The Astrophysical Journal, 949:40 (13pp), 2023 June 1 https://doi.org/10.3847/1538-4357/acc8d5
© 2023. The Author(s). Published by the American Astronomical Society.
OPEN ACCESS

A Machine Learning–Based Approach to Time-series Wave Identification in the Solar Wind

Samuel Fordin[1], Michael Shay[1], Lynn B. Wilson III[2], Bennett Maruca[1], and Barbara J. Thompson[2]
[1] University of Delaware, 104 The Green, Newark, DE 19711, USA; sfordin@udel.edu
[2] NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
Received 2023 January 13; revised 2023 March 15; accepted 2023 March 28; published 2023 May 25

Abstract

The Wind spacecraft has yielded several decades of high-resolution magnetic field data, a large fraction of which displays small-scale structures. In particular, the solar wind is full of wavelike fluctuations that appear in both the field magnitude and its components. The nature of these fluctuations can be tied to the properties of other structures in the solar wind, such as shocks, that have implications for the time evolution of the solar wind. As such, having a large collection of wave events would facilitate further study of the effects that these fluctuations have on solar wind evolution. Given the large volume of magnetic field data available, machine learning is the most practical approach to classifying the myriad small-scale structures observed. To this end, a subset of Wind data is labeled and used as a training set for a multibranch 1D convolutional neural network aimed at classifying circularly polarized wave modes. Using this algorithm, a preliminary statistical study of 1 yr of data is performed, yielding about 300,000 wave intervals out of about 5,000,000 solar wind intervals. The wave intervals come about more often in the fast solar wind and at higher temperatures, and the number of waves per day is highly periodic. This machine learning–based approach to wave detection has the potential to be a powerful, inexpensive way to catalog waves throughout decades of spacecraft data.

Unified Astronomy Thesaurus concepts: Solar wind (1534); Interplanetary physics (827); Space weather (2037); Space plasmas (1544); Neural networks (1933); Convolutional neural networks (1938); Alfven waves (23)

1. Introduction

The solar wind is a turbulent supersonic plasma streaming away from the Sun that has a major impact on the near-space region of Earth. Understanding the important dynamics and morphology of the solar wind, especially the dissipation of fluctuations/discontinuities and subsequent heating, is a major unsolved problem (e.g., Viall & Borovsky 2020). Many of these fluctuations have been classified in terms of their wavelike properties. The exact role that waves play in influencing solar wind evolution is an active area of research, with solar wind studies examining a broad class of waves including but not limited to Alfvén and kinetic Alfvén (Unti & Neugebauer 1968; Hasegawa 1976), magnetosonic whistler (Wilson et al. 2012; Verniero et al. 2020), and ion cyclotron (Jian et al. 2009; Wicks et al. 2016). The large amount of spacecraft data available from in situ sources in the solar wind present an excellent opportunity to investigate the nature of wave modes and the roles that they play. One such data source is the Wind spacecraft, which has taken continuous measurements of solar wind properties for 28 yr, mostly at 1 au (e.g., Wilson et al. 2021). High-resolution magnetic field data from the Magnetic Field Investigation (MFI; Lepping et al. 1995) contain a large number of wave modes that originate from different physical processes.

The cadence of the Wind MFI instrumentation allows it to measure modes with spacecraft-frame frequencies up to ∼5.5 Hz. In this range of frequencies, circularly or elliptically polarized modes are frequently observed in the solar wind, and we focus on these types of modes for this study. One such mode is the magnetosonic-whistler (MSW) mode. The MSWs are an extension of the fast magnetosonic wave above the ion cyclotron frequency (Marsch & Chang 1983). They are right-hand polarized with respect to the quasi-static magnetic field, and the plasma density oscillates in phase with the magnitude of the magnetic field. They have been seen at many locations in the solar wind from 0.1 to 1 au in data from many satellites, such as the Parker Solar Probe (Jagarlamudi et al. 2021; Mozer et al. 2021), ARTEMIS (Stansby et al. 2016; Tong et al. 2019), Wind (Wilson et al. 2017), and Cluster (Lacombe et al. 2014), among numerous others. The MSWs can be generated by a range of instabilities (Tong et al. 2019; Jagarlamudi et al. 2021). They are also generated near shocks and are associated with dissipation in collisionless shocks (Wilson et al. 2012). Based on spectral analysis of ARTEMIS data, MSWs seem to be present in the solar wind ∼10% of the time at 1 au (Tong et al. 2019), though they seem to be less common near the Sun based on Parker Solar Probe measurements (Jagarlamudi et al. 2021; Cattell et al. 2022).

Besides MSW waves, other circularly polarized wave modes near the ion cyclotron frequency are generally identified in the solar wind. Circularly polarized Alfvén/ion cyclotron waves are present in many parts of the solar wind (Jian et al. 2009; Wicks et al. 2016). Alfvén wave frequencies tend to be below the ion cyclotron frequency, lying in the frequency range of $\sim 10^{-4}$–10^{-2} Hz in the plasma frame. Alfvén waves tend to be mainly antisunward-propagating (Bruno & Carbone 2013) and may heat protons (Telloni et al. 2019). Ion cyclotron waves have been observed in solar wind data (Jian et al. 2009), and the ion cyclotron resonance is thought to provide a major source of energy dissipation (Telloni et al. 2019).

Given the long duration of the Wind mission, the Wind data set represents an unprecedented opportunity to study circularly

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

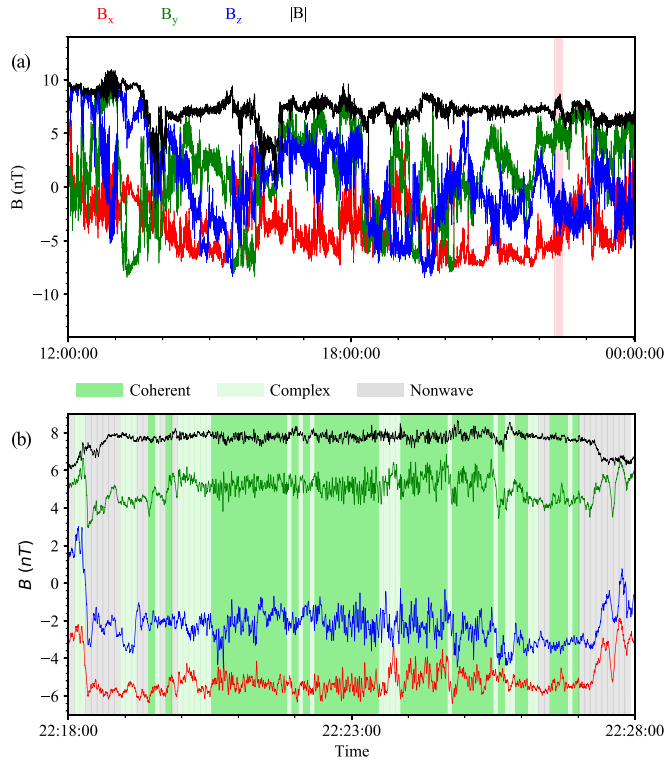


Figure 1. (a) 12 hr interval of Wind magnetic field data from 2004 April 28. The spacecraft is in the solar wind above the ecliptic plane. (b) 10 minute subinterval split into 66 point (∼6 s) time windows manually classified using minimum variance techniques. Dark green time windows correspond to coherent wave time windows, light green time windows correspond to complex wave time windows, and gray time windows correspond to nonwave time windows.

polarized wave modes at 1 au across multiple solar cycles. To examine these wave modes, statistical analysis could be used to examine the general properties of the solar wind, but given the breadth of the data, traditional methods of examining these wave modes are overly expensive and require manual validation. For example, frequency filtering together with minimum variance analysis (MVA) is an effective approach for identifying circularly polarized wave modes, but it requires manual and visual analysis and is time-expensive (e.g., Wilson et al. 2009). Similarly, spectral analysis using wavelet transforms is robust and effective but time-consuming and not a realistic approach to apply to several decades of data (e.g., Wilson et al. 2010). Machine learning is the most practical approach to classify the myriad small-scale structures observed in the magnetic field data.

With that being said, the choice of machine-learning algorithm to use in this context is not an obvious one. Numerous machine-learning algorithms exist and have different applications. In space physics alone, various machine-learning approaches have been used to analyze diverse phenomena. For example, 2D convolutional neural networks (CNNs) have been used to identify flux rope signatures in the solar corona (dos Santos et al. 2020), generative adversarial networks have been used to examine the coronal magnetic field (Jeong et al. 2020), and self-organizing maps have been used to classify magnetic field spectra (Vech & Malaspina 2021). For wave classification, the range of wave periods and highly varying wave durations poses a challenge, as not all machine-learning algorithms generalize well to classifying time-series structures with a high degree of variance in their size and duration.

One candidate for the problem of time-series wave identification is the 1D CNN (1DCNN). Artificial neural networks pass and feed information forward using sequential layers of neurons, and 1DCNNs include one or more 1D convolutional layers in their layout. The 1D convolutional layers within a neural network are given a kernel size, which defines the length of the 1D convolution operation performed by that layer upon a given input data series (LeCun 2015). When these 1D convolutional layers are placed before feed-forward layers in a neural network, the convolutional layers effectively perform a smoothing operation on the input data series and create additional features that are passed to the feed-forward layers. In this way, convolutional layers perform feature extraction by learning their own structures corresponding to a designated kernel size (Kiranyaz et al. 2021).

In space physics, CNNs have been used to process solar wind data for space weather forecasting models that predict the SYM-H geomagnetic index (Siciliano et al. 2021) and the rate of change of the ground magnetic field (Smith et al. 2021; Pinto et al. 2022). In other fields, particularly biology, CNNs have been used in various contexts. For example, they have been used to classify heartbeats from ECG data (Acharya et al. 2017; Wu et al. 2020), automatically diagnose brain wave activity from EEG signals (Acharya et al. 2018), and diagnose structural damage in several different civil, mechanical, and aerospace systems for structural health monitoring (Abdeljaber et al. 2017). The types of signals analyzed in these examples share many similarities with Wind magnetic field data in that they are noisy time series with periodic features of interest that are of varying length.

