

Estudio de algoritmos de inteligencia artificial más utilizados para el diagnóstico de diabetes mellitus tipo 2

Study of artificial intelligence algorithms most used for the diagnosis of type 2 diabetes mellitus

Gabriela Cuenca-Andrade¹, René Guamán-Quinche¹, Francisco Álvarez-Pineda¹, Wilman Chamba-Zaragocin¹ y José Guamán-Quinche^{1,*}

¹ Carrera de Ingeniería en Sistemas, Universidad Nacional de Loja, Loja, Ecuador

* Autor para correspondencia: jose.o.guaman@unl.edu.ec

Fecha de recepción del manuscrito: 01/03/2023

Fecha de aceptación del manuscrito: 06/04/2023

Fecha de publicación: 30/06/2023

Resumen—La diabetes es la segunda causa de muerte a nivel mundial sobre todo en países de bajos recursos. En el Ecuador uno de cada diez personas es diagnosticado con diabetes mellitus tipo 2, esto es debido a factores de riesgo como: antecedentes familiares con diabetes, medicación, sedentarismo o mala alimentación. Por lo tanto, es indispensable hacer una Revisión Sistemática de Literatura sobre el estado del uso de técnicas o algoritmos de Inteligencia Artificial para el diagnóstico de diabetes mellitus tipo 2; con este fin de responder la pregunta: ¿Cuáles son las técnicas de inteligencia artificial aplicadas al diagnóstico de diabetes mellitus tipo 2?

Palabras clave—Diabetes tipo 2, Técnicas de inteligencia artificial, Predicción de diabetes

Abstract—Diabetes is the second leading cause of death worldwide, especially in low-income countries. In Ecuador, one in ten people is diagnosed with type 2 diabetes mellitus, this is due to risk factors such as: family history of diabetes, medication, sedentary lifestyle or poor diet. Therefore, it is essential to carry out a Systematic Literature Review on the state of the use of Artificial Intelligence techniques for the diagnosis of type 2 diabetes mellitus; in order to answer the question: What are the artificial intelligence techniques applied to the diagnosis of type 2 diabetes mellitus?

Keywords—Type 2 diabetes, Artificial intelligence techniques, Diabetes prediction

INTRODUCCIÓN

Según la OMS (Organización mundial de la salud) la diabetes es “una enfermedad crónica que aparece cuando el páncreas no produce insulina suficiente o cuando el organismo no utiliza eficazmente la insulina que produce” (S. OMS, 2020), siendo una de las patologías metabólicas con más prevalencia en el mundo aumentando con mayor rapidez en países de ingresos bajos o medios (*Organización Mundial de la Salud, Informe Mundial Sobre la diabetes*, 2016). En 2015, el Atlas de la Diabetes de la Federación Internacional de Diabetes (FID) estima que uno de cada once adultos tiene diabetes representado por una población de 415 millones de personas entre las edades de 20 y 79 años; además estima que en el año 2040 unos 642 millones de personas, entre uno de diez adultos tendrá diabetes (International Diabetes Federation y Séptima, 2015). De acuerdo a Harrison (Sataloff *et al.*, 2015), la clasificación de la diabetes se contempla en cuatro grupos: (i) Diabetes tipo 1 (DMT1) que se caracteriza por la destrucción de las células betas provocando una deficiencia absoluta de insulina, (ii) Diabetes tipo

2 (DMT2) se presenta en personas con grados variables de insulina o deficiencia de ella que puede ser o no predominante, (iii) Diabetes gestacional (DMG) que se caracteriza por la alteración del metabolismo de los hidratos de carbono en el primer trimestre de embarazo y (iv) otros tipos de diabetes que se caracteriza por defectos genéticos, inducida por drogas o químicos o por factores inmunológicos.

En Ecuador, la diabetes afecta a la población con las tasas cada vez más elevadas, siendo la segunda causa de muerte en el país, según la encuesta ENSANUT 2017, la prevalencia de diabetes en la población de 10 a 59 años es de 1.7%, esa proporción va subiendo a partir de los 30 a 50 años de edad, donde uno de cada diez ecuatorianos ya tiene diabetes, lo que a menudo es por consecuencia del estilo de vida que lleva la población (INEC, 2017). Según la revista A.L.A.D. (2019) en el Ecuador el número de casos que padecía diabetes mellitus tipo 2 es 554.500 personas entre una población de 20 y 79 años. Además el diagnóstico de la DMT2 siempre ha presentado inconvenientes en sus resultados, debido a la falta de precisión de los exámenes de laboratorio y la conexión del cuestionario Findrisk (Bello, 2016), relacionado

al estilo de vida del paciente (Toro *et al.*, 2020), a esto también se suma el análisis de los estudios epidemiológicos donde se identifica tres retos a resolver: el número creciente de casos en riesgo, un porcentaje alto de casos no diagnosticados y la efectividad del tratamiento aplicado es insuficiente (A.L.A.D., 2019). Algunas definiciones importantes vemos en anexo 3.

TRABAJOS RELACIONADOS Y MÉTODOS

Según (Shetty y Katkar, 2019a), (Chen *et al.*, 2017), propusieron un modelo que ayuda a los médicos a detectar la enfermedad en etapas tempranas, lo que ayuda a reducir la probabilidad de contraer la diabetes de tipo 2 (Diabetes de Inicio Adulto). El modelo propuesto se construye sobre un conjunto de datos de tiempo real y tres tipos diferentes de algoritmos de árboles de decisión como Simple Cart, J48 y NB Tree, utilizando la herramienta WEKA para el entrenamiento.

