Radio Frequency Interference Detection in Microwave Radiometry Using Support Vector Machines

Imara Mohamed Nazar and Mustafa Aksoy

Abstract – Herein, a machine-learning–based approach for radio frequency interference (RFI) detection in microwave radiometry is discussed. Features of RFI-free and RFI-contaminated radiometer measurements are different in nature; thus, a classification framework utilizing support vector machines is proposed to detect RFI contamination. The framework has been evaluated with a set of simulations in which RFI-free radiometer measurements are modeled as Gaussian noise, and RFI contamination is simulated by injecting pulsed sinusoidal signals with variable amplitude and duty cycle into the radiometer measurements. The results of these simulations demonstrated the ability of the proposed approach to detect RFI accurately even in low-interference-to-noise-ratio and very short duty-cycle cases, a significant improvement over traditional techniques.

1. Introduction

Although the use of the radio frequency spectrum is regulated in a way that certain frequency bands are solely allocated for microwave radiometry applications, the amount of radio frequency interference (RFI) observed in microwave radiometer measurements is significantly increasing over time [1]. Thus, it is important to design interference detection techniques while keeping track of frequency allocations. So far detection algorithms in spatial, temporal, spectral, statistical, and polarimetric domains have been developed and implemented in radiometer systems [2–6]. The performance of some of these detection algorithms against pulsed sinusoidal interference has also been discussed [7–11]. However, the variability of spectral, temporal, spatial, and statistical properties of RFI signals requires multi-domain approaches for proper RFI detection and mitigation [12]. The National Aeronautics and Space Administration's Soil Moisture Active Passive (SMAP) mission operates one of the first radiometers that implements such a multi-domain RFI detection procedure by combining the outputs of several algorithms running simultaneously in the temporal, spectral, statistical, and polarimetric domains with a logical OR operator [13]. However, SMAP products also have been reported to be susceptible to RFI, especially when the contamination is wideband and continuous [14]. Thus, new approaches, possibly in new domains, are required for RFI detection and mitigation in microwave radiometry.

In our previous work [15], we took an initiative to propose a support vector machine classifier to address the challenging problem of RFI detection in the feature domain through a set of simulations with pulsed sinusoidal interference. In this paper, we analyze the applicability to a radiometer system such as SMAP. Further, we compare our approach to traditional RFI detection methods. Sections 2–5 describe the simulations and operation of the classifier. Then, in Sections 6 and 7, the results are summarized, and conclusions and future work are discussed.

Manuscript received 28 August 2020.

Imara Mohamed Nazar and Mustafa Aksoy are with the Department of Electrical and Computer Engineering, University at Albany–State University of New York, Albany, New York, 12222, USA; e-mail: imohamednazar@albany.edu, maksoy@albany.edu.

2. Simulation Setup

In this study, RFI in the radiometer baseband was considered to be a pulsed sinusoidal signal with 12 MHz frequency, 1 V amplitude, 300 μs pulse repetition interval, and variable duty cycle (DC). To simulate RFI-contaminated radiometer measurements, this RFI was injected into a Gaussian noise with variable standard deviation, which represents radiometer measurements. In order to analyze how the performance of the proposed algorithm changes with DC and interference-to-noise-ratio (INR) cases, the DC of the interference signal was varied from 1% to 100% in increments of 1%, and the INR was increased from –20 dB to 10 dB in steps of 1 dB. Notice that since the amplitude of the interference signal was kept constant at 1 V, the INR was changed by altering the standard deviation of the Gaussian noise. Furthermore, it was assumed that the baseband radiometer measurements were sampled at a 96 MSPS rate and the integration time was 300 μs, similar to that of the SMAP radiometer [13].

RFI-free and RFI-contaminated radiometer measurements were generated as explained above, and parameters that characterize them, which are known as "features," were extracted from each 300 μs radiometer integration periods at multiple instances. Features are the root-mean-square power, mean, variance, ratio between the mean and the variance, skewness, kurtosis, maximum amplitude, minimum amplitude, and the number of peaks, which are the positive voltage samples with zero gradients.

