

**MÉTODO IMPLEMENTADO EN COMPUTADOR Y SISTEMA PARA ESTIMAR
DAÑOS EN EDIFICACIONES ANTE SISMOS**

5 CAMPO TÉCNICO

La presente divulgación se refiere a sistemas y métodos relacionados con la estimación de daños en edificaciones. En particular, se relaciona con métodos implementados por computador y sistemas para el procesamiento de datos mediante aprendizaje automático para predecir el alcance de los daños estructurales en edificaciones causados por fenómenos naturales, por ejemplo, sismos.

DESCRIPCIÓN DEL ESTADO DE LA TÉCNICA

15 La estimación de daños en edificaciones ante sismos es una valiosa herramienta que permite evaluar el riesgo y la vulnerabilidad de las estructuras ante terremotos. Determinar el grado de daño que experimenta una estructura cuando se ve sometida a un sismo de determinada intensidad, permite proporcionar información útil para la prevención de desastres, la planificación y la ordenación del territorio.

20 Adicionalmente, estimar el daño el daño que experimenta una estructura cuando se ve sometida a un sismo, permite tener un importante punto de partida para la toma de decisiones relacionadas con la rehabilitación o demolición de edificios peligrosos, la ubicación de hospitales, la ubicación de puestos de socorro en una determinada zona o, y el diseño de las
25 vías alternativas de evacuación y rápido acceso de las ayudas hacia las zonas más vulnerables ante determinados eventos sísmicos.

Otra de las aplicaciones que tiene la estimación de daños en edificaciones ante sismos es la industria de seguros. Las solicitudes de seguros, especialmente para pólizas que abarcan
30 objetos de valor como viviendas y mejoras en bienes inmuebles, suelen requerir una evaluación de riesgo para determinar los límites de cobertura proporcionados por la aseguradora y las tarifas de pago que el posible asegurado debe abonar por esa cobertura. Este proceso de evaluación de riesgos, conocido como proceso de suscripción, generalmente se ocupa de diversas fuentes importantes de riesgo, algunos de los cuales son los riesgos
35 relacionados con la propiedad en sí, uno de los cuales es, por ejemplo, el riesgo ante sismos y la estimación de daños en edificaciones ante sismos. En ocasiones, se necesita llevar a cabo una inspección física de la edificación que contribuye a medir el riesgo asociado a la propiedad misma.

En el estado de la técnica se conocen diferentes métodos y sistemas para estimar daños, daños estructurales causados por desastres naturales en general y estimar el daño sísmico en edificaciones.

- 5 El documento **US20210398227A1** divulga un sistema y método para detectar daños, pérdidas y/u otras condiciones asociadas con un edificio, estructura u otra propiedad inmueble usando un sistema de monitoreo de propiedad y el uso de aprendizaje automático mediante redes neuronales para facilitar la determinación de los niveles de riesgo de bienes inmuebles. El documento **US20210398227A1** no divulga un método explícito y definido para obtener datos
- 10 específicos y detallados sobre el tipo estructural de las edificaciones que se quieren evaluar. Adicionalmente, es sabido que en ocasiones las redes neuronales no pueden manejar adecuadamente variaciones dadas por características dinámicas por lo que podría perder información importante sobre la evolución del riesgo con el tiempo.
- 15 Esto puede ser especialmente problemático en áreas con una historia sísmica limitada o en el caso de edificaciones donde no se conoce su tipología estructural o son poco comunes o únicas.

Por otro lado, el documento **US10909647B2** divulga un sistema y un método para predecir y

20 propagar datos de daños estructurales en tiempo real durante un desastre natural, como un terremoto, para ayudar a los equipos de respuesta de emergencia a priorizar y coordinar sus esfuerzos de rescate y recuperación. El sistema incluye un procesador y una memoria que almacena un modelo de aprendizaje automático entrenado para predecir el daño estructural en edificios y bloques de edificios. Particularmente, el documento divulga como modelos de

25 aprendizaje automático Random Forest (RF), Redes Neuronales (NN) y Máquinas de Vectores de Soporte (SVM). No obstante, este sistema y método no tiene en cuenta la posible variación de los datos de sacudida de los bloques de los edificios ante un terremoto real, ni la influencia de otros factores ambientales o estructurales que puedan afectar al daño sísmico. Esto puede ser problemático si se requiere hacer una previsión o anticipación del impacto

30 que podría tener un evento sísmico antes de que ocurra como medida de prevención, ya que estos factores podrían hacer que la estimación de daño por bloque sea menos precisa o consistente, y que el coeficiente de propagación entre los bloques relacionados sea menos confiable.

- 35 El documento **US11004001B1** presenta un método para analizar las predicciones de daños estructurales causados por un terremoto con el fin de identificar áreas con niveles de daño elevados. Además, el documento también incluye información sobre una simulación en tiempo real. El método divulgado abarca los siguientes pasos: identificar las características asociadas a los edificios en una región, incluyendo al menos la antigüedad del edificio, el
- 40 material de construcción y el número de pisos; entrenar un programa de aprendizaje

automático a través de uno o más procesadores, el cual se basa en las características identificadas y los datos de daños por terremotos para generar un modelo que permita predecir los daños causados por un terremoto en función de la intensidad sísmica, utilizando datos de daños por terremotos correspondientes a varios terremotos y valores relacionados a las características identificadas. Este proceso evalúa las relaciones entre las características y el daño causado. Posteriormente, se estima el valor de daño para un bloque a partir de una serie de valores de daño en la región provocados por el terremoto. Se identifica un área de daño crítico en la región, que incluye múltiples bloques agrupados geográficamente con valores de daño por encima de un umbral predefinido. Por último, se presenta esta área de daño crítico en un mapa de la región. No obstante, el documento **US11004001B1** puede presentar deficiencias en situaciones en que no se cuentan con datos de eventos sísmicos o valores asociados con las características de edificios, el método podría no predecir el nivel de riesgo de un inmueble. Además de esto, el método divulgado menciona el uso de un proceso de identificación del área crítica de daño (CDA), la definición de parámetros para identificar el área crítica de daño CDA, como tamaño mínimo o máximo, puede afectar la precisión. del modelo siendo esto clave para la selección de bloques geográficamente *clusterizados*. Esto puede ser insuficiente en áreas con una historia sísmica limitada o en el caso de edificaciones donde no se conoce su tipología estructural o son poco comunes o únicas.

De acuerdo con lo anterior, los métodos y sistemas anteriormente descritos no permiten estimar correctamente los daños que podría sufrir una edificación específica durante un evento sísmico. Existe la necesidad de métodos y sistemas para estimar daños en edificaciones ante sismos.

BREVE DESCRIPCIÓN

La presente divulgación describe modalidades de un método implementado en computador para estimar daños en edificaciones ante sismos que comprende, ejecutar en una unidad de cómputo (1) las etapas de: a) de recibir un dato de georreferenciación de la ubicación de una edificación y una etapa b) de extraer de una base de datos conectada a una unidad de cómputo un parámetro estructural de la edificación. El método además incluye una etapa c) de obtener un dato de tipo estructural mediante un método de clasificación de aprendizaje supervisado no paramétrico que se ejecuta en la unidad de cómputo, dicho método toma como entrada el parámetro estructural, donde el dato de tipo estructural representa el sistema constructivo de la edificación. También, el método tiene una etapa d) de obtener un dato de estimación de daño sísmico mediante la unidad de cómputo aplicando una función de vulnerabilidad obtenida de la modelación computacional del comportamiento del dato de tipo estructural y el parámetro estructural de la edificación bajo los efectos adversos de una amenaza sísmica en una zona de proximidad a la demarcada por el dato de georreferenciación.

En algunas realizaciones del método implementado por computador aquí divulgado, el parámetro estructural de la edificación se puede seleccionar del grupo de datos que incluye: valor de rango de altura de la edificación, valor en rango de edad de la edificación, dato de características de construcción y combinaciones de estos. En los casos de combinación de datos significa que en parámetro estructural puede en algunas realizaciones contener una pluralidad de datos. El dato de tipo estructural de la etapa c) se selecciona del grupo conformado por las variables muros, pórtico, dual, prefabricado, reticular, confinada, reforzada, no reforzada, pórticos arriostrados, pórticos no arriostrados, bahareque, adobe.

En algunas modalidades del método aquí divulgado, La etapa d) que obtiene un dato de estimación de daño sísmico del método implementado en computador para estimar daños en edificaciones ante sismos que comprende, ejecutar en una unidad de cómputo de la presente divulgación un método estimación de vulnerabilidad estructural (8) que comprende las siguientes subetapas que se ejecutan en una unidad de cómputo:

d1) generar datos de combinaciones del parámetro estructural extraído de la etapa b) y del dato de tipo estructural obtenido en la etapa c) y almacenarlos en una base de datos conectada a la unidad de cómputo;

d2) generar una matriz de puntos de coordenadas que cubra una región geográfica de interés para generar microzonificación dentro de la cual se incluye un dato de georreferenciación;

d3) asociar en un base de datos la matriz de puntos de coordenadas de la etapa d2) con las combinaciones generadas en la subetapa d1) y con los valores de daño dados por una función de vulnerabilidad para las diferentes clases de dato de tipo estructural bajo los efectos adversos de una amenaza sísmica determinista, obteniendo una matriz de tres parámetros; combinación de parámetro de la subetapa d1) por los puntos de la matriz en coordenadas de la subetapa d2) y el daño esperado según la función de vulnerabilidad para la amenaza sísmica determinista;

d4) encontrar el punto en coordenadas más próximo entre las coordenadas de localización de la edificación a estimar el daño, indicadas con el dato de georreferenciación, con el punto más cercano de la matriz de puntos de coordenadas (10) que cubre una región geográfica de la subetapa d2);

d5) encontrar a qué combinaciones de parámetro de un dato de tipo estructural corresponde la edificación a estimar;

d6) obtener un dato de estimación de daño sísmico buscando en la matriz generada en la etapa d3) el daño esperado para la combinación de parámetro encontrado en la etapa d1) para el punto el punto en coordenadas encontrado en la etapa d4).

