Artículo original/ Original article
ISSN 2953-6553

RUTEA: Sistema Predictivo de Tráfico Vehicular para Ciudades Intermedias mediante arquitectura Hibrida, Deep Learning y Procesamiento de Lenguaje Natural

RUTEA: A Predictive Vehicular Traffic System for Intermediate Cities Using Hybrid Architecture, Deep Learning and Natural Language Processing.

Andrés Mejía [1] 0009-0009-5242-495X, Alexis Matute [2] 0009-0000-6159-2095, Lady Sangacha [3] 0000-0002-5169-8918

lady.sangacha@tecazuay.edu.ec, jose.mejia.est@tecazuay.edu.ec, lady.sangacha@tecazuay.edu.ec
[1 2 3] Instituto Superior Tecnológico del Azuay con Condición de Universitario Cuenca-Ecuador

DOI 10.36500/atenas.5.006

Resumen

La congestión vehicular en ciudades intermedias latinoamericanas representa un desafío creciente agravado por la ausencia de datos históricos continuos e infraestructura de sensórica dedicada. El presente trabajo tiene como objetivo integrar las fuentes de datos, clasificar incidentes, predecir el flujo vehicular e implementar la interfaz, describiendo a RUTEA como un sistema predictivo de tráfico que integra redes neuronales recurrentes LSTM con un clasificador de Procesamiento de Lenguaje Natural (PLN) en una arquitectura de base de datos híbrida (PostgreSQL/MongoDB). Ante la disponibilidad de un único día de aforo vehicular provisto por el GAD Municipal de Cuenca, se diseñó un esquema de simulación en base de datos que replica variaciones temporales mediante factores calibrados del Highway Capacity Manual (HCM 2016) y ajustes de calendario escolar local. El modelo LSTM recibe un tensor de entrada de dimensión (12, 4); ventana de 3 horas, con 4 variables exógenas y produce predicciones de flujo vehicular a 15 minutos. El clasificador PLN, entrenado con 324 noticias viales etiquetadas extraídas de medios locales, clasifica la severidad de incidentes en tres categorías e inyecta este valor como señal de ajuste dinámico en la ventana de predicción. Los resultados en entorno sintético controlado reportan un MAE promedio de 68.65 vehículos/15 min y un accuracy global del clasificador PLN de 89%. El sistema se despliega mediante una interfaz interactiva en Streamlit. RUTEA establece una línea base metodológica y tecnológica replicable para ciudades intermedias con recursos limitados de datos.

Abstract

Traffic congestion in medium-sized Latin American cities is a growing challenge, exacerbated by the lack of continuous historical data and dedicated sensing infrastructure. This study aims to integrate data sources, classify incidents, predict traffic flow, and implement a user interface through RUTEA, a predictive traffic system that combines long short-term memory (LSTM) recurrent neural networks with a natural language processing (NLP) classifier within a hybrid database architecture (PostgreSQL/MongoDB). Given the availability of only one day of traffic counts provided by the Municipal Government of Cuenca, a database-based simulation framework was designed to reproduce temporal variations using calibrated factors from the Highway Capacity Manual (HCM 2016) and adjustments based on the local school calendar. The LSTM model receives an input tensor of shape (12, 4), representing a three-hour window with four exogenous variables, and generates traffic flow forecasts for the next 15 minutes. The NLP classifier, trained on 324 labelled traffic news articles collected from local media, assigns incident severity to three categories and incorporates this value into the prediction window as a dynamic adjustment signal. Evaluation in a controlled synthetic environment yielded an average mean absolute error (MAE) of 68.65 vehicles per 15-minute interval and an overall NLP classifier accuracy of 89%. The system is deployed through an interactive Streamlit interface. RUTEA establishes a replicable methodological and technological baseline for medium-sized cities with limited data resources.

Palabras Claves – gestión de tráfico, inteligencia artificial, redes LSTM, aprendizaje profundo, procesamiento de lenguaje natural.
Keywords – traffic management, artificial intelligence, LSTM networks, Deep learning, Natural language processing.

