



UNIVERSIDAD DE CONCEPCIÓN

Dirección de Postgrado

Facultad de Ingeniería – Programa de Magister en Ciencias de la Ingeniería con Mención
en Ingeniería Civil

**IDENTIFICACIÓN DE FACTORES DE RIESGO DE COLISIÓN DE VEHÍCULOS
EN INTERSECCIONES URBANAS SEÑALIZADAS MEDIANTE
HERRAMIENTAS DE APRENDIZAJE AUTOMÁTICO**

Tesis para optar al grado de Magister en Ciencias de la Ingeniería con mención en
Ingeniería Civil

HERNÁN ANDRÉS PEÑA PALTA
CONCEPCIÓN -CHILE
2024

Profesor Guía: Dr. Tomás Echaveguren Navarro
Dpto. de Ingeniería Civil, Facultad de Ingeniería
Universidad de Concepción

RESUMEN

El Informe sobre la situación mundial de la seguridad vial 2023 señala que en 2021 murieron 1,19 millones de personas en accidentes de tráfico, de los cuales el 40% se produjeron en vías urbanas. Las colisiones mortales en intersecciones de calles representan entre el 25% y el 40%. Los modelos disponibles correlacionan los índices o la gravedad de las colisiones con factores como la geometría de la vía, las características del conductor, el tráfico y los atributos ambientales de cada lugar de colisión. El efecto de las características de la intersección y del entorno se trata por separado. En la bibliografía predominan los análisis basados en el tipo de intersección, el flujo de vehículos, el número de carriles disponibles y la presencia de una mediana. Se han estudiado poco los factores del uso del suelo y las instalaciones urbanas cerca de las intersecciones, así como la interacción entre los factores operativos, de infraestructura y del entorno construido.

La tesis identifica los componentes significativos de los factores de riesgo de colisión en intersecciones urbanas señalizadas mediante un proceso sistemático basado en herramientas de aprendizaje automático. A partir del estado del arte, se especifican 13 factores de riesgo relacionados con la infraestructura, 19 con los atributos operativos de las intersecciones y 4 con el entorno construido. Se utilizaron datos de la ciudad de Santa María de Los Ángeles. En esta ciudad, entre 2009 y 2019, el 60% de las colisiones de tráfico tuvieron lugar en intersecciones controladas por señales. Murieron 27 personas y 1053 resultaron gravemente heridas. La base de datos preparada contiene datos continuos, ordinales y nominales que describen en 128 intersecciones los 36 factores de riesgo identificados. La interacción entre la infraestructura, el entorno operativo y construido y 24 factores de riesgo se identificó mediante vectores de apoyo (SVM), k-nearest Neighbors (KNN), Descend Gradient (XGBoost) y Random Forest (RF). La distancia de visibilidad, los carriles de las calles principales y secundarias, el ángulo de la intersección, el número de pasos de peatones, el número de paradas de autobús y el uso del suelo explican el 60% de las colisiones.

ABSTRACT

The Global Status Report on Road Safety 2023 states that 1.19 million people were killed in road traffic crashes in 2021, 40% of which occurred on urban roads. Fatal crashes at street intersections accounted for 25% to 40%. Available models correlate crash rates or crash severity with factors such as roadway geometry, driver characteristics, traffic, and environmental attributes of each crash location. The effect of intersection and environmental characteristics is treated separately. The literature is dominated by analyses based on intersection type, vehicle flow, number of available lanes, and the presence of a median. Little study has been made of land use factors and urban facilities near intersections, as well as the interaction between operational, infrastructure, and built environment factors.

The thesis identifies the significant components of collision risk factors at urban signalized intersections through a systematic process based on machine learning tools. From the state of the art, 13 risk factors related to infrastructure, 19 to intersection operational attributes and 4 to the built environment are specified. Data from the city of Santa Maria de Los Angeles were used. In this city, between 2009 and 2019, 60% of traffic crashes occurred at signal-controlled intersections. Twenty-seven people were killed and 1053 were seriously injured. The prepared database contains continuous, ordinal and nominal data describing at 128 intersections the 36 identified risk factors. The interaction between infrastructure, operational and built environment and 24 risk factors was identified by support vectors (SVM), k-nearest Neighbors (KNN), Descend Gradient (XGBoost) and Random Forest (RF). Sight distance, primary and secondary street lanes, intersection angle, number of pedestrian crossings, number of bus stops, and land use explain 60% of collisions.

AGRADECIMIENTOS

En primer lugar, me gustaría agradecer al profesor patrocinante, Tomás Echaveguren, por guiarme en este estudio, siempre ayudarme cuando lo necesité y sus buenas conversaciones formales e informales.

Además, me gustaría agradecer a todas las personas de diferentes instituciones que me permitieron con sus aportes la recopilación de antecedentes de esta memoria, César Ibacache y Camila Galdames de SECTRA, Antonio Michela y Cecilia Aldana de la UOCT, Victoria Barrera del SIGLA de la Municipalidad de Los Ángeles y Pablo Muñoz Ingeniero Civil.

También muchas gracias a mi familia, Andrea Flores y Millaray Peña por permitirme la oportunidad de cumplir este anhelo y ayudarme incondicionalmente en este proceso.

ÍNDICE

1	INTRODUCCIÓN	9
1.1	Motivación.....	9
1.2	Hipótesis de investigación.....	11
1.3	Objetivo general	11
1.4	Objetivos específicos	11
1.5	Metodología de trabajo	11
1.6	Principales resultados y conclusiones	12
1.7	Organización del informe.....	13
2	FACTORES DE RIESGO DE ACCIDENTES EN INTERSECCIONES URBANAS	14
2.1	Introducción.....	14
2.2	Técnicas de modelación de accidentes.....	14
2.3	Herramientas de Machine Learning para modelación de accidentes.....	16
2.3.1	Aprendizaje No Supervisado.	17
2.3.2	Aprendizaje Supervisado.	20
2.4	Factores de Riesgo de Infraestructura.....	22
2.5	Factores de Riesgo de Operación	24
2.6	Factores de Riesgo de Entorno	26
2.7	Bases de Datos Utilizadas en Diferentes Investigaciones de Factores Contribuyentes a Accidentes en Intersecciones Urbanas.....	28
2.8	Conclusiones.....	30
3	CONFIGURACION DE LA BASE DE DATOS PARA EL ANÁLISIS DE FACTORES DE RIESGO DE ACCIDENTES	31
3.1	Introducción.....	31
3.2	Datos de Accidentes.....	32
3.3	Variables independientes	32

3.3.1	Datos de Factores de Riesgo de Infraestructura.....	32
3.3.2	Datos de Factores de Riesgo Operacional	33
3.3.3	Datos del Factor de Riesgo del Entorno	35
3.4	Conclusiones.....	36
4	DESCRIPCIÓN DE LA BASES DE DATOS DE INTERSECCIONES.....	37
4.1	Introducción.....	37
4.2	Análisis Exploratorio de Datos.....	37
4.3	Reducción dimensional mediante el análisis de componentes principales	43
4.4	Análisis de Datos	44
4.5	Conclusiones.....	47
5	CONCLUSIONES	49
6	REFERENCIAS	52

ÍNDICE DE FIGURAS

Figura 1 Técnicas de modelación de accidentes de tránsito.....	14
Figura 3 Algoritmos del ML No Supervisadas asociadas al Clustering.	17
Figura 4 Ejemplo de Algoritmos del ML No Supervisadas.	18
Figura 5 Ejemplo de Algoritmos de aprendizaje Supervisadas.	20
Figura 6. Estructura de la Base de Datos31.....	31
Figura 7. Histograma de Frecuencia de Accidentes por Periodo	38
Figura 8. Histograma de Variables que Describen el Factor de Riesgo de Infraestructura ..	38
Figura 9. PDF Empíricas de Variables que Describen el Factor de Riesgo de Operación ...	40
Figura 10. PDF Empíricas de Variables que Describen el Factor de Riesgo del Entorno Construido	42
Figura 11: Precisión media de disminución de los factores de riesgo del experimento	46
Figura 12: Incremento de la Pureza de los Factores de Riesgo en los nodos del experimento (en %).....	46

ÍNDICE DE TABLAS

Tabla 1. Componentes del Factor de Riesgo de Infraestructura.....	23
Tabla 2. Componentes del Factor de Riesgo de Operación.	24
Tabla 3. Componentes del Factor de Riesgo de Entorno.	27
Tabla 4. Base de datos utilizadas en Diferentes Investigaciones.	29
Tabla 5. Características del Factor de Riesgo de Infraestructura	32
Tabla 6. Características del Factor de Riesgo Operativo	33
Tabla 7. Características del Factor de Riesgo del Entorno Construido	35
Tabla 8: Resumen de los factores de riesgo identificados	43
Tabla 9: Experimentos de ajuste del factor de riesgo y del índice de colisiones en las intersecciones.....	45

1 INTRODUCCIÓN

1.1 Motivación

En el Informe sobre la situación mundial de la seguridad vial 2023 se afirma que en 2021 murieron 1,19 millones de personas en accidentes de tráfico, de los cuales, por término medio, el 60% se produjeron en vías rurales y el 40% en vías urbanas (OMS, 2023; ITF, 2023). Los accidentes mortales en las intersecciones de las calles representan entre el 25% y el 40% de los accidentes de tráfico mortales (Abdel-Aty y Keller, 2005; Anjana y Andjaneyulu, 2015; Roshandeh et al., 2016). La mayoría de los países están de acuerdo con la segunda década de acción, 2020 - 2030, promovida por las Naciones Unidas. El objetivo es reducir las muertes por colisión en un 50 %. Paralelamente, varios países han puesto en marcha políticas de seguridad vial, como Visión Cero en Suecia y seguridad sostenible en los Países Bajos, que aportan lecciones para el marco del sistema seguro (Fahlquist, 2006; Weijermars y Wegan, 2011; Elvik, 2023). Un aspecto clave del marco de sistema seguro es la responsabilidad del efecto de las medidas de seguridad para reducir las colisiones mortales en términos de las relaciones entre conductores, infraestructuras, vehículos y el entorno natural y construido (Islam et al., 2020). Significa, en primer lugar, comprender los factores causales de las colisiones mortales y, a continuación, aplicar contramedidas rentables para reducir las colisiones mortales.

Históricamente, la investigación se ha centrado en comprender el factor que explica la frecuencia y gravedad de las colisiones utilizando herramientas como regresiones, modelos logit/probit, bayesianos, logísticos y de eventos discretos (Barbosa et al., 2020; Islam et al., 2020; Jamal et al., 2021; Ali et al., 2024) denominados por Lord y Mannering (2010) modelos tradicionales. Estos modelos permitían calibrar las relaciones causales entre las tasas de colisiones o la gravedad de las mismas y factores causales como la geometría de la carretera, las características de los conductores, el tráfico y los atributos ambientales presentes en cada lugar de colisión. Sin embargo, los resultados obtenidos con los modelos tradicionales dependen en gran medida de la calidad e integridad de la base de datos y, simultáneamente, del cumplimiento de los supuestos de cada modelo estadístico (Barbosa et al., 2020; Jamal et al., 2021). Ali et al. (2024) introducen el problema del desequilibrio de datos en las bases de datos combinado con la diversidad de funciones de densidad probabilística de las variables explicativas, lo que limitaba la aplicación de los modelos tradicionales. Además, en

muchos casos, la disponibilidad de datos y los costes limitan la selección adecuada de las variables explicativas. El gran número de variables hace que sea complejo analizar cómo contribuye cada una por separado o combinada a explicar los accidentes, debido a la necesidad de muestras de gran tamaño, la posibilidad de multicolinealidad entre variables y la dificultad de identificar interacciones y asociaciones entre variables.

La aparición de herramientas de aprendizaje automático (ML) permitieron a los investigadores superar algunas desventajas de los modelos tradicionales (Jamal et al., 2021). Las herramientas de ML se han utilizado principalmente para correlacionar los factores explicativos de la frecuencia y la gravedad de los accidentes, pero también para resolver el problema de la clasificación binaria y multiclase de los accidentes en términos de gravedad, utilizando algoritmos de clasificación e híbridos (Barbosa et al., 2020; Assi, 2020; Al-Misarehi et al., 2022). Los algoritmos más típicos reportados por Ali et al. (2024) son árboles de decisión combinados con técnicas de ensemble (CART, Random Forest, AdaBoost, XgBoost, entre otros), redes neuronales artificiales (redes neuronales normales, convolucionales y recurrentes) y otros algoritmos como, máquinas vectoriales soportadas, Naïve Bayes, vecindades más cercanas y redes bayesianas. Algoritmos similares de ML correlacionan los factores humanos, del vehículo y del entorno vial con la frecuencia de los accidentes. Sin embargo, la mayoría de las aplicaciones de las herramientas de ML se realizan en entornos rurales. Por ejemplo, Barbosa et al. (2020), entre 2000 y 2020, contabilizaron 26 trabajos 7 en los que se aplicaron herramientas de ML para explicar colisiones en entornos urbanos y tráfico ininterrumpido. En estos casos, las variables más utilizadas fueron atributos del conductor, tipo de colisión, volumen de tráfico, geometría de la carretera y normas de tráfico, desagregadas en características específicas de segmentos de carretera. Comparativamente, la investigación en intersecciones urbanas se centra principalmente en correlacionar la gravedad de la colisión, la frecuencia y el tipo de uso implicado con variables operativas y de infraestructura utilizando modelos tradicionales.

El principal desafío identificado y abordado en este artículo es la identificación de los componentes significativos de los factores de riesgo de colisión en intersecciones urbanas señalizadas a través de un proceso sistemático basado en herramientas de aprendizaje automático. Se utilizaron datos de la ciudad de Santa María de Los Ángeles (SMLA). En esta ciudad, entre 2009 y 2019, el 60% de las

colisiones de tráfico tuvieron lugar en intersecciones controladas por señales. En este período, 27 personas murieron y 1053 resultaron gravemente heridas.

