UNIVERSITY OF LAGOS · FACULTY OF ENGINEERING · DEPARTMENT OF SYSTEMS ENGINEERING

IoT-Enabled Real-Time Logistics Tracking and Delay Prediction Using Graph Neural Network–LSTM for Supply Chain Visibility

Paper CMT-30 · Revised submission · A graph–sequence delay predictor with an acquisition pathway for corridors that carry no fixed instrumentation.

Presented by

David Dickson

Matriculation No. 190407031

Supervisor

Dr. John Ogbemhe

B.Sc. Final Year Project

THE PROBLEM

The corridor that most needs a delay forecast has no data to build one from

The operational cost is real

Freight through the Lagos–Apapa corridor is subject to queueing, terminal throttling and access-road failure. Dwell is measured in hours and sometimes days, and it is not predictable to the operators who must plan around it.

But the corridor is uninstrumented

There are no inductive-loop detectors, no corridor-wide speed feed, no public historical record. A predictor cannot be trained on it, and — just as importantly — it cannot be **tested** on it either.

So the project has two problems, not one

Predicting delay is a modelling question. Obtaining the measurements in the first place is an engineering question. Most published work answers only the first.

This project separates those two questions, answers each against evidence, and joins them at a defined interface.

THE CONCEPTUAL MOVE

Visibility is not predictability

What visibility gives you

Control towers, IoT tracking and digital twins answer **“where is the consignment now, and what state is it in?”** The literature shows this is now largely a solved engineering problem.

What it does not give you

None of it answers **“how will this disruption propagate, and when will it reach me?”** Observing a queue forming is not the same as forecasting the delay it will cause downstream in an hour.

A system can observe every event on a network and still forecast badly, because delay is produced by **interaction between locations over time** — not by any single location’s current state.

That is why this work uses a graph model. The corridor is not a set of independent sensors; it is a chain in which congestion at one location constrains its neighbours.

RESEARCH GAP

Fourteen predictive studies reviewed. All fourteen assume the data already exists.

Literature	Models the network?	Acquisition assumed	Fits an uninstrumented corridor?
Classical forecasting (ARIMA, Kalman, SVR)	No	Fixed detectors	No
Spatio-temporal GNNs (DCRNN, STGCN, MTGNN)	Yes	Dense detector networks (METR-LA, PEMS-BAY)	No
Freight / intermodal ETA	No	Carrier event feeds	No
Developing-country logistics	No predictive architecture	Documents the data gap	Describes, does not solve

The studies that **model the network** all run on already-instrumented networks. The studies that address **under-instrumented settings** contribute no predictive architecture. **No reviewed study does both — and that conjunction is the gap.**

AIM AND OBJECTIVES

Aim

To develop and evaluate a GNN–LSTM framework for corridor delay prediction, together with a source-agnostic edge-IoT acquisition pathway for freight corridors that carry no fixed instrumentation — of which Lagos–Apapa is the motivating case.

---

1 · Review

Identify the gap across IoT tracking, supply chain visibility and graph learning in developing-country settings.

2 · Acquire

Design an architecture able to fetch and fuse heterogeneous IoT data along the corridor.

3 · Model

Construct a GNN–LSTM to forecast spatio-temporal corridor-level delay.

4 · Implement

Realise it in Python and PyTorch against a representative logistics dataset.

5 · Evaluate

Assess by MAE and RMSE against ARIMA, Random Forest and a standard LSTM.

SYSTEM ARCHITECTURE

Six layers, with one substitutable interface

1 · Acquisition	**2 · Feature contract**	3 · Preprocessing	4 · Graph	5 · Prediction	6 · Visibility
Loop detectors where they exist, or the edge-IoT probe where they do not. Both paths terminate at layer 2.	Node × 15-min bin × {flow, occupancy, speed} plus a quality tuple. **Every layer above consumes this, never a sensor.**	Quality-flag-aware imputation, outlier screening, scaling fitted on the training partition only.	Twelve corridor nodes; edge weights from road-network distance under an exponential kernel.	GNN–LSTM producing a delay estimate per node at 15, 30 and 60 minutes.	Forecast, acquisition health and realised error surfaced as they arrive.

