, despite its more complex architecture, the CNN-LSTM not only provided significantly improved accuracy but also converged faster than the other two methods. This efficiency can be attributed to the convolutional layers' ability to effectively extract spatial features, thereby simplifying subsequent temporal analysis by the LSTM layers.

An intriguing observation is that despite the theoretical advantage of the adaptive sampling approach employed by VCAS—expected to enhance accuracy by intelligently selecting training samples—its performance was comparable to the simpler LSTM model. Both LSTM and VCAS displayed similar error values and correlation coefficients, suggesting that the adaptive sampling strategy did not significantly benefit prediction accuracy within the context of the analyzed ionospheric data. Furthermore, LSTM systematically underestimated (MBE −0.228), while VCAS exhibited slight systematic overestimation (MBE 0.226). CNN-LSTM achieved the lowest systematic bias (MBE: 0.135), indicating more balanced predictions.

From a practical perspective, the enhanced predictive capability of CNN-LSTM is particularly valuable for forecasting dynamic ionospheric phenomena that rapidly fluctuate due to solar and geomagnetic activities. These findings also suggest further research directions, including investigating attention-based CNN-LSTM architectures, which have demonstrated potential in recent studies for improving accuracy and interpretability of predictions. Additionally, exploring the effectiveness of these models across different ionospheric conditions or incorporating external data sources might lead to further advancements in prediction quality and reliability.

In summary, the results of this study underscore the significant advantages of hybrid deep learning architectures in ionospheric prediction, with CNN-LSTM emerging as the most effective approach. Its ability to capture both spatial and temporal dependencies enables more precise forecasting of ionospheric variations, making it a valuable tool for space weather monitoring and communication systems. While the potential of adaptive sampling methods like VCAS warrants further exploration, the findings suggest that its current implementation does not provide a substantial advantage over traditional LSTM models. Future research should focus on refining deep learning architectures, integrating external geophysical data, and evaluating model performance under extreme ionospheric conditions to further enhance prediction accuracy and reliability.

5. Acknowledgements

The research has been partially supported by the Faculty of Informatics and Management UHK specific research project 2107 Addressing Current Challenges in the Field of Smart Cities and Cybersecurity

References

[1] S. Mao, H. Li, Y. Zhang, and Y. Shi, "Prediction of Ionospheric Electron Density Distribution Based on CNN-LSTM Model," *IEEE Geoscience and Remote Sensing Letters*, vol. 21, pp. 1–5, 2024, doi: https://doi.org/10.1109/lgrs.2024.3437650.

[2] J. Tang, Y. Li, M. Ding, H. Liu, D. Yang, and X. Wu, "An Ionospheric TEC Forecasting Model Based on a CNN-LSTM-Attention Mechanism Neural Network," *Remote Sensing*, vol. 14, no. 10, p. 2433, May 2022, doi: https://doi.org/10.3390/rs14102433.

[3] X. Chen, H. Yang, and C. Yu, "Prediction of Ionospheric F2 Layer Height with Bi-parametric Deep Learning Network Based on HFSWR Data," *2022 Photonics & Electromagnetics Research Symposium (PIERS)*, pp. 1–6, Apr. 2024, doi: https://doi.org/10.1109/piers62282.2024.10617844.

[4] W. Li, H. Zhu, S. Shi, D. Zhao, Y. Shen, and C. He, "Modeling China's Sichuan-Yunnan's ionosphere based on multi-channel WOA-CNN-LSTM algorithm," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–18, Jan. 2024, doi: https://doi.org/10.1109/tgrs.2024.3403684.

[5] M. Saqib, M. A. Adil, and M. Freeshah, "Pre-earthquake Ionospheric Perturbation Analysis using Deep Learning Techniques," vol. 1, no. 1, pp. 48–67, Oct. 2023, doi: https://doi.org/10.5281/zenodo.10202326.

[6] T.-R. Tsai et al., "Deep Learning of Detecting Ionospheric Precursors Associated With M ≥ 6.0 Earthquakes in Taiwan," Earth and Space Science, vol. 9, no. 9, Sep. 2022, doi: https://doi.org/10.1029/2022ea002289.

[7] P. Ban et al., "A Regional Ionospheric Storm Forecasting Method Using a Deep Learning Algorithm: LSTM," Space Weather, vol. 21, no. 3, Mar. 2023, doi: https://doi.org/10.1029/2022sw003061.

[8] T. O. Hodson, "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not," Geoscientific Model Development, vol. 15, no. 14, pp. 5481–5487, Jul. 2022, Available: https://gmd.copernicus.org/articles/15/5481/2022/

[9] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature," *Geoscientific Model Development*, vol. 7, no. 3, pp. 1247–1250, Jun. 2014, doi: https://doi.org/10.5194/gmd-7-1247-2014.

[10] A. Botchkarev, "Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology." Accessed: Aug. 08, 2023. [Online]. Available: https://arxiv.org/pdf/1809.03006