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

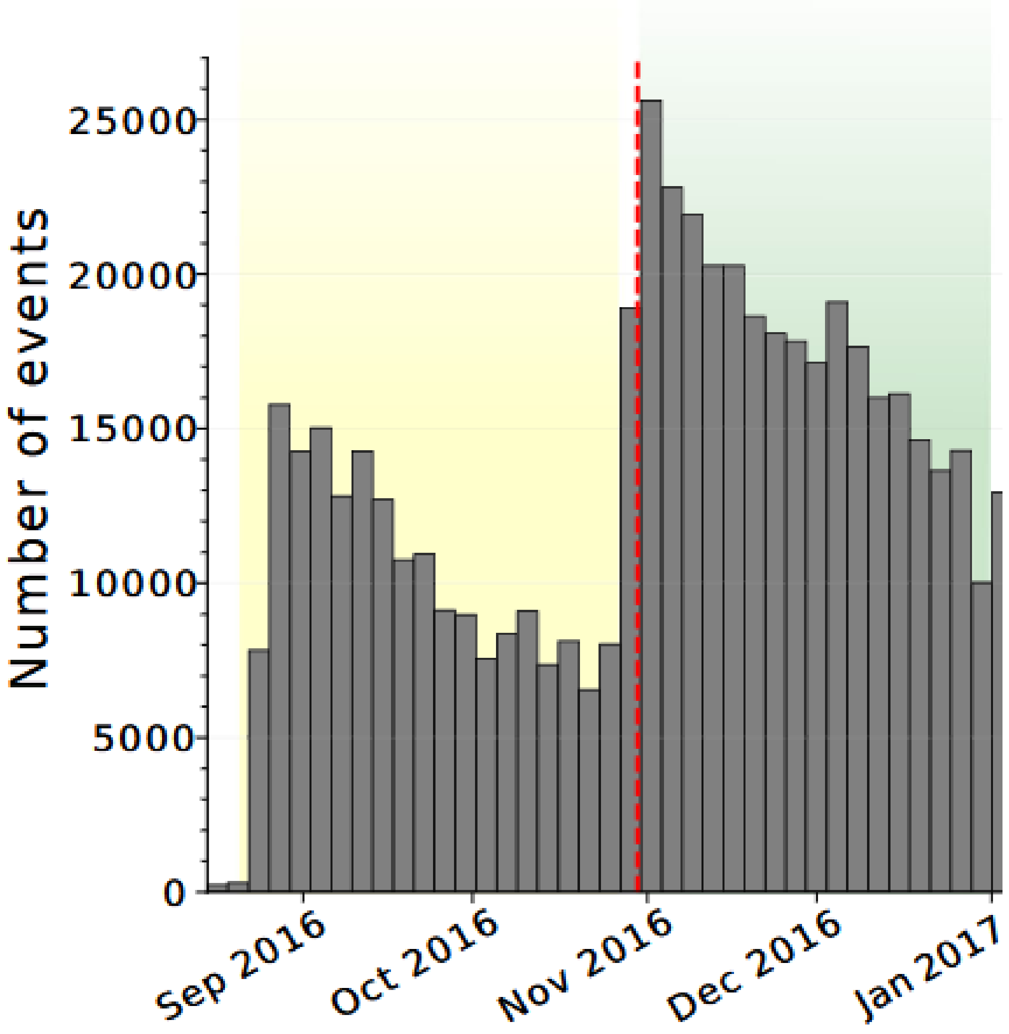
Michele Magrini (2066963)
Francesco Marrocco (2062223)
SMIA 24/25 EQML Final Project

First experiment:

CNN Explainability for the classification of Foreshocks and Aftershocks

What are Aftershocks and Foreshocks?

