

USING DEEP LEARNING TO RECOVER PULSAR SIGNAL

A DISSERTATION SUBMITTED TO THE UNIVERSITY OF MANCHESTER
FOR THE DEGREE OF MASTER OF PHYSICS
IN THE DEPARTMENT OF PHYSICS AND ASTRONOMY

2022

Andrea Sante
(10305586)

Performed in Collaboration with *Sergio Belmonte Diaz*
under the supervision of *Prof. Rene Breton*

Department of Physics And Astronomy

Contents

Abstract

1	Introduction	1
2	Artificial Neural Networks	5
2.1	Classification Model	5
2.1.1	Convolutional Neural Networks	5
2.1.2	Model Architecture	6
2.2	Regression Model	6
2.2.1	Residual Neural Networks	6
2.2.2	Model Architecture	7
3	Training	8
3.1	Training Data Set	8
3.2	Classification Model	9
3.3	Regression Model	10
4	Test	10
4.1	Test Data Set	11
4.2	Classification Model	13
4.3	Regression Model	17
5	Conclusion	18

Abstract

Technological advances have largely increased the amount of data generated by modern astrophysical surveys. In the pulsar field, this has translated to the generation of millions of pulsar candidates per survey making the identification of pulsar signals impossible without the use of automated classification techniques. In this report, a new deep learning approach to the standard pulsar searching pipeline is proposed for the identification and characterisation of single-pulse events in radio pulsar observations. The advantage of this new approach would be to skip computationally-expensive steps such as the de-dispersion of the pulse profile. A convolutional neural network has been developed for distinguishing dispersed pulse signals from radio frequency interference, whereas a residual neural network has been created to estimate the dispersion measure value associated to each identified pulse profile. The two models have been trained and tested on simulated data. The classification model was tested to identify single pulses in RFI-excised data with a precision of 75% and a recall of 88%; the regression model could predict DM values with a difference smaller than 10 from the true for 84% of the cases.

1 Introduction

Pulsars are compact objects whose signal is characterised by a periodic sequence of peaks in intensity called pulses (Lorimer, Kramer, 2004). Physically, pulsars arise as the remnants of massive stars undergoing core-collapse supernovae explosions and are supported in a state of hydrostatic equilibrium by neutron degeneracy pressure (Pacini, 1968). Because of conservation of magnetic field flux and angular momentum, pulsars are generated as extremely fast rotating bodies (periods ranges from milliseconds to seconds) with strong magnetic fields ($10^8 - 10^{15}$ G). Although the emission mechanism is yet to be understood (Lorimer, Kramer, 2004), pulsars are thought to produce a beam of radiation from their magnetic poles.

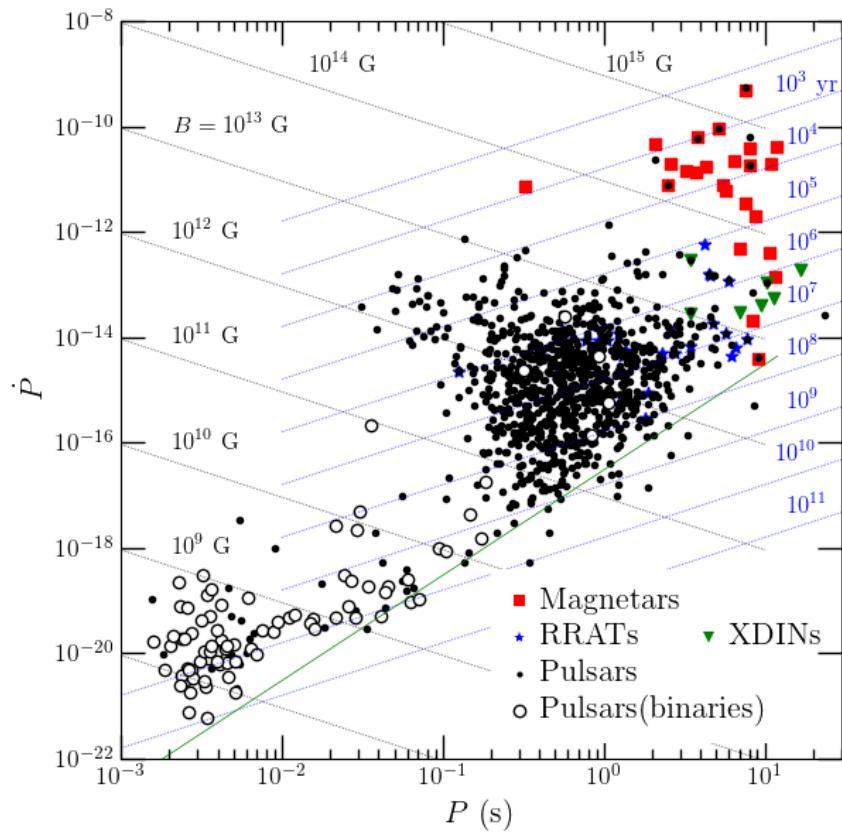


Figure 1.1: Plot of the spin-down rate (\dot{P}) and spin (P) for part of the known pulsar population. Millisecond pulsars are indicated as pulsar binaries. Figure created by Andrew W. Steiner ¹.

Based on the period of rotation (spin) and spin-down rate, the pulsar population can be broadly divided in four classes (Lyne, Graham-Smith, 2012): i) millisecond pulsars, characterised by periods shorter than 30 ms caused by the interaction with a companion star; ii) normal pulsars, which comprise the vast majority of known pulsars and have periods between 100 ms and few seconds; iii) rotating radio transients (RRATs), observed to emit isolated sporadic pulses and to have higher magnetic fields and longer periods than normal pulsars; iv) magnetars, characterised by strong magnetic field and long periods and

spin-down rates. The properties of the pulsar population are schematically summarised in the $P - \dot{P}$ diagram shown in Figure 1.1.

Like the beacon of a lighthouse, the radiation emitted by a pulsar intersects the line-of-sight of an aligned observer once per rotation. By measuring the time-of-arrival of these pulses, pulsar periods can be measured with a precision of a fraction of a microsecond (Manchester, 2017). As pulsars are extremely stable rotators, variations in the period residuals are used by astronomers to detect sensitive quantities such as the gravitational radiation emitted in pulsar binary systems and relativistic effects in strong gravitational fields (Kramer et al., 2021), or even the orbit of extra-solar planets (Wolszczan, 1994). Despite the numerous applications, there is a limited knowledge of the physics of pulsars

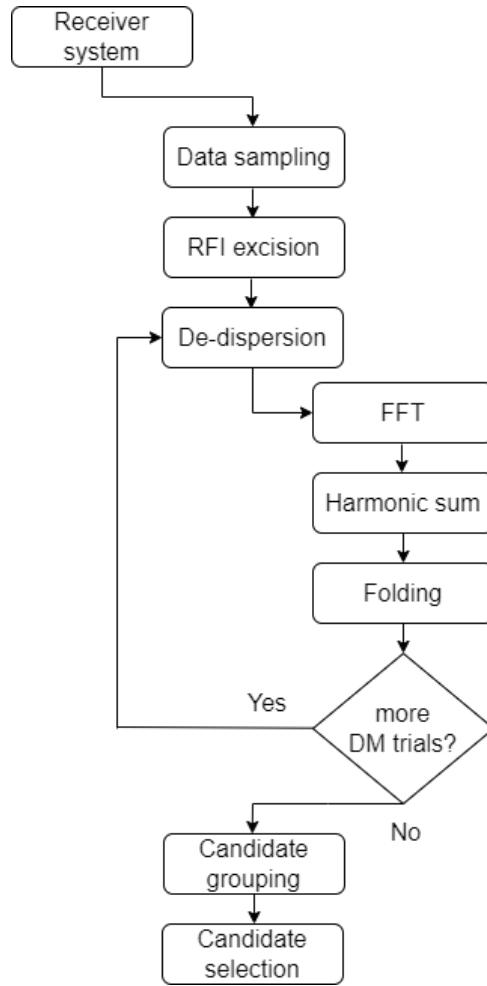


Figure 1.2: Schematic diagram showing the main steps involved in the standard pulsar searching pipeline.

