Predicting Solar X-ray Flux Using Deep Learning Techniques

1st Sumi Dey
Computational Science
University of Texas at El Paso
El Paso, USA
sdey2@miners.utep.edu

2nd Olac Fuentes
Computer Science
University of Texas at El Paso
El Paso, USA
ofuentes@utep.edu

***Abstract*—The accurate prediction of solar X-ray flux is a difficult problem due to noise and miscalibration of sensors, missing data, and the effects of the Earth's position relative to the line of sight from the sensing satellite to the Sun. Most work on this problem has focused on predicting large sudden increases in flux known as solar flares that can have severe detrimental effects on human activities. Since solar flares are relatively rare and have heterogeneous characteristics, it has been difficult to train machine learning models for their prediction. In this work we approach the problem as regression rather than classification, addressing it as a time series prediction problem. We use X-ray data from the Geostationary Operational Environmental Satellite (GOES) gathered at one-minute intervals to predict the flux after time intervals ranging from one to five hours. We evaluate three deep neural network architectures — a convolutional neural network, Long Short-Term Memory, and a hierarchical dense residual network and compare with two conventional baselines. Our experiments show that all three architectures provide better results than the baseline algorithms, with the hierarchical residual network providing the best results.**

***Index Terms*—Convolutional neural network, Long Short-Term Memory, residual network.**

I. Introduction

Space weather refers to the time-varying conditions in the space surrounding the Earth due to Solar activity. A variety of physical phenomena are associated with space weather, some of which have significant impacts on human activities. In particular, solar flares, which are large sudden increases in X-ray flux, can damage satellite infrastructure, disable power grids, disrupt Global Positioning Systems (GPS), and hinder long-distance communication [25]. A recent study estimated that a single solar flare event of severe magnitude could cause an economic loss of between 0.5 trillion and 2.7 trillion USD [7].

Due to the potential economic impact, efforts to continually monitor solar activity from satellite sensors have been undertaken for nearly two decades. The series of Geostationary Operational Environmental Satellites (GOES), for example, has been used to record the solar X-ray flux since 2001 at one-minute intervals, almost uninterruptedly. Availability of large datasets like GOES has driven a growing interest in automated analysis for event detection and prediction in the last few years. Some approaches for exploiting these data have been proposed [14], [23], [34], mainly focusing on predicting solar flares from X-ray flux. While these works have shown promising results, they have also highlighted the inherent difficulty of the problem. A major challenge is the fact that the physical phenomena giving rise to solar flares are still not well-understood, so physical models have limited applicability. Thus purely data-driven approaches are required. However, since the signal being measured is weak, it presents low signal to noise ratios and quantization effects. Another issue is that as time passes the satellite sensors degrade in ways that are difficult to model. Moreover, as old satellites are replaced, new sensors with different characteristics are used, reducing data consistency. Additionally, about 2% of the data are missing, due to sensor or storage failure and to solar occlusion by the Earth. Lastly, there are absorption effects due to the Earth's atmosphere, which vary depending on time of day and season.

Most work aimed at forecasting solar weather is focused on predicting solar flares from X-ray time series data. Normally this is posed as a classification problem, where a sequence of measurements is classified as a precursor or not of a solar flare within a given time frame. In [3], fluxes above 10^{-5} watts per square meter were considered positive instances (corresponding to X-class and M-class flares, the two strongest types) while the remaining fluxes are regarded as negative instances. Consequently, the binary labeling results in an unbalanced classification problem, as solar flares are relatively rare, with 50 X-class and 742 M-class flares since 2001. In addition, while binarization may help reduce the effects of noise and outliers, it also results in information loss, which may be especially harmful when predicting rare events. Moreover, since flares are heterogeneous and have no clear identified precursors, machine learning models have had only moderate success in this prediction task, performing only slightly better than probabilistic random guessing. In this paper, we propose to tackle the problem of predicting X-ray flux in continuous target space. To the best of our knowledge, our work is the first effort focused on predicting the magnitude of X-ray flux from GOES X-ray flux that poses the problem as a regression rather than a classification problem.

We experimented with three deep learning architectures, namely a 1D Convolutional Neural Network (CNN), a Long

We acknowledge the support of NVIDIA, which donated hardware to perform most of the experiments described in this paper.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

Short-Term Memory (LSTM) and a recently proposed hierarchical dense residual architecture named N-BEATS [24] to assess the potential of deep learning algorithms to forecast X-ray flux. We evaluate the performance of the different networks using the hard X-ray (0.1 to 0.8 nanometers) data from the GOES-15 satellite from 2010 to 2019 at 1-minute cadence. Our experimental results show that all three architectures outperform commonly-used baseline methods, with the hierarchical dense residual architecture providing the best results.