3. Support Vector Machine

We consider a set S of integration periods. Each period $s \in S$ is denoted by a d-dimensional feature vector $\mathbf{x} = \{z_1, \ldots, z_d\}$ called the support vector where d is the number of features. Furthermore, each radiometer integration period s belongs to one of N number of classes. In this study, N is two, namely, RFI-free and RFI-contaminated measurements. A set of class labels $y = \{-1, 1\}$ is defined to denote these RFI-free and RFI-contaminated classes, respectively. The support vector machine (SVM) [16] denotes the class of hyperplanes of $\langle \mathbf{w}, \mathbf{x} \rangle + b = 0$, where $\mathbf{w}, \mathbf{x} \in R^d$ and $b \in R$. It is possible to prove that the optimal hyperplane is the one with a maximal margin of separation between the two classes. The optimization problem for a maximal margin can be formulated as follows:

$$\min \langle \mathbf{w}, \mathbf{w} \rangle, \quad \text{s.t } y_s[\langle \mathbf{w}, \mathbf{x}_s \rangle + b] \geq 1 \tag{1}$$

where $\mathbf{x}_s$ and y_s denote the feature vector and the class label of integration period s, respectively. Consequently, the final decision function is given by

$$f(x) = \text{sgn}(\langle \mathbf{w}, \mathbf{x} \rangle + b) \tag{2}$$

4. Feature Selection

In order to reduce the computational cost, a smaller subset of aforementioned features can be selected without undermining the detection performance. In this work, a similarity-based filter-type feature selection method, the Fisher score, was used [17]. The Fisher score is easy to implement and computationally efficient. This method selects features considering that the feature values of integration periods that belong to the same class are similar, whereas the feature values of integration periods from different classes are dissimilar. The Fisher score for each feature $z_i, i = 1, \ldots, d$ is defined as follows:

$$\text{FS}_{z_i} = \frac{\sum_{j=1}^{N} n_j (\mu_{ij} - \mu_i)^2}{\sum_{j=1}^{c} n_j \sigma_{ij}^2} \tag{3}$$

where N, n_j, μ_i, μ_{ij}, and σ_{ij}^2 denote the number of classes, number of samples in class j, mean value of feature z_i, mean value of feature z_i for samples in class j, and variance of feature z_i for samples in class j, respectively. A larger Fisher score value for a feature implies greater distinctiveness.

5. RFI Detection

The process of the proposed RFI detection approach is shown in Figure 1. Initially, the features listed in Section 2 were extracted from the radiometer measurements within each integration period and stored in a data set. Then these data were divided into training and testing data sets. The training data set was used to train the SVM model, after applying the feature selection process described in Section 4. Finally, the corresponding features were taken from test data and then imported into the trained SVM model to label the integration periods as RFI-free or RFI-contaminated.

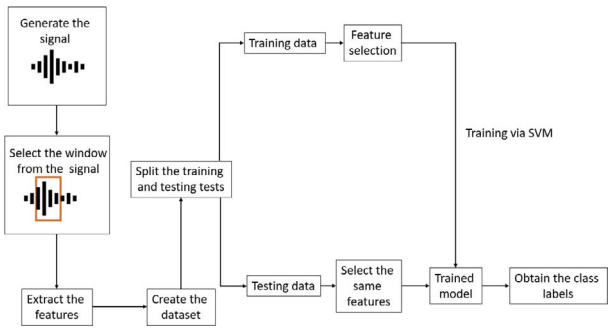

Figure 1. Interference detection process.

6. Results

We created 3100 data sets to analyze the effect of INR and DC on RFI detection performance. Each data set, associated with a pair of INR and DC values, consisted of 400 radiometer integration periods, which belong to either RFI-free or RFI-contaminated classes. In order to avoid over-fitting, fivefold cross-validation was implemented in which the data set was divided into five folds, and each fold was tested by the model trained by the remaining folds. Thus, in each fold, the SVM model was trained on 320 integration periods and tested on 80 integration periods. In each period, relevant features were selected based on the algorithm discussed in Section 4. When multiple features had the same Fisher score, only one them was selected for RFI detection. Also, features that produced indefinite Fisher score values were discarded.

The RFI detection performance was evaluated based on the accuracy, precision, and recall parameters, which are defined as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{4}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{6}$$

where TP is the number true positives (the number of RFI-contaminated integration periods identified as RFI-contaminated); TN is the number of true negatives (number of RFI-free integration periods identified as RFI-free); FP is the number of false positives (number of RFI-free integration periods identified as RFI-contaminated); and FN is the number of false negatives (number of RFI-contaminated integration periods identified as RFI-free).