which is mainly caused by the difficulties in finding these compact objects and conducting population studies. In about fifty years of Pulsar Astronomy, only about three thousand of the over a hundred thousand potentially observable pulsars have been discovered (Faucher-Giguère, Kaspi, 2006). Searching for pulsars is a cumbersome procedure as pulse signals are distorted by the interstellar medium (ISM) and can be as weak as a fraction of Jansky (Lorimer, Kramer, 2004). Figure 1.2 shows the main steps involved in the standard pulsar searching pipeline. At first, the recorded signal is sampled into s_{tot} time intervals and split into n_f frequency channels in order to create an $n_f \times s_{tot}$ matrix of data. Then,

frequency channels corresponding to known radio frequency interference (RFI) are removed. After that, the data is corrected for the frequency-dependent delay affecting the pulse radiation as it travels through the interstellar medium (ISM). This is caused by intervening ionized material which makes the ISM dispersive; hence, the pulse radiation experience a frequency-dependent group velocity which causes high frequency pulses to be observed earlier than their lower frequency counterparts. The amount of delay observed between two frequency channels (CH_1 and CH_2) of frequency f_1 and f_2 (with $f_1 < f_2$) can be quantified as

$$\Delta t \simeq 4.15 \times \text{DM} \times (f_1^2 - f_2^2) \text{ ms},$$

where the frequencies are expressed in MHz and the dispersion measure (DM) is defined as the integrated column density of free electrons along the line-of-sight to the pulsar (Lorimer, Kramer, 2004). The amount of delay observed at each frequency is corrected during the de-dispersion step. This is one of the most computationally-expensive steps of searching for pulsars as the amount of delay is proportional the dispersion measure, which cannot be known a priori. Hence, a brute-force approach is usually used and the data is corrected for several trial dispersion measure values before the right one is found. An example of a pulse signal before and after the de-dispersion step is shown in Figure 1.3a and 1.3b respectively.

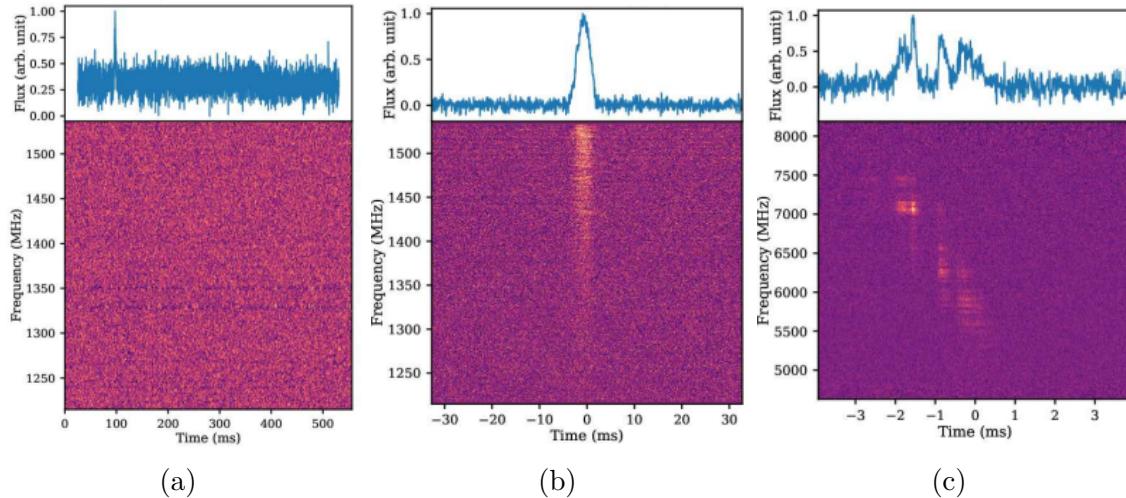


Figure 1.3: Time-frequency images of Fast Radio Burst spectra (bottom panels) with the corresponding band-integrated pulse profiles (top panels). The images represent the FRB121102 before (a) and after (b) being de-dispersed and when observed at a different time and frequency resolution (c) showing burst sub-structures. Images from Cordes, Chatterjee (2019).

After the de-dispersion step, periodicity in the data are searched in the frequency-domain by performing a Fourier analysis of the data; at the end a number of pulsar candidates are produced. For each candidate, diagnostic plots and statistics are computed to decide whether it is the signal of a pulsar or just noise. The increased sensitivity of new telescopes eased both the detection of pulsars and faint background sources (Lyon et al., 2016) leading to the production of millions of pulsar candidates for each survey. As can be seen in Table 1.1, while candidate selection could be performed manually before the early 2000s, the implementation of automated approaches is the only viable option for modern surveys.

Survey	Year	Candidates
2nd Molonglo Survey (Manchester et al., 1978)	1973	2500
Phase II survey (Stokes et al., 1986)	1983	5405
Parkes Southern Pulsar Survey (Manchester et al., 1996)	1991	40,000
Parkes Multibeam Pulsar Survey (Manchester et al., 2001)	1997	8,000,000
Arecibo P-Alfa all configurations(Cordes, et al., 2006)	2004	>5,000,000
Southern HTRU (Keith, et al., 2010)	2010	55,434,300
Northern HTRU (Barr, et al., 2013)	2010	>80,000,000

Table 1.1: Number of pulsar candidates produced by seven pulsar surveys conducted during the past fifty years.

However, machine learning techniques do not have to limit to candidate selection but have the potentiality to be extended to the whole pulsar searching pipeline. This could be achieved dividing the data collected during the observation of a pulsar in a specific frequency band into a sequence of snapshots that could be individually analysed using a deep learning pipeline. Each snapshot would represent a time-frequency image of the spectrum of a pulsar candidate that could be elaborated using convolutional neural networks, state-of-the-art mathematical models for the manipulation of multidimensional data (Lecun et al., 2015). This could be applied to identify the pulse signal directly from the raw time-frequency spectrum skipping the computationally-expensive de-dispersion.

The same technique could also be extended to the search of Fast Radio Bursts (FRBs), which are short-duration (sub-second), broad-band pulses characterised by a high dispersion measure which implies an extragalactic origin (Petroff et al., 2022). The first FRB was discovered in 2007 (Lorimer et al., 2007) searching for single-pulse events in an archival data from a pulsar survey conducted in the Magellanic Clouds. At the moment about 600 unique sources and 24 repeating sources are known (Petroff et al., 2022). The burst morphology is a powerful probe for understanding the emission and propagation of FRBs (Pleunis, Chime/Frb Collaboration, 2021); however, this is complicated by the presence of time, frequency-dependent fine structures (sub-bursts), caused by Galactic diffractive interstellar scintillation (Hessels, et. al, 2019), which make the radiation signal similar to RFI. An example of time-frequency image of an FRB containing sub-structures is shown in Figure 1.3c. An efficient deep learning pipeline would be able to distinguish FRBs from RFI despite the presence of discrete sub-structures and has the potential to be applied to the large amount of archival data collected in 50 years of pulsar surveys expanding our understanding of the FRB emission and population.

Hence, the purpose of the Master project hereby reported is to create a deep learning pipeline that, given an $n_f \times s_{tot}$ image of raw data is able to: i) directly classify the observation as containing the signal of a pulse or not; ii) estimate the dispersion measure associated to each detected pulse signal. For both tasks, specific artificial neural networks were implemented. A more detailed description of the deep learning techniques used in this analysis is provided in Section 2. The data set and procedure used for training the networks are discussed in Section 3, whereas the performance on the test data sets is reported in Section 4.

2 Artificial Neural Networks

Artificial neural networks consist in a series of parameterised functions, called neurons, arranged in an ordered sequence, a layer (Goodfellow et al., 2016). As data enters in the network, it is processed such that the value returned by each neuron is used as the input of the neurons in the subsequent layer. Before reaching the next layer, the output of each neuron is passed through an "activation" function which performs a non-linear operation. At the end of the network, the output of the final layer is compared to the expected value (called *label*) associated to the initial input through a *loss*" function which quantifies the discrepancy between the two values. The information provided by the loss function is then used to update the parameters of each neuron in the network through an optimisation technique called backpropagation with the aim to minimise the loss function. In brief, artificial neural networks can be considered as "models" attempting to represent an unknown function relating the input data and the label. This process is completely general and can be used for multiple purposes (Goodfellow et al., 2016). During the project, two different artificial neural networks (or models) were developed:

- a classification model, used for classifying the images of the radio observation as pulse signal or noise;
- a regression model, which returns an estimate of the dispersion measure value associated to a certain pulse signal.

