Automated track infrastructure recognition using vibration analysis

Natalie Platilová

Problem:

The main goal of this project is to develop a robust algorithm capable of detecting track infrastructure events, specifically turnouts, joints, and bridges, directly from the rail vibration signals. This capability would facilitate more effective and automated monitoring of track conditions in real time.

This consists of:

- mapping: visualizing GPS tracks with infrastructure overlay
- labelling: associating vibration segments with nearby infrastructure
- classification: using Machine Learning algorithms to identify infrastructure types form vibration patterns

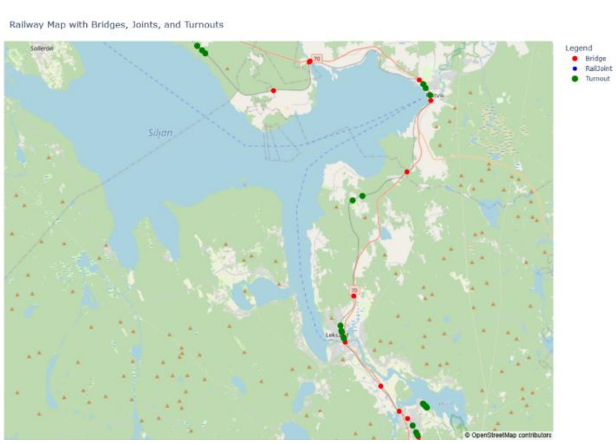
Method:

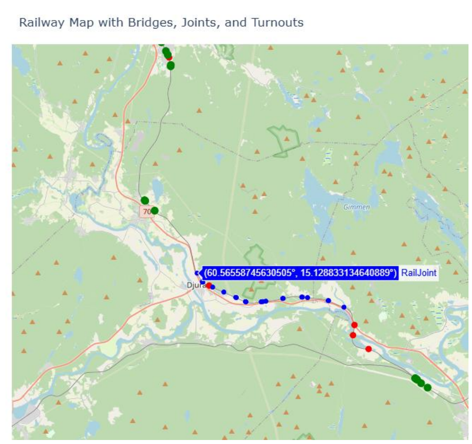
Mapping

Starting with Code 1, I loaded the three CSV files:

- converted_coordinates_Resultat_Bridge.csv
- converted_coordinates_Resultat_RailJoint.csv
- converted_coordinates_Resultat_Turnout.csv

This allowed me to render an interactive map showing the spatial distribution of Bridges (red), Rail Joints (blue), and Turnouts (green) across the railway network around the Siljan region.





This is a screenshot from Code 1. This figure illustrates various data points, including the turnout, the joint and the bridge. The geographic coverage of the visualization spans: Latitude range 60.51° to 61.01°, Longitude range 14.52° to 15.35°.

As we can see, Bridges are primarily concentrated in areas where the railway crosses water bodies and terrain features around Lake Siljan. Rail Joints are distributed more uniformly along the track, indicating regular maintenance points and rail segment connections. Turnouts appear at strategic locations, likely corresponding to stations, sidings, and junction points.

Data Selection and Preprocessing

To analyze vibration patterns aligned with the mapped track, I selected vibration data from Data 2, focusing on a subset that matched the GPS coordinates from Code 1.

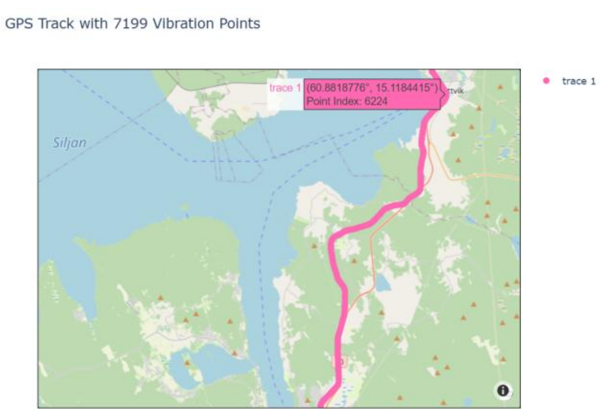
Because Data 2.zip was around 130 GB, I explored potential subsets using folder-level metadata and GPS overlap. After comparing coverage across several timestamped folders, I selected Data 2 / 2024-12-10 14-00-00 (1), which had the most comprehensive overlap with the Siljan region track.

Another option would be to combine several of the folders and merge datasets but that does become more computationally heavy.

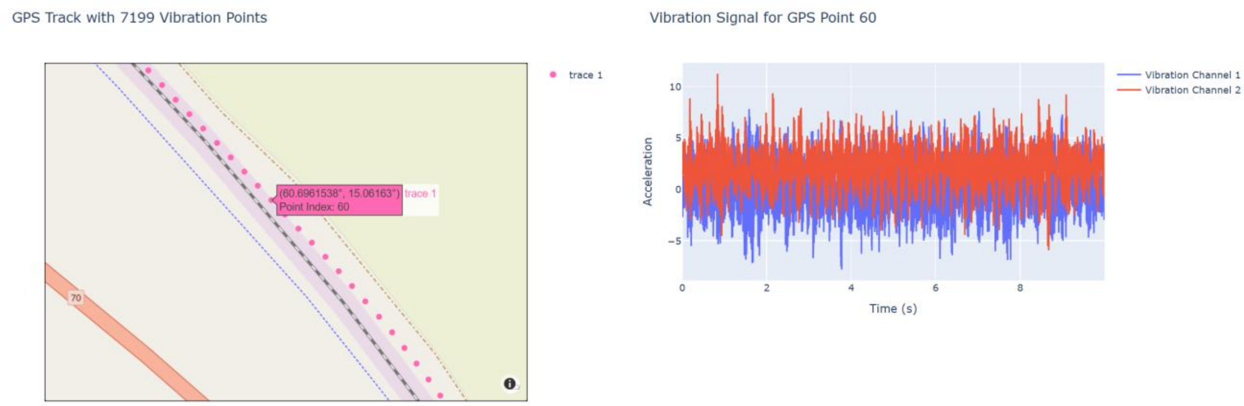
The dataset contained missing values and corrupted entries that caused parsing issues. I removed rows with missing data, validated coordinate ranges to ensure alignment with the mapped area, confirmed the presence of meaningful vibration signals for each GPS point.

Vibration Visualization

With the cleaned data, I was able to display vibration signal plots tied to individual GPS coordinates. This enabled inspection of how specific segments behaved dynamically.



Here we can see a stretch going around Siljan. The coordinate you see on the left in pink is currently showing the specific vibration data for that single point.



For example, selecting a single point along the route allowed inspection of its vibration waveform, which is useful for identifying features of nearby infrastructure like a turnout or joint.

The next phase involves labelling each vibration segment with its corresponding infrastructure type (bridge, joint, turnout, or none). This labelled dataset will serve as training data for supervised machine learning models aimed at classification.

Observations & Reflections:

Data management was a bit of a challenge. With so much data and some missing metadata, finding the right subset took a lot of trial and error, some geospatial guesswork, and quite a bit of manual digging.

GPS alignment wasn't always perfect. Points didn't line up with the actual tracks on the map. This was likely due to time mismatches or GPS noise, which is why the filtering and validation was needed.

Working with the data was also heavy on my machine. Because of the dataset's size, handling different folders took lots of time.