The design decision that matters is **layer 2**. Because every layer above it consumes the contract rather than any particular sensor, the acquisition path can be replaced without touching the model.

CORE CONTRIBUTION

The source-agnostic node feature contract

What a node must report

For each node, in each 15-minute bin:

- **flow** — vehicles per bin
- **occupancy** — fraction of time occupied
- **mean speed** — km/h

…accompanied by a quality tuple: **source**, **error estimate**, **interpolation flags**, **sample count**.

The quality tuple is what makes the contract real rather than decorative.

A detector reporting 300 vehicles and a probe inferring speed from 4 passes are **not** equal evidence. The tuple carries that difference forward, so preprocessing can weight, flag or reject a reading on its provenance.

Without it, substituting the sensor would silently change what the model is learning from. With it, substitution is an engineering operation with stated consequences.

Why this answers the gap

It is the mechanism by which a model trained on an instrumented network can be fed by a device on an uninstrumented one.

ACQUISITION · BUILT AND MEASURED

The edge-IoT probe is characterised hardware, not a block diagram

Build

ESP32 microcontroller · NEO-M8N GNSS receiver · SIM800L cellular modem · microSD store-and-forward · MQTT uplink to a Python subscriber.

Writes to the SD card **first**, transmits second — so a dropped cellular link loses nothing.

13,872
valid fixes logged over 4 h 22 min, 88.1% availability

4.51 m
CEP50 positional scatter (CEP95 = 12.70 m)

1.35×
geofence hysteresis multiplier, **derived from this measurement**

A static receiver inside a geofence produced **≈76 false entry/exit pairs in four hours** at a 1.00× boundary. At 1.35× for radii ≥ 100 m, that count is zero.

THE PREDICTIVE MODEL

Graph aggregation, then recurrence, then a head per horizon

What is predicted

Excess travel time per kilometre — the extra minutes to traverse 1 km relative to an empirically estimated free-flow reference.

Reciprocal in speed, so it is additive along a route. Signed and unclipped.

How the corridor is represented

Twelve nodes over **8.44 km**. Edge weights from road-network distance under an exponential kernel, thresholded, then symmetrically normalised.

108 of 132 off-diagonal entries survive.

The model

Two-layer GraphSAGE-style encoder → shared LSTM → three linear heads.

2,867 trainable parameters. Pure PyTorch; no geometric library.

The corpus

GO-MO, derived from the Madrid City Council Open Data Portal. 24 months at 15-minute resolution.

842,112 records. Chronological split, no shuffling.

Madrid supplies what Lagos–Apapa cannot: a measured, continuous, publicly auditable record against which a predictor can be both trained **and falsified**.

EVALUATION PROTOCOL

The test partition was opened once

Sealed test

Chronological 70/10/20 split. Every model fitted on training alone, tuned on validation alone, evaluated **once** on the held-out partition. No model was retrained or reselected after it was opened.

Dependence-aware testing

Test samples are serially and cross-sectionally dependent, so ordinary resampling understates uncertainty. Significance uses a **day-block paired bootstrap** over 147 calendar blocks.

Multiplicity controlled

Holm–Bonferroni correction within each horizon family, so that testing three horizons against four baselines does not manufacture a result.

This matters because it changed a conclusion. Under ordinary resampling one comparison reached **p = 0.034** and looked significant. Its standard error was understated by a factor of **1.40**. Under day-block resampling the same comparison gives **p = 0.153**, and is reported as not established.

162,022 test samples per horizon, across 147 days

RESULTS · SEALED TEST, 162,022 SAMPLES PER HORIZON

Mean absolute error, min/km

Model	15 min	30 min	60 min	Capacity
Historical average (floor)	0.0973	0.0973	0.0973	—
ARIMA	0.0474	0.0625	0.0822	one model per node per horizon
Standard LSTM (no graph)	0.0441	0.0517	0.0622	1,395 parameters
GNN–LSTM (proposed)	0.0422	0.0487	0.0586	2,867 parameters
Random Forest	0.0378	0.0464	0.0550	≈ 6.45×10^8 decision nodes

The graph term works

GNN–LSTM beats the otherwise identical sequence model at every horizon, **$p \leq 0.0005$** adjusted, and the margin **grows with forecast reach**.