Given the complexity of using these techniques to identify solar wind fluctuations, we focus first on identifying circularly polarized wave modes in a single high-pass filter (referred to here as coherent wave modes). We utilize a machine-learning approach based on CNNs to identify waves in time-series data from Wind magnetic field data. To this end, a subset of current Wind data is labeled and used as a training set for a machine-learning algorithm aimed at classifying small-scale structures. Eventually, the goal is to generalize to wave modes that originate in intervals that require narrower bandpass filter ranges to be resolved (referred to here as complex wave modes).

2. Building the Training Set

2.1. Collecting Real Data Events

An example solar wind interval is given in Figure 1(a), in which 12 hr of magnetic field data are shown. In Figure 1(b), a 10 minute zoom of the magnetic field from the Wind MFI is shown. The magnetic field in the 10 minute zoom is primarily in the xy plane in geocentric solar ecliptic coordinates (GSE), with a B_z smaller than the other two components. At the beginning of the interval, the magnetic field undergoes a sharp rotation with a small increase in the magnetic field magnitude. After the rotation, there are periods of high-frequency oscillations, eventually followed by a modest reduction in the field magnitude. The high-frequency oscillations during this interval are typically characterized by comparable fluctuations in all magnetic field components with the magnitude of B relatively constant.

As a first step toward using machine learning to classify waves, we manually divide the data into equal-length time intervals, classifying each interval as having "coherent" waves, "complex" waves, or no waves ("nonwave"). Note that for this study, we are limiting the search to circularly polarized waves, with elliptical and planar polarizations characterized as nonwave. In addition, the study is limited to waves with amplitudes larger than 0.1 nT.

To begin the analysis, a high-pass filter is performed on ∼10–20 hr of Wind magnetic field data with a frequency cutoff of 0.2 Hz. In most cases, a wavelet transform of the data is examined by eye for intervals with significant fluctuation energy. In a smaller number of cases, the subintervals were chosen randomly from solar wind data to eliminate bias. Whichever of the two methods is used, the chosen periods are then divided into intervals with 66 time values, which represents about 6 s of spacecraft data.

An MVA is performed on each of these 6 s intervals (Khrabrov & Sonnerup 1998). Multiple MVAs are performed on various subintervals chosen by hand that span at least one period of fluctuations in the fields. Circularly polarized waves are characterized by having the two largest MVA eigenvalues similar in value and much larger than the minimum MVA eigenvalue. Coherent events are those that exhibit the property that the two largest eigenvalues are within a factor of 3 of each other and at least 10 times greater than the minimum eigenvalue. Ultimately, we will use machine-learning algorithms to find these coherent wave events in the solar wind.

Intervals that do not exhibit coherent waves are further analyzed to determine if more complex wavelike structure is present. First, multiple bandpass filters are applied to the intervals. For each resultant bandpass output, an MVA is run and analyzed. If at least one bandpass interval has good eigenvalue ratios, as described in the previous paragraph, that interval is classified as a "complex wave interval" or "complex." If no such good eigenvalue ratios exist, then the event is classified as a "nonwave interval" or "nonwave." Note that in all cases, a visual inspection is performed of the MVA results to verify the classification.

In panel (b) of Figure 1, this wave classification method has been applied to 99 subintervals totaling 10 minutes of magnetic field data. Green intervals (numbering 49) have coherent waves, light green intervals (numbering 26) have complex waves, and gray intervals (numbering 24) are nonwave intervals. Examples of the three types of intervals are shown in Figure 2. Panels (a)–(c) are the raw time-series data. Panels (d)–(f) are the same intervals but high-pass-filtered. Panels (g)–(i) are the high-pass-filtered components in the MVA frame between the dotted gray boundaries. Panels (j) and (k) are the bandpass-filtered components in the frequency range 0.6–1 Hz for complex and nonwave intervals, respectively. Panels (l) and (m) are the bandpass-filtered coordinates in the MVA frame in the associated pink boundaries for the complex and nonwave intervals, respectively. Panels (n)–(p) are the hodograms associated with the final MVA panel in each column.

For the coherent wave in panel (a) of Figure 2, the MVA frame in panel (g) clearly shows an oscillation in B_y and B_z, with little oscillation in B_x. The MVA eigenvalue ratios for this case are $\lambda_{max}/\lambda_{mid} \approx 1.8$, $\lambda_{mid}/\lambda_{min} \approx 33.6$. These parameters, together with visual identification of a nearly circular ellipse in the hodogram in panel (n), are enough to classify this interval as a coherent wave interval.

For the complex wave in panel (b) of Figure 2, there appears to be a potential wave packet, but the high-pass-filtered signal is noisy in panel (e). The MVA eigenvalue ratios for this case are $\lambda_{max}/\lambda_{mid} \approx 1.4$, $\lambda_{mid}/\lambda_{min} \approx 6.8$, with the MVA components plotted in panel (h). In order to isolate the wave modes in this window, a bandpass filter covering a narrower frequency range is required. For this case, the passband 0.6–1 Hz is chosen, and the bandpass-filtered components are plotted in panel (j). The time interval between the pink bounds has MVA eigenvalue ratios $\lambda_{max}/\lambda_{mid} \approx 1.8$, $\lambda_{mid}/\lambda_{min} \approx 71.6$. The hodogram in panel (l) also displays an ellipse with radii that are close enough to satisfy our cutoff for circular polarization, which corresponds to a complex wave interval.

For the nonwave interval in panel (c) of Figure 2, no circularly polarized wave modes are present for any choice of bandpass filter. The high-pass-filtered components plotted in panel (f) and the corresponding MVA components in panel (i) do not show any circular polarization. We choose the same passband as for the complex wave. The bandpass-filtered components are plotted in panel (k), and we evaluate the eigenvalue ratios associated with the MVA components plotted in panel (m). We find $\lambda_{max}/\lambda_{mid} \approx 10.4$, $\lambda_{mid}/\lambda_{min} \approx 19.1$ for the whole interval. If we take a shorter section of data, such as the first third of the window, we find $\lambda_{max}/\lambda_{mid} \approx 4.7$, $\lambda_{mid}/\lambda_{min} \approx 69.4$, which does not meet our definition for a circularly polarized wave interval. Furthermore, the hodogram in panel (p) does not display any notable wave polarization.

By analyzing several thousand ∼6 s solar wind intervals using this time-intensive manual selection process on data spread over the date range 1995–2018, 1274 coherent, 715 complex, and 682 nonwave examples were found. Figure 3 contains histograms of total counts by year.

2.2. Generating Synthetic Time-series Data

Due to the relatively small size of the data set obtained by the above methods, the data set used to train the neural network is augmented with synthetic time-series data that approximate both positive and negative training examples of the solar wind magnetic field. The creation of synthetic data is more time-efficient than the manual labeling method and thus is an attractive approach for quickly increasing the number of intervals available for training. Classifying by hand the 2671 data intervals displayed in Figure 3 required approximately 400 person-hours; to obtain an order of magnitude more data by hand is not feasible, as this would represent several thousand person-hours. Synthetic data have been used in other machine-learning contexts to bolster training sets, such as identifying flux rope signatures (dos Santos et al. 2020) and ECG classification (Acharya et al. 2017). Appendix A provides a detailed description of how the synthetic data are generated, including plots of representative examples. The synthetic data are meant to approximate certain structures in the solar wind, not emulate them exactly; in many cases, the noise background in the synthetic time-series data is nonphysical and does not satisfy $\nabla \cdot B = 0$. However, this property is present across both positive and negative synthetic data and is not expected to bias network training. The ultimate goal as the number of real data events grows is to eventually fully replace the synthetic data in the training set.

The synthetic data consist of several different classes, all of which have a background of noise. Synthetic positive examples have a randomly oriented, circularly polarized wave mode

