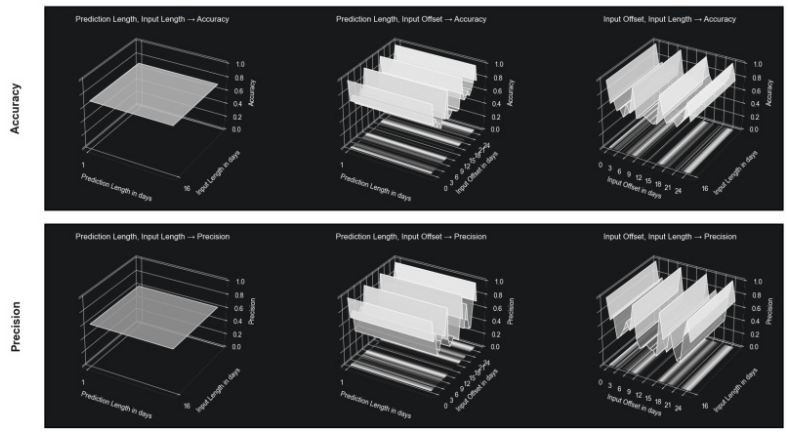
Hessisches Kompetenzzentrum
für Hochleistungsrechnen

Feature Engineering for Deep Spatiotemporal Prediction in Last-Mile Logistics



Project Manager
Fayong Jiao

Principal Investigator
Prof. Dr. Heinz Koeppl

Project Term
2025 - 2025

Clusters
Lichtenberg II Cluster Darmstadt

Software
PyTorch

Additional Software
CUDA Toolkit, SLURM Workload Manager

Institute
Fachbereich Elektrotechnik und Informatik

University
Technische Universität Darmstadt

Introduction

Accurately predicting recipient presence at their delivery address is a critical challenge in last-mile logistics, as failed delivery attempts due to recipient unavailability create substantial economic and environmental costs , inflate logistics expenses, and produce additional carbon emissions from repeated trips . This project addresses the challenge of forecasting individual presence at specific locations using historical GPS trajectory data from 13 users. The computational demands of this task necessitate High Performance Computing infrastructure due to several factors: processing large-scale temporal datasets with thousands of time steps per user, training multiple complex deep learning architectures simultaneously , conducting extensive feature engineering experiments across spatial, temporal, weather, and holiday dimensions , and performing computationally intensive model interpretability analysis. Each model training iteration requires significant GPU memory and computational time, while systematic hyperparameter optimization and feature selection experiments require parallel job execution that would be infeasible on standard workstations.

Methods

We developed a modular time series prediction framework based on PyTorch Lightning, implementing three architectures: PatchTST, PewLSTM, Time-Shift Transformer. The feature engineering pipeline generates over 30 features organized into four groups: temporal features using cyclical encoding for hour, day, week, and season patterns; spatial features including displacement, velocity, and zone transitions; weather conditions encompassing temperature, humidity, wind, and precipitation;

and holiday indicators with proximity measures. We employed SLURM workload manager on TU Darmstadt HPC cluster with CUDA-enabled GPU nodes, utilizing batch job submission for systematic experiments across different model- feature combinations. Model interpretability was achieved through SHAP analysis using Kernel Explainer to identify feature importance and understand prediction mechanisms. The data procurer system handles dynamic multi-user data loading with configurable sequence lengths and time shifts.

Results

The PatchTST model achieved 71.56 % accuracy and 64.49 % precision for one-day-ahead presence prediction using 16-day input sequences, evaluated across 10 users and 260 prediction variations. We successfully trained and compared LSTM, Transformer, and TFT variants across multiple feature configurations, generating comprehensive evaluation metrics and visualizations stored in the evaluations directory. The SHAP analysis produced six types of interpretability visualizations including global feature importance rankings, beeswarm plots showing feature value distributions, waterfall diagrams explaining individual predictions, and heatmaps revealing cross-sample patterns. Feature ablation studies comparing basic, temporal, spatial, weather, and holiday feature groups were completed, with results indicating that spatial movement features and temporal cyclical encodings contribute most significantly to prediction accuracy. Over 78 experimental configurations were executed through automated HPC batch processing, accumulating detailed training logs and model checkpoints for subsequent analysis. The computational efficiency enabled rapid iteration cycles, with each full training experiment completing within 24 hours using single GPU nodes.

