



**UNIVERSIDAD POLITÉCNICA SALESIANA
SEDE GUAYAQUIL
CARRERA DE INGENIERÍA DE SISTEMAS**

**Modelo de predicción para enfermedades de insuficiencia cardiaca mediante razonamiento
basado en casos Machine Learning**

Trabajo de titulación previo a la obtención del
Título de Ingeniero de Sistemas

AUTOR: KERLY JOHANA RODRÍGUEZ LOPEZ

TUTOR: JOE LLERENA IZQUIERDO

Guayaquil – Ecuador

2024

**CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE
TITULACIÓN**

Yo, Kerly Johana Rodríguez López con documento de identificación N° 0931421721 manifiesto que:

Soy el autor y responsable del presente trabajo; y, autorizo a que sin fines de lucro la Universidad Politécnica Salesiana pueda usar, difundir, reproducir o publicar de manera total o parcial el presente trabajo de titulación.

Guayaquil, 21 de febrero del año 2024

Atentamente,



Kerly Johana Rodríguez López

0931421721

**CERTIFICADO DE CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE
TITULACIÓN A LA UNIVERSIDAD POLITÉCNICA SALESIANA**

Yo, Kerly Johana Rodríguez López con documento de identificación No. 0931421721, expreso mi voluntad y por medio del presente documento cedo a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que soy autor(a) del Artículo Académico: “Modelo de predicción para enfermedades de insuficiencia cardiaca mediante razonamiento basado en casos con Machine Learning” , el cual ha sido desarrollado para optar por el título de: Ingeniero de Sistemas, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En concordancia con lo manifestado, suscribo este documento en el momento que hago la entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Guayaquil, 21 de febrero del año 2024

Atentamente,



Kerly Johana Rodríguez López

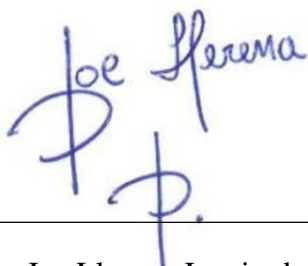
0931421721

CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN

Yo, Joe Llerena Izquierdo con documento de identificación N° 0814884879 docente de la Universidad Politécnica Salesiana, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: **Modelo de predicción para enfermedades de insuficiencia cardiaca mediante razonamiento basado en casos con Machine Learning**, realizado por Kerly Johana Rodríguez López con documento de identificación N° 0931421721, obteniendo como resultado final el trabajo de titulación bajo la opción Artículo Académico que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Guayaquil, 21 de febrero del año 2024

Atentamente,



Joe Llerena-Izquierdo

0914884879

DEDICATORIA

A Dios, por estar presente en cada uno de mis pasos, A mi Madre por su amor incondicional, a mi Padre orientarme por el camino del bien desde que tengo uso de razón, a mi Hermano por ser un apoyo fundamental a lo largo de este camino.

AGRADECIMIENTO

A la Universidad Politécnica Salesiana por todos los conocimientos adquiridos a lo largo de este camino, que espero poder ponerlos en práctica en mi etapa como profesional.

A la Centro Médico Redima “Red De Dispensarios Médicos de la Arquidiócesis” por su colaboración en la implementación del algoritmo de Razonamiento Basado en Casos

RESUMEN

En la actualidad Machine Learning es un campo importante para la Inteligencia Artificial (IA) en la que por medio de un algoritmo permite que los dispositivos aprendan sin que haya la necesidad de ser programadas explícitamente para la solución de una tarea en específica. El Razonamiento Basado en Casos (RBC), es aquel proceso que trata de simular el comportamiento humano referente a la toma de decisiones en base a un caso y con ello aprender de la experiencia de ese caso , y a su vez ha demostrado ser una herramienta útil para el campo médico ya que ayuda diagnóstico enfermedades como lo establece (Escobar, 2019)

La presente investigación es de tipo cuasiexperimental, ya que, busca un control total de las variables de estudio, usando grupos experimentales y de control. En la cual el enfoque, se considera una investigación mixta, ya que, se integrará, de manera sistemática, los datos cuantitativos y cualitativos, debido a que las historias clínicas necesarias para registrar el historial del paciente contienen datos numéricos y descriptivos. Para mejorar el algoritmo y probar, se tomarán datos reales a los pacientes de Cardiología del Centro Médico REDIMA (Red de Dispensarios Médicos de la Arquidiócesis).

Adicional a esto se comprobó que los resultados que surgieron por el algoritmo implementando en el software Google Colab es de 98%, lo que contribuye de manera positiva al campo de la medicina, ya que mediante este ayuda a los médicos a obtener con exactitud un diagnóstico y a su vez aplicar un tratamiento eficaz.

Palabras claves: Inteligencia Artificial, Razonamiento Basado en Casos ,Diagnóstico médico, Software

ABSTRACT

Machine Learning is an important field for Artificial Intelligence (AI) in which, through an algorithm, it allows devices to learn without the need to be explicitly programmed to solve a specific task. Based Reasoning in Cases (RBC) is that process that tries to simulate human behavior regarding decision making based on a case and thereby learn from the experience of that case. It has proven to be an important tool for the medical field since it helps diagnose patients through a classification of presence or attention of disease.

The present investigation is of a quasi-experimental Type, since its weeks totals control of the study variables, using experimental and control groups. In which the approach is considered a mixed investigation, since quantitative and qualitative data will be systematically integrated, because the medical records necessary to record the patient's history contain numerical and descriptive data, algorithm and test, real data will be taken from the Cardiology patients of the REDIMA medical center.