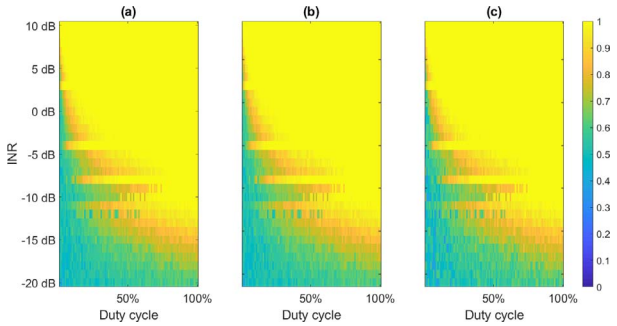

Figure 2. (a) Accuracy, (b) precision, and (c) recall of the proposed RFI detection method as a function of the INR and DC.

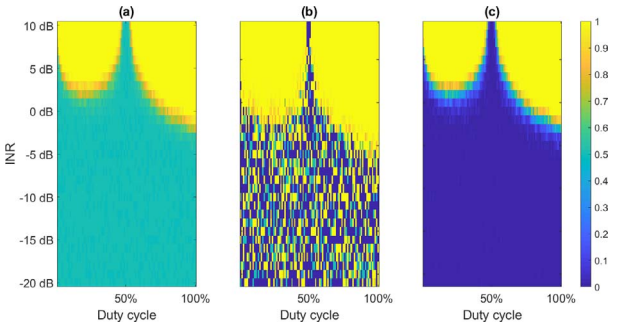

Figure 4. (a) Accuracy, (b) precision, and (c) recall of the kurtosis-detection method as a function of the INR and DC.

Figure 2 illustrates the accuracy, precision, and recall of the proposed RFI detection method as a function of INR and DC. It can be observed that, in general, the accuracy, precision, and recall values increased with increasing INR and DC. More than 95% accuracy was achieved even for very low (e.g., –10 dB) INR rates provided that the DC was larger than 50%. This indicates the performance of the proposed algorithm against continuous RFI contamination. For a lower DC, i.e., short-duration RFI contamination, cases, the accuracy diminished, but an acceptable detection performance still could be reached if the INR was larger than 0 dB. Slightly higher recall values, i.e., the fraction of detected RFI-contaminated periods out of all RFI-contaminated integration periods, especially for low INR cases, meant that failure to detect RFI contamination happened less often. Finally, high-precision values, i.e., the ratio of the number of RFI-contaminated integration periods to the number of periods labeled RFI-contaminated, implied low false alarm rates for the proposed approach when accuracy and recall rates were high. Thus, the trade-off between RFI detection and false alarm rates was found to be minimal.

Figure 3 shows the number of data sets for which each feature was used for RFI detection in this study. As seen from the figure, the minimum amplitude, maximum amplitude, and number of the peaks were used more frequently than the other features. It is also observed that the root-mean-square power and the variance were used mostly in extreme (too low or too high) INR cases. Furthermore, the kurtosis was utilized in high-INR cases.

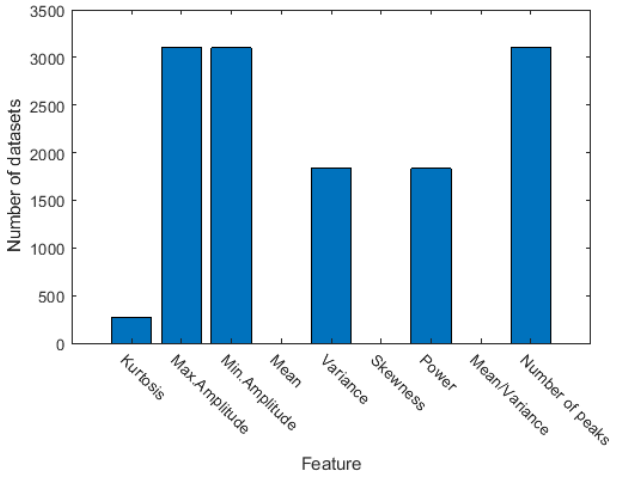

Figure 3. Frequency of features used in data sets.