En los estudios (V. Kumari *et al.*, 2013), (M. Kumari *et al.*, 2014), (Kadhm *et al.*, 2018), (Sampieri *et al.*, 1997), proponen distintos algoritmos de inteligencia artificial para la predicción/diagnóstico de diabetes tales como SVM, Clasificación Bayesiana, Naive Bayes y KNN (D. Barrios y Infantes, 2018), (Benítez *et al.*, 2013) con base en el análisis de algunas variables tales como número de embarazos, presión arterial diastólica, espesor cutáneo del tríceps, índice de masa corporal, herencia y edad manteniendo la misma proporción entre individuos diabéticos y no diabéticos. Dicho estudio identifica a la diabetes como una enfermedad crónica de mayor riesgo en la salud mundial.

Sin embargo (Sampieri *et al.*, 1997), también realizó un enfoque de clasificación basado en el Árboles de Decisión (Rodríguez, 2017), (Lantz, 2019) para asignar a cada muestra de datos una clase apropiada, es decir realiza una clasificación para la predicción, dicho estudio utilizo como herramienta de software a Matlab.

En otros estudios se han usado la regresión logística (Benítez *et al.*, 2013), máquinas de soporte vectorial (Rodríguez, 2017), (Lantz, 2019), k-menús (Freddy y Viera, 2015), (Ram y Christian, 2016), redes neuronales (D. Barrios y Infantes, 2018), (Wang, 2013), redes bayesianas (Verónica *et al.*, 2015), (Mosquera *et al.*, 2018) y en si otras técnicas de inteligencia artificial para predicción y diagnóstico (Kopitar *et al.*, 2019), (O. Barrios *et al.*, 2017a).

MATERIALES Y MÉTODOS

El objetivo principal planteado es de realizar una revisión sistemática en la aplicación de técnicas o algoritmos de inteligencia artificial para el diagnóstico de diabetes mellitus tipo 2, y, se plantea la siguiente pregunta de investigación: ¿Cuáles son las técnicas de Inteligencia Artificial aplicadas al diagnóstico de diabetes mellitus tipo 2? Para alcanzar el objetivo general, se siguió el protocolo definido por Barbara Kitchenham (Barbara, 2007) (Petticrew y Roberts, 2008), se usó la investigación bibliográfica (Kitchenham, 2004) , aunque existen otras técnicas como bola de nieve (Bocco *et al.*, 2014), (Baltar y Gorjup, 2014), además se debe tener en cuenta las amenazas que puedan existir en las RSL (Carrizo y Moller, 2018), (Bocco *et al.*, 2014), (G, 2005).

A. Formulación de preguntas de investigación

Para establecer las preguntas de investigación en la RSL, se realizó una búsqueda exploratoria que consistió en analizar cuatro artículos seleccionados al azar para analizar los problemas que se presentan al diagnosticar diabetes tipo 2, las soluciones propuestas por los autores y algunas palabras claves. Para finalizar se planteó 3 preguntas del mapeo sistemático (MQ) y 3 preguntas de la revisión sistemática (RQ):

Preguntas del Mapeo Sistemático (MQ)

- MQ1: ¿Cuántos estudios se publicaron en los últimos cinco años en el área de inteligencia artificial aplicada a la detección de la diabetes mellitus tipo 2?
- MQ2: ¿En qué publicaciones se han encontrado estudios relacionados con el tema?
- MQ3: ¿Quiénes son los autores más relevantes y activos en esta área? Artículos más citados.

Preguntas de la Revisión Sistemática (RQ)

- RQ1: ¿Cuáles son las técnicas o algoritmos de inteligencia artificial significativos para la predicción/diagnóstico de la diabetes mellitus tipo 2?
- RQ2: ¿Cuáles son las variables que se tomaron en cuenta para la predicción diagnóstico de la diabetes mellitus tipo 2 en los resultados?
- RQ3: ¿Qué herramientas de software se han aplicado para la predicción/diagnóstico de la diabetes mellitus tipo 2 en los estudios?

B. Estrategia de búsqueda

Se determinó los términos base aplicando el método PICOC propuesto por (Libed *et al.*, 2020) para definir el ámbito de la RSL y se usa para definir el ámbito de la revisión sistemática (Bocco *et al.*, 2014), (G, 2005) Sus componentes son la población, intervención, comparación, resultados y contextos. Este método permitió definir las expresiones que compusieron la cadena de búsqueda. Los términos clasificados se detallan en la siguiente lista:

- Población (P): "Type 2 Diabetes Mellitus".
- Intervención (I): "Intelligence artificial".
- Comparación (C): No aplica.
- Resultados (O): "Algorithms" OR "Classification" OR "Methods" OR "Predictive modeling" OR "Techniques".
- Contexto (C): "Intelligence artificial".

Para completar los términos identificados en PICOC se utilizó las palabras claves detalladas en la Tabla 1. Además, a través del 2020 IEEE Thesaurus se obtuvieron las sinonimias de cada término para la construcción final de la cadena de búsqueda.

El conjunto de términos finales se detalla en la Tabla 2.

Tabla 1: Definición de palabras claves

Título	Palabras claves candidatas
An Accurate Diabetes Prediction System Based on K-means Clustering and Proposed Classification Approach. (Sampieri <i>et al.</i> , 1997).	Classification, Algorithms.
Prediction of Diabetes Using Bayesian Network (Kadhm <i>et al.</i> , 2018)	Technique, Algorithms, Machine learning, Classification.
Type II Diabetes Analysis using Naïve Bayesian Classification Algorithm (Vidhya y Shanmugalakshmi, 2020a)	Type II Diabetes, Machine learning, Algorithms, Prediction.
Deep learning based big medical data analytic model for diabetes complication prediction (Brereton <i>et al.</i> , 2007)	Deep Learning, Diagnostic diabetes, Neural Networks.

Tabla 2: Conjunto de términos

Palabra clave	Sinónimos	Relación
Algorithms.	Classification, Methods, Techniques, Predictive Modeling	Resultados.
Artificial Intelligence	Automated detection, Deep learning, Machine learning, Neural networks Predictive analysis	Intervención.
Type 2 diabetes mellitus	Diagnostic Type II, Diabetes, Prediction Type-2 diabetes Type-II Diabetes	Población.