In addition to this, it was found that the results that arose from the algorithm implemented in the Google Colab software is 98%, which contributes positively to the field of medicine, since it helps doctors to accurately obtain a diagnosis, and in turn apply an effective treatment.

Key words: Artificial Intelligence, Based Reasoning in Case, Medical Diagnosis, Software.

ÍNDICE DE CONTENIDO

1.	INTRODUCCIÓN.....	11
2.	REVISIÓN DE LITERATURA.....	12
2.1.	Razonamiento basado en Casos	12
2.1.1.	Razonamiento basado en casos en la vida cotidiana	14
2.1.2.	Elementos del razonamiento basado en casos	14
2.1.2.1.	El caso	14
2.1.2.2.	Librería de Casos.....	15
2.1.2.3.	Similitud	16
2.1.2.4.	Ciclo del Razonamiento Basado en Casos	17
2.1.2.5.	Fase de Recuperación	17
2.1.2.6.	Fase de Reutilización.....	18
2.1.2.7.	Fase de Revisión.....	19
2.1.2.8.	Fase de Retención.....	19
2.2.	Descripción del modelo propuesto	20
2.2.1.	Descripción de datos	20
2.2.2.	Sistema de clasificación	20
2.2.2.1.	Algoritmo k-NN	21
2.2.2.2.	Recuperación inductiva	21
2.2.2.3.	Software para el razonamiento basado en casos.....	21
2.3.	Enfermedad de insuficiencia cardíaca	22
2.3.1.	Concepto.....	22
2.3.2.	Causas de la Insuficiencia Cardiaca	22
2.3.3.	Diagnósticos estandarizados.....	23
2.3.4.	Estadísticas de salud.....	24
3.	METODOLOGÍA	25
3.1.	Métodos y técnicas de Recopilación de datos empleadas	25
3.2.	Métodos y técnicas de Análisis de datos	26
4.	RESULTADOS.....	26
5.	DISCUSION.....	29
6.	CONCLUSION	31
7.	REFERENCIAS	32

ANEXO.....	34
Anexo 1	34
Anexo 2	35
Anexo 3	35

1. INTRODUCCIÓN

El aprendizaje surge a partir de la experiencia en la cual se le da de forma natural a los seres humanos, mediante esto el conocimiento conseguido se convierte en un apoyo para soluciones a problemas habituales como lo indica (Serrano Buenano, 2023). En un estudio de (Rashid, 2023) para aparentar copiar esta capacidad del ser humano, nace el Razonamiento Basado en Casos conocido por siglas en español como RBC, que se basa primordialmente en el proceso de la imitación de la toma de decisiones con respecto algún tema específico y aprender de la experiencia de casos anteriores.

En base al estudio de (Sanchez-Romero & Llerena-Izquierdo, 2023) el RBC cuando el sistema se enfrenta a una situación, analiza soluciones que se aplicaron en casos anteriores y mediante esto las utiliza como punto de partida en la solución. Los nuevos casos se resuelven adaptando soluciones que ya fueron utilizadas con éxito, por lo tanto, es importante usar el razonamiento ya que permite resultados y conclusiones para la comprensión de una situación refiriéndose a otra.

En el campo de la salud, el sistema de RBC sirve para todas las tareas en especial para el diagnóstico, mediante una librería de casos, por medio de características relevantes de un caso, el RBC muestra el estado del paciente y guarda los casos anteriores como ejemplos para resolver casos nuevos, el sistema reconoce la igualdad de los casos nuevos con los casos antiguos, en donde la información se registra mediante el dominio que se asociará el caso como lo expresa (Barreto Vite, 2022; Macias Vera, 2023).

Es importante el análisis de datos para este tipo de operaciones, tal como lo expresa (Aguilar et al., 2021), menciona que el análisis inteligente de datos se define como un proceso complejo de descubrimiento de información o conocimiento útil de los datos. En otras palabras, el análisis inteligente de datos es un proceso no trivial de identificación válida, novedosa y potencialmente útil de los patrones comprensibles en los datos.

Esta investigación tiene como objetivo de aportar con sus hallazgos en el ámbito del uso de algoritmos de *machine learning* a los especialistas de la salud, ya que el uso de sistemas expertos, en este caso el RBC, permiten dar soporte en el diagnóstico médico procurando servir como prevención.

2. REVISIÓN DE LITERATURA

El aprendizaje de máquina mejor conocido como Machine Learning, forma parte de uno de los campos de la Inteligencia Artificial, en la que, mediante el desarrollo de algoritmos, permite a las computadoras y dispositivos, aprender por medio de datos, sin que exista la necesidad de ser programadas explícitamente para la resolución de una tarea en particular

Como lo indica un estudio de (Cedeño Pinargote & Chinga Quimis, 2020; Valdiviezo et al., 2023) El Razonamiento Basado en Casos, es aquel sistema que se basa en el funcionamiento mediante experiencias, ya sea, desde el punto de vista humana o propio del sistema, y mediante este, se relaciona la extracción de una solución para resolver el problema

El RBC, forma parte de una de las técnicas de Machine Learning, que puede trabajar en conjunto con datos previos de entrenamiento para la resolución de problemas y tomas de decisiones.

Para el ingreso de datos al algoritmo, es necesario contar con un ‘dataset’ de historias clínicas que contengan ciertos factores determinantes para emitir un diagnóstico más cercanos a la realidad.