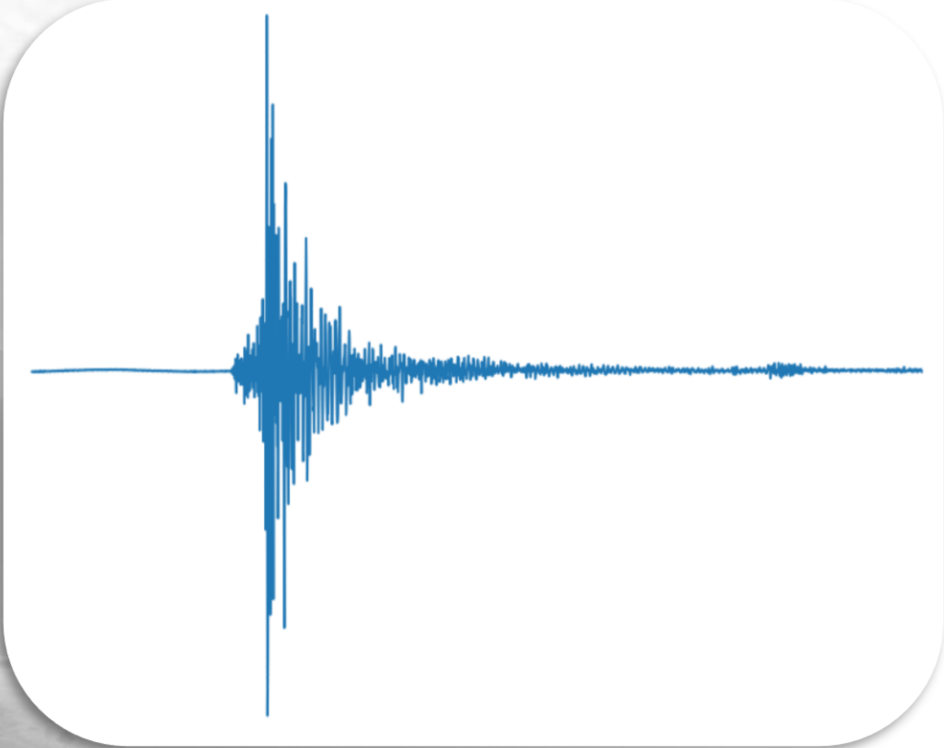
- **Foreshocks**: Smaller earthquakes occurring before a mainshock, often signaling stress build-up along a fault.
- **Aftershocks**: Smaller earthquakes following the mainshock, caused by the crust adjusting to new stress levels
- Both are critical stages in the seismic cycle, reflecting fault behaviour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

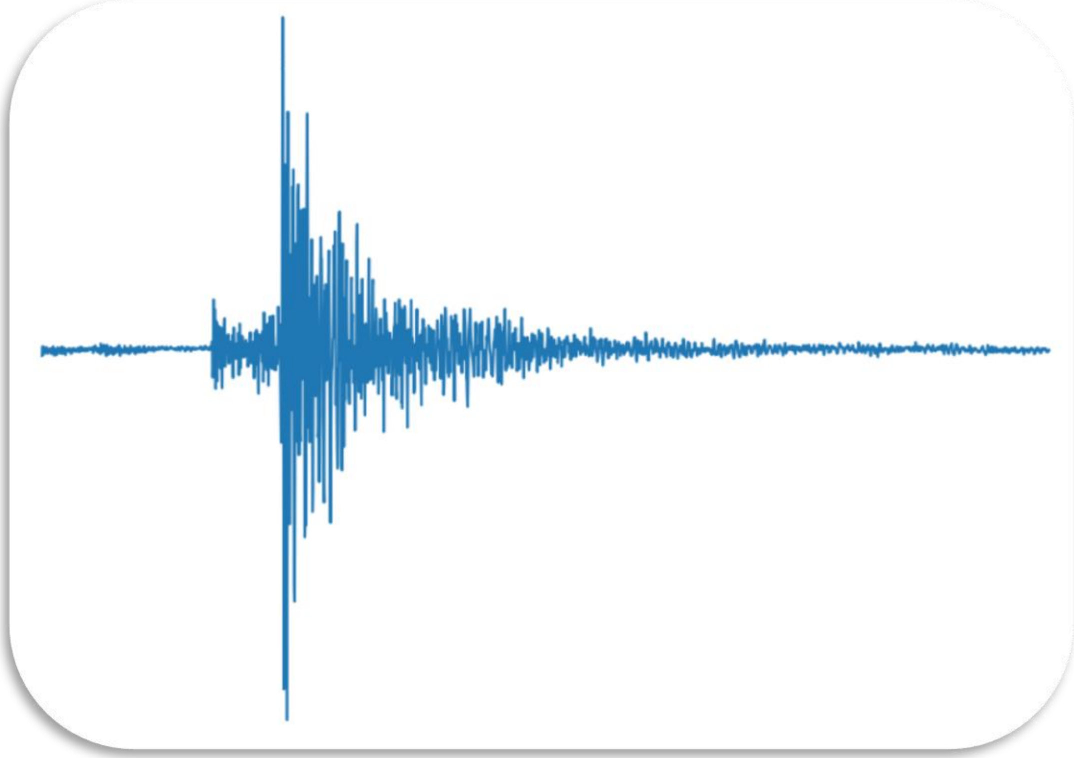


Why is classifying them important?

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.