En algunas realizaciones de la presente divulgación si el dato de estimación de daño sísmico (7) es mayor un valor predeterminado se genera un dato de alarma.

La presente divulgación además describe modalidades de un sistema de procesamiento de datos que comprende al menos una unidad de cómputo configurada para ejecutar cualquiera de las modalidades del método implementado por computador aquí divulgadas.

Asimismo, la presente divulgación además describe modalidades de un programa de computador y un medio de almacenamiento legible por computadora que comprenden instrucciones que, cuando son ejecutadas por una computadora o una unidad de cómputo, hacen que ésta realice las etapas de cualquiera de las modalidades del método implementado por computador aquí divulgadas.

BREVE DESCRIPCIÓN DE LAS FIGURAS

La FIG. 1 ilustra un diagrama de flujo de una modalidad del método implementado en computador para estimar daños en edificaciones ante sismos.

La FIG. 2 ilustra un diagrama de bloques de los componentes de un sistema para un ejemplo particular de la presente divulgación.

La FIG. 3 ilustra un ejemplo de fallo determinista con falla frontal a 50 kilómetros de la ciudad de Bogotá, Colombia con magnitud de 7,5.

La FIG. 4 ilustra un ejemplo de matriz de puntos en coordenadas que cubra una región geográfica de interés, para el caso del ejemplo la ciudad de Bogotá, Colombia y que permite generar microzonificación dentro de la cual se incluye un dato de georreferenciación.

La FIG. 5. ilustra una función de vulnerabilidad.

DESCRIPCIÓN DETALLADA

La presente divulgación describe modalidades de un sistema, un medio de almacenamiento legible por computadora y un método implementado en computador para estimar daños en edificaciones ante sismos.

Haciendo referencia a la FIG.1, el método implementado en computador para estimar daños en edificaciones ante sismos, de la presente divulgación, comprende, ejecutar en una unidad de cómputo (1) una etapa a) de recibir un dato de georreferenciación (2) de la ubicación de una edificación; y una etapa b) de extraer de una base de datos (3) conectada a una unidad de

cómputo (1) un parámetro estructural (4) de la edificación. También, el método incluye una etapa c) de obtener un dato de tipo estructural (5) mediante un método de clasificación de aprendizaje supervisado no paramétrico (6) que se ejecuta en la unidad de cómputo, dicho método toma como entrada el parámetro estructural (4), donde el dato de tipo estructural (5) representa el sistema constructivo de la edificación. El método además incluye una etapa d) de obtener un dato de estimación de daño sísmico (7) mediante la unidad de cómputo aplicando un método estimación de vulnerabilidad estructural (8) que puede aplicar una función de vulnerabilidad obtenida de la modelación computacional del comportamiento del dato de tipo estructural (5) y el parámetro estructural (4) de la edificación bajo los efectos adversos de una amenaza sísmica en una zona de proximidad a la demarcada por el dato de georreferenciación (2).

Algunas de las ventajas y efectos técnicos que presentan las realizaciones de la presente descripción, por ejemplo, permiten evaluar la vulnerabilidad sísmica de una edificación en función de su estructura y ubicación geográfica. Esto, por ejemplo, puede ser útil para tomar decisiones informadas sobre la seguridad de las edificaciones en áreas propensas a sismos y planificar medidas de mitigación de riesgos. Por su parte la predicción del sistema estructural de una edificación ante la ausencia de alguno de los datos para su cálculo por medio de modelos de aprendizaje automático puede ser ventajoso para ampliar la cobertura en los análisis de seguridad y riesgos de determinada zona o ubicación. Además, puede ser ventajoso utilizar un método de clasificación de aprendizaje supervisado no paramétrico para obtener el dato de tipo estructural (5) de la edificación, que representa el sistema constructivo de la misma, ya que permite clasificar las edificaciones sin asumir una distribución previa de los datos ni requerir una gran cantidad de datos de entrenamiento. Asimismo, es beneficioso aplicar una función de vulnerabilidad obtenida de la modelación computacional del comportamiento del dato de tipo estructural (5) y el parámetro estructural (4) de la edificación bajo los efectos adversos de una amenaza sísmica, ya que permite estimar el daño sísmico de forma rápida y precisa, sin necesidad de realizar inspecciones o mediciones in situ.

La vulnerabilidad sísmica de una estructura se refiere a su predisposición intrínseca a sufrir daño ante la ocurrencia de un movimiento sísmico y está asociada directamente con sus características físicas y estructurales de diseño. Por otro lado, el daño sísmico se refiere al grado de degradación o destrucción causado por un terremoto o movimiento sísmico. Por tanto, un factor importante en la estimación del daño sísmico es determinar la vulnerabilidad sísmica. La evaluación de la vulnerabilidad sísmica de una estructura permite determinar su capacidad para resistir los efectos de un terremoto, lo que a su vez permite estimar el grado de daño que sufrirá en caso de un evento sísmico. Por lo tanto, la evaluación o cálculo de la vulnerabilidad sísmica es un paso fundamental en la estimación de daños en edificaciones ante sismos.

Para estimar daños en edificaciones durante sismos mediante la evaluación de la vulnerabilidad sísmica, se siguen varios pasos. En primer lugar, se recopilan datos sobre la edificación, como su diseño, materiales y ubicación. Luego, se modela la respuesta de la estructura a las fuerzas sísmicas y se realiza un análisis sísmico que simula diferentes escenarios de terremotos. La evaluación de la vulnerabilidad se basa en estos análisis y evalúa la capacidad de la edificación para resistir las fuerzas sísmicas. A partir de esta evaluación, se estima el posible daño en la edificación en caso de un sismo. Finalmente, se pueden hacer acciones como por ejemplo emitir recomendaciones de mitigación para mejorar la seguridad de la edificación. Este proceso puede servir para garantizar la seguridad de las edificaciones y reducir el riesgo de daños en caso de terremotos.

Ahora bien, el cálculo de la vulnerabilidad sísmica está condicionado por varios factores, tales como, por ejemplo, el tipo de daño que se pretende evaluar, el nivel de amenaza existente en la zona y la información disponible sobre las estructuras. No obstante, las estimaciones de daños en edificaciones ante sismos, divulgadas en el estado de la técnica, se establecen con referencia a una región geográfica amplia y no a una ubicación geográfica específica. En términos más precisos, se comienza por definir o seleccionar una región geográfica y luego se calcula el riesgo sísmico asegurado en esa área sin tener en cuenta los factores de riesgo individuales específicos de la ubicación, realizando estimaciones uniformes para todas las ubicaciones particulares dentro de esa zona geográfica. Este enfoque no refleja la variabilidad en las proyecciones de daños que puede existir dependiendo de los diferentes sistemas constructivos presentes y se realiza de esta manera, entre otras cosas, por la dificultad de saber los sistemas constructivos que puede tener una edificación en una posición geográfica específica. Además, los resultados obtenidos a través de las técnicas de investigación de campo suelen ser aplicables solo a una muestra de la población y no a una edificación en particular. Lo anterior puede ser considerado una dificultad que soluciona la presente divulgación.

En algunas realizaciones de la presente divulgación el parámetro estructural (4) de la edificación se selecciona del grupo de datos que incluye: valor de rango de altura de la edificación, valor en rango de edad de la edificación, dato de características de construcción y combinaciones de estos.

El dato de tipo estructural (5) puede incluir una o más características que describen el sistema constructivo o estructural de una edificación. Estas características se utilizan para categorizar y describir los elementos de construcción diseñados e interconectados para dar forma a la construcción de las edificaciones. Cada sistema estructural se subdivide según los tipos de elementos verticales utilizados para resistir las fuerzas sísmicas y el grado de capacidad de disipación de energía del material estructural empleado.

En una realización particular el dato de tipo estructural (5) de la etapa c) se selecciona del grupo conformado por las variables muros, pórtico, dual, prefabricado, reticular, confinada, reforzada, no reforzada, pórticos arriostrados, pórticos no arriostrados, bahareque, adobe.

5 El método de clasificación de aprendizaje supervisado no paramétrico (6) de la etapa c) del método implementado en computador para estimar daños en edificaciones ante sismos de la presente divulgación permite obtener un dato de tipo estructural (5). Un dato de tipo estructural (5) está conformado por las variables de vulnerabilidad definidas por ejemplo por los datos de ubicación geográfica de la edificación, materiales, altura, edad y tipo de
10 estructura que posee, ya que en función de esto se determina el grado de seguridad estructural y el grado de funcionalidad que tendrá la estructura, evaluando el comportamiento dinámico de las estructuras ante diferentes niveles de intensidad sísmica. Por tanto, el desconocimiento de este dato de tipo estructural (5) incide directamente en la función de vulnerabilidad impidiendo un estimado del daño en un punto de localización de una edificación.