C. Criterios de inclusión y exclusión

En la búsqueda de artículos se seleccionó aquellos que cumplan los criterios de inclusión y se desechará aquellos que cumplan por lo menos uno de los criterios de exclusión, los mismos que se presentan a continuación.

Criterios de inclusión

- IC1. Artículos provenientes de librerías digitales y fuentes indexadas.
- IC2. Artículos que contengan estudios de técnicas de inteligencia artificial para el diagnóstico y predicción o resultados de análisis comparativos de los algoritmos para el diagnóstico de diabetes mellitus tipo 2.
- IC3. Los artículos que contengan el área de inteligencia artificial y relacionados.

- IC4. Artículos que hayan sido publicado desde el año 2016 al 2021.
- IC5. Artículos que hayan sido publicados en revistas científicas y conferencias.
- IC6. Los artículos que se hayan publicado en el idioma inglés y español.

Criterios de exclusión

- EC1. Artículos duplicados serán excluidos.
- EC2. Artículos que no estén escritos en el idioma inglés.
- EC3. Artículos que hayan sido publicados antes del 2016.
- EC4. Se excluyen artículos donde el contenido sea similar a otros estudios quedando solo estudios de contenido más completo.
- EC5. Se excluyen capítulos de libro, libros, artículos técnicos y resúmenes.
- EC6. Se excluyen artículos cuyo título no tenga relación con el objeto de estudio.
- EC7: Todos los artículos que no pertenecen al área de Ciencias y Computación.

D. Definición de fuentes Bibliográficas o utilizar

Para que la selección de artículos sea de calidad y de rigor, las búsquedas de artículos se realizaron en bases de datos científica en función a su relevancia y al acceso. Por ello, las bases de datos seleccionados se detallan a continuación:

- ACM Digital Library.
- IEEE Digital Library.
- Science@Direct.
- Scopus.

E. Preguntas de calidad

Las siguientes preguntas se establecieron para evaluar la calidad de los artículos preseleccionados:

- QA1: ¿El autor realiza un análisis comparativo entre diversas técnicas de inteligencia artificial para la predicción/diagnóstico de diabetes mellitus tipo 2?.
- QA2: ¿En los estudios se menciona alguna herramienta de software utilizada para la predicción/diagnóstico de diabetes mellitus tipo 2?.
- QA3: ¿En los estudios utilizan conjunto de datos para la predicción/diagnóstico de diabetes mellitus tipo 2?.

Para finalizar, se determinaron los parámetros de puntuación para definir qué artículos serán seleccionados y rechazados. Los parámetros se detallan a continuación:

- Si la respuesta es Si su puntuación será de 1,0.
- Si la respuesta es Parcialmente su puntuación será de 0,5.
- Si la respuesta es No su puntuación será de 0,0.

F. Cadenas de búsqueda

Para la creación de las cadenas de búsqueda, se consideró sinónimos de palabras claves, mediante el Thesaurus de IEEE, de igual manera se utilizó los operadores lógicos "OR/AND", con la finalidad de potenciar la búsqueda. Tomando en cuenta los estudios como: artículos de revistas y conferencias. La cadena de búsqueda se la definió de la siguiente manera:

('Type 2 Diabetes Mellitus' OR 'Diagnostic Type II Diabetes' OR 'Type-II Diabetes') AND ('Intelligence artificial' OR 'machine learning' OR 'Neural Networks') AND ('Algorithms' OR 'Methods' OR 'Techniques')

- ACM Digital Library: [[All: "type 2 diabetes mellitus"] OR [All: "diagnostic type ii diabetes"] OR [All: "prediction type-2 diabetes"] OR [All: "type-ii diabetes"]] AND [[All: "artificial intelligence"] OR [All: "automated detection"] OR [All: "deep learning"] OR [All: "machine learning"] OR [All: "neural networks."] OR [All: "predictive analysis"]] AND [[All: "algorithms"] OR [All: "classification"] OR [All: "methods"] OR [All: "predictive modeling"] OR [All: "techniques"]] AND [Publication Date: (01/02/2016 TO 01/02/2021)].
- IEEE Digital Library: (('Type 2 Diabetes Mellitus' OR 'Diagnostic Type II Diabetes' OR 'Prediction Type-2 diabetes' OR 'Type-II Diabetes') AND ('artificial intelligence' OR 'automated detection' OR 'Deep Learning' OR 'machine learning' OR 'Neural Networks' OR 'Predictive Analysis') AND ('Algorithms' OR 'Classification' OR 'Methods' OR 'Predictive modeling' OR 'Techniques'))).
- Cadena en Scopus: (TITLE-ABS-KEY(('Type 2 Diabetes Mellitus' OR 'Diagnostic Type II Diabetes' OR 'Prediction Type-2 diabetes' OR 'Type-II Diabetes')) AND TITLE-ABS-KEY (('Artificial intelligence' OR 'automated detection' OR 'Deep Learning' OR 'machine learning' OR 'Neural Networks' OR 'Predictive Analysis')) AND TITLE-ABS-KEY (('Algorithms' OR 'Classification' OR 'Methods' OR 'Predictive modeling' OR 'Techniques')) AND (LIMIT-TO (LANGUAGE , 'English')) AND (LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR, 2020) OR LIMIT-TO (PUBYEAR, 2019) OR LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2016)) AND (LIMIT-TO (SUBJAREA , 'COMP'))).
- Cadena en Science@Direct: ('Type 2 Diabetes Mellitus' OR 'Diagnostic Type II Diabetes' OR 'Type-II Diabetes') AND ('artificial intelligence' OR 'machine learning' OR 'Neural Networks') AND ('Algorithms' OR 'Methods' OR 'Techniques').