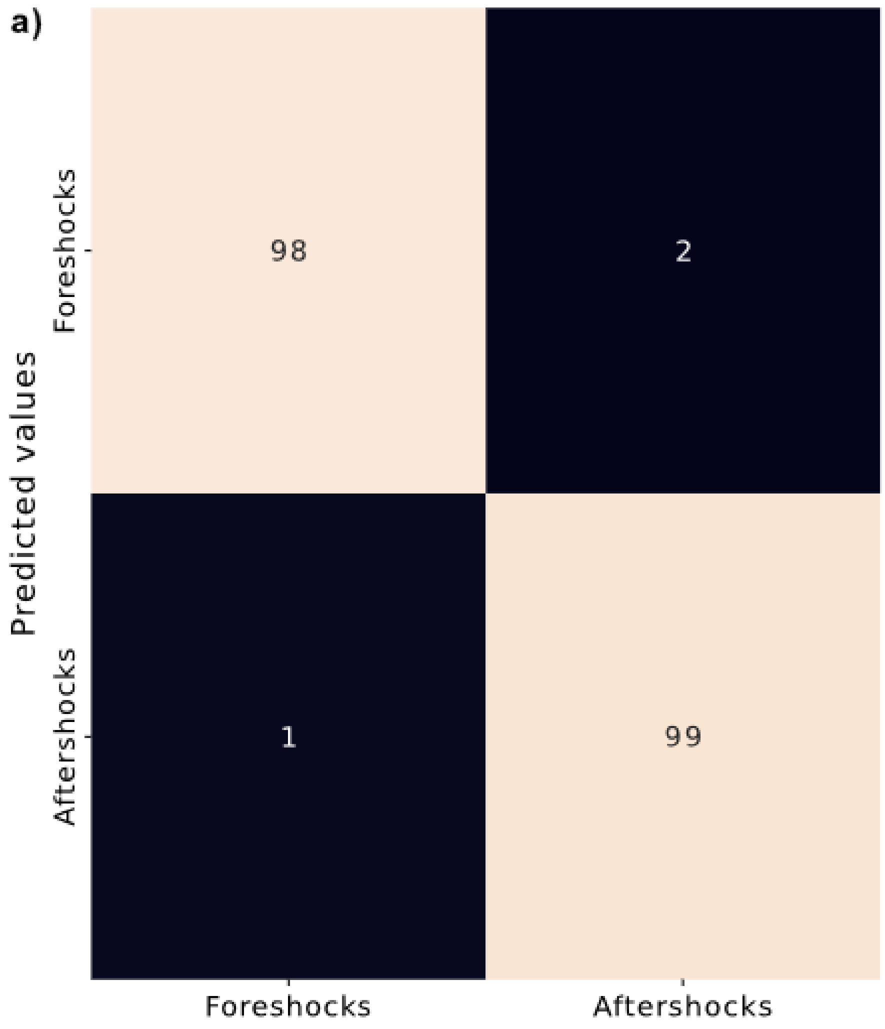
Can you spot the differences?

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.

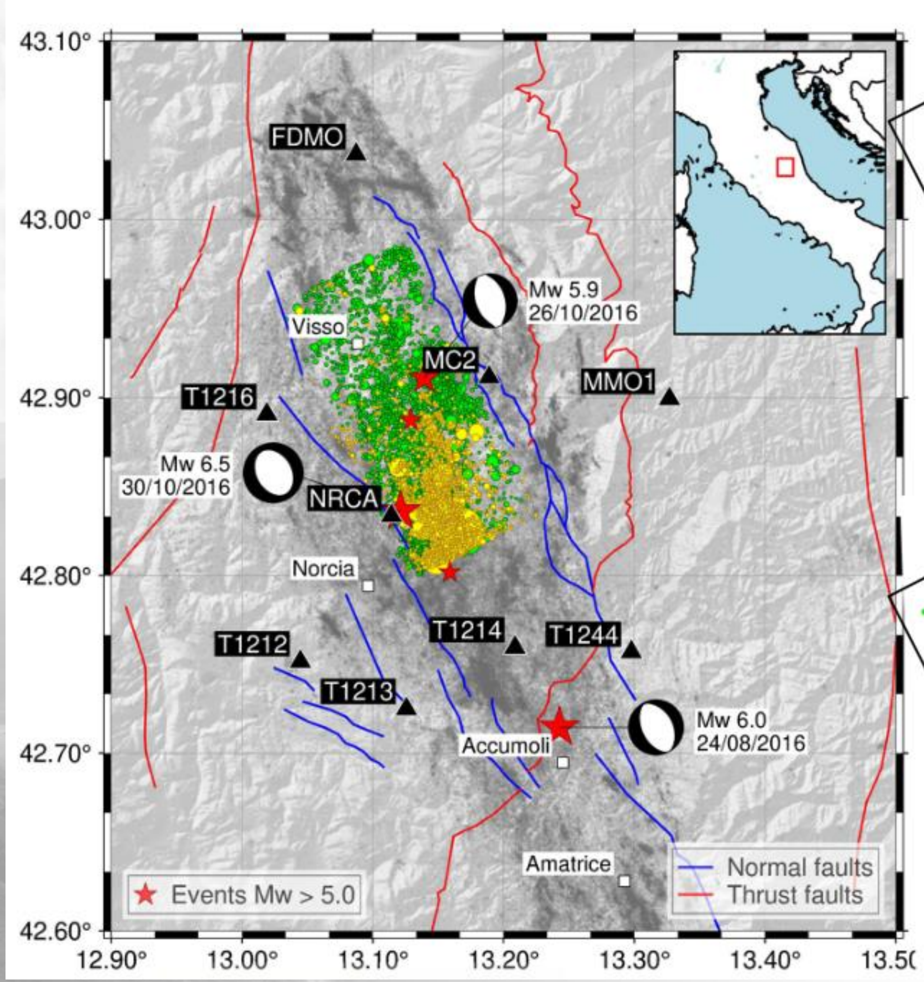
But what are these DL models really learning?