15

En algunas realizaciones de la presente divulgación el método de clasificación de aprendizaje supervisado no paramétrico (6) de la etapa c) del método implementado en computador para estimar daños en edificaciones ante sismos de la presente divulgación se selecciona del grupo conformado por K-Nearest Neighbors (K-NN), Árboles de Decisión, Random Forest, Support Vector Machines (SVM), Naive Bayes, Redes Neuronales Artificiales (ANN), Máquinas de Soporte Vectorial, Vecinos Más Cercanos por Densidad (DBSCAN), Máquinas de Aprendizaje Extremo (Extreme Learning Machines, ELM), Gradient Boosting, XGBoost, LightGBM, Adaboost, Máquinas de Aprendizaje Basadas en Instancias (Instance-Based Learning), Clasificadores Bayesianos Locales (Local Bayesian Classifiers), Máquinas de
20 Aprendizaje Basadas en Reglas (Rule-Based Learning), Clasificador de Vecinos Más Cercanos Ponderados (Weighted K-NN), Clasificación por Regresión Logística Local (Local Logistic Regression), Clasificadores por Máxima Verosimilitud (Maximum Likelihood Classifiers), Clasificación por Máxima Entropía (Maximum Entropy Classification), Máquina de Aprendizaje de Votación (Voting Machine Learning), Clasificador de Bayes No
25 Paramétrico (Non-Parametric Bayesian Classifier), Clasificador de Árbol de Decisión No Paramétrico (Non-Parametric Decision Tree Classifier), Clasificador de Árboles Aleatorios No Paramétricos (Non-Parametric Random Forests), Clasificación por Regresión Local (Local Regression Classification), Clasificación por Redes Bayesianas (Bayesian Networks Classification) y combinaciones de los anteriores.

35

Adicionalmente en algunas realizaciones de la presente divulgación, el método de clasificación de aprendizaje supervisado no paramétrico (6) de la etapa c) del método implementado en computador u otros modelos, se puede aplicar en la obtención datos característicos asociados con el edificio o edificación objetivo, que permiten estimar o
40 generar datos vinculados con factores de riesgo y/o indicadores de riesgo del edificio. Las

técnicas pueden incluir el análisis de los factores de riesgo generados y/o indicadores de riesgo para determinar un nivel de riesgo del edificio. Los factores de riesgo, los indicadores de riesgo y/o el nivel de riesgo se pueden utilizar para propósitos, como por ejemplo la fijación de precios, cotizaciones, suscripción o renovación de pólizas de seguro.

5

En algunas realizaciones del método aquí divulgado, el método de clasificación de aprendizaje supervisado no paramétrico (6) puede ser un método de procesamiento k vecinos más cercanos (KNN, K-Nearest-Neighbours, en inglés) con una métrica M, y un valor K correspondiente al número de vecinos más cercanos. La métrica M se puede seleccionar, por ejemplo, del grupo conformado por distancia Minkowski, distancia Euclidiana, distancia Manhattan, distancia de hamming, distancia de Chebyshev, distancia de similitud coseno.

10

El método de procesamiento K vecinos más cercanos KNN comprende, ejecutar en una unidad de cómputo (1) las subetapas:

15

c1. leer con una unidad de cómputo un conjunto de datos de entrenamiento almacenados en una base de datos (3) que consta de pares de entrada y etiquetas de clase correspondientes;

20

c2. seleccionar un valor para K, que representa el número de vecinos más cercanos entre en el rango de valores dado por [3, 5, 7, 9];

c3. seleccionar una métrica de distancia "M" para calcular las distancias entre la nueva instancia y todos los puntos de datos en el conjunto de entrenamiento.

25

c4. cuando se presenta una nueva instancia para la clasificación:

i. calcular las distancias entre la nueva instancia y todos los puntos de datos en el conjunto de entrenamiento utilizando la métrica de distancia seleccionada "M".

30

ii. identificar los "K" puntos de datos más cercanos en función de las distancias calculadas.

iii. contar cuántos de los "K" vecinos pertenecen a cada clase de etiqueta y asignar la etiqueta de clase más común como la predicción.

35

iv. calcular el promedio de los valores objetivo (etiquetas) de los "K" vecinos y usar el resultado como la predicción.

e. proporcionar el resultado de la predicción como la clasificación de la nueva instancia.

El conjunto de datos de entrenamiento de la subetapa c1 del método de procesamiento k vecinos más cercanos KNN donde de las entradas pueden ser información histórica de edificaciones, número de pisos especificados a partir del rango de altura, edad de la edificación (rango de edad), materiales y sistema estructural) bajo etiquetas de clase correspondientes y dicha información puede estar almacenada en una base de datos.

En algunas realizaciones antes de la ejecución de ejecutar la subetapa c1 del método de procesamiento k vecinos más cercanos KNN se ejecuta una etapa c0 que consiste en preparar un conjunto de datos de entrenamiento, que consta de pares de entrada y etiquetas de clase correspondientes, y realizar cualquier procesamiento necesario mediante las siguientes subetapas:

a. Recopilar de Datos de Entrenamiento: En esta etapa, se reúnen los datos de entrenamiento que consisten en pares de entrada y las etiquetas de clase correspondientes. Cada par de entradas representa un conjunto de características o atributos, y la etiqueta de clase indica la categoría a la que pertenece la entrada. En un ejemplo particular de la siguiente divulgación los datos de entrada son edad de construcción, material y altura, la etiqueta de clase correspondiente es el sistema estructural, usando como datos de entrada los avalúos históricos realizados los 106.000 registros de una base de datos recolectada a partir de información histórica de avalúos realizados por peritos in situ a lo largo del tiempo.

b. Seleccionar el Valor de K: El primer paso en el proceso es seleccionar el valor de "K", que representa el número de vecinos más cercanos que se utilizarán para tomar decisiones de clasificación. La elección de un valor apropiado de K puede afectar la calidad de las predicciones. En un ejemplo particular de la siguiente divulgación se realiza de esta manera Se prueban varios centroides o vecinos cercanos, y se prueban los rangos de valores [3, 5, 7, 9] y se selecciona el centroide que produce la suma más baja de distancias al cuadrado.

En algunas realizaciones de ejemplo, una fracción de los datos relacionados con el nivel de daño no se emplea durante la etapa de entrenamiento; en su lugar, se reserva con el propósito de evaluar la precisión del método de clasificación de aprendizaje supervisado no paramétrico (6). Por ejemplo, se podría destinar el 80% de los datos disponibles para entrenar al método de procesamiento k vecinos más cercanos KNN, mientras que el restante 20% se guarda para evaluar el rendimiento del método de procesamiento k vecinos más cercanos KNN. Es posible reservar distintos porcentajes de datos para estas pruebas, como por ejemplo el 10%, el 30%, y así sucesivamente, y se pueden apartar distintos segmentos de datos con el mismo fin.

c. Almacenar los Datos de Entrenamiento: Los datos de entrenamiento, que consisten en los pares de entrada y etiquetas de clase, se almacenan en una base de datos en una estructura de datos adecuada, por ejemplo, Excel o txt formato csv, para facilitar la búsqueda y recuperación de vecinos cercanos durante el proceso de predicción. Estos datos de entrenamiento reposan en bases de datos alojadas en servidores y se consulta esta base de datos construida para el proceso de entrenamiento, una vez la base es extraída, misma es guardada de forma temporal en los servidores, por ejemplo, en bases de datos postgresQL.

d. Calcular Distancias: Para determinar los vecinos más cercanos de una nueva entrada, se calcula la distancia entre la entrada de prueba y todas las entradas de entrenamiento. Las medidas de distancia comunes incluyen la distancia euclidiana, la distancia de Manhattan y la distancia de Minkowski. En un ejemplo particular se calcula la diferencia entre los valores de los atributos de los dos objetos. La Distancia de Minkowski se calcula a partir de la siguiente fórmula:

$$D(x,y) = (\sum_{i=1}^n |x_i - y_i|^p)^{1/p}$$

En esta fórmula, x y y son los dos objetos que se están comparando, n es el número de atributos de los dos objetos y p es un parámetro que se utiliza para controlar la cantidad de influencia que cada atributo tendrá en el resultado. Para este caso en particular "p" se estableció en 2 y es equivalente a la distancia euclidiana, que es apropiada para datos numéricos.

e. Ordenar Vecinos: Después de calcular las distancias, los pares de entrada de entrenamiento se ordenan en función de su proximidad a la entrada de prueba. Los K vecinos más cercanos se seleccionan como aquellos con las distancias más cortas. En un ejemplo particular se calcula minimizando el error de la suma de las distancias en este caso la euclidiana (p=2 en el paso anterior) al cuadrado de cada punto a su centroide más cercano

f. Votar Clases: Una vez identificados los K vecinos más cercanos, se realiza una votación para determinar la etiqueta de clase más común entre estos vecinos. La etiqueta que obtiene la mayoría de los votos se asigna como la etiqueta de clase de la entrada de prueba. En un ejemplo particular de la siguiente divulgación, una vez todos los puntos se han asignado a un clúster inicial, se recalculan los centroides de cada grupo como la media de todos los puntos del segmento. Este proceso se itera varias veces hasta que no se presenten variaciones. se itera varias veces hasta que no se presenten variaciones.

g. Entrenamiento Completo: En el caso de KNN, el proceso de entrenamiento se almacenan los datos de entrenamiento y no implica la construcción de un modelo matemático o estructura interna para hacer predicciones. Es decir, que el método KNN almacena los datos de entrenamiento y los utiliza directamente en el proceso de predicción. Por lo tanto, una vez que se han almacenado los datos de entrenamiento y se ha seleccionado el valor de K, el entrenamiento se considera completo. En un ejemplo particular de la siguiente divulgación El entrenamiento completo se realiza en un servidor de acuerdo a los parámetros definidos anteriormente y se guarda el resultado en una base de datos temporal por ejemplo postgresQL, para su uso en el cálculo del nivel de riesgo En una realización de la presente divulgación para seleccionar los mejores hiper parámetros para el modelo KNN se basó en la técnica de búsqueda exhaustiva en cuadrícula (*Grid Search*) para determinar los valores óptimos para un modelo determinado.