Recibido: 202-09-30 Aprobado tras revisión: 2026-10-19

I. INTRODUCCIÓN

La movilidad urbana enfrenta presiones crecientes a nivel global: se estima que para 2030 el 60% de la población mundial residirá en zonas urbanas, intensificando la demanda sobre sistemas viales diseñados para cargas menores (Openvia, 2015). Esta realidad es especialmente crítica en ciudades intermedias latinoamericanas, donde el crecimiento del parque automotor supera la capacidad de expansión de la infraestructura vial. Cuenca, Ecuador, ciudad declarada Patrimonio de la Humanidad por la UNESCO con aproximadamente 380.000 habitantes, ilustra este escenario: entre 2018 y 2025 el parque automotor creció cerca del 18% anual, presionando 1.200 km de vías que operan en niveles de servicio críticos (Aguagallo y Becerra, 2023). Corredores como la Av. de las Américas registran flujos superiores a 3.200 vehículos/h con relaciones volumen/capacidades mayores a 1 (HCM, 2016). La morfología urbana agrava el problema: los ríos Yanuncay, Tomebamba, Machángara y Tarqui concentran el tráfico en 28 puentes que actúan como nodos de estrangulamiento, mientras el Centro Histórico limita la expansión vial con calles menores a 6 m (Orellana, 2015).

Se presentan las consecuencias económicas: pérdidas cercanas al 2,3% del PIB local, ambientales: con un incremento del 14% en emisiones de CO_2 y sociales: la reducción del 17% en satisfacción ciudadana justifican el desarrollo de herramientas predictivas de gestión del tráfico (PADECO, 2013; Aguagallo y Becerra, 2023).

Los LSTM, que son modelos algorítmicos, han demostrado superioridad sobre métodos estadísticos clásicos como ARIMA y SVM en series temporales con dependencias complejas (Zheng et al., 2017; Wei et al., 2019). Paralelamente, el Procesamiento de Lenguaje Natural (PLN) ha emergido como herramienta eficaz para la detección automática de incidentes viales a partir de fuentes textuales (Melhem et al., 2024). Sin embargo, persiste una brecha metodológica relevante: la mayoría de estos enfoques presupone la disponibilidad de datos históricos continuos de al menos un año y una infraestructura densa de sensores, condiciones ausentes en ciudades intermedias. Estudios previos en Cuenca evidencian que el HCM sobrestima la capacidad vial local en un 58% debido a diferencias en parámetros de conducción (Carpio y Avilés, 2018), lo que señala la necesidad de soluciones adaptadas al contexto local.

El presente trabajo propone RUTEA (Ruteo Urbano Temporal Estratégico Azuayo), un sistema que aborda esta brecha mediante tres contribuciones: (1) un esquema de simulación de patrones de tráfico usando una base de datos relacional basado en factores CAF del HCM 2016 adaptados al contexto

Recibido: 202-09-30 Aprobado tras revisión: 2026-10-19

cuencano, que permite entrenar modelos de series temporales con un único día de aforo disponible; (2) una arquitectura híbrida PostgreSQL/MongoDB que integra datos estructurados de tráfico con noticias viales no estructuradas; y (3) la inyección de la severidad PLN como señal de ajuste dinámico en la ventana de entrada del modelo LSTM. El artículo se organiza como sigue: la Sección 2 presenta el marco teórico; la Sección 3 describe la metodología; la Sección 4 reporta los resultados; la Sección 5 discute los hallazgos; y la Sección 6 presenta las conclusiones.

II. MARCO TEÓRICO

Redes LSTM para predicción de tráfico

Las redes de memoria a corto y largo plazo (LSTM) constituyen una variante especializada de las redes neuronales recurrentes (RNN) diseñada para abordar el problema del gradiente desvaneciente (Goodfellow et al., 2016). Su arquitectura incorpora celdas de memoria controladas por tres puertas: entrada, olvido y salida, que regulan el flujo de información a través del tiempo, permitiendo capturar dependencias de largo alcance en series temporales. Zheng et al. (2017) demostraron que las LSTM superan a ARIMA (MRE=11.23), SVM (MRE=10.52) y RNN estándar, alcanzando un MRE de 6.22 en predicción de tráfico a 15 minutos para 500 estaciones en Beijing. Chauora et al. (2024) extendieron este resultado combinando LSTM con optimización por enjambre de partículas (PSO), logrando reducciones adicionales de MAE y RMSE en intersecciones urbanas. Estas evidencias sustentan la elección de LSTM como núcleo predictivo de RUTEA.

