The RFI 2022 Workshop at ECMWF, 14 – 18 February 2022
2022 URSI International Union of Radio Science

PSRFINET: Radio Frequency Interference Detection in Pulsar Data with Deep Residual Networks

A. Hamid(1), W. van Straten(1), and A. Griffin(2)
(1) Institute of Radio Astronomy and Space Research, Auckland University of Technology
(2) High Performance Computing Research Laboratory, Auckland University of Technology

Abstract

Radio Frequency Interference (RFI) is a hindrance to high-precision pulsar timing experiments aimed at detecting the stochastic gravitational wave background. Thresholds set by linear combinations of statistical quantities are among the most common approaches to RFI flagging of folded pulse profiles. We propose a deep convolutional neural network approach to RFI flagging called PSRFINET that treats two-dimensional arrays of pulse profiles (rotational phase versus radio frequency) as images and performs feature learning on labelled RFI samples. We train and validate multiple deep residual neural networks on many hours of pulsar observations (thousands of 8 second sub-integrations) of MeerKAT L-band data where the ground truth is generated from Clfd and Coastguard software packages for RFI mitigation. A method of combining the separate ground truths aimed at enhancing the RFI mitigation capabilities of the networks is also explored. The performance of the networks was evaluated by examining the classification metrics of area under the curve of the receiver operating characteristic (AUROC), Precision-Recall (PR) and F1 scores. Our preliminary results show an AUROC of more than 0.91 and PR of 0.67 which indicates that although the neural networks are capable of distinguishing between clean and corrupted frequency channels, precision and recall scores are limited by a class imbalance of a small amount of RFI with respect to clean channels. We also discuss our approach to develop a statistical objective figure of merit for evaluating and comparing the effectiveness of different RFI flagging approaches in the data.

1 Introduction

Radio pulsar astronomy is capable of providing incredible insight on various astrophysical phenomena such as gravitational waves and the behaviour of diffuse interstellar matter [1, 2]. The next generation of radio telescope arrays—with a prime example being the Square Kilometre Array (SKA)—is set to improve our understanding of these phenomena through large high quality datasets. However, unwanted emission from local sources of Radio Frequency Interference (RFI) degrade the overall quality of the data. The precision timing of pulsars—an experiment which demands high sensitivity—can be extremely vulnerable to bursts of RFI on the order of microseconds [3]. RFI can be emitted intentionally from artificial sources such as communication satellites, GSM, and air traffic control towers and unintentionally from onsite electrical equipment. Natural RFI sources include emission from lightning and solar weather [4]. RFI can be either impulsive or periodic in nature, however, in reality their signals can take a wide range of frequency and time dependencies lasting from a few nanoseconds to several hours [5].

The most straightforward and effective method of countering the effects of RFI is to construct radio observatories within a dedicated radio quiet zone (RQZ) where frequency bands relevant to radio astronomy are protected by legislation. One such RQZ is located in the Karoo region where the mid frequency portion of the SKA will be hosted [6]. However, a terrestrial RQZ cannot protect an observatory from interference generated by constellations of satellites in orbit. RFI mitigation techniques are henceforth considered in the pre-detection and post-detection stages of a radio astronomical pipeline. Although pre-detection methods are powerful, such processes typically must be performed online and require specialized hardware or additional computing resources [7]. Residual RFI is usually handled in the post-detection stage where flagging or masking of frequency channels is done. Manual flagging of RFI in post-detection data artefacts becomes difficult if not impossible to carry out as datasets increase in size. This paper proposes a method we call PSRFINET used to identify and flag RFI in pulsar data using a supervised deep learning approach based on a variant of a deep convolutional neural network (CNN). The data examined in this work are from the MeerTime Key Science Project on pulsar timing on MeerKAT [8].

