Earthquake AI forecast system.
Invest your money where your values are.

Multi-station cosmo-seismic detection

Presentation by: Noemi Zabari
Prepared by: Kamil Zielinski

GOALS

TOOL
Machine learning

FOR?
To predict whether an earthquake would occur

HOW?
Defined timeframe
Defined distance from cosmic ray stations
Based on the data from three stations.

DATA

WDCCR
World Data Center for Cosmic Rays

USGS
science for a changing world

Cosmic rays data:

- The corrected cosmic rays rate
- for the years 1974-2010
- three selected stations:
  Mt. Norikura, Alma B, and Tbilisi.

Earthquakes data:

- for the years 1950-2023
- via USGS API

DATA

WDCCR

World Data Center for Cosmic Rays

USGS

science for a changing world

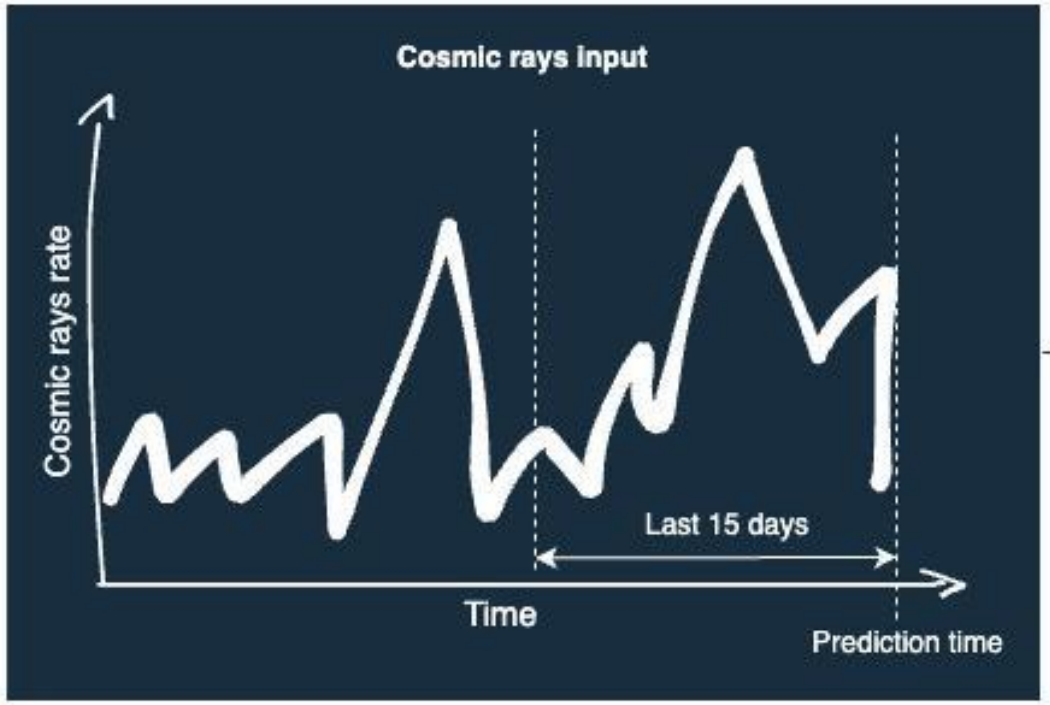
MODEL INPUT

Input consists of the past 360 hours (approximately 15 days) of cosmic ray data from the three stations, along with information about the time elapsed since the last earthquake.

MODEL OUTPUT

The probability of an earthquake with a magnitude of 6 or greater occurring within the next 14 days, at a distance of 1000 km from any of the three stations.