2.1. Razonamiento basado en Casos

Como lo expresa (Homem et al., 2020; Reinoso Ordóñez, 2021) desde épocas anteriores se ha estudiado la forma en la que piensa el hombre, y posteriormente, se toma beneficio de las experiencias en las actividades cotidianas y los procesos que ayudan a razonar basadas en experiencias. Por medio de un estudio de (Leake et al., 2021; Llerena-Izquierdo & Merino-Lazo, 2021) el RBC va de la mano con este concepto, pero la diferencia radica en que, dichas experiencias y el aprendizaje lo realiza una máquina, como parte de la Inteligencia Artificial

El Razonamiento basado en Casos, tiene numerosos casos de éxito en su aplicación, los mismos que han sido revisados para el objeto del marco teórico del presente trabajo, entre ellos, existe un proyecto inicial en RBC, que se basa en un sistema llamado *Lockheed's Clavier* que consiste en presentar piezas compuestas que se funden en un horno industrial de convección como lo indica (Flor Alvarado & Rodríguez Hurtado, 2021). En dicho trabajo, se ha probado con conjuntos de datos reales, y se aplican a diferentes sistemas de análisis de datos, para brindar

un diagnóstico muy cercano a lo que podría predecir un médico mediante la respectiva revisión clínica.

Por lo tanto, estos trabajos sirven de referencia por su manera de retroalimentar los datos e ingresarlos para la perfección del algoritmo.

La estructura de un RBC se basa principalmente en:

- **ENTRADA DEL PROBLEMA:** se basa en el caso en el que se enfrentará el sistema.
- **SOLUCIÓN:** aplicación del proceso en el cual se adapta y se reutiliza el caso anterior para la resolución el problema actual.
- **BASE DE CASOS:** es utilizada por el mecanismo de razonamiento para obtener la solución correcta o aproximada.

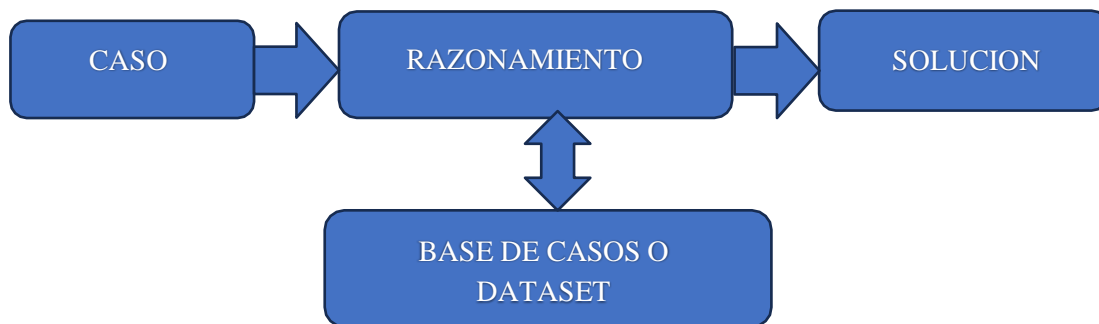


Figura 1. Estructura de Razonamiento Basado en Casos

Ventajas de utilizar RBC

Las ventajas que se da al utilizar RBC son las siguientes:

- Permite la reducción de tareas
- No se necesita de todos los pasos para una solución concreta
- Éxitos en las soluciones planteadas
- Razonamiento en un dominio con un conocimiento pequeño.

Con esta técnica se resuelven los casos y nuevas problemáticas de entrada, mediante la recuperación de datos basados en uno o más casos, reutilizando el caso de una manera,

revisando la solución dada y guardando la nueva experiencia en la base de casos o dataset, para mejorar el algoritmo y tener datos más precisos con cada retroalimentación.

2.1.1. Razonamiento basado en casos en la vida cotidiana

El RBC se basa en la utilización de experiencias para resolver problemas, cuando un sistema quiere solucionar un problema recuerda problemas similares en las que las soluciones han funcionado bien y esta las utiliza como punto de partida en la resolución.

El RBC se puede utilizar la experiencia previa para:

- Originar soluciones a problemas nuevos por medio de soluciones que funcionaron bien con otros casos parecidos.
- Evitar errores cometidos en el pasado, utilización de casos pasados para soluciones propuestas y si es necesario, corregirlas.

2.1.2. Elementos del razonamiento basado en casos

2.1.2.1. El caso

Se lo representa por medio de una situación que se ha suscitado en el pasado y que se lo puede utilizar como experiencia para la resolución de problemas similares en el futuro

Un caso brinda información sobre el problema, la solución que se le encontró, y el contexto por lo que se aplicó la solución. Esta información es almacenada en una base de casos y es utilizada para guiar el proceso de resolución de problemas en un futuro.

Por ejemplo, en el campo médico, el caso puede ser la historia clínica en el que un paciente presentó síntomas similares al de nuevo paciente. Al utilizar la información del caso anterior, el médico puede diagnosticar más rápido y con un enfoque más preciso al nuevo paciente.

Un caso se compone de los siguientes componentes:

Descripción de problema: problema que se desea resolver, y de este, se determina la similitud con otros.

Descripción de solución: estrategias que permiten mejorar la eficiencia y precisión de la búsqueda de casos.

Resultados: en este se determina si la solución elegida es una estrategia para la solución del problema.

En el presente trabajo investigativo se tomará referencia a un determinado paciente con determinados síntomas para una enfermedad específica, en el cual, se va a establecer el estudio, para que la solución dada sea precisa y refleje lo que realmente sucede en un contexto determinado.

2.1.2.2. Librería de Casos

Los beneficios que brindan RBC:

Aprendizaje y Desarrollo del sistema: proporciona casos que el sistema puede usar para aprender y desarrollar estrategias, al momento del análisis y adaptación de casos, este puede mejorar para la toma de decisiones más precisas y eficientes.