Supervised deep learning has been previously used to mask RFI in radio data and continues to show much promise. The first published work was seen in [10] which implements a CNN architecture with a downsampling path and upsampling path known as the U-net used in the task of semantic segmentation of RFI on a 2D plane of time and frequency. The approach was followed by RFI-Net for the FAST radio telescope using a network with two residual units to improve RFI detection with less training data and to reduce the number of false positives [11]. A U-net with dilated convolution has been explored with the aim of expanding the receptive field of the network without changing the size of the output [12]. An approach called the R-Net used transfer learning on a network trained on MeerKAT data before training the last few layers on human labelled KAT 7 data to surpass the performance of the original U-Net and AOFlagger [7, 13]. According to the authors, transfer learning solves the problem of benchmarking the performance of the R-Net [13]. CNNs have also been used to classify transient RFI in time domain signals [14]. Supervised deep learning of neural networks have been used for candidate selection in pulsar searches [15, 16]; however, to date there has been no published attempt to use CNNs for RFI flagging in pulsar data.

This paper is organized as follows: in Sec. 2 we describe the data preparation, the network architecture, and the training of the network used to learn the features of RFI, in Sec. 3 the performance of the network on a test set of data is presented in terms of the standard practice metrics of binary classification as well as introducing our own metric based on dimensionality reduction and Gaussian mixture modelling to compare the effectiveness of the trained networks against the original ground truth; and in Sec. 4 we present our conclusions and proposed future work.

2 Methodology

2.1 Data Preparation

The type of pulsar data examined in this work are folded data where Stokes parameters of hundreds to thousands of individual pulses are averaged as a function of pulse longitude to create a time-frequency-pulse phase and polarisation data cube. The MeerTime project has already recorded a large dataset of hundreds of thousands of folded pulsar archives from more than 1500 pulsars that we draw upon as training data for PSRFINET. Since RFI is not likely to be weighted to any pulsar, a random selection of data from all pulsars is used. In order to capture a homogeneous sample space for the training data, the selected archives are initially constrained to single sub-integrations of 8-second time duration, L-band receiver data with 1024 frequency channels, and signal-to-noise ratio (SNR) between 4 and 14.

To generate labelled samples of channels flagged as RFI, we use two RFI mitigation software packages: Clfd [17] and Coastguard [18]. Clfd detects outliers depending on whether pulse profiles possess features outside an acceptable range. The profile feature parameters we select are; standard deviation, peak-to-peak difference, and amplitude of the second bin of the Fourier transform of the profile. The free parameter q that is mapped to a false rejection probability when most of the data follow a normal distribution is set to 1.75. For Coastguard, only the surgical and bandwagon cleaners are used. The surgical cleaner evaluates 4 metrics from pulsar free residuals calculated from the integrated pulse profile. The metrics are standard deviation, mean, range, and maximum amplitude of the Fourier transform of the mean-subtracted residuals computed for each sub-integration and or channel pair. Any profiles with two or more of the metrics greater than $N\sigma$ from the median of either their sub-integration or channel are flagged as RFI. The threshold N is set to 5 and the edge cut parameter is set to 10%. The ground truth used as target data are binary masks of length 1024 where a flag of 1 is a clean channel and a flag of 0 is for RFI. Figure 1 shows an example and comparison between channel masks of Coastguard and Clfd for a sample sub-integration.

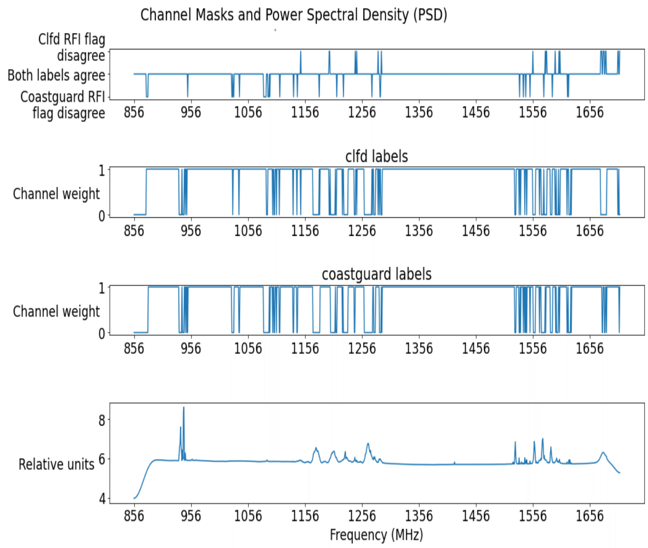

Figure 1. Power spectral density of the pulsar J1651-7642 recorded on 2020-08-15 at 16:45:28 UTC (bottom) with the Clfd and Coastguard masks of RFI channels (two middle). A breakdown of the channel masks (top) shows the frequency channels where both ground truth are in agreement and disagreements where Clfd detects RFI and Coastguard does not and where Coastguard detects RFI and Clfd does not. For this example, Coastguard is the more aggressive ground truth which may not hold for all cases and is subject to the tuning of the input of both algorithms. The L band of MeerKAT (856 - 1712 MHz) is mostly impacted by global positioning and mobile communications systems [9].