ARCHITECTURE OF THE SOLUTION

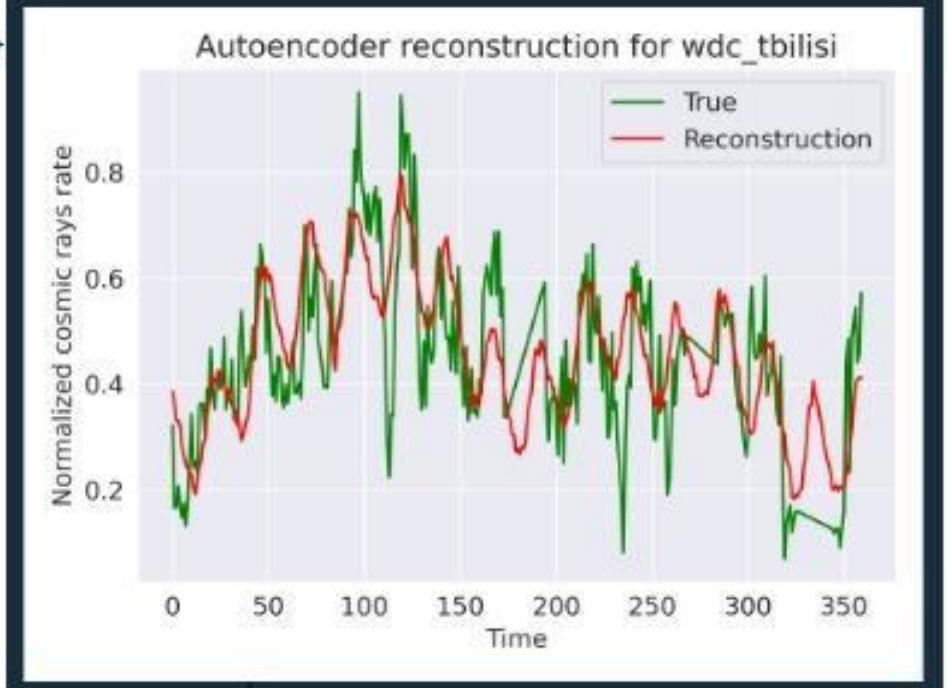


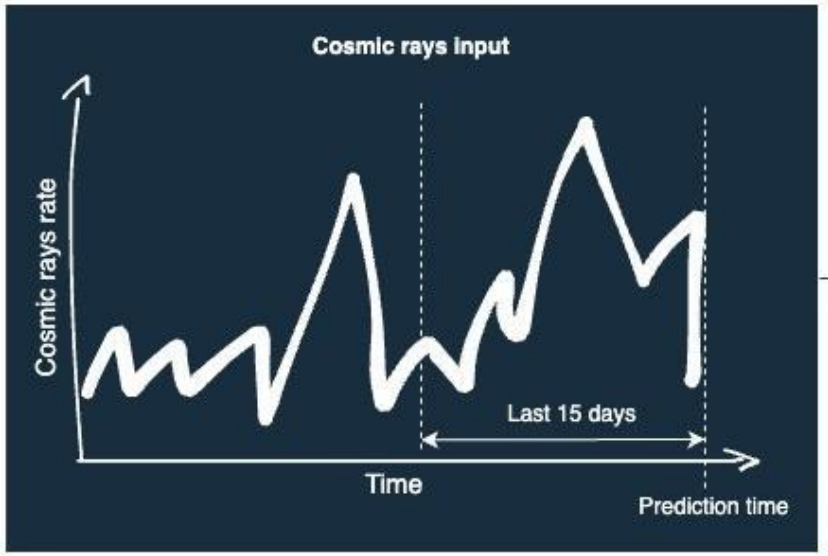
3 stations x 15 days x 24 hours x 1 cosmic rays rate measurement = 1080 numbers

Autoencoder

Attempts to condense the 1080 numbers into 20 features and then accurately reconstruct the original cosmic ray rate sequences. By ignoring unimportant information, it enhances the model's performance.

Reconstruction example:



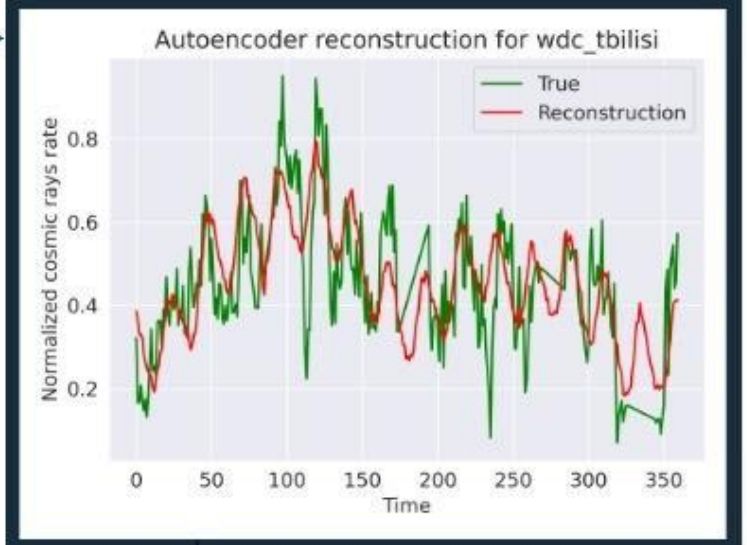


3 stations x 15 days x 24 hours x 1 cosmic rays rate measurement = 1080 numbers

Autoencoder

Attempts to condense the 1080 numbers into 20 features and then accurately reconstruct the original cosmic ray rate sequences. By ignoring unimportant information, it enhances the model's performance.

Reconstruction example:



"Days since last earthquake" feature

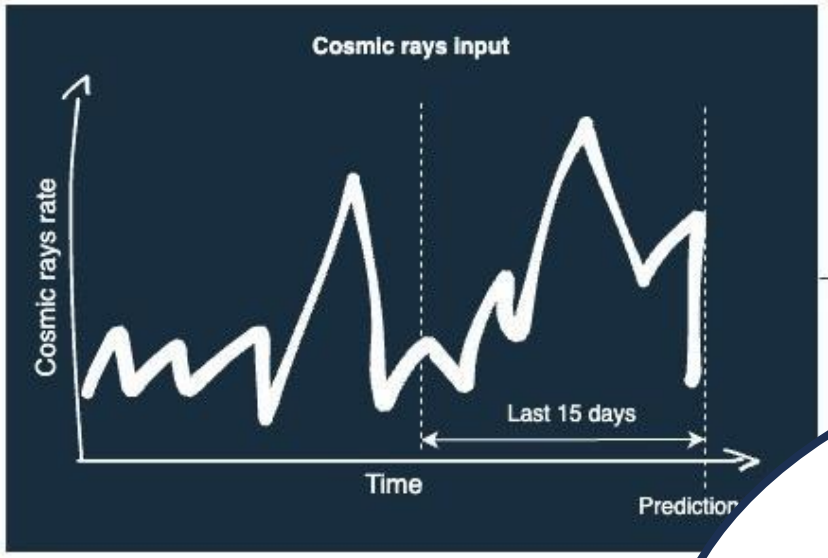
20 numbers

1 number

Classifier

1 number

Estimated probability

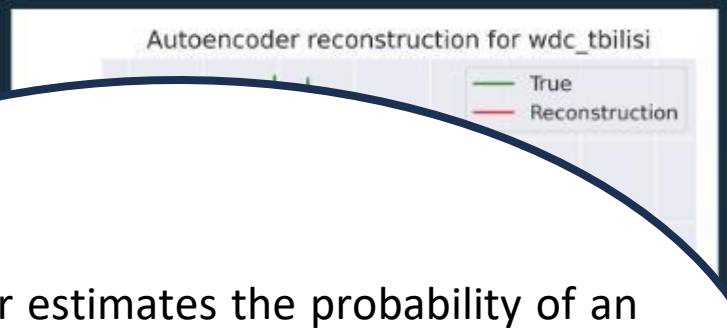


3 stations x 15 days x 24 hours x 1 cosmic rays rate measurement = 1080 numbers

Autoencoder

Attempts to condense the 1080 numbers into 20 features and then accurately reconstruct the original cosmic ray rate sequences. By ignoring unimportant information, it enhances the model's performance.

Reconstruction example:



"Days since last earthquake" feature

1 number

The classifier estimates the probability of an earthquake with a magnitude 6+ occuring within the next 14 days, within 1000 km distance of any station. This estimation is based on compressed cosmic rays data and the time elapsed since the last earthquake.