1.2 Hipótesis de investigación

La interacción entre los factores causales asociados a la infraestructura, operación y entorno, explican los siniestros de tránsito en intersecciones urbanas-

1.3 Objetivo general

Proponer un modelo que explique la variación de accidentes de tránsito en intersecciones urbanas, en función de variables de Infraestructura, Operación y Entorno, utilizando herramientas asociadas a Inteligencia Computacional.

1.4 Objetivos específicos

- Caracterizar modelos existentes de predicción de accidentes en intersecciones urbanas y sus variables.
- Elaborar una base de datos de accidentes de tránsito y de las características físicas, operacionales y entorno de intersecciones en que existen accidentes de tránsito.
- Calibrar y validar un modelo de predicción de accidentes utilizando algoritmos del Machine Learning.
- Aplicar el modelo a un caso de estudio en la Ciudad de Los Ángeles.

1.5 Metodología de trabajo

La metodología de trabajo se dividió en cinco etapas clave para desarrollar un modelo de predicción de accidentes en intersecciones urbanas: la primera parte identifica los componentes de los factores de riesgo de accidentes en intersecciones urbanas señalizadas. En la segunda parte, se describen las bases de datos utilizadas en la literatura existente. La estructura y contenido de la base de datos configurada para el análisis se reportan en la tercera parte, explicando los procedimientos aplicados

para determinar los accidentes de tráfico y las características físicas, operativas y ambientales de las intersecciones analizadas. El análisis se reporta en la cuarta sección. Consiste en las estadísticas descriptivas de accidentes y factores de riesgo de accidentes, reducción dimensional, clasificación, regresión y predicción. Finalmente, la discusión de los resultados y las conclusiones se incluyen en la quinta parte.

1.6 Principales resultados y conclusiones

Se identificaron 13 variables relacionadas con la infraestructura, 19 variables de operación de la intersección y 4 variables relacionadas con el entorno. Se sintetizó una base de datos integrando la información de accidentes de la comuna de Los Ángeles entre los años 2011 y 2019 (considerando los ocurridos en intersecciones urbanas) con la base de datos de las características de las intersecciones, la cual fue elaborada con mediciones en terreno y estudios de base que aportaron con la información requerida.

Al analizar individualmente la relación entre el número de accidentes y las variables, aquellas que mejor se relacionaron con la ocurrencia de accidentes fueron las de operación, de las cuales 16 resultaron significativas, entre ellas el flujo vehicular, la visibilidad y el número de fases de los semáforos. En segundo lugar, las variables relacionadas con la infraestructura, tales como el número de pistas, la presencia de mediana, el ángulo de la intersección, la presencia de pistas exclusivas de giro y el número de calles unidireccionales.

Finalmente, aquellas variables relacionadas con el entorno de la intersección no generan efectos significativos en el número de accidentes, siendo el número de paraderos la única variable significativa. Con estos resultados, se obtuvieron diversos modelos, tanto de regresiones lineales como no lineales. En este sentido, un modelo no lineal de forma exponencial es el que mejores ajustes presenta para la estimación de los accidentes, resultando como la variable con mayor peso estadístico e importancia, el flujo vehicular y en segundo plano, características físicas y operativas de las intersecciones, tales como, el número de pistas, el ángulo de la intersección, la densidad peatonal, visibilidad, número de paraderos y número de cruces peatonales.

Otro resultado importante fue el nivel de ajustes de los modelos predictivos en base a las características de las intersecciones, los cuales alcanzaron una capacidad explicativa de 50% (R^2 ajustado). Esto quiere decir que el número de accidentes es posible relacionarlo con factores relacionados a la infraestructura, la operación y el entorno de las intersecciones. No obstante, existe una importante porción del fenómeno que no resulta del todo explicada por estas características y, por consiguiente, son otros aspectos y causalidades que también influyen en el número de accidentes, estas se encuentran fuera del alcance de esta investigación. Las cuales podrían ser características inherentes de los conductores, de los vehículos, así como condiciones aleatorias del entorno.

1.7 Organización del informe

Esta Tesis consta de cinco capítulos. En el capítulo dos se identifican los factores de riesgo de accidentes en intersecciones urbanas, a su vez, se hace una revisión de las herramientas de machine learning existentes. En el capítulo tres se definen las variables a utilizar, dependientes e independientes, la elaboración y el proceso de síntesis de la base de datos. Luego en el capítulo cuatro, se procede a utilizar herramientas de machine learning, de tal modo de analizar patrones de comportamiento. Finalmente, en el capítulo seis se presentan las conclusiones de esta investigación.

2 FACTORES DE RIESGO DE ACCIDENTES EN INTERSECCIONES URBANAS

2.1 Introducción

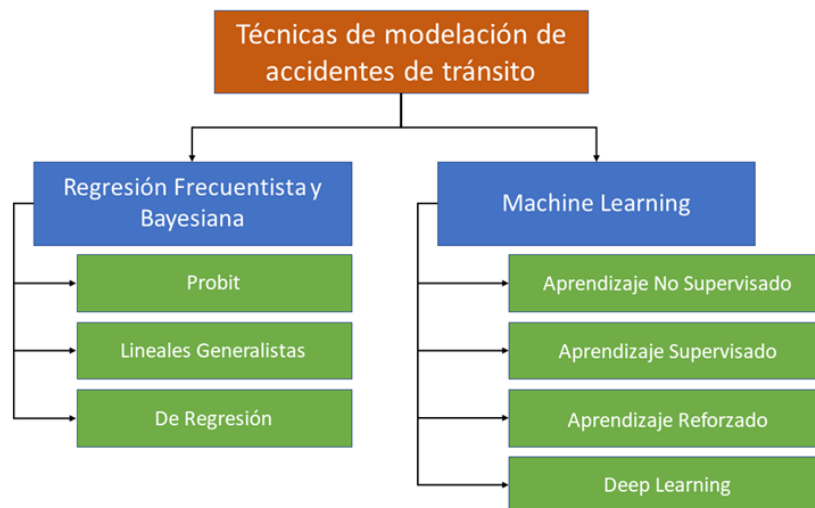
El conocimiento de los elementos del espacio vial urbano que rodea las intersecciones y su relación con los índices de colisión es un paso previo para aplicar contramedidas que reduzcan el número y la gravedad de las colisiones. Estos elementos se denominan factores de riesgo de colisión (FRC).

Dufeu (2018) identificó 40 FRC utilizados por diferentes autores, clasificados en variables de infraestructura, operacionales y de entorno construido resumidos en las Tablas 1, 2 y 3 (Greibe, 2003; Kumara y Chin, 2003; Salifu, 2004; Abdel-Aty y Keller, 2004; Mitra y Washington, 2007, Vieira Gomes et al., 2012, Vieira Gomes, 2013; Cunto y otros, 2015; Tasic y Porter, 2016, Chen y Xie, 2016; Xu y otros, 2017; Islam y otros, 2020; Merlin y otros, 2020).

2.2 Técnicas de modelación de accidentes.

Para la predicción de accidentes de tránsito existen dos grupos de técnicas de modelación: La Regresión Frecuentista y Bayesiana y las herramientas Machine Learning. Las cuales se detallan en la figura 1.

Figura 1 Técnicas de modelación de accidentes de tránsito.



Las técnicas de modelación de accidentes de tránsito más utilizadas son las regresiones frecuentistas y bayesianas (Úbeda, 2017). Las que se dividen principalmente en 3 tipos:

- Modelos Probit (Greibe, 2003; Kumara y Chin, 2003; Abdel-Aty y Keller, 2004),
- Modelos Lineales Generalizados (GLM) (Mitra y Washington, 2007; Vieira Gomes, 2013; Chen y Xie, 2016); y
- Modelos de Regresión (Cunto et. al., 2014; Tasic y Porter, 2016; Dufeu, 2019).

Sin embargo, presentan limitaciones, debido a que cada modelo requiere sus supuestos y contienen relaciones predefinidas de la variable dependiente con las independientes. También, requieren asumir cierta distribución de los datos, definir a priori una distribución de los errores, la posibilidad de presentar multicolinealidad y la dificultad para tratar con outliers y datos faltantes (Úbeda, 2017).

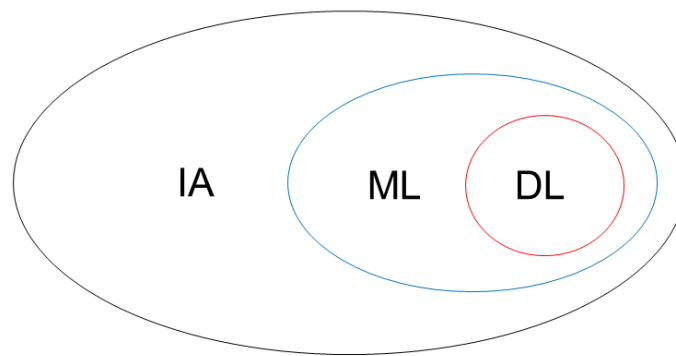
Las anteriores premisas pueden llevar a errores en las estimaciones o a inferencias incorrectas. Es en ese sentido, la modelación con Machine Learning se abre a una gran oportunidad, ya que estos funcionan bajo el mecanismo de “caja negra” en los cuales se desconoce qué es lo ocurre y como se infieren los resultados (Martínez, 2020).

El grupo de técnicas de modelación de accidentes de tránsito basadas en ML, se han utilizado para identificar zonas con altas tasas de accidentes o áreas potencialmente riesgosas, mediante análisis multivariado (Gutiérrez-Osorio, et al 2019), K vecino más cercano (KNN) (Iranitalab y Khattak, 2017; Zhang, et al. 2018), Máquina de soporte Vectorial (SVM) (Montt, et al. 2011), Deep Learning (DL) (Essien, et al. 2019), redes neuronales artificiales (ANN) (Zeng, et al. 2016; Vélez, 2018) y Redes Neuronales Bayesianas (BNN) (Molina, 2019).

2.3 Herramientas de Machine Learning para modelación de accidentes.

La Inteligencia Artificial son la combinación de algoritmos planteados con el propósito de crear máquinas que presenten las mismas capacidades que el ser humano. Un subconjunto de la IA es el ML, que se define como un método analítico que permite a un sistema a través de un aprendizaje detectar patrones, tendencias y relaciones en los datos de un estudio, para ser utilizarlos para la predicción de datos futuros o para realizar otro tipo de decisiones bajo cierta incertidumbre (Sancho, 2020). Por último, el Deep Learning (DL), es un subgrupo del ML que se define como un algoritmo automático estructurado o jerárquico, el cual trata de emular el aprendizaje humano, con el fin de obtener algún tipo de conocimiento. La anterior relación queda expuesta en la Figura 1.2.

Figura 2 Relación entre Inteligencia Artificial (IA), Machine Learning (ML) y Deep Learning (DL).



El Machine Learning (ML) es una herramienta que se centra en como la Inteligencia Artificial (IA) aprende y utiliza datos, mediante algoritmos que posibilitan el aprendizaje (Fernández y Guzmán del Río, 2020).

Los algoritmos de Machine Learning (AML) se dividen en cuatro grupos: Aprendizaje Supervisado, No Supervisado, Reforzado y Aprendizaje Profundo (Zapata, 2021).

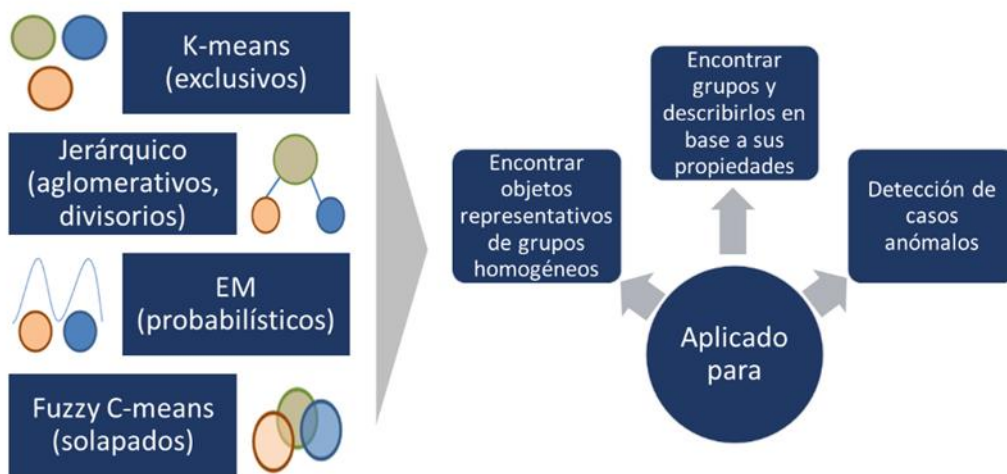
2.3.1 Aprendizaje No Supervisado.

El Aprendizaje No Supervisado (ANS) no requiere que los datos a utilizar tengan un etiquetado previo y son algoritmos destinados a determinar relaciones de similitud, diferencia o asociación (Quituisaca-Samaniego y Álvarez, 2017).

Los algoritmos de ANS trabajan con datos de estructura desconocida. Con estos se puede explorar la estructura principal de los datos y extraer información significativa, sin la utilización de una variable de resultado conocida.

Una de las técnicas de aprendizaje no supervisado más utilizadas son los algoritmos de Agrupación en Conglomerado (Clustering). En estos se agrupan los valores de entrada según determinados criterios de clasificación, pero sin tener un conocimiento previo de los elementos del grupo. La clasificación puede basarse en distancias, centroides, modelos paramétricos, entre otros criterios de semejanza. Algunos de los algoritmos utilizados son K-means, Random Forest, Fuzzy C-means y Dbscan, entre otros.

Figura 3 Algoritmos del ML No Supervisadas asociadas al Clustering.



La reducción dimensional es otra de las aplicaciones del ANS. El objetivo de este tipo de algoritmo es reducir la cantidad de dimensiones que explican un conjunto de datos. La reducción dimensional permite reducir los recursos necesarios para utilizar estos datos por un algoritmo supervisado o depurar los datos a utilizar (Quituisaca-Samaniego y Álvarez, 2017).