The performance of our proposed RFI detection method was compared against traditional RFI detection algorithms such as kurtosis-detection, time-domain pulse-blanking, and the kurtosis-detection OR time-domain pulse-blanking. For geophysical emissions, the kurtosis estimate itself was a Gaussian random variable with a mean of three. On the other hand, for non-Gaussian RFI-contaminated signals the kurtosis value deviated from three except for RFI with 50% DC where the kurtosis was exactly three [3]. In order to detect RFI-contaminated signals, we calculated the mean and standard deviation of the RFI-free signals and set the range of the RFI-contaminated signals to be three standard deviations apart from the mean while allowing only 0.3% of false alarm. Figure 4 shows the accuracy, precision, and recall of the kurtosis detection method as a function of the INR and DC. As seen in the figure, the kurtosis detection method was efficient for pulsed sinusoidal interference with INR $>$ 0 dB except interference with 50% DC.

In the time-domain pulse-blanking method, received powers that exceed a certain range above or below the estimated RFI-free power levels were labeled as RFI-contaminated. In order to decide the threshold for our data set, the mean and the standard deviation of the received power for RFI-free integration periods were calculated. Power levels that fell from the mean by three standard deviations were labeled as RFI-contaminated, again allowing only 0.3% of false alarm. Figure 5 shows the performance of the time-domain pulse-blanking detection method as a function of the INR and DC. The figure shows that the method performed well in detecting RFI with the INR between –10 dB and 5 dB depending on the DC.

The "kurtosis-detection OR time-domain pulse-blanking" method combined the kurtosis-detection and the time-domain pulse-blanking detector outputs. This was a direct result of the labels retrieved from both

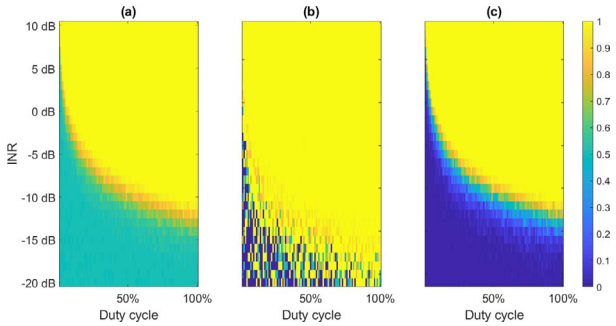

Figure 5. (a) Accuracy, (b) precision, and (c) recall of the pulse-blanking detection method as a function of INR and DC.

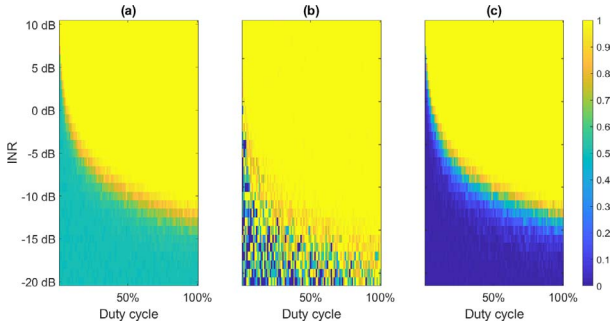

Figure 6. (a) Accuracy, (b) precision, and (c) recall of the OR method as a function of the INR and DC.

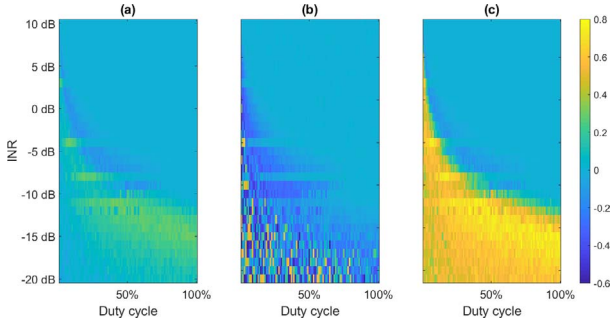

Figure 7. (a) Accuracy, (b) precision, and (c) recall difference between the proposed RFI detection method and the OR method as a function of the INR and DC.

