

What Is Machine Learning Teaching Us? Explainable AI for Seismic Models

Michele Magrini
Francesco Marrocco
SMIA 24/25 EQML Final Project



SAPIENZA
UNIVERSITÀ DI ROMA

First experiment:

CNN Explainability for the classification of Foreshocks and Aftershocks

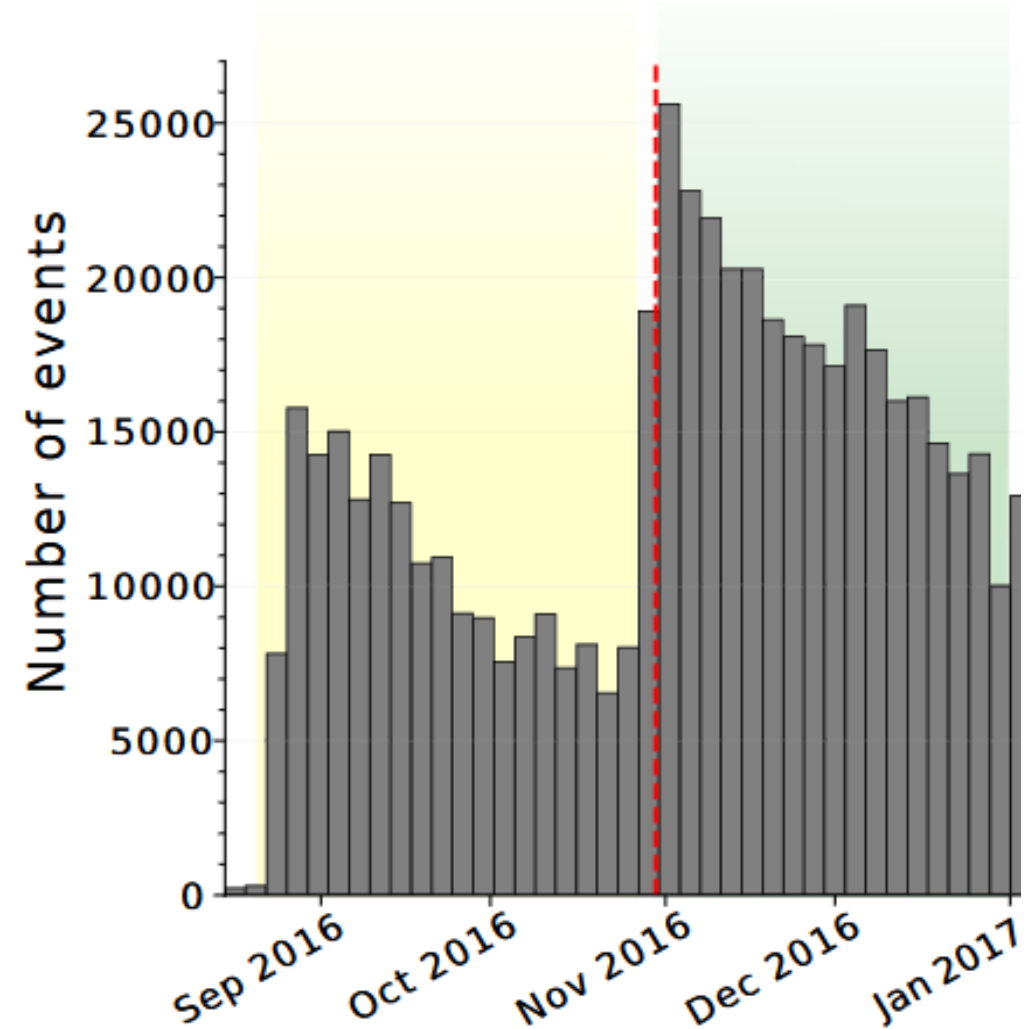


What are Aftershocks and Foreshocks?