En una realización de la presente divulgación el método de clasificación de aprendizaje supervisado no paramétrico (6) se buscaron parámetros para el modelo KNN mediante una técnica de búsqueda exhaustiva en cuadrícula (*Grid Search*). Este proceso tiene como objetivo determinar los valores más adecuados para un modelo específico para lo cual se procedió mediante los siguientes pasos:

Paso 1: Definir y ajustar parámetros:

a. K, representa el número de vecinos más cercanos (también llamados *n_neighbors*): Este parámetro determina cuántos vecinos se considerarán al realizar una predicción para un nuevo punto de datos. Se exploraron varios valores de K en el rango de valores [3, 5, 7, 9];

b. Métrica M (también llamada Métrica de distancia): La métrica de distancia utilizada para calcular las similitudes entre los puntos es de vital importancia en el método KNN. En una realización de la presente divulgación se exploraron dos métricas: 'euclidiana' y 'Manhattan';

c. W Función de Peso (también llamada *weights*): El parámetro W determina cómo se ponderan los vecinos cercanos. En una realización de la presente divulgación se consideraron dos opciones: 'uniforme' y 'distance'. Con 'uniforme', todos los vecinos tienen el mismo peso en la votación, mientras que con 'distance', los vecinos más cercanos tienen un peso mayor.

Paso 2: Evaluación del modelo para cada combinación utilizando el método de validación cruzada K-FOLD-CROSS-VALIDATION consiste en las siguientes etapas:

- a. División del conjunto de datos: En primer lugar, se divide el conjunto de datos en K subconjuntos aproximadamente iguales. Estos subconjuntos a menudo se denominan "pliegues" (folds). El valor de K es un parámetro que se elige previa y generalmente varía de 5 a 10, aunque puede ser diferente según el problema;

5

- b. Iteración a través de los pliegues: Luego, se realiza un ciclo de K iteraciones. En cada iteración, un pliegue se utiliza como conjunto de prueba y los otros K-1 pliegues se utilizan como conjunto de entrenamiento. Esto significa que el modelo se entrena K veces, cada vez con una combinación diferente de pliegues de entrenamiento y prueba;

10

- c. Evaluación del rendimiento: Durante cada iteración, se entrena el modelo KNN en el conjunto de entrenamiento y se evalúa en el conjunto de prueba. Se registran las métricas de rendimiento, como la precisión, la puntuación F1, el error cuadrático medio (en problemas de regresión), etc;

15

- d. Promedio de resultados: Finalmente, se promedian las métricas de rendimiento de todas las iteraciones para obtener una estimación más robusta y confiable del rendimiento del modelo. Esta estimación se utiliza para evaluar qué tan bien se desempeñaría el modelo en datos no vistos.

20

La ventaja del método de validación cruzada K-Fold Cross-Validation es que ayuda a evitar problemas como el sobreajuste (cuando un modelo se ajusta demasiado a los datos de entrenamiento) y proporciona una mejor evaluación del rendimiento del modelo en comparación con una sola división de entrenamiento/prueba. Además, permite aprovechar al máximo el conjunto de datos disponible para el entrenamiento y evaluación.

25

Paso 3: Selección de los mejores parámetros encontrados para el modelo KNN basándose en su precisión o pérdida para cada combinación de parámetros que para una realización de ejemplo de la presente divulgación por ejemplo así:

30

- a. $K=5$: Esto indica que se utilizarán los 5 vecinos más cercanos para tomar decisiones de clasificación.
- b. W Función de Peso (weights) seleccionada: 'uniforme', lo que implica que todos los vecinos tienen el mismo peso en la votación.
- c. Métrica de distancia (Métrica M) elegida: 'Minkowski' con $p=2$, que es equivalente a la métrica 'euclidiana'. Esta métrica calcula la distancia euclidiana entre puntos.

35

Cada uno de los parámetros seleccionados desempeña un papel fundamental en el rendimiento y el comportamiento del modelo KNN:

40

El valor de K (`n_neighbors`) influye en la suavidad de la frontera de decisión del modelo. Un valor bajo puede hacer que el modelo sea más sensible al ruido, mientras que un valor alto puede hacerlo demasiado suave y menos capaz de capturar patrones complejos.

- 5 La función de peso (`weights`) afecta cómo se ponderan los vecinos cercanos. 'Uniforme' trata a todos los vecinos por igual, mientras que 'Distance' da más peso a los vecinos más cercanos, lo que puede ser beneficioso cuando algunos vecinos son más influyentes que otros.

- 10 La elección de la métrica de distancia (Métrica M) es esencial. 'Minkowski' con $p=2$ es equivalente a la distancia euclidiana y es apropiada para datos numéricos, mientras que 'Manhattan' es adecuada cuando las características tienen diferentes escalas o cuando se desea que el modelo sea más robusto ante valores atípicos.

- 15 La selección de los parámetros óptimos mediante la metodología de búsqueda en cuadrícula ha permitido ajustar el modelo KNN para que se adapte mejor a los datos y, en última instancia, mejorar su capacidad de clasificación. Los resultados indican que un modelo KNN con $K=5$, Función de Peso (W) 'uniforme', y Métrica de Distancia (Métrica M) 'Minkowski' con $p=2$ es la configuración en este contexto. Esta elección se basa en la maximización de la precisión y la capacidad de generalización del modelo, lo que sugiere que es la mejor configuración para clasificar datos similares en futuras aplicaciones.

- 25 En una realización de ejemplo la elección de los rangos específicos para los hiper parámetro, para el hiper parámetro K en este caso [3, 5, 7, 9], es importante adaptar los valores de los hiper parámetros a las características del conjunto de datos y los requisitos del problema, considerando ajustar estos valores según el conocimiento previo de los datos y realizar pruebas adicionales si es necesario para encontrar el valor óptimo de parámetros.

- 30 En una realización de ejemplo la exploración de K para valores impares: Se optó por considerar valores impares para K en el método KNN debido a que utilizar un número impar de vecinos puede prevenir empates en las votaciones. En situaciones de empate, un número impar garantiza la existencia de un voto ganador, lo que simplifica el proceso de clasificación. Esta estrategia es particularmente útil en problemas binarios donde se requiere una decisión clara.

- 35 Consideración de la paridad de clases: Si el conjunto de datos presenta un número par de clases, la elección de un valor impar para K puede resultar beneficiosa, ya que ayuda a evitar situaciones de empate en la votación. Esto puede ser relevante en contextos de clasificación multiclase, donde se busca obtener una decisión única.

Exploración de valores representativos: en el rango de valores [3, 5, 7, 9] representan opciones comunes y adecuadas para la exploración de parámetros en el método KNN. Estos valores ofrecen un rango razonable de alternativas para K sin necesidad de explorar un espacio de parámetros excesivamente extenso.

5

Equilibrio entre sesgo y varianza: La elección de K también está vinculada al balance entre sesgo y varianza. Un valor bajo de K, como 3, puede resultar en un modelo con bajo sesgo pero alta varianza, lo que implica que se ajusta de manera ajustada a los datos, pero es sensible al ruido. Por otro lado, un valor alto de K, como 9, puede reducir la varianza, pero aumentar el sesgo, generando un modelo más suave y menos sensible al ruido. Valores intermedios de K, como 5 y 7, buscan un punto intermedio entre estas dos consideraciones.

10

Tamaño del conjunto de datos: Además, el tamaño del conjunto de datos también puede influir en la elección de los valores de K. En conjuntos de datos pequeños, a menudo se prefieren valores más bajos de K para evitar que el modelo no pueda capturar la relación entre las variables de entrada y salida. En conjuntos de datos grandes, valores más altos pueden ser apropiados para suavizar las decisiones del modelo.

15

En algunas modalidades si el dato de estimación de daño sísmico (7) es mayor un valor predeterminado se genera un dato de alarma.

20

Finalmente, algunas realizaciones de ejemplo, el método de la presente divulgación también puedes incluir etapas que partiendo del dato de estimación de daño sísmico (7) pueden estimar la pérdida financiera en la edificación basándose en el daño estimado por terremoto a la edificación y un valor de reemplazo de la edificación.

