PUNCH 4NFDI

Deep Learning for real-time classification of astronomical radio signals

Andrei Kazantsev, Ramesh Karuppusamy, Yunpeng Men, Michael Kramer
Max Planck Institute for Radio Astronomy
Bonn, Germany

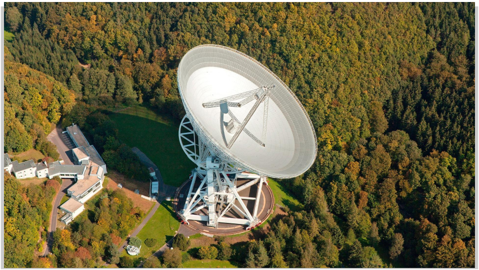
Max-Planck-Institut für Radioastronomie

Hamburg, Germany, June 15 – 16, 2023

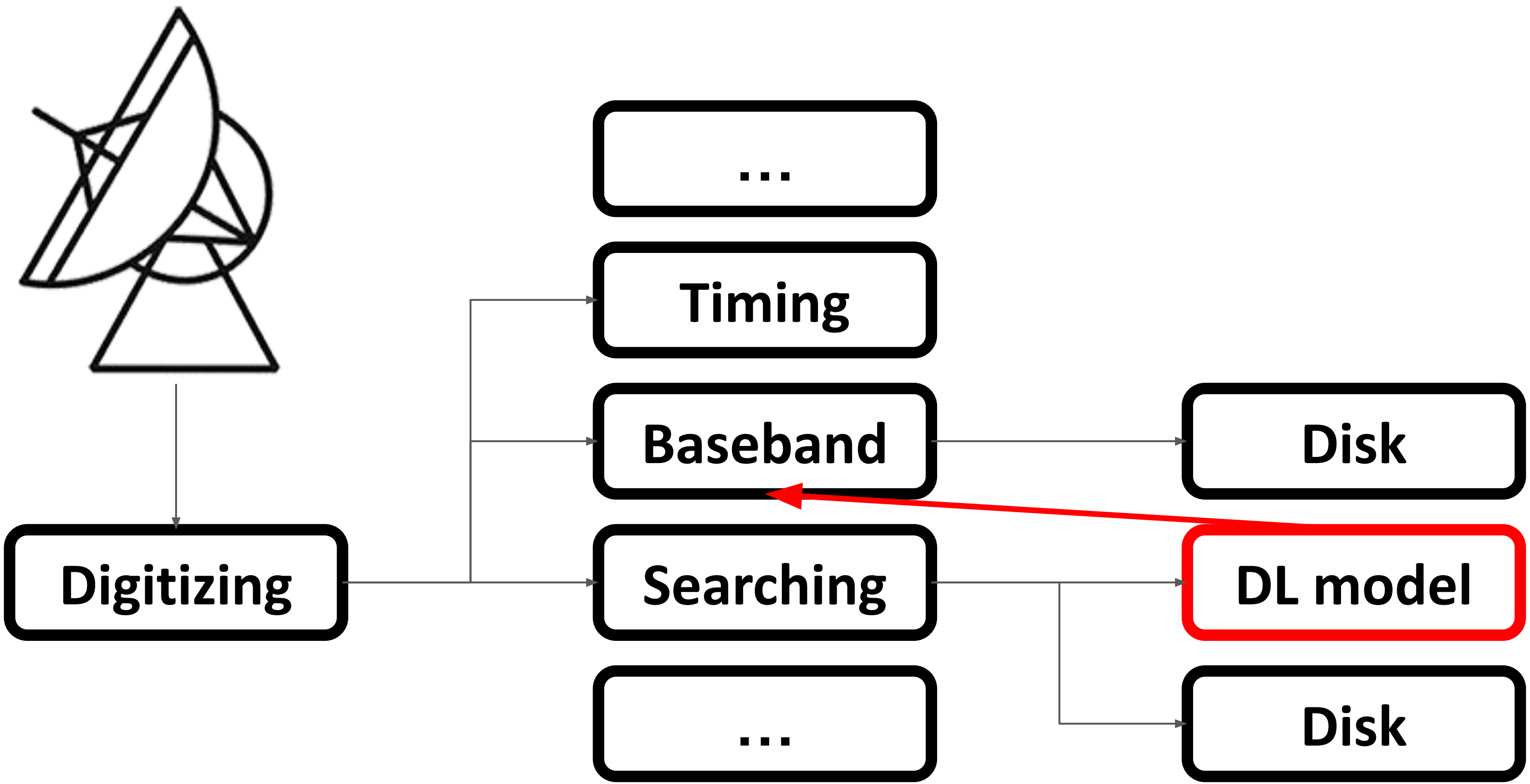
Outline of the talk

- Project recall
- Preliminary training
- Synthetic dataset for training and results of the training
- Sensitivity of the model in SNR and DM ranges.

Motivation of the project

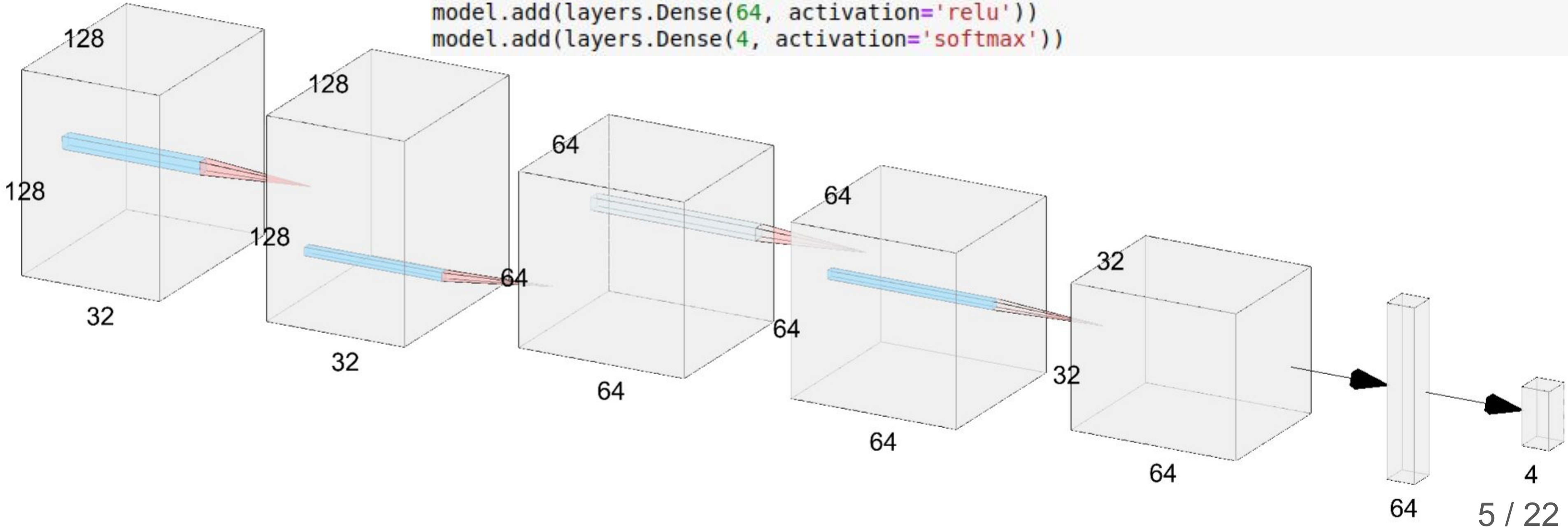
Radio telescope name	Radio telescope exterior	Bitrate per beam	Total bitrate
Effelsberg		P210-7: 11.04 Gb /s	77 Gb / s (7)
		UWB: 290 Gb / s	290 Gb / s (1)
MeerKAT		107 Mb / s	1.7 Tb / s (~1024 beams)
Square Kilometer Array		~ 1 Gb / s	20 Tb / s (>2200 beams)