2.1 Classification Model

The classification of images of radio observations was performed using a convolutional neural network. This is a specific type of artificial neural network which is optimal for modelling multi-dimensional data and performing classification tasks (Lecun et al., 2015).

2.1.1 Convolutional Neural Networks

Convolutional neural networks are designed to extract and detect patterns in data made of multiple arrays such as coloured images, audio spectrograms, or videos (Lecun et al., 2015). The application of convolutional neural networks in astrophysics are numerous and can range from the morphological classification of galaxies (Walmsley, et al., 2020) to the detection of gravitational waves (George, Huerta, 2018). The power of this technique lies in its mathematical structure. Each convolutional layer comprises three main stages (Goodfellow et al., 2016):

- Convolution stage, in which a convolution between the input data and a matrix of same dimension but smaller size called kernel is computed. The aim of this operation is to extract features while reducing the quantity of information that will be processed in the subsequent layers. It is common to perform the convolution over multiple *channels*, i.e. convolve the input data with more kernels in parallel. Each operation results in a single *feature map* whose dimension depends on the kernel size and the stride of the convolution, i.e. the amount of data elements by which the kernel is shifted across the input data.

- Detector stage, where the result of the convolution becomes the argument of an activation function which introduces non-linearity in the network.
- Pooling stage, where the elements of a region of the feature map are substituted to a summary statistics such as the average (AvgPooling) or maximum value (MaxPooling). The aim of this process is to further reduce the information processed in the network while providing translational invariance to the detected features.

2.1.2 Model Architecture

A schematic diagram of the artificial neural network used for classifying images of radio spectra is shown in Figure 2.1. It consists of four convolutional neural networks each with a 3×3 kernel size, a rectified linear unit (ReLU) activation function, and a maximum pooling. The convolutional layers are connected to three further layers with 4096, 1024 and 256 neurons all activated through ReLU functions. Dropout (Srivastava et al., 2014) regularization technique is adopted for each layer. Hence, neurons are randomly deactivated during training process reducing the risk of interdependent learning. The output layer comprises only two neurons, activated by a sigmoid function, which represent the probability of the image representing the signal of a pulse or noise.

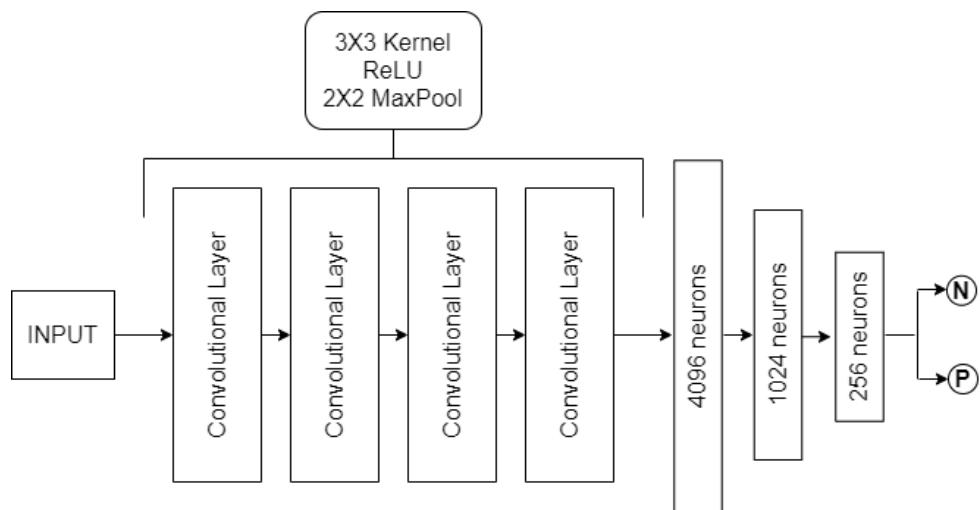


Figure 2.1: Schematic of the architecture of the classification model.

2.2 Regression Model

The dispersion measure value associated to a pulse signal was determined directly from the corresponding image of the dispersed spectrum by using a residual neural network.

2.2.1 Residual Neural Networks

Residual neural networks (ResNet) were developed by researchers at Microsoft Research in 2014 (He et al., 2015) to provide a solution to two problems affecting very deep neural networks:

- degradation, i.e. after a certain number of layers there is a saturation in the accuracy of the model which starts degrading as the number of layers increases (He, Sun, 2014);

- vanishing gradient, the information used to update the parameters of the neurons (as it consists in gradients propagating through the layers of the network by chain rule) converges to zero not allowing the performance of the network to improve (Glorot, Bengio, 2010).

Residual neural networks address these challenges by reformulating the working concept of hidden layers. Instead of approximating a model $H(x)$ mapping the input x to the output of the network, residual blocks are designed to model the *residual* mapping $F(x) := H(x) - x$, which is empirically found to be easier to optimise (He et al., 2015). The original model $H(x)$ is then retrieved by adding to the output of the residual block $F(x)$ the input x . Practically, this is done using a shortcut connection which performs an identity mapping of the input adding no further computational complexity and allowing the gradients to backpropagate directly from the deeper to the initial layers. In the original paper (He et al., 2015), five types of ResNets were defined based on the overall number of layers in the network: ResNet-18, ResNet-34, ResNet-50, ResNet-101, ResNet-152. The first layer is the same for all architectures and comprises a convolutional layer with a 7×7 kernel, stride of two and MaxPooling.

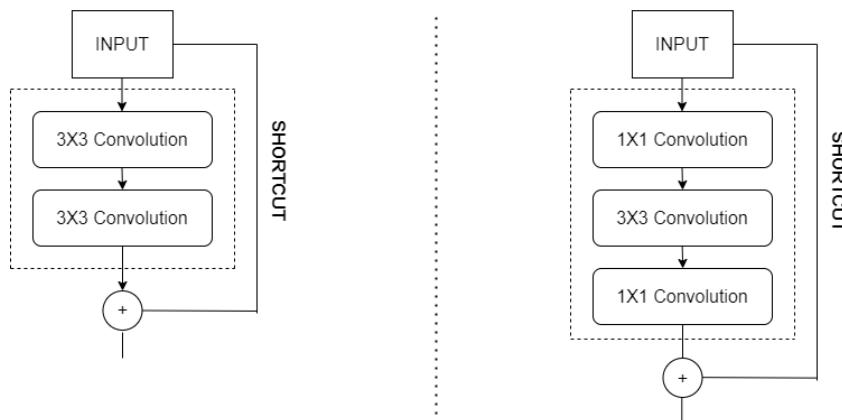


Figure 2.2: Diagram showing the operations involved in the basic (left) and bottleneck (right) residual blocks.

The difference between the two shortest models and the other is in the structure of the residual block. ResNet-18 and ResNet-34 have a residual block comprising two convolutional layers with 3×3 kernel size and MaxPooling. ResNet-50, ResNet-101 and ResNet-152 are composed by *bottlenecks* which consists of a stack of 1×1 , 3×3 , and 1×1 convolutional layers which make training over deep networks more time efficient. A comparison between the basic and bottleneck residual units is shown in Figure 2.2.