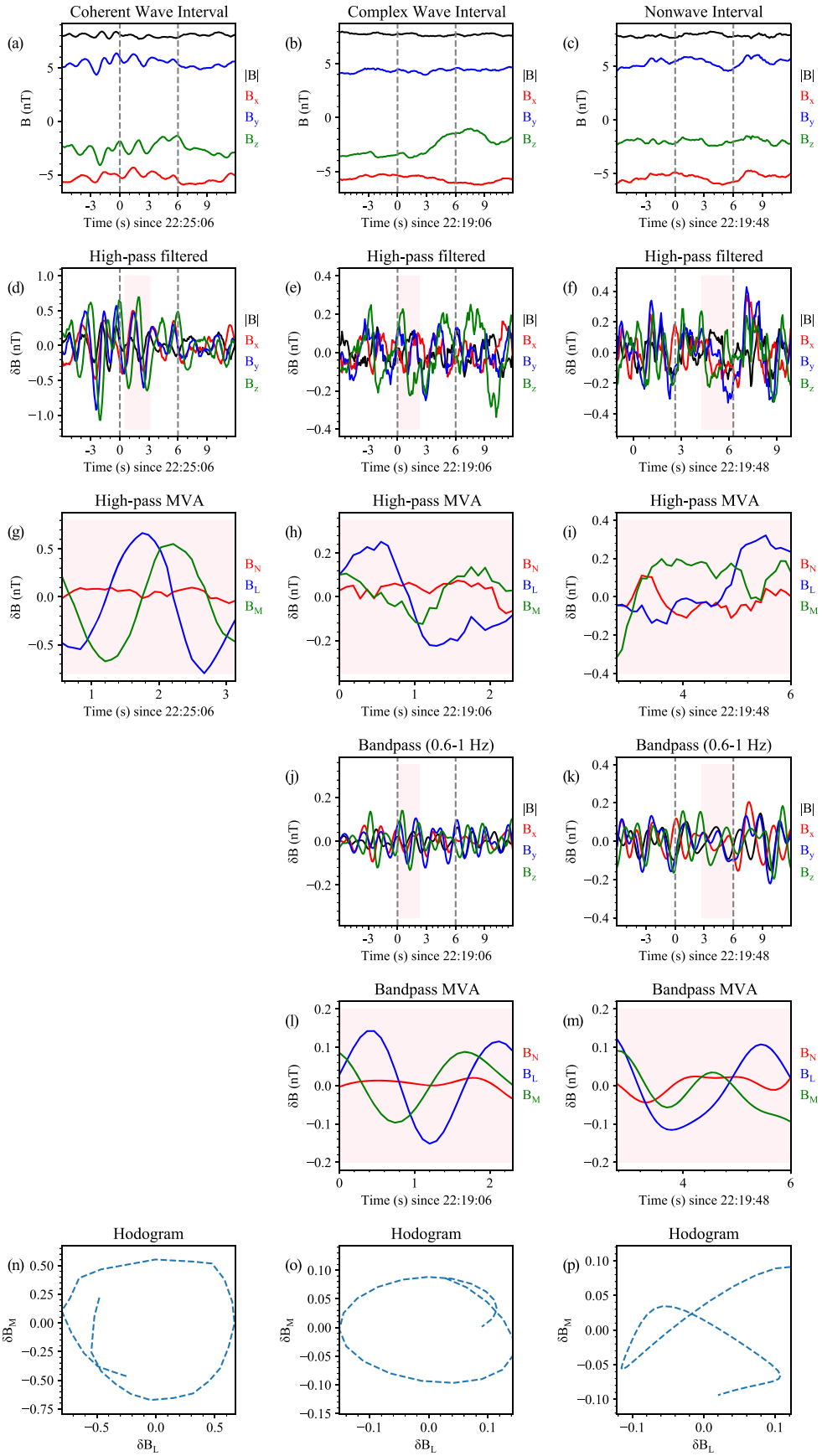


Figure 2. Examples of coherent wave, complex wave, and nonwave intervals. Panels (a)–(c) are the raw time-series data. Panels (d)–(f) are the same intervals but high-pass-filtered. Panels (g)–(i) are the high-pass-filtered coordinates in the MVA frame between the dotted gray boundaries. Panels (j) and (k) are the raw time-series bandpass-filtered components for 0.6–1 Hz. Panels (l) and (m) are the bandpass-filtered coordinates in the MVA frame in the associated pink shaded regions. Panels (n)–(p) are the hodograms associated with the final MVA panel in each column.

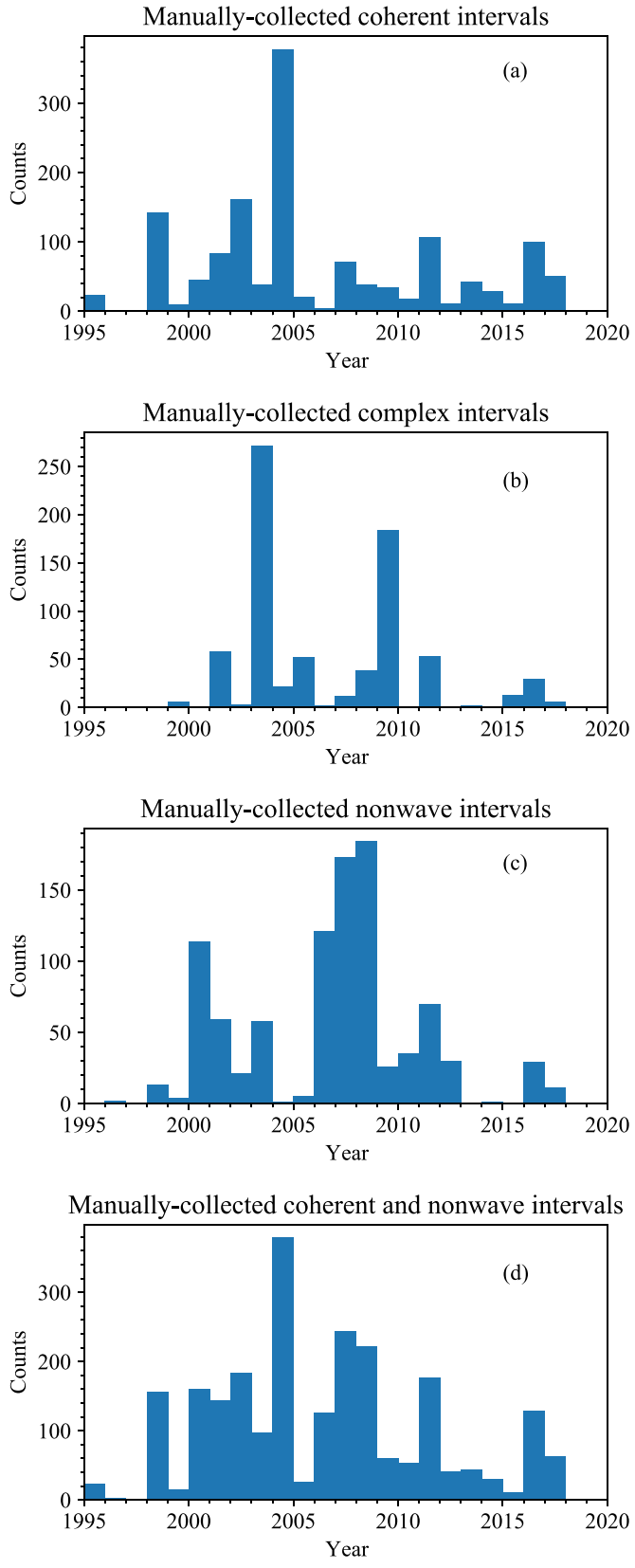


Figure 3. Counts of (a) coherent, (b) complex, (c) nonwave, and (d) total coherent and nonwave events collected manually. The data are drawn from most years between 1995 and 2018 and cover multiple solar cycles. The coherent and nonwave intervals represented in panel (d) are used to train and validate the networks.

combined with a wave envelope function. They may also contain a randomly oriented step function with a magnitude less than or equal to the amplitude of the wave. The circularly polarized wave mode may also be finite in extent inside the time window. Synthetic negative examples may have linearly polarized waves and/or a randomly oriented step function superposed on the noise background. A subset of examples contain no additional structure beyond the noise background.

In total, there are 16 classes of synthetic data with 4000 training examples in each class, for a total of 64,000 synthetic data examples. Half of this is reserved for training, and half is reserved for validation during training.

3. Network Architecture, Training, and Validation

As an initial proof of concept, we focus our effort around coherent waves in 66 point (∼6 s) intervals. The machine-learning problem therefore becomes a supervised binary classification problem with only two classes: coherent events and nonwave events. A single neuron output is sufficient to describe the classification problem, with an output of one corresponding to a coherent event and zero corresponding to a nonwave event. As such, complex wave events are excluded from the training and validation sets. However, complex events will be examined during testing in Section 4.

A priori, it is not known what type of neural network will be the most effective at finding waves. Therefore, we compare multiple types of neural networks with different layouts to classify coherent wave and nonwave intervals. As a simple test case, we consider a neural network with two fully connected and dropout layers and a single output neuron. This network achieved an average validation accuracy on real data of ∼63% across five training trials (detailed in Figure 5, described later), which is not sufficient for real-world applications. Clearly, such a simple network is not sufficient to diagnose waves in a time series where chronological progression plays a key role. To address this issue, it is necessary to add convolutional layers to the network.

In order to allow the network architecture to diagnose waves with many different periods and envelopes, we include several convolutional layers to ensure that the network architecture inherently has feature extraction at multiple scales (Gu et al. 2018). Each convolutional layer in the network has a different kernel size, which is the size of the 1D convolution operation performed over the input time series. Each convolutional layer generates a chosen number of feature maps. Feature maps contain various types of structure that differ based on the size and weight of the convolutional filter applied to the input data. A multibranch CNN implementation provides a structure that combines several convolution operations performed in parallel on the input. A visualization of an N-branch network architecture is shown in Figure 4. Each branch of the 1DCNN begins with two convolutional layers in series that perform convolution on the input magnetic field components. The pooling layers reduce the number of parameters from the convolutional layers to prevent overfitting. The output of each pooling layer is flattened and concatenated before being fed into two fully connected dense layers and two dropout layers. The final layer is a single neuron layer whose output is the network's prediction for the class of the input time series.

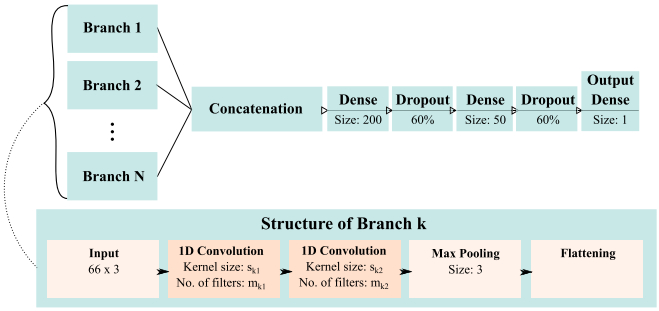


Figure 4. Layout of the five-branch neural network architecture used for coherent wave detection. Each branch contains two convolutional layers, followed by a pooling layer. Information from each branch is combined and fed into multiple fully connected and dropout layers. The choice $N = 5$ yields the best performance on real data and is simpler than $N = 6$ or 7.

We examine the performance of several iterations of an N-branch neural network. Each iteration has a different number of branches, N, with each branch utilizing different kernel sizes. To test the performance of each iteration, we make use of fivefold cross-validation, a technique used to rigorously examine the performance of different network architectures, parameter sets, and hyperparameter sets on a given data set that is commonly used in machine-learning applications (Stone 1974). Fivefold cross-validation involves splitting the training set into five partitions and performing five discrete training trials for a single network iteration, where each training trial uses one of the partitions as the training set and the other four partitions for validation.

For this fivefold validation, we keep the initialization of the random weights for each neuron in a given network iteration consistent across each of the five training trials. For each network variation and training trial, the same hyperparameters were used and are detailed in Appendix B. The fivefold cross-validation results for the neural network without convolutional layers described previously, as well as the results for networks with one, four, five, six, and seven convolutional branches, are tabulated in Figure 5, and the validation accuracy corresponding to real data in the validation set for each network iteration is plotted in Figure 6. Typical loss and accuracy plots are shown in panels (a) and (b) of Figure 17 in Appendix B.

The five-branch network iteration shown in Figure 5 has the highest average validation accuracy, with an average accuracy of 93.6%. The six- and seven-branch network iterations had slightly lower average accuracies of 93.2% and 92.7%, respectively. Given that the performances of the six- and seven-branch networks do not show accuracy improvements over the five-branch network, the convolutional branches up to the fifth branch contribute features that contain useful information, while the convolutional layers in the sixth and seventh branches do not contribute meaningful new features.