Single dish radio astronomical data in a nutshell

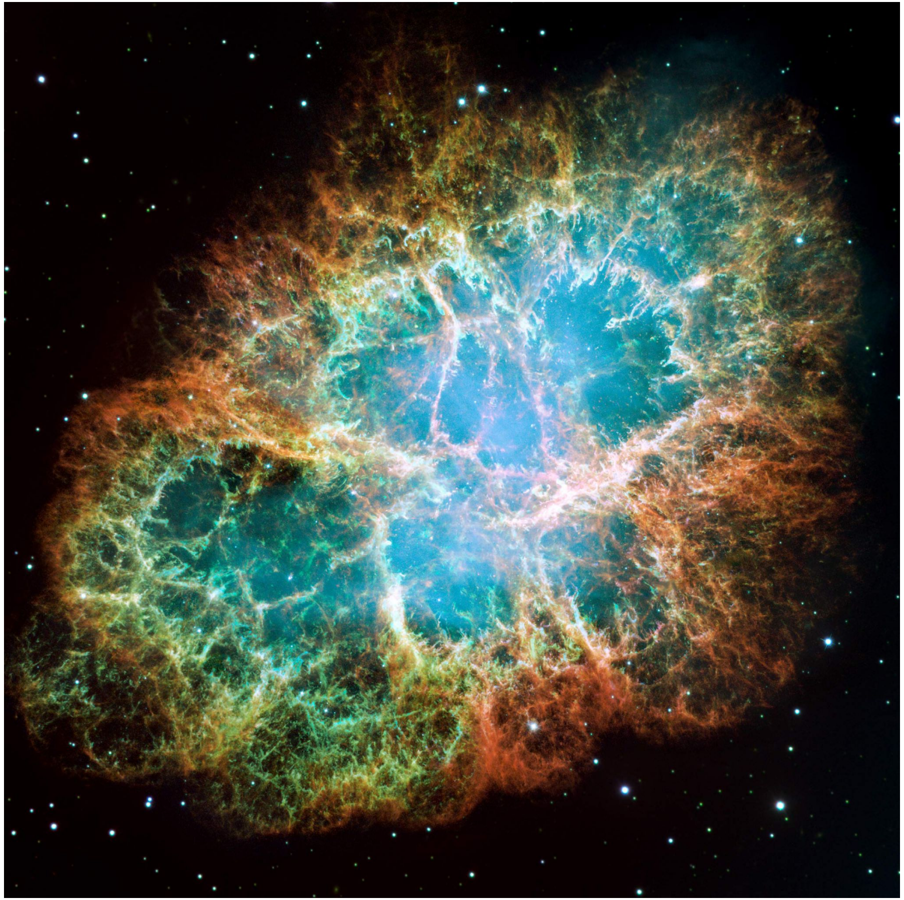


Prototype model

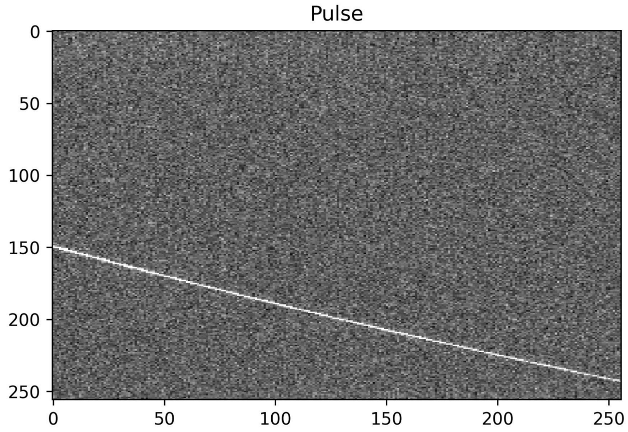
```
model = models.Sequential()

model.add(layers.Conv2D(32, (3, 3), activation='relu', input_shape=(resol, resol, 1)))
model.add(layers.MaxPooling2D((2, 2)))
model.add(layers.Conv2D(64, (3, 3), activation='relu'))
model.add(layers.MaxPooling2D((2, 2)))
model.add(layers.Conv2D(64, (3, 3), activation='relu'))

model.add(layers.Flatten())
model.add(layers.Dense(64, activation='relu'))
model.add(layers.Dense(4, activation='softmax'))
```