Generalización: Los casos que se encuentran en diferentes librerías, ya desarrolladas, o por desarrollar, pueden cubrir una variada amplitud de situaciones.

La librería en el RBC puede poseer un gran número de casos que se han recopilado en la historia o a su vez contiene las experiencias como lo indica (Torres Cordero et al., 2019) que permite tener una organización sistemática para facilitar su acceso y búsqueda. En el campo médico, las librerías contienen casos de pacientes con síntomas y diagnósticos iguales, por lo que la utilización de esta sirve para encontrar casos similares, mediante esto, se ayudaría a los profesionales de la salud con la toma de decisiones más precisas y acertadas.

Indexación de casos

La indexación de casos en RBC, es el procedimiento de etiquetado y organización de casos dentro una librería de casos, y es un proceso de suma importancia, ya que, en base a esto, se puede realizar la búsqueda y recuperación de casos cuando se requiere resolver un problema o situación.

Para realizar la indexación se debe hacer uso de distintas etiquetas o términos de búsqueda en la que se describen los aspectos claves del problema o situación particular.

Algunos ejemplos de campos para la base de caso pueden ser Edad, Género, Resultados, etc. Una vez realizada la indexación, esto ayuda al proceso de la resolución del problema, ya que, permite que los médicos en general encuentren con rapidez los casos relevantes y las soluciones que se han aplicado anteriormente.

La información que se obtenga de un caso, da una porción de conocimiento que será necesario guardarlo en la base de casos para determinar cuándo se aplica nuevamente ese conocimiento y de vital importancia para el proceso de mejora continua del algoritmo.

2.1.2.3. Similitud

Si los datos del problema proporcionado por el usuario no se encuentran en la base de datos se da la similitud como lo expresa en su estudio (Rachmawati et al., 2022), en la cual cuando hay dos casos en términos de características o atributos relevantes similares, permite que el caso actual se resuelva de una forma más rápida

Existen técnicas diferentes para la similitud de casos, en la que una de las más utilizada es la medida de distancia, que dice que los casos más similares se encuentran más cerca en un espacio de características, una de ella es la distancia euclidiana, que realiza la medición de la distancia directa entre dos puntos en un espacio de características. Otra más común, es el coeficiente de similitud de Jaccard, que mide la similitud de dos conjuntos de características, y se utiliza en los casos en que las características no son numéricas, sino binarias o categóricas.

Adaptación del RBC

Este proceso juega un rol importante en los sistemas RBC, ya que, identifica casos existentes que sean relevantes para el nuevo problema y los adapta para generar la solución del problema actual.

Para llegar a esto, el sistema de RBC realiza un conjunto de tareas que son: identificación de las partes relevantes del caso existente, modificación de la solución existente o la combinación de la solución existente con la solución generada por medio del razonamiento.

La adaptación de RBC es fundamental, ya que, permite que sea efectiva a medida que se enfrenta a los cambios en su entorno como nuevas circunstancias o variaciones. Uno de los aspectos claves de la adaptación es la detección de cambios, ya que, va a detectar los cambios

que ocurren o los casos que se están manejando, incluyendo cambios en las características de los casos, ya sean emergentes o situaciones inesperadas.

Aprendizaje del RBC

En el transcurso del aprendizaje, el sistema debe de realizar algún juicio para valorar el rendimiento, y después, el sistema se actualiza con la información adquirida sobre la solución, mediante esto, el sistema va aprendiendo. Al momento de almacenarse, se aumenta la probabilidad de obtener un caso similar o igual y se convertiría en un sistema más completo.

El aprendizaje no solo ocurre cuando el caso ha sido solucionado con éxito, sino cuando el sistema permita almacenar las causas del error.

2.1.2.4. Ciclo del Razonamiento Basado en Casos

Como lo establece un estudio realizado por (Barreto Vite, 2022) el Razonamiento Basado en Casos se da por medio de un ciclo en el cual consta de cuatro fases que se repiten de manera continua mientras se ingresen datos adicionales. A continuación, serán mencionadas las cuatro fases:

2.1.2.5. Fase de Recuperación

En esta fase, el RBC realiza la búsqueda por medio de la base de datos en el que se encuentran casos anteriores que sean similares al nuevo problema que se quiere resolver. Esta se basa en la selección de casos más relevantes para el problema actual, con atributos relevantes de los casos que han sido seleccionados mediante la representación de casos.

Una vez realizado esto, el sistema los utiliza para adaptarlos al caso actual y se genera la solución para el problema. Es de suma importancia destacar que, a medida que se sumen casos y se mejora la medida de similitud utilizada, la calidad de soluciones será más efectiva.

Las técnicas para medir la similitud:

- Técnica del **vecino más cercano** “k-NN” k-nearest neighbors
- Técnica de recuperación inductiva

La similitud permite recuperar uno o más casos, donde cada caso va entre 0 a 1, en la cual, 0, indica que no hay similitud entre casos y el 1, muestra la coincidencia exacta del caso nuevo

con los casos de la base de datos, por lo que, emplean varios algoritmos de recuperación en RBC, siendo el más conocido, el algoritmo k-NN o técnica del vecino más cercano.