II. Related Work

Solar flares and their associated phenomena have been the subject of intense research through observation and theoretical modeling in recent years. While a complete understanding remains elusive, progress has been made in modeling the formation of sunspots [2], [15], [28], flare eruption processes [19], [27] and coronal mass ejections [11].

Various works have attempted to forecast flares using conventional machine learning algorithms to exploit the large amounts of data available. This has included k-nearest neighbors, logistic regression, random forests, support vector machines, and neural networks, among others.

In [13], solar flares and proton events are predicted applying a classification algorithm called SVM-KNN, which is a combination of support vector machine and k-nearest neighbors. [30] proposed a logistic regression model to predict the probability for a given solar active region to produce X-, M-, or C- class flares in the next twenty-fours hours. Support vector machines have been used successfully to forecast M- and X-class solar flares from Solar Dynamics Observatory's Helioseismic and Magnetic Imager data [4]. Using data from the National Geophysical Data Center, [25] established a correlation between solar flares and sunspot groups. A combination of k-nearest neighbors and nearest centroid algorithms was proposed in [34] to predict solar flares based on the relation between the maximum ratio of the flare flux and no-flare background flux using the GOES X-ray flux dataset and the Space Weather Prediction Center flares catalog. [22] compared the performance of k-NN, SVM, and a random forest algorithm known as ERT [8] to forecast solar flares from solar magnetograms, showing that k-NN offers the best performance. A deep Convolutional Neural Network is developed in [20] to predict solar flares in a time window from 20 minutes to 120 minutes using GOES X-ray flux data. To calculate the probability of flares occurring in the next 24 hours using SDO images, a deep neural network named Deep Flare Net (DeFN) is proposed in [23].

Recurrent neural networks (RNNs) have been used to predict geomagnetic storms [35] and sunspot dynamics [17]. RNNs have also been used in combination with wavelets to forecast hourly and daily solar irradiation [6] and solar wind [21].

III. Proposed Architectures

A. 1D Convolutional Neural Network

CNN models are mainly used for image classification [33], object detection [26], and image segmentation [18], where the given inputs are two dimensional. Recently, 1D CNNs have been used successfully to identify patterns from fixed-length segments in sequential datasets, providing results that are competitive with those of recurrent neural networks with a much smaller computational cost. 1D CNNs have been applied in many challenging sequence-processing tasks, including activity recognition [12], sound classification [1], and several natural language processing tasks [29].

B. Long Short-Term Memory (LSTM)

LSTM is special kind of Recurrent Neural Network that can learn long-term dependencies [10]. The main difference between conventional RNNs and LSTMs is that the repeating module of RNNs has a single layer, while in LSTMs, it has four interacting layers, which allows them to model long-term dependencies effectively. The LSTM architecture provides a large degree of flexibility, allowing many to one and many to many modelling in both regression and classification tasks. LSTM has had great success in several problems, including language modeling [31], speech recognition [9], image captioning [32], handwriting recognition [5], and music generation [16].

C. N-BEATS

N-BEATS is the short form for Neural Basis Expansion Analysis for interpretable Time Series forecasting [24]. Essentially, the N-BEATS architecture consists of a basic block and a double residual stack. The input to the basic block is the history included in a lookback window of a particular length. There are two outputs of the basic block: the block's best estimate of the input, called backcast, and the block's forecast of a certain length.

This basic block contains two parts. The first part is a fully connected network that predicts the expansion coefficients for the backward and forward predictors. The second part receives the expansion coefficients and generates the corresponding backcast and forecast. The basis functions used to generate the backcast and forecast can be parametric functions with learnable parameters (usually low-degree polynomials), or completely unconstrained, implemented as a fully connected networks with ReLU non-linearities.

The outputs of a basic block are subtracted from its inputs and the residues are passed as the input to the next block. Several residual blocks are combined to form a stack. Stacks are combined in a similar manner, resulting in a hierarchical residual network, which can have a very large effective depth.