2.2.2 Model Architecture

The regression model used to estimate the dispersion measure for the images containing pulse signals was based on the architecture of the ResNet-50. This is composed of four units containing 3, 6, 4 and 3 residual blocks. When passing from one unit to the other the amount of processed data is halved and, in order to preserve the time complexity, the number of channels is doubled. Hence, each unit computes convolutions on 64, 128, 256 and 512 channels respectively. Contrary to the original architecture (He et al., 2015), the final unit is connected to four layers containing 128, 64 and 32 neurons activated using ReLU

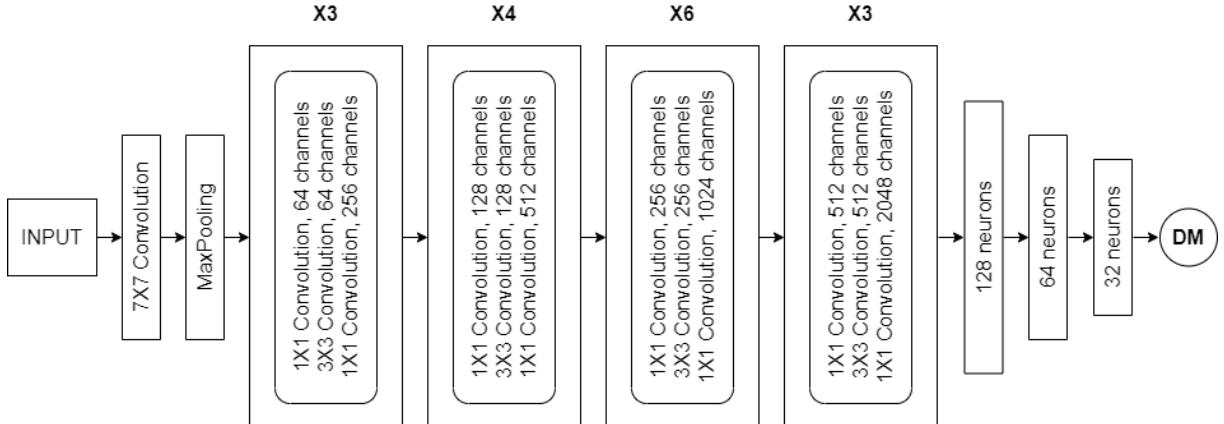


Figure 2.3: Diagram of the architecture of the artificial neural network used for determining the dispersion measure (DM) values associated to images of pulse spectra.

functions. The output layer consists of one neuron which returns the value of the dispersion measure predicted by the network. This addition to the ResNet-50 architecture was done to optimise the performance of the network. Even though convolutional neural networks are extremely successful at detecting the presence of features in multidimensional data, the quantity and length of dense layers are more decisive in improving the performance of the network for regression tasks (Florea, Andonie, 2020). In this case, the number of neurons and layers was determined using Ray Tune (Liaw et al., 2018), a hyperparameter tuning pipeline which, given a certain ensemble of models, determines the most efficient configuration. A schematic of the architecture of the regression model used in this analysis is shown in Figure 2.3.

3 Training

The training of the classification and regression models was performed using a data set of simulated pulse profiles and receiver noise. The hyperparameters used in the training procedure were different between the two models because of the different level of complexity between the two tasks.

3.1 Training Data Set

The training examples consisted of a 128x128 time-frequency images of radio observations generated simulating the statistical properties of radio receiver noise and the main effects affecting gaussian-shaped pulse profiles travelling through the ISM: pulse dispersion and broadening. The former is the frequency-dependent delay caused by ionized material in the line-of-sight to the pulsar, while the latter is due to scattering from inhomogeneities in the ISM causing radiation at different frequencies to travel through different path lengths. The effect of scattering on the pulse profile can be mathematically described by a convolution with a time-dependent, one-sided exponential with parameter τ , the *scattering time*. The effect of scattering is stronger at low frequencies as $\tau \propto f^{-4}$ (Lorimer, Kramer, 2004).

All the images used for the training of the two models were generated simulating radio observations in the 1200-1600 MHz frequency band and an exposure time of 1s. A sample rate of $50 \mu\text{s}$ and 512 frequency channels were assumed to create 512X20000 matrices of data. Before entering the network, each image was downsampled to a 128X128 shape. Pulse signals were simulated considering a gaussian-shaped pulse profile with a signal-to-noise ratio (S/N) between 5 and 100 and duration in the range 1-20ms; the DM value was chosen from the $[1;793] \text{ cm}^3 \text{ pc}^{-1}$ interval and the first time-of-arrival was calculated to ensure that the pulse profile fitted entirely in the observation window. The mean flux density of pulses was made to follow a frequency-dependent power-law with spectral index in the range $-4 < \beta < 0$ and mean estimated value of -1.8 as empirically studied by Maron et al. (2000). Pulse broadening was implemented convolving the gaussian pulse profile with a one-sided exponential with scattering time calculated from the DM value using the empirical relation found by Bhat et al.(2004)

$$\log(\tau) = -6.46 + 0.154 \log(\text{DM}) + 1.07 \log^2(\text{DM}) - 3.86 \log(f),$$

where τ is expressed in ms, DM in $\text{cm}^{-3} \text{ pc}$ and the frequency of the radiation f in GHz.

3.2 Classification Model

The training of the classification model was performed considering a data set of 120,000 images, half of which consisting of examples of pulses. These were divided equally in four S/N ranges: 5-15, 15-30, 30-50, 50-100. Each example was associated to a label consisting of 0 in case of a pulse and 1 for a noise image. A cross-entropy loss function was used to compare the predictions of the network to the labels. A validation data set comprising 10% of the training examples was used to control the quality of the training. Figure 3.1 shows an example of an image used for the training of the classification model.

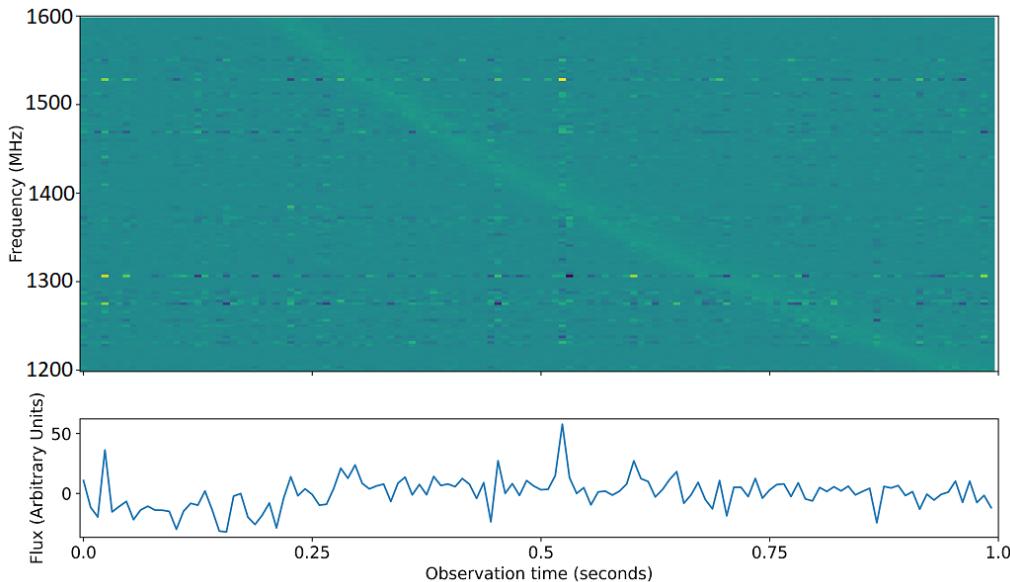


Figure 3.1: Example of the time-frequency images (top panel) used for the training process of the classification and regression models. The band-integrated pulse profile is shown in the bottom panel.

The network was trained for 20 epochs with mini-batches of 10 images. This means that the parameters of the model were update once every 10 examples and that the whole training data set was processed by the network 50 times. The learning rate, i.e. the fractional amount by which the parameters were updating, was set to 0.001 and Adam was chosen as the optimisation algorithm. The loss function chosen to compare the prediction Figure 3.2a shows a plot of the mean loss function per epoch for the training and validation data sets. No overfitting or underfitting was observed during the training.

3.3 Regression Model

The regression model was trained with the same training and validation data sets of the classification model. A mean-square-error function was used to compare the predictions of the model with the expected DM values. Because of the complexity of identifying a pulse feature and associating it to a specific numerical value, the regression model was trained for 50 epochs. The size of the mini-batches and the optimisation algorithm were the same used for the classification model. Considering the duration of the training, an algorithm was implemented to update the learning rate every time the mean loss function per epoch was observed to plateau. This was done to ensure the parameters of the model could be tuned by the amount needed for minimising the loss function. The learning rate was initially set to 3×10^{-5} (value indicated to be optimal by Ray Tune) and then made to decrease by 25% every time the loss function was increasing more or decreasing less than the standard deviation calculated from the previous five mean loss functions. Figure 3.2b shows a plot of the mean training and validation loss functions per epoch. Again, the training was not affected by overfitting or underfitting.