Hay tres métodos para realizar reducción de dimensionalidad:

1. Selección de características: De toda la muestra se selecciona un subconjunto, el cual, según algunos criterios, logra una representación del conjunto de datos.
2. Transformación Lineal: Se realiza un análisis de los componentes principales de los datos de un espacio de la muestra, de manera que la varianza de estos se maximice. El espacio original se reduce con la pérdida de datos, pero se espera que conserve la varianza más importante al espacio abarcado.
3. Transformación No Lineal: Se utilizan cuando los datos no se encuentran en un espacio lineal. Se fundamenta en la hipótesis que en una base de datos con una estructura de alta dimensión la información más relevante se centra en un pequeño grupo de datos de baja dimensión.

La Figura 4 muestra un ejemplo de clasificación de algoritmos de Aprendizaje No Supervisado.

Figura 4 Ejemplo de Algoritmos del ML No Supervisadas.



Depaire, et al (2008), analizó si los clusters pueden ser utilizados como una técnica de clasificación de los accidentes de tránsito. En este estudio se seleccionaron los clusters de clase lateral y ampliada, con estos se logró encontrar varios clusters en una base de datos de accidentes, determinando que algunos parámetros están asociados con una mayor gravedad de los accidentes, pero esta influencia no es siempre la misma y depende del tipo de vehículo involucrado. Como conclusión del estudio se obtuvo que el tipo de choque tiene un rol importante cuando se examina la gravedad del accidente.

Beshah (2010), centra su investigación en la relación entre la severidad del accidente y determinados factores de la vía mediante el uso de algoritmo de clustering PART. Como conclusión del estudio se obtuvo que la gravedad de los accidentes varió con diferentes combinaciones de factores relacionados a la carretera. Las variables que se escogieron para el análisis fueron: subciudad, área particular, separación de carretera, orientación de carretera, cruce de carretera, tipo de superficie de la carretera, condición de resaca en la carretera, condiciones climáticas, condiciones de iluminación y severidad de accidentes.

Vigneswaeen, et al (2014), basaron su investigación en Naive Bayes y Árboles de Decisión. Con estos desarrollaron modelos de predicción para clasificar de forma automática la severidad de los accidentes de tránsito ocurridos en Hong Kong en el 2008 e investigar el desempeño para predecir la precisión de la clasificación. Los resultados de esta investigación proponen los árboles de decisión superan al algoritmo de clasificación Naive Bayes en lugar de seleccionar todos los atributos para la clasificación.

Kashyap (2016), aplicó técnicas de minería de datos para vincular las características de la vía con la severidad de los accidentes. El objetivo del trabajo fue detectar las causas de los accidentes y como reducirlas. El clasificador utilizado fue Naive Bayes alcanzando una precisión de un 89%. El análisis se centró en las contribuciones de factores relacionados con la vía, las condiciones meteorológicas o anomalías de los vehículos.

2.3.2 Aprendizaje Supervisado.

Los algoritmos de Aprendizaje Supervisado se usan con datos etiquetados, intentado encontrar una función que, dadas los datos de entrada, les asigna la etiqueta de salida adecuada. El algoritmo se entrena con un registro histórico de datos y así aprende a asignar la etiqueta de salida adecuada a un nuevo valor, es decir, predice el valor de salida (Simeone, 2018).

Dentro de las categorías de algoritmos de aprendizaje supervisado se encuentran los de regresión y de clasificación.

Los algoritmos de regresión se usan con bases de datos etiquetadas, que permiten hacer predicciones sobre resultados (Cestero, 2019). El modelador entrega al AML una base de datos depurada y etiquetada que incluye la información de entrada y salida. Con este grupo de datos que se denominan datos de entrenamiento (calibración), se realiza el ajuste al modelo inicial. El algoritmo va “aprendiendo” a predecir las muestras de entrada comparando el resultado del modelo, realizando las compensaciones respectivas de acuerdo con cada error en la estimación del resultado.

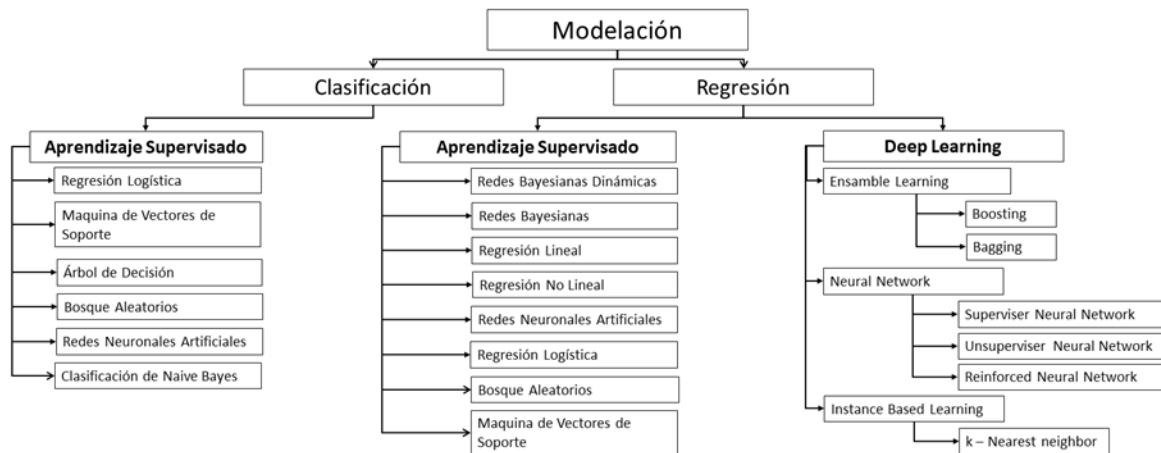
Los algoritmos de clasificación se usan cuando el resultado es una etiqueta discreta. Esto quiere decir, que se utilizan cuando la respuesta se fundamenta en conjunto finito de resultados (Sancho, 2020).

Existen dos procesos de clasificación:

- Clasificación binaria: El algoritmo de clasificación es capaz de asignarle la etiqueta correcta, entre 2 términos conocidos.
- Clasificación con probabilidades: El algoritmo asigna la clase con probabilidad más alta como resultado del proceso de clasificación.

La Figura 5 muestra una clasificación del algoritmo de aprendizaje supervisado.

Figura 5 Ejemplo de Algoritmos de aprendizaje Supervisadas.



Wu, et al (2009), analizó la severidad de los accidentes de un solo vehículo a través de un modelo basado en Random Forest, obteniendo una precisión de 73%. Los predictores utilizados fueron condiciones climáticas, límites de velocidad, iluminación, factores de colisión, sexo, edad, experiencia del conductor, tipo de vehículo y cinturón de seguridad.

Mohaymany, et al (2010), basó la investigación en la utilización de árboles de regresión y clasificación, obteniendo un modelo de estudio de severidad de los accidentes con una precisión de 72,5%, empleando una variable objetivo asociadas al conductor y 4 variables independientes asociadas al vehículo, la experiencia del conductor y sus condiciones legales para conducir. Sin embargo, consideró predictores a la infraestructura y su operación como son: el número de pistas, la existencia de berma, superficie de la vía y tipo de berma, causas del accidente, tipo de colisión, tipo de localización, iluminación y condiciones meteorológicas.

Ehsaei y Evdorides (2011), demostraron que es posible determinar la probabilidad que ocurra un accidente de tránsito con resultado de lesiones graves o fatales en condiciones específicas las que pueden ser reproducida usando Redes Bayesianas. Además, se demostró que es posible calcular la probabilidad de un accidente en diferentes puntos de una carretera, definiendo puntos de riesgos por accidentes a través de variables como detalles de las uniones o condiciones especiales. También, respecto de la alta probabilidad que ocurran una mayor cantidad de accidentes de tránsito durante el día que la noche. Adicionalmente, la probabilidad de accidentes mortales en tramos de la vía donde

no existe iluminación es el doble que la probabilidad de accidentes con heridos graves. Por último, se determinó que los accidentes en superficies húmedas disminuyeron durante un periodo de análisis de 5 años.

Elvik (2011), plantea la aplicación de criterios operacionales de causalidades a modelos estadísticos multivariantes; desarrollados para identificar fuentes de variaciones sistemáticas en los recuentos de accidentes, en particular los efectos de las variables que representan los procedimientos de seguridad.

Soler (2014), desarrolló un modelo para el estudio de accidentes de carreteras llamado Naive – Poisson. Este se basa en modelos de Redes Bayesianas para la estimación de la frecuencia en que suceden accidentes. De este estudio se obtuvo una metodología para que, a partir de cualquier conjunto de datos y una vez seleccionada la variable objetivo, se analicen los sucesos de baja probabilidad de ocurrencia.

2.4 Factores de Riesgo de Infraestructura

Los factores de riesgo de infraestructura son las variables que explican el diseño geométrico y los elementos de seguridad de la intersección (Vieira Gomes, 2013; Chen y Zhou, 2016). Los componentes de este factor de riesgo se resumen en la Tabla 1.

Tabla 1. Componentes del Factor de Riesgo de Infraestructura

Componentes	Greibe (2003)	Kumara and Chin (2003)	Salifu (2004)	Abdel-Aty and Keller (2004)	Mitra and Washington (2007)	Vieira Gomes <i>et al.</i> (2012)	Vieira Gomes (2013)	Cunto <i>et al.</i> (2015)	Tasic and Porter (2016)
% de vías expresa									•
Ancho de calle secundaria			•						
Ancho mediana		•	•		•		•		
Ancho de vereda					•				
Ancho de pistas							•		
Curvatura horizontal		•							
Longitud de pista de giro a la izquierda		•							
Longitud del cruce peatonal							•		
Presencia de mediana	•			•		•	•	•	
Facilidades para bicicletas	•								
Presencia de pistas exclusivas para giros a la derecha		•		•					
Presencia de pistas exclusivas para giros a la izquierda		•	•						
Ramas de la intersección	•						•		
Pistas exclusivas de viraje	•								
Pista exclusiva para giros a la izquierda				•					
Nº de pistas exclusivas para giros a la derecha				•				•	

Componentes	Greibe (2003)	Kumara and Chin (2003)	Salifu (2004)	Abdel-Aty and Keller (2004)	Mitra and Washington (2007)	Vieira Gomes <i>et al.</i> (2012)	Vieira Gomes (2013)	Cunto <i>et al.</i> (2015)	Tasic and Porter (2016)
Pistas en a la calle principal	•	•		•	•	•	•	•	
Pistas de aceleración		•							

Mitra y Washington (2007) determinaron que el número de pistas por rama reduce significativamente los accidentes. Kumara y Chin (2003) verificaron que esta variable es significativa solo en el tramo principal de la intersección. Abdel-Aty y Keller (2004) determinaron que la presencia de medianas en las vías principales y secundarias de la intersección reduce el número de accidentes más graves, aunque no afecta el número de accidentes menores. Vieira Gomes (2013) y Cunto et al. (2015) obtuvieron resultados similares. Salifu (2004), Abdel-Aty y Keller (2004), y Kumara y Chin (2003) establecieron que la presencia de pistas exclusivos para giros a la izquierda, así como giros canalizados a la derecha, reduce el número de accidentes. Greibe (2003) determinó que el número de accesos a la intersección impacta las tasas de accidentes en la intersección, mostrando que los accidentes aumentan si la configuración de la intersección es más compleja.

2.5 Factores de Riesgo de Operación

Los factores de riesgo operacionales describen el flujo de tráfico y las regulaciones de tráfico de la intersección, tales como el flujo de vehículos y peatones, la visibilidad, los semáforos y las señales de alto o ceda el paso (Wong et al., 2007). Los componentes de este factor de riesgo se resumen en la Tabla 2.

Tabla 2. Componentes del Factor de Riesgo de Operación.

Componentes	Greibe (2003)	Kumara and Chin (2003)	Salifu (2004)	Abdel-Aty and Keller (2004)	Mitra and Washington (2007)	Vieira Gomes <i>et al.</i> (2012)	Vieira Gomes (2013)	Cunto <i>et al.</i> (2014)	Tasic and Porter (2016)	Chen and Xie (2016)
Jerarquía de calles								•		
% de viajes peatonales									•	
% de viajes laborales en									•	

Componentes	Greibe (2003)	Kumara and Chin (2003)	Salifu (2004)	Abdel-Aty and Keller (2004)	Mitra and Washington (2007)	Vieira Gomes <i>et al.</i> (2012)	Vieira Gomes (2013)	Cunto <i>et al.</i> (2014)	Tasic and Porter (2016)	Chen and Xie (2016)
transporte público										
Km diarios/vehículo									•	
Flujo peatonal diario							•			
Giros a la derecha en las intersecciones						•	•			
Giros a la izquierda en las intersecciones						•	•			
Límite de velocidad			•	•						
N° de cruces peatonales							•			
N° de giros a la izquierda				•						
Presencia de cruce peatonal en cada tramo							•			
Relación de flujo entre la calle secundaria y la principal						•				•
Relación de flujo entre la calle secundaria y todas las calles de la intersección						•				
Tipo de control en la intersección (semáforo, Stop, Ceda el paso)	•		•		•	•	•	•		
Flujo de vehículos pesados			•							
Flujo en las calles principal y secundaria	•		•	•	•	•	•	•		•
Flujo de giros a la derecha		•								
Flujo de giros a la izquierda		•								

Componentes	Greibe (2003)	Kumara and Chin (2003)	Salifu (2004)	Abdel-Aty and Keller (2004)	Mitra and Washington (2007)	Vieira Gomes <i>et al.</i> (2012)	Vieira Gomes (2013)	Cunto <i>et al.</i> (2014)	Tasic and Porter (2016)	Chen and Xie (2016)
Flujo motorizado	•	•	•			•				
Flujo no motorizado	•									
Flujo opuesto			•							
Visibilidad		•						•		

Chen y Xie (2016) estudiaron la influencia del flujo vehicular en la tasa de accidentes utilizando dos modelos. Uno consideraba el flujo diario promedio en el tramo principal y el tramo secundario, y el otro el flujo total y la relación entre el flujo del tramo secundario y el del tramo principal en las intersecciones. En cuanto a la relación entre el flujo secundario y el principal en la intersección, establecieron que el número de accidentes aumentaba cuando el valor estaba entre -1.5 y -1. Por el contrario, a medida que el flujo del tramo secundario aumentaba en relación con el principal, el número de accidentes disminuía. Abdel-Aty y Keller (2004) y Salifu (2004) estudiaron el impacto de la velocidad operacional en la ocurrencia de accidentes, pero no encontraron efectos significativos. Salifu (2004) encontró que la variable de velocidad en el tramo principal no era estadísticamente significativa. Kumara y Chin (2003) concluyeron que la visibilidad es irrelevante como predictor de accidentes en intersecciones. Cunto et al. (2015) utilizaron la visibilidad como un factor de modificación de accidentes por la presencia de obstáculos, según AASHTO (2010).