Classifier

1 number

Estimated probability

DATA SPLIT

Time-aware data split into training, validation and test sets with proportions:

50%/25%/25%

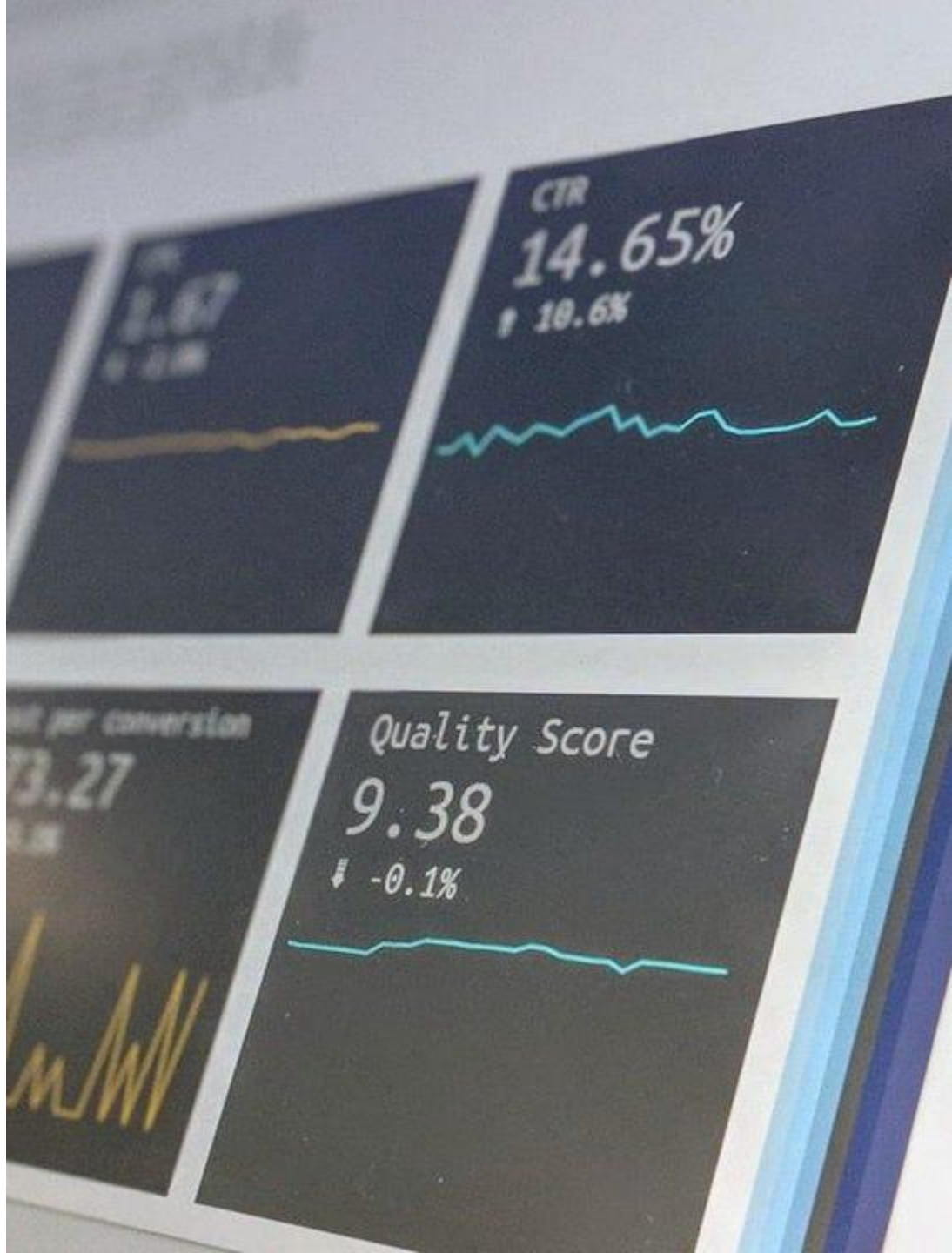
METRIC

Metric used for evaluation and statistical significance assessment was:

Average Precision (AP)

MODEL PERFORMANCE

Raw AP performance:
Model's test AP = **0.320**
Baseline (average random predictor) AP = **0.288**



Earthquake-focused performance

Average Precision (AP) is a commonly used Machine Learning (ML) metric for assessing classifier performance. It's useful for optimization and statistical significance testing but doesn't provide much insight into actual model performance in our context. AP evaluates every individual prediction (made on an hourly basis) without addressing questions such as:

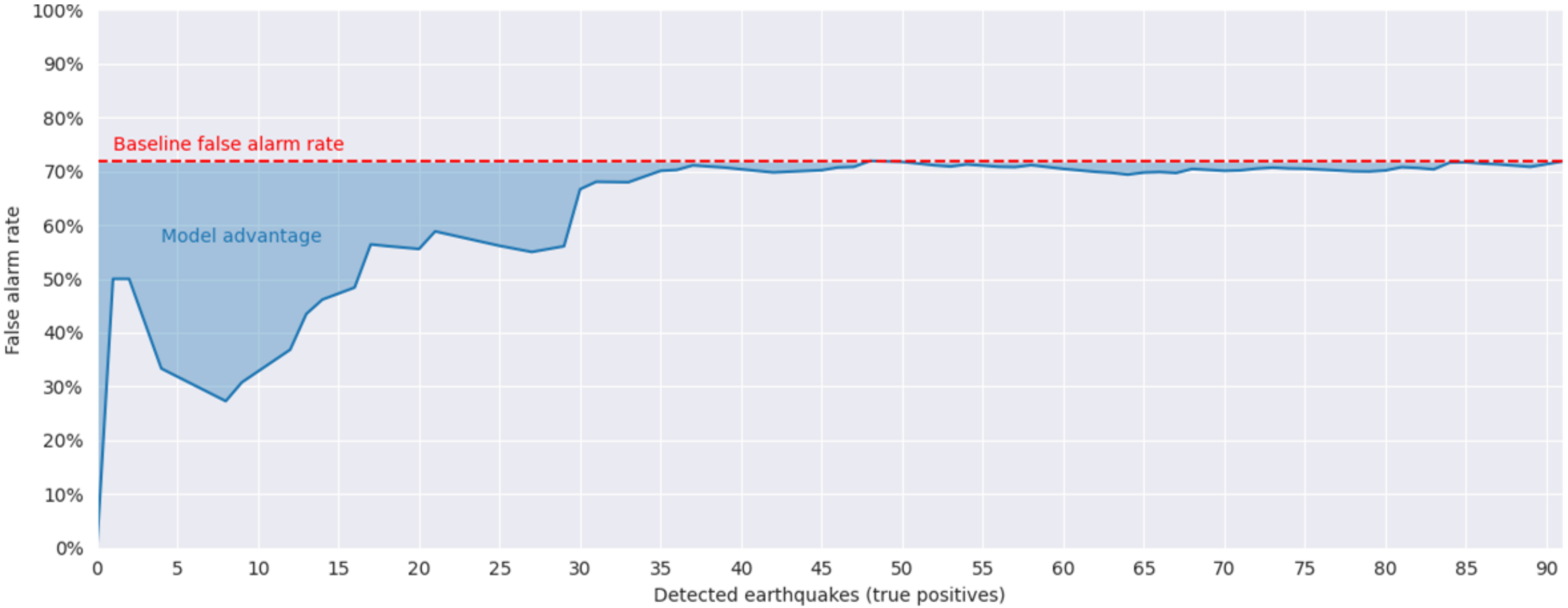
- How many earthquakes can we detect in advance?
- What is the false alarm ratio at a specific "detected earthquakes" level?

Earthquake-focused performance

To answer these questions, we've developed an earthquake-focused metric with these assumptions:

- We count false positives (FP) and true positives (TP) over a 14-day period (forecast horizon).
- A false positive (false alarm) is recorded if any false positive alert occurred within the 14-day period.
- A true positive (detected earthquake) is recorded if any alarm occurred within the 14 days preceding an earthquake.

Earthquakes detection trade-off curve



The baseline false alarm rate is the average false alarm rate expected from a random model.

Statistical significance test

- 300,000 experiments conducted
- each with a different seed.

- predictions were uniformly distributed random numbers between 0 and 1.
- None of these experiments could exceed the AP performance of the model,
- indicating a strong significance of 6 sigma or more.