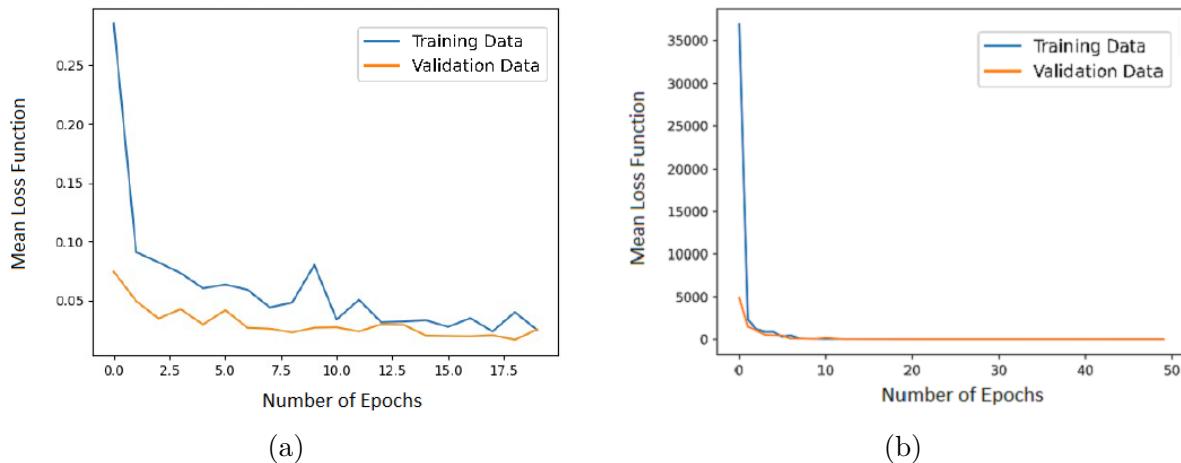


Figure 3.2: Plots of the mean loss function per epoch calculated for the training and validation data sets of the classification (3.2a) and regression (3.2b) model.

4 Test

In order to gain a more realistic perspective of the possible applications of the classification and regression models described in Section 3.2 and 3.3, an external data set was used for testing. Details of the adopted test data are provided in Section 4.1. The performance

of the classification model was quantified using two metrics: i) confusion matrices, which contain information on the number of images correctly classified as pulses (true positives, TP) and noise (true negatives, TN), as well as the number of images misclassified as pulses (false positives, FP) and RFI (false negatives, FN); ii) F1 score, which measures the accuracy of the model by considering both the precision (the fraction of actual pulses over the total images classified as pulses) and the recall (the fraction of correctly identified pulses over the total number of available pulse images) of the test

$$F_1 = 2 \frac{precision \times recall}{precision + recall} = \frac{TP}{TP + 0.5(FP + FN)}.$$

4.1 Test Data Set

The performance of the classification and regression models was assessed using time-frequency images of pulses and RFI spectra sourced from Agarwal et al. (2020). The data set consists of observations conducted using the Green Bank Telescope (GBT) and the 20m telescope of the Green Bank Observatory. The data from the GBT receiver was recorded using GREENBURST (Surnis et al., 2019) and FLAG (Rajwade et al., 2019) instruments, whereas the data from the 20m telescope was observed using the Skynet (Smith et al., 2016) and GBTrans (Golpayegani, et al., 2019) backends. The RFI examples comprise 2170, 2790 and 2359 observations from GREENBURST, FLAG, and Skynet respectively. The pulse images comprehends: 357 and 1931 single pulses observed through FLAG from PSR B2011+38 and PSR 1933+16; 1376 single pulses from PSR B0740-28 obtained with the GREENBURST backend; and 3000 Crab giant pulses recorded with GBTrans. Table 4.1 provides a summary of the composition of the test data set.

Instrument	RFI	Single Pulse
GREENBURST	2170	1376 (PSR B0740-28)
FLAG	2790	357 (PSR B2011+38) 1931 (PSR 1933+16)
Skynet	2359	-
GBTrans	-	3000 (Crab pulsar)

Table 4.1: Number of RFI and single pulse examples in the test data set observed with GREENBURST, FLAG, Skynet, and GBTrans backends. Inside the round-brackets there is the name of the observed pulsar.

The single pulse data sets contained images of dedispersed pulses centred in the observation window and with a S/N in the range [2; 50]. In order to be able to test the performance of the two models, the images were dispersed considering random DM values in the range [1, 793]. The pulse signal was also shifted by an amount selected such that each pulse would lie completely in the image even after the dispersion. The DM values used to disperse the pulse signals were collected and used as labels when testing the regression model. Figure 4.1 shows images of pulses from GREENBURST, FLAG, and GBTrans before and after the dispersion.

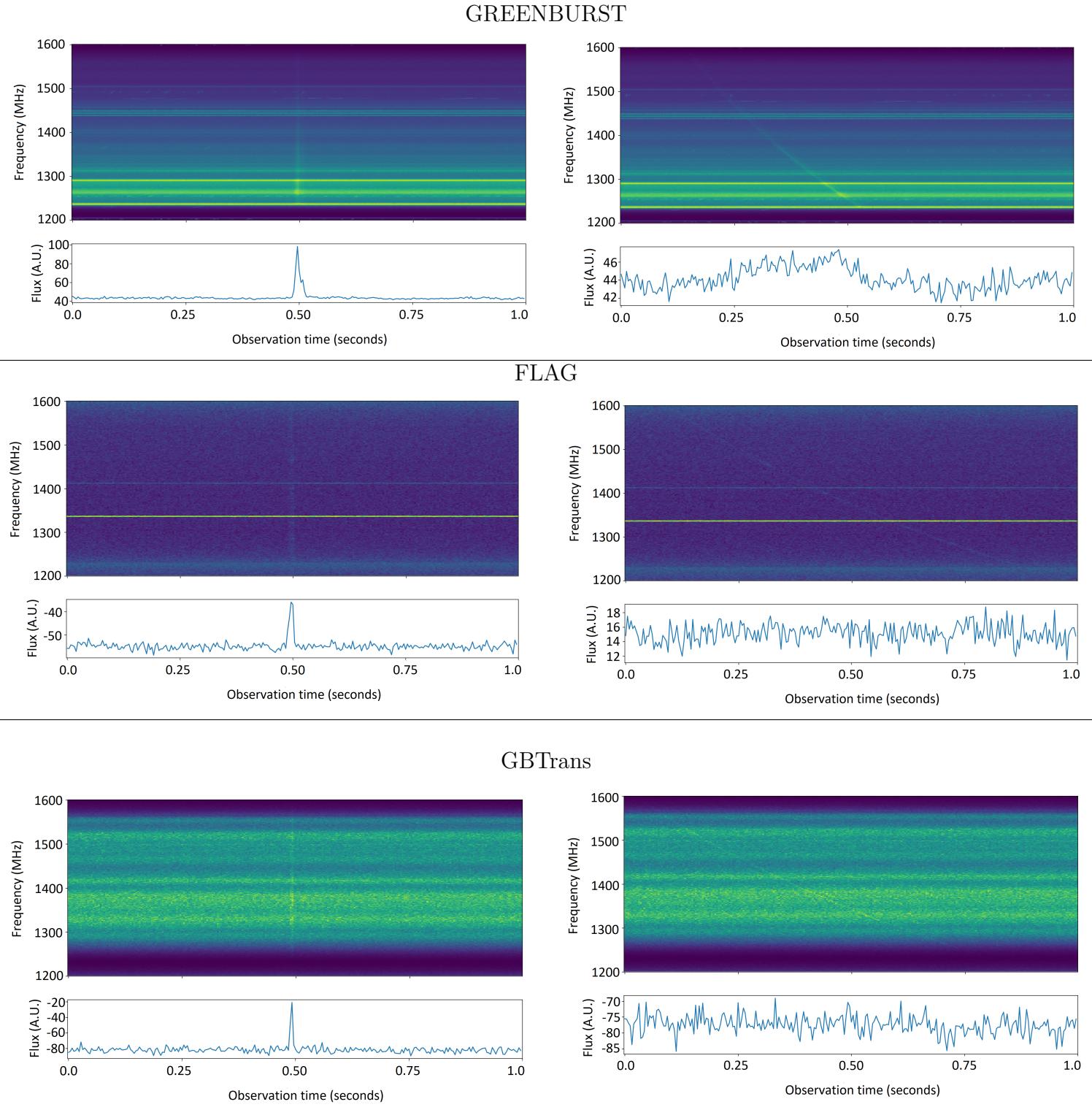


Figure 4.1: Time-frequency images (top panels) and corresponding band-integrated pulse profiles (bottom panels) of single pulses used for testing the classification model. The name of the instrument used to record the data is written at the top of each column. Images on the left column are the original examples as provided by Agarwal et al. (2020); the same images after being dispersed are shown on the right column.