Mitra y Washington (2007) y Vieira Gomes (2013) utilizaron el tipo de control en la intersección para discriminar patrones de comportamiento en intersecciones con y sin semáforos. Cunto et al. (2015) abordaron este tema estimando modelos independientes para intersecciones con y sin semáforos. Kumara y Chin (2003) determinaron que el número de fases de señalización es irrelevante para explicar los accidentes.

2.6 Factores de Riesgo de Entorno

Los factores de riesgo del entorno construido están compuestos por variables como la presencia de escuelas o centros de salud, edificios comerciales, paradas de transporte público o cualquier edificio, e instalaciones que atraigan el flujo peatonal hacia las intersecciones (Merlin et al., 2020). Los componentes de este factor de riesgo se resumen en la Tabla 3.

Tabla 3. Componentes del Factor de Riesgo de Entorno.

Componente	Greibe (2003)	Kumara and Chin (2003)	Salifu (2004)	Abdel-Aty and Keller (2004)	Mitra and Washington (2007)	Vieira Gomes <i>et al.</i> (2012)	Vieira Gomes (2013)	Cunto <i>et al.</i> (2015)	Tasic and Porter (2016)
Accidentes por comuna				•					
Densidad de población									•
Uso del suelo N°									• •
Estacionamiento de bicicletas									
N° Parada de autobús								•	•
N° de Centros de Educación								•	
N° de Centros de Salud								•	
N° de Locales de expendio de alcohol								•	
Presencia de estacionamiento								•	
Área total de veredas									•
Número de intersecciones por kilómetro de cale									•

Los factores de riesgo asociados al entorno construido están presentes en AASHTO (2010) y Tasic y Porter (2016); explican la ocurrencia de choques entre vehículos y peatones. Las principales variables del entorno construido se refieren al tipo de uso del suelo, la presencia de paradas de autobús, y escuelas o centros de salud. Tasic y Porter (2016) utilizaron el uso del suelo, la infraestructura de transporte multimodal, las paradas de autobús, los carriles para bicicletas, las rutas de transporte público y los carriles exclusivos para autobuses. Establecieron que solo las paradas de autobús eran significativas para explicar los accidentes de tráfico.

Como se muestra en las Tablas 1, 2 y 3, el factor de riesgo más analizado es la operación de las intersecciones, seguido por la infraestructura y el entorno construido. En cuanto a la operación, el componente dominante es el tipo de intersección y el flujo en las calles principales y secundarias. En cuanto a la infraestructura, los principales factores son la presencia y el ancho de las medianas y el número de carriles en las calles principales. El factor del entorno construido es el menos analizado, y solo la presencia de paradas de autobús se reconoce como un componente relevante que contribuye a los choques en las intersecciones.

Cada factor de riesgo de las Tablas 1, 2 y 3 es válido para el conjunto de datos específico (ver Tabla 4) utilizado por cada autor. Por lo tanto, los factores de riesgo significativos para explicar los accidentes en un estudio de investigación no necesariamente son significativos en otro. Esto lleva a evaluar todo el conjunto de factores de riesgo y discriminar estadísticamente cuáles son relevantes antes de modelar su efecto en los accidentes.

2.7 Bases de Datos Utilizadas en Diferentes Investigaciones de Factores Contribuyentes a Accidentes en Intersecciones Urbanas

Un aspecto relevante para caracterizar los factores de riesgo de accidentes es describir sus patrones de comportamiento. Se desarrolló una base de datos basada en esta caracterización. La confiabilidad de la base de datos está determinada por la calidad y precisión de los registros de accidentes, el tamaño de la muestra, la precisión de las variables que definen los factores de riesgo de accidentes y la precisión de la georreferenciación. La calidad y precisión de los registros de accidentes son características que no se pueden controlar, ya que dependen de los protocolos de recopilación de datos de encuestas de cada país para este tipo de información. Por lo general, la policía u otras agencias gubernamentales son típicamente responsables de la recopilación de datos de encuestas de accidentes y de los protocolos de recopilación de datos en cada país. En general, los análisis se realizan con datos tal como están y, en muchos casos, solo es posible filtrar los registros de accidentes eliminando errores o discrepancias referenciados. El tamaño de la muestra está relacionado con la calidad y cantidad de registros de accidentes. Por lo general, se consideran registros de accidentes de cinco años porque,

con períodos más largos, la infraestructura puede haber cambiado, y no siempre es posible confiar en datos de diseños geométricos anteriores (Castro et al., 2012).

La Tabla 4 resume las características generales de las bases de datos de accidentes utilizadas por diferentes autores para modelar el efecto de grupos de factores de riesgo de accidentes. Se observa que el tamaño de la muestra se obtiene a partir de la combinación del número de intersecciones, el número de accidentes en las intersecciones y el período de registros de accidentes utilizado. El número de intersecciones varía desde 44 (Vieira Gomes, 2012) hasta 832 (Abdel-Aty and Keller, 2004), y el número de accidentes varía desde 150 (Tasic and Porter, 2016) hasta 21,204 (Abdel-Aty and Keller, 2004). El período de registros de accidentes varía entre dos y nueve años, siendo más comúnmente de cuatro o cinco años. El tamaño de muestra óptimo dependerá del número de componentes de cada factor de riesgo de accidentes, la configuración del diseño factorial y el número y tipo de interacciones a modelar.

Tabla 4. Base de datos utilizadas en Diferentes Investigaciones.

Autores	Lugar	N° de Intersecciones	N° de Accidentes	Periodo de Tiempo de Registro de Accidentes
Greibe (2003)	Dinamarca	1036	3.343	1987 – 1991
Kumara and Chin (2003)	Singapur	104	809	1992 – 2000
Abdel-Aty and Keller (2005)	Florida, Estados Unidos	832	21.204	2000 – 2001
Salifu (2004)	Ghana	91	592	1996 – 1998
Mitra et al. (2002)	Singapur	52	832	1992 – 1999
Mitra and Washington (2007)	Georgia, Estados Unidos	165	837	1996 – 1997
Wang et al (2011)	Londres, Reino Unido	(**)	1310	2003 – 2007
Vieira Gomes et al. (2012)	Lisboa, Portugal	94	490	2004 – 2007
Zhang et al. (2012)	Houston, Estados Unidos	(**)	46.813	1996 – 2000
Vieira Gomes (2013)	Lisboa, Portugal	94	573	2004 – 2007
Anjana and Anjaneyulu (2015)	India	32	(*)	2009 – 2012
Cunto et al. (2015)	Fortaleza, Brasil	181	2756	2007 – 2009
Chen and Xie (2016)	Massachusetts, Estados Unidos	90	1587	2009 – 2012

Autores	Lugar	N° de Intersecciones	N° de Accidentes	Periodo de Tiempo de Registro de Accidentes
Tasic and Porter (2016)	Chicago, Estados Unidos	(**)	11.167	2005 – 2012
Islam et al. (2020)	Dhaka, Bangladesh	61	(*)	2008 – 2012
Gilani et al. (2021)	Rasht, Irán	(*)	738	2019 – 2020
Al-Mistarehi et al. (2022)	Zarqa, Jordania	(*)	97.900	2014 – 2018
Barman and Bandyopadhyaya (2023)	Patna, India	(*)	1.477	2015 – 2019

2.8 Conclusiones

Se identificaron tres clasificaciones de variables asociadas a las intersecciones; relacionadas a la infraestructura, la operación y el entorno.

De la literatura existente, se puede determinar un gran coeficiente asociado a las variables utilizadas, la impactará en su significancia estadística, dependiendo de la severidad del accidente.

Del análisis de los autores, se determina que, en promedio, utilizar datos de cinco años es lo recomendable. Esto debido a que se posee una base robusta de datos, pero sin generar sesgos debido a los problemas de utilizar muchos años de accidentes, ya que, al considerar muchos años, se obvian cambios que se pudiesen haber generado en las intersecciones.

3.2 Datos de Accidentes

Los datos de accidentes se obtuvieron de la base de datos de la policía chilena, incluyendo accidentes que ocurrieron en intersecciones entre 2011 y 2019, con víctimas fatales y lesiones graves. Los accidentes durante ese período fueron georreferenciados para identificar intersecciones señalizadas con registros de accidentes. Se utilizó un buffer de 50 metros para identificar accidentes en intersecciones, excluyendo accidentes entre intersecciones. Los registros de accidentes identificaron 128 intersecciones con al menos un accidente, de los cuales 20 correspondieron a intersecciones en forma de T y 108 eran intersecciones de cuatro vías. En las intersecciones en forma de T, se georreferenciaron 722 accidentes, y en las intersecciones de cuatro vías, 3.196.

3.3 Variables independientes

3.3.1 Datos de Factores de Riesgo de Infraestructura

Los datos de infraestructura se obtuvieron en el trabajo de campo. En comparación con la Tabla 1, se midieron nuevas variables: carriles exclusivos, número de brazos y carriles, medianas, cercas peatonales y ángulo de la intersección (reemplazando la curvatura horizontal). Las instalaciones para bicicletas y las autopistas no fueron detectadas, por lo que no se incluyeron en la base de datos. Los anchos de calles, medianas, aceras y carriles no se incluyeron en los datos porque están estandarizados en las directrices de diseño geométrico de calles urbanas chilenas y no varían entre intersecciones.

Tabla 5. Características del Factor de Riesgo de Infraestructura

ID	Variable	Valor	Descripción
IF1	Tipo de intersección	1	Tres Ramas
		2	Cuatro Ramas
IF2	Pista de giro a la derecha en las calles principales	0	Ninguno
		1	Uno o más pistas auxiliares
IF3	Pista de giro a la derecha en las calles secundarias	0	Ninguno
		1	Uno o más pistas auxiliares
IF4	Pista de giro a la izquierda en las calles principales	0	Ninguno
		1	Uno o más pistas auxiliares
IF5	Pista de giro a la izquierda en las calles secundarias	0	Ninguno
		1	Uno o más pistas auxiliares

IF6	Ramas de entrada a la intersección	1 – 4	Número de ramas por los cuales fluye el tráfico hacia la intersección
IF7	Pistas en las calles principales	2 – 8	Número total de pistas disponibles en las calles principales en ambas direcciones
IF8	Pistas en las calles secundarias	0 – 6	Número total de pistas disponibles en las calles secundarias en ambas direcciones
IF9	Mediana en las calles principales	0	Ninguna o, si existe, el ancho de la mediana es inferior a 1 m
		1	Existe y el ancho de la mediana es superior a 1 m
IF10	Mediana en las calles secundarias	0	Ninguna o, si existe, el ancho de la mediana es inferior a 1 m
		1	Existe y el ancho de la mediana es superior a 1 m
IF11	Vallas peatonales	0	Ninguna
		1	Al menos en una acera hay una valla peatonal
IF12	Ángulo de la intersección	0 – 90	Ángulo entre el eje de cada brazo (grados)
IF13	Calles unidireccionales	0 – 2	Número de calles con flujo unidireccional solamente

3.3.2 Datos de Factores de Riesgo Operacional

Los datos de operación se obtuvieron en el trabajo de campo y de estudios de ingeniería. En comparación con la Tabla 2, se incluyeron nuevas variables: giros a la izquierda y el flujo total se clasificaron por el tamaño de las calles. El flujo de tráfico se desagregó de manera más precisa que en la Tabla 2 para incluir el tráfico local. La densidad peatonal, las fases de señal y las distancias de visibilidad se incluyeron como nuevas variables. El porcentaje de viajes peatonales, de trabajo y la longitud de los viajes se descartaron debido a la falta de datos confiables. El tipo de control de la intersección no se consideró porque la investigación se centró en intersecciones señalizadas.