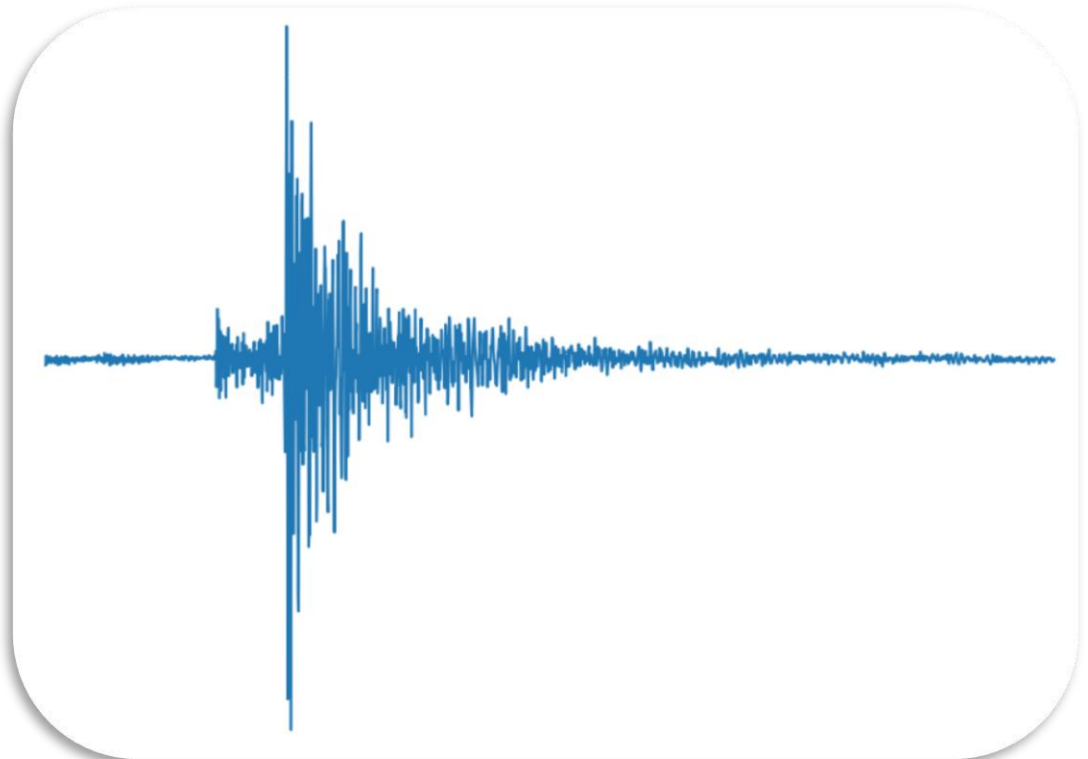
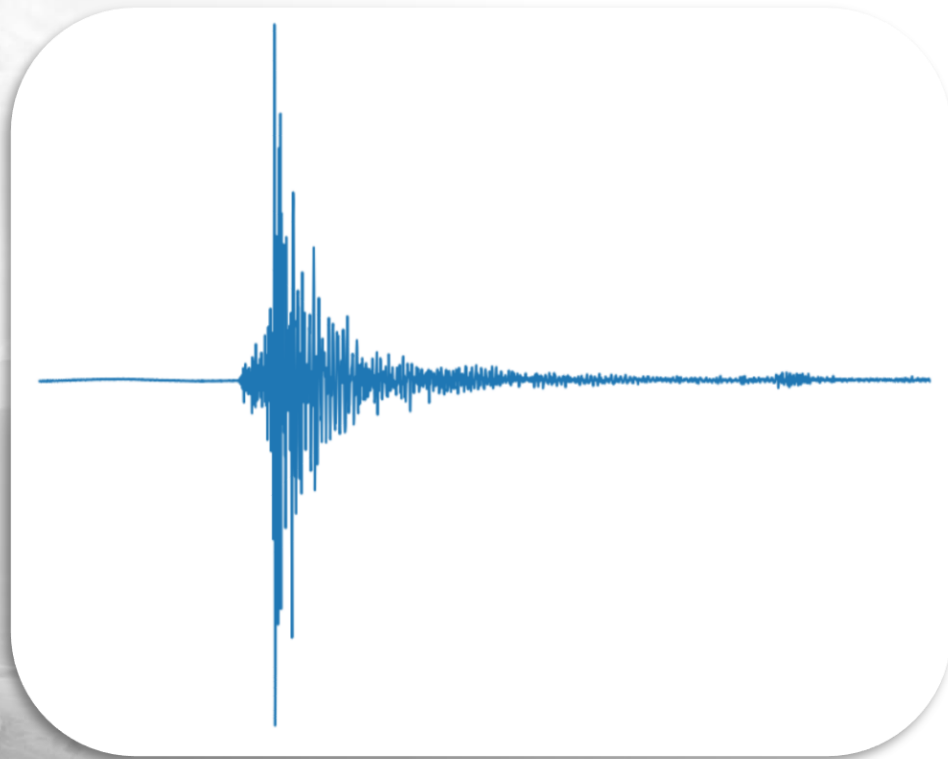
- **Foreshocks:** Smaller earthquakes occurring before a mainshock, often signaling stress build-up along a fault.
- **Aftershocks:** Smaller earthquakes following the mainshock, caused by the crust adjusting to new stress levels
- Both are critical stages in the seismic cycle, reflecting fault behaviour and stress redistribution

Why is classifying them important?