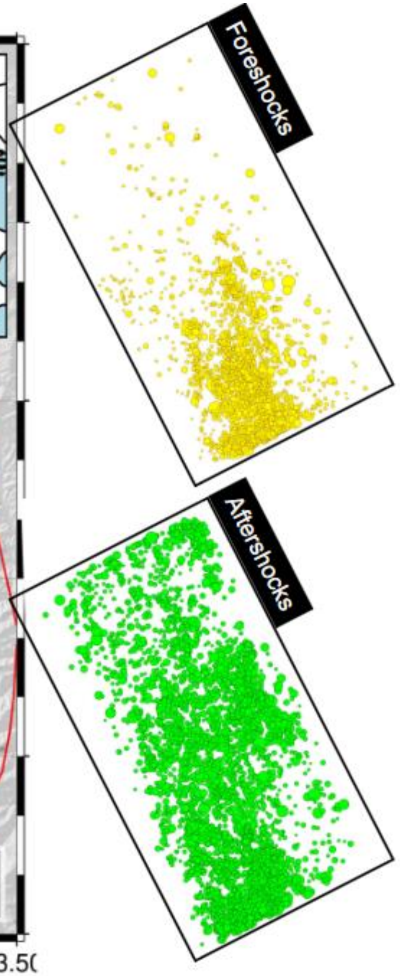
Can they teach us something useful?

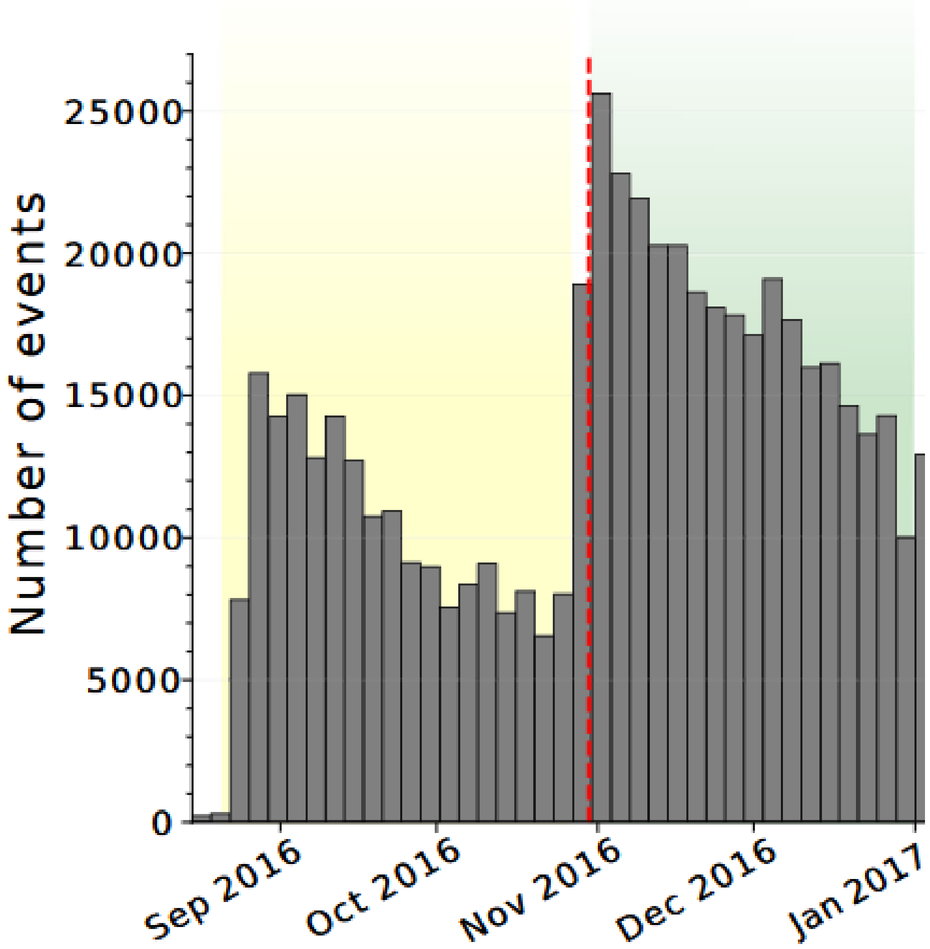
In our project we aim to build a new DL model to replicate the task from the paper and then use SHAP for model explainability.
This allow us to better understand which features drive the model’s predictions.