PLN para detección de incidentes viales

El Procesamiento de Lenguaje Natural (PLN) permite transformar texto libre (noticias, reportes policiales, publicaciones en redes sociales) en señales estructuradas para sistemas de predicción (Melhem et al., 2024), donde propusieron un framework de tres etapas para detección de incidentes: clasificación de relevancia textual, extracción de entidades mediante grafos de conocimiento y predicción integrada con modelos de lenguaje de gran escala, obteniendo mejoras significativas frente a métodos tradicionales. Para contextos con recursos computacionales limitados, la combinación de vectorización TFIDF con redes neuronales densas representa una alternativa eficiente y reproducible para clasificación de texto de severidad de eventos (Pedregosa et al., 2011).

Arquitectura de datos híbrida

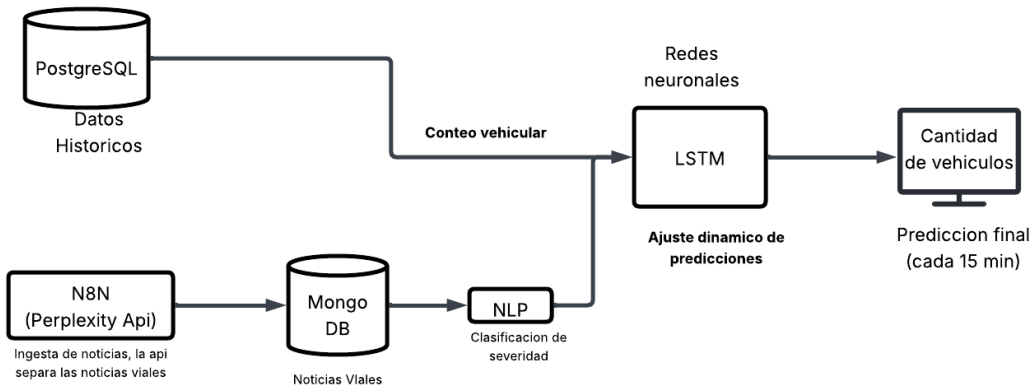
Es el modelado predictivo de tráfico que requiere gestionar simultáneamente series temporales estructuradas con alta frecuencia de actualización y datos no estructurados de fuentes textuales externas.

Recibido: 202-09-30 Aprobado tras revisión: 2026-10-19

Las bases de datos relacionales como PostgreSQL ofreciendo la integridad transaccional y soporte para consultas analíticas complejas mediante índices especializados, mientras que los sistemas documentales como MongoDB proveen flexibilidad de esquema para la ingesta ágil de datos heterogéneos (Achan 2024). La arquitectura híbrida SQL/NoSQL es reconocida como solución adecuada para sistemas de tráfico en tiempo real que requiere consultas históricas complejas junto con captura flexible de eventos externos (Fig.1).

Figura 1

Arquitectura de datos hibrida usando PostgreSQL y MongoDB



Nota. Diseño arquitectónico del proceso de RUTEA. 2026

III. METODOLOGÍA Y CÁLCULOS

El desarrollo de RUTEA siguió la metodología Cross Industry Standard Process for Data Mining (CRISPDM), adaptada al dominio de predicción de tráfico (Chapman et al., 2000)(Reyes et al., 2022). Las fases de comprensión del negocio y de los datos definieron el problema, las intersecciones piloto y las fuentes disponibles; la preparación de datos integró el esquema de simulación y el pipeline PLN; el modelado implementó los modelos LSTM y el clasificador de severidad; y la evaluación validó el desempeño en entorno sintético controlado.