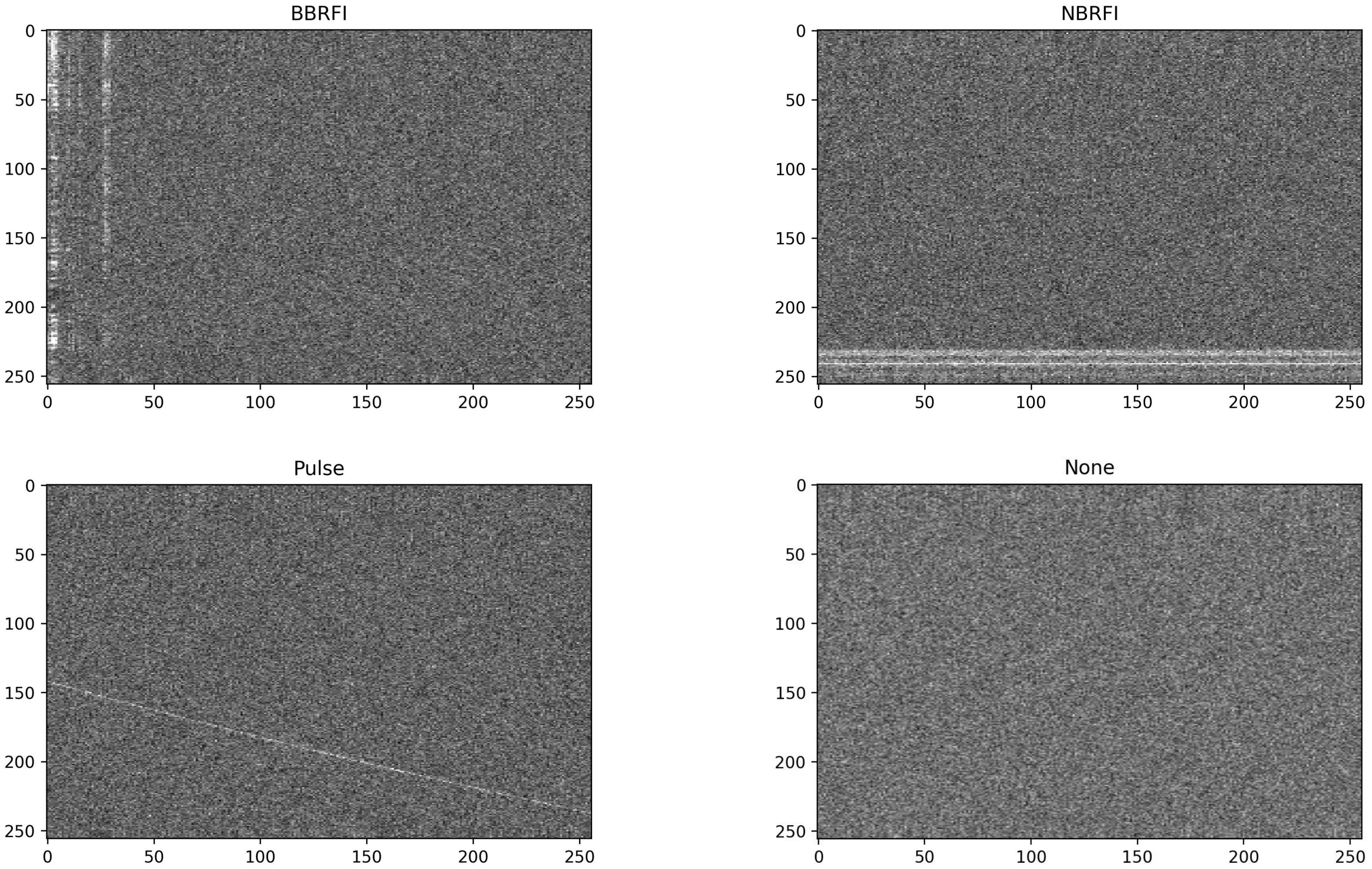
Dataset for prototype training



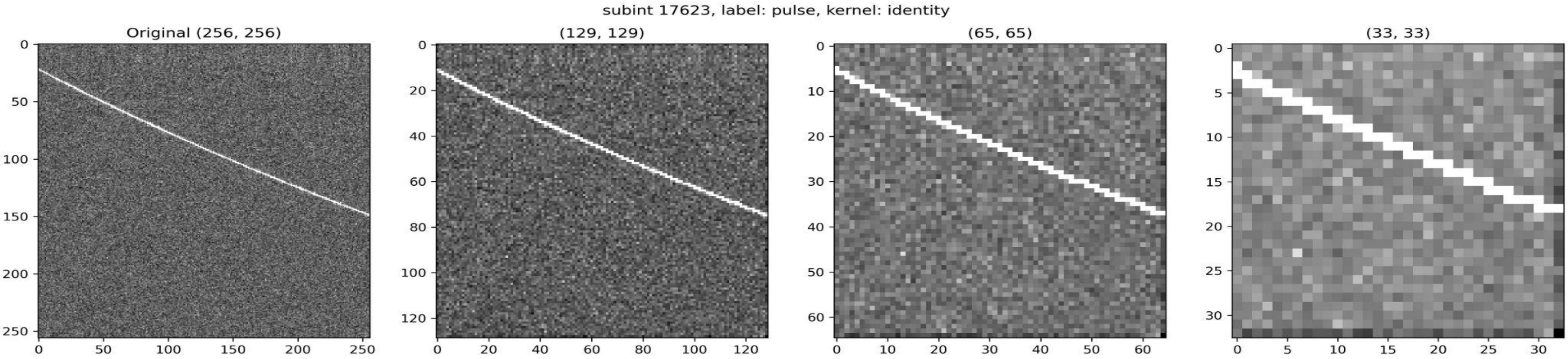
Object: **Crab pulsar (B0531+21)**
Data: **2020-05-31**
Time resolution: **0.1024 ms**
Telescope: **Effelsberg**
Number of subintergations: **50 000**
Number of labeled subintergations: **30 000**