2.2 Network Architecture

PSRFINET uses the residual network architecture [19]. Since deep neural networks become difficult to train because of the vanishing gradient problem, skip connections between layers were introduced to help backpropagate the gradient to earlier layers [20]. The ResNet50 model was chosen for its speed, simplicity, and known performance [21]. We change the network input by discarding 20 channels from the high frequency and low frequency band edges resulting in a 984 x 1024 x 4 sized input layer for 1024 phase bins and 4 polarizations. The Resnet50 is further composed of convolution, activation, and pooling layers as detailed in Figure 2. A dropout layer is applied within the dense, fully-connected output to reduce overfitting [22].

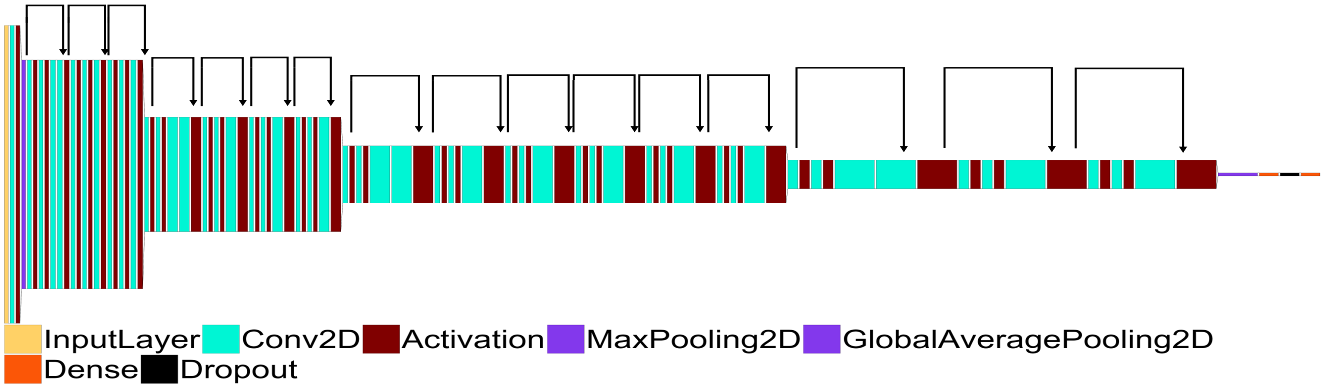

Figure 2. Architecture of PSRFINET adapted from the ResNet50. The black arrows represent the skip connections used to backpropagate the gradient. The image is down-sampled after the first block of 3 residual units. Zero padding, batch normalization, and add layers have been hidden for simplicity. The output of the network is 984 x 1.

2.3 Training the network

7056 pulsar archives are used for training, validation, and testing which corresponds to a time duration of more than 15 hours of observation. The data are split into 60 % training, 20 % validation, and a 20 % hold out test set. The batch size is set to 2 and training is done in mini batch mode to increase speed. The learning rate is 0.0045, a value that was acquired through a manual iterative process from an initial value of 0.01. The network is trained to a maximum of 100 epochs where one epoch is a single forward and backward pass of the training data. Early stopping is implemented such that training is stopped and the best model is selected when there is no change in the loss function of the validation data after 10 consecutive epochs. A binary cross entropy loss function is minimized with the Adam optimizer for stochastic gradient descent [23]. Since the distribution of RFI corrupted channels to clean channels is an imbalanced class problem with nearly one RFI channel for every 5 clean channels in the band, we must assign class weights to the neural network model. We use the following equation to compute the class weights to fit the model:

$$w_{i,j} = \frac{N_{\text{samp}}}{N_{\text{class}} N_{i,j}} \quad (1)$$

where N_{samp} is the number of samples, N_{class} is the number of classes, and $N_{i,j}$ is the number occurrences for each class. The class weights of the model are set to {0: 3.163, 1: 0.594} for Coastguard, and {0: 3.882, 1: 0.574} for Clfd.