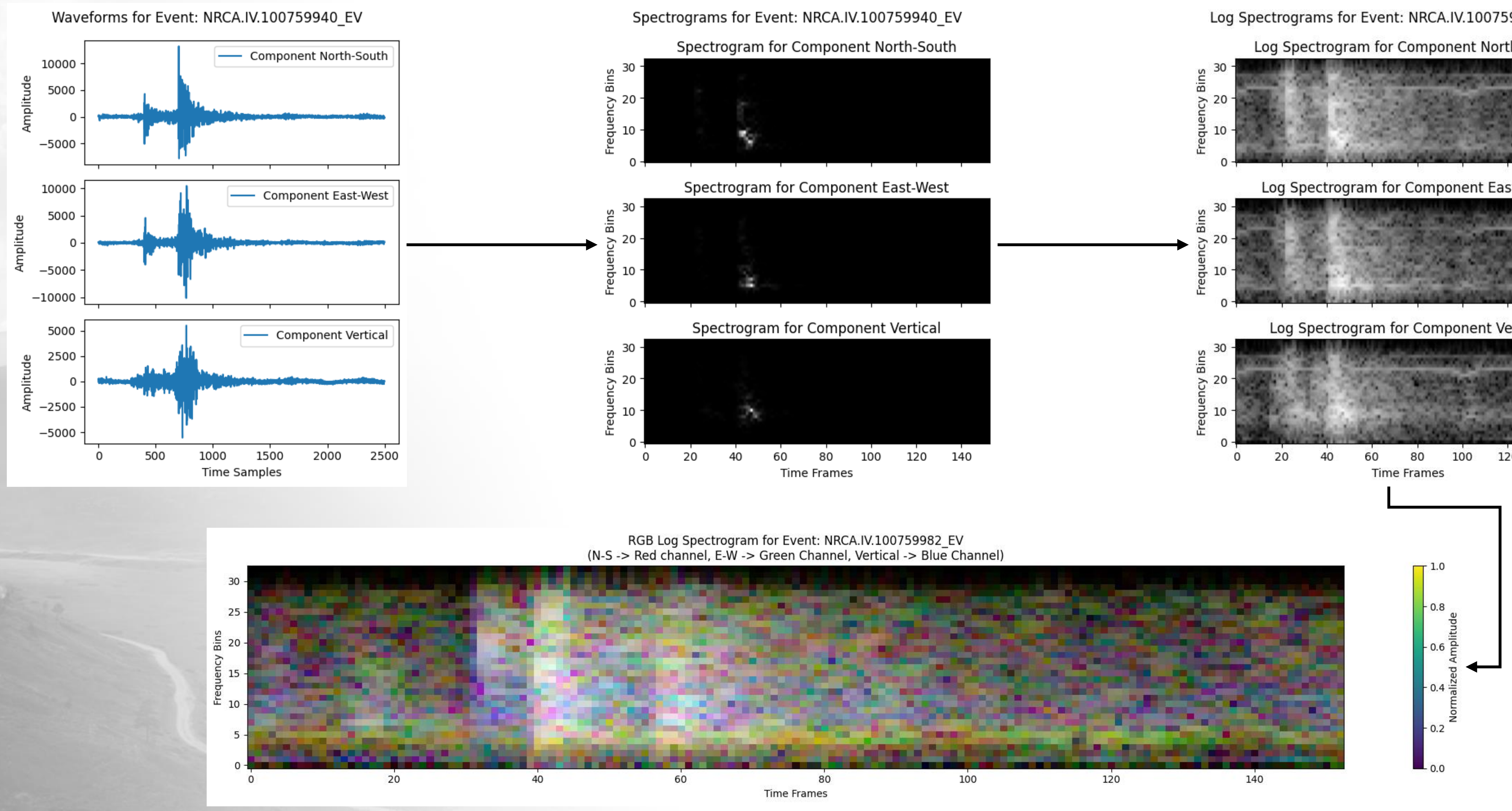
Dataset


43.10°
43.00°
42.90°
42.80°
42.70°
42.60°
12.90°
13.00°
13.10°
13.20°
13.30°
13.40°
13.50
FDMO
Visso
MC2
MMO1
T1216
Mw 5.9
26/10/2016
Mw 6.5
30/10/2016
NRCA
Norcia
T1212
T1213
T1214
T1244
Mw 6.0
24/08/2016
Accumoli