1. **Early Warning:** Identifying foreshocks can aid in predicting mainshocks, potentially saving lives.
2. **Risk Management:** Understanding aftershocks helps assess ongoing seismic risks
3. **Fault Dynamics:** Classifying these events provides insights into stress evolution during the seismic cycle



Shown here are two waveforms from the Norcia dataset, recorded during the M6.5 earthquake in Norcia, PG (Italy). One represents a foreshock, while the other is an aftershock of the same event.



Can you spot the differences?

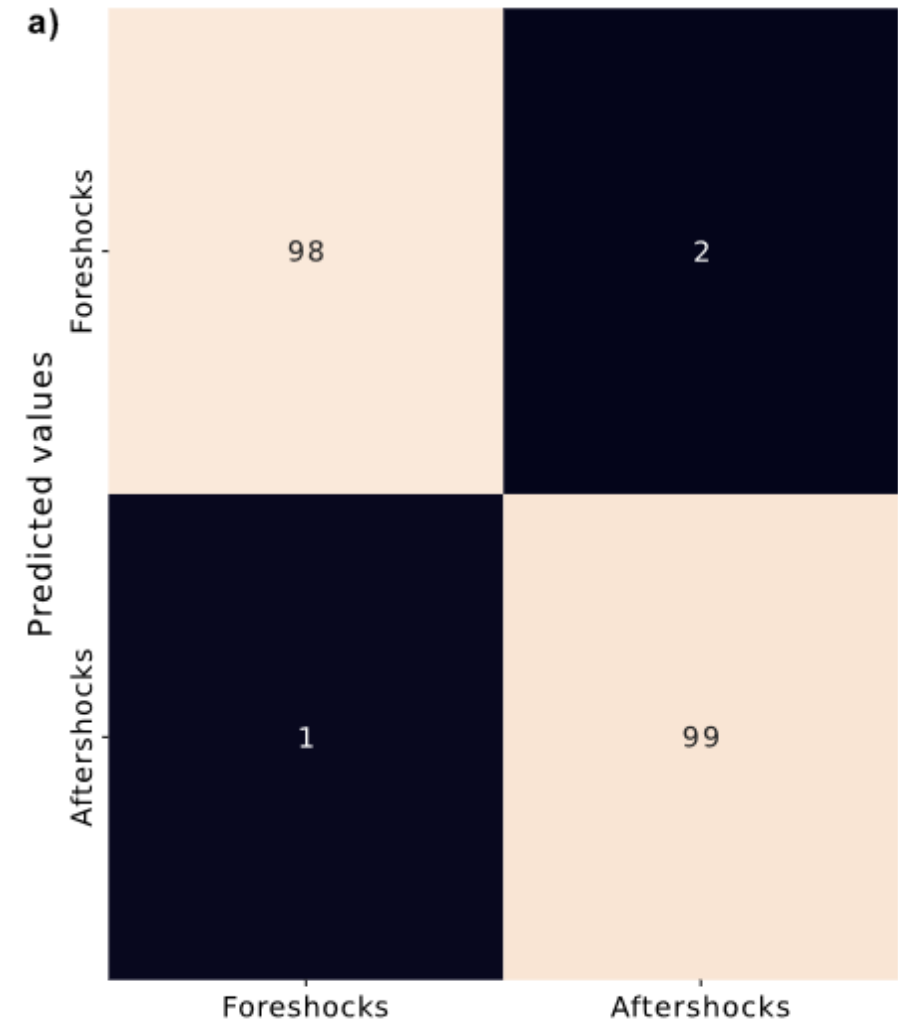
In their work, *Probing the evolution of fault properties during the seismic cycle with deep learning*, Laura Laurenti, Gabriele Paoletti, et al. demonstrated that DL models accurately distinguish seismic waves recorded before and after a mainshock.

But what are these DL models really learning?