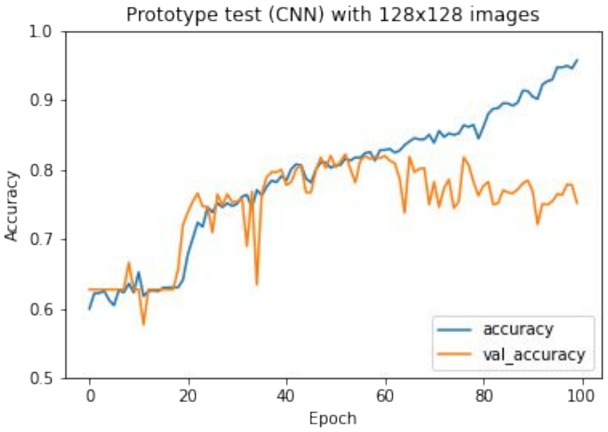
Preliminary classes for classification

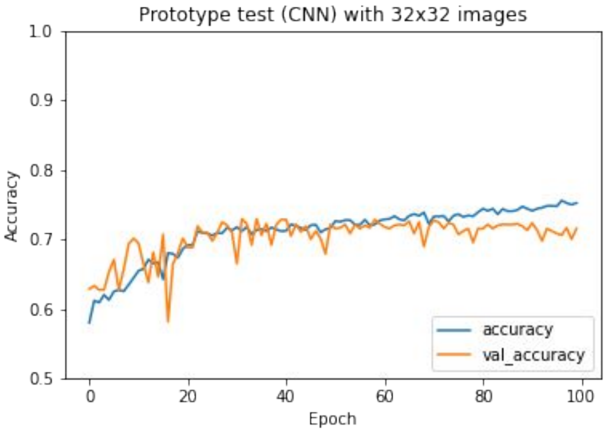


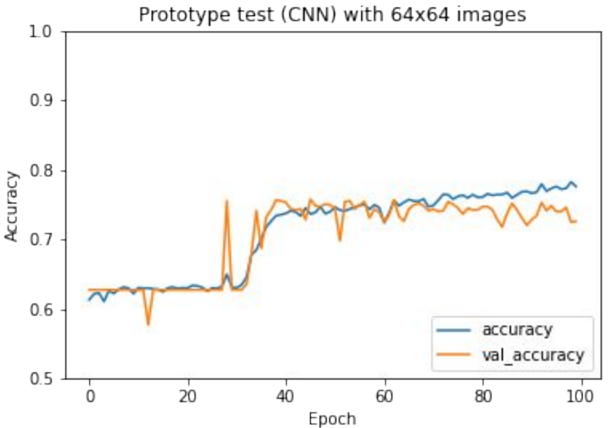
Resampling original images

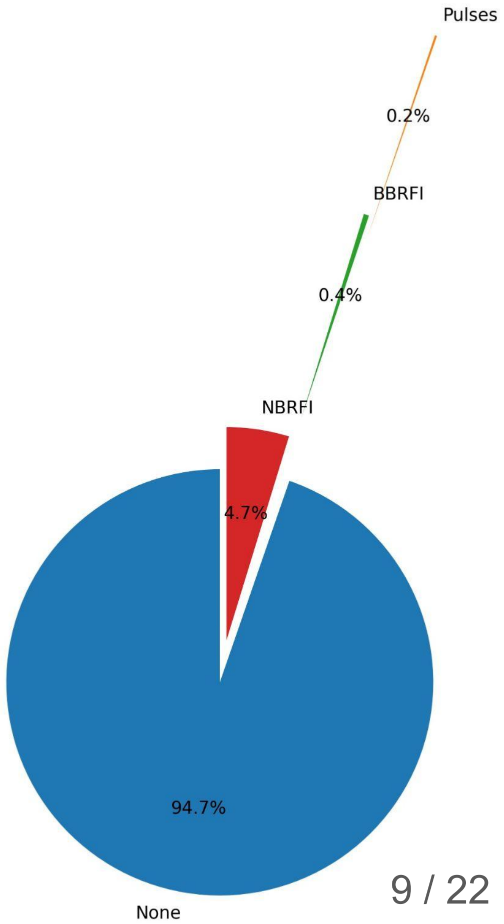


Accuracy of the prototype model

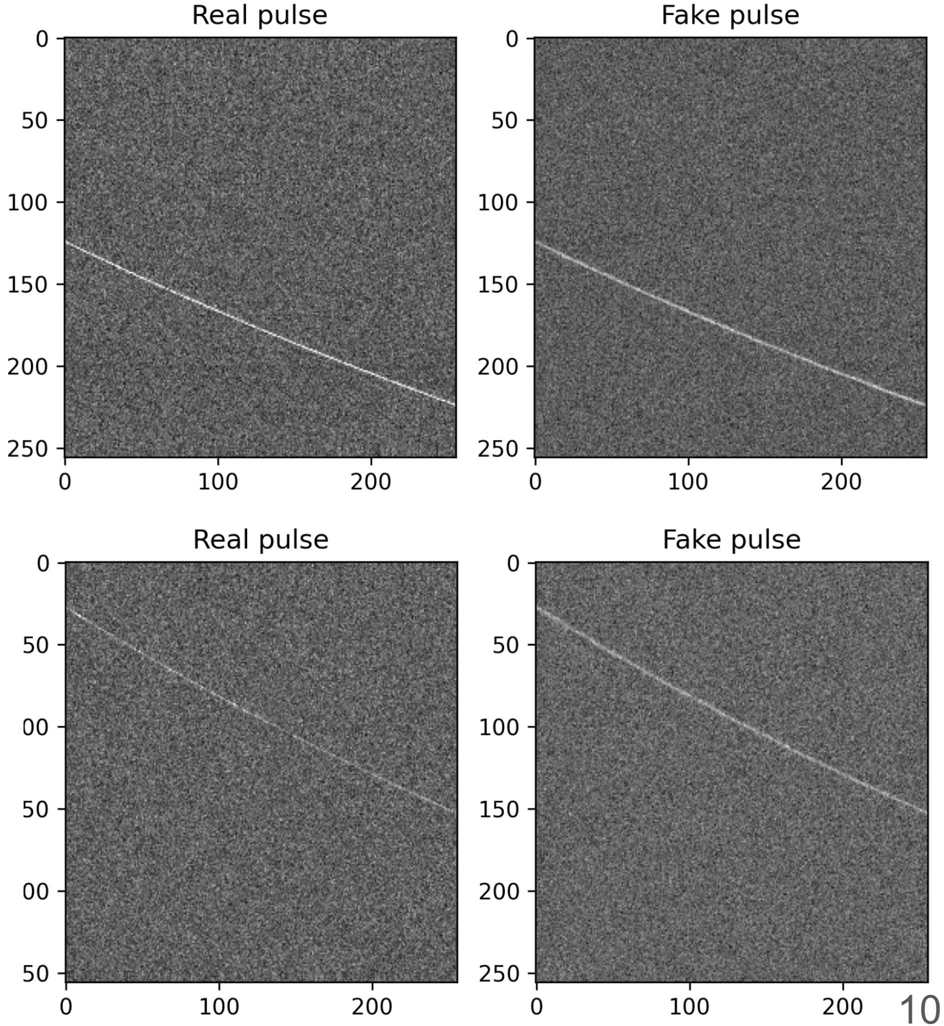








Real vs. synthetic



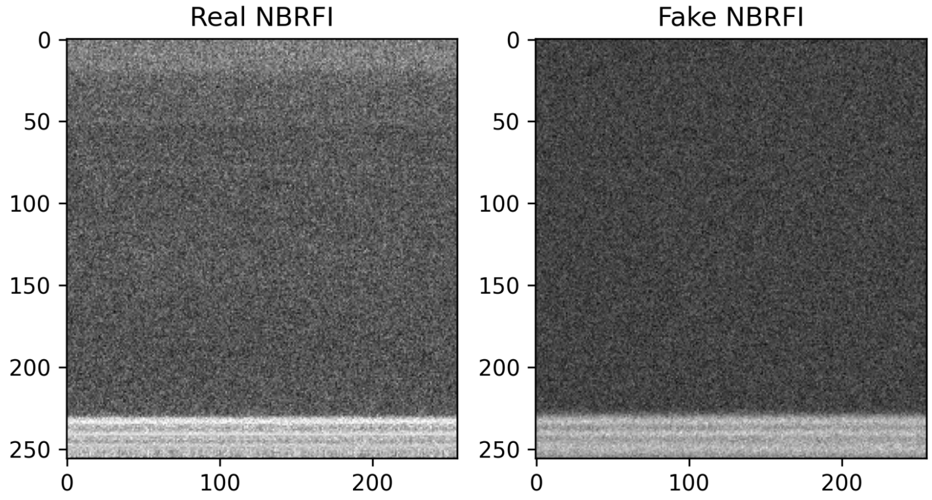
Included effects for pulses:

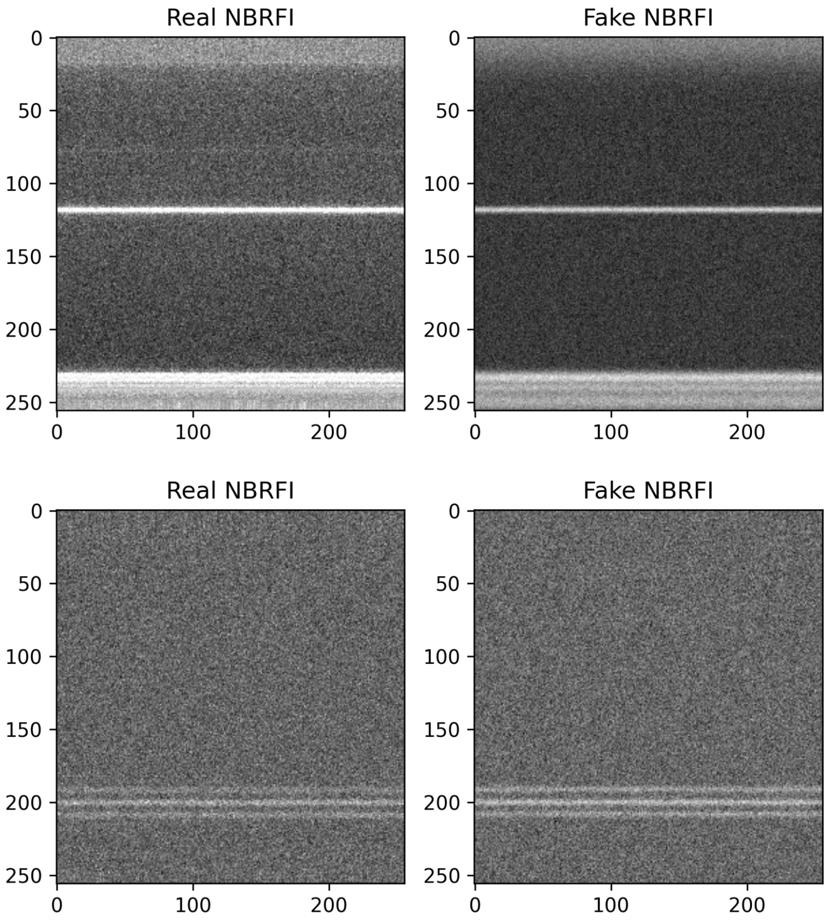
- Dispersion delay,
- Scattering,
- Spectra of a pulsar,

Background for synthetic data is uniform Gaussian noise.

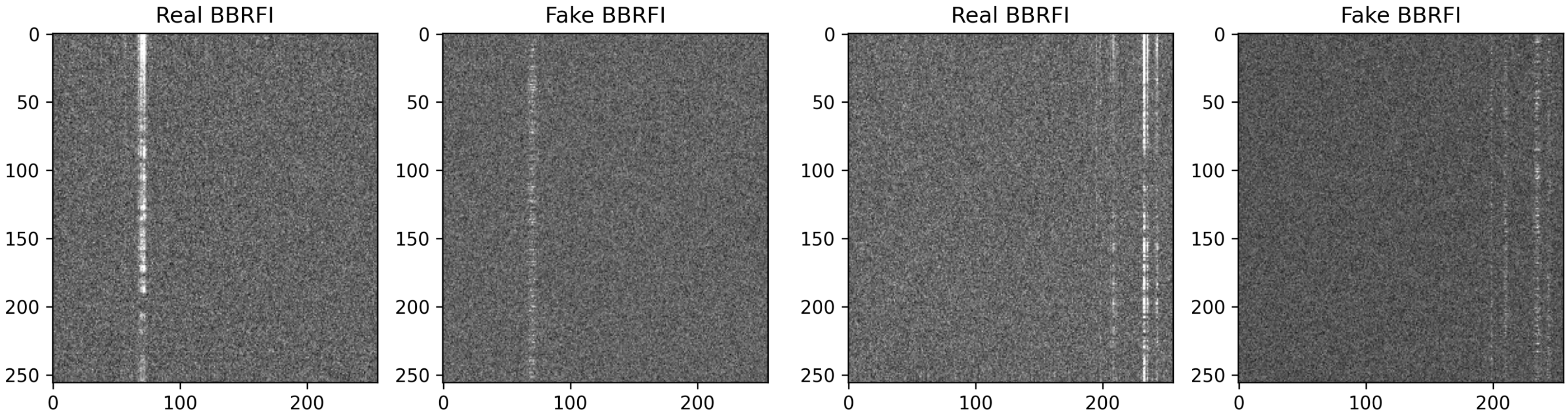
https://github.com/KazAndr/prfi_generator