Five-fold cross validation is used, where the training and validation data are split into 5 sets with the validation data taking a different fraction in each fold. The best-performing model from the 5 folds is chosen based on the lowest validation loss at the end of each fold. Models are trained and tested using different data sets as summarized in Table 1. Data augmentations are applied to increase the size of the training dataset [24]. The augmentations applied are; image contrast adjustment, addition of white noise to all pixels, a horizontal flip, and translation along the phase axis.

Table 1. Network name, ground truth, and test data.

Network model	Training data	Test data	Max F1
Clfd net	100% Clfd	100% Clfd	0.62
Cg net	100% Cg	100% Cg	0.69
mix net (Clfd)	50% Clfd 50% Cg	100% Clfd	0.61
mix net (Cg)	50% Clfd 50% Cg	100% Cg	0.69
mix net (mix)	50% Clfd 50% Cg	50% Clfd 50% Cg	0.65

3 Results

A hold out set of 1176 sub-integrations is allocated to evaluate the neural network models. Predictions made by each of the three trained models in Table 1 are evaluated using area under the curve of the receiver operating characteristic (AUROC), area under Precision-Recall (PR) curve, and the F1 score. AUROC plots the rate of true positives (TPR) versus the rate of false positives (FPR) which are given by:

$$\text{TPR} = \frac{\text{TP}}{\text{TP+FN}} \quad (2)$$

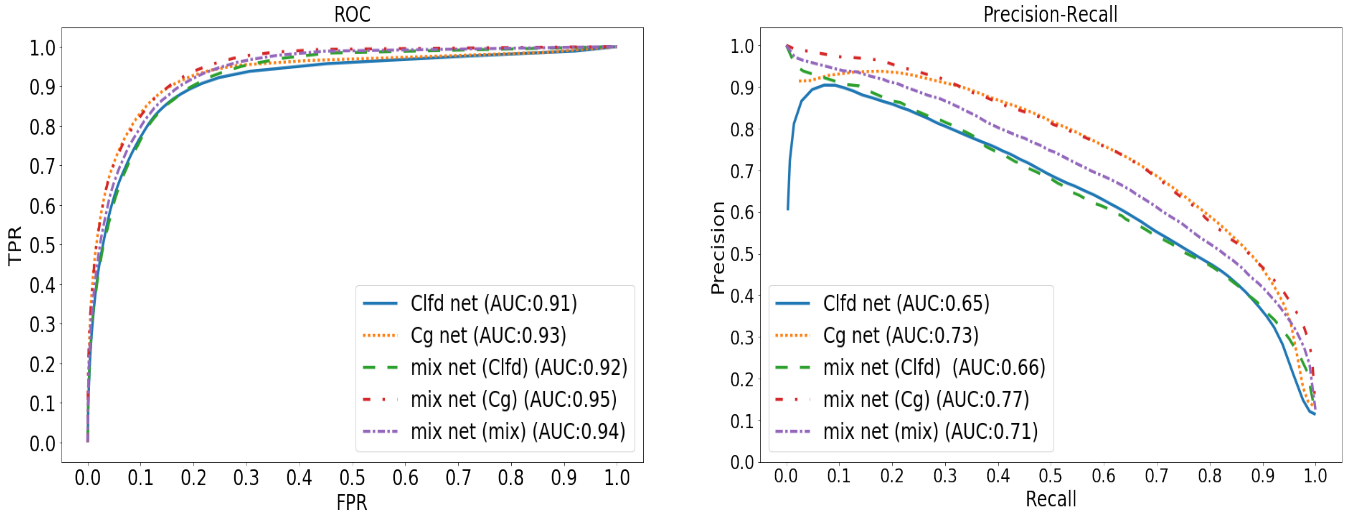

Figure 3. AUROC and PR curves for 3 different PSRFINET models with test data as described in Table 1. AUC is the area under the respective curve.

$$\text{FPR} = \frac{\text{FP}}{\text{TN+FP}} \tag{3}$$

where TP is the true positives and TN the true negatives, and FN is the false negatives and FP the false positives.