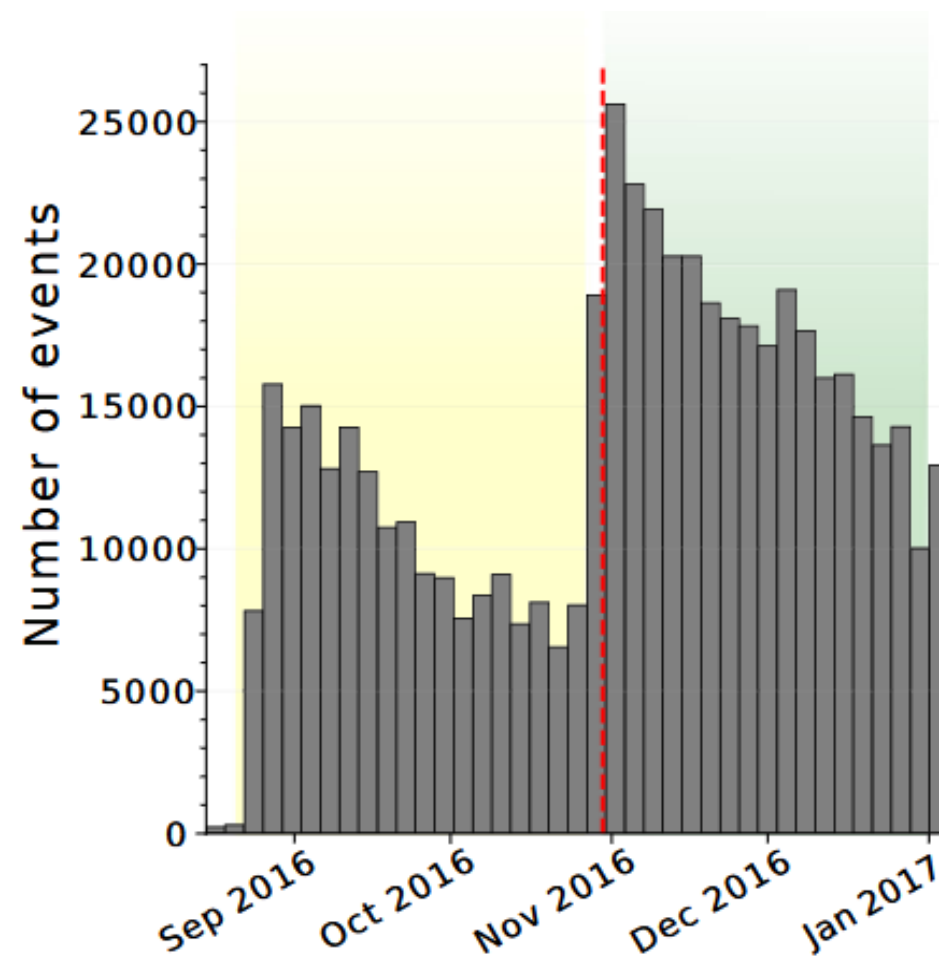
