

**INFRASTRUTTURE**

## **Reti Neurali per individuare anomalie nel segnale delle stazioni RSN**

*Casale P., Pignatelli A. e Lauciani V.*

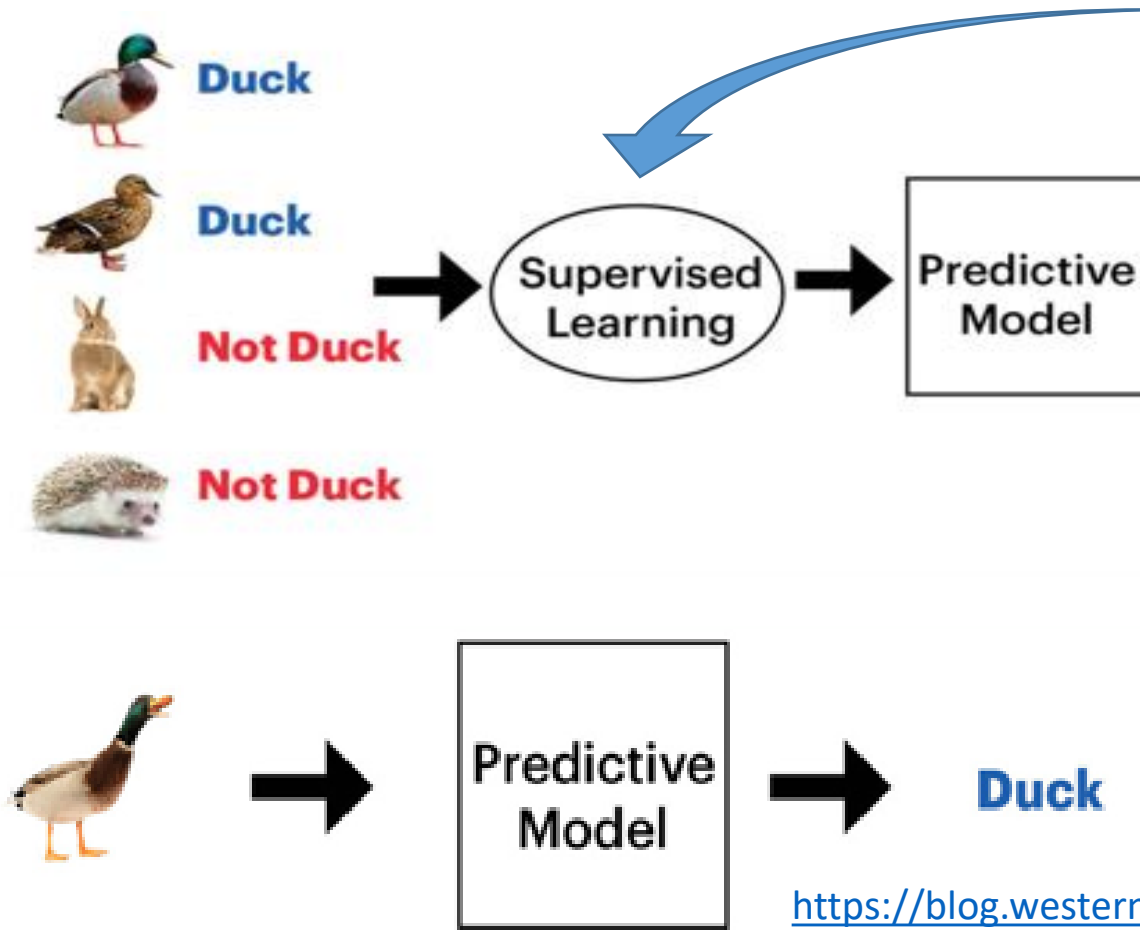
# Machine learning: definizione

Il termine **machine learning (ML)**, letteralmente “apprendimento macchina”, reso in italiano come **apprendimento automatico**, indica un ambito di ricerca all’interno dell’Intelligenza Artificiale volto a realizzare *sistemi in grado di imparare autonomamente senza la necessità di essere specificatamente programmati per svolgere tale compito.*