The PR curve is a more sensitive metric to imbalance between classes. The F1 score combines Precision and Recall into a single metric by taking their harmonic mean. The Precision, Recall and F1 score are given by:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{4}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{5}$$

$$\text{F1 score} = \frac{2\,(\text{Precision} \times \text{Recall})}{\text{Precision + Recall}} \tag{6}$$

The results of the AUROC and area under the PR curve for the three networks are shown in Figure 3. The AUROC metric produce near similar results for all networks, however, the PR curve shows more variation. From the PR scores we can see that networks trained and tested on Clfd score the lowest. We interpret this as Clfd being the more difficult ground truth to learn in comparison to Coastguard. The maximum F1 Scores for all the networks are shown in Table 1. Using the maximum of the F1 score we are able to find an ideal threshold that balances precision and recall. We use a prediction threshold value of 0.67 averaged from all maximum F1 scores to be used to zero weight channels containing RFI for further testing in the next section.

3.1 RFI mask metric

Motivated by the problem where some bright pulsars can be unintentionally zapped, we have developed a custom distance metric to score the efficacy of RFI masking by modelling the distribution of folded profile amplitudes and quantifying the presence of anomalous features. Computation of the metric starts by integrating the mean profile (over all frequency channels and sub-integrations), $\vec{S} = \langle\vec{P}\rangle$, and then (for each frequency channel and sub-integration) computing the residual profile

$$\vec{R} = \vec{P} - a\vec{S} + b\vec{1} \tag{7}$$

after an ordinary least squares fit for the scale a and offset b. From the residuals, the covariance matrix

$$\mathbf{C} = \langle\vec{R} \otimes \vec{R}^T\rangle\langle\vec{R}\rangle \otimes \langle\vec{R}^T\rangle \tag{8}$$

is calculated and decomposed to yield its eigenvalues λ_j and eigenvectors $\vec{v}_j$. To reduce the dimensionality of the data, only the n eigenvalue/vector pairs with an eigenvalue greater than the variance of the off-pulse phase bins are kept.

The distribution of the principal components of the n kept eigenvectors, $\rho_j = \vec{v}_j \cdot \vec{R}$ are then modelled as a mixture of n-dimensional Gaussians. The optimal number of Gaussians is determined using the Akaike information criterion (AIC) [25].

For each data file, the above procedure is repeated and a Gaussian Mixture Model (GMM) is produced for each of the five masks produced by Clfd, Coastguard, and the Resnet50 trained using Clfd, Coastguard, and a mixture of the two. The GMM with the smallest number of Gaussian components is selected based on the reasoning that the simplest model reflects the smallest number of undetected/unflagged outliers. Using the best GMM, the log-likelihood, log(L), of the five flag sets is computed.