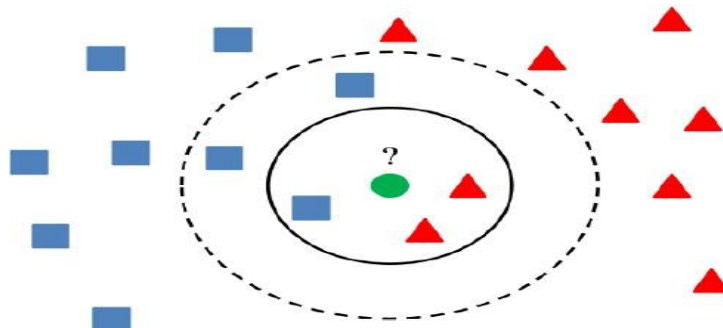


Figura 2. Algoritmo k-NN (vecino más cercano)

2.1.2.6. Fase de Reutilización

Por medio de la recuperación de casos relevantes, el sistema RBC maneja esta información para generar una solución al problema nuevo, este adapta los casos para ajustarlos al problema y se lo realiza mediante las similitudes y diferencias de los casos relevantes, el problema actual y la modificación de casos relevantes para adaptarse al problema actual.

Al momento de haber adaptado los casos relevantes, se utiliza la información para así poder generar la solución al problema actual y se lo realiza por medio de la transferencia de la solución, que puede ser directa o requerir la modificación de la solución para adaptarse al problema actual.

Esta es la más importante del ciclo de RBC, ya que, se aprovecha la experiencia adquirida en casos anteriores para la resolución de problemas nuevos, y al momento de reutilizar la solución, se ahorra tiempo y esfuerzo, mejorando la calidad de las soluciones dadas.

Tipos de adaptación:

- Adaptación Estructural: se aplica reglas de adaptación a la solución generada
- Adaptación Derivacional: reutiliza métodos que generaron la solución

2.1.2.7. Fase de Revisión

En esta parte se evalúa la solución propuesta, la cual, hace una comparación de la solución con las restricciones del nuevo problema y se verifica si cumple con ellas, así como también se evalúa la calidad.

Por ejemplo, se realizan las comparaciones con casos similares que se encuentran en la base de casos e indica si la solución es la mejor opción. De ser el caso, si la solución no es favorable, se realiza la fase de reutilización, adapta los casos relevantes y propone una solución mejor. Si la solución es favorable se registra como un nuevo caso para el uso en futuros problemas.

2.1.2.8. Fase de Retención

Es la etapa final del ciclo RBC y se basa en la actualización de la base de casos que contiene la información aprendida durante el proceso de resolución del problema actual. Consiste en el almacenamiento del caso resuelto en la base de casos con la solución generada, y cualquier otro dato relevante como la descripción del problema y pasos seguidos para resolverlo.

Es de suma importancia, ya que, permite que el RBC aprenda y mejore con el transcurso del tiempo, si se resuelven más problemas y se agreguen más casos, el sistema se vuelve más experto y con ello, se generan soluciones de alta calidad basándose en la experiencia.

En la siguiente figura se observa el ciclo RBC como lo indica (Barreto Vite, 2022).

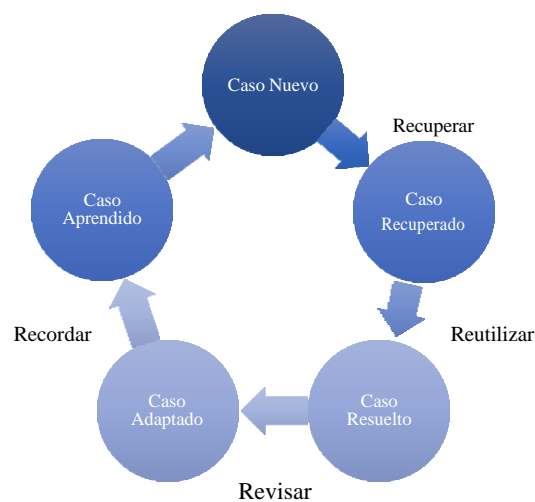


Figura 3. Ciclo de Razonamiento Basado en Casos

Existe 4 contenedores en el RBC que son:

Contenedor de base de conocimientos: que se encuentra en los casos.

Contenedor de las medidas de similaridad: mediante los algoritmos de similaridad.

Adaptación del conocimiento: contiene información para que un caso

Vocabulario: características de los casos

2.2. Descripción del modelo propuesto

Este modelo va dirigido al diagnóstico médico que está determinado en dos capas principales. Esto consiste principalmente, como primer punto, en una base de casos y sus relaciones, y el segundo punto se basa en el razonamiento para la toma de decisiones que se denomina máquina de interferencia, que tendrá acceso a la base evaluando todos los casos de pacientes por medio de sus síntomas.

2.2.1. Descripción de datos

Es de suma importancia mencionar que los datos recolectados se basan en la enfermedad de insuficiencia cardiaca, contenidos en una base de datos obtenida en la siguiente dirección (<https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>) que cuenta con más de 500 registros de pacientes, tanto femenino y masculino, en español, y dichas personas objetos de estudio cuentan con factores predisponentes como lo son: colesterol sérico que se encuentra en la sangre, glucemia en la sangre, electrocardiograma en reposo, entre otros datos relevantes para el estudio.

2.2.2. Sistema de clasificación

Para la extracción de conocimientos por medio de la base de datos existe la aparición de KDD conocido como Knowledge Discovery from Databases, en la cual, existe una etapa conocida como Minería de Datos, que consiste en crear un modelo basado en los datos para poder realizar predicciones y entender mejor los datos o explicar situaciones pasadas según (Berástegui & Galar, 2018).

El aprendizaje consiste en dos tipos:

- Aprendizaje controlado: se da por medio de conjuntos de instancias pre-etiquetadas, esto quiere decir que es un método para predecir instancias.
- Aprendizaje no supervisado: este método es capaz de conseguir una estructura de datos, que permita organizarlos de algún modo.