RESULTADOS

Los resultados para la selección de estudios se detallan en los siguientes pasos:

- Se procedió a realizar la ejecución de las cadenas de búsqueda en cada base de datos seleccionadas de los cuales dio como resultado 224 artículos (datos.bib).

- Se identificaron y eliminaron 23 artículos que estaban duplicados.
- Se procedió a analizar 201 artículos, de cada artículo se leyó los resúmenes, palabras claves y las conclusiones quedando un total de 47 artículos.
- Por último, se aplicó las preguntas de calidad quedando como resultado 35 artículos seleccionados, de los cuales se extrajo la información necesaria para responder las preguntas de investigación. Ver: enlace.

RQ1: ¿Cuáles son las técnicas de inteligencia artificial o algoritmos significativos para el diagnóstico/predicción de la diabetes mellitus tipo 2?

Se identificaron 154 técnicas o algoritmos de inteligencia artificial para la predicción/diagnóstico de la Diabetes Mellitus Tipo 2. Al encontrar una cantidad muy extensa de técnicas o algoritmos, se lista las 8 técnicas/algoritmos más utilizadas por los autores en sus experimentaciones o casos de estudio, como se ilustra en la Figura 1. Además, en la tabla 3 se detalla cada técnica con el número de artículos y el porcentaje que hacen mención de ellas.

Tabla 3: Técnicas o algoritmos más utilizados

Técnica	Nro artículos	Porcentaje
Artificial Neuronal Network (ANN).	17	11.04 %.
Random Forest (RF).	13	8.44 %.
Support Vector Machine (SVM).	13	8.44 %.
Decision Tree (DT).	12	7.79 %.
K-nearest neighbor (KNN).	11	7.14 %.
Logistic Regression (LR).	10	6.49 %.
K-means.	6	3.90 %.
Adaboost.	5	3.25 %.

El algoritmo ANN, en el estudio (Patil *et al.*, 2020) el autor utiliza el algoritmo para predecir y detectar la diabetes a través de variables de impacto del padecimiento de diabetes, dichas variables pertenecen al conjunto de datos de la Pima Indian Diabetes Dataset (PIDD).

Por otro lado, en (Lukmanto *et al.*, 2019) utilizaron al algoritmo ANN como modelo ensamblado con otros algoritmos tales como K-means, RF, DT, SVM y Naive Bayes (NB) debido a que mejora el tiempo de cálculo de la red obteniendo mejores resultados, la aplicación desde este algoritmo con otros obtuvo el 98% de aceptabilidad dicho conjunto de datos está comprendido por 400 personas de diferentes grupos de edad, hábitos alimenticios, culturas, fumadores, no fumadores, bebedores, no bebedores, etc. Un caso similar se presenta en (Derevitskii y Kovalchuk, 2019), que utiliza una técnica de clustering K-Means mejorada para producir centros de cluster (centros de clasificación) en un conjunto de entrenamiento para la red en lugar de todas las instancias del

conjunto de entrenamiento dando precisión de predicción en un 86 % de aceptabilidad dicha técnica se aplicó al conjunto del PIDD conformado por 768 personas.

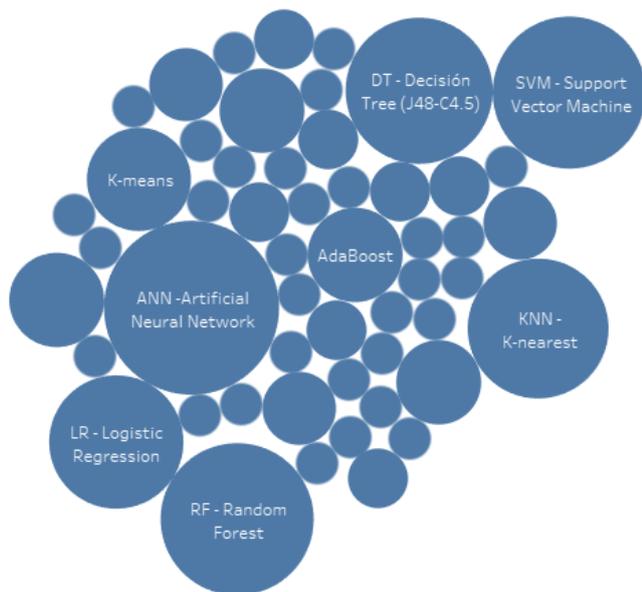


Fig. 1: Algoritmos usados para la detección de diabetes

En cuanto al estudio (Shetty y Katkar, 2019b) utilizo los algoritmos SVM, ANN y RF donde obtuvieron un 90.1 %, 88.02 %, y 83.59 % respectivamente de precisión en la dataset PIDD percibiendo en dicho estudio que la mejor predicción la realizó SVM, dicho estudio desarrolla una comparación de resultados de precisión la cual a su vez propone un modelo híbrido entre los tres algoritmos para una mejor predicción.

Para finalizar, la propuesta descrita en (Kazerouni *et al.*, 2020) realiza una comparación de cuatro algoritmos de Aprendizaje Automático: DT, KNN, ANN y DL en dos diferentes conjuntos de datos, la primera obtenida del hospital de Frankfurt (Alemania), y la segunda es el conocido conjunto de datos de la India Pima; encontrando en primera instancia resultados que usan diferentes métricas y técnicas; con respecto a las técnicas los Arboles de decisión obtiene mejores resultados antes del procesamiento con un 98 % y después del procesamiento 99.5 % KNN mejora la precisión.

RQ2: ¿Cuáles son las variables que se tomaron en cuenta para el diagnóstico/predicción de diabetes mellitus tipo 2 en los resultados?

En cuanto a QR2 se identificaron variables que determinan los factores de riesgo para predicción/diagnóstico de las diabetes definidas en la base de datos de Pima Indian Diabetes Dataset (PIDD) debido a que 21 de los 35 estudios la utilizaron en sus casos de estudios y experimentos. Ver: enlace.