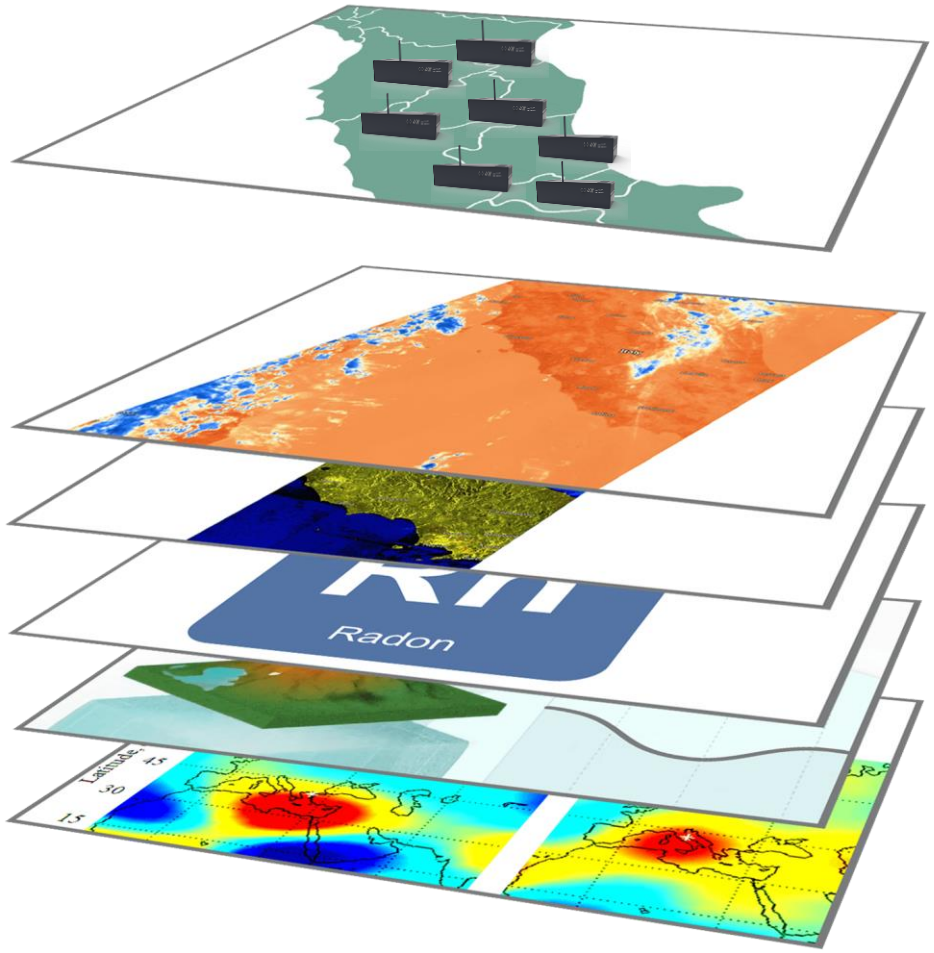
Conclusions

The Machine Learning (ML) approach appears **to confirm** the existence of cosmo-seismic correlation.

While the raw metrics might not seem impressive (AP 0.320 vs 0.288), they could translate into significant real-world effects. It seems we can forecast at least a few percent of massive earthquakes with a decent false alarm rate (below 30%).

- **It's important to remember that earthquakes are currently almost entirely unpredictable.**
- This result could be the **starting point** for more significant findings with the potential to save thousands of lives.

Multi-channel approach.



Multichannel approach = multiple observations for the same phenomena.

MVP

CHANNEL 1: cosmic ray rates

CHANNEL 2: land surface temperature

CHANNEL 3: ground displacement

CHANNEL 4: radon gas concentration

CHANNEL 5: groundwater level changes

CHANNEL 6: ionospheric disturbances

+ other channels

Purpose-driven DeepTech

esa

T

IAI
Israel Aerospace Industries

POLSA

hub:raum

STARBURST

Narodowe Centrum Badań Jądrowych
National Centre for Nuclear Research
ŚWIERK

CREDO
THE QUEST FOR THE UNEXPECTED

THE HENRYK NIEWODNICZAŃSKI
INSTITUTE OF NUCLEAR PHYSICS
POLISH ACADEMY OF SCIENCES

Institute of Geophysics
Polish Academy of Sciences

Politechnika Krakowska
im. Tadeusza Kościuszki

A. ALIKHANYAN
National Laboratory

Find out more: www.astroteq.ai
Email our founders: noemi@astroteq.ai
itamar@astroteq.ai
kevin@astroteq.ai