Amatrice
Events Mw > 5.0
Normal faults
Thrust faults


Foreshocks
Aftershocks

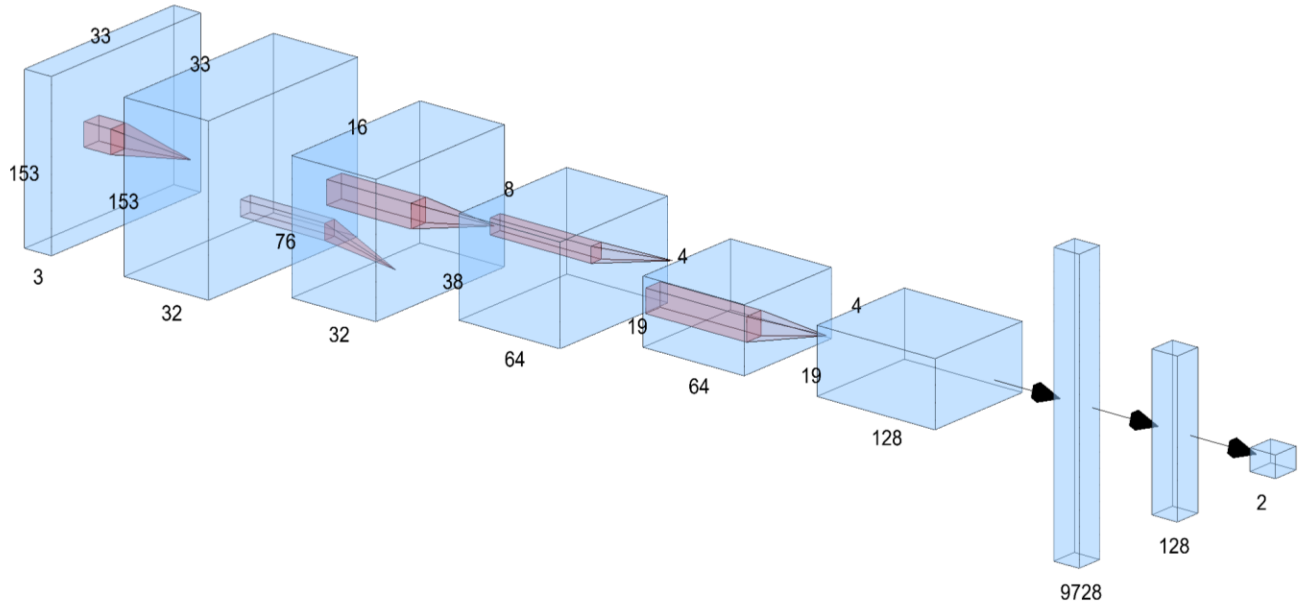

Number of events
25000
20000
15000
10000
5000
0
Sep 2016
Oct 2016
Nov 2016
Dec 2016
Jan 2017

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.