2.2.2.1. Algoritmo k-NN

El algoritmo k-NN por sus siglas en inglés k-nearest neighbors como lo expresa en su estudio (Berástegui Arbeloa, 2018), es el código más utilizado en comparación con el k-medidas mientras que el estudio de (Varshavskii et al., 2020) indica, que a través del uso de un conjunto de elementos que se encuentren ya seleccionados conocido como el conjunto de entrenamiento o train, permite que se pueda clasificar los nuevos ejemplos.

La variable K es utilizada para seleccionar la cantidad de ejemplos más cercanos del conjunto de entrenamiento que ha de tenerse en cuenta para clasificarlo.

Esta clasificación se da mediante el cálculo de la distancia del nuevo ejemplo con todos del conjunto de train. Por medio del estudio de se visualiza que la clase predicha para el nuevo ejemplo es dada por la clase que pertenezcan a los ejemplos más cercanos del conjunto de entrenamiento, en la que, el valor de k es el que indica cuantos vecinos debemos fijarnos para predecir la clase.

2.2.2.2. Recuperación inductiva

Implica en la búsqueda de casos similares al nuevo problema por medio de la extracción de las características de casos para la generación de la solución para el problema nuevo. Esta es la desarrolla cuando la solución no está definida. Esto quiere decir que, en lugar de dar una solución por medio del razonamiento, el RBC hace uso de la experiencia previa, almacenada en la base de casos para la solución y toma en cuenta la extracción y combinación de características relevantes.

2.2.2.3. Software para el razonamiento basado en casos

En la actualidad existen múltiples programas que se encuentran disponibles para el Razonamiento basado en casos, en el cual para el desarrollo del algoritmo se eligió **Google Colab** más conocido como Colaboraty, como lo indica (Potluri et al., 2023) es una plataforma

popular que cuenta con una serie de ventajas sobre las de su competencia, ya que su uso es totalmente gratuito y no es necesario instalar ningún tipo de software ni configurar un entorno de desarrollo local. Todo es accesible desde cualquier ordenador con conexión en internet y está alojado en los servidores de Google.

2.3. Enfermedad de insuficiencia cardíaca

2.3.1. Concepto

La insuficiencia cardíaca es la principal causa de hospitalización en personas mayores de 65 años establecido por el estudio de (Scagliusi et al., 2023), esta ocurre cuando el corazón no bombea sangre y oxígeno suficiente al resto del cuerpo, uno de la causa más común de esta enfermedad es el daño del músculo cardíaco, ya sea por medio por medio de una enfermedad coronaria, ataque cardíaco etc.

La prevalencia de la insuficiencia cardíaca (IC) oscila entre el 0,9% en las personas menores de 64 años y el 17% en los mayores de 85 como lo expresa (Grau & Pérez, 2022).

2.3.2. Causas de la Insuficiencia Cardíaca

Las causas más frecuentes que puede desarrollar una insuficiencia Cardíaca son:

- Tensión arterial alta más conocida como: “hipertensión arterial”
- Obstrucción de las arterias: “cardiopatía isquémica”
- Enfermedades de las válvulas: “valvulopatías”
- Enfermedades del músculo cardíaco: “cardiomiopatías”

La descompensación se debe a diversos factores precipitantes entre los que son:

- Arritmias
- Infecciones
- Anemia
- Crisis hipertensiva

2.3.3. Diagnósticos estandarizados

En el trabajo investigativo se tomará en cuenta la historia clínica del paciente que se basa en la recopilación de información sobre síntomas del paciente, entre las características más importantes son:

Edad: edad de la persona

Sexo: sexo de la persona

Tipo de dolor en el pecho: El dolor de pecho es una detección temprana de enfermedades relacionadas con el corazón (Ibrahim et al., 2022), en la cual puede ser:

- Angina típica: Ocurre cuando el corazón no recibe suficiente flujo sanguíneo y por lo tanto oxígeno durante esfuerzo físico
- Angina atípica: Síntomas que no encajan completamente con la presentación clásica de la angina en la que se basa en la disminución del flujo sanguíneo al corazón.
- Dolor no anginal: Dolor de pecho que está relacionado con la angina de pecho
- Asintomático: Que no posee síntomas de dolor en el pecho.

Presión sanguínea en reposo (mm HG): La presión sanguínea es la fuerza que se aplica contra las paredes de las arterias cuando el corazón bombea la sangre al cuerpo y esta oscilará entre el rango de 120 y 150 mm Hg.

Colesterol: es un tipo de grasa corporal o lípida que una persona puede representar en la sangre, por lo tanto puede predecir la mortalidad intrahospitalaria en la mayoría de los pacientes de 70 años (Bonilla-Palomas et al., 2016), se dividen en 3 tipos:

- LDL “Low-Density Lipoprotein”: es la baja cantidad de lipoproteínas
- HDL “High-Density Lipoprotein”: es la alta cantidad de lipoproteínas
- Triglicéridos

Glucemia en ayunas o Nivel de Azúcar en la sangre: se basa en medir la azúcar o la glucosa en la sangre, con el fin de indentificar si hay una enfermedad preexistente como la diabetes.

Resultados del electrocardiograma en sus abreviaturas ECG: permite la detección de anomalías, por la que es de gran ayuda al diagnóstico del estado cardíaco del paciente como lo expresa un estudio de (Malepati et al., 2020), estas se pueden clasificar en:

1. Normal: muestra que la actividad eléctrica del corazón se encuentra en un estado saludable.
2. ST Inversión Onda T: Despolarización completa del miocardio ventricular
3. LVH: Hipertrofia completa del miocardio ventricular, infarto actual o pasado adaptación del miocardio a una sobrecarga de presión en el ventrículo izquierdo

Frecuencia cardíaca máxima: Conocido como los pulsos cardíacos por minuto, se basa en la cantidad de veces que late el corazón en el período de un minuto. En personas de 10 años en adelante, el rango se debe encontrar entre 60-100 latidos por minuto (Zhang et al., 2018).