Real vs. synthetic



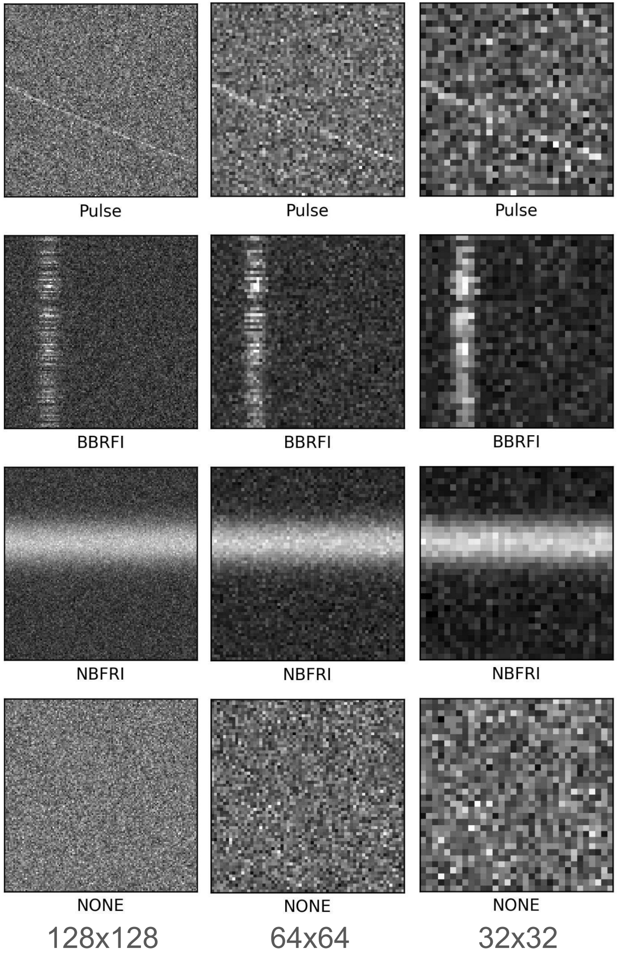


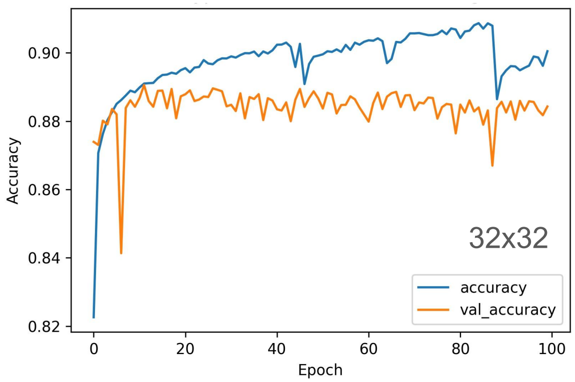
Real vs. synthetic

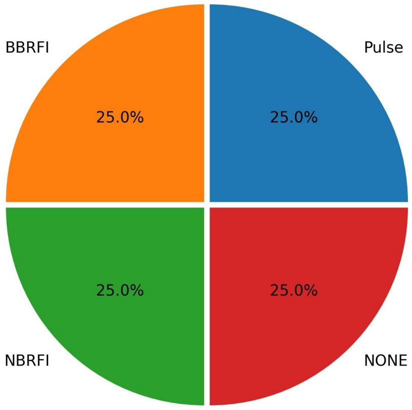


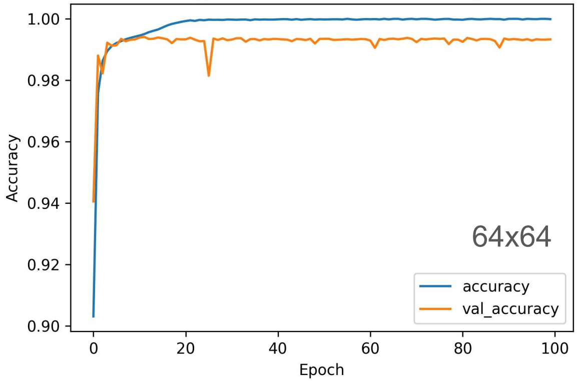
https://github.com/KazAndr/prfi_generator

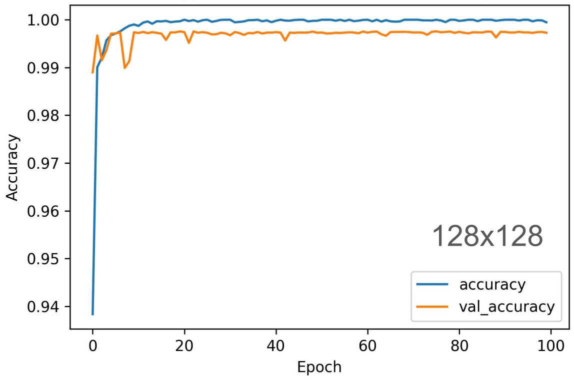
Accuracy of the model trained on synthetic data





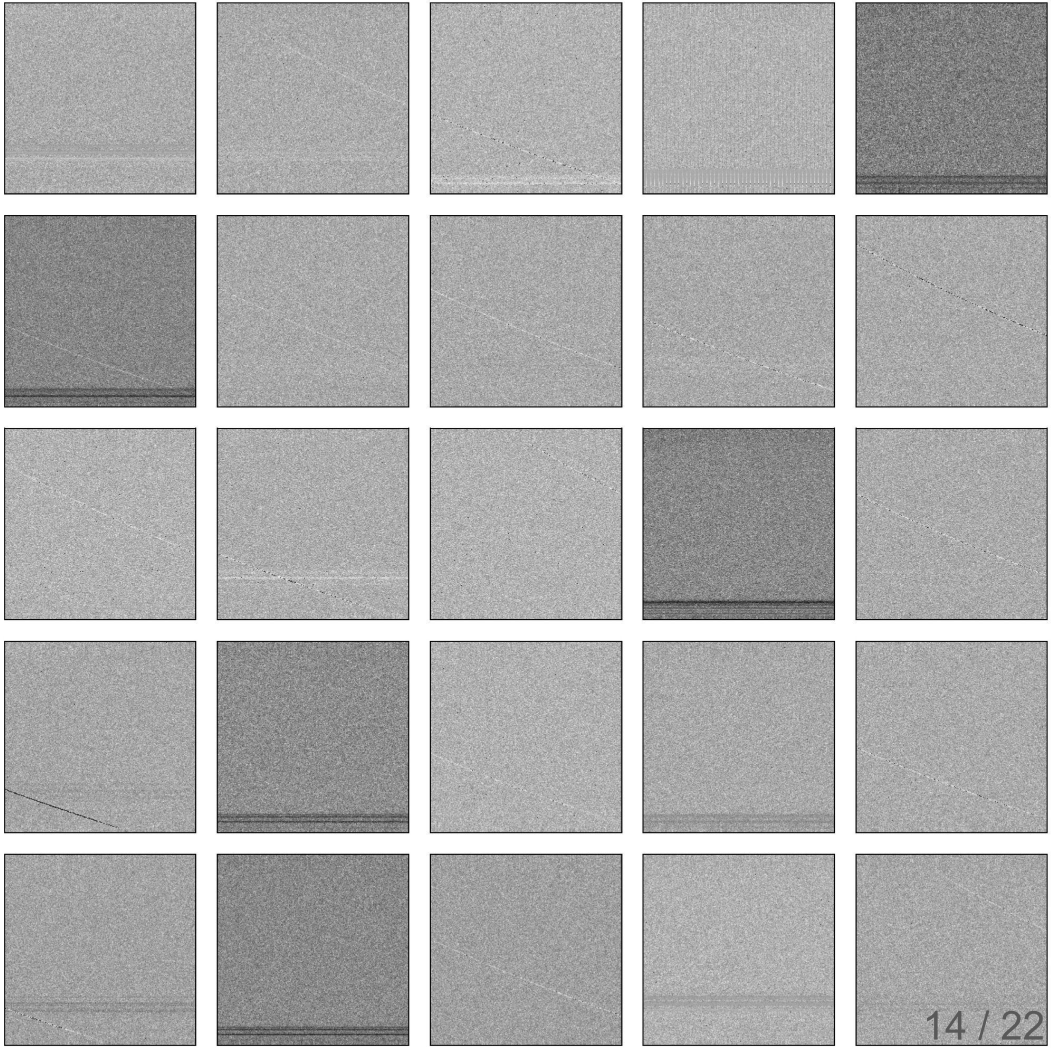






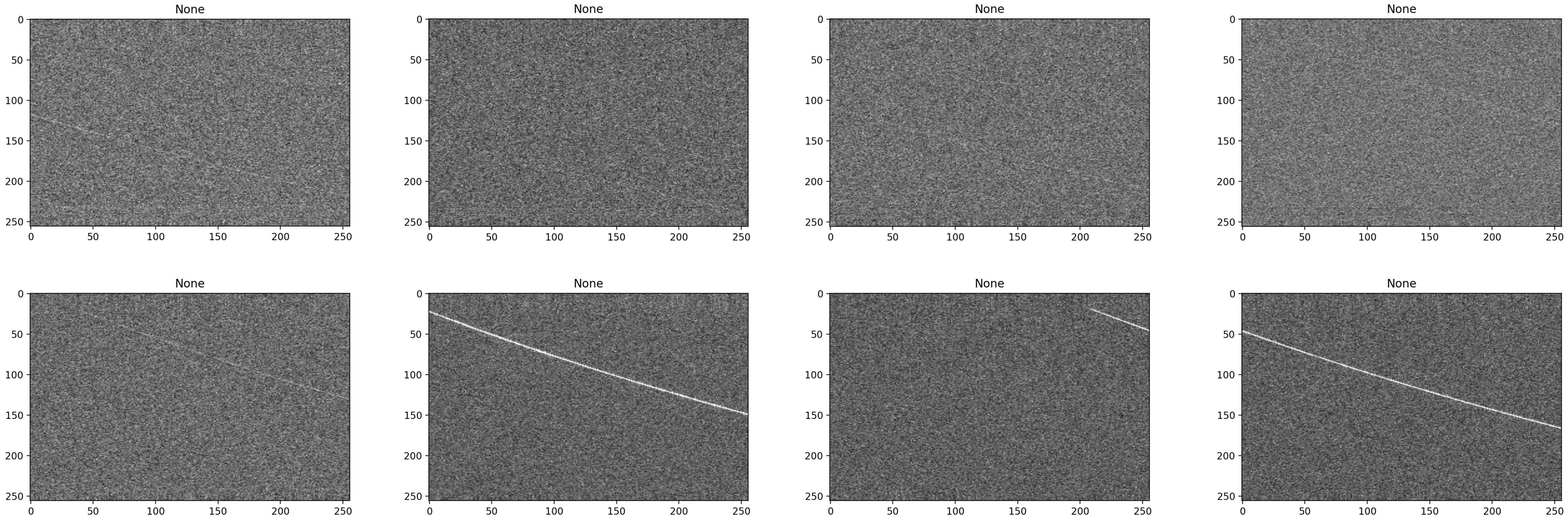
Classification of real data

Label	Accuracy	Recall	Precision	F1-Score
Pulse	0.99	0.28	0.48	0.35
NBRFI	0.72	0.47	0.98	0.64
BBRFI	1.00	0.09	0.26	0.13
None	0.71	1.00	0.62	0.77