Tabla 6. Características del Factor de Riesgo Operativo

ID	Variable	Valor	Descripción
OF1	Giro a la izquierda en la Rama Principal	0	No existe
		1	Existe al menos un giro a izquierda
OF2	Giro a la izquierda en la Rama Secundaria	0	No existe
		1	Existe al menos un giro a izquierda
OF3	Densidad peatonal	1	Densidad peatonal baja en la intersección
		2	Densidad peatonal media en la intersección

ID	Variable	Valor	Descripción
		3	Densidad peatonal alta en la intersección
OF4	Numero de Cruces Peatonales	0 – 4	Numero de Cruces peatonales sin considerar las medianas
		0 – 12	Numero de Cruces peatonales considerar las medianas
OF5	Numero de Fases del semáforo	2 – 3	Numero de fases del semáforo en la intersección
OF6	Visibilidad	5 – 110	Distancia de visibilidad, según método MOP (2023)
		1	Nivel de visión malo, obstáculos importantes
		2	Nivel de visión regular, obstáculos menores
OF7	Visibilidad del conductor en la intersección	3	Nivel de visión bueno, obstáculos casi imperceptibles
		4	Nivel de visión excelente, sin obstáculos
		1.806,45	
OF8	Flujo total de la intersección	– 19.410,5	Flujo total de la intersección en la intersección (Veq/hr)
		5	
OF9	Flujo en la rama principal	829,55 – 13.999	Flujo total en la rama principal en la intersección (Veq/hr)
OF10	Flujo en la rama secundaria	506,75 – 9.347,7	Flujo total en la rama secundaria en la intersección (Veq/hr)
OF11	Flujo que gira a la izquierda	71,6 – 5.888,95	Flujo total que gira a la izquierda en la intersección (Veq/hr)
OF12	Flujo que gira a la derecha	0 – 5.888,95	Flujo total que gira a la derecha en la intersección (Veq/hr)
OF13	Flujo de oposición	854,3 – 13.261,7	Flujo total de oposición en la intersección (Veq/hr)
OF14	Flujo de vehículos livianos	1.258 – 14.030	Flujo total de vehículos livianos en la intersección (Veq/hr)
OF15	Flujo de taxi colectivos	0 – 4.727,7	Flujo total de taxi colectivos en la intersección (Veq/hr)
OF16	Flujo de Taxi buses	0 – 2.709,3	Flujo total de Taxi buses en la intersección (Veq/hr)
OF17	Flujo de Camiones	2,5 – 3.391,5	Flujo total de Camiones en la intersección (Veq/hr)
OF18	Flujo de buses inter urbano	0 – 1.120	Flujo toral de buses inter urbano en la intersección (Veq/hr)
OF19	Flujo no motorizado	0 – 308,4	Flujo total no motorizado en la intersección (Veq/hr)

La densidad peatonal se obtuvo en el sitio utilizando las categorías establecidas por el Ministerio de Vivienda y Urbanismo de Chile (MINVU, 2009). El número de cruces peatonales en cada intersección se verificó en el sitio. La visibilidad en la intersección se obtuvo mediante el cálculo en el sitio del triángulo de distancia mínima de visión basado en el procedimiento descrito por el Ministerio de Obras Públicas de Chile (MOP, 2023). Los flujos de tráfico de vehículos se obtuvieron a partir de mediciones realizadas por la Secretaría de Planificación del Transporte (SECTRA, 2014-2018). Los datos se actualizaron un año después del estudio, utilizando tasas de crecimiento por tipo de vehículo, que variaron desde 0% para el transporte público hasta 4.5% para camiones (SECTRA, 2016). Estos datos se utilizaron para calcular el tráfico en hora pico en la mañana (PHM), al mediodía (PHN), por la tarde (PHA), en horas no pico (OPH), durante el día (D) y en todo el día (AD).

3.3.3 Datos del Factor de Riesgo del Entorno

Los datos del entorno construido se recopilaron en el trabajo de campo. La Tabla 3 muestra que los accidentes por país, la densidad de población, el estacionamiento de bicicletas, las tiendas de alcohol, la presencia de estacionamiento, el área total de aceras y el número de intersecciones por kilómetro no se incluyeron en la Tabla 7.

Tabla 7. Características del Factor de Riesgo del Entorno Construido

ID	Variable	Valor	Descripción
BEF1	Tipo de Zona	1	Zona principalmente del tipo residencial
		2	Zona principalmente del tipo comercial
		3	Zona principalmente del tipo mixto (residencial and comercial)
		4	Zona principalmente del tipo industrial
BEF2	Numero de Centro de Salud	0 – 1	Número de centros de salud en radio de 130 m en torno al centro de la intersección
BEF3			
BEF4	Numero de Paradas de Transporte Público	0 – 3	Número de paraderos en radio de 50 m en torno al centro de la intersección
BEF5	Numero de Centro de Educación	0 – 3	Número de centros de educación en radio de 130 m en torno al centro de la intersección

El uso del suelo en los alrededores de la intersección se obtuvo a partir de la zonificación de uso del suelo de la ciudad de SMLA en su Plan Maestro de Uso de Suelo (Municipalidad de Los Ángeles, 2011). Otras variables se obtuvieron directamente en el sitio, apoyadas por imágenes satelitales y trabajo de campo

3.4 Conclusiones

Se elaboraron seis bases, la que incorpora variables de entorno, operación e infraestructura, además de accidentes de todas las intersecciones semaforizadas, lo anterior para los periodos Punta Mañana, Punta Medio Dia, Punta Tarde, Fuera de Punta, Diurno y Todo el Dia, además de considerar las variables asociadas a infraestructura, entorno y operación, estos datos obtenidos de distintos estudios, Plan Regulador Comunal e información recopilada de terreno. Esto permitió generar una caracterización de las 128 intersecciones semaforizadas consideradas dentro del estudio.

Para lograr obtener una base de datos estructuradas y cumplir con lo propuesto en el diseño experimental, se debió depurar los datos entregados por Carabineros de Chile, esto ya que existen diversos tipos de nomenclaturas para la intersección. Respecto de los flujos se hace presente que se cuentan registros para las 128 intersecciones, las cuales se actualizaron en base a los factores obtenidos en los estudios de SECTRA.

4 DESCRIPCIÓN DE LA BASES DE DATOS DE INTERSECCIONES

4.1 Introducción

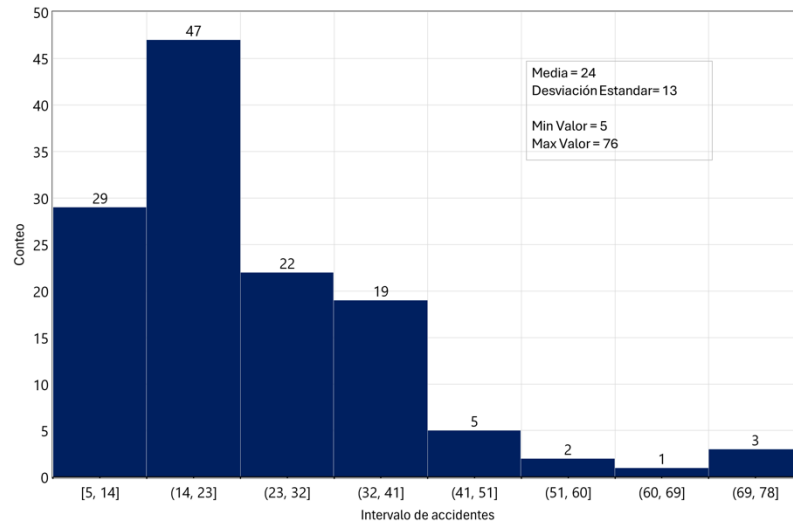
El efecto de los factores de riesgo de colisión se considera en cuatro pasos. El primer paso fue un análisis exploratorio del conjunto de datos, que incluía la media, la desviación típica, el intervalo y las funciones de densidad de probabilidad (FDP). En el segundo paso, se explora la reducción dimensional mediante el análisis de componentes principales (ACP). Teniendo en cuenta los resultados obtenidos con el APC, aplicamos algoritmos de aprendizaje automático de clasificación en el tercer paso. Finalmente, en el último paso, utilizamos modelos de regresión de aprendizaje automático para establecer predicciones de colisiones basadas en las variables explicativas.

en función de las variables explicativas.

4.2 Análisis Exploratorio de Datos

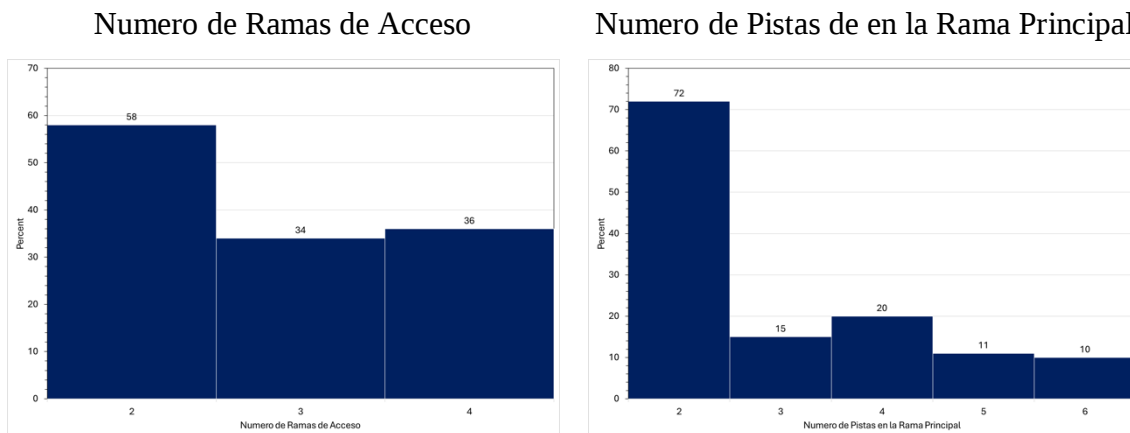
La Figura 7 muestra las funciones de densidad de probabilidad empíricas (PDF) de los accidentes para los periodos de mañana, mediodía, tarde, fuera de las horas pico, durante el día y en todas las horas pico. Los histogramas muestran una alta presencia de datos cercanos a cero en los periodos de horas pico. En otros momentos, el número de ceros disminuye sustancialmente en comparación con otros periodos. Asimismo, la forma de las distribuciones de frecuencia es asimétrica, con una sola cola explicada por la presencia de valores extremos de accidentes. Por lo tanto, la forma de la PDF empírica es similar a una PDF de Poisson o Binomial Negativa, lo cual es consistente con la evidencia del estado del arte.

Figura 7. Histograma de Frecuencia de Accidentes por Periodo

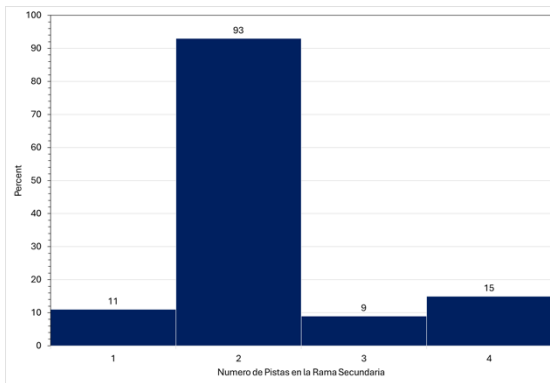


Las Figuras 8, 9 y 10 muestran las PDF empíricas de las variables asociadas con los factores de riesgo de infraestructura, operación y medio ambiente.

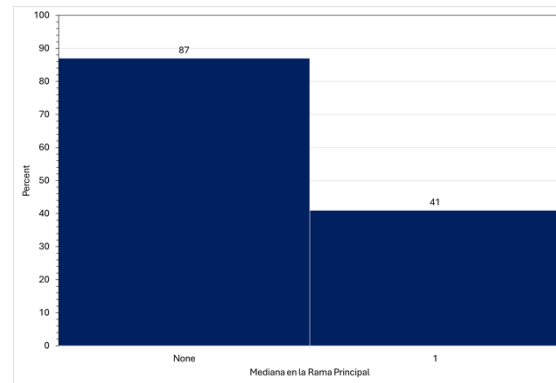
Figura 8. Histograma de Variables que Describen el Factor de Riesgo de Infraestructura



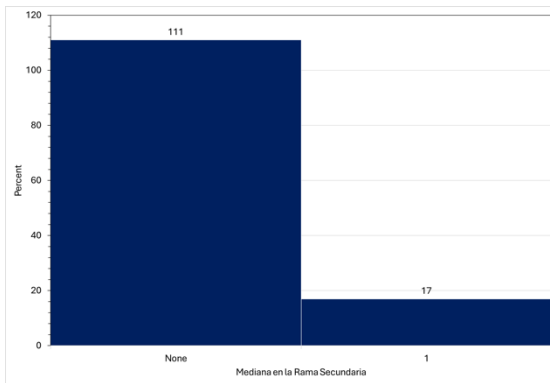
Numero de Pistas de en la Rama
Secundaria



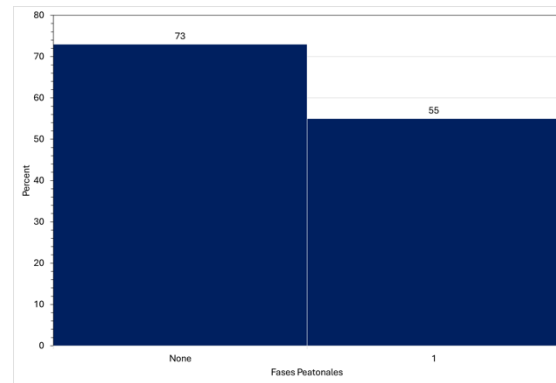
Mediana en la Rama Principal



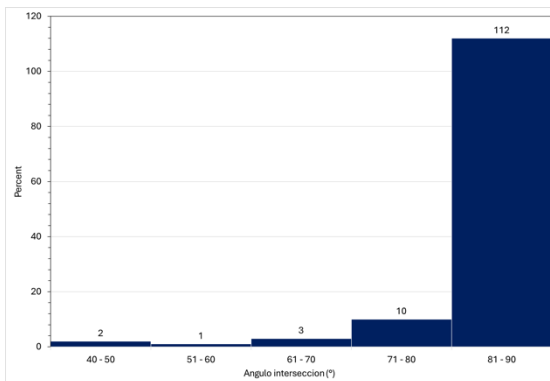
Mediana en la Rama Secundaria



Fases Peatonales



Angulo de la Intersección (°)