25

La etapa d) que obtiene un dato de estimación de daño sísmico (7) del método implementado en computador para estimar daños en edificaciones ante sismos de la presente divulgación comprende, ejecutar en una unidad de cómputo (1) un método estimación de vulnerabilidad estructural (8) las siguientes subetapas:

30

d1) generar datos de combinaciones (9) del parámetro estructural (4) extraído de la etapa b) y del dato de tipo estructural (5) obtenido en la etapa c) y almacenarlos en una base de datos (3) conectada a la unidad de cómputo;

35

d2) generar una matriz de puntos de coordenadas (10) que cubra una región geográfica de interés para generar microzonificación dentro de la cual se incluye un dato de georreferenciación (2);

d3) asociar en un base de datos (3) la matriz de puntos de coordenadas (10) de la etapa d2) con las combinaciones generadas en la subetapa d1) y con los valores de daño dados por una función de vulnerabilidad para las diferentes clases de dato de

40

tipo estructural (5) bajo los efectos adversos de una amenaza sísmica determinista, obteniendo una matriz de tres parámetros; combinación de parámetro de la subetapa d1) por los puntos de la matriz en coordenadas de la subetapa d2) y el daño esperado según la función de vulnerabilidad para la amenaza sísmica determinista; donde este dato del daño esperado se ubica en la intersección de cada punto de coordenadas con cada una de los datos de combinaciones (9);

d4) encontrar el punto en coordenadas más próximo entre las coordenadas de localización de la edificación a estimar el daño, indicadas con el dato de georreferenciación (2), con el punto más cercano de la matriz de puntos en coordenadas que cubre una región geográfica de la subetapa d2);

d5) para el dato de tipo estructural (5) encontrar la combinación de parámetro de dato de tipo estructural (5) corresponde la edificación a estimar;

d6) obtener un dato de estimación de daño sísmico (7) buscando en la matriz generada en la etapa d3) el daño esperado para la combinación de parámetro encontrado en la etapa d1) para el punto el punto en coordenadas encontrado en la etapa d4).

En un ejemplo etapa d) que obtiene un dato de estimación de daño sísmico (7) del método implementado en computador para estimar daños en edificaciones ante sismos de la presente divulgación:

Haciendo referencia a la FIG. 3 En un ejemplo del método implementado por computador para estimar daños en edificaciones ante sismos de la presente divulgación se quiere estimar los daños para una edificación con una amenaza de un sismo de magnitud 7.75 (13) que ocurre en una falla frontal, entre 40 kilómetros a 50 kilómetros de distancia de la ciudad de Bogotá, Colombia (15).

El daño esperado en un sismo también depende de la tipología constructiva, por este motivo se definieron las tipologías constructivas sobre las cuales se quería estimar el nivel de daño para un ejemplo particular de la subetapa d1) se genera la generar datos de combinaciones (9) del parámetro estructural (4) extraído de la etapa b) y del dato de tipo estructural (6) como se presenta en la siguiente Tabla 1:

Tipo estructural (5)	Parámetro estructural (4)			
SISTEMA Constructivo	MATERIAL	# RANGOS DE ALTURA	# Edad	Combinaciones
Pórticos	Concreto	4	3	12
Muros	Concreto	2	3	6
Dual	Concreto	3	3	9
Prefabricado	Concreto	3	3	9

Reticular	Concreto	5	3	15
Confinada	Mampostería	3	3	9
Reforzada	Mampostería	3	3	9
No reforzada	Mampostería	1	1	1
pórticos arriostrados	Acero	3	3	9
Pórticos No arriostrados	Acero	4	3	12
Bahareque	Madera	1	1	1
Adobe	Barro	1	1	1

Tabla 1

Para el ejemplo en particular de la tabla la altura se agrupó en cuatro grupos de rangos de altura (en la tabla designado como # RANGOS DE ALTURA) organizados según el número de pisos de la edificación así de entre 1 piso y 3 pisos, 4 pisos y 7 pisos, 8 pisos y 10 pisos, finalmente más de 10 pisos. Para la edad, básicamente se tiene en cuenta cuál es la fecha de construcción de la edificación, Para el ejemplo particular se definieron 3 intervalos de fechas de construcción. Para el caso del ejemplo, antes de 1984, entre 1985-1997 y posterior a 1998. Los anteriores intervalos se definen considerando las fechas en que se han publicado o actualizado los códigos de construcción sismo resistente en Colombia.

El número de combinaciones para este ejemplo sale de multiplicar los rangos de altura por cada uno de los intervalos de edad establecidos. En el ejemplo de la tabla se definió un total de 93 combinaciones finales.

En el ejemplo para el dato tipo estructural (5) pórticos combinados con el parámetro estructural (4) concreto tenemos 12 combinaciones porque son cuatro rangos de altura por tres posibles intervalos de edad del parámetro (4).

Así por ejemplo la combinación número tres estaría conformada por ejemplo por pórticos de concreto de rangos de altura de entre uno y tres pisos con una edad en el rango mayor al año 2010.

Se debe notar que el número de combinaciones no es el mismo para los diferentes datos tipo estructural (5), por ejemplo, para adobe solo se contempla un rango de altura, ya que generalmente son edificaciones entre 1 y 2 pisos. De igual manera, solo se contempla un intervalo de edad ya que las edificaciones en adobe no se encuentran reglamentadas en los códigos de construcción sismo resistente, y por tanto para el análisis solo se contempla un intervalo.

Haciendo referencia a la FIG. 4 Se tomó el perímetro de la ciudad Bogotá, Colombia (5) y se planteó matriz de puntos de coordenadas (10) con puntos espaciados cada de 250 metros para un total de 10987 puntos que cubren toda la ciudad para generar matriz de puntos de coordenadas (10) de la subetapa d2).

En algunos ejemplos la matriz de puntos de coordenadas (10) de la subetapa d2) se puede definir mediante una malla de puntos que pudiera tener una cobertura del predio a evaluar y la distancia del punto más cercano de la malla que consideren un buen tiempo computacional de acuerdo con la unidad de cómputo disponible para realizar el proceso.

En un ejemplo los puntos en coordenadas de la matriz que cubre una región geográfica de la subetapa d2) de los pasos para la obtención de un dato de estimación de daño sísmico (7) se encuentran en una distancia de un rango entre 250 metros y 250 metros esta selección proporciona como efecto técnico una buena resolución y tiempo computacional.

En otros ejemplos es posible realizar interpolación espacial para estimar los puntos coordinados que conforman la malla basándose en la distribución y las características de los puntos conocidos.

En cada uno de los puntos definidos en la matriz de puntos de coordenadas (10) se estiman cuál sería el nivel de daño que se podría generar. Dada la ocurrencia de este sismo, para cada uno de los puntos de coordenadas se realiza una estimación del nivel de daño para cada una de las 93 combinaciones.

Haciendo referencia a la FIG. 5 la estimación del daño se logra mediante funciones de vulnerabilidad (por ejemplo, usando un software como ERN en una modelación computacional).

En una realización de ejemplo se asocia la matriz de puntos en coordenadas que cubre una región geográfica de la etapa l) con combinaciones de los parámetros con una función de vulnerabilidad, para los parámetros sistema estructural predecibles por la etapa i) bajo los efectos adversos de una amenaza sísmica determinista y obteniendo una matriz de tres parámetros, 1) combinación de parámetro de los puntos en coordenadas que cubre una región geográfica y 3) daño esperado según la función de vulnerabilidad para la amenaza sísmica determinista de esta manera en el ejemplo la subetapa d3) la matriz de puntos de coordenadas (10) asocia en un base de datos (3) una a matriz de 10987 puntos por 93 combinaciones y la intercesión de puntos y combinaciones es la estimación de daño según la función de vulnerabilidad para un edificio de las características de dadas por la combinación en la coordenada con datos de combinaciones (9).

En la siguiente Tabla 2 se presenta como ejemplo una porción de la matriz para los primeros 5 puntos y las 5 primeras combinaciones.

NUMREG	LONGITUD	LATITUD	COL_1	COL_2	COL_3	COL_4	COL_5
1	-74.099618	4.455625	0.2171	0.1782	0.1381	0.2362	0.1897
2	-74.097368	4.455625	0.2182	0.1791	0.1388	0.2375	0.1908
3	-74.095118	4.455625	0.2193	0.18	0.1395	0.2388	0.1819
4	-74.092868	4.455625	0.2204	0.1809	0.1402	0.24	0.1929
5	-74.101868	4.457875	0.2171	0.1783	0.1384	0.2351	0.1888

5 Tabla 2.

En la tabla 2 la columna NUMREG corresponde al número del punto con coordenadas. Para el caso del ejemplo, se refiere a los cinco primeros puntos de los 10987. Cada punto especifica sus coordenadas por LONGITUD y LATITUD. Por su parte las columnas COL_1, COL_2, COL_3, COL_4 y COL_5 corresponden a las primeras cinco combinaciones de las 93 combinaciones totales.

Continuando con el ejemplo, suponiendo que se tiene una edificación para la cual se requiere estimar daños mediante un dato de estimación de daño sísmico (7), ante el sismo de magnitud 7.75 (13) que ocurre en una falla frontal, entre 40 kilómetros y 50 kilómetros de distancia de la ciudad de Bogotá, Colombia (15). Dicha edificación tiene un dato de georreferenciación (2) específico.