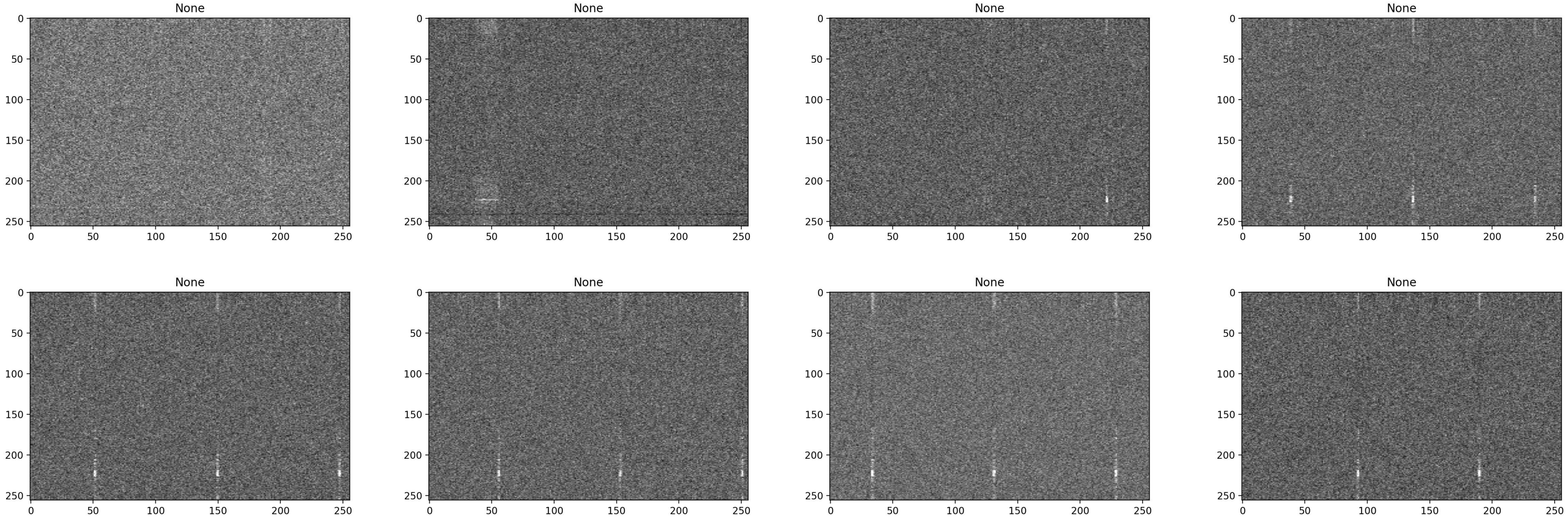
Classification of real data

FN for Pulses = 236

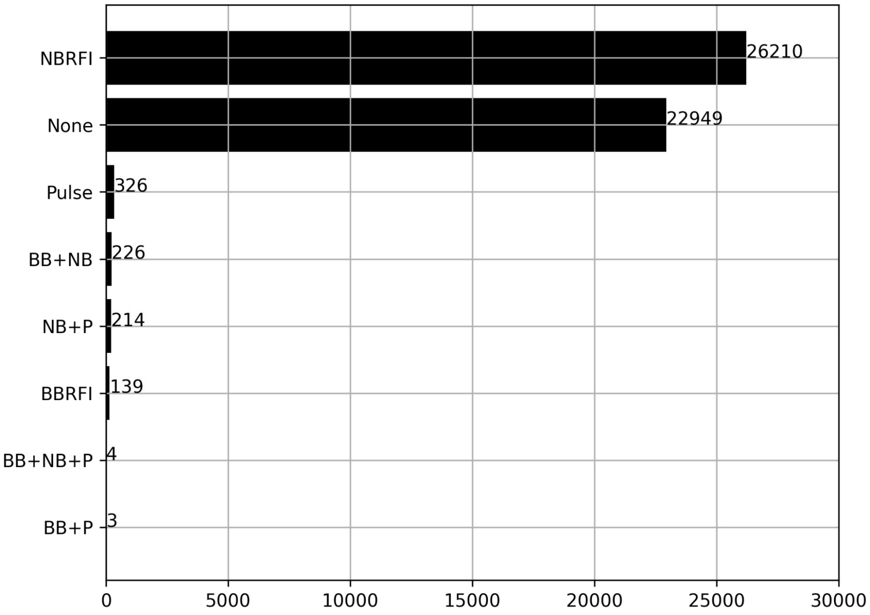


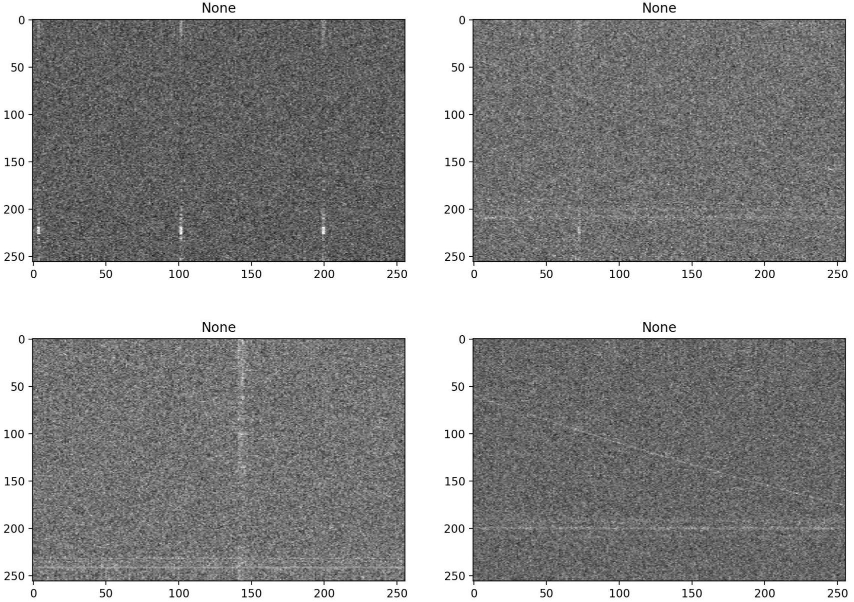
Classification of real data

FN for BBRFI = 127

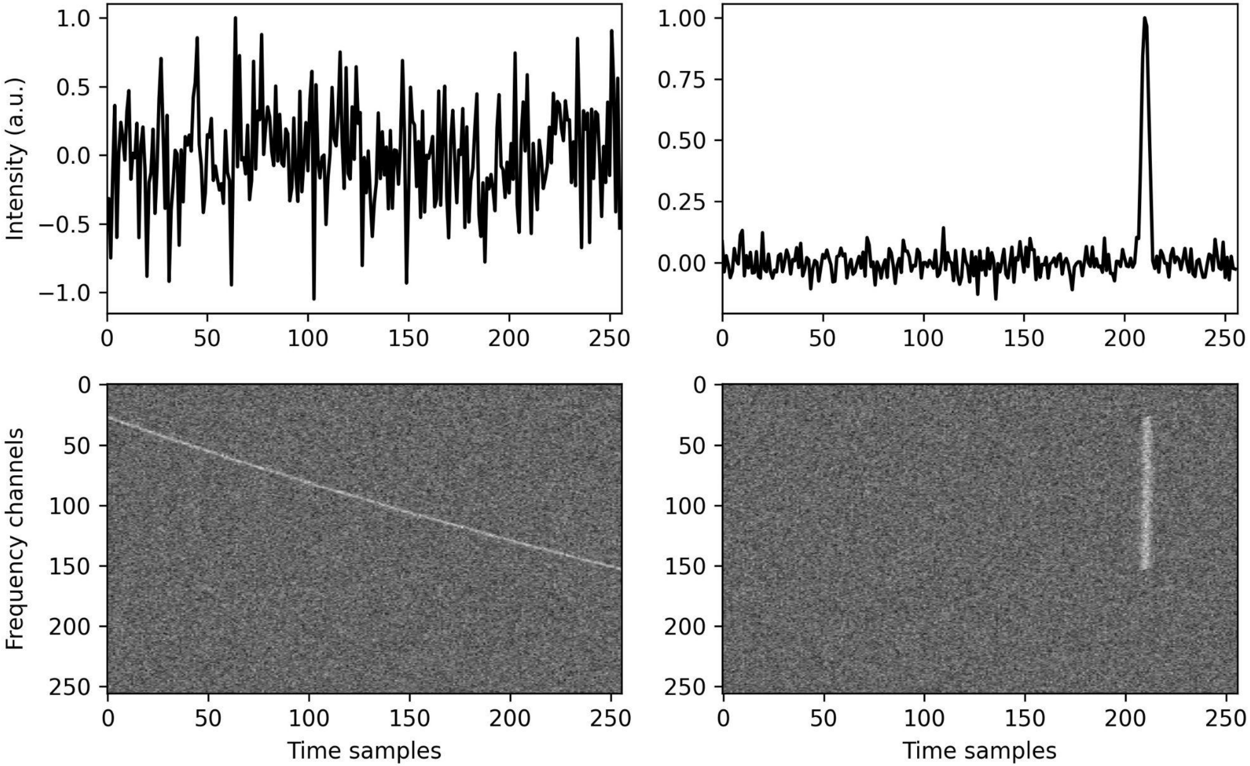


Classification of real data (mixed classes)

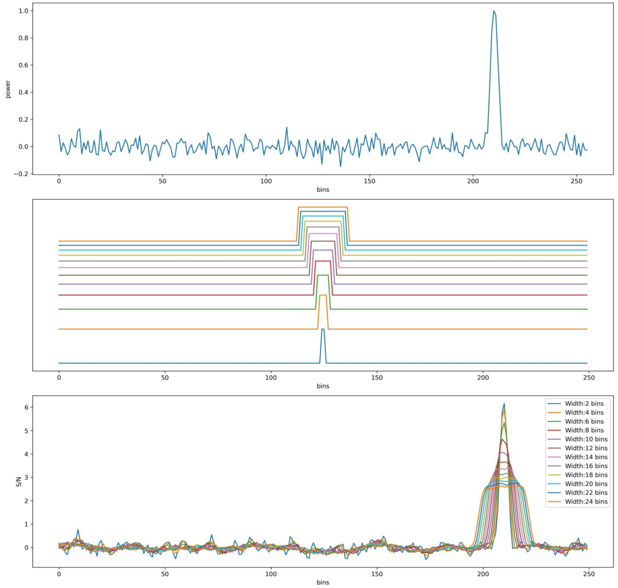




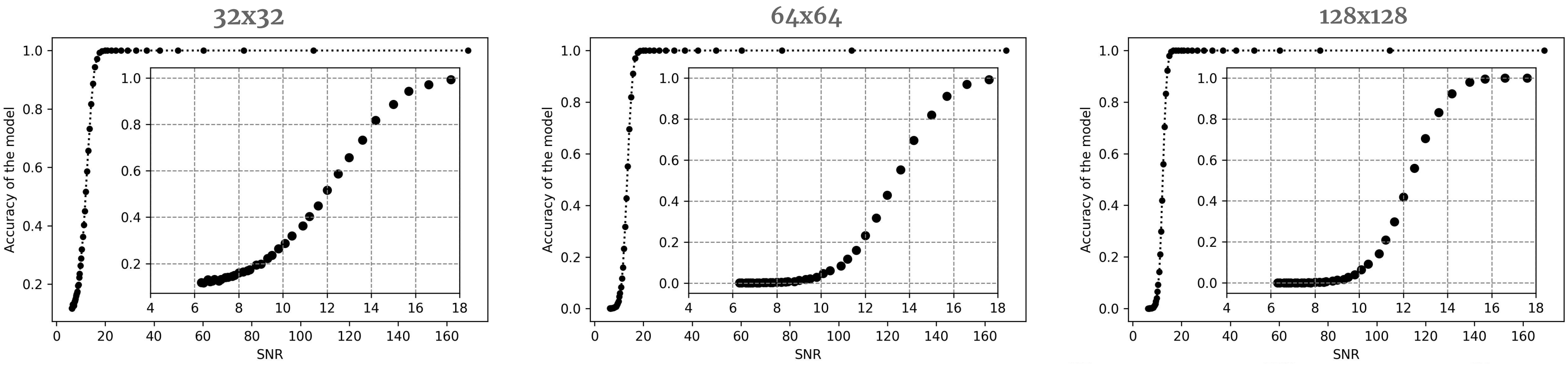
Classification of real data

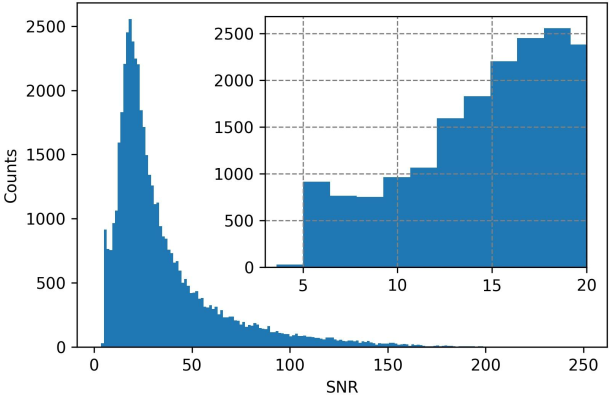