Por lo que se refiere a las variables más determinantes tanto en artículos que aplicaron casos de estudio y experimentaciones para la predicción/diagnóstico de diabetes son Edad (Age), Índice de Masa Corporal (BMI), concentración de glucosa en sangre (Plass), tal como se ilustra en la Figura 2

En estudios como (Ganesh y Sripriya, 2020) contienen AGE, BMI, PLASS estas variables pertenecen a conjuntos de datos distintos al PIDD, tales como Big Medical Data Analy-

Tabla 4: Variables de casos de estudio y experimentos

Descripción	Variable	Nro experimentos	Nro caso estudios
Número de veces embarazadas.	PREG	9	7.
La glucosa en plasma concentración 2 horas en una prueba de tolerancia.	PLASS	12	12.
La presión arterial diastólica (mm Hh).	PRESS	9	10.
Tríceps espesor del pliegue de la piel (mm).	SKIN	9	8.
2 horas suero de insulina (mu U/ml).	TEST	9	7.
Índice de masa corporal.	BMI	12	11.
Función de la diabetes pedigri.	PED	10	8.
Edad.	AGE	10	10.
Diabético.	CLASS	1	8.
Otras variables tomadas en cuenta			
Sexo.	GENDER	4	6.
Fumador.	SMOKING	2	5.
Alcohólico.	DRINKING	2	3.
Sed.	THIRST		3.
Altura.	HEIGHT		3.
Fatiga.	FATIGUE		3.
Peso.	WEIGHT		3.
Antecedentes Familiares.	FAMILY	2	4.
Mala Alimentación.	FOOD	2	4.
Sedentarismo.	STAYIN	2	6.
Depresión.	WEIGHT		3.
Colesterol.	CHOLESTEROL	3	2.

tics y Mostoles Univerity Hospital (Madrid, España), en este estudio excluye la variable glucosa en sangre porque la utilizan para etiquetar los datos (diabetes si/no), y a su vez es considerada como variable ruidosa, también excluyó la edad, ya que la tendencia habría sido invariablemente, utilizando variables como SMOKING, CHOLESTEROL, BMI, TEST para el entrenamiento de los datos con el algoritmo KNN obteniendo una precisión y sensibilidad de predicción del 95 %.

Otro estudio similar es (Vidhya y Shanmugalakshmi, 2020b), realiza una comparación entre dos bases de datos: caso 1) Pima Indian Diabetes que esta comprendido por 8 atributos de varios factores de riesgo clínicos, físicos y epidemiológicos y caso 2) Hippokrateion dataset la cual contiene información relacionada con la demografía, el estilo de vida, los exámenes de laboratorio, las complicaciones o comorbilidades y el tratamiento con un total de 27 atributos, ambos conjuntos de datos contienen edad, indice de masa corporal

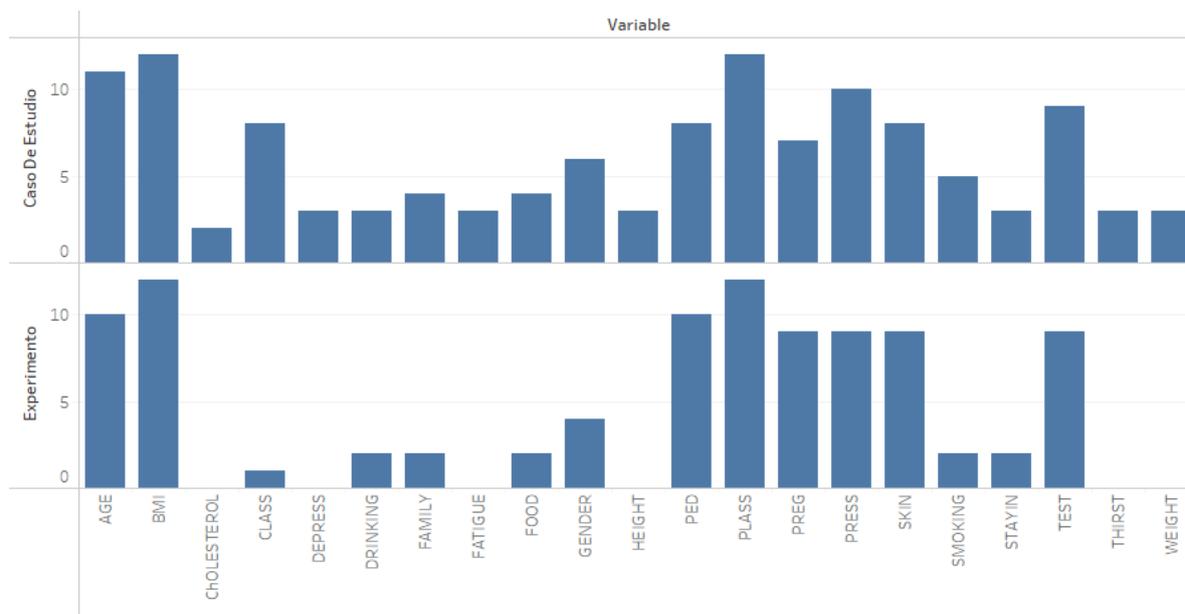


Fig. 2: Variables para la predicción/diabetes mellitus tipo 2

y concentración de glucosa sin embargo se obtuvo mejores resultados con los atributos del caso 1 con un clasificación e precisión del 92.86 % con el algoritmo ensamblado Badding.

RQ3: ¿Qué herramientas de software se han aplicado para el diagnóstico de la diabetes mellitus tipo 2 en los estudios?