IV. GOES X-ray Data

A. Data Description

The Geostationary Operational Environmental Satellites (GOES) are a series of geostationary satellites that have

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

been used to monitor space weather since 1975. Solar X-ray emissions are measured in two wavelength bands by the detectors on board the satellites in the GOES series (GOES-10, GOES-11, GOES-13, GOES-14 and GOES-15). The shorter wavelength is from 0.05 to 0.4 nanometers, which is called harder X-ray channel and the longer wavelength is from 0.1 to 0.8 nanometers, which is called softer X-ray channel. The data are made available at 1-minute and 5-minute cadences and are freely available online. For our research, we used 1-minute X-ray with longer wavelength. Figure 1 shows a typical observational sequence including 1000 data points from this dataset. Units are watts per square meter.

B. Data Preprocessing

Before feeding the data into the model we replaced the missing values in the original data with linearly interpolated data. We then clipped the data, setting all the readings below a certain threshold m (assumed to be the minimum signal detectable) to m. After that, we normalized the data to the 0-1 range and applied gamma correction (with $\gamma = 1/3$) to obtain more separation among low magnitude readings (which are the majority in the data set). Figure 2 shows the same 1000 observations of long X-ray flux after preprocessing.

V. Model Architectures

A. 1D CNN

A 1D CNN requires as input a 3D tensor - (number of samples, lookback, number of features), where lookback is the number of previous samples we are looking at for each sample. So we need to convert our input into a 3D tensor to feed into the 1D CNN. We can customize the lookback to analyze the behavior of the network. We used 32 filters in the first layer and each of the them has sliding window (kernel) size 7. Rectified Linear Unit (ReLU) is the activation function here. After the first layer we used maxpooling with region size and stride of 7. The output of the maxpooling layer was flattened and fed to a two-layer fully-connected network. The first fully connected layer has 250 units with ReLu as activation function. The second fully connected layer has a single output and a linear activation function. We used mean squared error as the loss function and the Adam optimizer. The 1D CNN architecture is shown in Figure 3.

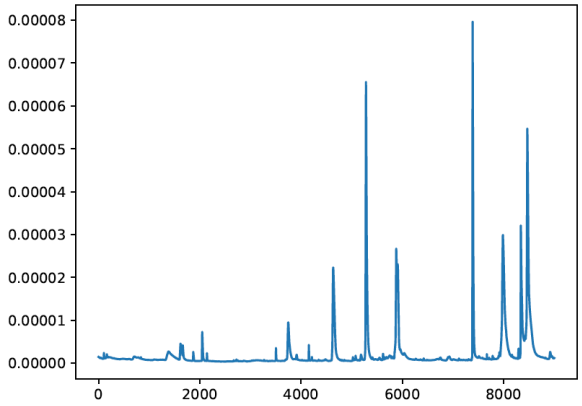

Figure 1: Samples of long X-ray flux before preprocessing

B. LSTM

LSTM also requires a 3D tensor as input - (number of samples, lookback, number of features). So to be fed into the network the data is converted into three dimensional data. We used a single LSTM layer with 200 units with hyperbolic tangent activation function. On top of this LSTM layer, we used a dropout layer with a rate of 0.2. Then we added fully connected layers similar to those used for the 1D CNN. The loss function and optimizer are also the same as the 1D CNN. Figure 4 shows the architecture of the LSTM network.

C. N-BEATS

We used a small version of the N-BEATS architecture. The number of hidden layers in the basic building block is 2. In each layer the number of units is same as the length of the prediction window (i.e. the number of timesteps we are predicting in future). We used a single stack with 2 basic blocks.

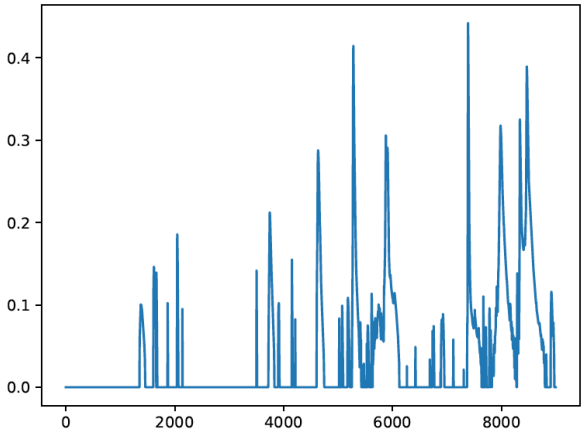

Figure 2: Samples of long X-ray flux after preprocessing

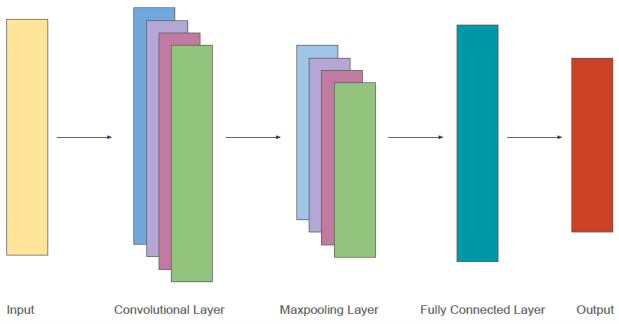

Figure 3: Architecture of 1D Convolutional Neural Network

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

D. Baselines

The three architectures described above were compared to two commonly-used baseline predictors.