4.2 Classification Model

The performance of the classification model tested on the data set generated by Agarwal et al. (2020) is shown in Table 4.2, which contains confusion matrices calculated for assessing the classification accuracy on the images recorded from GREENBURST, FLAG, and Skynet-GBTrans. The precision in the classification appears to depend strongly on the characteristics of the test images. The model fails completely in classifying images from GREENBURST as it shows a precision and recall of 0%, whereas it shows a strong bias towards the pulse category for both FLAG and Skynet-GBTrans, which show a precision of 45% and 60%, respectively, and recall of 100%. The overall performance of the classification model is described by a F1 score of 58%.

The most plausible cause of the poor performance of the classification model is the evident difference between the training (Figure 3.1) and test (Figure 4.1) examples. The model was trained to recognise the pulse signal in data where containing only that continuous feature; thus, it is not surprising the drastic drop of the performance in classifying images containing strong and lasting RFIs. The different behaviour showed in the classification of the images of GREENBURST and the ones of FLAG and Skynet-GBTrans could be caused by the intense RFIs present in multiple channels for the first instrument. The signal of the pulse, which appears as the only continuous feature in the training examples, is obscured by the RFIs and the classifier is not able to detect any strong single feature: hence, all the data recorded with GREENBURST is classified as noise. In the case of FLAG, a single frequency channel is strongly affected by RFI. The intensity of the RFI dominates the spectrum and causes the model to confuse it for a pulse signal. Because the RFI is related to the operation of the instrument, all FLAG images were classified as containing pulses. In order to verify whether the presence of RFI was the actual cause of the poor performance of the classification model, an excision procedure was conducted for all the test images. The adopted RFI excision consisted in subtracting the mean intensity value in every frequency channel to all the data observed at that frequency. Figure 4.2 shows some of the images of both pulses and RFI from GREENBURST, FLAG and Skynet after the RFI excision.

The classification model was tested on the RFI-excised test data set. The results are shown in Table 4.3. In contrast to what observed with the raw data, the model appears to have a strong bias towards the pulse category for the images from GREENBURST and towards the RFI category for the images from FLAG. The precision and recall for the GREENBURST data set increased consistently to 39% and 100% as the model is able to correctly classify all the pulse images; however, a consistent misclassification of

Instrument		Predicted Pulse	Predicted Noise
GREENBURST	Real Pulse	0	1376
	Real RFI	1425	745
FLAG	Real Pulse	2288	0
	Real RFI	2790	0
Skynet-GBTrans	Real Pulse	3000	0
	Real RFI	1978	381

Table 4.2: Confusion matrix calculated testing the classification model described in Section 3.2 on a data set generated by Agarwal et al. (2020) (examples shown in Figure 4.1).

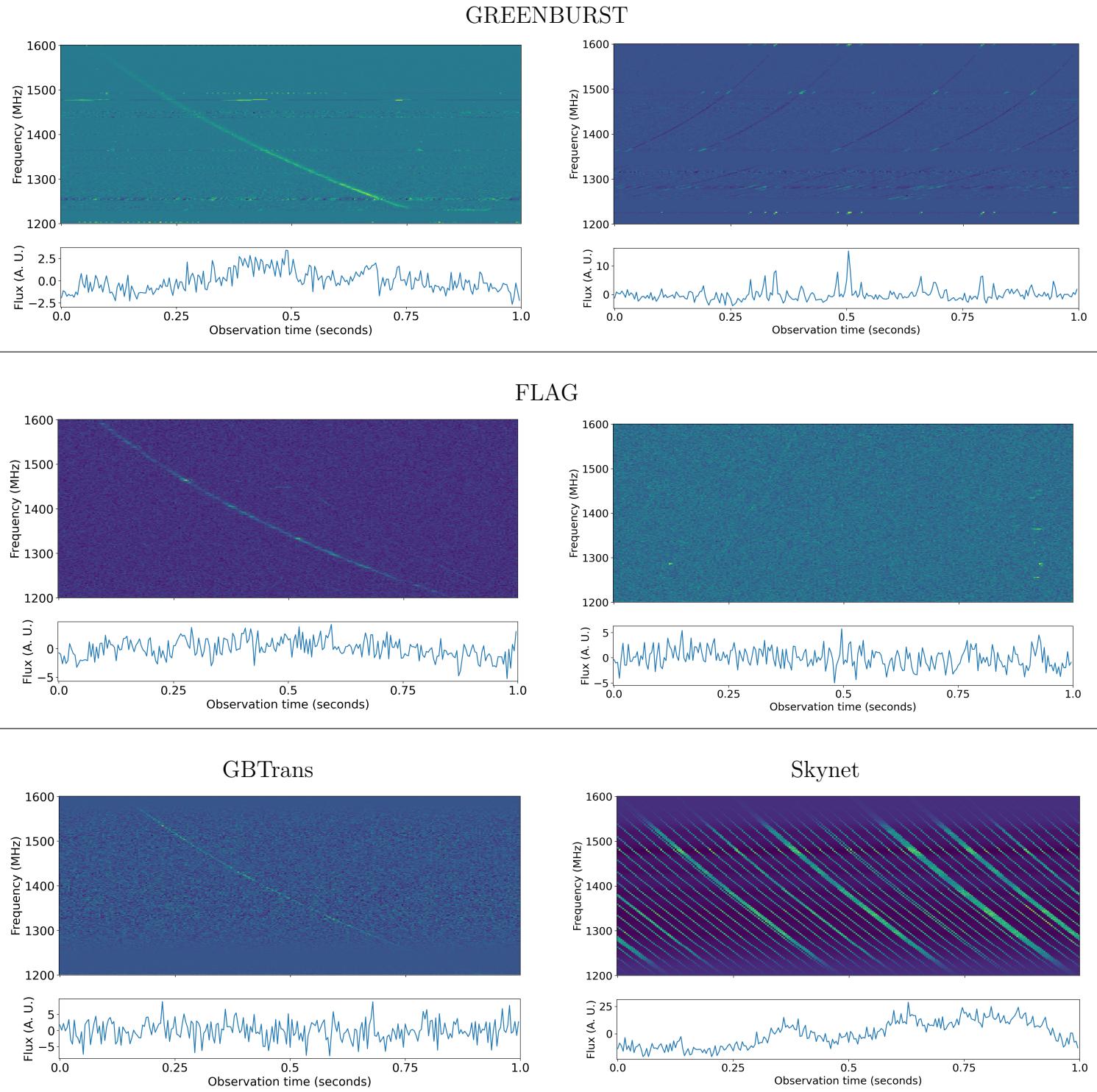


Figure 4.2: RFI-excised time-frequency images (top panels) and corresponding band-integrated pulse profiles (bottom panels) of single pulses and images of RFI spectra used for testing the classification model. The name of the instrument used to record the data is written at the top of each column. Pulse images are shown on the left column, whereas the RFI images are displayed shown on the right column.

RFI examples is observed. This is probably a consequence the pulse-looking features which appeared in the RFI examples after the RFI excision procedure. The precision for the FLAG data set improved to 96% while the recall worsen considerably to 10%. The GBTrans data set maintained the same level of precision while worsening its recall to 85%. For both instruments, the difficulty in identifying the pulse-feature in the pulse images is probably due to the strong suppression of the pulse signal in the frequency channels with a high mean intensity value. As for GREENBURST, the RFI excision caused pulse-like feature to arise in the RFI images of the Skynet data set resulting in the model failing to correctly classifying the examples.

Instrument		Predicted Pulse	Predicted Noise
GREENBURST	Real Pulse	1376	0
	Real RFI	2170	0
FLAG	Real Pulse	231	1967
	Real RFI	9	2781
Skynet-GBTrans	Real Pulse	2541	459
	Real RFI	1632	717

Table 4.3: Confusion matrix calculated testing the classification model described in Section 3.2 on the RFI-excised data set from Agarwal et al. (2020) (examples shown in Figure 4.2).