Entre las tecnologías identificadas se mencionan las siguientes: Java, Python, Weka, Matlab, SAP y SPSS sin embargo, con un porcentaje del 43 % los artículos seleccionados no mencionan las tecnologías aplicadas, seguido de un porcentaje del 3 % en Java, 6 % Python y R, 11 % para Python y finalmente un 20 % para Weka. También se identifica que en algunos de los estudios utilizan en conjunto dos herramientas de software para aplicar la predicción tal como es el caso de Java más Weka o Matlab más Weka que representan el 3 % (ver Figura 3).

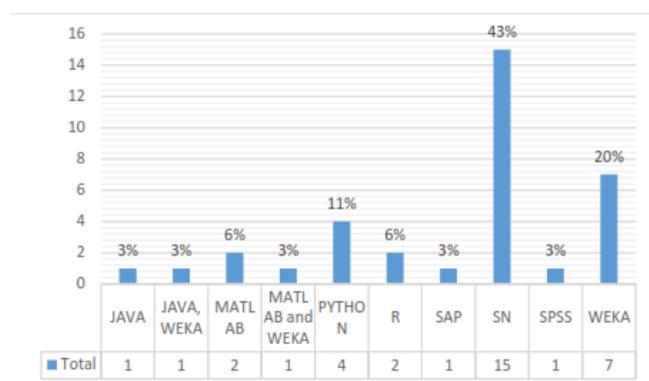


Fig. 3: Herramientas de software

El uso del software WEKA para evaluar la efectividad del modelo propuesto en (Daanouni *et al.*, 2019) están programados en el lenguaje C++, en la que compara el rendimiento de cinco enfoques de clasificación, Antminer, CN2, ANN, Adaboost y Bagging para la predicción de la diabetes mellitus, estos algoritmos se han probado con tres conjuntos de datos

de diabetes de tipo 2 (PIDD, US, AIM'94) donde el algoritmo Antminer ha alcanzado el valor kappa más alto de 0,982, dicho estudio menciona la gran cantidad de métodos que posee la herramienta e indica la personalización de acuerdo a los requisitos del estudio como una de sus mayores ventajas.

Por otra parte (Dong *et al.*, 2019), (Patil *et al.*, 2020) y (P *et al.*, 2020) explicaron que eligieron Python por su sencillez y eficacia, así como por el enorme apoyo de las bibliotecas que permiten realizar estimaciones y predicciones. En los tres artículos antes mencionados utilizan el algoritmo ANN, LR, KNN, SVM, debido a su capacidad para predecir y calcular valores aproximados y funciones interconectadas a partir de un gran número de entradas que pueden ser desconocidas, cabe recalcar que los conjuntos de datos utilizado son distintos a la base de datos PIDD donde se aprecia en los tres estudios realizan la división del conjunto en datos para entrenamiento en un 70 % y el 30 % para pruebas.

Para (Lukmanto *et al.*, 2019) la combinación de Weka y Matlab en algoritmos como ANN, SVM, KNN, Naive Bayes y algoritmos ensamblados permiten entrenar, probar y validar la predicción de diabetes a través de un sistema experto donde utiliza la validación cruzada para dar a conocer el diagnóstico, dicho sistema fue aplicado un conjunto de 400 personas que contenía la información de 11 atributos a través del cuestionario del Test de Findrisk.

En (Durgadevi y Kalpana, 2017) se utiliza el software SPSS para el análisis de los resultados, mientras (Loku *et al.*, 2020) utiliza SAP para predecir la diabetes ya que permite la entrada y transformación de datos para cada variable definida y así aplicar en distintos escenarios, en este estudio realiza la recolección de datos a través de un formulario de Google capturando 656 casos. Por otra parte, en (Singh y Singh, 2020) se usa software RStudio con el lenguaje R donde implementa el algoritmo LR para verificar la relevancia de las variables para una mejor predicción. A pesar que 15 de los artículos no describen la herramienta utilizada en su contenido, 8 de ellos pertenecen a la base datos PIDD como por ejemplo (O. Barrios *et al.*, 2017b), (Wu *et al.*, 2018),

(Alehegn y Joshi, 2019), (Shetty y Katkar, 2019b) donde dichos autores utilizan las 8 variables del conjunto para realizar la predicción a través de la validación cruzada utilizada comúnmente en la herramienta de software WEKA.

CONCLUSIONES

Los algoritmos identificados como mas significativos para la predicción y/o diagnóstico de Diabetes Mellitus Tipo 2 son los siguientes: Artificial Neuronal Network (ANN), Random Forest (RF) Support Vector Machine (SVM), Decision Tree o Árbol de decisión (J48, C4.5), K-nearest neighbor's o vecino más cercano (KNN) y Logistic Regression (LR), de los cuales el mejor algoritmo para detección de diabetes mellitus tipo 2 es el algoritmo ANN debido que realiza el procesamiento de las variables de entrada por capas en un gran conjunto de datos comprendido con datos cuantitativos como cualitativos, mientras que el algoritmo SVM no trabaja con grandes volúmenes de información y requiere de clases dependientes para realizar una correcta predicción, finalmente el algoritmo Deep Learning aún se encuentra en fase de experimentación para la detección de la diabetes mellitus tipo 2, ya que los datos deben ser cuantitativos, sin embargo cabe recalcar que tiene gran aceptación con grandes volúmenes de datos.

Las variables más utilizadas dentro de los casos de estudio como experimentación son: Índice de masa corporal (BMI), concentración de glucosa en sangre (PLASS) y edad sin embargo el ampliar las variables mejora la predicción.

Para finalizar los lenguajes de programación más utilizadas en los estudios seleccionadas fueron: Java y Python, las herramientas más utilizadas dentro del estudio son Weka y Matlab.