1) Baseline-prev.: In this baseline, the latest available flux in the lookback window is used as prediction for the whole forecast window.

2) Baseline-avg.: In this baseline, the average flux in the lookback window is used as prediction for the whole forecast window.

E. Training

We split the dataset into train and test subsets. 80% of the data is used as training set and the remaining 20% is used as test set. The number of features in each sample is determined by the size of the lookback window we are using for a particular experiment. All the models are trained for $29k$ batches of size 128. Adam optimizer is used with a learning rate of 0.001.

VI. Results and Discussion

We evaluated the performance of the proposed architectures with different lookback periods and prediction window lengths and compared them with the two baselines. Figure 6 shows the mean squared error in prediction of long X-ray fluxes for prediction window lengths going from one to five hours using 1D CNN, LSTM and N-BEATS in comparison with the two baselines. For all experiments, we used a lookback window of the same length as the prediction window.

It is clear from the result that the three deep learning methods - 1D CNN, LSTM and N-BEATS outperform both of the baselines while forecasting X-ray flux. Among the learning methods, N-BEATS is the best model for forecasting X-ray flux. The results also show that, as expected, for all methods the error increases as the size of the prediction window increases. Among these two baselines, the performance of baseline-avg is better for all prediction windows larger than one hour. Figure 7 shows the percentage of improvement of our proposed models to forecast X-ray flux for different prediction window lengths over the baseline Baseline-avg. It can be seen that the relative improvement over the baseline remains roughly constant for every architecture as the length of the prediction length increases.

In addition to comparing the performance of the different architectures over the full prediction windows, we analyzed their behavior on a minute-by-minute basis. Figure 8 shows the mean-squared prediction error of the networks trained with 60-minute lookback and a 60-minute prediction window. It can be seen that for very short term prediction (less than