(Da <https://www.simonefavaro.it/2017/04/07/introduzione-machine-learning/> )

**DEFINIZIONE PRATICA: algoritmi che "imparano" dai dati già acquisiti ad effettuare "previsioni" su dati non ancora acquisiti**

Esistono algoritmi di *machine learning* che simulano il comportamento del cervello.  
Es. *Reti Neurali* che imparano a "classificare" le immagini sulla base dell'**esperienza** fornita da altre immagini **già classificate**



Immagini **già classificate** Con l'**esperienza** (training) si costruisce il modello predittivo

Il modello predittivo ricavato è successivamente utilizzato per "indovinare" la classificazione di altre (nuove) immagini

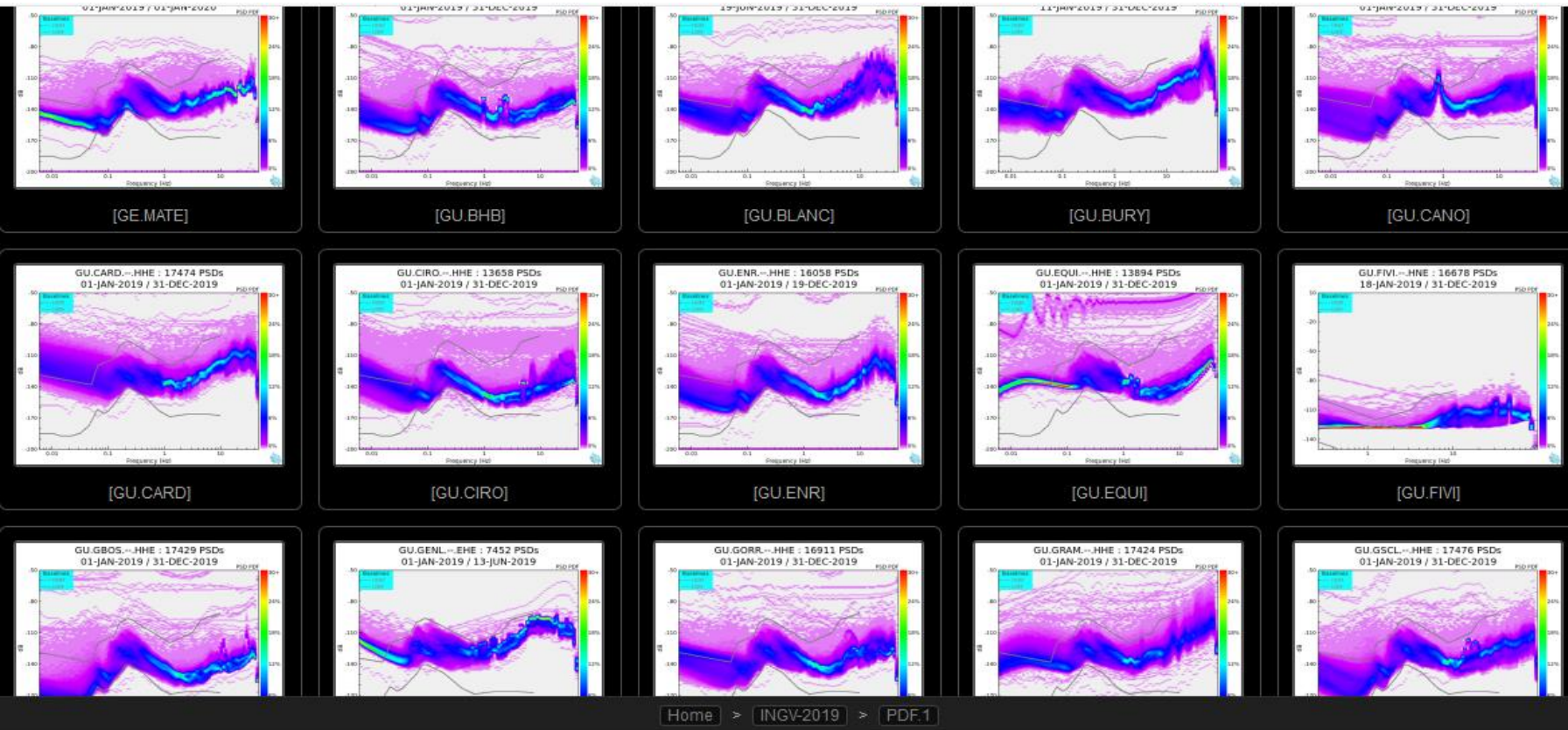
<https://blog.westerndigital.com/wp-content/uploads/2018/05/supervised-learning-diagram.jpg>