In order to understand in more details the reasons behind the poor performances of the classification model, variations in the model architecture were considered. In this analysis only the RFI-excised images from FLAG were considered as the RFI examples do not contain pulse-like features. The main problem observed in this data set consisted in the misclassification of pulse images. By considering the feature maps generated by the model at the end of each convolutional layer for the misclassified pulse examples, it can be noticed that the pulse-feature becomes more indistinguishable from the background at the end of each layer. This is probably caused by an exaggerated use of MaxPooling and by a poor choice of the parameters of the convolutional layers. To verify this hypothesis, three simpler architectures were considered:

- a one-layer model, containing only one convolutional layer with a 4×4 kernel, 16 channels, stride of 2 and a 2×2 MaxPooling;
- a one-layer model, containing only one convolutional layer with a 4×4 kernel, 32 channels, stride of 2 and a 2×2 MaxPooling;
- a two-layer model, consisting of two consecutive layers performing convolutions with a 4×4 kernel, 16 channels and a stride of 2, and followed by a 2×2 MaxPooling.

In all cases, the convolutional layers were connected to the same fully-connected layers as for the model in Section 3.2. Dropout was used for regularization. The three alternative models were trained using the same examples and hyperparameters as for the original classification model. The results obtained after testing the models on the RFI-excised FLAG data set are shown in Table 4.4.

It is clear that simplifying the architecture of the original classification model led to a considerable improvement in the classification performance, especially with regards to

Classification Model		Predicted Pulse	Predicted Noise
one conv. layer, 16 channels	Real Pulse	2031	257
	Real RFI	678	2112
one conv. layer, 32 channels	Real Pulse	2006	282
	Real RFI	602	2188
two conv. layers, 16 channels	Real Pulse	1729	559
	Real RFI	323	2467
four conv. layers	Real Pulse	231	1967
	Real RFI	9	2781

Table 4.4: Confusion matrix calculated testing three variations of the classification model described in Section 3.2 on RFI-excised images from FLAG. The confusion matrix from the original model is also reported as reference.

the identification of pulse signals. The F1 scores are 81%, 82% and 80% for the one-layer, 16-channels, one-layer, 32-channels, and two-layers model respectively. The recall increased steeply to 88% for the two one-layer models and to 76% for the two-layer one. The precision appears to be worse for the two one-layer models (75% for the 16-channels model and 77% for the 32-channels one) than the two-layer model (84%) which in turns, it is worse than the original model (96%). These results seem to suggest that deeper models struggle more to identify pulse signals than their shallower counterparts but are less likely to misclassify RFI as pulses. This could be explained because at the end of each convolution some information of the pulse signals are filtered-out as can be seen in Figure 4.3 which shows the output of the convolutional layers for the one-layer, two-layer, and four-layer classification models. The number of channels in the convolution seems to affect positively the precision of the classification model; however, the improvement does not match the increase in the required computational power as doubling the number of kernels led to a mere 2% increment in precision.

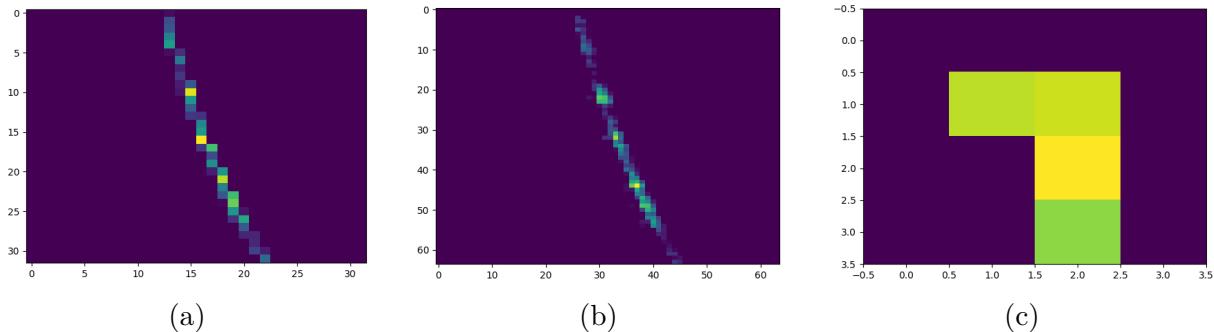


Figure 4.3: Feature maps obtained for a pulse image at the end of the last convolutional layer for the one-layer model (a), two-layer model (b), and the model described in Section 3.2 (c).

The one-layer, 16-channels model identified correctly the largest number of pulse images. Only 257 of pulses were identified as RFI. All the misclassified images contained pulses with an S/N smaller than 8; of these, 240 have an S/N smaller than 5 for which the classification model was not even trained. The performance could be improved by providing more examples of pulse with low S/N but the detection of such FRBs or pulses is not realistic.

4.3 Regression Model

The regression model was conceived to be applied after the identification of the images containing pulse signals; thus, the model was tested considering the images predicted to contain a pulse signal by the most precise and accurate classification model. As the model described in Section 3.2 failed to distinguish between pulse and RFI examples in both GREENBURST and Skynet-GBTrans data sets, the regression model was tested on the RFI-excised FLAG images predicted to contain a pulse signal by the one-layer, 16-channels model introduced at the end of Section 4.2. The model was able to retrieve the exact DM value for 274 of the 2709 available images, whereas for 1448 of them the DM was predicted with a difference from the actual value smaller than 10. If only the 2031 images actually containing pulse signals are considered, 84% of the DM values are estimated exactly or within a range of 10 from the true value. For all the considered examples, the fractional difference between the actual and predicted DM values were calculated and compared to the respective S/N . Results are shown in Figure 4.4.

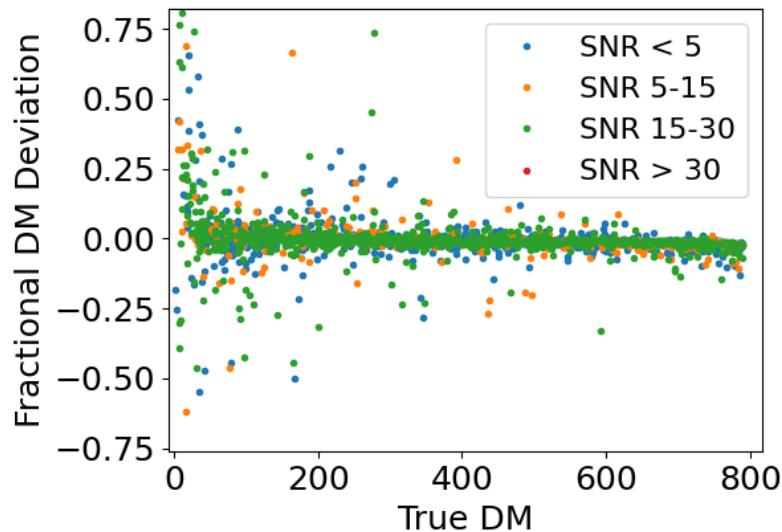


Figure 4.4: Fractional difference between predicted and actual DM values calculated for the RFI-excised FLAG images classified as pulses. The images were divided in four ranges based on the S/N of the pulse profile. The fractional DM deviation can reach -2 and +12 for true DM values smaller than 50.

In general, the regression model appears to be less effective at retrieving the DM value of low dispersed pulses ($DM < 100$); however, this trend seems to be more accentuated for pulse images with an S/N smaller than 15. Although the predictions are not exact, the regression model could be still be applied in the pulsar search pipeline as a preliminary step of the de-dispersion process. The estimate calculated by the model could be used to reduce the range of trial DM values associated to a pulse profile.

5 Conclusion

In this Master thesis project a new deep learning approach has been developed to improve the automation of signal identification of pulsars from the enormous amount of data recovered with modern astrophysical surveys. The developed classification and regression model can be applied to the analysis of single-pulse observational data as an alternative to the standard pulsar searching pipeline. An efficient RFI-excision procedure is required prior to the application of the classification model in order to obtain a precision and recall higher than 50%. The best model considered in this analysis consists of one convolutional layer with a 4×4 kernel and 16 channels, followed by three fully-connected layers of 4096, 1024 and 256 neurons. The model is able to categorise correctly RFI-excised time-frequency images of pulses and RFI with a precision of 75%, recall of 88% and F1 score of 81%. The model has difficulties identifying pulses with S/N smaller than 5 for which it was not trained. The regression model is on the 80% of the time able to provide an estimate of the DM associate to a pulse profile within a difference smaller than 10 from the true value. Predictions for pulses with a S/N smaller than 15 and a true DM smaller than 100 tend to be highly inaccurate. Nevertheless, the regression model could be potentially be used as a preliminary step in de-dispersion procedures, where it would reduce the range of trial DM values, and the identification of FRBs, for which the associated DM value is so large that the inaccuracy of the model predictions does not represent a limitation.