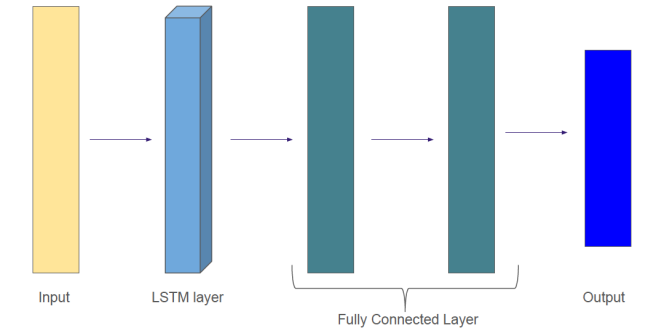

Figure 4: Architecture of Long-Short Term Memory Network

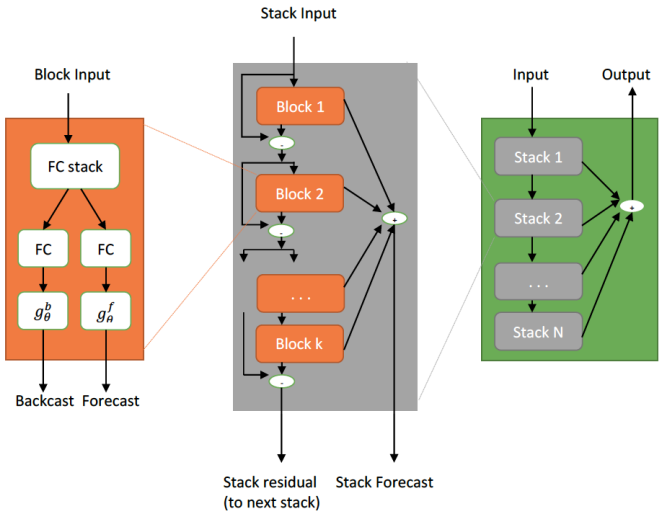

Figure 5: Architecture of the N-BEATS network

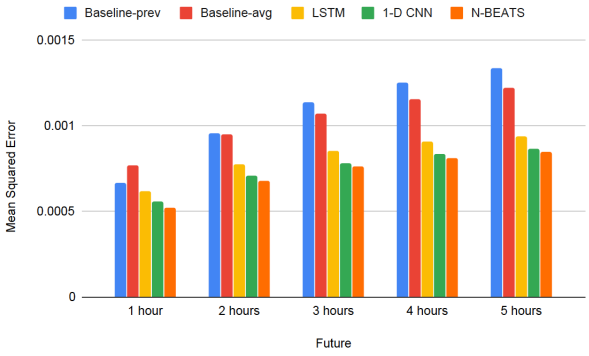

Figure 6: Forecasting result of long X-ray flux

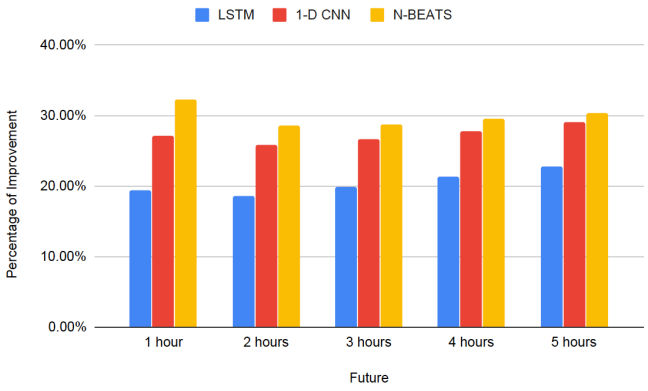

Figure 7: Percentage of improvement compared to Baseline-avg

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

about 15 minutes), LSTM outperforms 1D CNN, but that for delays greater than that 1D CNN is better. On the other hand, the prediction error of N-BEATS is always smaller than both LSTM and 1D CNN.

In Figure 9, we show the minute-by-minute behavior of N-BEATS when trained with different lookbacks and prediction windows; as before, the length of the lookback is the same as the length of the prediction window. It can be seen that the performance is surprisingly uniform for different parameters, showing that optimizing for longer term predictions does not result in an increase in shorter term error.

We also conducted experiments to assess how the length of the lookback window affects the predictions. Figures 10, 11, and 12 show the mean-squared error of 1D CNN, LSTM and N-BEATS, respectively, to forecast X-ray flux for different history windows, where L is the length of the history window that is same as the future window. We can see that a longer lookback results in slight performance improvements for all three architectures with prediction windows of three hours or more, while for shorter prediction windows it appears not to help (in the case of 1D CNN and N-BEATS) and even to hurt performance (in the case of LSTM).

Numerical data summarizing the results described above are presented in Figure 13.

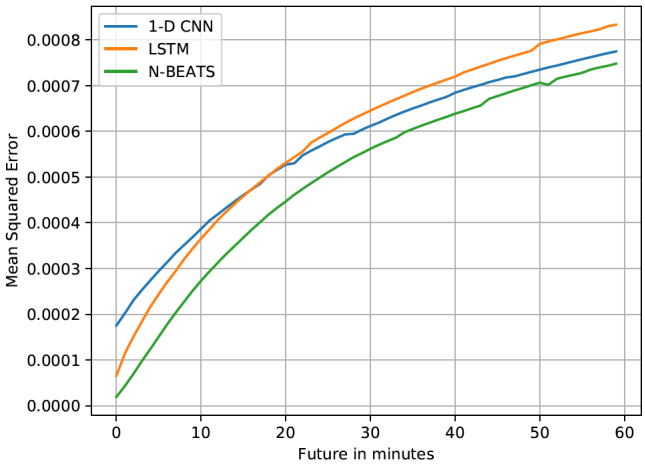

Figure 8: Minute-by-minute X-ray flux forecasting

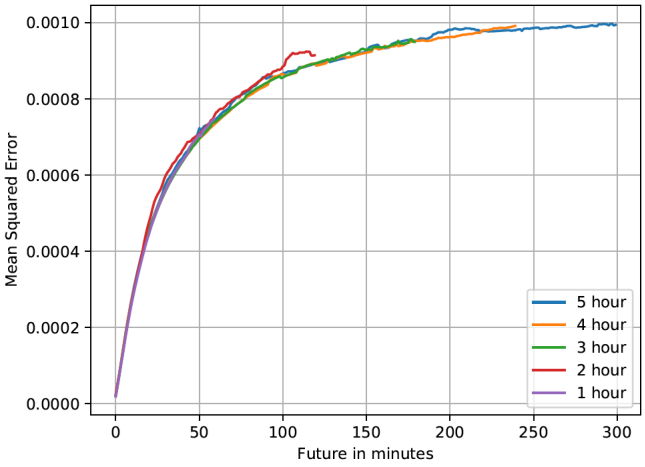

Figure 9: Performance of N-Beats for long X-ray flux prediction with different prediction window sizes