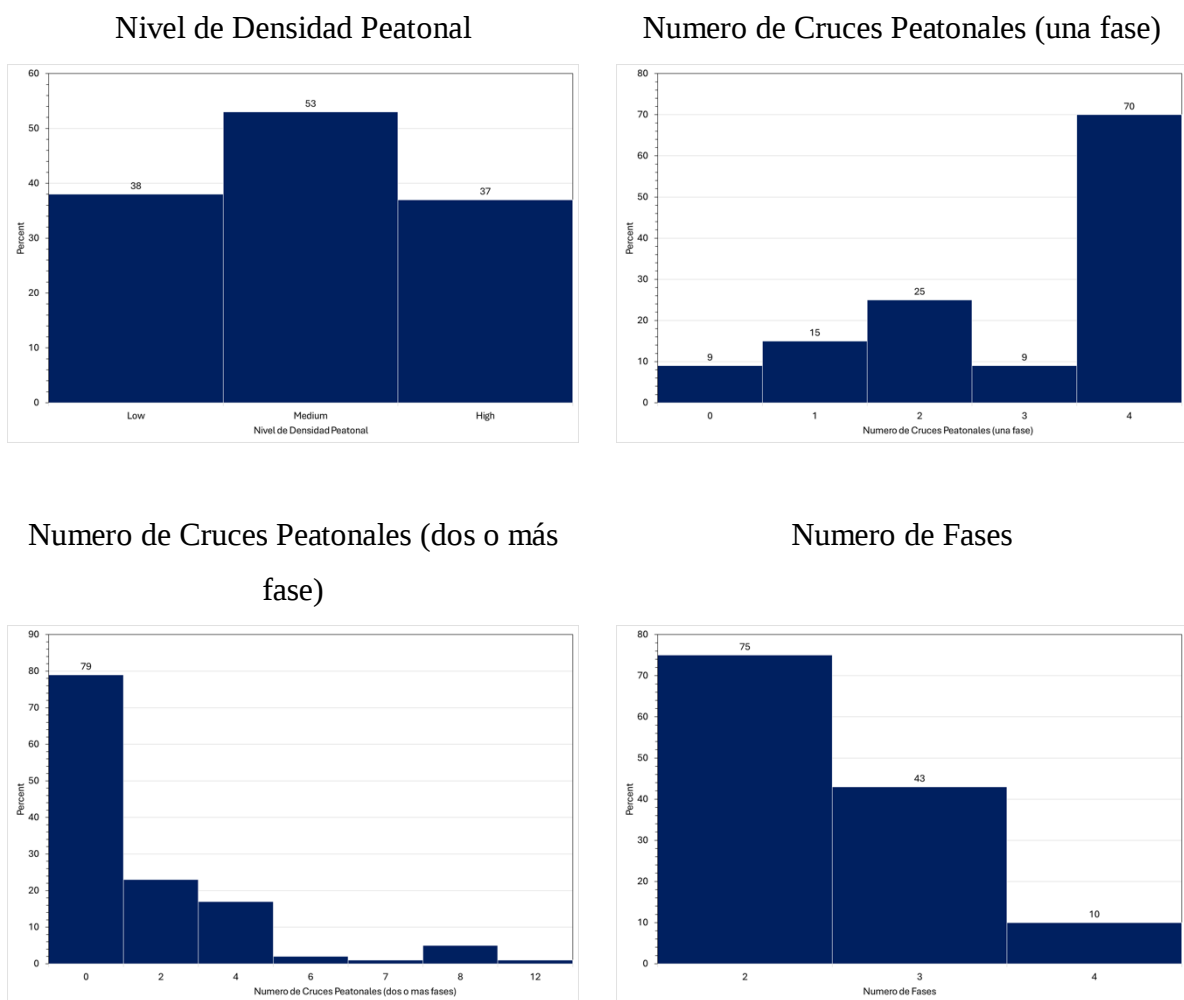


La Figura 8 muestra que el 27% de las intersecciones tienen tres brazos de aproximación, el 45% tienen dos y el 28% tienen cuatro. El 6% de las intersecciones tienen solo un brazo de aproximación. El 56% de la muestra tiene seis carriles en la calle principal. El 12% de las intersecciones tienen dos

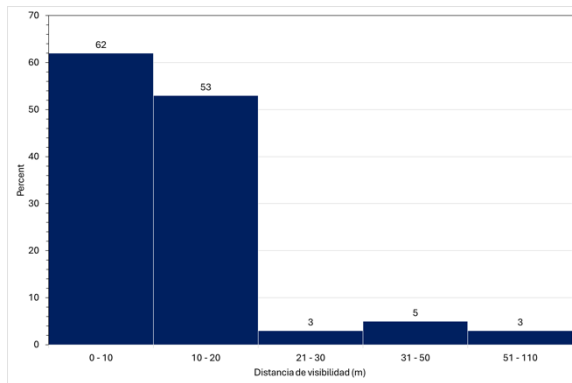
carriles en la calle principal, el 32% tienen cuatro o más carriles. El 11% de las intersecciones tienen dos carriles en sus calles secundarias, y el 73% tienen tres carriles. El 95% de las intersecciones tienen una configuración perpendicular, y el 5% tienen un ángulo inferior a 90°.

Estos datos muestran un predominio de intersecciones de cuatro brazos y en T, donde las calles principales y secundarias convergen homogéneamente. La distribución homogénea de carriles en las calles principales es evidencia de que, en la mayoría de las intersecciones con accidentes graves, falta gradualidad en la interconexión de las calles principales y secundarias.

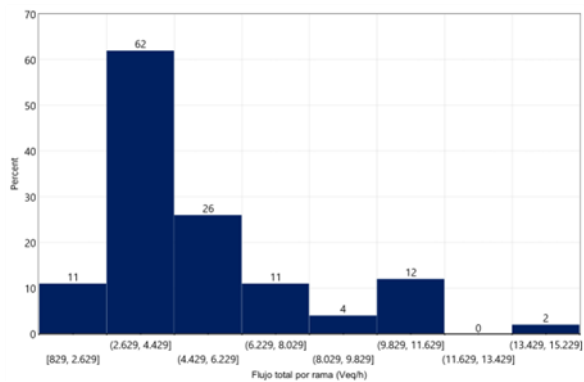
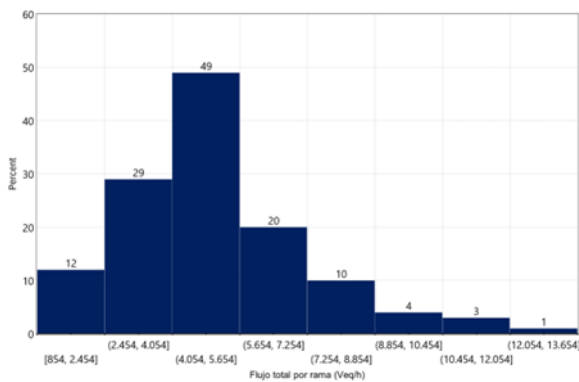
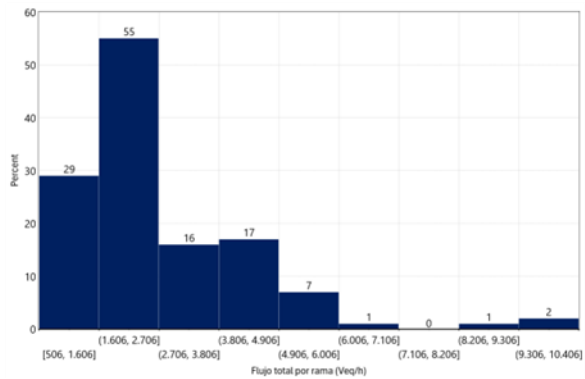
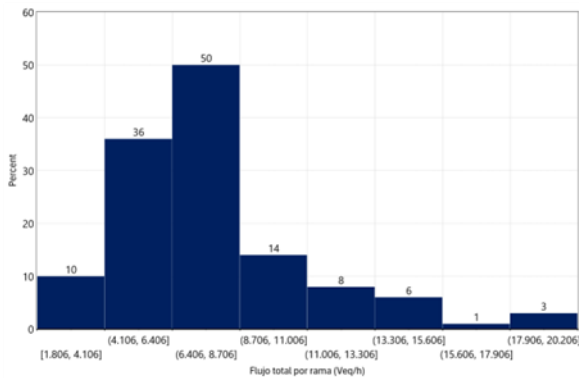
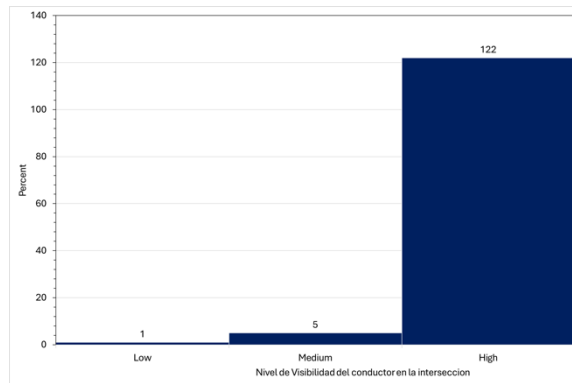
Figura 9. PDF Empíricas de Variables que Describen el Factor de Riesgo de Operación



Distancia de Visibilidad (m)



Nivel de Visibilidad del Conductor en la Intersección

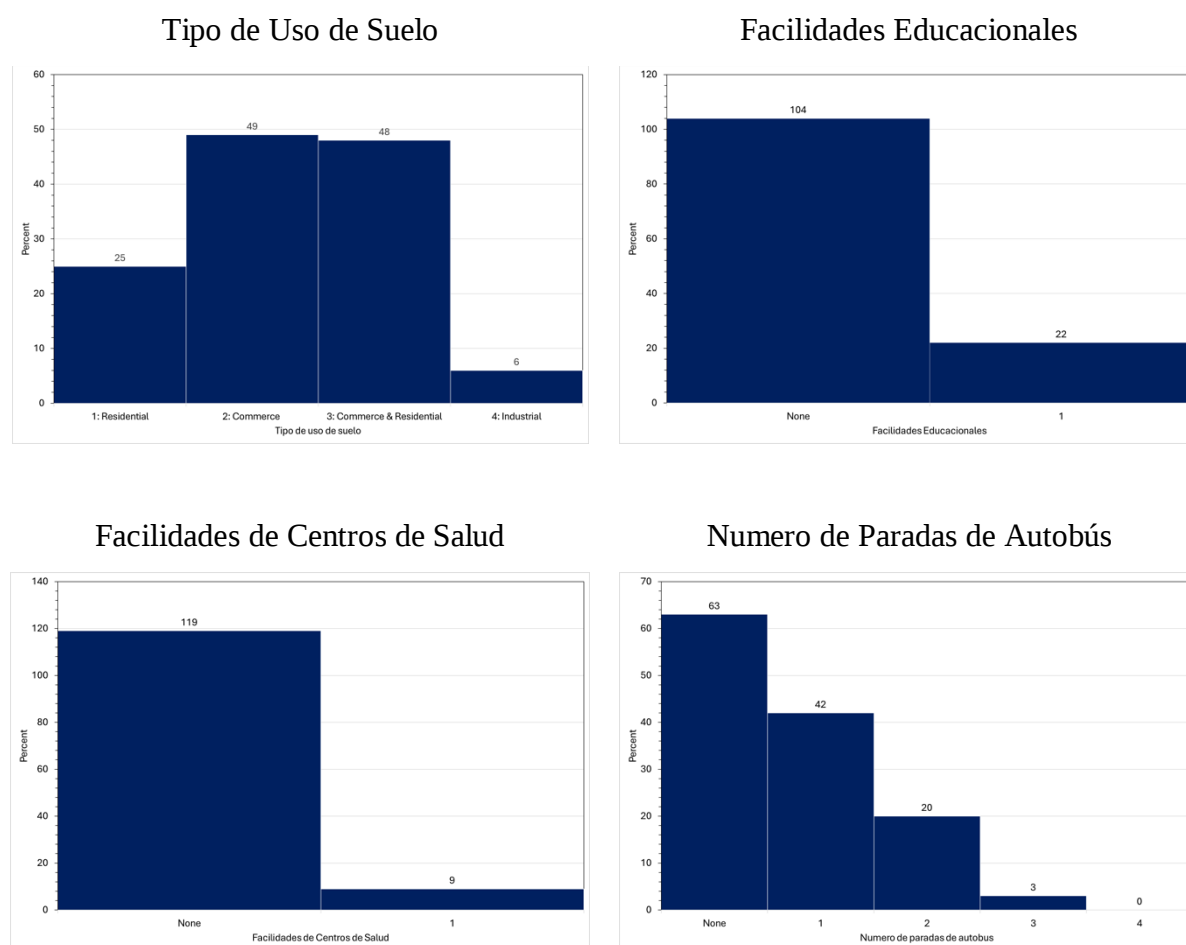


La Figura 9 muestra evidencia de que el 30% de las intersecciones tienen una baja densidad peatonal, el 41% una densidad media y el 29% una alta densidad. La mayor proporción de intersecciones (59%) tiene señales de dos fases. El 95% de las intersecciones no tienen obstáculos que impidan al conductor ver la intersección. El 81% de las intersecciones tienen una distancia mínima de visibilidad de menos

de 20 m. El 63% de las intersecciones tienen cuatro cruces peatonales de una etapa, y el 99% tienen al menos dos. En cuanto a los cruces peatonales de dos o más etapas, el 20% tiene cuatro cruces, y el 40% tiene al menos dos cruces peatonales.

Las intersecciones predominantes son aquellas con baja densidad peatonal, buena visibilidad, cruces peatonales de una y dos etapas y señales de dos fases. Por lo tanto, es necesario explorar la relación entre la visibilidad y la circulación peatonal como factor de riesgo de accidentes.

Figura 10. PDF Empíricas de Variables que Describen el Factor de Riesgo del Entorno Construido



La Figura 10 muestra que las intersecciones están distribuidas homogéneamente en tres tipos de uso del suelo. El 93% de las intersecciones están lejos de los centros de salud, y el 81% están lejos de los edificios educativos. El 19% de las intersecciones están a menos de 130 metros de las escuelas. El 49% de las intersecciones no tienen paradas de autobús a menos de 50 metros, y el 51% restante tiene

una o más paradas de autobús a menos de 50 metros. Se observa que la mayoría de las intersecciones están lejos de los centros de salud y las escuelas, y el uso del suelo y la presencia de paradas de autobús están distribuidos homogéneamente. En consecuencia, estos cuatro componentes del factor ambiental podrían no contribuir significativamente a la ocurrencia de accidentes. La Tabla 8 muestra estadísticas descriptivas de los datos de accidentes y de las variables continuas de las Tablas 5, 6 y 7.