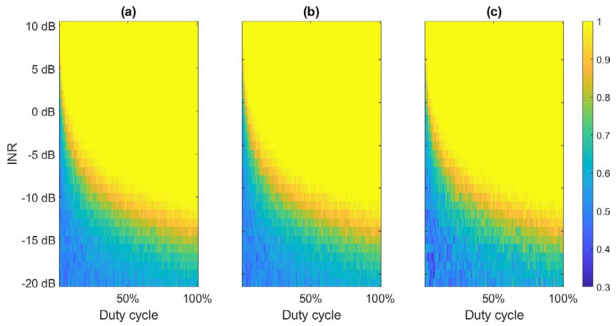

Figure 8. (a) Accuracy, (b) precision, and (c) recall of the proposed method that uses kurtosis and the received signal power as features as a function of the INR and DC.

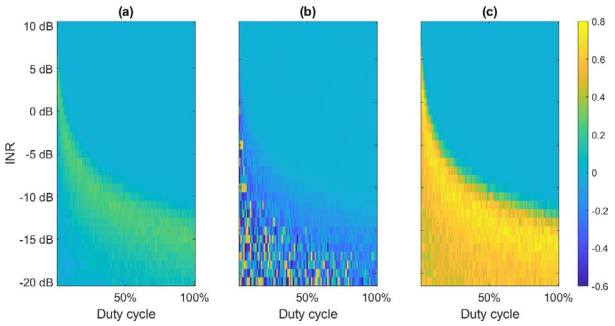

Figure 9. (a) Accuracy, (b) precision, and (c) recall difference between modified classifier and the OR method as a function of the INR and DC.

detectors combined with a logical OR operator (from here onward, we will refer this algorithm as the "OR method"). Figure 6 shows the performance of the OR method as a function of the INR and DC. The figure shows that the performance of the method was similar to that of the time-domain pulse-blanking algorithm.

Among all three baselines, the OR method performed better than kurtosis detection and pulse blanking. Therefore, we compared the accuracy, precision, and recall differences between our method and the OR method. Here the difference between our method and the OR method was calculated by subtracting the performance metrics of the OR method from our method. Figure 7 shows this difference as a function of the INR and DC. As shown in the figure, when compared with the OR method, our method, in general, achieved higher accuracy, precision, and recall, especially for the low-INR and high-DC RFI cases. Notably, higher recall values, i.e., the fraction of detected RFI-contaminated periods out of all RFI-contaminated integration periods, in low-INR cases indicated that our method performs better in detecting RFI contamination from all the available RFI-contaminated integration periods than the OR method. Here we need to mention that the OR method may outperform our method for high-INR and low-DC cases because the kurtosis and pulse-blanking algorithms are specifically designed to detect such RFI.

Furthermore, we modified our method so that it used only kurtosis and the received signal power as features. Figure 8 illustrates the performance of the modified classifier. We also compared this modified classifier with the OR method, and the performance differences between them are shown in Figure 9. As seen in the figure, the modified classifier performed even better by eliminating the weaknesses against the OR method in low-DC and high-INR cases. Thus, the performance of the classifier can be improved by an informed selection of features.

7. Conclusions and Future Work

In this study, an SVM classifier is introduced as a new approach for RFI detection in microwave radiometry through a set of simulations. The results illustrate

the ability of this method in detecting pulsed sinusoidal interference. Furthermore, the distinctive features for RFI detection are demonstrated for the proposed classifier. The performance comparison of the proposed approach with traditional RFI detection algorithms shows the greater ability to detect RFI, particularly in low-INR cases, and the importance of intelligent feature selection. In the future, we plan to implement this approach in real radiometer measurements. Further, this study focuses on pulsed sinusoidal RFI and relevant features, and we plan to expand our investigations by adding more features that can characterize other types of RFI.