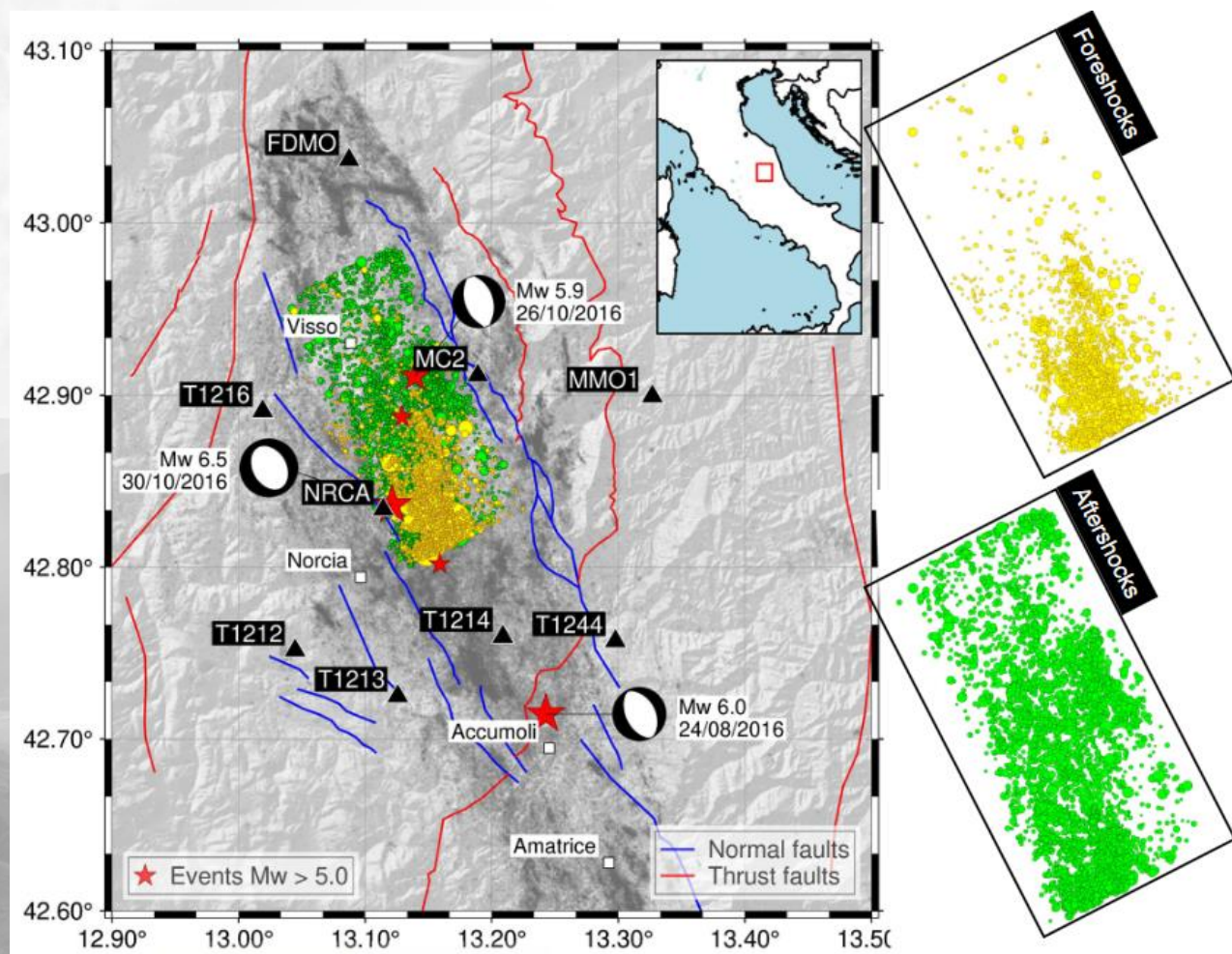
Can they teach us something useful?

