Running performance prediction Using Deep Learning

Deep Learning Course Project - CentraleSupélec

Xavier Plantier et Riccardo Castellano

1. Problem Statement

Fitness tracking platforms are increasingly focusing on understanding and predicting users' athletic performance based on their activity data. Accurately modeling performance progression requires analyzing different type of signals, including GPS trajectories, heart rate dynamics, speed variations, and elevation profiles. Our project aims to develop a deep learning system capable of predicting individual performance levels. This approach can enhance personalized training recommendations, optimize goal setting, and provide deeper insights into endurance and fitness trends over time.

2. Dataset

We utilize the Endomondo fitness tracking dataset containing 253,020 workouts with heart rate data. Each workout includes time-series measurements of speed, altitude, GPS coordinates, heart rate (BPM), and activity labels (run, bike, bike transport, etc.). The dataset naturally provides ground truth labels for fraud detection by treating bike/scooter activities as fraudulent runs. Our exploratory data analysis on 1,000 samples revealed that heart rate patterns are the strongest fraud indicators, with genuine runners maintaining significantly higher average heart rates (144 BPM) compared to bikers (128 BPM).

3. Proposed Deep Learning Approach

We propose a multi-phase implementation strategy:

- **Phase 1 - Baseline:** Random Forest classifier using 6 statistical features (speed mean/std, HR mean/std, HR-speed correlation, speed variability). Achieved: 84.3% accuracy.
- **Phase 2 - LSTM/GRU:** Recurrent neural networks to leverage full time-series sequences rather than aggregated statistics. Target: >90% accuracy.
- **Phase 3 – Fine-tunning :** Testing with existing pretrained model, like fondation model, with transformers architecture
- **Phase 4 - Multi-Modal Fusion:** Advanced architecture combining separate encoders for speed, heart rate, and elevation sequences, with late fusion for final classification.

endomondoHR.json :(253,020 workouts), contains detailed time-series data WITH heart rate information. Each line is one workout with:

Time-series data (arrays synchronized by index):

- - speed: Speed measurements (km/h)
- - altitude: Elevation data (meters)
- - heart_rate: Heart rate (BPM) - KEY for fraud detection!
- - latitude / longitude: GPS coordinates
- - timestamp: Unix timestamps for each measurement