REFERENCIAS

A.L.A.D. (2019). Guías alad sobre el diagnóstico, control y tratamiento de la diabetes mellitus tipo 2 con medicina basada en evidencia edición 2019". *Rev. la Asoc. Latinoam. diabetes*, 2019. Descargado de www.revistaalad.com. (En línea]. Disponible en:)

Alehegn, M., y Joshi, R. (2019). Type ii diabetes prediction using combo of svm". *Int.J. Eng. Adv. Technol*, 8, núm. 6, 712–715,. doi: 10.35940/ijeat.F7974.088619.

Baltar, F., y Gorjup, M. (2014). Muestreo mixto online: Una aplicación en poblaciones ocultas". *Intang. Cap*, núm. 1, 2012–2020,. doi: 10.3926/ic.294.

Barbara, K. (2007). *Guidelines for performing systematic literature reviews in software engineering*".

Barrios, D., y Infantes, E. (2018). Modelo predictivo para el diagnóstico de la diabetes mellitus tipo 2 soportado por sap predictive analytics". *UNIVERSIDAD PERUANA DE CIENCIAS APLICADAS*.

Barrios, O., Alberto, D., Infantes, V., Raphael, E., Aguirre, A., y Alexander, J. (2017a). *Predictive modeling for presumptive diagnosis of type 2 diabetes mellitus based on symptomatic analysis*". doi: 10.1109/INTERCON.2017.8079667.

Barrios, O., Alberto, D., Infantes, V., Raphael, E., Aguirre, A., y Alexander, J. (2017b). Predictive modeling for presumptive diagnosis of type 2 diabetes mellitus based on symptomatic analysis". *en 2017 IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing (INTERCON)*, 1–4,. doi: 10.1109/INTERCON.2017.8079667.

Bello, L. (2016). El riesgo de los que cuidan el riesgo: Fin-drisk en personal de blanco". *Rev. virtual Soc. Parag. Med. Int*, 3, núm. 2, 71–76,. doi: 10.18004/rvspmi/.

Benítez, R., Escudero, G., y Kanaan, S. (2013). *Inteligencia artificial avanzada, editorial*. Barcelona.

Bocco, M., Lemus, J., y Velthuis, M. (2014). *Métodos de investigación en ingeniería del software*".

Brereton, P., Kitchenham, B., Budgen, D., Turner, M., y Khalil, M. (2007). Lessons from applying the systematic literature review process within the software engineering domain". *J. Syst. Softw*, 80, núm. 4, 571–583,. doi: 10.1016/j.jss.2006.07.009.

Carrizo, D., y Moller, C. (2018). *Estructuras metodológicas de revisiones sistemáticas de literatura en ingeniería de software: un estudio de mapeo sistemático methodological structures of systematic literature review in software engineering: a systematic mapping study*".

Chen, W., Chen, S., Zhang, H., y Wu, T. (2017). A hybrid prediction model for type 2 diabetes using k-means and decision tree". *Proc. IEEE Int. Conf. Softw. Eng. Serv. Sci. ICSESS*, núm. 61272399, 386–390,. doi: 10.1109/ICSESS.2017.8342938.

Daanouni, O., Cherradi, B., y Tmiri, A. (2019). *Predicting diabetes diseases using mixed data and supervised machine learning algorithms*". doi: 10.1145/3368756.3369072.

Derevitskii, I., y Kovalchuk, S. (2019). Analysis course of the disease of type 2 diabetes patients using markov chains and clustering methods". *Procedia Comput. Sci*, 156, 114–122,. doi: 10.1016/j.procs.2019.08.186.

Dong, Y., Wen, R., Li, Z., Zhang, K., y Zhang, L. (2019). Clu-rnn: A new rnn based approach to diabetic blood glucose prediction". *En en 2019 ieee 7th international conference on bioinformatics and computational biology (icbcb (p. 50–55).* doi: 10.1109/ICBCB.2019.8854670.

Durgadevi, M., y Kalpana, R. (2017). Performance analysis of classification approaches for the prediction of type ii diabetes". *En en 2017 ninth international conference on advanced computing (icoac (p. 339–344).* doi: 10.1109/ICoAC.2017.8441197.

Freddy, , y Viera, G. (2015). *Técnicas de aprendizaje de máquina utilizadas para la minería de texto*" (Vol. 31).

G, (2005). *Revisiones sistemáticas de la literatura*".

Ganesh, P., y SriPriya, P. (2020). A comparative review of prediction methods for pima indians diabetes dataset". *Adv. Intell. Syst. Comput(C)*, 735–750,. doi: 10.1007/978-3-030-37218-7_83.

INEC. (2017). Instituto nacional de estadísticas y censos, diabetes, segunda causa de muerte después de las enfermedades isquémicas del corazón. *INEC*. Descargado de <https://www.ecuadorencifras.gob.ec/diabetes-segunda-causa-de-muertedespues-de-las-enfermedades-isquemicas-del-corazon/> (consultado mar. 15, 2020).

International Diabetes Federation, A., y Séptima. (2015).

Kadhm, M., Ghindawi, I., Of, D.-I., y U. (2018). An accurate diabetes prediction system based on k-means clustering and proposed classification approach". *ripublication.com*, 13, 4038–4041,.