Next, an important consideration is the best ratio of real to synthetic data to use in the training and the efficacy of the synthetic data in replacing real data during training. Several random training trials using a five-branch architecture were performed using three different training sets, with each containing the same synthetic data and a different amount of real data. Respectively, the three training sets contained 0% of the real data, 35% of the real data (446 coherent and 239 nonwave intervals), and 70% of the real data (892 coherent and 477 nonwave intervals). A holdout set containing 360 coherent and 205 nonwave intervals was designated and used across all three training sets to test network performance. A classification threshold of 0.5 was used to differentiate between coherent and nonwave predictions. The validation results for each network are shown in Figure 7, and loss and accuracy plots associated with the highest-accuracy network are given in Figure 17 in Appendix B. Averaged across three training experiments, the lowest-accuracy networks were trained using the fully synthetic training set, with an average accuracy of 77.8%. The second-highest accuracy set was trained with half of the real training data training set, with an average accuracy of 88.1%. The networks trained with all of the real training data performed the best, with an average accuracy of 94.0% on the validation set. The highest-accuracy network has an accuracy of 95%. Histograms of network predictions from the highest-accuracy network on the holdout set are shown for coherent events in Figure 8. It is promising that higher proportions of real data noticeably improve network accuracy; this is a sign that the network is learning meaningful features about the real data. As the proportion of real data in the training set grows, a larger number of relevant features are captured by the network, improving the accuracy of the network.

4. Testing Using Contiguous Data

While the validation results are very promising, there is still significant uncertainty concerning the utility of the network for isolating coherent wave events in a typical contiguous solar wind interval. Specifically, complex wave events were excluded from the training and holdout sets, but of course, they would be present in a real solar wind analysis. To study the response of the network to raw solar wind data, all 99 6 s intervals of panel (b) of Figure 1 were classified by the highest-accuracy network from among the random training trials. These events included a mix of coherent, complex, and nonwave events.

The results of this testing trial are tabulated in Figure 9. Using an output value of 0.5 as the boundary between coherent and nonwave, all 49 coherent events were classified as wavelike. Similarly, all 24 nonwave events were classified as nonwave. Twelve of 26 complex events were classified as wavelike.

The network output of complex waves may be problematic depending on our goals for the network in classifying solar wind data. If we desire to find all complex and coherent waves and discard nonwave intervals, clearly our network is not sufficient; there is no threshold that both includes the majority of complex events and at the same time discards most nonwave events. On the other hand, the more modest goal of finding coherent events is reasonable. A receiver operating characteristic (ROC) curve is shown in panel (a) of Figure 10. Here the true-positive rate (TPR) is defined as

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{1}$$

and the false-positive rate (FPR) is defined as

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}, \tag{2}$$

where TP, FP, FN, and TN are the numbers of true-positive, false-positive, false-negative, and true-negative events, respectively. In the ROC curve, we find that a threshold of 0.57 yields the closest point to (FPR = 0, TPR = 1).

ANN	# TP	# FN	# FP	# TN	% Accuracy
Fold 1	101	139	17	120	58.6%
Fold 2	115	125	19	118	61.8%
Fold 3	114	126	10	127	63.9%
Fold 4	104	136	16	121	59.7%
Fold 5	121	118	11	125	65.6%
Average	111	129	15	122	**63.1%**
1DCNN, 1B	**# TP**	**# FN**	**# FP**	**# TN**	**% Accuracy**
Fold 1	173	67	37	100	72.4%
Fold 2	164	76	25	112	73.2%
Fold 3	153	87	29	108	69.2%
Fold 4	151	89	27	110	69.2%
Fold 5	167	73	27	110	73.5%
Average	162	78	29	108	**71.5%**
1DCNN, 4B	**# TP**	**# FN**	**# FP**	**# TN**	**% Accuracy**
Fold 1	199	41	5	132	87.8%
Fold 2	228	12	16	121	92.6%
Fold 3	224	16	5	132	94.4%
Fold 4	211	29	2	135	91.8%
Fold 5	208	31	6	130	90.1%
Average	214	26	7	130	**91.3%**
1DCNN, 5B	**# TP**	**# FN**	**# FP**	**# TN**	**% Accuracy**
Fold 1	216	24	10	127	91.0%
Fold 2	232	8	13	124	94.4%
Fold 3	230	10	12	125	94.2%
Fold 4	225	15	6	131	94.4%
Fold 5	220	19	4	132	93.9%
Average	225	15	9	128	**93.6%**
1DCNN, 6B	**# TP**	**# FN**	**# FP**	**# TN**	**% Accuracy**
Fold 1	221	19	12	125	91.8%
Fold 2	238	2	22	115	93.6%
Fold 3	231	9	16	121	93.4%
Fold 4	226	14	7	130	94.4%
Fold 5	222	17	10	126	92.8%
Average	228	12	13	123	**93.2%**
1DCNN, 7B	**# TP**	**# FN**	**# FP**	**# TN**	**% Accuracy**
Fold 1	216	24	12	125	90.5%
Fold 2	231	9	13	124	94.2%
Fold 3	221	19	3	134	94.2%
Fold 4	218	22	5	132	92.8%
Fold 5	213	26	4	132	92.0%
Average	220	20	7	129	**92.7%**

Figure 5. Fivefold cross-validation results for the five neural network configurations, detailing the total number of true positives (TP), false negatives (FN), false positives (FP), and true negatives (TN) for each training trial. The accuracy is the highest for five branches, with an average accuracy of 93.6% across the five cross-validation folds.

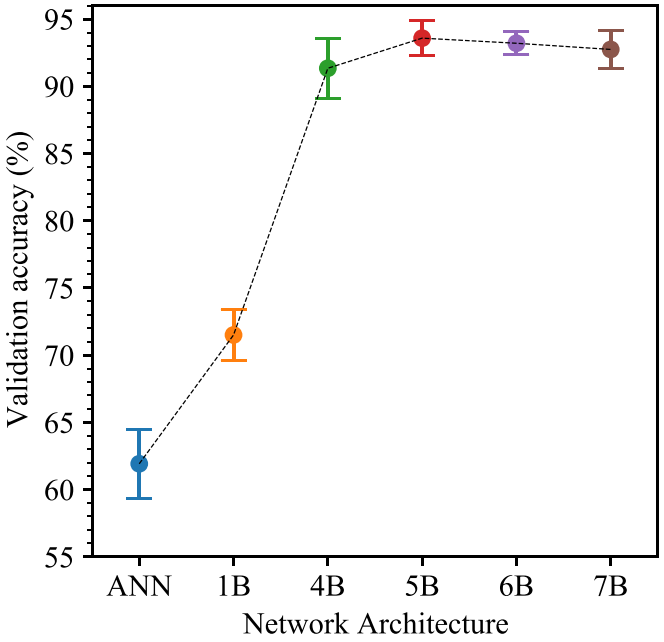


Figure 6. Fivefold cross-validation results for the five network configurations plotted by number of branches. The five-branch configuration has the highest average accuracy.

Synthetic Only	# TP	# FN	# FP	# TN	% Accuracy
Trial 1	255	105	22	183	77.5%
Trial 2	271	89	39	166	77.3%
Trial 3	257	103	19	186	78.4%
Average	261	99	27	178	**77.8%**
Hybrid (35%)	**# TP**	**# FN**	**# FP**	**# TN**	**% Accuracy**
Trial 1	315	45	12	193	89.9%
Trial 2	318	42	10	195	90.8%
Trial 3	279	81	12	193	83.5%
Average	304	56	11	194	**88.1%**
Hybrid (70%)	**# TP**	**# FN**	**# FP**	**# TN**	**% Accuracy**
Trial 1	334	26	8	197	94.0%
Trial 2	343	17	11	194	**95.0%**
Trial 3	335	25	14	191	93.1%
Average	337	23	11	194	**94.0%**

Figure 7. Validation results for the nine randomly trained networks, detailing the total number of true positives (TP), false negatives (FN), false positives (FP), and true negatives (TN) for each network. Hybrid 35% refers to the networks trained with synthetic data and 35% of the real data set. Hybrid 70% refers to the networks trained with synthetic data and 70% of the real data set. A classification threshold of 0.5 was used to determine the class prediction. The training set with the highest accuracy is the hybrid 70% set, and the highest-accuracy network within that set achieved 95.0% validation accuracy.

However, this threshold does not consider the reality that for typical long-duration solar wind measurements, it is expected that the nonwave events will outnumber the wave events by roughly a factor of 10. A threshold of 0.57 that optimizes the ROC curve would likely give comparable numbers of false-positive and true-positive events, which is not acceptable for the type of statistical studies planned with this data set. For that reason, we choose to optimize the threshold to minimize the ratio of the number of false-positive to true-positive events while at the same time still finding a large number of true positives.

This goal can be achieved by increasing the classification threshold of the network. In panel (b) of Figure 10, the ratio of the TPR to FPR for varying choices of threshold is shown. The ratio is highest near $\approx$1, but a threshold at or near 1 would exclude almost all coherent intervals. The ratio near $\approx$0.95 is more than double the ratio associated with 0.57. In Figure 8, pertaining to the holdout set, 80% of the coherent wave intervals had network predictions above 0.95, while only 2% of the nonwave intervals had network predictions above 0.95. In Figure 9, pertaining to the 10 minute data interval in Figure 1, 94% of the coherent intervals had network predictions above 0.95, and 0% of the nonwave intervals had network predictions above 0.95. Assuming that the holdout set and 10 minute data interval together provide an adequate representation of the whole set of coherent and nonwave intervals in Wind data, we extrapolate that a threshold value of $\approx$0.95 will capture nearly all coherent events, some complex events, and a small number of nonwave events relative to coherent events.