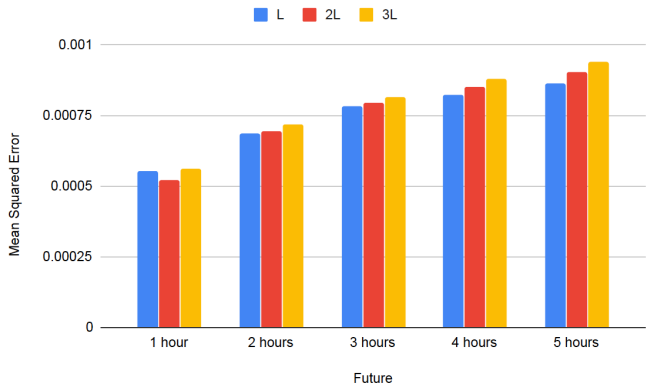

Figure 10: Long X-ray flux forecasting using 1D CNN with different lookback lengths

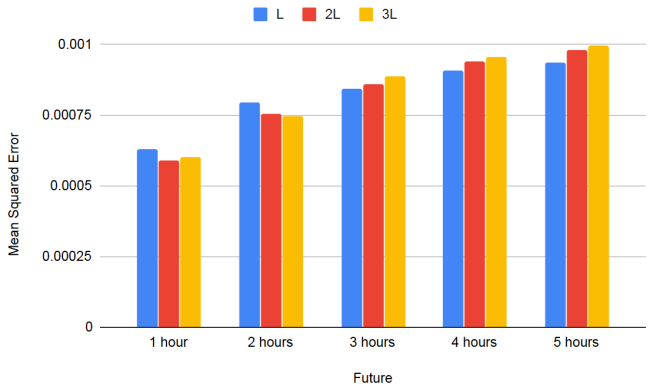

Figure 11: Long X-ray flux forecasting using LSTM with different lookback lengths

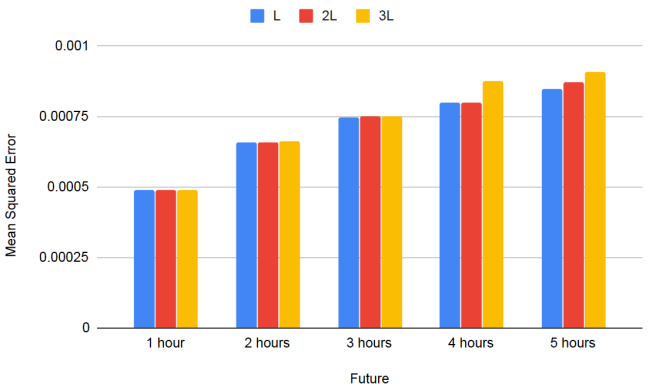

Figure 12: Long X-ray flux forecasting using N-BEATS with different lookback lengths

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

VII. Conclusion and Future Work

We have shown that deep learning models provide a promising approach to the important and very challenging problem of predicting solar X-ray flux. In contrast to previous approaches, we posed the problem as a regression problem and evaluated three types of neural architectures: convolutional (1D CNN), recurrent (LSTM), and dense-hierarchical (N-BEATS) using a real-world dataset with noisy and missing data. All these models outperform both of the baselines, with N-BEATS providing the best results.

We have also analyzed the effect of lookback window sizes in prediction and we have shown that increasing the lookback beyond the size of the prediction window results in little improvement in prediction quality. This is most likely due to the networks' inability to exploit such data because of the chaotic or semi-chaotic behavior of the Sun's surface, where events beyond the most recent past convey little information.

While results are promising, much work remains to be done before reliable forecasting can be achieved. For future research, we will work on combining data from several sources, including proton fluxes and ultraviolet radiation, and combining time series data with image data at various wavelengths.

References

[1] S. Abdoli, P. Cardinal, and A. L. Koerich, "End-to-end environmental sound classification using a 1D convolutional neural network," *Expert Systems with Applications*, vol. 136, pp. 252–263, 2019.