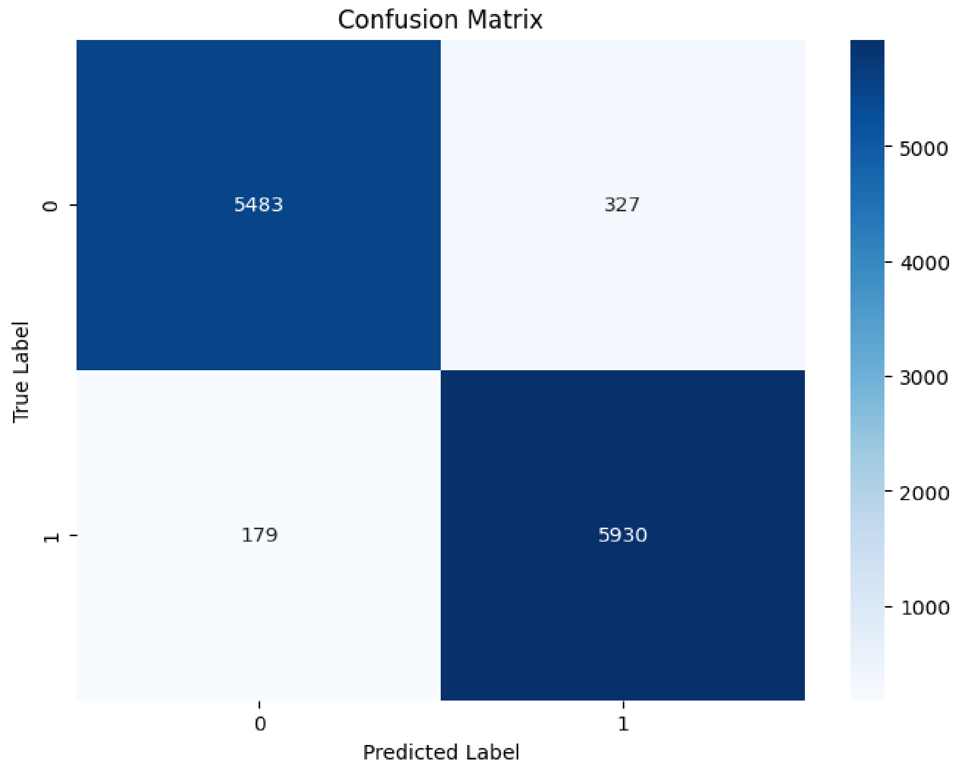


Model Evaluation

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%.

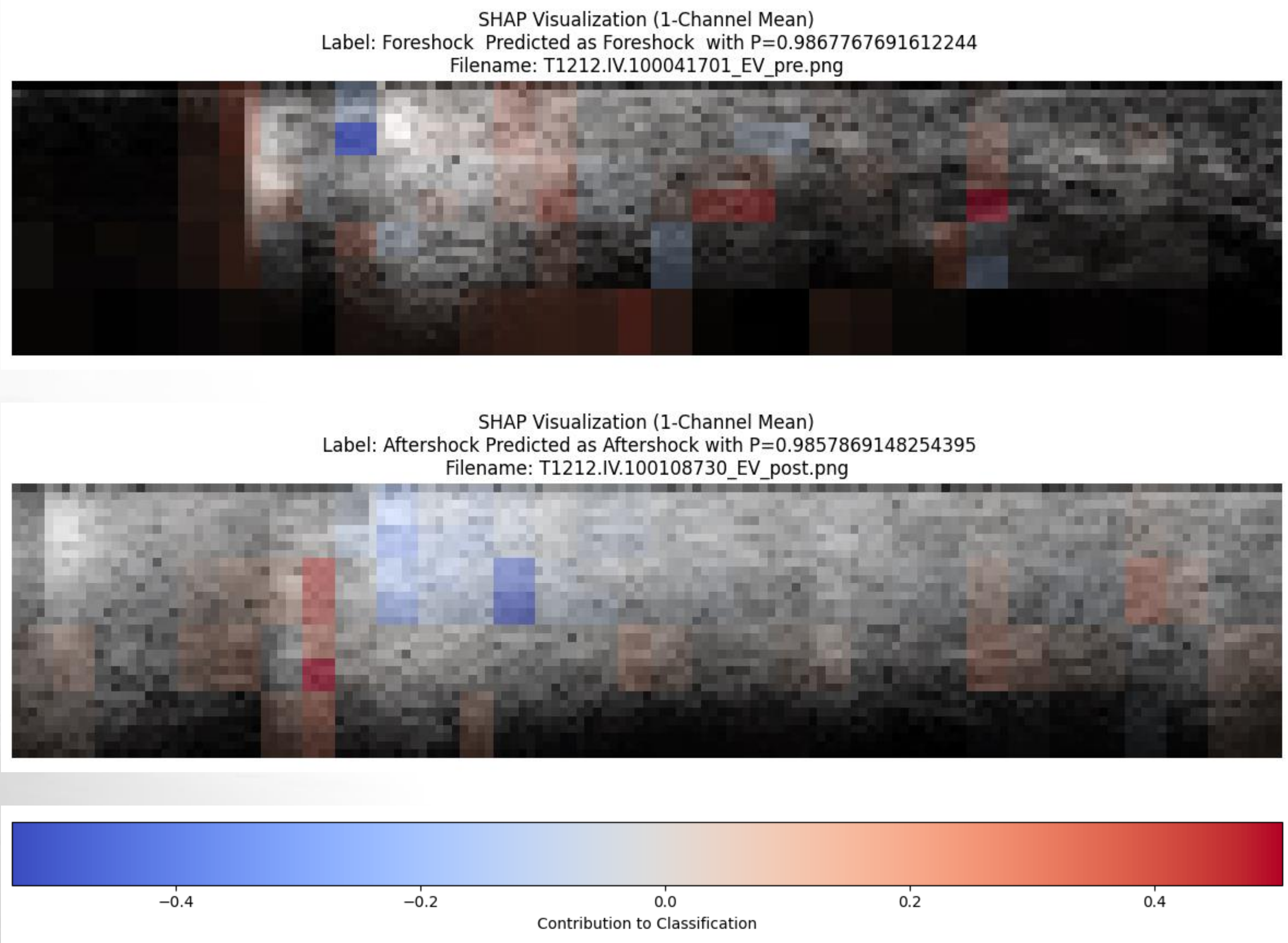
The confusion matrix shows minimal misclassifications, highlighting its reliability.