5. Classifying One Year

Having demonstrated that the network can find coherent waves, the next natural step is to apply the network to a much longer time interval and examine any patterns that emerge regarding circularly polarized waves in the solar wind. As an

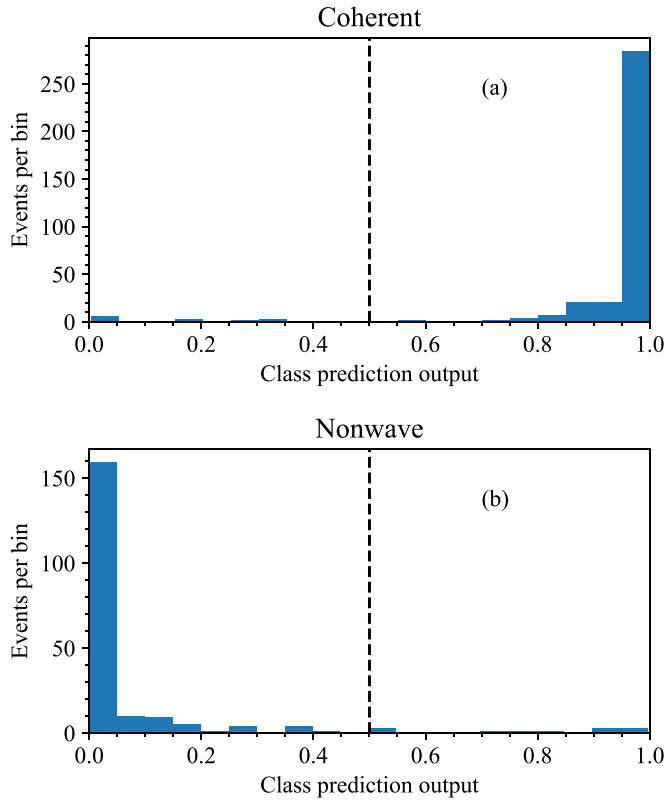


Figure 8. Class predictions of the highest-accuracy network for (a) coherent and (b) nonwave intervals in the validation set. Using a classification threshold of 0.5, 343 out of 360 coherent and 194 out of 205 nonwave intervals were classified correctly.

initial case, we study Wind magnetic data for a single year; 2005 marks the first full year during which Wind operated at the L1 point. The year has few corresponding data examples included in the training set (see Figure 9), which makes it particularly attractive for testing the network's performance on data on which it has not been trained. The year is in the declining phase of the solar cycle about halfway between maximum and minimum.

An overview of this year of data is shown in Figure 11. Averaged by day, the magnetic field components cover a wide range of values, fluctuating in magnitude between 1 and 26 nT. Similarly, the daily averages of the magnitude of the velocity components fluctuate regularly, with values corresponding to both the fast and slow solar wind. The proton density also varies substantially through the year, with averages ranging from ∼1 to ∼24 cm^{-3}. The antisunward velocity component and the proton density notably fluctuate with a periodicity of ∼9 days. These fluctuations appear to be oscillations between the fast and slow solar wind, possibly related to a corotating interaction region.

The analysis of this year of data is the same as described in Section 2.1; the magnetic field components for the year are high-pass-filtered above 0.2 Hz and separated into 66 point intervals. Each interval was analyzed by the best-performing network from Figure 7, the results of which are shown in Figure 12. As discussed in Section 4, we choose a classification threshold of 0.95 (denoted by the vertical dashed line) to minimize false positives; this choice leads to 227,991 intervals classified as coherent out of a total of 5,040,592 intervals (∼4.5%).

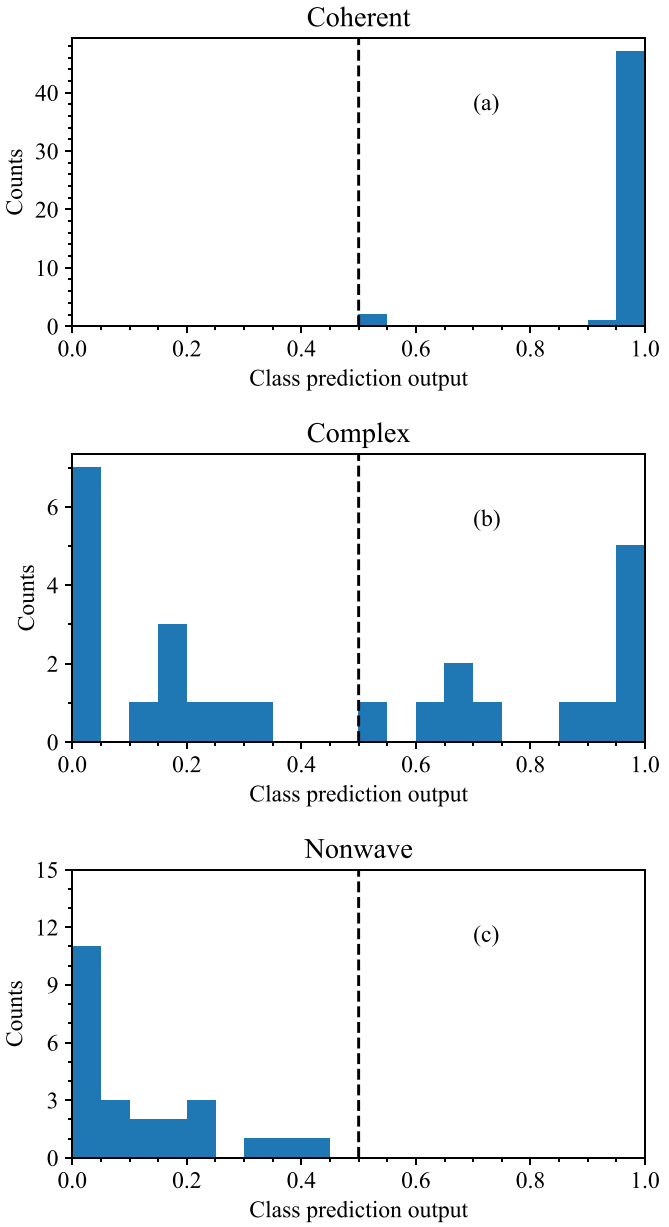


Figure 9. Network prediction results for the 10 minute period shown in Figure 1, grouped by (a) coherent, (b) complex, and (c) nonwave intervals. The classification threshold used is 0.5 and denoted by the thick black dashed lines. All 49 coherent intervals and all 24 nonwave intervals were classified correctly by the network, while 12 of 26 complex events were marked as wavelike.

To ensure that the neural network successfully found waves, we examined and classified by hand a random selection of 100 intervals scored 0.95 and above by the network; of these, 45 were coherent, 44 were complex, and 11 were nonwave. Of these nonwave intervals, nine of 11 were either elliptically polarized waves that did not meet our definition for circular polarization or circularly polarized waves with amplitudes less than 0.1 nT. We extrapolate that ∼90% of the predictions above the classification threshold of 0.95 contain circularly polarized waves. Any discussion of statistics should be considered carefully, therefore, as a relatively large number of false-positive intervals dilute the positive predictions made by the network. However, due to the large number of positive predictions in total, the statistics of the whole set of wave

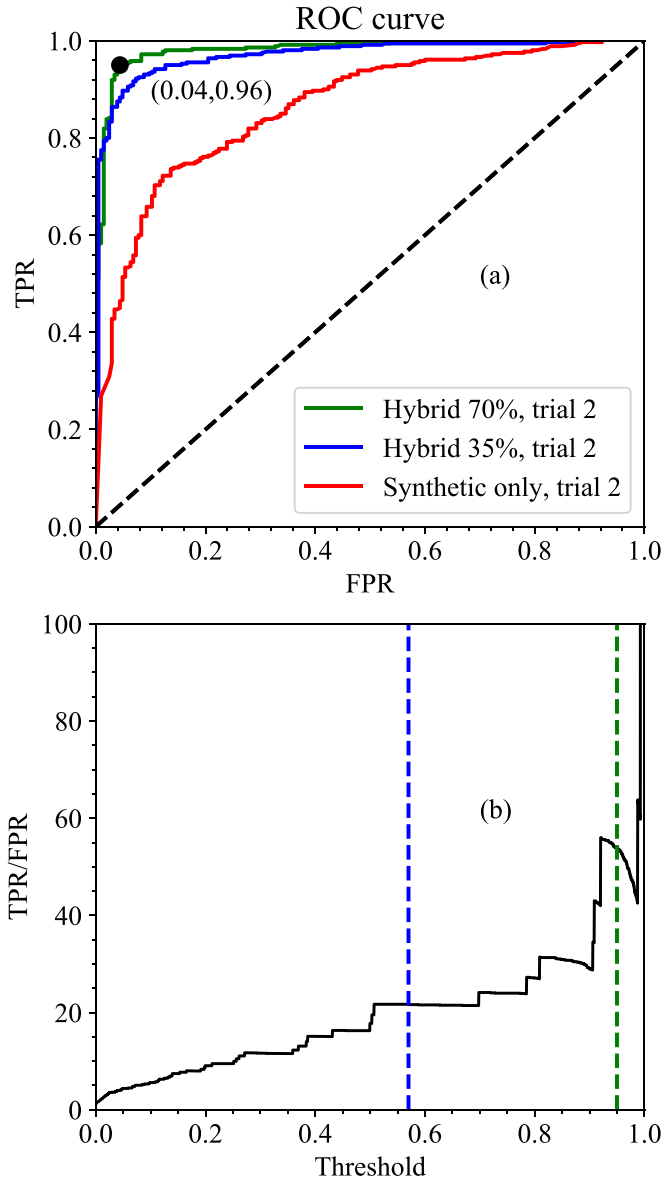


Figure 10. (a) ROC curve associated with three network iterations given in Figure 7. (b) Ratio of the TPR vs. FPR for varying choices of classification threshold for the highest-accuracy network (hybrid 70%, trial 2) given in Figure 7. The capability of the five-branch network to function as a binary classifier improves as real data are added to the training set, and the optimal threshold based on validation data is marked by the black dot in panel (a). However, the ratio of true-positive to false-positive events for the 0.57 threshold (denoted by the blue dotted line in panel (b)) is not sufficiently high when considering the distribution of wave vs. nonwave events in the solar wind. Instead, a threshold of 0.95 (green dotted line) yields a much higher ratio of true-positive to false-positive events while still preserving a large number of true-positive events.

predictions can be examined with a relatively high degree of confidence.