Tabla 8: Resumen de los factores de riesgo identificados

Factor de Riesgo	Componente	Factor de Riesgo	Componente
Operacional	Flujo total de la Intersección	Operacional	Paso Peatonal en dos Etapas
	Flujo Rama Principal		Giro a la izquierda utilizando una pista exclusiva en la Rama Principal
	Flujo Rama Secundaria		Densidad Peatonal
	Flujo que gira a la Derecha		Distancia de Visibilidad
	Flujo Vehículos Livianos		Fases de la Semaforización
	Flujo Vehículos Pesados		Número de calle unidireccionales que llegan a la intersección
	Flujo Buses Interurbanos		Angulo de la Intersección
	Flujo Taxi Buses		Numero de Pistas de la Rama Principal
	Flujo de No Motorizados		Numero de Pistas de la Rama Secundaria
	Relación de Flujo entre la Rama Principal y la Rama Secundaria		Pista Exclusiva de giro a la Izquierda en la Rama Principal
	Relación de Flujo de la Rama Secundaria y el Flujo Total de la Intersección		Mediana en la Rama Principal

4.3 Reducción dimensional mediante el análisis de componentes principales

El PCA se utilizó para reducir el número de variables explicativas. La idoneidad del PCA se comprobó mediante la prueba de Kaiser-Meyer-Olkin (KMO) (Dziuban y Shirley, 1974), la prueba de esfericidad de Bartlett (Bartlett, 1954) y la prueba de Mardia (Mardia, 1970). El valor KMO global obtenido fue de 0,6. El valor p de la prueba de Bartlett fue inferior a 0,05, lo que significa que se viola la esfericidad. El valor p de la prueba de Mardia fue inferior a 0,05, lo que significa que se rechaza la hipótesis nula sobre distribuciones normales multivariantes. Este último resultado fue visualizado previamente en el análisis exploratorio de datos, donde varias variables tienen histogramas como una

distribución de probabilidad normal. En conclusión, el resultado de la aplicación del PCA no fue concluyente, y no se utilizó la reducción dimensional.

4.4 Análisis de Datos

La base de datos contiene distintos tipos de datos: variables continuas, como las colisiones y el flujo de tráfico; variables binarias, como la presencia/ausencia de medianas; datos de recuento que codifican atributos ordinales, como el uso del suelo; y variables discretas, como el número de tramos en cada intersección. Por lo tanto, estas variables pueden clasificarse como escalares, ordinales o nominales, condicionando el método utilizado para el análisis de los datos. Debido a ello, utilizamos modelos de regresión de aprendizaje automático que admiten todas estas variables para determinar el comportamiento y la importancia relativa de la variable independiente, en lugar de construir un modelo predictivo.

La estrategia de modelización fue la siguiente: a) Análisis de los factores de riesgo agrupados por factores de infraestructura, operativos y de uso del suelo; b) Análisis de las interacciones entre estas variables; y c) Análisis integrado de todas las variables. Estos análisis pretendían determinar la importancia de cada factor de riesgo en las colisiones en intersecciones urbanas. Teniendo en cuenta esta estrategia, se configuraron 15 experimentos para probar el efecto relativo de los factores de riesgo de infraestructura, operacionales y del entorno construido, y sus interacciones en los índices de colisiones. En cada experimento se probaron la máquina de vectores de apoyo (SVM), k-nearest Neighbors (KNN), Descend Gradient (XGBoost) y Random Forest (RF). Tras un análisis exploratorio, se seleccionó el algoritmo RF para obtener las ponderaciones de los factores de riesgo.

La Tabla 9 muestra los resultados obtenidos en cada experimento. Los experimentos 1 a 5 muestran el efecto de cada grupo de factores de riesgo sobre los índices de colisión, los experimentos 6 a 12 representan interacciones por pares entre grupos de factores de riesgo, y los experimentos 13 a 15 representan la interacción entre los tres grupos de factores de riesgo. El experimento 16 representa la interacción entre los tres grupos de factores de riesgo, incluyendo las variables significativas identificadas en el experimento anterior. Principalmente, el grupo operativo se separó en dos subconjuntos. El primer subconjunto considera los factores de riesgo OF1 a OF7 de la Tabla 6, y el segundo los OF8 a OF19. El primer subconjunto incluía las propiedades operativas de cada intersección, y el segundo incluía el flujo de tráfico en cada intersección.

Tabla 9: Experimentos de ajuste del factor de riesgo y del índice de colisiones en las intersecciones

Experimento	Conjunto de datos sobre factores de riesgo	MAPE (%)	R ²
1	Infraestructura	38.5	0.23
2	Entorno construido	43.3	0.25
3	Operacional (OF1 a OF7)	43.7	0.14
4	Operacional (OF8 a OF19)	41.1	0.31
5	Operacional (todos los OF)	67.3	0.00
6	Infraestructura y operacional (OF1 a OF7)	42.2	0.05
7	Infraestructura y operacional (OF8 a OF19)	49.5	0.02
8	Infraestructura y operacional (todos los OF)	56.4	0.08
9	Infraestructura y entorno construido	38.0	0.16
10	Entorno construido y operacional (OF1 a OF7)	60.1	0.00
11	Entorno construido y operacional (OF8 a OF19)	49.3	0.27
12	Entorno construido y operacional (todos los OF)	53.5	0.03
13	Entorno construido, infraestructuras y operacional (OF1 a OF7)	58.3	0.46
14	Entorno construido, infraestructuras y operacional (OF8 a OF19)	41.5	0.00
15	Entorno construido, infraestructuras y operacional (todos los OF)	44.1	0.11

Los resultados muestran que el porcentaje medio de error absoluto (MAPE) varía entre el 38 y el 44 % si los factores de riesgo operativo se dividen en dos grupos, el efecto operativo. Por el contrario, todos los factores de riesgo operativos no afectan a la tasa de accidentes. Los efectos cruzados de los grupos de factores de riesgo muestran que el impacto más significativo es la interacción entre el entorno construido, la infraestructura y los factores operativos considerados como la intersección de las condiciones operativas (Experimento 13).

Figura 11: Precisión media de disminución de los factores de riesgo del experimento

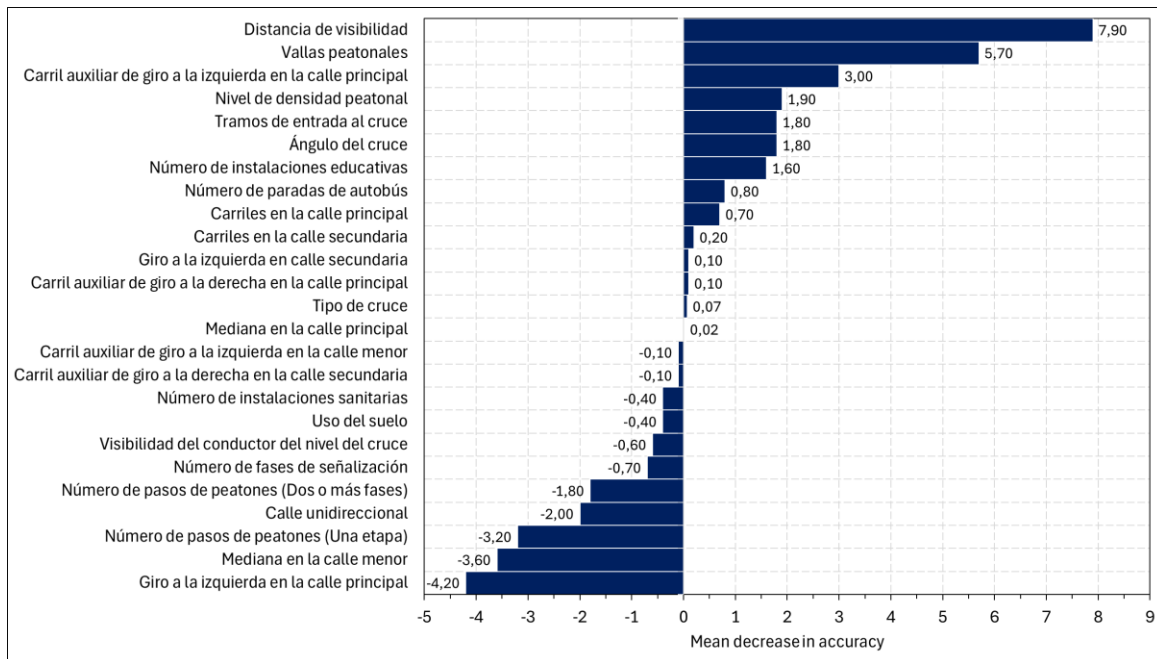
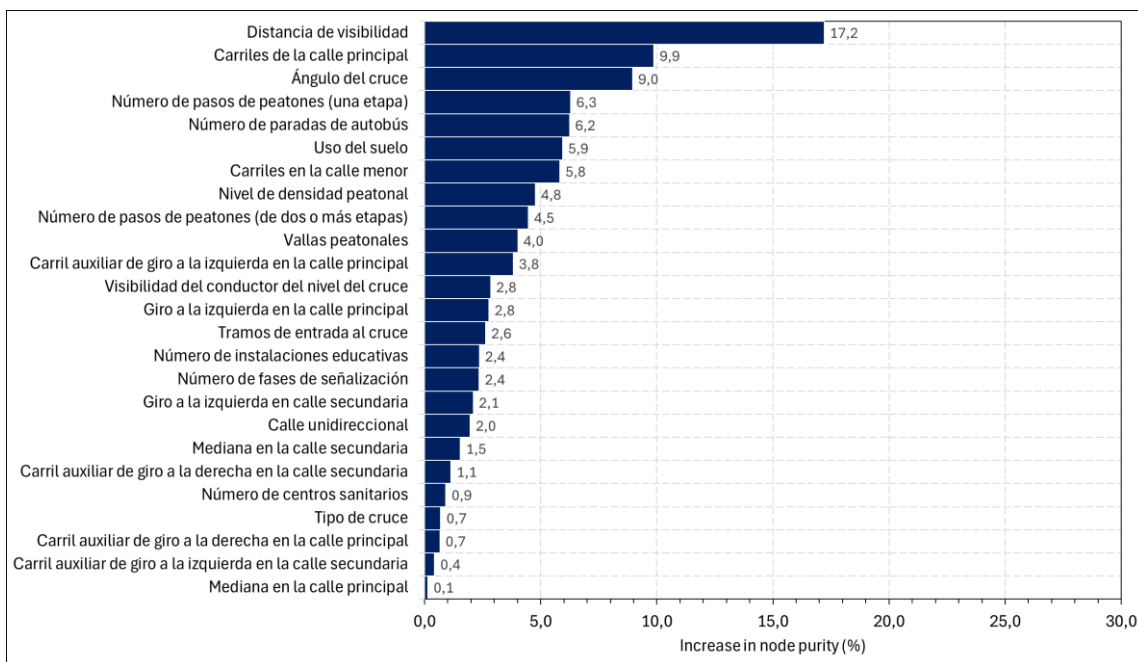


Figura 12: Incremento de la Pureza de los Factores de Riesgo en los nodos del experimento (en %)



Las figuras 11 y 12 muestran el peso de cada factor de riesgo en términos de la disminución media de la precisión y el aumento de la pureza de los nodos. La primera explica la relevancia y el signo del efecto de cada factor de riesgo sobre el índice de colisiones, y la segunda explica el peso relativo del

factor de riesgo sobre el índice de colisiones en la precisión de la predicción. La figura 6 muestra que la "distancia de visibilidad", las "vallas peatonales", el "carril auxiliar de giro a la izquierda en la calle principal", la "mediana en la calle secundaria" y el "giro a la izquierda en las calles principales" contribuyen a la precisión y deben tenerse en cuenta para predecir los índices de colisión en las intersecciones señalizadas. Tres factores se consideran infraestructuras y dos factores de riesgo operativo. Por el contrario, "carriles en la calle principal y secundaria", "giro a la izquierda en la calle secundaria", "carril auxiliar de giro a la derecha en la calle principal", "tipo de intersección", "mediana en la calle principal", "carril auxiliar de giro a la izquierda en la calle secundaria", "carril auxiliar de giro a la derecha en la calle secundaria", "número de instalaciones sanitarias", "uso del suelo", "visibilidad del conductor a nivel de la intersección" y "número de fases de señalización" no son relevantes para explicar la precisión de la predicción del índice de colisiones. Se consideran seis factores de riesgo para la infraestructura, tres para el funcionamiento de la intersección y dos para el entorno construido.

En la Figura 12 se identifican seis grupos que explican la importancia relativa de cada factor de riesgo en la predicción del índice de colisiones. Los tres grupos representan el 60% de los factores de riesgo que afectan al índice de colisiones: "Distancia de visibilidad", "carriles en la calle principal", "ángulo del cruce", "número de pasos de peatones (una etapa)", "número de paradas de autobús", "uso del suelo", "carriles en la calle secundaria", "nivel de densidad peatonal", "número de pasos de peatones (dos o más etapas)", "vallas peatonales", "carril auxiliar de giro a la izquierda en la calle principal". Seis de estos factores se consideran infraestructuras, tres al funcionamiento de la intersección y dos al entorno construido. Este resultado coincide con los obtenidos en la figura 7.

4.5 Conclusiones

En este análisis, se aplicó una estrategia de modelización exhaustiva para comprender los factores que influyen en los accidentes en intersecciones urbanas. Mediante la agrupación de factores de riesgo relacionados con la infraestructura, operaciones y uso del suelo, se evaluó su impacto en los índices de accidentabilidad.

Del análisis exploratorio se obtuvo el comportamiento en la base de datos de las variables independientes, en donde se pudo ver cuáles tenían mayor variabilidad. Del mismo modo se planteó el PCA de estas donde se observó que, analizada la prueba de Kaiser-Meyer-Olkin (KMO), la prueba

de esfericidad de Bartlett y la prueba de Mardia, sobre las variables estas no son concluyentes por lo que no se hace una reducción de las mismas.

Se desarrollaron una serie de experimentos los cuales permitieron determinar la relevancia de la “distancia de visibilidad”, las “vallas peatonales” y el “giro a la izquierda en las calles principales”. Como los grupos con más relevancia dentro de la incidencia de accidentes de tránsito. Los cuales reúnen el 60% de factores de riesgo que explican la ocurrencia de estos.