En el ejemplo, aplicando la subetapa d4), se encuentra que el punto en coordenadas más próximo entre las coordenadas de localización de la edificación, dado por un dato de georreferenciación (2), para estimar el daño es, por ejemplo, el punto 3 de la tabla, es decir el punto con coordenadas LONGITUD -74.095118 y LATITUD 4.454625.

En el ejemplo, tomando como entrada el dato de tipo estructural (5), por ejemplo, obtenido mediante un método de clasificación de aprendizaje supervisado no paramétrico (6), de la edificación dada en dato de georreferenciación (2) y aplicando al ejemplo la subetapa d5), se encuentra que corresponde, por ejemplo, a la combinación COL_3. Según la tabla 1. Finalmente, en la subetapa d6), buscando, mediante la unidad computo (1), en la matriz generada en la etapa d3) el daño esperado para la combinación de parámetros encontrados en la etapa d1) para el punto de coordenadas encontrado en la etapa d4), se encuentra que el dato de estimación de daño sísmico (7) es 0.2193, lo que corresponde a un daño esperado del 21.93%.

En una realización de la presente divulgación del método implementado en computador para estimar daños en edificaciones ante sismos, de la presente divulgación, comprende, ejecutar en una unidad de cómputo (1) después de la etapa d, una etapa donde si el daño esperado es mayor un valor predeterminado se genera un dato de alarma.

5

En algunas realizaciones de ejemplo el valor predeterminado de daño esperado es de menor o igual al treinta y cinco por ciento, 35%.

10 A continuación, se presenta una posible escala de daño sísmico en edificaciones basada en porcentajes:

Menos del 10%: puede representar por ejemplo daño superficial, como grietas leves en paredes. No se requiere intervención estructural significativa.

15 Entre el 10% - 20% puede representar por ejemplo grietas más extensas y posiblemente daño en elementos no estructurales. Puede requerir reparaciones menores.

20 Entre el 20% - 30% puede representar por ejemplo daño moderado con grietas importantes en paredes y posibles deformaciones en elementos estructurales. Puede requerir intervención para reparaciones sustanciales.

Entre el 30% - 40%: puede representar, por ejemplo, daño severo con grietas extensas y deformaciones significativas. Puede implicar la necesidad de demolición y reconstrucción.

25 Más del 40% puede representar por ejemplo daño crítico con colapso parcial o total de la estructura. Puede requerir demolición y reconstrucción.

30 Para evaluar el porcentaje de daño estructural, es común utilizar escalas de daño específicas, como la Escala de Daño Sísmico (EDS) o la Escala de Evaluación de Daños Sísmicos (SEEDS), que pueden variar en su diseño y en los criterios específicos de evaluación, es posible realizar equivalencias entre estas escalas y la estimación de daños en edificaciones ante sismos obtenida mediante el método de la presente divulgación.

35 Se cuenta además con un sistema de procesamiento de datos que comprende medios para llevar a cabo las etapas del método y un medio de almacenamiento legible por computadora o una unidad de cómputo que comprende instrucciones que, cuando son ejecutadas por una computadora o una unidad de cómputo y hacen que ésta realice los pasos del método.

40 El sistema incluye al menos la unidad de cómputo (1) que ejecuta cualquiera de las modalidades del método aquí divulgado.

La unidad de cómputo (1) se puede seleccionar entre el grupo que incluye microcontroladores (v.g., arduino, Raspberri-pi), micro procesadores, DSCs (Digital Signal Controller por sus siglas en inglés), FPGAs (Field Programmable Gate Array por sus siglas en inglés), CPLDs (Complex Programmable Logic Device por sus siglas en inglés), ASICs (Application Specific Integrated Circuit por sus siglas en inglés), SoCs (System on Chip por sus siglas en inglés), PSoCs (Programmable System on Chip por sus siglas en inglés), computadores, servidores, tabletas, celulares, celulares inteligentes, generadores de señales y otras unidades de cómputo similares o equivalentes conocidas por una persona medianamente versada en la materia.

10

La unidad de cómputo (1) puede incluir un módulo de memoria que se puede seleccionar entre el grupo que incluye , memorias RAM (memoria caché, SRAM, DRAM, DDR), memoria ROM (Flash, Caché, discos duros, SSD, EPROM, EEPROM, memorias ROM extraíbles (v.g. SD (miniSD, microSD, etc), MMC (MultiMedia Card), Compact Flash, SMC (Smart Media Card), SDC (Secure Digital Card), MS (Memory Stick), entre otras)), CD-ROM, discos versátiles digitales (DVD por las siglas en inglés de Digital Versatile Disc) u otro almacenamiento óptico, casetes magnéticos, cintas magnéticas, almacenamiento o cualquier otro medio que pueda usarse para almacenar información al que pueda acceder la unidad de cómputo (1) que sean conocidos por una persona medianamente versada en la materia.

20

Adicionalmente, la unidad de cómputo (1) puede conectarse a un dispositivo de entrada/salida que puede incluir un dispositivo de almacenamiento, dispositivo de visualización y/o un Dispositivo de Interfaz Humana (HID, por las siglas en inglés de Human Interface Device). Ejemplos del dispositivo de visualización se seleccionan entre monitor CRT (por las siglas en inglés de Cathode Ray Tube), pantalla plana, pantalla de cristal líquido LCD (por las siglas en inglés de Liquid Crystal Display), pantalla LCD de matriz activa, pantalla LCD de matriz pasiva, pantallas LED, proyectores de pantallas, TV (4KTV, HDTV, TV de plasma, Smart TV), pantallas OLED (por las siglas en inglés de Organic Light Emitting Diode), pantallas AMOLED (por las siglas en inglés de Active Matrix Organic Light Emitting Diode), Pantallas de puntos cuánticos QD (por las siglas en inglés de Quantic Display), pantallas de segmentos, entre otros dispositivos capaces de mostrar datos a un usuario, conocidos por los expertos en la técnica, y combinaciones de estos.

25

30

Ejemplos del dispositivo HID (por las siglas en inglés de Human Interface Device) pueden ser, sin limitación, teclado, mouse, trackball, touchpad, dispositivo apuntador, joystick, pantalla táctil, entre otros dispositivos capaces de permitir que un usuario ingrese datos en la unidad de cómputo (1) del dispositivo y combinaciones de estos.

35

Haciendo referencia a la FIG. 2 los componentes de un sistema, para un ejemplo particular de la presente divulgación consta de una unidad de cómputo (1) que se encuentra interconectada con una base de datos (3). Esta base de datos (3) almacena por ejemplo información de georreferenciación mediante el dato de georreferenciación (2) y datos estructurales (5).

Los datos estructurales (5) son obtenidos mediante un método de clasificación de aprendizaje supervisado no paramétrico (6) almacenado en por ejemplo un primer medio de almacenamiento legible por computadora, y que toma como entrada un parámetro estructural (4). La Unidad de Cómputo (1) realiza una estimación de daño sísmico (7) utilizando un método de estimación de vulnerabilidad estructural (8) almacenado en un segundo medio de almacenamiento legible por computadora. En algunas realizaciones el primer medio de almacenamiento legible por computadora y el segundo medio de almacenamiento legible por computadora son un mismo medio de almacenamiento legible por computadora.

Opcionalmente, el sistema tiene la capacidad opcional de conectarse a una red (16), como por ejemplo internet, y otras unidades de cómputo (18) lo que permite integración con otros sistemas y fuentes de datos, como, por ejemplo, la conexión a Internet. Esta conexión adicional permite mejorar y complementar la información almacenada en la base de datos (3).

El término "medio de almacenamiento legible por computadora" debe considerarse en el entendimiento de la presente divulgación para incluir un medio único o múltiples medios (por ejemplo, una base de datos centralizada o distribuida, o cachés y servidores asociados) capaces de almacenar instrucciones. El término "medio de almacenamiento legible por computadora" también se considerará para incluir cualquier medio, o combinación de múltiples medios, capaz de almacenar instrucciones para el método implementado en un computador para estimar daños en edificaciones ante sismos y realizar su ejecución por un dispositivo o sistema. De este modo, las instrucciones o etapas que comprende el método, cuando son ejecutadas por uno o más procesadores de un dispositivo o sistema, hacen que el dispositivo o sistema pueda llevar a cabo una o más de las metodologías descritas en la presente divulgación, en su totalidad o en parte.

En consecuencia, un "medio de almacenamiento legible por computadora" se refiere a un único aparato o dispositivo de almacenamiento, así como a sistemas de almacenamiento basados en la nube o redes de almacenamiento que incluyen múltiples aparatos o dispositivos de almacenamiento. Por lo tanto, se considerará que el término "medio de almacenamiento legible por computadora" incluye, entre otros, uno o más depósitos de datos tangibles (por ejemplo, no transitorios) en forma por ejemplo de una memoria, una memoria de estado sólido, un medio óptico, o cualquier combinación adecuada de los mismos.