# PERCHÉ UTILIZZARE RETI NEURALI NELL'ANALISI DELLA QUALITÀ' DEI DATI ?

Spesso si devono analizzare immagini (sismogrammi, spettri...)

Dalle 500 stazioni RSN, il pacchetto SQLX produce tantissimi diagrammi

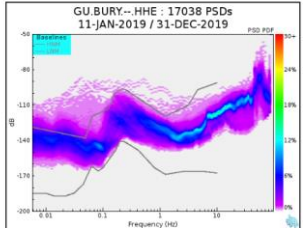
<http://sqlx-srvr.int.ingv.it/>



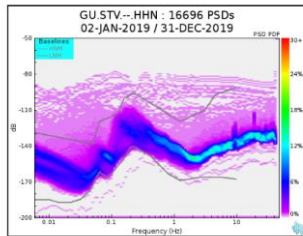
...di cui quasi 3000 da analizzare !!! →

→ Proviamo a istruire una rete neurale...

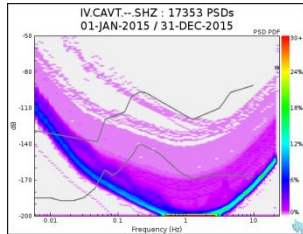
## Immagine già classificate



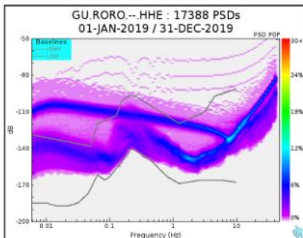
OK signal



OK signal



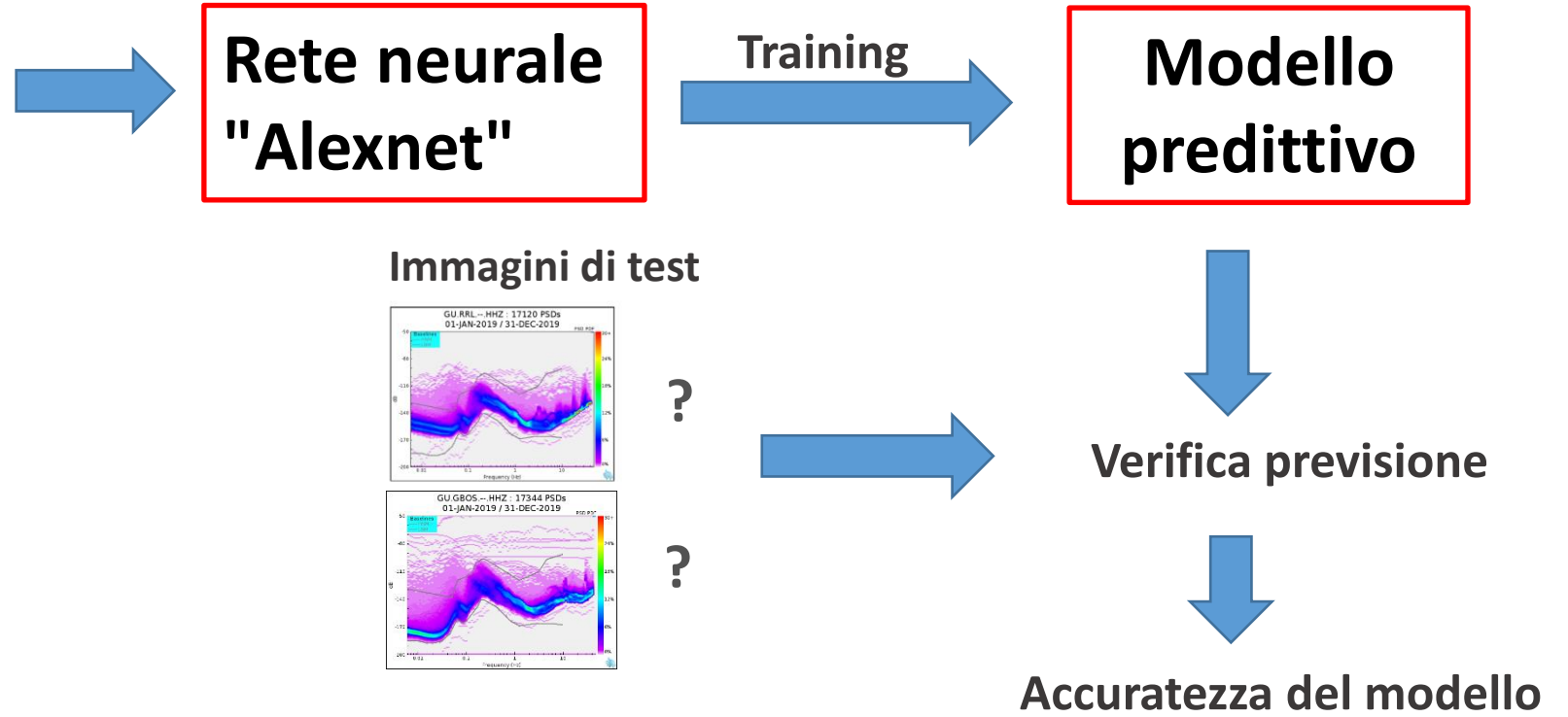
BROKEN signal



BROKEN signal

Diagrammi SQLX: come prima

→ classificazione «*ok signal*» o «*broken signal*»  
invece di «*duck*» o «*not duck*»



## COME ABBIAMO OPERATO

TRAINING  
RETE NEURALE



Software matlab implementato per adattarlo alla rete neurale.  
Input= immagini classificate. Prodotto= modello predittivo

INPUT PER  
IL TRAINING



- 300 spettri annuali del **2019** già classificati: **150 «OK»** e **150 «Broken»**
- 219 usati per il training