Los datos de conteo vehicular provienen del repositorio de datos abiertos del GAD Municipal de Cuenca (año 2023), que comprende registros de 43 intersecciones estratégicas en horario 06:00–20:00. Los registros incluyen el tipo de conteo (VEHICULAR, CICLISTA, PEATONAL), el código de maniobra por intersección y la clasificación por tipo de vehículo (livianos, buses, camiones). En este proyecto se priorizó el conteo vehicular, ya que el conteo de ciclistas y peatones no es consistente en

Recibido: 202-09-30 Aprobado tras revisión: 2026-10-19

todas las intersecciones, dejando las dos categorías mencionadas anteriormente para futuras versiones. Luego, a partir de un análisis del volumen promedio diario (TPDA), se seleccionaron 6 intersecciones de mayor criticidad, considerando flujos vehiculares y su rol como nodos en las líneas de transporte público 100, 28, 12 y 4, y como conectores hacia el Parque Industrial de Cuenca, considerando que son los lugares de criticiad de horas pico (Tabla 1).

Tabla 1

Intersecciones piloto seleccionadas por densidad vehicular promedio diaria

ID	Intersección	Latitud	Longitud	Vol. Prom. (veh/15 min)	Criticidad
7	Av. Las Américas y De las Laderas	2.8815	78.9848	1553	Alta
8	Av. González Suárez y Guapondelig	2.8991	78.9906	1115	Alta
9	Sebastián de Benalcázar y Del Chorro	2.8928	78.9944	933	Media
10	Sebastián de Benalcázar y Av. España	2.8935	78.9937	910	Media
11	Calle Vieja y Av. Elia Liut	2.8870	78.9893	672	Media
12	Puente Fabián Alarcón	2.8821	78.9742	2530	Alta

Nota. Vol. Prom. = volumen promedio de vehículos por intervalo de 15 minutos.

La disponibilidad de un único día de aforo impide capturar la variabilidad temporal necesaria para entrenar modelos de series temporales. Para resolver esta limitación, se implementó la vista vista_entrenamiento_lstm_2 en PostgreSQL, que genera registros en intervalos de 15 minutos para el período 2024-2025 mediante combinación temporal espacial. El conteo simulado se calcula aplicando factores multiplicativos al conteo base según prioridad: primero se evalúa la ocurrencia de eventos (severidad_pln 1–3), y en su ausencia se aplican ajustes de calendario (Tabla 2). Esta parametrización es explícitamente experimental y está sujeta a calibración con datos reales multimensuales.

Recibido: 202-09-30 Aprobado tras revisión: 2026-10-19

Tabla 2

Factores multiplicativos aplicados en la simulación de datos de entrenamiento

Variable	Condición	Factor aplicado	Fuente
severidad_pln = 1	Congestión / embotellamiento leve	0.80	HCM 2016 (adaptado)
severidad_pln = 2	Desvíos / cierres parciales	0.43	HCM 2016 (adaptado)
severidad_pln = 3	Siniestro / accidente grave	0.19	HCM 2016 (adaptado)
es_fin_de_semana = 1	Sábado o domingo	0.80	Parra (2017)
hay_clases = 0	Meses marzo y agosto (vacaciones)	0.90	Calendario escolar Ecuador

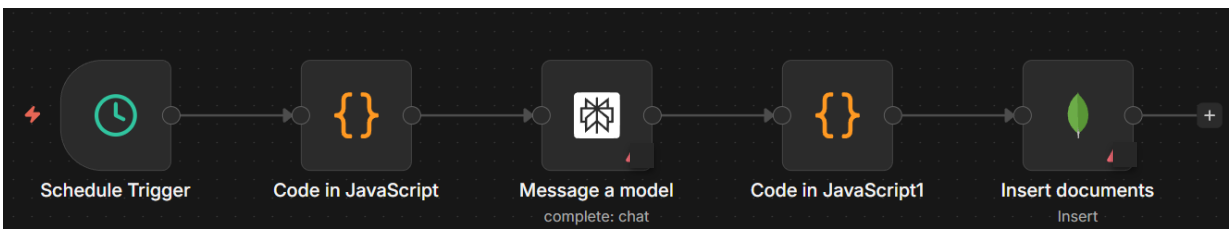
Nota. Los factores de severidad se derivan de los CAF del HCM (2016), empleados como aproximación proporcional ante la ausencia de factores calibrados localmente (Carpio y Avilés, 2018). Los factores de calendario se sustentan en evidencia indirecta de calidad del aire en Cuenca (Parra, 2017).