El medio legible por computador puede seleccionarse entre archivos ejecutables, archivos instalables, discos compactos, memorias RAM (memoria caché, SRAM, DRAM, DDR), memoria ROM (Flash, Caché, discos duros, SSD, EPROM, EEPROM, memorias ROM extraíbles (v.g. SD (miniSD, microSD, etc), MMC (MultiMedia Card), Compact Flash, SMC (Smart Media Card), SDC (Secure Digital Card), MS (Memory Stick), entre otras, CD-ROM, discos versátiles digitales (DVD por las siglas en inglés de Digital Versatile Disc) u otro almacenamiento óptico, casetes magnéticos, cintas magnéticas, almacenamiento o cualquier otro medio que pueda usarse para almacenar información y a la que se puede acceder por una unidad de procesamiento.

Por otro lado, la presente divulgación también se relaciona con un programa de computador que comprende instrucciones, las cuales cuando se ejecuta el programa en un dispositivo de acuerdo con cualquiera de las modalidades anteriormente descritas causa que dicho dispositivo lleve a cabo las etapas de un método de acuerdo con cualquiera de las modalidades de los métodos anteriormente descritos en esta divulgación.

El programa de computador puede estar escrito en cualquier lenguaje de programación o marco de trabajo (framework, en inglés) conocido por una persona medianamente versada en la materia. Por ejemplo, el programa de computador puede estar escrito en Java, JavaScript, Perl, PHP y C++, #C, Python, R-studio SQL, Swift, Ruby, Delphi, Visual Basic, D, HTML, HTML5, CSS, NodeJs, Angular, React, Go, RPA, .NET, Scala, Drupal, Ember, Labview y otros lenguajes de programación y/o marcos de trabajo (framework, en inglés) conocidos por una persona medianamente versada en la materia.

La presente divulgación también presenta un sistema de procesamiento de datos que comprende medios para llevar a cabo las etapas del método:

- a) recibir un dato de georreferenciación (2) de la ubicación de una edificación;
- b) extraer de una base de datos (3) conectada a una unidad de cómputo (1) un parámetro estructural (4) de la edificación;
- c) obtener un dato de tipo estructural (5) mediante un método de clasificación de aprendizaje supervisado no paramétrico (6) que toma como entrada el parámetro estructural (4) de la edificación, donde el dato de tipo estructural (5) representa el sistema constructivo de la edificación;
- d) obtener un dato de estimación de daño sísmico (7) con un método estimación de vulnerabilidad estructural (8) que toma como entrada el dato de tipo estructural (5) y el dato de georreferenciación (2);
donde el método de clasificación (8) aplica una función de vulnerabilidad obtenida a partir de una modelación computacional del comportamiento de la estructura de la edificación.

5 donde la modelación computacional toma como entrada el dato de tipo estructural (5) y el dato de georreferenciación (2) y estima los efectos adversos de una amenaza sísmica en una zona de proximidad localizada en una vecindad del dato de georreferenciación (2) y el medio de almacenamiento legible por computadora que comprende instrucciones que, cuando son ejecutadas por una computadora, hacen que ésta realice los pasos del método de la presente divulgación.

GLOSARIO

10 **Nueva instancia:** es el punto de datos para el cual deseamos realizar una predicción utilizando el modelo K-NN, y no tiene una etiqueta de clase o valor objetivo conocidos en el momento de la predicción.

15 El término "**memoria**" se refiere a un medio legible por máquina capaz de almacenar datos de forma temporal o permanente y puede considerarse que incluye, entre otros, memoria de acceso aleatorio (RAM), memoria de solo lectura (ROM).), memoria intermedia, memoria flash y memoria caché

20 **K-FOLD-CROSS-VALIDATION** es una técnica fundamental para evaluar modelos de aprendizaje automático, incluido KNN. Funciona dividiendo el conjunto de datos en K partes iguales, llamadas "pliegues". Luego, se realiza un ciclo de K iteraciones, donde en cada iteración se entrena el modelo en K-1 de estos pliegues y se evalúa en el pliegue restante. Esto se repite K veces, de modo que cada pliegue actúa como conjunto de prueba en una de las iteraciones.

25 Durante cada iteración, el modelo se ajusta a diferentes conjuntos de entrenamiento y se evalúa en diferentes conjuntos de prueba. Las métricas de rendimiento, como la precisión o el error cuadrático medio, se registran en cada iteración. Después de las K iteraciones, se promedian estas métricas para obtener una estimación general del rendimiento del modelo.

30 La ventaja clave de esta técnica es que proporciona una evaluación más confiable del rendimiento del modelo en comparación con una única división de entrenamiento/prueba. Ayuda a detectar si el modelo está sobre ajustado o sub ajustado y permite una mejor utilización del conjunto de datos en el proceso de entrenamiento y evaluación.

35 **Técnica de búsqueda exhaustiva en cuadrícula (Grid Search):** La técnica de búsqueda exhaustiva en cuadrícula, conocida como "Grid Search," es un método fundamental en el contexto de KNN (K Nearest Neighbors) y otros métodos de clasificación de aprendizaje supervisado no paramétrico (6) de aprendizaje automático. Su propósito es encontrar la mejor combinación de parámetros también llamado hiperparametros para un modelo KNN. Los

parámetros son configuraciones ajustables que influyen en cómo el modelo toma decisiones, y deben establecerse antes de entrenarlo.

5 El proceso comienza identificando qué parámetros del modelo KNN se desean ajustar. Por ejemplo, se pueden considerar el número de vecinos (K), la métrica de distancia (como la distancia euclidiana o la distancia de Manhattan), y la función de peso (que determina cómo se ponderan los vecinos cercanos).

10 Una vez identificados los hiper parámetros, se establecen rangos o valores posibles para cada uno. Por ejemplo, para K, se podría definir un rango [3, 5, 7, 9]. Luego, se crea una cuadrícula que incluye todas las posibles combinaciones de valores de parámetros.

15 El siguiente paso es entrenar y evaluar el modelo KNN para cada combinación en la cuadrícula. Esto implica entrenar el modelo con un conjunto de datos de entrenamiento y evaluar su rendimiento en un conjunto de prueba. Para medir su desempeño, se utilizan técnicas como la validación cruzada o métricas específicas.

20 Al completar la evaluación de todas las combinaciones posibles, se selecciona la que resulta en el mejor rendimiento según la métrica de evaluación elegida. Esta combinación de hiper parámetros se considera la óptima para el modelo KNN en ese conjunto de datos y problema específico.

25 **K-Nearest Neighbors (K-NN):** K-Nearest Neighbors es un método de clasificación de aprendizaje supervisado no paramétrico (6) y regresión. Para clasificación, se basa en asignar una etiqueta a un punto de datos en función de la mayoría de las etiquetas de los "k" puntos de datos más cercanos en un espacio de características. En el contexto del desarrollo de software, K-NN es utilizado para problemas de clasificación y búsqueda de similitud, donde se deben comparar datos en función de sus características numéricas.

30 **Árboles de Decisión:** Los árboles de decisión método de clasificación de aprendizaje supervisado no paramétrico que se utilizan en problemas de clasificación y regresión. Estos árboles dividen los datos en función de las características en nodos y hojas, permitiendo tomar decisiones basadas en condiciones lógicas. En desarrollo de software, se implementan árboles de decisión para automatizar la toma de decisiones en aplicaciones como asistentes virtuales y sistemas de recomendación.

Random Forest: Random Forest es un conjunto de árboles de decisión que se utiliza para mejorar la precisión y reducir el sobreajuste. Cada árbol se entrena con una muestra aleatoria de datos y características, y luego se combinan las predicciones de todos los árboles para

obtener una predicción final. Random Forest es valioso en problemas de clasificación y regresión donde se requiere robustez y generalización.

Support Vector Machines (SVM): Es un método de clasificación de aprendizaje supervisado no paramétrico que encuentra un hiperplano de separación óptimo en un espacio de características de alta dimensión. Su objetivo es maximizar el margen entre las clases de datos. En desarrollo de software, SVM es útil en aplicaciones de clasificación, como detección de spam, reconocimiento de escritura a mano y diagnóstico médico.

Naive Bayes: Naive Bayes es un método de clasificación de aprendizaje supervisado no paramétrico que se basa en el teorema de Bayes y asume independencia condicional entre las características. Este enfoque calcula la probabilidad de que un dato pertenezca a una clase dada y elige la clase con la probabilidad más alta. En desarrollo de software, Naive Bayes es comúnmente utilizado en aplicaciones de procesamiento de lenguaje natural, como clasificación de texto y filtrado de spam.

Redes Neuronales Artificiales (ANN): las redes neuronales artificiales son métodos de clasificación de aprendizaje supervisado no paramétrico inspirados en el funcionamiento del cerebro humano. Consisten en múltiples capas de nodos (neuronas) interconectados que procesan datos a través de funciones de activación. En desarrollo de software, las ANN se utilizan para una amplia gama de aplicaciones, incluyendo visión por computadora, procesamiento de lenguaje natural y juegos.

Vecinos Más Cercanos por Densidad (DBSCAN): es un método de clasificación de aprendizaje supervisado no paramétrico de agrupación que identifica grupos de datos basados en la densidad de los puntos en el espacio, en lugar de asumir una forma específica de agrupación. Esto es útil para detectar grupos de datos en formas no necesariamente esféricas o uniformes. En desarrollo de software, DBSCAN se utiliza en la segmentación de imágenes y la detección de anomalías.