It replicates

The same pattern reproduces on the independent PEMS-BAY benchmark — 2.1–4.1% improvement on the corridor, 2.5–4.4% on PEMS-BAY.

But it is not the most accurate model

An unconstrained Random Forest attains lower MAE at every horizon. **This is reported without qualification.**

READING THAT RESULT CORRECTLY

The Random Forest's advantage is bought with capacity, not architecture

The two models are not comparable in size. So three further baselines were fitted — with the proposed model frozen and verified to reproduce its sealed figures exactly beforehand.

Capacity-matched baseline	Relative capacity	MAE outcome	RMSE outcome
Per-node Ridge	0.5× (1,332 weights)	GNN–LSTM wins ×3	GNN–LSTM wins ×3
Compact Random Forest	3.1× (9,000 leaves)	tie, tie, GNN wins at 30 min	GNN–LSTM wins ×3
Per-node MLP (no graph)	11.7× (33,636 weights)	GNN–LSTM wins ×3	GNN–LSTM wins ×3
Unconstrained Random Forest	224,974×	Random Forest wins ×3	no significant difference

The corridor is not linearly predictable. Ridge loses on both metrics at every horizon, so the non-linear model class is warranted by evidence rather than assumed.

The graph term, not just non-linearity, is doing the work. An MLP with 11.7× the parameters but no network representation still loses at every horizon.

And it is faster. 54.97 ms vs 184.20 ms per batch — the proposed model is **3.35× quicker** at 1/225,000 of the size.

LIMITATIONS · STATED, NOT DISCOVERED IN QUESTIONING

What this work does not establish

1 · The motivating corridor is still unmeasured

The model is trained and tested on Madrid. Transfer to Lagos–Apapa is demonstrated **architecturally and operationally, not empirically**. No accuracy claim is made for Lagos.

2 · The device is characterised stationary

Pilot size: one tracker, no vehicle deployment. In-motion accuracy, multi-device operation, obstructed sky, battery endurance, vehicle power and installation, and cellular coverage along the route are all **untested**.

3 · Governance is specified, not discharged

Operator agreement, driver consent, a legal basis for processing, a retention period and the monitoring-purpose boundary are all **stated requirements, none of them met**. A public test broker with no authentication was used.

4 · The squared-error advantage is indicated, not established

On the full sealed partition it is not significant. It becomes significant only when 92 physically impossible detector records are excluded — a **post-hoc** rule, however principled.

5 · Twelve nodes is a small graph

Against METR-LA's 207. Legitimate as a demonstration — STGCN's own BJER4 benchmark is twelve roads — but **no city-scale claim is made**.

6 · The sampling interval hides propagation

Queue spillback is present at all seven junctions, but a queue transits a 400–850 m gap in 1.5–3.5 min while bins are 15 min. Propagation collapses into within-bin covariance.

CONTRIBUTIONS AND NEXT STEPS

What the project contributes

A source-agnostic feature contract

An interface at which a detector feed and an edge-IoT probe are substitutable, with provenance carried forward so the substitution is accountable rather than silent.

Evidence that the graph term earns its cost

Significant at every horizon against an identical sequence model, replicated on an independent benchmark, and not displaced by any capacity-comparable baseline.

A measured acquisition device

Built, logged and characterised — with the geofence hysteresis multiplier **derived from its own error distribution** rather than assumed.

A verification record

Including a contraction-ordering fault found by implementing the graph layer explicitly, and a result of my own that the correct statistical test overturned.

Next: close the empirical gap

A labelled field campaign on the corridor itself — multi-probe deployment, in-motion characterisation, and re-evaluation at finer temporal resolution so propagation becomes observable.

Next: make the error profiles work together

The two model classes fail differently. A regime-aware combination — graph model in free flow, tree ensemble in breakdown — is the obvious unexploited result.

THANK YOU

Questions

Presenter

David Dickson · 190407031

Department of Systems Engineering, University of Lagos

In one sentence

A graph–sequence delay predictor, and the acquisition interface that would let an uninstrumented corridor feed it.