In our project we aim to build a new DL model to replicate the task from the paper and then use SHAP for model explainability.

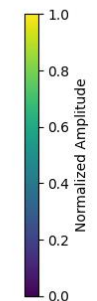
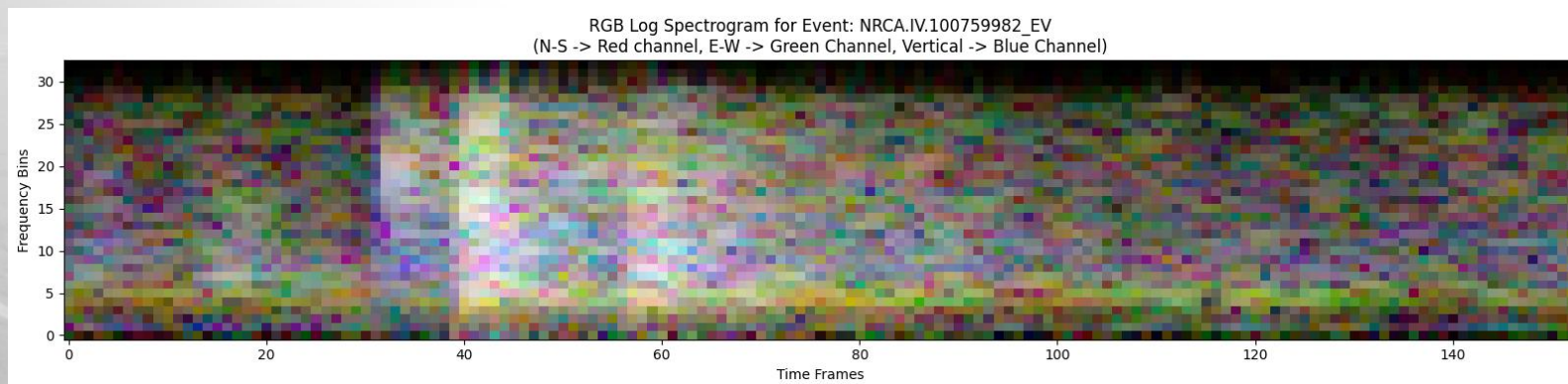
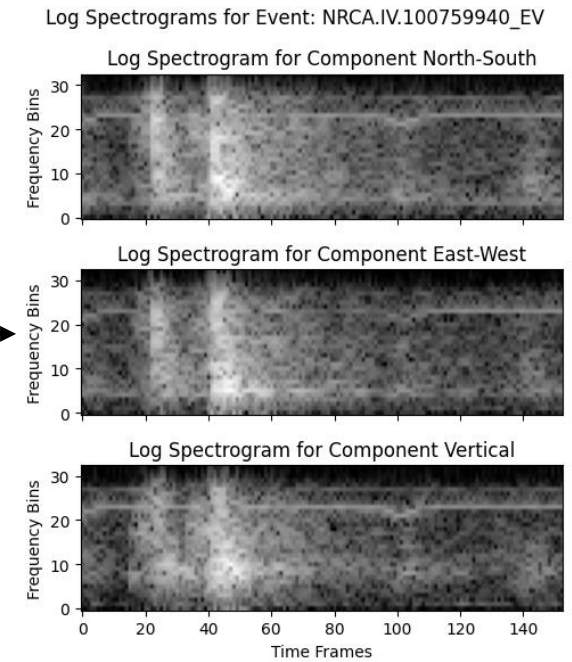
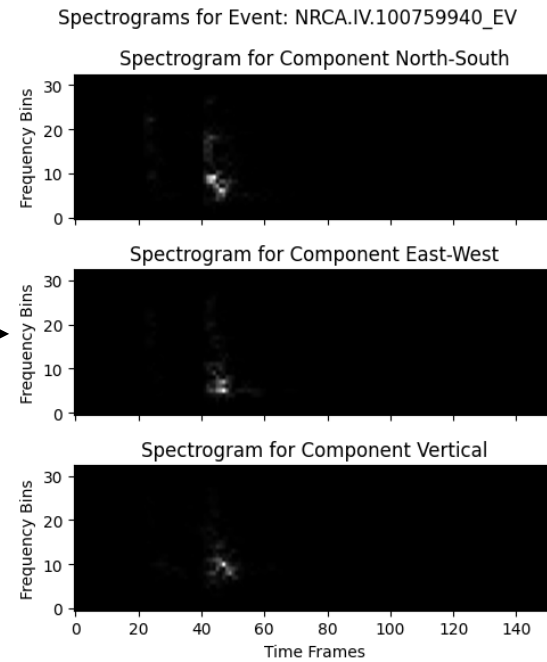
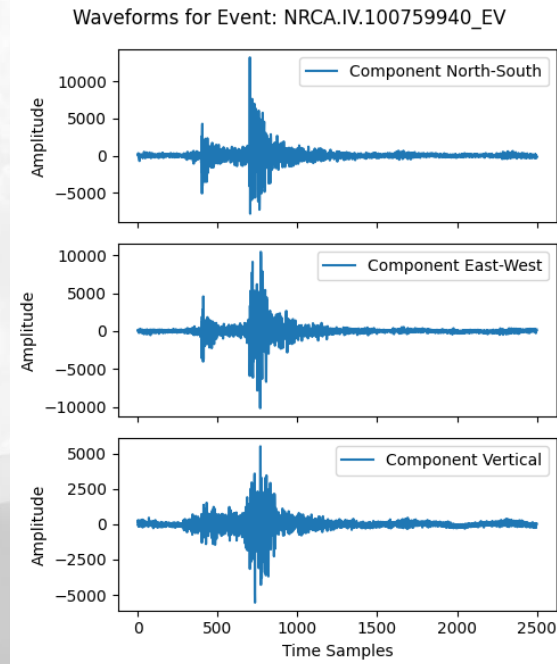
This allow us to better understand which features drive the model's predictions.