Máquinas de Aprendizaje Extremo (Extreme Learning Machines, ELM): es un enfoque de aprendizaje automático que utiliza redes neuronales de una sola capa oculta y entrena los pesos de forma aleatoria. Esto permite un rápido entrenamiento y se utiliza en tareas de clasificación y regresión. En desarrollo de software, ELM puede ser beneficioso cuando se requiere eficiencia computacional y rapidez en el entrenamiento de modelos.

Gradient Boosting: Un método de clasificación de aprendizaje supervisado no paramétrico que construye un modelo de predicción en forma de conjunto de árboles de decisión, donde cada árbol corrige los errores del anterior.

XG Boost: Una implementación de Gradient Boosting optimizada y eficiente que es ampliamente utilizada en competiciones de aprendizaje automático.

LightGBM: Otro método de clasificación de aprendizaje supervisado no paramétrico de Gradient Boosting que utiliza una técnica de crecimiento de árboles en hojas, lo que lo hace rápido y eficiente en términos de memoria.

Adaboost: Un método de clasificación de aprendizaje supervisado no paramétrico de boosting que ajusta iterativamente el peso de los ejemplos de entrenamiento para enfocarse en los ejemplos mal clasificados.

Máquinas de Aprendizaje Basadas en Instancias (Instance-Based Learning): Son métodos de clasificación de aprendizaje supervisado no paramétrico que toman decisiones basadas en ejemplos individuales en lugar de modelos generales.

Clasificadores Bayesianos Locales (Local Bayesian Classifiers): Método de clasificación de aprendizaje supervisado no paramétrico que aplica el teorema de Bayes a cada dato en un espacio local de características.

Máquinas de Aprendizaje Basadas en Reglas (Rule-Based Learning): Método de clasificación de aprendizaje supervisado no paramétrico que utilizan reglas lógicas para tomar decisiones de clasificación.

Clasificador de Vecinos Más Cercanos Ponderados (Weighted K-NN): Una variante de K-NN que asigna diferentes pesos a los vecinos cercanos.

Clasificación por Regresión Logística Local (Local Logistic Regression): Un método de clasificación de aprendizaje supervisado no paramétrico que combina la regresión logística con un enfoque local para la clasificación.

Clasificadores por Máxima Verosimilitud (Maximum Likelihood Classifiers): método de clasificación de aprendizaje supervisado no paramétrico que asignan una etiqueta basada en la clase más probable según la verosimilitud de los datos.

Clasificación por Máxima Entropía (Maximum Entropy Classification): Un método que asigna una etiqueta basada en la entropía máxima, dadas las restricciones de los datos.

Máquina de Aprendizaje de Votación (Voting Machine Learning): Un enfoque que combina múltiples modelos de aprendizaje para tomar decisiones de clasificación basadas en votos.

Clasificador de Bayes No Paramétrico (Non-Parametric Bayesian Classifier): Un enfoque que utiliza modelos Bayesianos no paramétricos para la clasificación.

Clasificador de Árbol de Decisión No Paramétrico (Non-Parametric Decision Tree Classifier): Una variante de los árboles de decisión que no impone restricciones en la estructura del árbol.

Clasificador de Árboles Aleatorios No Paramétricos (Non-Parametric Random Forests): Una variante de Random Forest que no impone restricciones en la estructura de los árboles.

Clasificación por Regresión Local (Local Regression Classification): Un enfoque que combina técnicas de regresión con un enfoque local para la clasificación.

Clasificación por Redes Bayesianas (Bayesian Networks Classification): Un enfoque que utiliza redes Bayesianas para modelar las relaciones entre las características y tomar decisiones de clasificación.

Dato de georreferenciación (2): La georreferenciación es el proceso mediante el cual se asignan coordenadas geográficas a datos, estableciendo una vinculación entre información específica y ubicaciones precisas en la superficie terrestre. Este procedimiento facilita la representación y análisis espacial de datos, permitiendo su visualización precisa en mapas y su integración efectiva con sistemas de información geográfica. El Dato de georreferenciación (2) es el resultado de este proceso para una coordenada geográfica particular.

Vecindad del dato de georreferenciación (2): se define como un conjunto de datos relacionados con puntos de ubicaciones geográficas que comparten proximidad espacial. son datos que están físicamente cerca unos de otros en el mundo real. La proximidad espacial puede medirse en términos de distancia euclidiana, distancia geodésica u otras métricas geoespaciales.

Base de datos: Una base de datos es un sistema organizado y estructurado para almacenar datos de manera temporal o permanente, siendo un medio legible por máquina. Este concepto abarca diversas formas de almacenamiento, incluyendo bases de datos relacionales y no relacionales. Engloba sistemas como sistemas de gestión de bases de datos (DBMS por sus siglas en inglés), los cuales facilitan la creación, consulta y modificación de datos de manera eficiente y segura.

Entre las variantes se encuentran por ejemplo las bases de datos SQL y NoSQL, cada una con sus propias características y aplicaciones específicas.

Dato de estimación de daño sísmico (7): Es un dato informático que proporciona información evaluativa sobre la posible magnitud de los daños causados por un evento sísmico en estructuras específicas.

5 **Función de vulnerabilidad:** La función de vulnerabilidad describe la relación matemática o analítica entre un peligro específico, como un terremoto o eventos sísmicos, y el grado anticipado de daño o pérdida en una estructura o área determinada. En el contexto de desastres naturales, esta función cuantifica cómo la vulnerabilidad de un sistema o comunidad puede influir en el impacto de un evento, siendo esencial para la evaluación y
10 gestión de riesgos.

Modelación computacional: La modelación computacional se refiere al proceso de utilizar herramientas y técnicas informáticas para crear representaciones virtuales de sistemas o fenómenos del mundo real. En este enfoque, se emplean algoritmos y software especializado
15 para simular el comportamiento, la evolución o la interacción de componentes en un entorno controlado. Este método es ampliamente utilizado en disciplinas como la ingeniería, la ciencia, la meteorología y la investigación para comprender fenómenos complejos, realizar predicciones y tomar decisiones informadas.

Datos de combinaciones (9): Los datos de combinaciones se refieren a conjuntos de información que representan diversas combinaciones de elementos, variables u otros factores. Estos datos implican una serie de combinaciones entre diferentes categorías, valores
20 o condiciones, y su análisis puede proporcionar insights sobre patrones, relaciones o resultados en un conjunto diverso de situaciones o contextos.

25 Para el entendimiento de presente divulgación se entenderá por:

Muro: Un sistema estructural que no dispone de un pórtico esencialmente completo y en el cual las cargas verticales son resistidas por los muros de carga y las fuerzas horizontales son resistidas por muros estructurales o pórticos con diagonales.

30 **Pórtico:** Se define así el conjunto estructural conformado por vigas y columnas unidas en forma rígida y reticular.

Dual: Son estructuras que tienen pórticos combinados con muros estructurales o pórticos arriostrados mediante diagonales, que restringen su deformación lateral en caso de cargas
35 laterales.

Prefabricado: Es una estructura conformada por elementos individuales o paneles previamente contruidos y llevados al sitio, que se conectan conformando entramados o sistemas tridimensionales.

40

Reticular: Es un sistema estructural tridimensional a porticado en el cual las losas cumplen la función de las vigas.

Confinada: Construcción de muros de mampostería de ladrillo o cemento con elementos perimetrales de concreto reforzado de pocas dimensiones (viguetas y columnetas), contruidos alrededor de las paredes conformando anillos que confinan las piezas de mampostería.

Reforzada: Construcción de muros de mampostería con piezas de perforación vertical que se refuerzan horizontalmente en los sitios de pega e internamente con barras de acero en concreto.

No reforzada: Es una construcción de mampostería usualmente de bloques o piezas de ladrillo o cemento que no tiene ningún tipo de refuerzo ni confinamiento mediante elementos estructurales.

Pórticos arriostrados: Es un conjunto estructural constituido por vigas y columnas de alma llena cuya estabilidad lateral se proporciona por medio de riostras diagonales o muros.

Pórticos no arriostrados: Son aquellos cuya estabilidad lateral depende de la rigidez a flexión de las vigas y columnas conectadas rígidamente.

Bahareque: Paredes que son usualmente paneles de madera (o guadua), con o sin relleno de tierra y con recubrimientos diversos.

A lo largo de la presente divulgación, se pueden encontrar situaciones en las que diversas entidades y etapas puedan incorporar elementos, procesos o configuraciones que se describen como una entidad única o etapa. Aunque las operaciones individuales de uno o más métodos se presentan y detallan de manera independiente, es posible que una o varias de estas operaciones individuales se lleven a cabo de forma simultánea, sin que exista una restricción para que se siga el orden presentado. Las estructuras y capacidades que se presentan como entidades separadas en ejemplos de configuración podrían ser implementadas como una entidad o estructura conjunta. De la misma manera, aquellas estructuras y capacidades que se presentan como un solo conjunto podrían ser tratadas como entidades separadas. Todas estas variaciones, ajustes, incorporaciones y mejoras, junto con otras posibles, se encuentran dentro del ámbito de la presente divulgación.

Se debe entender que la presente invención no se halla limitada a las realizaciones descritas e ilustradas, pues como será evidente para una persona versada en el arte, existen variaciones

y modificaciones posibles que no se apartan del espíritu de la presente divulgación, el cual solo se encuentra definido por las siguientes reivindicaciones

.