Predicting problematic internet use in children and adolescents

Using physical activity and fitness data to flag early signs

CLATIN Thibaut / OLABODE Jens-David / EISENGARTEN Laurits / TEAM #7

POSTECH · IMEN574_PFDS · June 2026

The problem, and why it matters

- Heavy internet use in kids often shows up in the body first: worse sleep, less movement, shifting fitness.
- Measuring it directly needs clinical interviews, which are slow, costly, and out of reach for many families.
- The idea is to use cheap, widely available fitness and activity data as a stand-in.
- The task: predict each child's Severity Impairment Index (sii), from 0 (none) to 3 (severe).

Scored on QWK

Quadratic weighted kappa rewards being close, not just exactly right.

0 random → **1** perfect

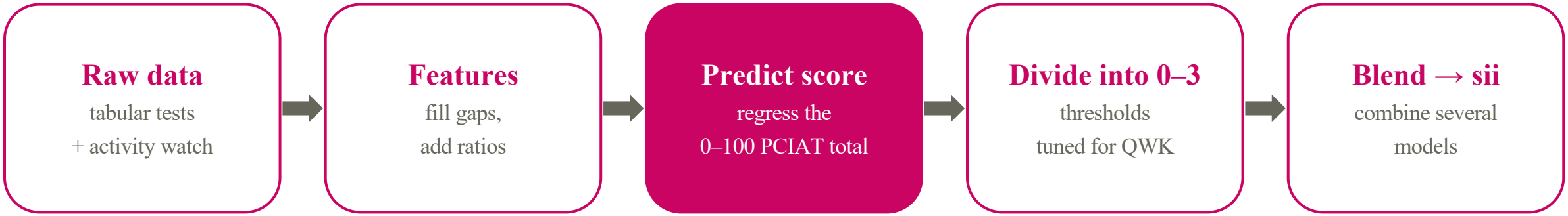
The data, and a problem with the test set

- About 3,960 children, each described by fitness tests and questionnaires; roughly 1,000 also wore an activity watch.
- The real Kaggle test set is hidden, so you cannot see the answers or score a model offline.
- About 1,224 children in the training data have no sii label at all.
- Our fix: split the labelled data 80/20 into our own train and test set, and judge progress on that.
- The classes are very uneven. None dominates; Severe is rare, with only 34 children.