The counts of wave intervals per day, along with the Fourier transforms of the set of counts, are shown in Figure 13. In panel (a), the number of wave intervals varies greatly by day, from a minimum of seven to a maximum of 2641 intervals. The fast Fourier transform (FFT) in panel (b) reveals several peaks. The strongest is associated with the ∼9 day oscillation period discussed in the previous paragraph, with smaller peaks corresponding to periodicities of ∼6.6, ∼13.5, and ∼27 days. Some of these peaks may be harmonics of the solar rotation

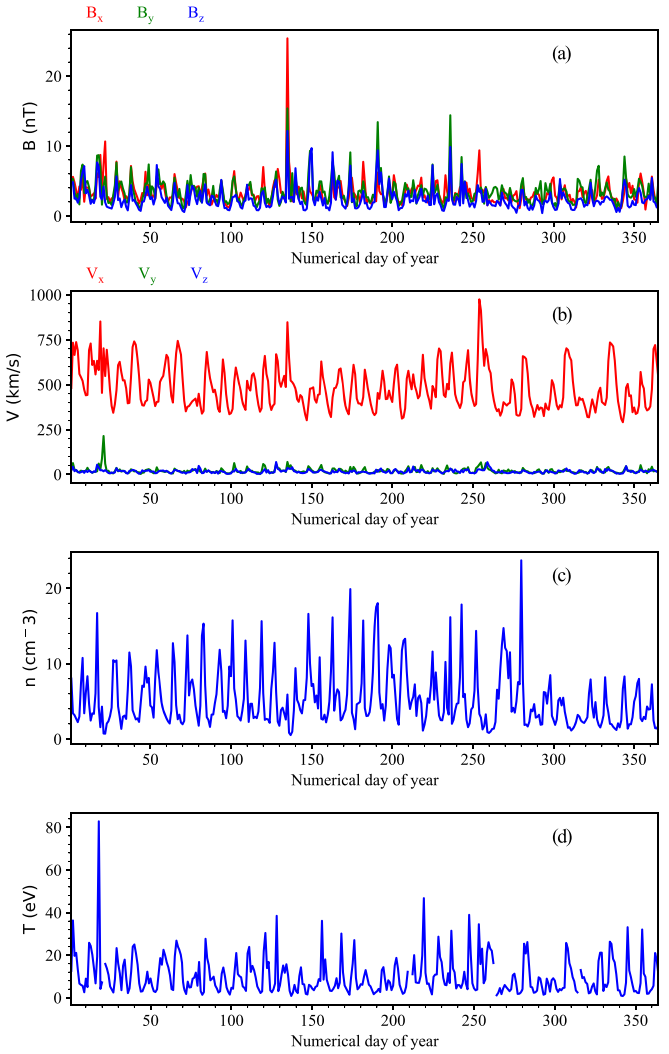


Figure 11. Averages of solar wind parameters by day of the year for 2005. Shown are the (a) magnitude of magnetic field components, (b) solar wind velocity components, (c) proton density, and (d) core proton temperature.

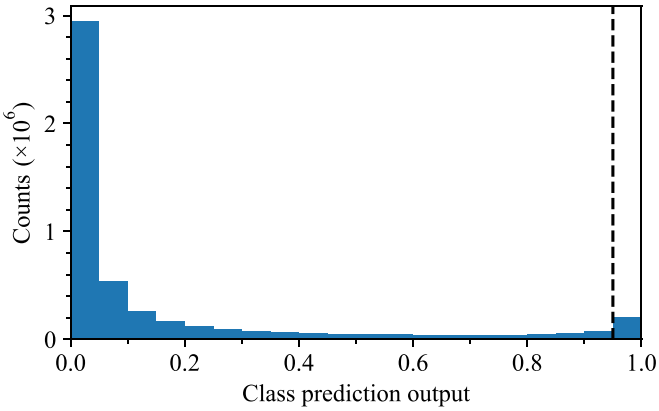


Figure 12. Network prediction results for 2005. The classification threshold used is 0.95, denoted by the black dashed line. Out of a total of 5,040,592 intervals, 227,991 were marked as wavelike.

frequency or spacecraft incidence with the heliospheric current sheet. Beyond ∼0.18 day^{-1} (∼5.5 days), there is a steady decrease in the FFT.

An important question is in regard to the kinds of solar wind conditions that are more likely to have wave modes. Due to the

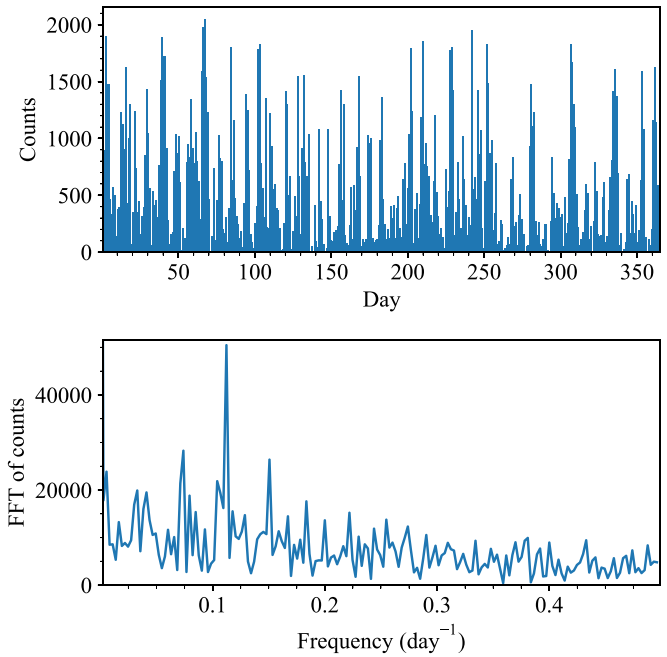


Figure 13. (a) Histogram of counts of coherent events in 2005 binned by day of the year. (b) Associated FFT of the histogram. The total counts per day has a strong ∼9 day periodicity, along with smaller peaks corresponding to ∼6.6, ∼13.5, and ∼27 day periodicities.

relatively small number of wave intervals compared to the total number of solar wind intervals, we study probability density functions (PDFs) of solar wind parameters for both wave intervals and all intervals. Figure 14 shows PDFs and medians of several key solar wind parameters: solar wind speed, magnetic field strength, proton temperature, and proton density. The median solar wind velocity of the solar wind is ∼455 km s^{-1}, and that of the wave population is ∼540 km s^{-1}, an increase of ∼20% for wave intervals. Similarly, the median proton temperature of the solar wind is ∼4.1 eV, and that of the wave population is ∼9.4 eV, an increase of ∼130%. The magnetic field strength for coherent intervals is ∼0.5 nT higher than for all intervals, an ∼10% increase in the magnetic field strength. On the other hand, the proton density shows a much smaller difference between the wave events and all events, with the median proton density for each population being less than 5% larger than the median for all intervals. Relative to all intervals, wave intervals tend to occur more often in the fast solar wind and at higher temperatures and magnetic field strengths.

6. Discussion

The initial analysis presented here is a strong starting point for the automatic detection of wave modes in the solar wind. We have demonstrated that a multibranch convolutional neural network (CNN) is effective at finding coherent circularly polarized waves that can ultimately be used to create a database of such waves in the solar wind. A cursory analysis of 1 yr of solar wind data has yielded intriguing results, in particular that coherent waves are measured more frequently in fast solar wind and at higher temperatures. Whether this is due to selection effects such as Doppler shifting or a basic change in the generating mechanism of these waves remains to be seen.

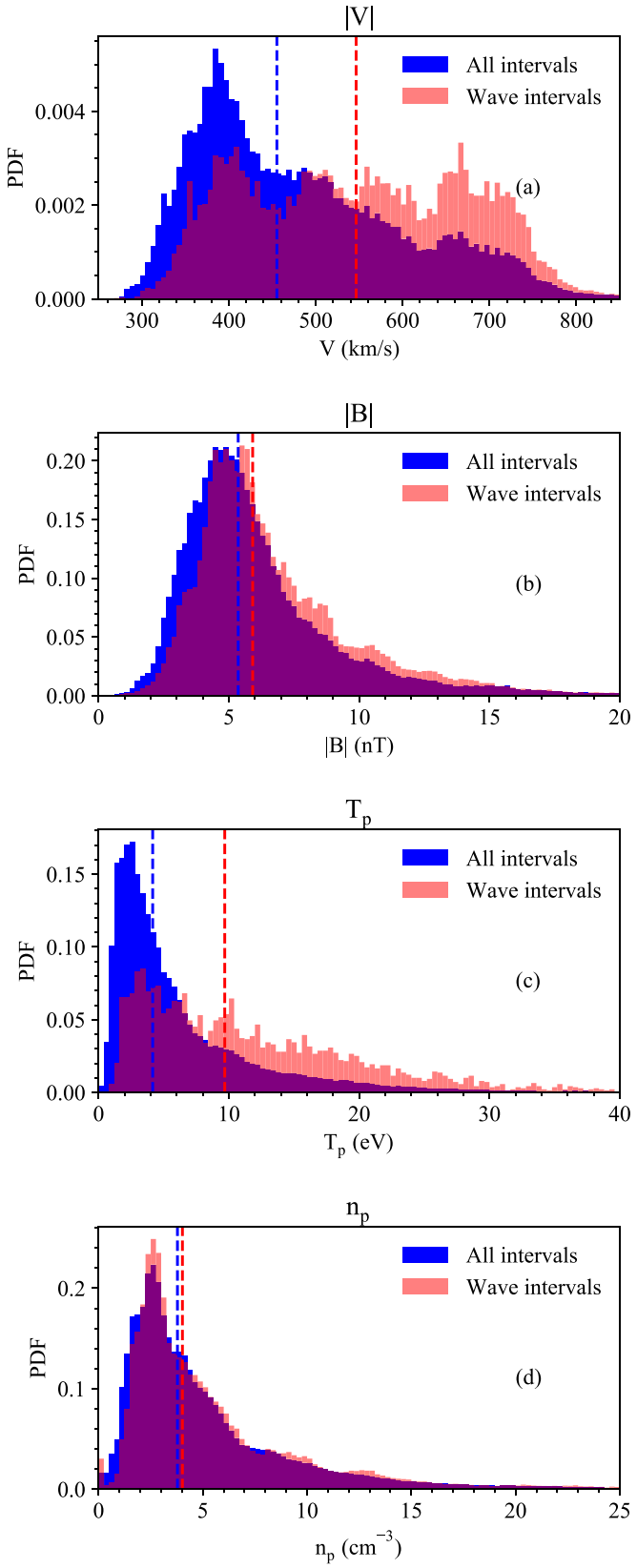


Figure 14. The PDFs for (a) solar wind speed, (b) magnetic field strength, (c) proton temperature, and (d) proton density. The blue distribution corresponds to all intervals; the red distribution corresponds to coherent intervals. Dashed lines correspond to the median of each distribution. Wave intervals tend to occur more often in the fast solar wind and at higher temperatures and magnetic field strengths compared to the rest of the solar wind.

Clearly, one next step will be to apply the neural network to a much longer time span of solar wind data.