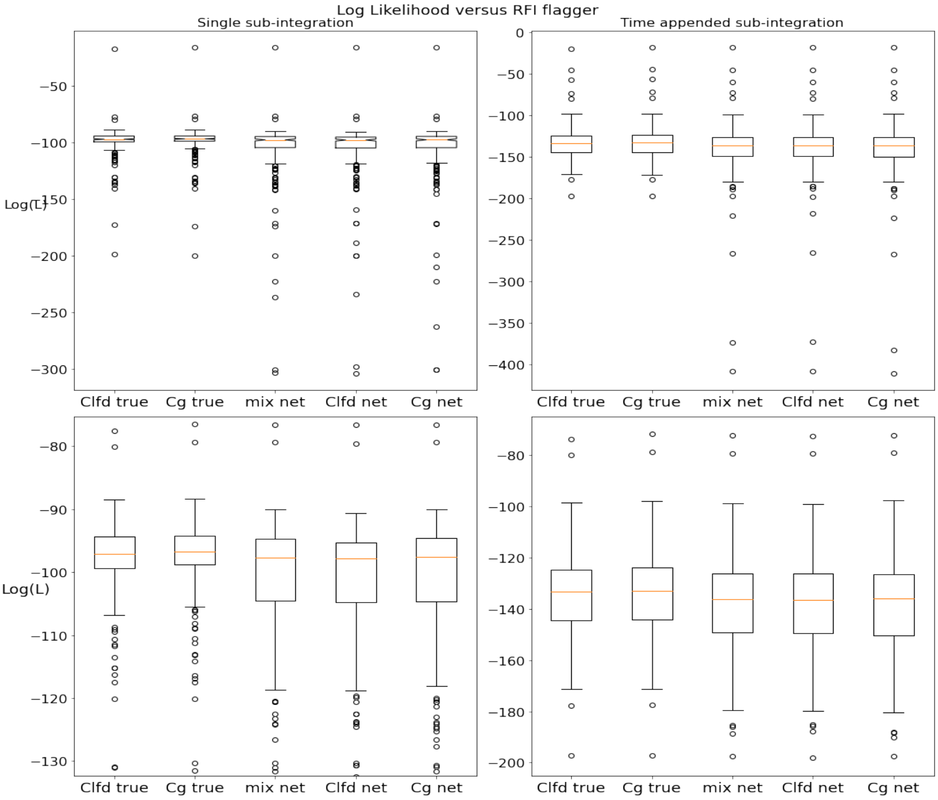

Figure 4. Box plots comparing Log (L) distributions of data with RFI channels flagged by ground truth and PSRFINET models for single sub-integration data (left column) and longer sub-integrations with data of the same observation time appended (right column). The top row shows all of the statistics while the bottom row is a zoomed in view of the box.

The order of magnitude of the log-likelihood depends on the number of kept principle components n; therefore, we do not define an absolute likelihood for a given data set. In Figure 4, we compare distributions of log(L) for RFI masked by the neural network models and the original ground truths of the hold out test. From Figure 4 we can see that the median of each distribution is within the interquartile range of every other distribution; however, compared to the ground truth, the neural network flagged data have distributions with more outliers and longer tails. The distributions with sub-integrations appended in time are more consistent with fewer outliers because computing the mean log(L) is more accurate over longer time sub-integrations.

4 Conclusions

We have trained and evaluated the performance of several neural networks capable of identifying and flagging RFI in folded pulsar data. Based on the AUROC and PR classifier metrics, Clfd is the more difficult ground truth for the network to learn given the same amount of data. Based on the GMM-based metrics displayed in Figure 4, we conclude that the networks are currently not as effective as the original ground truth. Their medians are close but they are affected by outliers. Deeper networks with different layer configurations and larger training sample sizes may improve the log(L) distributions to become as good as or even surpass the original ground truth.

For future work, we plan to experiment with training networks on an intersection (instead of a union) of the two ground truths, and on a selection of high-SNR data. It could be interesting to add observation time to the training data, as RFI is known to vary with time of day [10]. It may also be possible to apply reinforcement learning using the GMM-based metric as the loss function to be optimized with gradient descent.

5 Acknowledgements

We would like to thank the RFI 2022 Scientific Organising Committee, the European Centre for Midrange Weather Forecast (ECMWF), and the International Union of Radio Science (URSI) for organisation. Thanks to Auckland University of Technology for the funding. Thanks to Chris Rapson for the initial network design framework. A personal thanks to Lukman Hadi J. for some Linux command line operations that saved the main author from a lot of trouble.

References

[1] M. Bailes, “Gravitational-wave physics and astronomy in the 2020s and 2030s”, *Nature Reviews Physics*, vol. 3, no. 5, pp. 344–366, 2021. doi:10.1038/s42254-021-00303-8.

[2] J. W. McKee *et al.*, "Probing the local interstellar medium with scintillometry of the bright pulsar B1133+16" *The Astrophysical Journal*, vol. 927, no. 1, pp. 99, 2022, doi: 10.3847/1538-4357/ac460b.

[3] M. Bailes, "The art of precision pulsar timing", *Proceedings of the International Astronomical Union*, vol. 5, no. S261, pp. 212–217, 2009.

[4] B. T. Indermuehle *et al.*, "The ASKAP RFI environment as seen through BETA", *Radio Frequency Interference 2016*, 2016, pp. 43-48, doi: 10.1109/RFINT.2016.7833529.

[5] P. A. Fridman, and W. A. Baan, "RFI mitigation methods in radio astronomy", *Astronomy and Astrophysics*, vol. 378, pp. 327–344, 2001. doi:10.1051/0004-6361:20011166.

[6] P. E. Dewdney *et al.*, "The Square Kilometre Array", *IEEE Proceedings*, vol. 97, no. 8, pp. 1482–1496, 2009. doi:10.1109/JPROC.2009.2021005.