Angina por ejercicio inducido: Es cuando una persona presenta dolor en el pecho al momento de realizar una actividad física.

Oldpeak: Es la depresión del ST: se refiere a una posible anomalía en la actividad eléctrica del corazón, más específica con la depresión del segmento ST al momento de realizar una prueba de esfuerzo.

Pendiente ST: el segmento ST puede ser ascendente “U”, plano “flat”, descendente “Down”, en la que, si se presenta una elevación en el gráfico, puede presentar un signo de infarto de miocardio, y si desciende puede ser una isquemia miocárdica.

Numero de vasos principales coloreado por fluorcopsia: categorizada por 3 niveles, este procedimiento se realiza para examinar los vasos sanguíneos del miocardio.

Prueba de Thailum: categorizada en 3 niveles, “N” normal, “DF” Defecto fijo, DR “defecto reversibe”, diseñada para pacientes que no pueden realizar ningún esfuerzo físico.

2.3.4. Estadísticas de salud

La insuficiencia cardiaca es una enfermedad que afecta a millones de personas alrededor del mundo. Algunas estadísticas importantes son:

- La OMS “Organización Mundial De La Salud”, dice que la enfermedad de Insuficiencia Cardíaca afecta a unos 26 millones de personas en todo el mundo

- Se considera como la causa principal de hospitalización de personas mayores de los 65 años
- En Estados Unidos se estima que 6.5 millones de personas padecen de Insuficiencia cardiaca

3. METODOLOGÍA

Se considera de tipo cuasiexperimental, ya que, se busca un control total de las variables de estudio, tanto dependiente como independiente, usando grupos experimentales y de control.

Según su alcance o evolución del fenómeno a estudiar, se considera un proyecto de diseño transversal, ya que, se tomarán conjuntos de datos obtenidos con la recopilación de historias clínicas, en un solo momento de tiempo, con el afán de describir las variables y analizar su incidencia. Adicional a la medición, se efectuarán las pruebas, para mejorar el algoritmo, con los datos de entrenamiento, lo que va a implicar el seguimiento de las variables involucradas para medir su evolución en el tiempo. La frecuencia de mediciones, al ser un tema enfocado a estudio de enfermedades del corazón, se va a realizar cada día, con el registro de los pacientes encontrados a diario, tanto de cardiólogos como médicos internistas involucrados directamente en este tipo de patología.

Enfoque

Su enfoque, se considera una investigación mixta, ya que, se integrará, de manera sistemática, los datos cuantitativos y cualitativos, debido a que las historias clínicas necesarias para registrar el historial del paciente contienen datos numéricos y descriptivos, los mismos que ingresarán al algoritmo para hacer más preciso el análisis numérico y el resultado final, será un dato descriptivo de una característica particular de esta determinada patología.

3.1. Métodos y técnicas de Recopilación de datos empleadas

Para la información se utilizará una base de datos como secuencia de entrenamiento, obtenida por medio de la página <https://www.kaggle.com/code/tanmay111999/heart-failure-prediction-cv-score-90-5-models#Dataset-Information>, ya que, almacena datos reales de una determinada enfermedad como lo es la insuficiencia cardíaca, y contiene más de 200 casos coleccionando pacientes que poseen información como Edad, Sexo, Colesterol ,etc.

Adicional a ello, para mejorar el algoritmo y probar, se tomarán datos reales a los pacientes de Cardiología del Centro Médico REDIMA “Red de Dispensarios Médicos de la Arquidiócesis”.

3.2. Métodos y técnicas de Análisis de datos

Para el presente trabajo investigativo se basó en la aplicación de un algoritmo por medio de Google Colab que permitirá resolver los casos actuales teniendo como base casos anteriores. La función fundamental es la de automatizar, sintetizar la búsqueda, selección y la adaptación de casos para poder aplicarlos.

4. RESULTADOS

El propósito del artículo es de implementar un modelo de Razonamiento Basado en Casos con la finalidad de detectar el diagnóstico médico de una enfermedad, en la que el algoritmo es desarrollado por medio de Google Colab de Python, para comprobar si este tipo de RBC que es un área de la Inteligencia Artificial conocida por sus siglas “IA” pueda ser de ayuda en el campo médico.

Aparte de los casos de pacientes que se encuentra en el Datasets que son más de 500, se agregó tres pacientes más con sus respectivos síntomas para luego ejecutarlo en el algoritmo desarrollado para saber la precisión exacta del diagnóstico.

En la siguiente tabla (ver tabla 1), se presentan los datos de los pacientes que se sometieron a los exámenes médicos para la prueba del algoritmo

	Edad	Sexo	Tipo de Dolor en el Pecho	Presión Sanguínea	Colesterol	Glucemia en Ayunas	Resultados de ECG	Frecuencia Alcanzada	Angina por ejercicio	Oldpeak	Pendiente St	Numero de vasos	Prueba de Thallium
Paciente A	80	M	Asintomático	120	110	110	Normal	110	No	1.2	1	0	1
Paciente B	80	F	Asintomático	115	190	102	Normal	90	No	0.0	0	0	0

Paciente C	71	F	Asintomá tíco	1	149	170	Normal	85	No	1.6	1	0	2
---------------	----	---	------------------	---	-----	-----	--------	----	----	-----	---	---	---