$$SNR = (I_{max} - I_{mean}) / \sigma$$

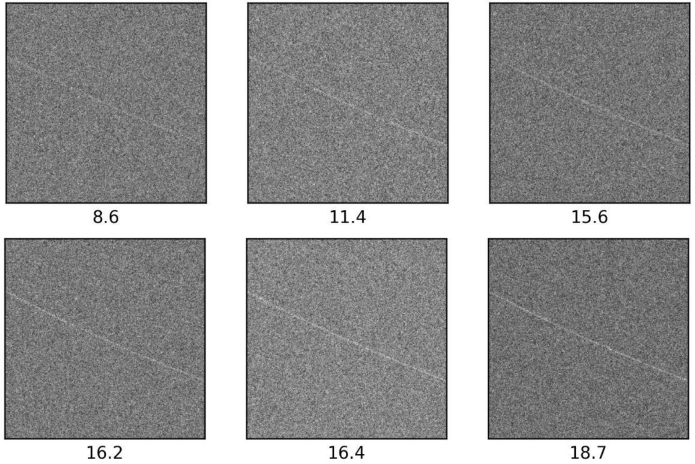


Signal-to-Noise Ratios (SNR) to which the model is sensitive



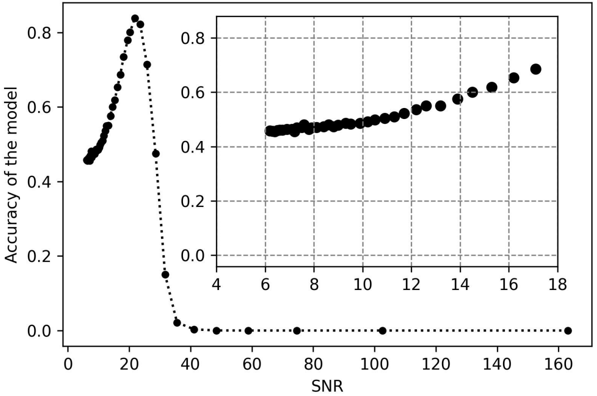


The model trained on synthetic data including pulsar pulses (also synthetic) with DM 57 and the SNR distribution shown on the right image has an efficiency of about 100% with a signal-to-noise ratio of about 16-18, depending on the resolution of the input image.