Kazerouni, F., Bayani, A., Asadi, F., Saeidi, L., Parvizi, N., y Mansoori, Z. (2020). Type2 diabetes mellitus prediction using data mining algorithms based on the longnoncoding rnas expression: A comparison of four data mining

- approaches". *BMC Bioinformatics*, 21, núm. 1. doi: 10.1186/s12859-020-03719-8.
- Kitchenham, B. (2004). Procedures for performing systematic reviews". *Br. J. Manag.*, 14, núm. 0, 207–222,. doi: 10.1111/1467-8551.00375.
- Kopitar, L., Cilar, L., Kocbek, P., y Stiglic, G. (2019). Local vs. global interpretability of machine learning models in type 2 diabetes mellitus screening". *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics, 11979 LNAI)*, 108–119,. doi: 10.1007/978-3-03037446-4_9.
- Kumari, M., Vohra, R., y Arora, A. (2014). Prediction of diabetes using bayesian network". *Int. J. Comput. Sci. Inf. Technol.*, 5, núm. 4, 5174–5178,. Descargado de <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.640.3573>. (En línea]. Disponible en:)
- Kumari, V., Chitra, R., Castrillón, O., Sarache, W., y Castaño, E. (2013). Classification of diabetes disease using support vector machine". *Int. J. Eng. Res. Appl.*, 3, núm. 2, 1797–1801,. doi: 10.4067/S0718-07642017000600017
- Lantz, B. (2019). *Machine learning with r* (Vol. 3). Packt Publishing.
- Libed, J., Perreras, R., y Carpio, J. (2020). Type ii diabetes analysis using naïve bayesian classification algorithm". *ACM Int. Conf. Proceeding Ser.*, 35–39,. doi: 10.1145/3424311.3424327.
- Loku, L., Fetaji, B., y Fetaji, M. (2020). *Prevention of diabetes by devising a prediction analytics model*". doi: 10.1109/HORA49412.2020.9152894.
- Lukmanto, R., Suharjito, A., y Akbar, H. (2019). Early detection of diabetes mellitus using feature selection and fuzzy support vector machine". *Procedia Comput. Sci.*, 157, 46–54,. Descargado de <https://doi.org/10.1016/j.procs.2019.08.140>. doi: 10.1016/j.procs.2019.08.140.
- Mosquera, R., Castrillón, O., y Parra, L. (2018). Máquinas de soporte vectorial, clasificador naïve bayes y algoritmos genéticos para la predicción de riesgos psicosociales en docentes de colegios públicos colombianos". *Inf. tecnológica*, 29, núm. 6, 153–162,. doi: 10.4067/s0718-07642018000600153.
- Organizacion mundial de la salud, informe mundial sobre la diabetes.* (2016).
- P, B., R, S., K, N., y K, A. (2020). Type 2: Diabetes mellitus prediction using deep neural networks classifier". *Int. J. Cogn. Comput. Eng.*, 1, 55–61,. Descargado de <https://doi.org/10.1016/j.ijcce.2020.10.002>. doi: 10.1016/j.ijcce.2020.10.002.
- Patil, R., Tamane, S., y Patil, K. (2020). Self organising fuzzy logic classifier for predicting type-2 diabetes mellitus using aco-ann". *Int. J. Adv. Comput. Sci. Appl.*, 11, núm. 7, 348–353,. doi: 10.14569/IJACSA.2020.0110746.
- Petticrew, M., y Roberts, H. (2008). *Systematic reviews in the social sciences: A practical guide.*
- Ram, G., y Christian, S. (2016). *Agrupamiento de textos cortos en dominios cruzados cross-domain clustering for short texts*" (Vol. 115).
- Rodríguez, F. (2017). *Métodos de clasificación estadística y machine learning con r*". en *Estadística y Machine Learning con R.*
- Sampieri, R., Collado, C., y Pilar Baptista Lucio, M. (1997). *Metodología de la investigación* (5ta Ed" ed.).
- Sataloff, R., Johns, M., y Kost, K. (2015). *Obesidad, diabetes mellitus y síndrome metabólico*". En *en Harrison principios de medicina interna, novena* (Vol. 2, p. 2392–2449).
- Shetty, G., y Katkar, V. (2019a). *Type-ii diabetes detection using decision-tree based ensemble of classifiers*". doi: 10.1109/ICCUBEA47591.2019.9129348.
- Shetty, G., y Katkar, V. (2019b). *Type-ii diabetes detection using decision-tree based ensemble of classifiers*". En *en 2019 5th international conference on computing, communication, control and automation (iccubea* (p. 1–5). doi: 10.1109/ICCUBEA47591.2019.9129348
- Singh, N., y Singh, P. (2020). A stacked generalization approach for diagnosis and prediction of type 2 diabetes mellitus". *Adv. Intell. Syst. Comput.*, 990, 559–570,. doi: 10.1007/978-981-13-8676-3_47.
- S. OMS, O. (2020). *Diabetes*". Descargado de <https://www.who.int/es/news-room/fact-sheets/detail/diabetes>
- Toro, J., Giraldo, S., y Plaza, J. (2020). Utilización de algoritmos de clasificación bayesiana y sistemas inteligentes para el desarrollo de un prototipo software para el diagnóstico de diabetes mellitus tipo 2". en *Investigación Formativa en Ingeniería, Cuarta Edi., Medellín*, 295.
- Verónica, N., Pérez, R., Estrada, M., Miriam, A., y Tovar, D. (2015). Aplicación métodos de inteligencia artificial en el área médica". *Pist. Educ. Inst. Tecnológico Celaya.*, núm, 111, 124–130..
- Vidhya, K., y Shanmugalakshmi, R. (2020a). Deep learning based big medical data analytic model for diabetes complication prediction". *J. Ambient Intell. Humaniz. Comput.*, 11, núm. 11, 5691–5702,. doi: 10.1007/s12652-020-01930-2.
- Vidhya, K., y Shanmugalakshmi, R. (2020b). Deep learning based big medical data analytic model for diabetes complication prediction". *J. Ambient Intell. Humaniz. Comput.*, 11, núm. 11, 5691–5702,. doi: 10.1007/s12652-020-01930-2.
- Wang, C. (2013). Evaluating the risk of type 2 diabetes mellitus using artificial neural network: An effective classification approach". *Diabetes Res. Clin. Pract.*, 100, núm. 1, 111–118,. doi: 10.1016/j.diabres.2013.01.023.
- Wu, H., Yang, S., Huang, Z., He, J., y Wang, X. (2018). Type 2 diabetes mellitus prediction model based on data mining". *Informatics Med. Unlocked*, 10, 100–107,. doi: 10.1016/j.imu.2017.12.006.