Dataset

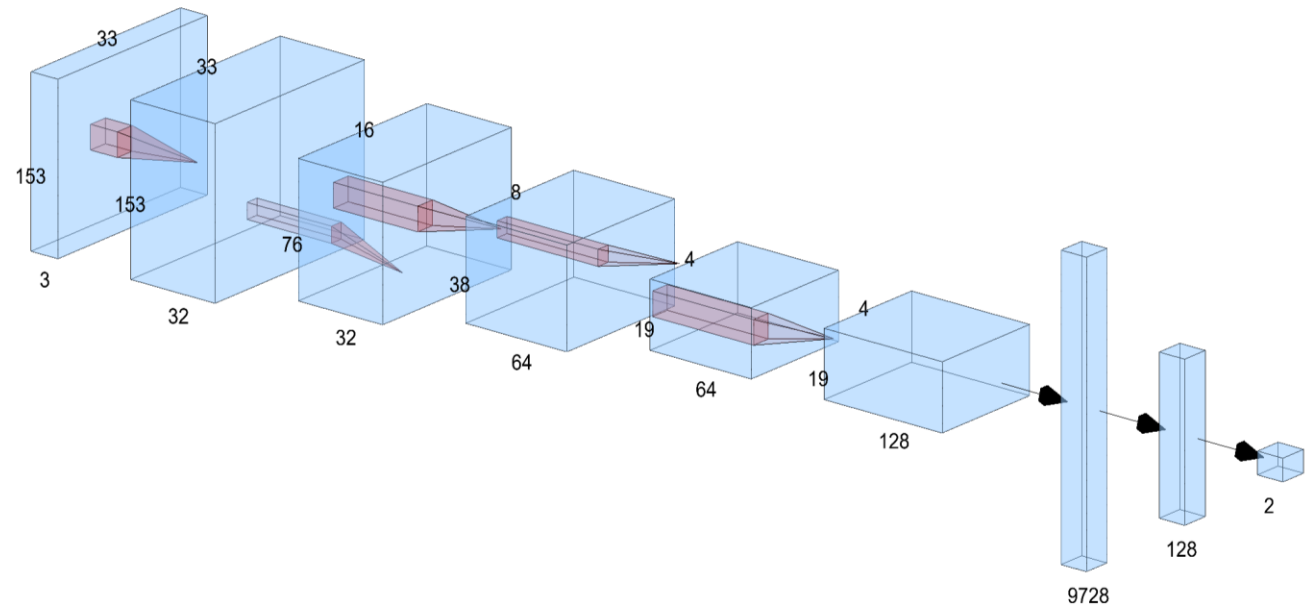


Data Preprocessing



Model Architecture

The model is a 7-layer convolutional neural network (CNN). It takes a 3-channel log spectrogram as input and processes it through three convolutional layers with batch normalization and max pooling. The output is a tuple with the raw probabilities for the event being a foreshock or aftershock, without SoftMax applied.

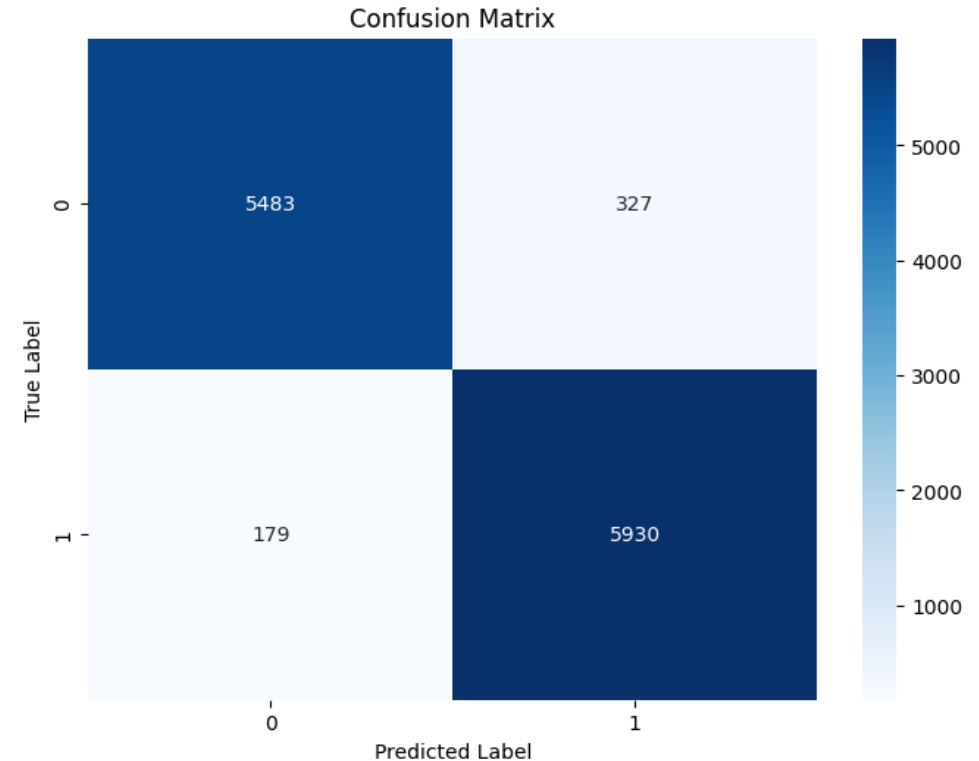


Model Evaluation

The model demonstrates excellent performance with 95.75% **accuracy**. It maintains a **high precision** (94.77%) and **recall** (97.07%), resulting in a balanced **F1 score** of 95.91%.

The confusion matrix shows minimal misclassifications, highlighting its reliability.

It **slightly underperforms** compared to the model presented in the paper, which achieved **>98% accuracy**.