Discussion

The results demonstrate that the systematically engineered feature set, after selection via SHAP analysis, provided consistent performance improvements across all three evaluated model architectures—PatchTST, Transformer, and PewLSTM. The SHAP analysis revealed that location-based features such as distance to home zones and movement velocity are more predictive than external factors like weather conditions, suggesting that habitual movement patterns dominate individual presence behavior. Future work should explore ensemble methods combining multiple architectures, investigate attention mechanisms to understand which historical time steps influence predictions most, extend the framework to handle irregular sampling intervals common in real-world GPS data, and deploy the trained models for real-time prediction in production environments. The modular design facilitates integration of additional feature groups and model architectures, while the HPC infrastructure enables scaling to larger user populations for improved generalization.

Figures

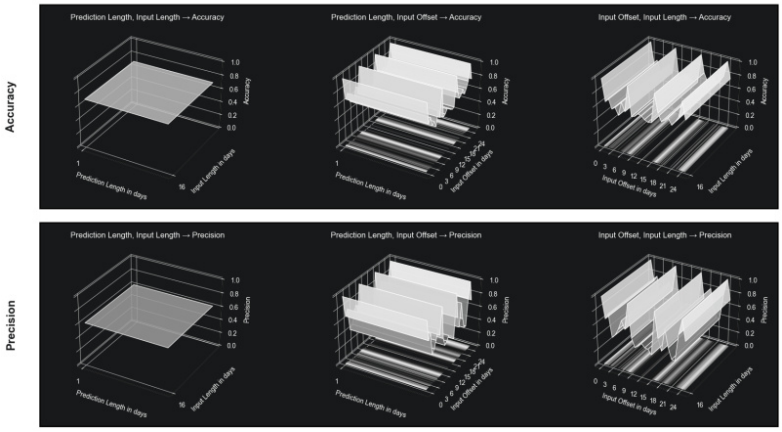


Figure 1: Baseline evaluation metrics for the 1-day forecast PatchTST model

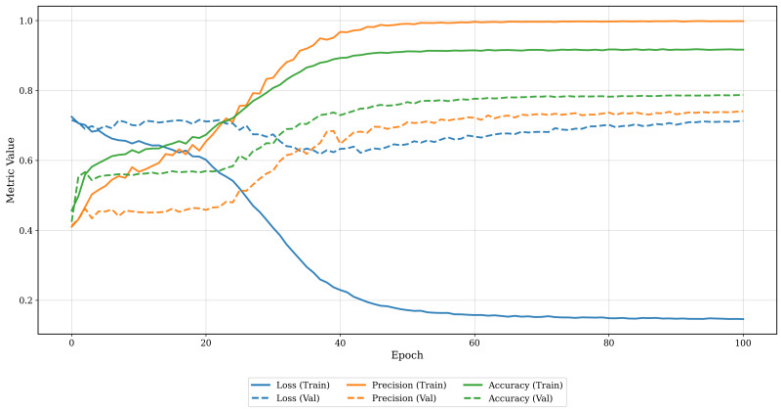


Figure 2: Training metrics for the basic features plus temporal features

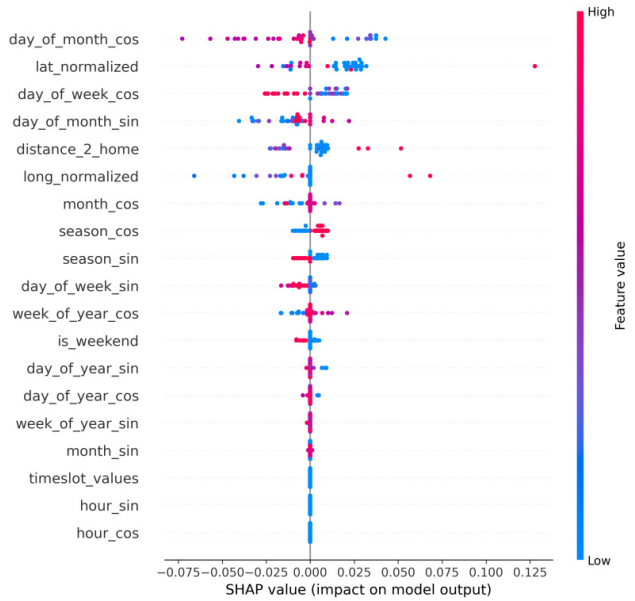


Figure 3: SHAP beeswarm summary plot for temporal features

Last Update: 2026-01-20 08:42

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN
TECHNISCHE UNIVERSITÄT DARMSTADT
UNIKASSEL VERSITÄT
Philipps Universität Marburg