Of course, many interesting waves are complex, having multiple frequencies associated with a wave packet. We are hopeful that extension of the neural network framework from this study to allow diagnosis of complex waves will be straightforward for two reasons. First, even with a single bandpass filter, the neural network managed to identify a portion of the complex modes that were outside the scope of the implemented machine-learning problem. At the same time, the network had a low FPR for nonwave examples. This is a sign that the network is learning robustly and in a way that captures circularly polarized wave behavior in a fundamental way. Second, the method used in this paper can be generalized in a straightforward manner to complex waves by passing the time series through narrow bandpass filters. The use of these additional bandpass filters effectively isolates more coherent wave modes that can then be detected by the existing machine-learning approach.

Ultimately, many waves in the solar wind are not circularly polarized. In principle, with a large training set of such waves, there is no obvious reason why a CNN trained to identify other wave polarizations would fail to find such waves. This approach can also be extended to other solar wind data sets. The CNNs are highly adaptable, and starting with a primarily synthetic training set allows for iterative improvement on the network without an overwhelmingly expensive manual approach.

This research was supported by NASA grant Nos. 80NSSC22K1728 and 80NSSC20K0198. We thank Brian A. Thomas for helpful discussions concerning validation of neural networks. We acknowledge high-performance computing support from Cheyenne provided by NCARs CISL, sponsored by the NSF. This research also used NERSC resources, a U.S. DOE Office of Science User Facility operated under contract No. DE-AC02-05CH11231.

Appendix A
Synthetic Data Parameters

To adequately train the neural net to find circularly polarized waves, it is necessary to augment the observational data with synthetic data. To that end, it is important that the synthetic data have a representation of important factors present in the observational data.

Synthetic data generation combines several magnetic field sources. All synthetic intervals have a noise background generated from a power law with a randomized phase. To this noise is added one of three possibilities: a circularly polarized wave, a linearly polarized wave, or no wave. The circularly and linearly polarized waves either have a double Gaussian envelope or a "broken envelope" (a double Gaussian interrupted by a discontinuous change in amplitude, phase, and wave direction). Finally, two additional signals may be added: a logistic step function and random high-frequency noise ("spiky noise"). Note that none, one, or both of these signals may be added. Several examples of synthetic data are shown in Figure 15. These examples contain a range of structures and include both positive and negative training examples.

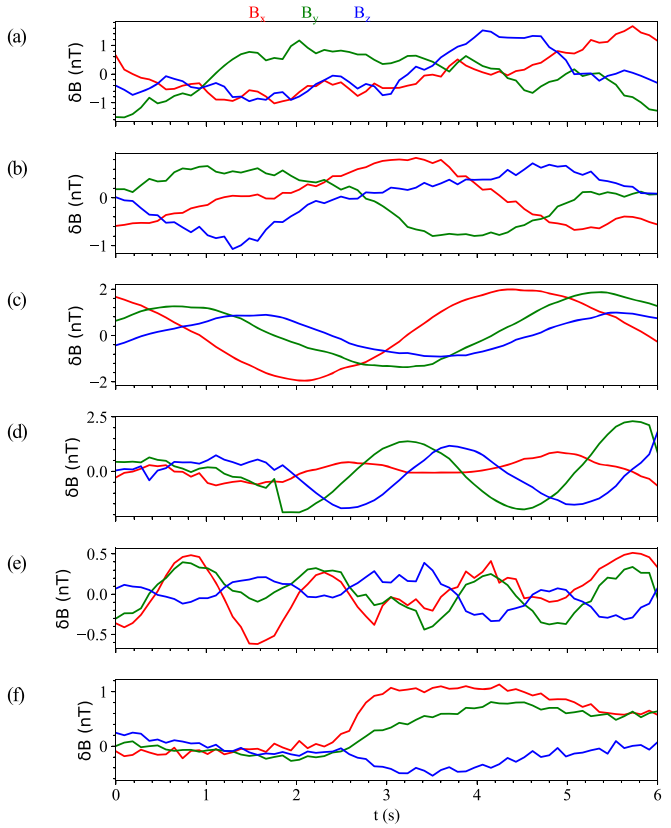


Figure 15. Examples of different classes of synthetic data: (a) noise without additional features, (b) spiky noise, (c) circularly polarized wave mode, (d) circularly polarized wave mode with interruption, (e) linearly polarized wave with interruption, and (f) logistic step function.

The three main pieces of synthetic time-series generation—the noise background, the wave structure, and a step function—are implemented as follows.

(i) The noise background is first generated in frequency space, and then an inverse discrete Fourier transform (DFT) is applied. A power-law noise background is applied with a randomized slope given by

$$\tilde{B}_{n,\alpha}(f) = \begin{cases} \tilde{B}_0 (10\,|f|)^{-p}\ [\cos\theta_\alpha(f) + i\sin\theta_\alpha(f)], \\ \qquad\qquad |f| > 2/11 \\ 10^{-3}\ \text{nT s}, \quad |f| = 2/11 \\ 0\ \text{nT s}, \quad |f| = 0 \end{cases} \tag{A1}$$

for each $\alpha \in \{x, y, z\}$, where $\tilde{B}_n$ is the DFT of the noise piece of the magnetic field, $\tilde{B}_0 \in [0.08, 60]$ nT Hz^{p-1} is the amplitude of the lowest-frequency bin, $f \in [-5.5, 5.5]$ Hz is the range of frequency bins for the DFT, $p \in [-2, -1]$ is the randomized slope of the power law for a given generated example, and $\theta_\alpha(f) \in [0, 2\pi)$ rad is the randomized phase associated with each frequency; note that $\theta_\alpha(-f) = -\theta_\alpha(f)$ to ensure that the inverse DFT results in a real function. To avoid introducing circularly polarized wave modes by chance through choices of random phase, we check the MVA eigenvalue ratios for the noise and omit examples that satisfy the criteria for coherent waves.

A subset of noise examples has additional power added to several higher-frequency bins,

$$\tilde{B}_{s,\alpha}(f) = \begin{cases} \tilde{B}_{n,\alpha}(f), & |f| \notin [f_{s,1}, f_{s,2}] \\ \tilde{B}_{s0,\alpha} + \tilde{B}_{n,\alpha}(f), & |f| \in [f_{s,1}, f_{s,2}] \end{cases}, \quad \text{(A2)}$$

for each $\alpha \in \{x, y, z\}$, where $B_{s,0} \in [0.1, 0.75]$ nT Hz^{p-1} is the randomly determined amplitude of the added power, and $f_{s,1}$, $f_{s,2} \in [3.4, 5.5]$ Hz form the randomly determined frequency range with added power.

(ii) If appropriate to the class of synthetic training data, a wave structure is added to the noise: circularly polarized for positive wave examples or linearly polarized for a subset of nonwave examples. Circularly polarized wave modes are generated with

$$\boldsymbol{B}_{\rm circ}(t) = g_{\rm env}(t)\left(\overleftrightarrow{\boldsymbol{R}}\begin{bmatrix} A_w \cos(2\pi f_w t - \phi_w) \\ A_w \sin(2\pi f_w t - \phi_w) \\ 0 \end{bmatrix}\right), \quad \text{(A3)}$$

where a 3×3 unitary rotation matrix $\overleftrightarrow{\boldsymbol{R}}$ is used to apply a random rotation to the vector in square brackets, $A_w \in [0.1, 2]$ nT is the randomly determined amplitude of the circularly polarized wave, $f_w \in [0.2, 5.5]$ Hz is the randomly determined frequency of the wave, $\phi_w \in [0, 2\pi)$ rad is the randomly determined phase of the wave, and $g_{\rm env}(t)$ is an envelope function, described shortly.

Linearly polarized wave modes $\boldsymbol{B}_{\rm lin}(t)$ are generated with

$$\boldsymbol{B}_{\rm lin}(t) = g_{\rm env}(t)\left(\overleftrightarrow{\boldsymbol{R}}\begin{bmatrix} A_w \cos(2\pi f_w t - \phi_w) \\ 0 \\ 0 \end{bmatrix}\right), \quad \text{(A4)}$$

with the same definitions as in Equation (A3).

The time envelope of the wave $g_{\rm env}(t)$ takes different forms depending on the class of example,

$$g_{\rm env}(t) = g_{\rm env,0}\left(e^{\frac{-(t-\mu_1)^2}{4\sigma_1^2}} + e^{\frac{-(t-\mu_2)^2}{4\sigma_2^2}}\right), \quad \text{(A5)}$$

where $g_{\rm env,0} = e^{(\mu_2-\mu_1)^2/4\sigma_2^2}$ normalizes the function at $t = \mu_1$; μ_1, $\mu_2 \in [0, 6]$ s are the randomly determined envelope means; and σ_1, $\sigma_2 \in [3, 9]$ s are the randomly determined envelope widths.

For broken wave packets,

$$B_{\rm circ}(t) = \begin{cases} B_{\rm circ1}(t), & t \notin [t_{p1}, t_{p2}] \\ A_p\, B_{\rm circ2}(t), & t \in [t_{p1}, t_{p2}], \end{cases} \quad \text{(A6)}$$

where t_{p1}, t_{p2} define the time range of the wave packet break, $B_{\rm circ1}$ and $B_{\rm circ2}$ are waves defined with the same amplitude and frequency but different phase and direction, and $A_p \in [0, 1]$ gives the amplitude of the wave in the vicinity of the break.

(iii) In a subset of both wave and nonwave examples, a logistic step function is added to the signal. Step functions are generated with

$$\boldsymbol{B}_{\rm step}(t) = \frac{A_{\rm step}}{1 + e^{-(t-t_{\rm step})/\sigma_{\rm step}}}\left(\overleftrightarrow{\boldsymbol{R}}\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right), \quad \text{(A7)}$$

where $A_{\rm step} \in [0.1, 3]$ nT is the step amplitude, $t_{\rm step} \in [0, 6]$ s is the center of the step function, $\sigma_{\rm step} \in [0.1, 0.3]$ s determines the width of the step function, and $\overleftrightarrow{\boldsymbol{R}}$ is a randomized unitary rotation matrix.