8. References

1. D. McKague, J. Puckett, and C. Ruf, "Characterization of K-Band Radio Frequency Interference From AMSR-E, WindSat and SSM/I," 2010 IEEE International Geoscience and Remote Sensing Symposium, Honolulu, HI, USA, July 25–30 2010, pp. 2492-2494.
2. S. Misra and P. De Matthaeis, "Passive Remote Sensing and Radio Frequency Interference (RFI): An Overview of Spectrum Allocations and RFI Management Algorithms [Technical Committees]," *IEEE Geoscience and Remote Sensing Magazine*, **2**, 2, June 2014, pp. 68-73.
3. R. D. De Roo and S. Misra, "Effectiveness of the Sixth Moment to Eliminate a Kurtosis Blind Spot in the Detection of Interference in a Radiometer," IGARSS 2008—2008 IEEE International Geoscience and Remote Sensing Symposium, Boston, MA, USA, July 7–11, 2008, pp. II-331-II-334.
4. B. Guner, J. T. Johnson, and N. Niamsuwan, "Time and Frequency Blanking for Radio-Frequency Interference Mitigation in Microwave Radiometry," *IEEE Transactions on Geoscience and Remote Sensing*, **45**, 11, November 2007, pp. 3672-3679.
5. C. Ruf and S. Misra, "Detection of Radio-Frequency Interference for the Aquarius Radiometer," *IEEE Transactions on Geoscience and Remote Sensing*, **46**, 10, October 2008, pp. 3123-3128.
6. J. M. Tarongi and A. Camps, ''Radio Frequency Interference Detection Algorithm Based on Spectrogram Analysis," 2010 IEEE International Geoscience and Remote Sensing Symposium, Honolulu, HI, USA, July 25–30, 2010, pp. 2499-2502.
7. R. D. De Roo, S. Misra, and C. S. Ruf, "Sensitivity of the Kurtosis Statistic as a Detector of Pulsed Sinusoidal RFI," *IEEE Transactions on Geoscience and Remote Sensing*, **45**, 7, July 2007, pp. 1938-1946.
8. J. T. Johnson and L. C. Potter, "A Study of Algorithms for Detecting Pulsed Sinusoidal Interference in Microwave Radiometry," IGARSS 2008—2008 IEEE International Geoscience and Remote Sensing Symposium, Boston, MA, USA, July 7–11 2008, pp. II-161-II-164.
9. B. Guner, M. T. Frankford, and J. T. Johnson, "A Study of the Shapiro–Wilk Test for the Detection of Pulsed Sinusoidal Radio Frequency Interference," *IEEE Transactions on Geoscience and Remote Sensing*, **47**, 6, June 2009, pp. 1745-1751.
10. R. D. De Roo and S. Misra, "A Moment Ratio RFI Detection Algorithm That Can Detect Pulsed Sinusoids of Any Duty Cycle," *IEEE Geoscience and Remote Sensing Letters*, **7**, 3, July 2010, pp. 606-610.
11. S. Misra and C. Ruf, "Inversion Algorithm for Estimating Radio Frequency Interference Characteristics Based on Kurtosis Measurements," 2009 IEEE International Geoscience and Remote Sensing Symposium, Cape Town, South Africa, July 12–17, 2009, pp. II-162-II-165.
12. M. Aksoy, J. T. Johnson, S. Misra, A. Colliander, and I. O'Dwyer, "L-Band Radio-Frequency Interference Observations During the SMAP Validation Experiment 2012," *IEEE Transactions on Geoscience and Remote Sensing*, **54**, 3, March 2016, pp. 132-1335.
13. J. R. Piepmeier, J. T. Johnson, P. N. Mohammed, D. Bradley, C. Ruf, et al., "Radio-Frequency Interference Mitigation for the Soil Moisture Active Passive Microwave Radiometer," *IEEE Transactions on Geoscience and Remote Sensing*, **52**, 1, January 2014, pp. 761-775.
14. P. N. Mohammed, M. Aksoy, J. R. Piepmeier, J. T. Johnson, and A. Bringer, "SMAP L-Band Microwave Radiometer: RFI Mitigation Pre–Launch Analysis and First Year On-Orbit Observations," *IEEE Transactions on Geoscience and Remote Sensing*, **54**, 10, October 2016, pp. 6035-6047.
15. I. M. Nazar and M. Aksoy, "Radio Frequency Interference Detection in Microwave Radiometry Using Support Vector Machines," URSI GASS 2020, Rome, Italy, August 29–September 5, 2020, https://www.ursi.org/proceedings/procGA20/papers/PID6352419.pdf (Accessed 15 November 2020).
16. C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning*, **20**, 1995, pp. 273-297.
17. J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu, "Feature Selection: A Data Perspective," *ACM Computing Surveys*, **50**, 6, December 2017, pp. 1-45.