Severity classes in the training data

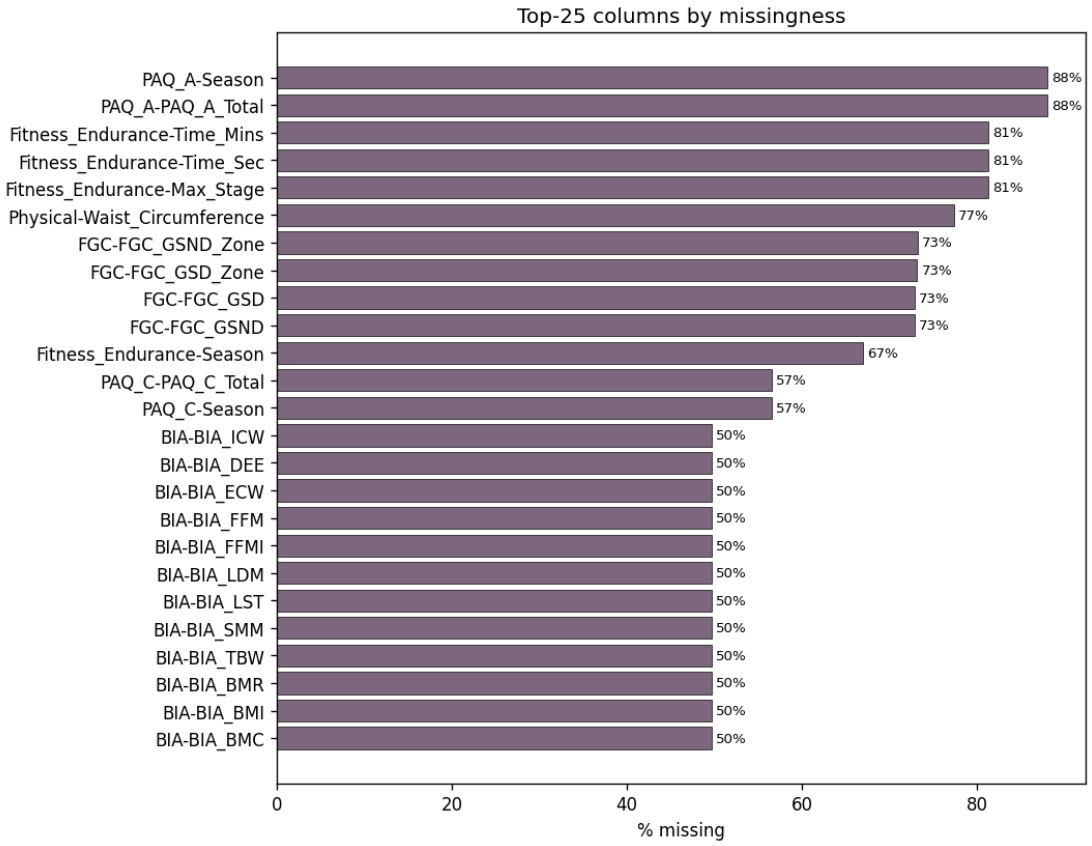
How the working method fits together



Key move: instead of guessing the class directly, we predict the underlying 0–100 questionnaire score, then chose the thresholds that score best on QWK.