[2] S.-I. Akasofu, "A new consideration on the formation of sunspots," *Physics & Astronomy International Journal*, vol. 2, pp. 408–418, 2018.

[3] R. A. Angryk, P. C. Martens, B. Aydin, D. Kempton, S. S. Mahajan, S. Basodi, A. Ahmadzadeh, S. F. Boubrahimi, S. M. Hamdi, M. A. Schuh *et al.*, "Multivariate time series dataset for space weather data analytics," 2019.

[4] M. G. Bobra and S. Couvidat, "Solar flare prediction using SDO/HMI vector magnetic field data with a machine-learning algorithm," *The Astrophysical Journal*, vol. 798, no. 2, p. 135, 2015.

[5] T. M. Breuel, A. Ul-Hasan, M. A. Al-Azawi, and F. Shafait, "High-performance ocr for printed english and fraktur using LSTM networks," in *2013 12th International Conference on Document Analysis and Recognition*. IEEE, 2013, pp. 683–687.

[6] J. Cao and X. Lin, "Study of hourly and daily solar irradiation forecast using diagonal recurrent wavelet neural networks," *Energy Conversion and Management*, vol. 49, no. 6, pp. 1396–1406, 2008.

Future	Lookback	Baseline-prev	Baseline-avg	LSTM	1-D CNN	N-BEATS
1 hour	L	0.00067	0.00077	0.00063	0.00055	**0.00049**
2 hours	L	0.00096	0.00095	0.00079	0.00068	0.00066
3 hours	L	0.00113	0.00107	0.00084	0.00078	0.00074
4 hours	L	0.00125	0.00115	0.00091	0.00082	0.00079
5 hours	L	0.00134	0.00122	0.00094	0.00086	0.00085
1 hour	2L	0.00067	0.00067	0.00059	0.00052	0.00049
2 hours	2L	0.00096	0.00095	0.00075	0.00069	0.00066
3 hours	2L	0.00113	0.00107	0.00086	0.00079	0.00075
4 hours	2L	0.00125	0.00115	0.00094	0.00085	0.00079
5 hours	2L	0.00134	0.00122	0.00098	0.0009	0.00087
1 hour	3L	0.00067	0.00077	0.0006	0.00056	0.00049
2 hours	3L	0.00096	0.00095	0.00074	0.00072	0.00066
3 hours	3L	0.00113	0.00107	0.00089	0.00081	0.00075
4 hours	3L	0.00125	0.00115	0.00096	0.00088	0.00087
5 hours	3L	0.00134	0.00122	0.00099	0.00094	0.00091

Figure 13: Forecasting result of long X-ray flux using LSTM, 1D CNN and N-BEATS with different lookback lengths

[7] J. Eastwood, E. Biffis, M. Hapgood, L. Green, M. Bisi, R. Bentley, R. Wicks, L.-A. McKinnell, M. Gibbs, and C. Burnett, "The economic impact of space weather: Where do we stand?" *Risk Analysis*, vol. 37, no. 2, pp. 206–218, 2017.

[8] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine learning*, vol. 63, no. 1, pp. 3–42, 2006.

[9] A. Graves, N. Jaitly, and A.-r. Mohamed, "Hybrid speech recognition with deep bidirectional LSTM," in *2013 IEEE workshop on automatic speech recognition and understanding*. IEEE, 2013, pp. 273–278.

[10] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[11] J. Jing, V. B. Yurchyshyn, G. Yang, Y. Xu, and H. Wang, "On the relation between filament eruptions, flares, and coronal mass ejections," *The Astrophysical Journal*, vol. 614, no. 2, p. 1054, 2004.

[12] S.-M. Lee, S. M. Yoon, and H. Cho, "Human activity recognition from accelerometer data using convolutional neural network," in *2017 IEEE International conference on big data and smart computing (BigComp)*. IEEE, 2017, pp. 131–134.

[13] R. Li, Y. Cui, H. He, and H. Wang, "Application of support vector machine combined with K-nearest neighbors in solar flare and solar proton events forecasting," *Advances in Space Research*, vol. 42, no. 9, pp. 1469–1474, 2008.

[14] R. Li, H.-N. Wang, H. He, Y.-M. Cui, and Z.-L. Du, "Support vector machine combined with K-nearest neighbors for solar flare forecasting," *Chinese Journal of Astronomy and Astrophysics*, vol. 7, no. 3, p. 441, 2007.