Dentro de sustentar el supuesto referente al menor flujo en los fines de semana, ha sido considerado para este estudio en el centro histórico de Cuenca (2013-2015), donde se reportan concentraciones menores de CO y PM2.5 en sábados y domingos asociadas a menor actividad vehicular (Parra, 2017). Se ha identificado que los meses de vacaciones escolares de primaria, secundaria y universitaria son marzo y agosto, con impacto documentado en la reducción del flujo vehicular en corredores urbanos.

En cuanto a la recolección de noticias viales (embotellamientos, trabajos, accidentes) la herramienta de automatización N8N juntamente con el api de Sonar 1.5(Perplexity) permitieron crear un sistema automático de búsqueda de noticias que contenga palabras clave como accidentes, congestión, obras viales (Fig.2) y los campos recopilados en la base de datos se detallan en la Tabla 3.

Figura 2

Automatización creada en N8N para la recolección de noticias viales



Nota. Recolección de noticias viales con el N8N con el api de sonar. 2026

Recibido: 202-09-30 Aprobado tras revisión: 2026-10-19

Tabla 3

Nombre y descripción de los campos almacenados en la base no relacional MongoDB

Campo	Descripción
fecha	Fecha de publicación de la noticia
Título	Titular de la nota con la que fue publicado en los medios
Sector	Lugar geográfico donde ocurrió el evento
Resumen	Sintesis de la nota periodística para entender el contexto del evento ocurrido
Fuente	Nombre del medio de comunicación que publicó la nota
URL	Enlace web hacia la noticia

Nota. Los campos considerados en la tabla de la base no relacional. 2026.

Clasificador PLN

El clasificador fue evaluado sobre un conjunto de prueba de 65 instancias (split 80/20). En la Tabla 3 se presentan las métricas por clase.

Tabla 4

Métricas de evaluación del clasificador de severidad PLN por clase

Clase	Precisión	Recall	F1Score	Soporte
1 Congestión habitual	1.00	0.75	0.86	8
2 Desvíos/cierres parciales	0.88	0.91	0.89	23
3 siniestros / accidentes	0.89	0.91	0.90	34
Accuracy global	0.89			65

Nota. El soporte indica el número de noticias de cada clase en el conjunto de prueba. La clase 1 presenta el menor soporte, con ocho instancias, por lo que sus métricas deben interpretarse con cautela. La exactitud global corresponde a la proporción de noticias clasificadas correctamente.

El desempeño del clasificador evidencia que la representación TFIDF combinada con una red neuronal densa captura patrones textuales útiles para diferenciar la severidad de eventos viales. La clase 3 (siniestros) presenta las métricas más estables dado su mayor soporte en el conjunto de prueba. Los errores de clasificación se concentran principalmente en confusiones entre clases 1 y 2, atribuibles al solapamiento semántico en el lenguaje periodístico cuando se describen eventos de mantenimiento versus congestión habitual.

Modelo LSTM

El modelo LSTM fue evaluado por intersección, reportando MAE y RMSE en unidades de vehículos por intervalo de 15 minutos. A continuación, observemos la tabla 5, donde denota el rendimiento del modelo LSTM por intersección en entorno sintético controlado de los lugares escogidos y además, el resumen del rendimiento por punto piloto.

Recibido: 202-09-30 Aprobado tras revisión: 2026-10-19

Tabla 5

Rendimiento del modelo LSTM por intersección en entorno sintético controlado

ID	Intersección	Vol. Prom. (veh/15 min)	MAE (veh)	RMSE (veh)
7	Avenida de las Américas y De las Laderas	1553	84.15	154.92
8	Avenida Gonzales Suarez y Guapondelig	1115	60.64	112.88
9	Sebastián de Benalcázar y del Chorro	933	50.96	92.49
10	Sebastián de Benalcázar y Avenida España	910	44.16	115.33
11	Calle Vieja y Elia Liut	672	34.56	64.83
12	Puente Fabian Alarcón	2530	137.41	249.60
	Promedio global		**68.65**	**131.68**

Nota. Resultados obtenidos en entorno sintético controlado; la magnitud del error varía con el volumen vehicular de cada intersección.