DATI PER IL  
TEST  
SPEDITIVO



- 81 spettri già classificati non usati nel training ma solo per il test

VERIFICA  
GENERALE



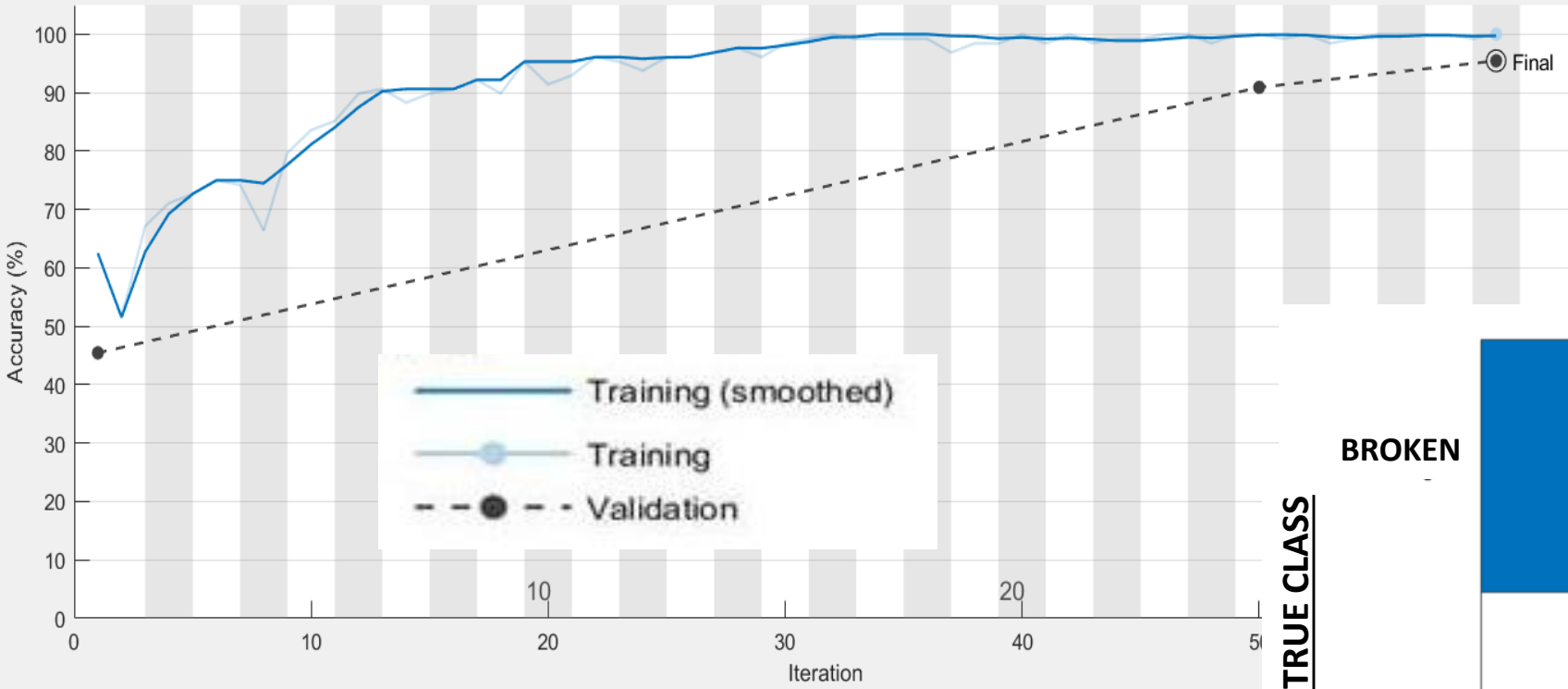
- Download di 1800 diagrammi (900 del 2018 e 900 del 2019) da intranet INGV tramite *web-scraping* (implementata in Python )
- Applicazione del modello predittivo su 1800 diagrammi →
- report pdf per il 2019 e 2018 indicanti la classificazione di ciascuno spettro (procedura matlab)
- Verifica delle previsioni: controllo (per ora) di **390** spettri **2018** (130 stazioni)



# RISULTATI DEL TEST SPEDITIVO

## TRAINING & TEST

Training Progress (22-May-2020 14:05:37)



← **ACCURATEZZA = 95%**  
**(77 successi su 81)**

TRUE CLASS

BROKEN

45

4

OK signals

32

BROKEN

OK signals

PREDICTED CLASS

Risultato del test:  
"matrice di confusione"

# RISULTATI DEL TEST FINALE

Dei 390 spettri (circa 130 stazioni) verificati: **93%** azzeccati e **7%** sbagliati

Differenza tra le reti sismiche IV+MN e GU:

IV+MN: **95%** azzeccati (**5%** sbagliati)

GU : **84%** azzeccati (**16%** sbagliati) Perché ?

## SVILUPPI FUTURI

Gli spettri delle stazioni a Larga Banda (canali HH\*) sono simili tra loro e molto diversi dalle stazioni a corto periodo (canali EH\*, SH\* ...) →

→ **Prima fase: solo le stazioni Larga Banda (canali HH\*): circa 60% delle stazioni.**

→ **Stiamo attualmente valutando come allargare l'analisi alle stazioni a corto periodo**



# CONCLUSIONI

**Le reti neurali si stanno rivelando utili anche nelle discipline geofisiche ed in particolare quando si devono eseguire classificazioni di grandi moli di dati come ad esempio immagini o diagrammi**

**I diagrammi di rumore sismico SQLX molto utili per l'Analisi della Qualità ma la loro grande mole (in crescita) non può essere controllata da un singolo operatore.**

**Invece si prestano per istruire una rete neurale. →**

**→ DOPO che la Rete Neurale ha selezionato un numero (si spera esiguo) di stazioni con problemi o «sospette», l'operatore può approfondire l'analisi concentrandosi su poche stazioni.**

**I buoni risultati (95% dei successi nel test speditivo e 93% nel test finale) ci stimolano a sviluppare ulteriormente questa tecnica per affinarla ed estenderla a tutte le stazioni**