Appendix B
Network Parameters, Hyperparameters, and Training Plots

The training process itself is dependent on a set of hyperparameters that determine every aspect of training. The set of network parameters and hyperparameters was found by manually modifying each value and testing hundreds of such sets by hand. A learning rate of 8×10^{-4} and a batch size of 10 are used for each network. A rectified linear unit activation function is used for each layer except the output layer. A sigmoid activation function is used for each output layer. The parameters associated with each branch are outlined in Figure 16. For a given N-branch network, each of the branches from 1 to N is included in the network and structured as outlined in Figure 4. The filter sizes s_k take a range of odd values between 1 and 33, and the number of feature maps m_k is either 4, 16, or 32, depending on the kernel size. A kernel size of 1 corresponds to a rescaling of the input magnetic field, while higher kernel sizes correspond to smoothing of the input magnetic field.

Loss and accuracy plots are shown in panels (a) and (b) of Figure 17 for the five-branch network trained with the second fold of the training set (labeled 1DCNN 5B, Fold 2 in Figure 5). The training and validation loss are similar and decrease with training epoch before the validation loss reaches a plateau, indicating that early stopping can be implemented to reduce training time without sacrificing performance.

Loss and accuracy plots are shown in panels (c) and (d) of Figure 17 for the highest-performing network in Figure 7 as functions of training epoch. Based on the results in panels (a) and (b) of Figure 17, early stopping was implemented and conditioned on the training loss with a patience of three training epochs. The patience was determined through trial and error.

Branch #	s_{k1}	s_{k2}	m_{k1}	m_{k2}
1	1	3	4	16
2	5	7	32	32
3	9	11	32	32
4	13	17	32	32
5	19	23	32	32
6	29	31	32	32
7	31	33	32	32

Figure 16. Table of kernel sizes s_k and number of filters m_k for each convolutional branch used in the neural network. A given N-branch network incorporates all branches from 1 to N.

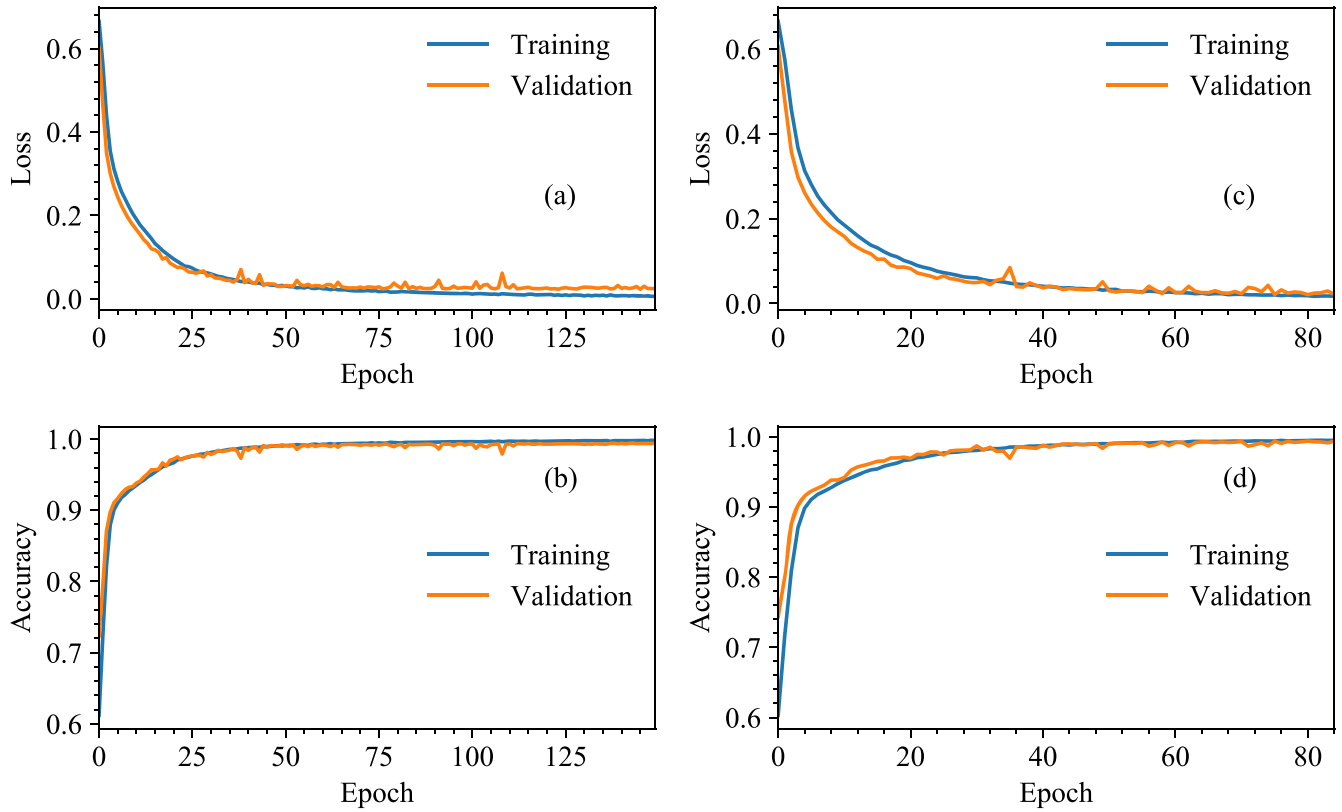


Figure 17. (a) Training/validation loss and (b) training/validation accuracy as functions of training epoch for the five-branch network trained with the second fold of the training set (labeled 1DCNN 5B, Fold 2 in Figure 5) and (c) training/validation loss and (d) training/validation accuracy as functions of training epoch for the five-branch network trained with the highest-accuracy network in Figure 7. The training/validation losses in panel (a) are similar and smoothly decrease with training epoch before reaching a plateau in performance. Similarly, the training/validation accuracy in panel (b) smoothly increases before reaching a plateau. The plateaus in network performance indicate that the network was not overtrained, but further training beyond a certain point does not improve network performance. The random training trials represented in panels (c) and (d) incorporate early stopping. The training/validation losses in panel (c) are similar and smoothly decrease with training epoch, indicating that the network was not overtrained, and the training/accuracy in panel (d) smoothly increases, indicating that the network's classification capability improved with training time.

The training and validation losses are similar and smoothly decrease with training epoch, indicating that the network was not overtrained.

ORCID iDs

Samuel Fordin https://orcid.org/0000-0002-1634-9122
Michael Shay https://orcid.org/0000-0003-1861-4767
Lynn B. Wilson III https://orcid.org/0000-0002-4313-1970
Bennett Maruca https://orcid.org/0000-0002-2229-5618
Barbara J. Thompson https://orcid.org/0000-0001-6952-7343

References

Abdeljaber, O., Avci, O., Kiranyaz, S., Gabbouj, M., & Inman, D. J. 2017, JSV, 388, 154
Acharya, U. R., Oh, S. L., Hagiwara, Y., et al. 2017, Comput. Biol. Med., 89, 389
Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., & Adeli, H. 2018, Comput. Biol. Med., 100, 270
Bruno, R., & Carbone, V. 2013, LRSP, 10, 2
Cattell, C., Breneman, A., Dombeck, J., et al. 2022, ApJL, 924, L33
dos Santos, L. F. G., Narock, A., Nieves-Chinchilla, T., Nuñez, M., & Kirk, M. 2020, SoPh, 295, 131
Gu, J., Wang, Z., Kuen, J., et al. 2018, PatRe, 77, 354
Hasegawa, A. 1976, JGR, 81, 5083
Jagarlamudi, V. K., Dudok de Wit, T., Froment, C., et al. 2021, A&A, 650, A9
Jeong, H.-J., Moon, Y.-J., Park, E., & Lee, H. 2020, ApJ, 903, L25
Jian, L. K., Russell, C. T., Luhmann, J. G., et al. 2009, ApJ, 701, L105
Khrabrov, A. V., & Sonnerup, B. U. 1998, JGR, 103, 6641
Kiranyaz, S., Avci, O., Abdeljaber, O., et al. 2021, MSSP, 151, 107398
Lacombe, C., Alexandrova, O., Matteini, L., et al. 2014, ApJ, 796, 5
LeCun, Y., Bengio, Y., & Hinton, G. 2015, Natur, 521, 436
Lepping, R. P., Acũna, M. H., Burlaga, L. F., et al. 1995, SSRv, 71, 207
Marsch, E., & Chang, T. 1983, JGRA, 88, 6869
Mozer, F. S., Bonnell, J. W., Halekas, J. S., et al. 2021, ApJ, 908, 26
Pinto, V. A., Keesee, A. M., Coughlan, M., et al. 2022, FrASS, 9, 869740
Siciliano, F., Consolini, G., Tozzi, R., et al. 2021, SpWea, 19, e2020SW002589
Smith, A. W., Forsyth, C., Rae, I. J., et al. 2021, SpWea, 19, e2021SW002788
Stansby, D., Horbury, T. S., Chen, C. H. K., & Matteini, L. 2016, ApJ, 829, L16
Stone, M. 1974, J. R. Stat. Soc., B: Stat. Methodol., 36, 111
Telloni, D., Carbone, F., Bruno, R., et al. 2019, ApJ, 885, L5
Tong, Y., Vasko, I., Artemyev, A., Bale, S., & Mozer, F. 2019, ApJ, 878, 41
Unti, T. W. J., & Neugebauer, M. 1968, PhFl, 11, 563
Vech, D., & Malaspina, D. M. 2021, JGRA, 126, e29567
Verniero, J. L., Larson, D. E., Livi, R., et al. 2020, ApJS, 248, 5
Viall, N. M., & Borovsky, J. E. 2020, JGRA, 125, e26005
Wicks, R. T., Alexander, R. L., Stevens, M., et al. 2016, ApJ, 819, 6
Wilson, L., III, Koval, A., Szabo, A., et al. 2012, GeoRL, 39, L08109
Wilson, L., III, Koval, A., Szabo, A., et al. 2017, JGRA, 122, 9115
Wilson, L. B., III, Brosius, A. L., Gopalswamy, N., et al. 2021, RvGeo, 59, e2020RG000714
Wilson, L. B., III, Cattell, C. A., Kellogg, P. J., et al. 2009, JGRA, 114, A10106
Wilson, L. B., III, Cattell, C. A., Kellogg, P. J., et al. 2010, JGRA, 115, A12104
Wu, Q., Sun, Y., Yan, H., & Wu, X. 2020, Comput. Biol. Med., 121, 103800

View publication stats