Bibliography

- Agarwal D., Aggarwal K., Burke-Spolaor S., Lorimer D. R., Garver-Daniels N.* FETCH: A deep-learning based classifier for fast transient classification // Monthly Notices of the RAS. IX 2020. 497, 2. 1661–1674.
- Barr E. D., et al.* The Northern High Time Resolution Universe pulsar survey - I. Setup and initial discoveries // Monthly Notices of the RAS. XI 2013. 435, 3. 2234–2245.
- Bhat N. D. R., Cordes J. M., Camilo F., Nice D. J., Lorimer D. R.* Multifrequency Observations of Radio Pulse Broadening and Constraints on Interstellar Electron Density Microstructure // Astrophysical Journal. IV 2004. 605, 2. 759–783.
- Cordes J. M., et al.* Arecibo Pulsar Survey Using ALFA. I. Survey Strategy and First Discoveries // Astrophysical Journal. I 2006. 637, 1. 446–455.
- Cordes James M., Chatterjee Shami.* Fast Radio Bursts: An Extragalactic Enigma // Annual Review of Astron and Astrophys. VIII 2019. 57. 417–465.
- Faucher-Giguère C., Kaspi V. M.* Birth and Evolution of Isolated Radio Pulsars // Astrophysical Journal. V 2006. 643, 1. 332–355.
- Florea A., Andonie R.* Weighted Random Search for Hyperparameter Optimization // arXiv e-prints. IV 2020. arXiv:2004.01628.
- George D., Huerta E. A.* Deep Learning for real-time gravitational wave detection and parameter estimation: Results with Advanced LIGO data // Physics Letters B. III 2018. 778. 64–70.
- Glorot X., Bengio Y.* Understanding the difficulty of training deep feedforward neural networks // Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. 9. Chia Laguna Resort, Sardinia, Italy: PMLR, 13–15 May 2010. 249–256. (Proceedings of Machine Learning Research).
- Golpayegani G., et al.* GBTrans: a commensal search for radio pulses with the Green Bank 20-m telescope // Monthly Notices of the Royal Astronomical Society. 09 2019. 489, 3. 4001–4006.
- Goodfellow I., Bengio Y., Courville A.* Deep Learning. 2016. <http://www.deeplearningbook.org>.
- He K., Sun J.* Convolutional Neural Networks at Constrained Time Cost // arXiv e-prints. XII 2014. arXiv:1412.1710.
- He K., Zhang X., Ren S., Sun J.* Deep Residual Learning for Image Recognition // arXiv e-prints. XII 2015. arXiv:1512.03385.
- Hessels J. W. T., et al.* FRB 121102 Bursts Show Complex Time-Frequency Structure // Astrophysical Journal, Letters. V 2019. 876, 2. L23.
- Keith M. J., et al.* The High Time Resolution Universe Pulsar Survey - I. System configuration and initial discoveries // Monthly Notices of the RAS. XII 2010. 409, 2. 619–627.
- Kramer M., Stairs I. H., Manchester R. N. et al.* Strong-Field Gravity Tests with the Double Pulsar // Physical Review X. X 2021. 11, 4. 041050.
- Lecun Y., Bengio Y., Hinton G.* Deep learning // Nature. V 2015. 521, 7553. 436–444.
- Liau R., Liang E., Nishihara R., Moritz P., Gonzalez J. E., Stoica I.* Tune: A Research Platform for Distributed Model Selection and Training // arXiv preprint arXiv:1807.05118. 2018.

- Lorimer D. R., Bailes M., McLaughlin M. A., Narkevic D. J., Crawford F.* A Bright Millisecond Radio Burst of Extragalactic Origin // *Science*. XI 2007. 318, 5851. 777.
- Lorimer D. R., Kramer M.* Handbook of Pulsar Astronomy. 4. 2004.
- Lyne Andrew, Graham-Smith Francis.* Pulsar Astronomy. 2012. 4. (Cambridge Astrophysics).
- Lyon R. J., Stappers B. W., Cooper S., Brooke J. M., Knowles J. D.* Fifty Years of Pulsar Candidate Selection: From simple filters to a new principled real-time classification approach // *Monthly Notices of the RAS*. VI 2016. 459. 1104.
- Manchester R. N.* Pulsar timing and its applications // *Journal of Physics Conference Series*. 932. XII 2017. 012002. (Journal of Physics Conference Series).
- Manchester R. N., Lyne A. G., Camilo F., Bell J. F., Kaspi V. M., D'Amico N., McKay N. P. F., Crawford F., Stairs I. H., Possenti A., Kramer M., Sheppard D. C.* The Parkes multi-beam pulsar survey - I. Observing and data analysis systems, discovery and timing of 100 pulsars // *Monthly Notices of the RAS*. XI 2001. 328, 1. 17–35.
- Manchester R. N., Lyne A. G., D'Amico N., Bailes M., Johnston S., Lorimer D. R., Harrison P. A., Nicastro L., Bell J. F.* The parkes Southern pulsar Survey — I. Observing and data analysis systems and initial results // *Monthly Notices of the Royal Astronomical Society*. 04 1996. 279, 4. 1235–1250.
- Manchester R. N., Lyne A. G., Taylor J. H., Durdin J. M., Large M. I., Little A. G.* The second Molonglo pulsar survey - discovery of 155 pulsars. // *Monthly Notices of the RAS*. XI 1978. 185. 409–421.
- Maron O., Kijak J., Kramer M., Wielebinski R.* Pulsar spectra of radio emission // *Astronomy and Astrophysics, Supplement*. XII 2000. 147. 195–203.
- Pacini F.* Rotating Neutron Stars, Pulsars and Supernova Remnants // *Nature*. 1968. 219. 145–146.
- Petroff E., Hessels J. W. T., Lorimer D. R.* Fast radio bursts at the dawn of the 2020s // *Astronomy and Astrophysics Reviews*. XII 2022. 30, 1. 2.
- Pleunis Z., Chime/Frb Collaboration .* Fast radio burst morphology in the first CHIME/FRB catalog // *American Astronomical Society Meeting Abstracts*. 53. I 2021. 236.03. (American Astronomical Society Meeting Abstracts).
- Rajwade K M, Agarwal D, Lorimer D R, Pingel N M, Pisano D J, Ruzindana M, Jeffs B, Warnick K F, Roshi D A, McLaughlin M A.* A 21 cm pilot survey for pulsars and transients using the Focal L-Band Array for the Green Bank Telescope // *Monthly Notices of the Royal Astronomical Society*. 08 2019. 489, 2. 1709–1718.
- Smith A. B., Caton D.l B., Hawkins R. L.* Implementation and Operation of a Robotic Telescope on Skynet // *Publications of the ASP*. V 2016. 128, 963. 055002.
- Srivastava N., Hinton G. E., Krizhevsky A., Sutskever I., Salakhutdinov R.* Dropout: a simple way to prevent neural networks from overfitting. // *Journal of Machine Learning Research*. 2014. 15, 1. 1929–1958.
- Stokes G. H., Segelstein D. J., Taylor J. H., Dewey R. J.* Results of Two Surveys for Fast Pulsars // *Astrophysical Journal*. XII 1986. 311. 694.
- Surnis Mayuresh P., Agarwal D., Lorimer D. R., Pei X., Foster G., Karastergiou A., Golpayegani G., Maddalena R. J., White S., Armour W., al. et.* GREENBURST: A commensal Fast Radio Burst search back-end for the Green Bank Telescope // *Publications of the Astronomical Society of Australia*. 2019. 36. e032.
- Walmsley M., et al.* Galaxy Zoo: probabilistic morphology through Bayesian CNNs and active learning // *Monthly Notices of the RAS*. I 2020. 491, 2. 1554–1574.
- Wolszczan A.* Confirmation of Earth-Mass Planets Orbiting the Millisecond Pulsar PSR B1257+12 // *Science*. IV 1994. 264, 5158. 538–542.