La heterogeneidad en el MAE entre intersecciones reportada con valores de 50 a 160 vehículos/15 min en evaluaciones previas refleja la diferencia de escala en los volúmenes vehiculares: intersecciones de alto flujo (~2.500 veh/15 min) presentan MAE absoluto mayor, aunque su en comparación al MAE al de intersecciones de bajo flujo (~65 veh/15 min). Esta variabilidad es esperada y metodológicamente coherente con el dominio del problema.

Es importante contextualizar que los resultados reportados corresponden a un entorno de validación sintético controlado, en el que los datos de evaluación comparten el mismo proceso generador que los datos de entrenamiento. El MAE promedio de 68.65 veh/15 min establece una línea base de consistencia interna del modelo bajo los supuestos de simulación definidos. La validación con datos reales multi-mensuales objetivo de la siguiente fase del proyecto permitirá estimar la capacidad de generalización real del sistema.

Interfaz de usuario

La interacción con el sistema se implementó mediante una aplicación web en Streamlit desplegada llamada ruteacuenca. La interfaz permite seleccionar una intersección piloto y visualizar la predicción del flujo vehicular a 15 minutos mediante el botón 'Calcular predicción', que ejecuta: (a) consulta a PostgreSQL para recuperar la ventana temporal de entrada (lookback=12) mediante la función obtener_datos_prediccion; (b) lectura de noticias recientes desde MongoDB para la intersección seleccionada; (c) clasificación de severidad PLN; y (d) inferencia del modelo LSTM con la ventana actualizada (Fig.3)

Recibido: 202-09-30 Aprobado tras revisión: 2026-10-19

Figura 3

Interfaz de usuario de RUTEA



Nota. Aplicativo RUTEA, presentación del resultado del monitoreo y predicción de tráfico en intersecciones claves.

IV. RESULTADOS Y DISCUSIÓN

Los resultados se han basado en los objetivos; la primera es referente a las fuentes de datos, integrando los aforos del GAD Municipal de Cuenca con un conjunto de 324 noticias viales etiquetadas, de las cuales se seleccionaron 43 intersecciones disponibles, con 6 puntos de piloto, con sus volúmenes vehiculares de su conectividad urbana y transporte público. Se almacenó en PostgreSQL para sus consultas respectivas en los registros estructurados de flujo vehicular. En cambio, MongoDB organizó las noticias con los campos de fecha, título, sector, resumen, fuente y URL. En el día de aforo en disponibilidad, se generaron series sintéticas para el periodo 2024/2025 en intervalos de 15 minutos; además, se incorporaron factores experimentales de severidad de incidentes y calendario. Esta integración ofreció una base de datos temporal para el desarrollo del modelo con la información textual al proceso del modelo de predicción.

La segunda es la clasificación de incidentes viales; se evaluó con 65 noticias la clasificación de PLN basado en TF-IDF y una red neuronal densa, dando una exactitud global aproximada del 89%. Sus categorías han sido congestión o embotellamiento leve, desvíos o cierres parciales y siniestros o accidentes graves. La clase de congestión leve obtuvo una precisión de 1,00, un recall de 0,75 y un F1-score de 0,86, con ocho instancias. La clase de desvíos o cierres parciales alcanzó un F1-score de 0,89,

Recibido: 202-09-30 Aprobado tras revisión: 2026-10-19

con 23 instancias, mientras que los siniestros o accidentes graves registraron un F1-score de 0,90, con 34 instancias. Estos resultados muestran capacidad de diferenciación entre las categorías evaluadas, aunque el reducido número de ejemplos de congestión leve limita la interpretación de sus métricas. La categoría predicha se incorporó como señal de ajuste en la entrada del componente de predicción.

En la tercera es la predicción del flujo vehicular. El modelo LSTM utilizó una entrada de dimensión (12, 4), correspondiente a una ventana temporal de tres horas con cuatro variables por intervalo, para estimar el flujo vehicular a 15 minutos. En las seis intersecciones piloto, el promedio aritmético del MAE fue de 68,65 vehículos por intervalo, mientras que el promedio de los RMSE fue de 131,68 vehículos. El menor MAE absoluto se registró en Calle Vieja y Av. Elia Liut, con 34,56 vehículos, y el mayor en el Puente Fabián Alarcón, con 137,41 vehículos. Estas diferencias deben interpretarse considerando la escala del flujo de cada intersección: un error absoluto mayor no implica necesariamente un peor desempeño proporcional.

Finalmente, dentro de la implementación de la interfaz, fue interactiva con una evaluación de no usuario, subida en STREAMLIT, denominada Ruteacuenca, permitiendo seleccionar una intersección piloto y ejecutar la predicción mediante el botón «Calcular predicción». El proceso recupera la ventana temporal desde PostgreSQL, consulta las noticias recientes almacenadas en MongoDB, clasifica el incidente y ejecuta la inferencia LSTM con la entrada actualizada. La implementación permitió disponer de un flujo integrado de consulta, clasificación y predicción en una misma aplicación. RUTEA constituye así un prototipo funcional y una base metodológica para continuar su validación con datos vehiculares reales de varios meses; la evaluación de usabilidad y del impacto sobre la gestión del tráfico permanece pendiente.

Discusión

Los resultados de RUTEA se enmarcan en la línea de sistemas híbridos para predicción de tráfico en entornos con disponibilidad limitada de datos. Zheng et al. (2017) y Wei et al. (2019) demostraron la superioridad de LSTM en contextos con cientos de estaciones de sensores y datos históricos continuos. RUTEA extiende este paradigma hacia ciudades intermedias mediante un mecanismo de simulación calibrado que permite operar con un único día de aforo, diferenciador metodológico respecto a la literatura existente. La integración de PLN como señal de ajuste dinámico en lugar de un módulo aislado de detección de incidentes sigue la propuesta de Melhem et al. (2024) y la implementa con recursos computacionales accesibles, demostrando viabilidad en contextos con restricciones de infraestructura.

Recibido: 202-09-30 Aprobado tras revisión: 2026-10-19

Las limitaciones del sistema son explícitas y metodológicamente declaradas. Los factores CAF del HCM 2016 fueron desarrollados para el contexto norteamericano; su aplicación en Cuenca es una aproximación proporcional ante la ausencia de factores locales calibrados, consistente con la evidencia de Carpio y Avilés (2018) que documenta diferencias del 58% en la sobrestimación de capacidad. La validación sobre datos sintéticos implica que el modelo aprende los patrones definidos por las reglas de simulación, lo que limita la interpretación del MAE como métrica de generalización real. Ambas limitaciones están explícitamente señalizadas en el diseño del sistema y constituyen la agenda de trabajo futuro. Las líneas de trabajo futuro incluyen: (1) reentrenamiento del modelo LSTM con datos reales multimensuales del GAD de Cuenca una vez disponibles; (2) integración de APIs de cámaras de tránsito para alimentar el sistema con vision de computadora(usando algoritmos de reconocimiento de accidentes/congestionamiento vehicular); (3) calibración local de los factores de reducción de capacidad mediante estudios de aforo continuo; (4) extensión del sistema a otras ciudades intermedias del Ecuador y Latinoamérica con características similares; y (5) desarrollo de la funcionalidad de rutas alternativas planificada como Fase 2 del proyecto

V. CONCLUSIONES

RUTEA demuestra que es técnicamente viable construir sistemas de predicción de tráfico vehicular en ciudades intermedias latinoamericanas sin disponer de series históricas continuas ni infraestructura de sensórica dedicada. El esquema de simulación basado en factores CAF del HCM 2016 y ajustes de calendario escolar local resuelve la limitación de datos de forma reproducible, documentada y recalibrable, estableciendo un precedente metodológico replicable para contextos similares en la región. La arquitectura híbrida PostgreSQL/MongoDB, combinada con el pipeline PLN para clasificación de severidad de incidentes, demuestra que la integración de fuentes de datos heterogéneas,series temporales estructuradas y texto no estructurado es abordable con herramientas de código abierto y despliegue accesible. La inyección de la señal de severidad PLN en la ventana de entrada del modelo LSTM constituye un mecanismo pragmático y efectivo para incorporar información contextual de eventos no recurrentes en la predicción de tráfico. Los resultados en entorno sintético controlado, MAE promedio de 68.65 veh/15 min para el modelo LSTM y accuracy global de 89% para el clasificador PLN establecen una línea base cuantificable para la calibración progresiva del sistema con datos reales. RUTEA sienta las bases tecnológicas y metodológicas para su evolución hacia un sistema de gestión predictiva del tráfico en tiempo real alineado con los objetivos de movilidad urbana inteligente.

Recibido: 202-09-30 Aprobado tras revisión: 2026-10-19

Para trabajos futuros, se recomienda que se amplíe el número de intersecciones y con más fechas en la base de datos para la identificación de futuros puntos y así crear una recomendación de rutas para una mejor movilidad.

REFERENCIAS BIBLIOGRÁFICAS

Achanta, S. (2024). Hybrid NoSQL and SQL architecture for realtime data processing in traffic systems. Journal of Data Engineering & Applications, 12(3), 45–63.

Aguagallo, J. y Becerra, M. (2023). Análisis de la movilidad y saturación vial en Cuenca. Editorial Universitaria.

Carpio, F. y Avilés, J. (2018). Aplicación del Highway Capacity Manual en intersecciones de Cuenca, Ecuador. Revista de Ingeniería Civil, 5(1), 23–34.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. y Wirth, R. (2000). CRISPDM 1.0: Stepbystep data mining guide. SPSS Inc.

Reyes, G., Roldán, J., Macias, A., Cordova, F., & León, O. (2022). Identificación de patrones de congestionamiento vehicular utilizando algoritmos de agrupamiento de trayectorias basados en densidad. *Investigación, Tecnología e Innovación, 14*(17), 1–15. https://doi.org/10.53591/iti.v14i17.1531

Chauora, J., Srivastava, S. y Bhatt, N. (2024). Hybrid LSTMPSO model for traffic flow prediction at urban intersections. Expert Systems with Applications, 238, 121952.

Goodfellow, I., Bengio, Y. y Courville, A. (2016). Deep learning. MIT Press. https://www.deeplearningbook.org/

HCM. (2016). Highway Capacity Manual (6th ed.). Transportation Research Board.

Melhem, A., Chen, X. y Wang, Y. (2024). NLPbased traffic incident detection from textual sources: A threestage framework. Transportation Research Part C, 158, 104412.

Openvia. (2025, 1 de agosto). *The major challenges of traffic management up to 2030*. https://www.openvia.io/the-major-challenges-of-traffic-management-up-to-2030/

Orellana, A. (2015). Cuenca, ciudad intermedia y sus desafíos de movilidad urbana. FLACSO Ecuador.

PADECO. (2013). Plan Maestro de Movilidad para el Cantón Cuenca. GAD Municipal de Cuenca.

Parra, R. (2017). Efecto fin de semana en la calidad del aire del Centro Histórico de Cuenca, 2013–2015. Revista de Salud Ambiental Ecuador, 4(1), 12–21.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830. https://jmlr.org/papers/v12/pedregosa11a.html

Wei, W. y Chen, L. (2019). Shortterm traffic flow prediction based on LSTM with enhanced feature extraction. IEEE Transactions on Intelligent Transportation Systems, 21(8), 3303–3312.

Zheng, H., Lin, F., Feng, X. y Chen, Y. (2017). A hybrid deep learning model with attentionbased convLSTM networks for shortterm traffic flow prediction. IEEE Transactions on Intelligent Transportation Systems, 22(11), 6910–6920.

Recibido: 202-09-30 Aprobado tras revisión: 2026-10-19