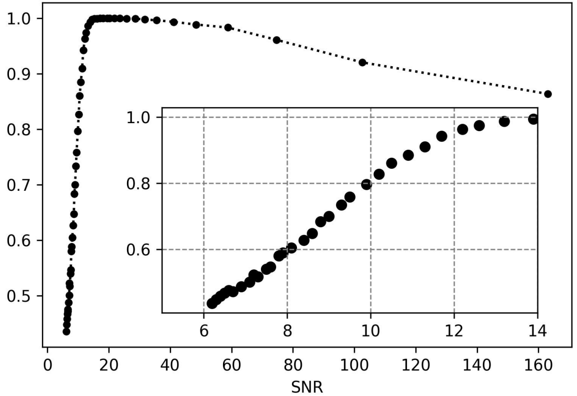


Signal-to-Noise Ratios (SNR) to which the model is sensitive

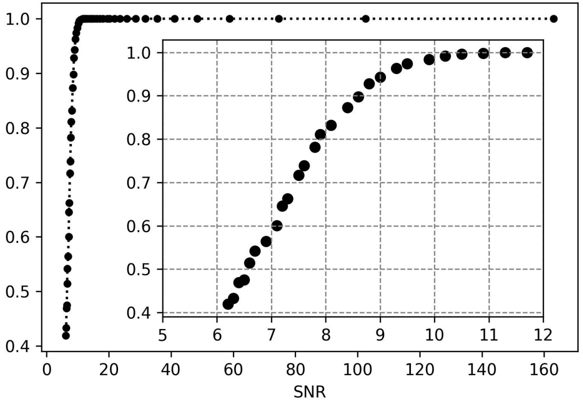
32x32

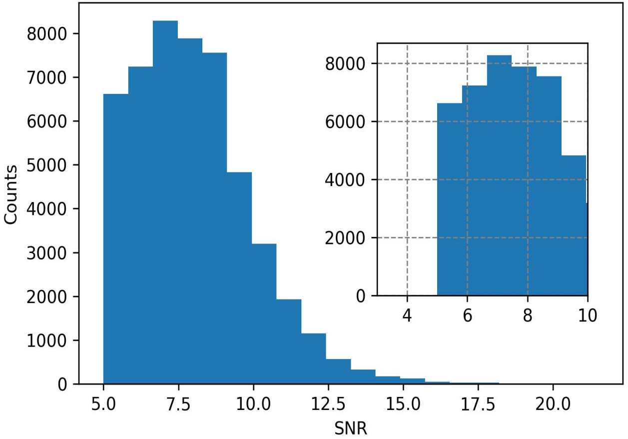


64x64

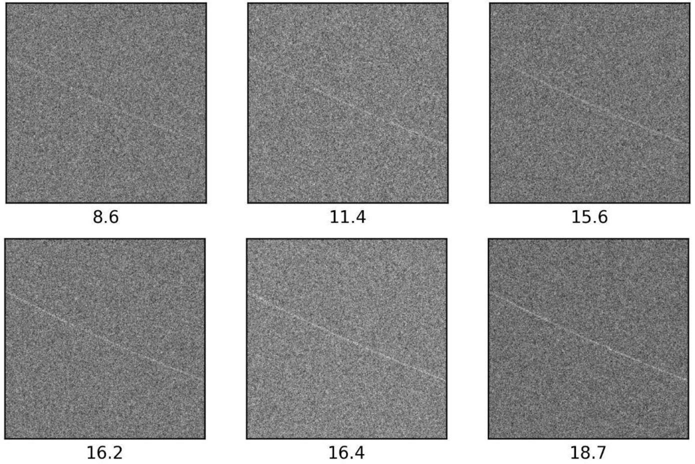


128x128

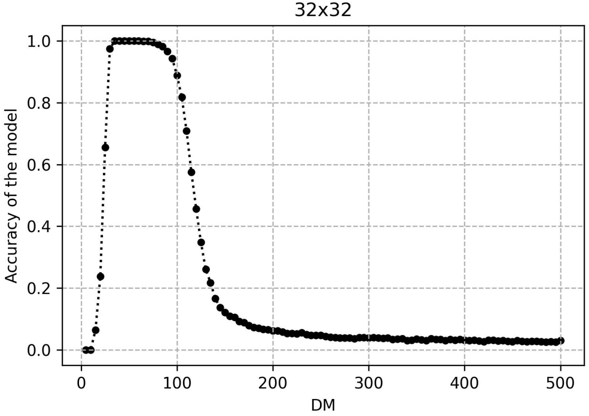


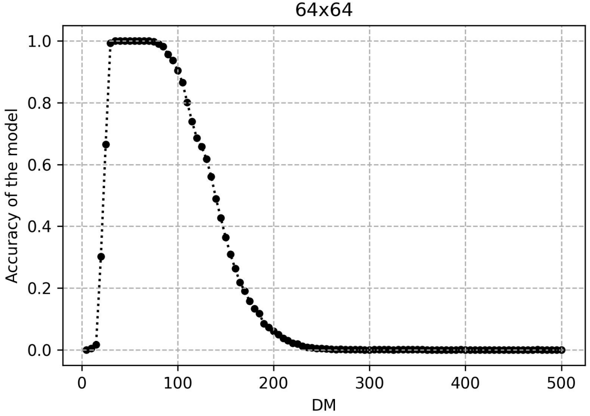


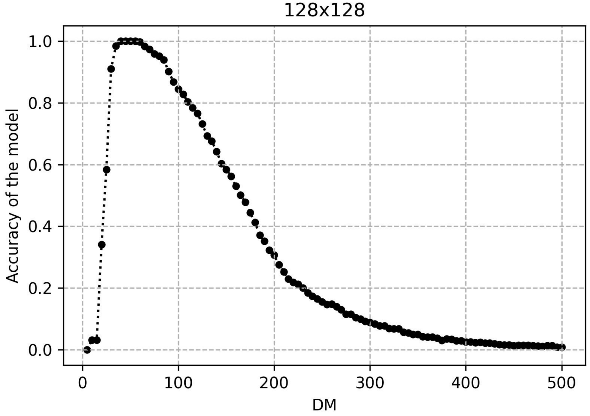
When we change the distribution of the mentioned pulses of the pulsar in the direction of reducing the ratio of their signals to noise, the sensitivity of the model is expected to change, although for 32x32 and 64x64 images we see artifacts in the model accuracy plots.



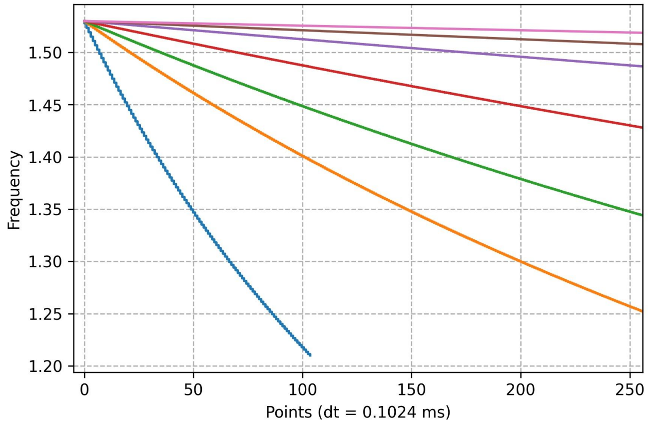
Dispersion measures to which the model is sensitive







To train the model, synthetically pulses of a pulsar with a dispersion measure of 56.758 (the dispersion measure of a pulsar in the Crab Nebula) were used. However, after analyzing the sensitivity of the model to pulses with different measures of dispersion (5 - 500 pc cm-3), it was obtained that the model trained on pulses with one measure of dispersion remains sensitive to some fairly wide range of DM.



DM_list = [10, 30, 50, 100, 250, 500, 1000]

Conclusions

- A program for generating synthetic pulses of pulsars and synthetic radio frequency interferences has been developed.
- A prototype deep-learning model was trained on a sample of synthetic pulses and radio frequency interferences.
- The model exhibits 100% efficiency for pulses with a Signal-to-Noise Ratio (SNR) greater than 18.
- It is shown that the prototype model trained on synthetic pulses with a measure of the dispersion of 57.758 pc cm-3 is able to classify as a pulse, pulses in a sufficiently wide range from the above measure of dispersion.
- Testing on real data has shown that the prototype is quite successful in classifying the corresponding data.