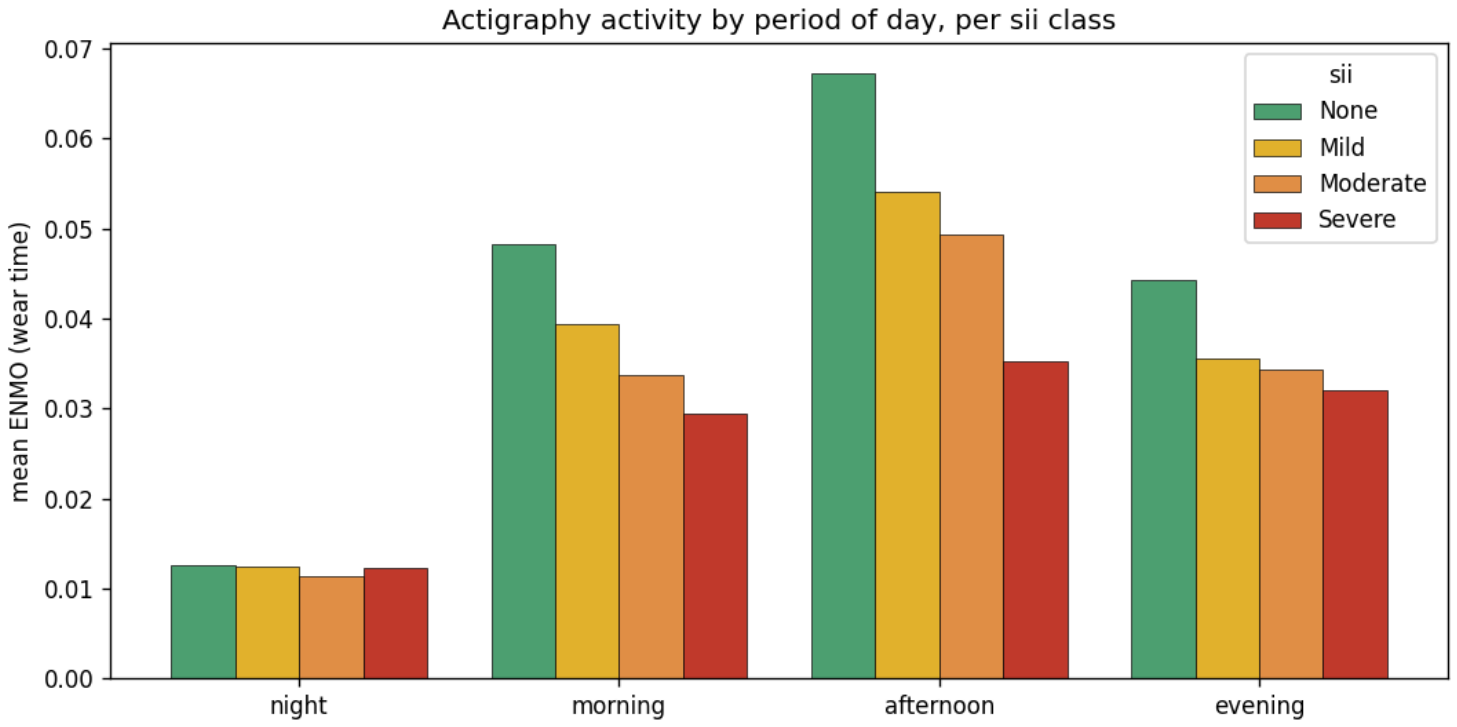
Cleaning the data and choosing features

- Most measurements are missing for most children. Which ones are missing is itself a clue, so we flagged the gaps instead of hiding them.
- We added simple, meaningful combinations: body and blood-pressure ratios, age × BMI, age × internet hours.
- We kept the signals that actually track severity (see right) and filled the rest with median values, so every model sees the same columns.



How often each measurement is missing (top 25)

Why the activity data carries the signal



Mean daytime movement (ENMO), by time of day and severity

- The watch records how much a child moves through the day.
- Lower-severity kids move more in daytime, and the afternoon gap is the clearest.
- At night everyone is still, so the daytime difference is the useful part.
- This is the behavior the whole competition is built on.

The models we used

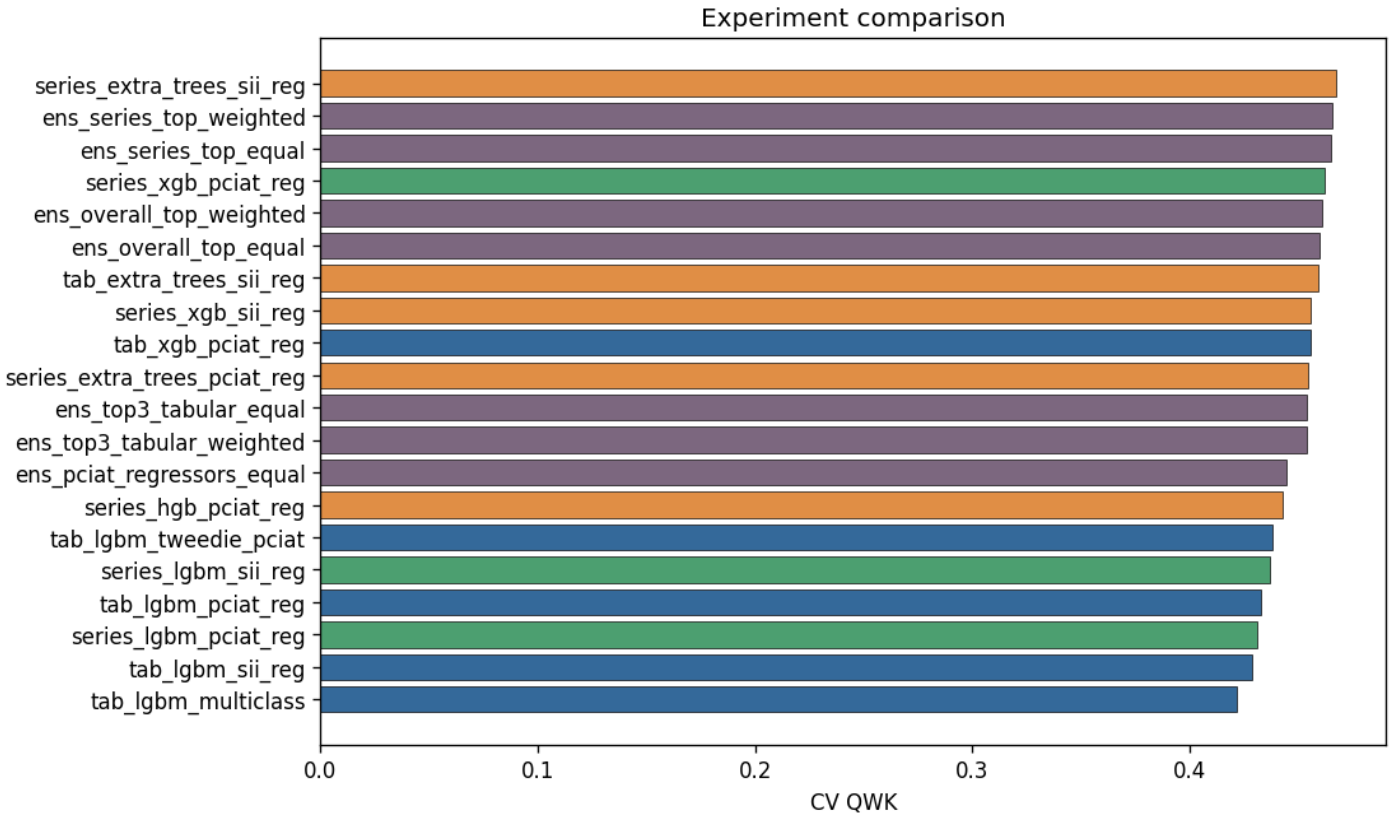
We leaned on gradient-boosted trees. They handle messy, partly missing tabular data well and need little tuning.