[7] A. R. Offringa *et al.*, "Post-correlation radio frequency interference classification methods", *Monthly Notices of the Royal Astronomical Society*, vol. 405, no. 1, pp. 155–167, June 2010.

[8] M. Bailes, "MeerTime - the MeerKAT Key Science Program on Pulsar Timing", *MeerKAT Science: On the Pathway to the SKA*, 2016.

[9] M. Bailes *et. al.*, "The MeerKAT telescope as a pulsar facility: System verification and early science results from MeerTime", *Publications of the Astronomical Society of Australia*, vol. 37, July 2020. doi:10.1017/pasa.2020.19

[10] J. Akeret *et. al.*, "Radio frequency interference mitigation using deep convolutional neural networks", *Astronomy and Computing*, vol. 18, pp. 35–39, Sept 2017. doi:10.1016/j.ascom.2017.01.002.

[11] Z. Yang *et. al.*, "Deep residual detection of radio frequency interference for FAST", *Monthly Notices of the Royal Astronomical Society*, vol. 492, no. 1, pp. 1421–1431, Feb 2020. doi:10.1093/mnras/stz3521.

[12] R. Yan, C. Dai *et. al.*, "Radio frequency interference detection based on the AC-UNet model", *Research in Astronomy and Astrophysics*, vol. 21, no 5, pp 119, 2021.

[13] A. Vafaei Sadr *et. al.*, "Deep learning improves identification of Radio Frequency Interference", *Monthly Notices of the Royal Astronomical Society*, vol. 499, no 1, bll 379–390, Oct 2020. doi:10.1093/mnras/staa2724

[14] D. Czech *et. al.*, "A CNN and LSTM-based approach to classifying transient radio frequency interference", *Astronomy and Computing*, vol. 25, bll 52–57, 2018.

[15] L. Künkel *et. al.*, "Detecting pulsars with neural networks: a proof of concept", *Monthly Notices of the Royal Astronomical Society*, vol. 506, no. 1, pp. 1111–1127, 2021. doi:10.1093/mnras/stab1628.

[16] I. Tariq *et. al.*, "Adaboost-DSNN: an adaptive boosting algorithm based on deep self normalized neural network for pulsar identification", *Monthly Notices of the Royal Astronomical Society*, vol. 511, no. 1, pp. 683–690, 2022. doi:10.1093/mnras/stac086.

[17] V. Morello *et al.*, "The High Time Resolution Universe survey - XIV. Discovery of 23 pulsars through GPU-accelerated reprocessing", *Monthly Notices of the Royal Astronomical Society*, vol. 483, no. 3, pp. 3673–3685, 2019. doi:10.1093/mnras/sty3328.

[18] Lazarus, P. *et al.*, "Prospects for high-precision pulsar timing with the new Effelsberg PSRIX backend", *Monthly Notices of the Royal Astronomical Society*, vol. 458, no. 1, pp. 868–880, 2016. doi:10.1093/mnras/stw189.

[19] C. Szegedy *et al.*, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning", *Association for the Advancement of Artificial Intelligence*, vol. 31, no. 1, Feb. 2017 .doi:10.48550/ARXIV.1602.07261

[20] J. F. Kolen, S. C. Kremer, "Gradient Flow in Recurrent Nets: The Difficulty of Learning LongTerm Dependencies", *A Field Guide to Dynamical Recurrent Networks*, IEEE, 2001, pp. 237-243. doi: 10.1109/9780470544037.ch14.

[21] K. He *et al.*, "Deep Residual Learning for Image Recognition," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[22] N. Srivastava *et al.*, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting", *Journal of Machine Learning Research*, Vol. 15, No. 1, pp. 1929–1958, Jan. 2014.

[23] D. P. Kingma, J. Ba, "Adam: A Method for Stochastic Optimization", *International Conference on Learning Representations*, Dec 2014. doi:10.48550/ARXIV.1412.6980

[24] C. Shorten, T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning", *Journal of Big Data*, vol. 6, No. 1, pp. 60, Jul 2019. doi:10.1186/s40537-019-0197-0

[25] J. Cavanaugh, A. Neath, "The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements", *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 11, No. 1, pp. e1460 , Mar 2019. doi: https://doi.org/10.1002/wics.1460