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

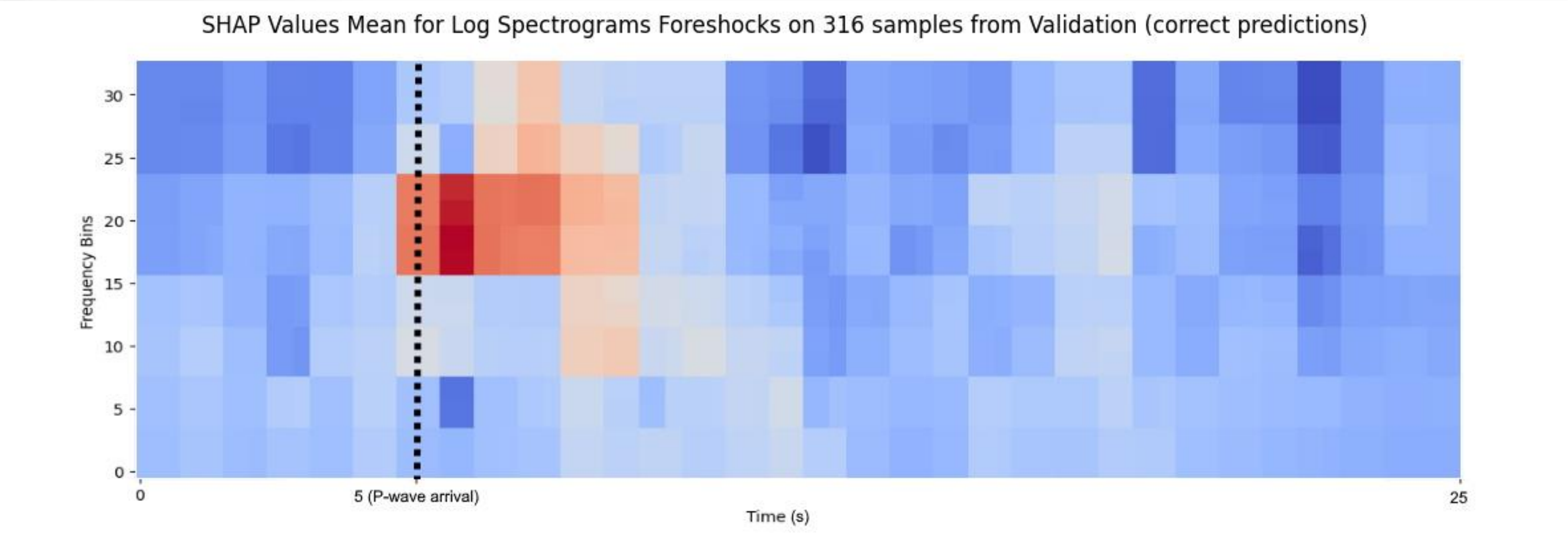


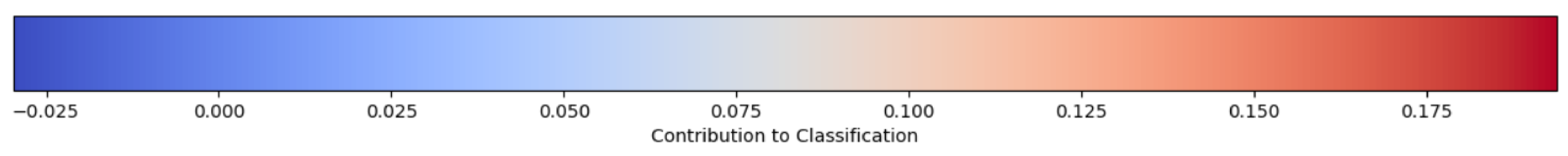

```
Test Accuracy: 0.9575
    Precision: 0.9477
       Recall: 0.9707
     F1 Score: 0.9591
```

SHAP Example



Average SHAP on Foreshocks





Average SHAP on Aftershocks