5 CONCLUSIONES

El objetivo de este trabajo fue identificar los factores de riesgo de colisión y su peso relativo en las intersecciones señalizadas. Incluimos como caso de estudio la ciudad de Santa María de Los Ángeles, Chile. A partir de la revisión bibliográfica, combinada con mediciones en terreno y estudios de ingeniería en 128 intersecciones, permitió identificar 13 componentes asociados a la infraestructura, 19 a la operación y cuatro al entorno construido. Las variables y las colisiones de tráfico se georreferenciaron y sintetizaron en una base de datos, a partir de la cual se estudió su importancia relativa para explicar las colisiones.

El efecto de las características de las intersecciones y del entorno circundante suele tratarse por separado. A partir de una revisión del estado de la técnica desde 2003, fue posible clasificar los atributos de las intersecciones en factores operativos, de infraestructura y de uso del suelo. En la bibliografía predominan los análisis basados en el tipo de intersección, el flujo de vehículos, el número de carriles disponibles y la presencia de una mediana. Se han estudiado poco los factores de uso del suelo y las instalaciones urbanas cerca de las intersecciones, así como la interacción entre los factores operativos, de infraestructura y de uso del suelo.

La base de datos creada para analizar los factores de riesgo de accidentes tiene datos de distintos orígenes. Se realizó un esfuerzo considerable para recopilar datos de diversas fuentes, normalizarlos, referenciarlos y configurarlos para su posterior análisis. Para cada una de las 128 intersecciones estudiadas, se recopilaron datos utilizando imágenes de satélite para inferir la geometría, datos de campo para estimar las características de la intersección, como los factores de uso del suelo, y datos de estudios de ingeniería y mediciones de tráfico.

El volumen de datos, la diversidad de orígenes de los mismos y la incertidumbre presente principalmente en las variables de flujo constituyen una de las principales limitaciones en la formación de bases de datos para este tipo de análisis, y no siempre es posible disponer de todo este tipo de información. Por lo tanto, los resultados obtenidos en este artículo permiten orientar la búsqueda de datos que contribuyan significativamente a explicar las colisiones en intersecciones urbanas, reduciendo así el esfuerzo en la recogida de información y, al mismo tiempo, integrando factores que, hasta ahora, en la literatura, habían sido considerados por separado.

El análisis PCA reduce el número de variables independientes aplicando el criterio de la varianza explicada. Los resultados no fueron concluyentes, ya que la prueba de Kaiser-Meyer-Olkin (KMO), la prueba de esfericidad de Bartlett y la prueba de Mardia sugirieron que no se cumplían todas las hipótesis del PCA y que los datos no eran adecuados para realizar un análisis de PCA. Este resultado es coherente con la gran diversidad de datos y la combinación de datos continuos y discretos.

A partir del estado del arte, se identificaron 19 factores de riesgo asociados al funcionamiento de las intersecciones, 13 asociados a la infraestructura y cuatro asociados al entorno construido. Tras georreferenciar cada uno de estos factores, además de los accidentes ocurridos en 5 años en las intersecciones estudiadas, se probaron varias herramientas de aprendizaje automático, principalmente porque las variables de interés son continuas, ordinales y cardinales, lo que invalida los análisis tradicionales con modelos lineales generalizados deterministas o bayesianos. Los resultados obtenidos mediante herramientas de aprendizaje automático muestran que los factores de riesgo asociados a la infraestructura dominan sobre los factores de riesgo operativos y del entorno construido a la hora de explicar las colisiones en intersecciones señalizadas.

Aunque estos resultados son locales y dependientes del tamaño muestral disponible (que no siempre se corresponde con el óptimo), muestran que mediante el procedimiento utilizado es posible identificar las variables relevantes que explican las colisiones en intersecciones urbanas y que son útiles para el diseño geométrico y operativo de las intersecciones teniendo en cuenta el contexto del entorno construido próximo.

Asimismo, los estudios disponibles sobre el estado del arte muestran que los índices de colisión en las intersecciones se basaban en factores de riesgo clasificados según la infraestructura, la operación y el entorno construido cercano a cada intersección por separado, sin prestar atención a las inevitables interacciones. En este sentido, la contribución de este trabajo se orienta a la identificación de interacciones entre estas variables y a la detección del peso relativo de cada factor de riesgo, considerando que cada variable es de naturaleza diferente (continua, ordinal y cardinal), lo que limita los análisis estadísticos tradicionales y que los patrones de interacción pueden detectarse mejor con herramientas de aprendizaje automático. Por último, los resultados obtenidos en la investigación constituyen una base para calibrar modelos predictivos de colisiones, para lo cual es necesario incluir

aspectos adicionales no abordados en este trabajo. Por ejemplo, el tratamiento del exceso de ceros requiere una diferenciación más detallada de los tipos de colisión y la consideración de usuarios vulnerables y no vulnerables.

Finalmente, existen líneas de investigación en las cuales ahondar. Por ejemplo, determinar cómo la interacción del clima influye directamente en la ocurrencia de accidentes de tránsito y como esta variable se vincula con las que se utilizaron para este análisis. Además, si se complementara la base de datos generada con información de intersecciones no semaforizadas, se podrían analizar otro tipo de variables y así ver cómo afecta al número de accidentes. En la misma línea identificar parámetros sobre alguna variable que presenten una menor preponderancia dentro de la base y parametrizarla a fin de que se determine una mejor modelación. A futuro con el avance tecnológico identificar nuevas herramientas. Respecto de las variables utilizadas, se propone analizar nuevas opciones no consideradas y como estas afectan a la ocurrencia de accidentes de tránsito.

6 REFERENCIAS

- AASHTO (2010). Highway Safety Manual. 1st Edition. American Association of State Highway and Transportation Officials. Washington D.C.
- Al-Mistarehi, B.W., Alomari, A.H., Imam, R., and Mashaqba, M. (2022). Using Machine Learning Models to Forecast Severity Level of Traffic Crashes by R Studio and ArcGIS.
- Abdel-Aty, M. and Keller, J. (2005). Exploring the overall and specific crash severity levels at signalized intersections. *Accident Analysis and Prevention*. 37. 417 – 425. *Frontiers in Built Environment*, 8, 860805.
- Ali, Y., Hussain, F., and Haque, M. (2024). Advances, challenges, and future research needs in machine learning-based crash prediction models: A systematic review. *Accident Analysis and Prevention*, 194, 107378.
- Anjana, S. and Anjaneyulu, M.V.L.R. (2015). Safety analysis of urban signalized intersections under mixed traffic. *Journal of Safety Research*. 52. 9 – 14.
- Assi, K. (2020). Traffic crash severity prediction—A synergy by hybrid principal component analysis and machine learning models. *Int. J. Environ. Res. Public Health* 17, 7598.
- Barbosa Silva, P., Andrade, M., and Ferreira, S. (2020). Machine learning applied to road safety modeling: A systematic literature review. *Journal of Traffic and Transportation Engineering*, 7(6), 775 – 790.
- Barman, S. and Bandyopadhyaya, R. (2023). Modelling crash severity outcomes for low speed urban roads using back propagation – Artificial neural network (BP – ANN) – A case study in Indian context. *IATSS Research*, 47, 382–400.
- Bartlett, M.S. (1950). Test of Significance in Factor Analysis, *British Journal of Statistical Psychology*, 3, 77 – 85.
- Castro, M., Paleti, R. and Bhat, C. (2012). A latent variable representation of count data models to accommodate spatial and temporal dependence: application to predicting crash frequency at intersections. *Transportation Research Part B*. 46. 253 – 272.
- Chen, Ch. and Xie, Y. (2016). Modeling the effects of AADT on predicting multiple-vehicle crashes at urban and suburban signalized intersections. *Accident Analysis and Prevention*. 91. 72 – 83.
- Chen, Ch. and Zhou, J. (2016). Effects of the built environment on automobile-involved pedestrian crash frequency and risk, *Journal of Transport & Health*, 3, 448 – 456.

- Cunto, F., Sobreira, L. and Ferreira, S. (2015). Assessing the transferability of the Highway Safety Manual predictive method for urban roads in Fortaleza City, Brazil. *Journal of Transportation Engineering*. 141(1). 04014072-1.
- Dohoo, I.R., Ducrot, C., Fourichon, C., Donald, A. and Hurnik, D. (1996). An overview of techniques for dealing with large numbers of independent variables in epidemiologic studies. *Preventive Veterinary Medicine*. 29. 221 – 239.
- Dufeu, R. (2018). Modeling crash risk factors in urban intersections. (In Spanish). MSc. Thesis. Civil Engineering, Universidad de Concepción. Chile.
- Dziuban, C.D. and Shirly, E.C. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin*, 81(6), 358–361
- Elvik, R. (2023). What would a road safety policy fully consistent with safe system principles mean for road safety? *Accident Analysis and Prevention*, 193, 107336.
- Fahlquist, J.N. (2006). Responsibility ascriptions and Vision Zero. *Accident Analysis and Prevention*, 38, 1113 – 1118.
- Ferreira, S. and Couto, A. (2013). Traffic flow-accidents relationships for urban intersections on the basis of the translog function. *Safety Science*. 60. 115 – 122.
- Gilan, V.N.M., Hosseinian, S.M., Ghasedi, M. and Nikookar, M. (2021). Data-Driven Urban Traffic Accident Analysis and Prediction Using Logit and Machine Learning-Based Pattern Recognition Models. *Mathematical Problems in Engineering*, 2021, 9974219.
- Greibe, P. (2003). Accident prediction models for urban road. *Accident Analysis and Prevention*. 35. 273 – 285.
- Islam, M.R., Barua, S., Akter, S., Hadiuzzaman. M. and Haque, N. (2020) Impacts of nongeometric attributes on crash prediction at urban signalized intersections of developing countries, *Journal of Transportation Safety & Security*, 12(5), 671-696.
- ITF (2023). Road Safety Annual Report 2023. OECD Publishing, Paris.
- Jamal, A., Zahid, M., Rahman, M.T., Al-Ahmadi, H.M., Almoshaogeh, M., Farooq, D. and Ahmad, M. (2021). Injury severity prediction of traffic crashes with ensemble machine learning techniques: a comparative study. *International Journal of Injury Control and Safety Promotion*, 28(4), 408-427
- Kumara, S.S.P. and Chin, H.C. (2003). Modeling accident occurrence at signalized tee intersections with special emphasis on excess zeros. *Traffic Injury Prevention*. 4(1). 53 – 57.

- Lord, D., Mannering, F. (2010). The statistical analysis of crash frequency data: a review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*, 44(5), 291 – 305.
- Loo, B.P.Y. and Anderson, T.K. (2016). *Spatial analysis methods of road traffic collisions*. 1st Ed. CRC Press. Boca Raton.
- Mardia, K.V. (1970). Measures of Multivariable skewness and kurtosis with applications. *Biometrika*, 57(3), 519 – 530.
- Merlin, L., Guerra, E. and Dumbergh, E. (2020). Crash risk, crash exposure, and the built environment: A conceptual review. *Accident Analysis and Prevention*. 134. 105244.
- MINVU (2009). Instructions for design of urban infrastructure. *Streets Manual*. (In Spanish). Ministry of Housing and Urbanism. Chile.
- MINVU (2011). *Land Use Zoning of Santa María de Los Ángeles City*. (In Spanish). Ministry of Housing and Urbanism. Chile.
- Mitra, S., Chin, H.C. and Quddus, M. A. (2002). Study of intersection accidents by maneuver type. *Transportation Research Record*. 1784. 43 – 50.
- Mitra, S. and Washington, S. (2007). On the nature of over-dispersion in motor vehicle crash prediction models. *Accident Analysis and Prevention*. 39. 459 – 468.
- MOP (2023). *Recommendations for Road Design. Vol 3. Highways Manual* (In Spanish). Ministry of Public Works. Chile.
- Roshandeh, A., Agbelie, B.R.D.K. and Lee, Y. (2016). Statistical modeling of total crashes frequency at highway intersections. *Journal of Traffic and Transportation Engineering*. 3(2). 166 – 171.
- Salifu, M. (2004). Accident prediction models for unsignalized urban junctions in Ghana. *IATSS Research*. 28(1), 68 – 81.
- SECTRA (2014). *Update of Signals Programming at the Bio Bio Region* (In Spanish). Ministry of Transport and Telecommunications. Chile.
- Tasic, I. and Porter, R.J. (2016). Modeling spatial relationships between multimodal transportation infrastructure and traffic safety outcomes in urban environments. *Safety Science*. 82(2016), 325 – 337.
- Vieira Gomes, S. (2013). The influence of the infrastructure characteristics in urban road accident occurrence. *Accident Analysis and Prevention*. 60, 289 – 297.
- Vieira Gomes, S., Reddy S. and Lord, D. (2012). Estimating the safety performance of urban intersections in Lisbon, Portugal. *Safety Science*. 50, 1732 – 1739.

- Wang, C., Quddus, M. and Ison, S. (2011). Predicting accident frequency at their severity levels and its application in site ranking using a two-stage mixed multivariate model. *Accident Analysis and Prevention*. 43, 1979 – 1990.
- Weijermars, W. and Wegman, F. (2011). Ten Years of Sustainable Safety in the Netherlands. *Transportation Research Record*, 2213, 1 – 8.
- WHO (2023). Global status report on road safety 2023. World Health Organization, Geneva.
- Wong, S.C., Sze, N.N. and Li, Y.C. (2007). Contributory factors to traffic crashes at signalized intersections in Hong Kong. *Accident Analysis and Prevention*. 39, 1107 – 1113.
- Xu, X., Wong, S.C., Zhu, F., Pei, X., Huang, H. and Liu, Y. (2017). A Heckman selection model for the safety analysis of signalized intersections. *PLoS ONE* 12(7), e0181544.
- Zhang, Y., Xie, Y. and Li, L. (2012). Crash frequency analysis of different types of urban roadway segments using generalized additive model. *Journal of Safety Research*, 43, 107 – 114.