[15] I. R. Losada, J. Warnecke, K. Glogowski, M. Roth, A. Brandenburg, N. Kleeorin, and I. Rogachevskii, "A new look at sunspot formation using theory and observations," *Proceedings of the International Astronomical Union*, vol. 12, no. S327, pp. 46–59, 2016.

[16] H. H. Mao, T. Shin, and G. Cottrell, "DeepJ: style-specific music generation," in *2018 IEEE 12th International Conference on Semantic Computing (ICSC)*. IEEE, 2018, pp. 377–382.

[17] S. Marra and F. C. Morabito, "A new technique for solar activity forecasting using recurrent elman networks," *International Journal of Computational Intelligence*, vol. 3, no. 1, pp. 8–13, 2006.

[18] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, 2016, pp. 565–571.

[19] P. K. Mitra, B. Joshi, A. Prasad, A. M. Veronig, and R. Bhattacharyya, "Successive flux rope eruptions from δ-sunspots region of NOAA 12673 and associated X-class eruptive flares on 2017 September 6," *The Astrophysical Journal*, vol. 869, no. 1, p. 69, 2018.

[20] T. A. Nagem, R. S. Qahwaji, S. S. Ipson, Z. Wang, and A. S. Al-Waisy, "Deep learning technology for predicting solar flares from (Geostationary Operational Environmental Satellite) data," 2018.

[21] C. Napoli, F. Bonanno, and G. Capizzi, "An hybrid neuro-wavelet approach for long-term prediction of solar wind," *Proceedings of the International Astronomical Union*, vol. 6, no. S274, pp. 153–155, 2010.

[22] N. Nishizuka, K. Sugiura, Y. Kubo, M. Den, S. Watari, and M. Ishii, "Solar flare prediction model with three machine-learning algorithms using ultraviolet brightening and vector magnetograms," *The Astrophysical Journal*, vol. 835, no. 2, p. 156, 2017.

[23] N. Nishizuka, K. Sugiura, Y. Kubo, M. Den, and M. Ishii, "Deep flare net (DeFN) model for solar flare prediction," *The Astrophysical Journal*, vol. 858, no. 2, p. 113, 2018.

[24] B. N. Oreshkin, D. Carpov, N. Chapados, and Y. Bengio, "N-BEATS: neural basis expansion analysis for interpretable time series forecasting," *arXiv preprint arXiv:1905.10437*, 2019.

[25] R. Qahwaji and T. Colak, "Automatic short-term solar flare prediction using machine learning and sunspot associations," *Solar Physics*, vol. 241, no. 1, pp. 195–211, 2007.

[26] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[27] C. J. Schrijver, "Driving major solar flares and eruptions: A review," *Advances in Space Research*, vol. 43, no. 5, pp. 739–755, 2009.

[28] M. Schüssler, "The formation of sunspots and starspots," *Astronomische Nachrichten*, vol. 323, no. 3-4, pp. 377–382, 2002.

[29] N. Sheikh, Z. T. Kefato, and A. Montresor, "Semi-supervised heterogeneous information network embedding for node classification using 1D-CNN," in *2018 Fifth International Conference on Social Networks*

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

Analysis, Management and Security (SNAMS). IEEE, 2018, pp. 177–181.

[30] H. Song, C. Tan, J. Jing, H. Wang, V. Yurchyshyn, and V. Abramenko, "Statistical assessment of photospheric magnetic features in imminent solar flare predictions," *Solar Physics*, vol. 254, no. 1, pp. 101–125, 2009.

[31] M. Sundermeyer, R. Schlüter, and H. Ney, "LSTM neural networks for language modeling," in *Thirteenth annual conference of the international speech communication association*, 2012.

[32] C. Wang, H. Yang, C. Bartz, and C. Meinel, "Image captioning with deep bidirectional LSTMs," in *Proceedings of the 24th ACM international conference on Multimedia*, 2016, pp. 988–997.

[33] Y. Wei, W. Xia, M. Lin, J. Huang, B. Ni, J. Dong, Y. Zhao, and S. Yan, "HCP: a flexible CNN framework for multi-label image classification," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 9, pp. 1901–1907, 2015.

[34] L. M. Winter and K. Balasubramaniam, "Using the maximum X-ray flux ratio and X-ray background to predict solar flare class," *Space Weather*, vol. 13, no. 5, pp. 286–297, 2015.

[35] J.-G. Wu and H. Lundstedt, "Geomagnetic storm predictions from solar wind data with the use of dynamic neural networks," *Journal of Geophysical Research: Space Physics*, vol. 102, no. A7, pp. 14 255–14 268, 1997.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE