Learning to fingerprint : physical layer identification

Cyrille Morin[1] Thibaud Vial Leonardo S. Cardoso[1] Jakob Hoydis[2] Jean Marie Gorce[1]

CITI Lab Inria

Nokia Bell Labs

02/07/2018

Motivations

In IOT Context

- Constrained resources
- High cost of overhead

Implicit Identification

- Increase payload capacity
- ID spoofing made difficult
- May be privacy issue

Identify source of signal

Possible identification sources

- Transmitter "voice" (RF layer noise)
- Channel characteristics

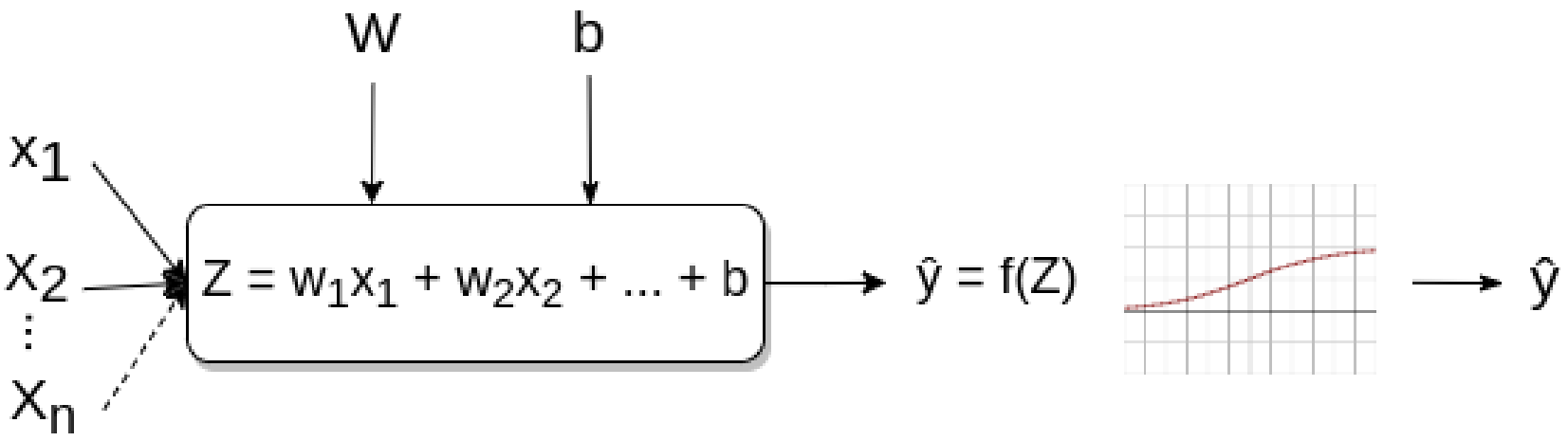
Why neural networks

- Task close to Voice recognition problems
- Predominance of Deep learning in modern audio processing

The benefits of CorteXLab

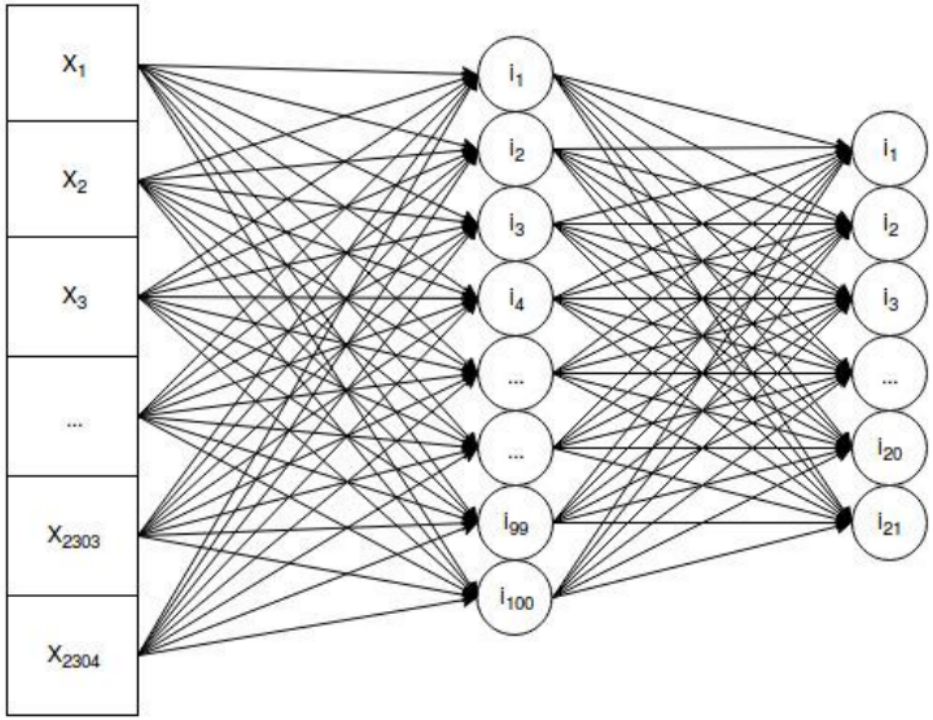
- Online access to several nodes
- Promise of reproducible experiments
- Automatable experiments

Basic building block : the unit

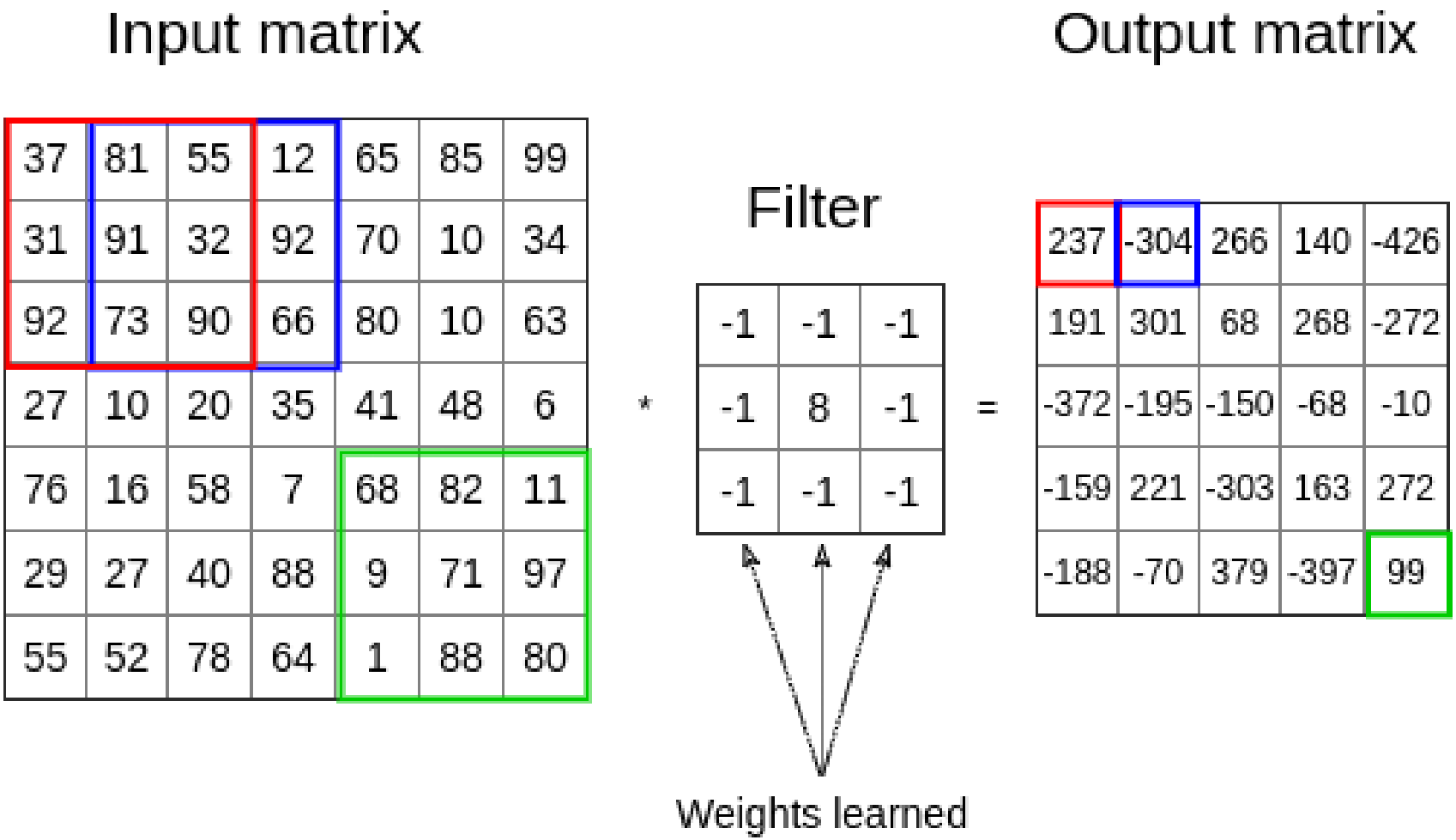


Approximate $f(x_1,x_2,...) = y$ by $\hat{y} \approx y$

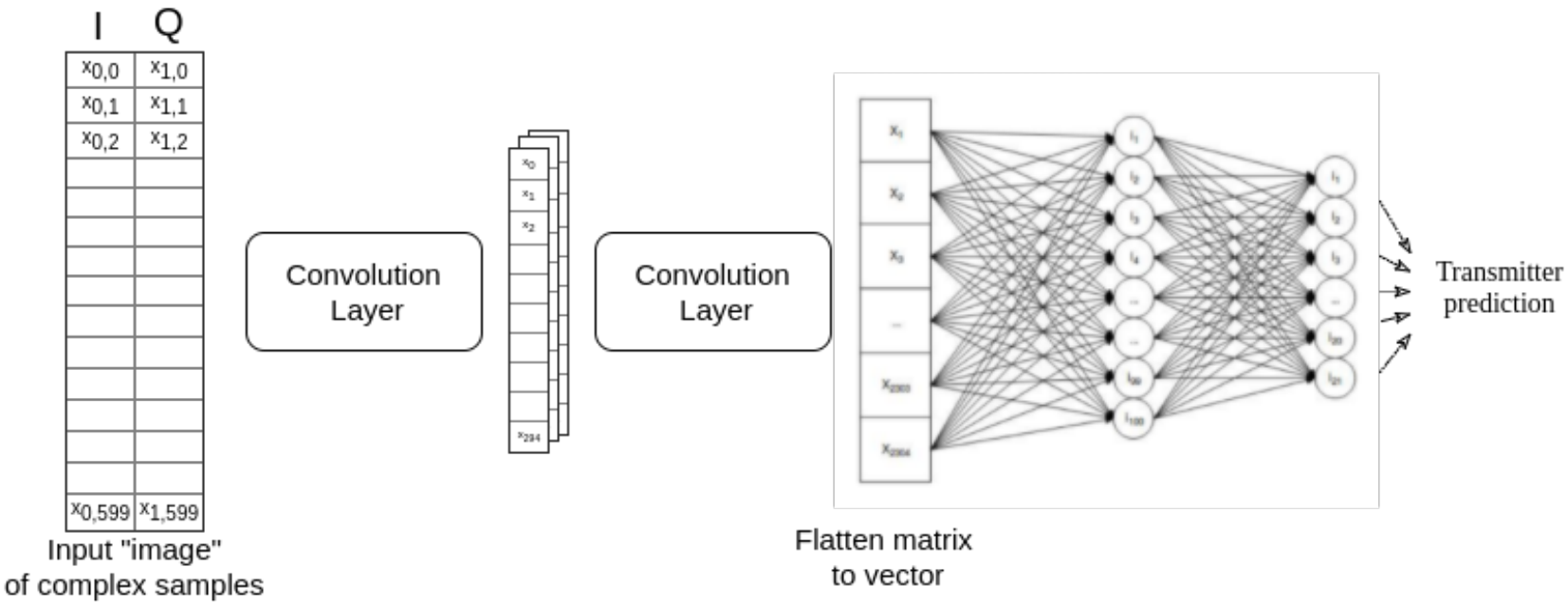
Multi Layer Perceptron (MLP) or plain neural network



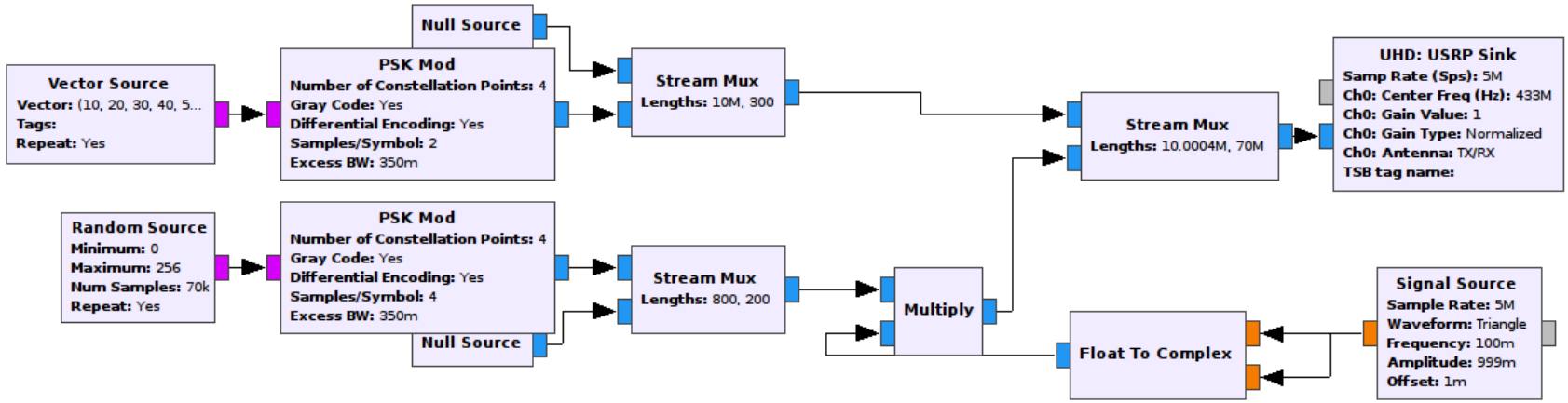
Convolutionnal layer



Neural network architecture



Generation of training data



Datasets

- Raw I/Q samples recorded by receiver for offline processing
- 21 transmitters and 50000 packets per transmitter
- $50000 * 21 = 1050000$ examples in one dataset
- Dataset examples split in 70% training, 10% validation and 20% test sets

Training and testing process

Training process

- Offline
- 100 packet examples per batch
- At least 30000 batches per training

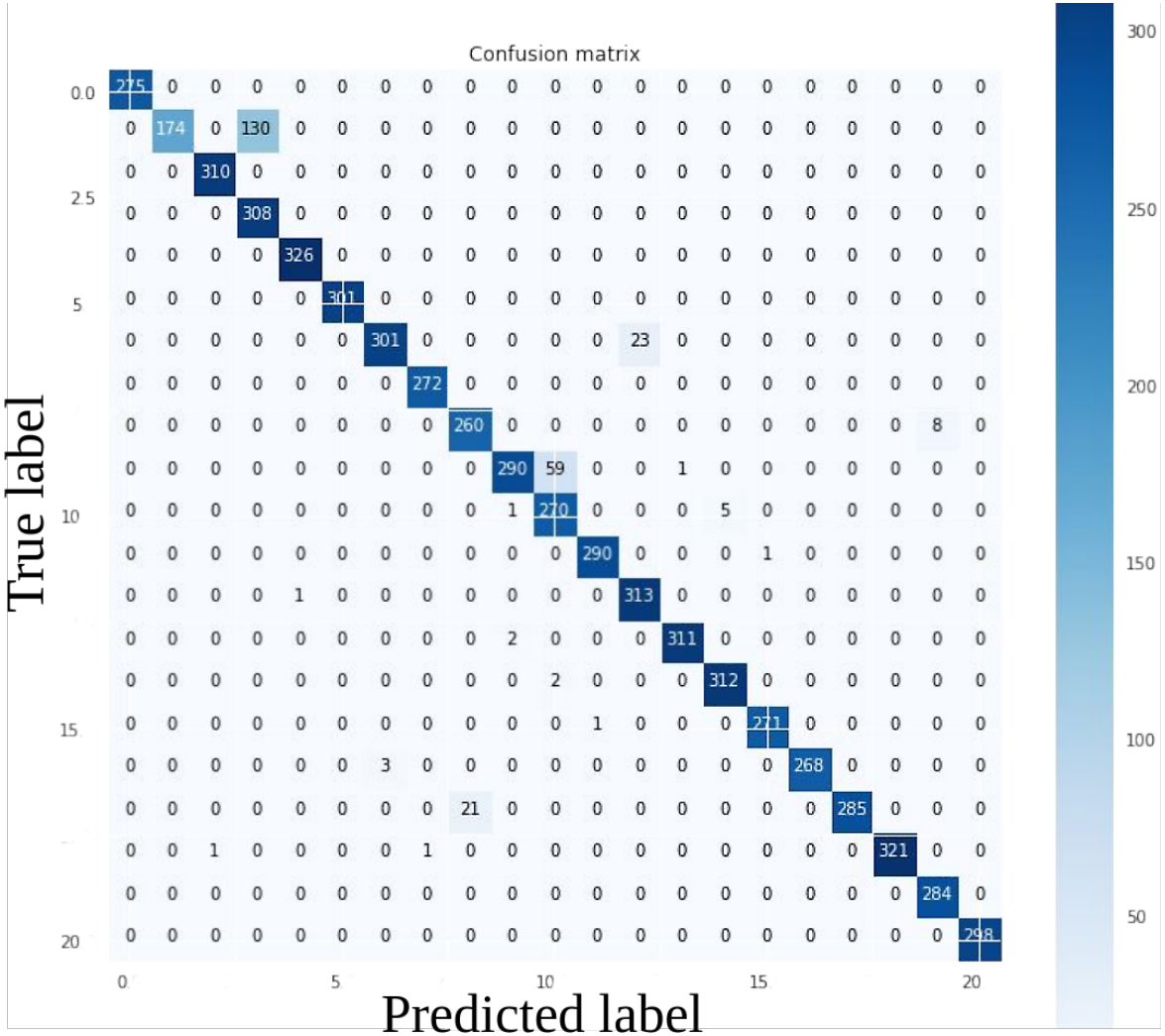
Testing

- Over same dataset — Show learning capability
- Across datasets — Show generalisation capability

Learning ability is good

Precision obtained by training and testing on same dataset					
Dataset ID	DS1	DS2	DS3	DS4	DS5
Precision	94.23%	92.08%	96.33%	97.24%	95.03%

Learning ability is good

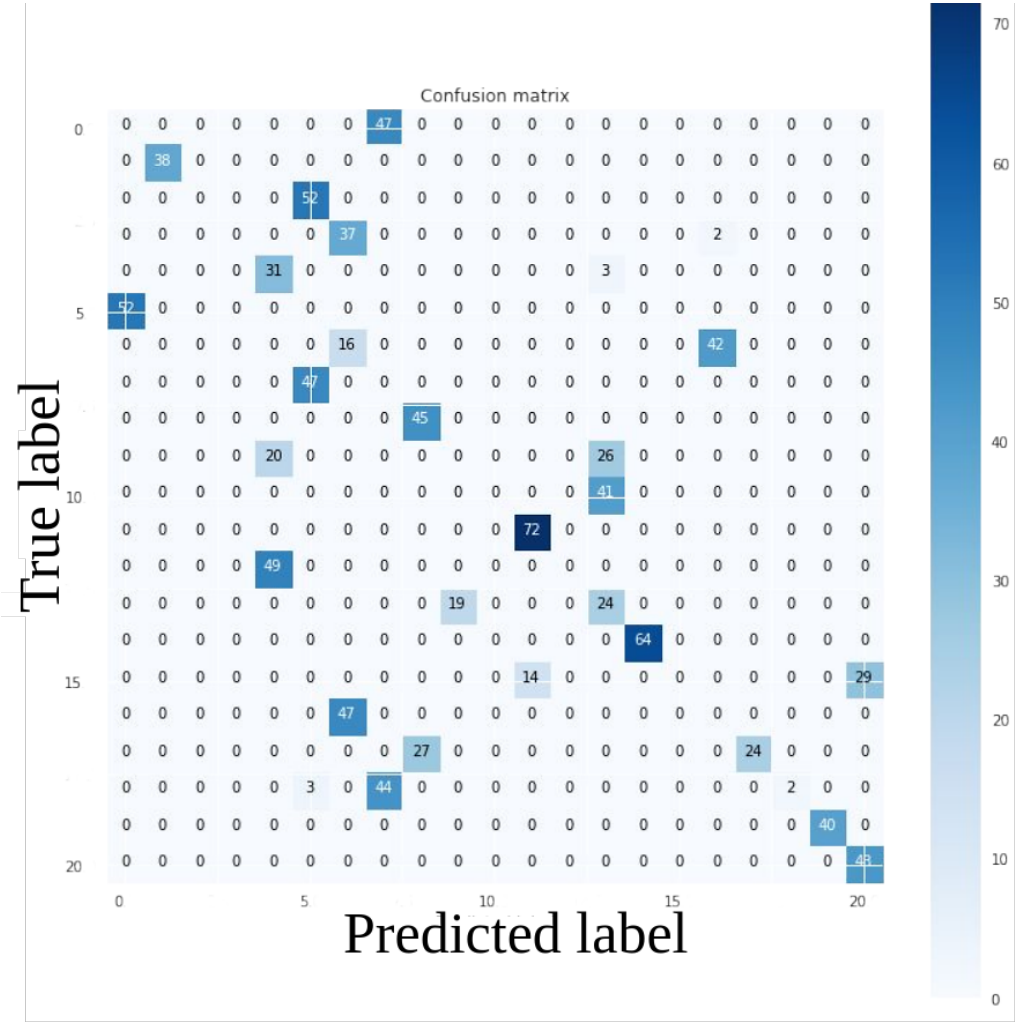


So is generalisation ..... Or is it ?

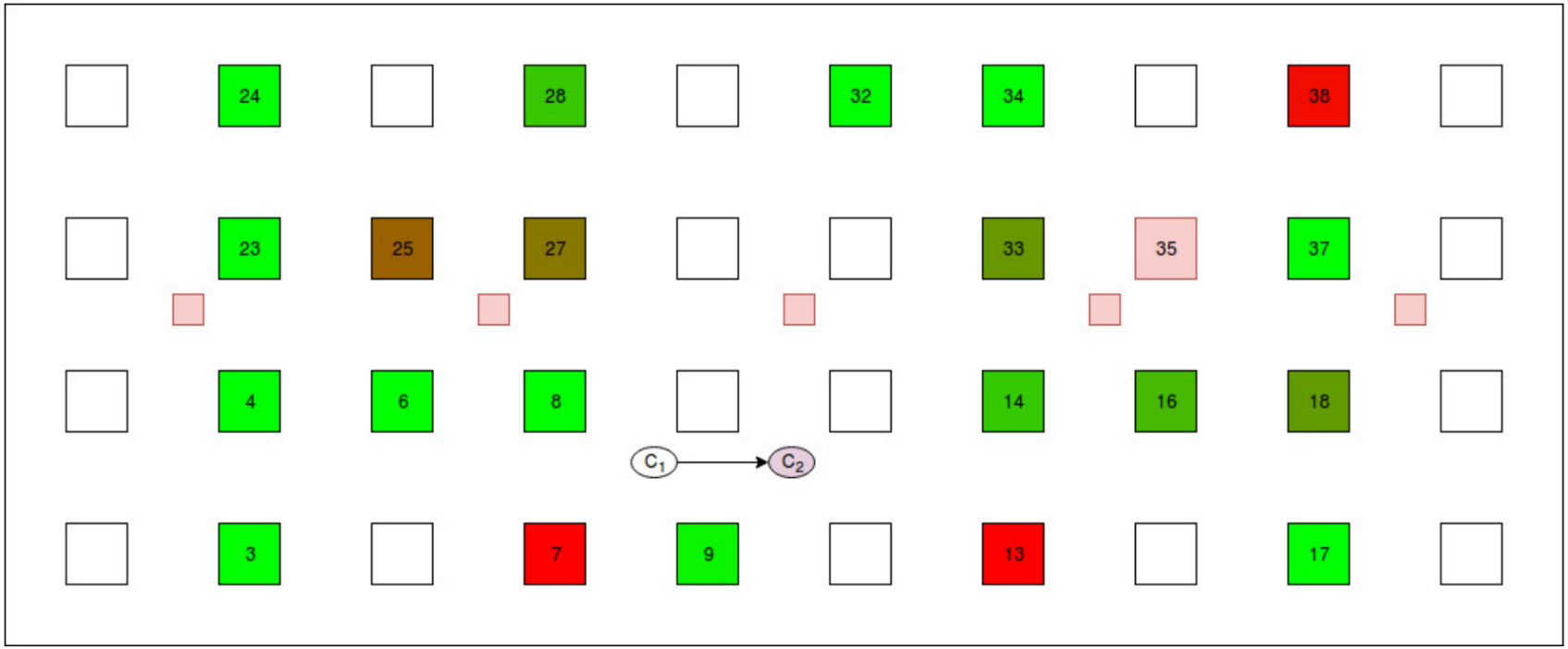
Dataset	Creation date
DS1	25/07
DS2	25/07
DS3	26/07
DS4	27/07
DS5	16/08

		Test				
		DS1	DS2	DS3	DS4	DS5
Train	DS1	97.94	98.41	97.99	97.21	35.85
	DS2	96.22	97.02	96.32	96.28	36.79
	DS3	93.34	93.97	95.18	93.45	36.77
	DS4	96.62	97.10	95.22	97.38	38.53
	DS5	42.99	43.48	43.16	40.65	94.25

The chair crisis

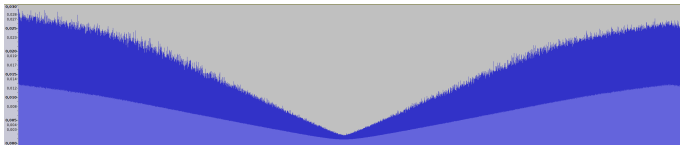


The chair crisis



We need more complete datasets

- With Power variation



80% learning accuracy
70% generalisation

- With Channel variation

- Introduce robot in the room
- Combined with power variation

Conclusion

For the neural network approach

- Reliable detection of transmitter
- Highly sensitive to radio channel characteristics

On CorteXLab's side

- Automatable experiment
- Stable channel over time

Perspectives

Improvements

- RF "Voice" identification
- Online identification

Beyond

- Extension of Deep Learning to other telecommunication aspects