- **LightGBM:** fast, strong baseline
- **XGBoost:** best single tabular score in our runs
- **Extra Trees:** best overall, especially with the watch features
- **HistGradientBoosting:** a useful fourth voice in the blend

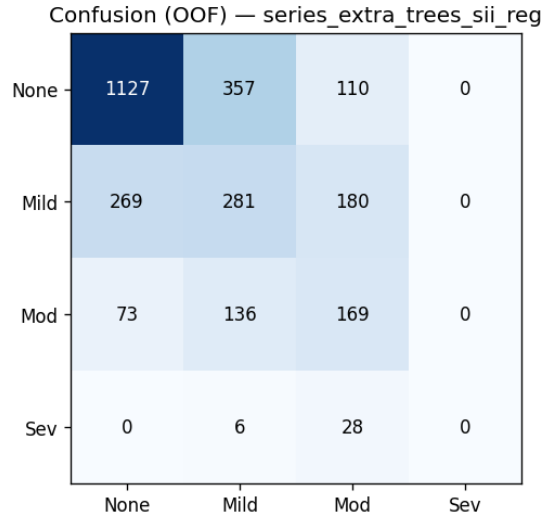
14

experiments across five learner types, then combined into ensembles

Results, and why we missed the top score



Cross-validated QWK for every experiment and ensemble



0.467

best local QWK

- Best model: Extra Trees on the watch features.
- Strong on None, fair on Mild and Moderate, never predicts Severe.
- Why: only 34 severe cases, and the hidden test set is a different sample. Read it as a screen, not a diagnosis.

Using coding agents to move faster

- We used Claude Code and Codex to build and compare many model variants quickly.
- They wrote the experiment scaffolding, ran the cross-validation loops, and summarised what came back.
- That let us try 14+ setups and several ensembles in the time one or two would usually take.
- Our effort shifted from writing boilerplate to deciding what to test next.

Claude Code

scaffold + run + summarise

Codex

second pair of hands

What we learned, and what's next

What worked

- Predicting the 0–100 score, then cutting it, beat predicting the class directly.
- Both the missingness flags and the watch features earned their place.
- Boosted trees stayed ahead of fancier options on this data size.

What's next

- Handle the rare Severe class better (resampling, class weights).
- Use more of the raw watch signal instead of hand-made summaries.
- A small neural net (MLP) we tried lost to the trees; revisit it with more data.
- Line up our validation more closely with the hidden test set.