Test Accuracy: 0.9575

Precision: 0.9477

Recall: 0.9707

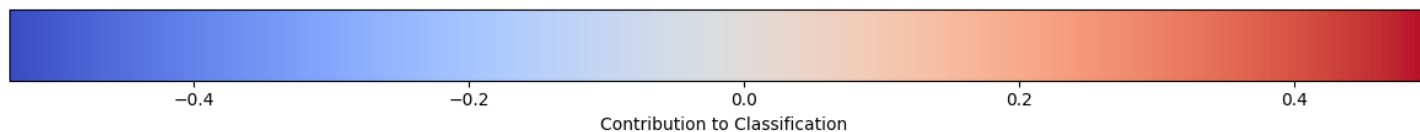
F1 Score: 0.9591

SHAP Example

SHAP Visualization (1-Channel Mean)
Label: Foreshock Predicted as Foreshock with $P=0.9867767691612244$
Filename: T1212.IV.100041701_EV_pre.png

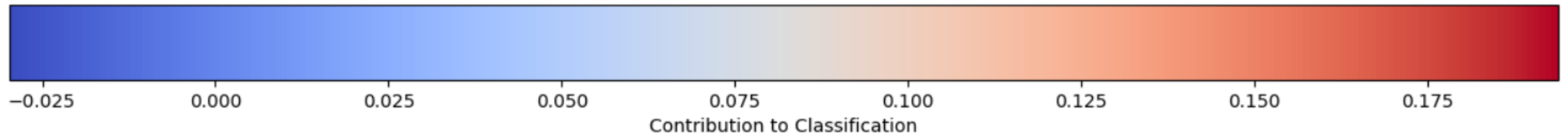
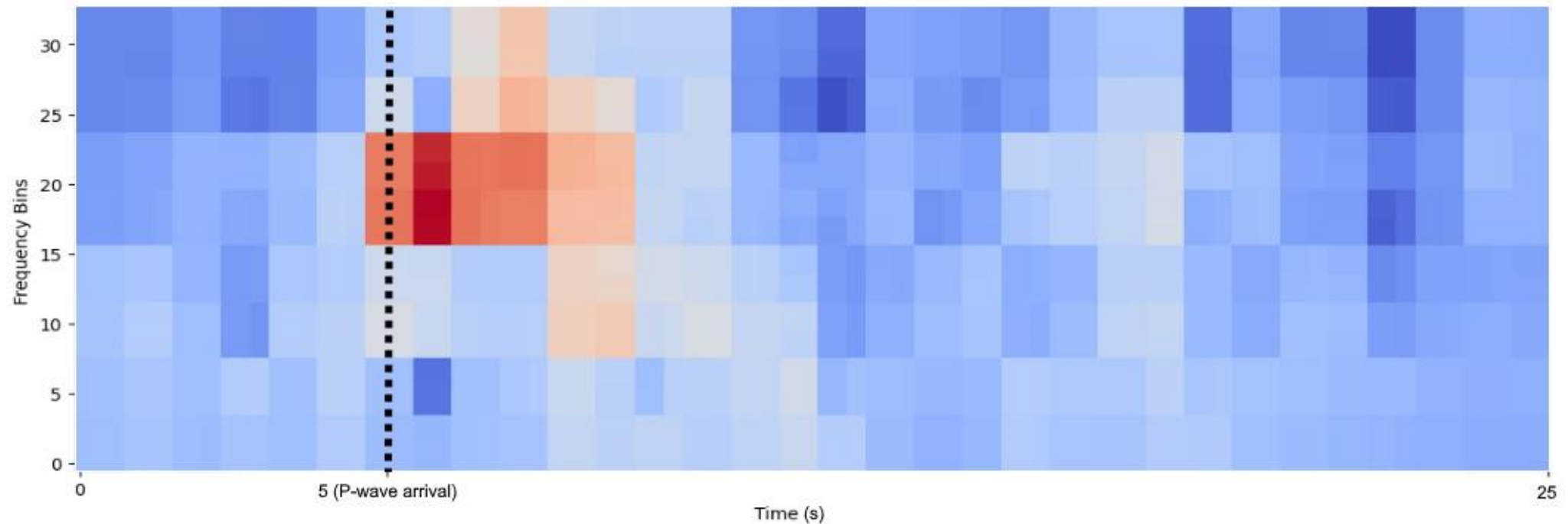


SHAP Visualization (1-Channel Mean)
Label: Aftershock Predicted as Aftershock with $P=0.9857869148254395$
Filename: T1212.IV.100108730_EV_post.png



Average SHAP on Foreshocks

SHAP Values Mean for Log Spectrograms Foreshocks on 316 samples from Validation (correct predictions)



Average SHAP on Aftershocks

SHAP Values Mean for Log Spectrogram Aftershocks on 301 samples from Validation (correct predictions)