Tabla 1. Pacientes Adicionales a la Base de Datos

Para probar el algoritmo (ver Fig. 4), se tomarán datos reales a los pacientes de Cardiología del Centro Médico REDIMA “Red de Dispensarios Médicos de la Arquidiócesis”



Figura 4. Prueba del algoritmo a los pacientes de Cardiología del Centro Médico Redima paciente “A”

En la siguiente figura (ver Fig.5) se observan los resultados obtenidos del paciente A en el cual se refleja que su estado es “Sano”, en base al ingreso de los parámetros medidos y necesarios para el algoritmo.

```

▶ prediccion = knn.predict([[80, 1, 0, 120, 110, 0, 0, 118, 0, 1.2, 1, 0,1]])
  resultado[prediccion[0]]
  'sano'

```

Figura 5. Resultado del algoritmo del RBC con relación al Paciente “A”

En la siguiente figura (ver Fig.6) se recopila los datos personales del Paciente “B”



Figura 6. Prueba del algoritmo a los pacientes de Cardiología d Centro Médico Redima paciente “B”

En la siguiente figura (ver Fig.7) se observan los resultados obtenidos del paciente “B” en el cual se refleja que su estado es “Sano”, en base a lo ingresado en el algoritmo

```

▶ prediccion = knn.predict([[80, 0, 0, 115, 190, 0, 102, 0, 1.6, 0, 0, 0,0]])
  resultado[prediccion[0]]

'sano'

```

Figura 7. Resultado del algoritmo del RBC con relación al Paciente “B”

En la siguiente figura (ver Fig.8) se recopila datos personales del Paciente “C”



Figura 8. Prueba del algoritmo a los pacientes de Cardiología d Centro Médico Redima paciente “C”

En la siguiente figura (ver Fig.9) se observan los resultados obtenidos del paciente “C” en el cual se refleja que su estado es “Enfermo”, referente a lo ingreso en el algoritmo.

```

▶ prediccion = knn.predict([[71, 0, 0, 112, 149, 0, 1, 125, 0, 1.6, 1, 0, 2,]])
  resultado[prediccion[0]]

'enfermo'

```

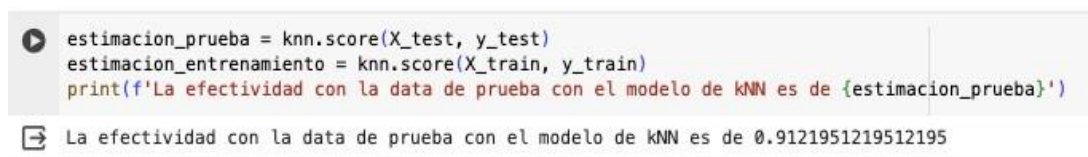
Figura 9. Resultado del algoritmo del RBC con relación al Paciente “C”

5. DISCUSION

La implementación y evaluación del modelo de predicción para enfermedades de insuficiencia cardiaca basado en razonamiento basado en casos con Machine Learning ha arrojado resultados prometedores que destacan la eficacia y la utilidad potencial de esta aproximación en el campo de la medicina predictiva. Se han encontrado importantes hallazgos que se contextualizan dentro del marco de la investigación existente, tales como:

Precisión y Desempeño del Modelo

En cuanto a la precisión y desempeño del modelo, se observó una alta precisión, con tasas de aciertos superiores al 91,21% para los datos de prueba y entrenamiento, tal como se indica en la gráfica siguiente (ver Fig.10), que es parte del algoritmo.



```
estimacion_prueba = knn.score(X_test, y_test)
estimacion_entrenamiento = knn.score(X_train, y_train)
print(f'La efectividad con la data de prueba con el modelo de kNN es de {estimacion_prueba}')
```

La efectividad con la data de prueba con el modelo de kNN es de 0.9121951219512195

Figura 10. Se obtiene aciertos mayores del 91,21% para los datos de prueba

Estos resultados sugieren que el modelo es capaz de identificar patrones relevantes en los datos médicos y realizar predicciones con una precisión considerable.

Importancia de las Características Seleccionadas

Durante el análisis de características, se identificó que ciertos factores, como la edad, antecedentes familiares, y datos relacionados con la presión arterial, tienen una influencia significativa en las predicciones del modelo. Esta información es crucial para los profesionales de la salud y puede ayudar a mejorar la comprensión de los factores de riesgo asociados a la insuficiencia cardiaca, datos con los cuales, se obtuvieron sus índices estadísticos, tal como se presenta en la siguiente gráfica (ver Fig.11):



```
df.describe().T
```

	count	mean	std	min	25%	50%	75%	max
age	1025.0	54.434146	9.072290	29.0	48.0	56.0	61.0	77.0
sex	1025.0	0.695610	0.460373	0.0	0.0	1.0	1.0	1.0
cp	1025.0	0.942439	1.029641	0.0	0.0	1.0	2.0	3.0
trestbps	1025.0	131.611707	17.516718	94.0	120.0	130.0	140.0	200.0
chol	1025.0	246.000000	51.592510	126.0	211.0	240.0	275.0	564.0
fbs	1025.0	0.149268	0.356527	0.0	0.0	0.0	0.0	1.0
restecg	1025.0	0.529756	0.527878	0.0	0.0	1.0	1.0	2.0
thalach	1025.0	149.114146	23.005724	71.0	132.0	152.0	166.0	202.0
exang	1025.0	0.336585	0.472772	0.0	0.0	0.0	1.0	1.0
oldpeak	1025.0	1.071512	1.175053	0.0	0.0	0.8	1.8	6.2
slope	1025.0	1.385366	0.617755	0.0	1.0	1.0	2.0	2.0
ca	1025.0	0.754146	1.030798	0.0	0.0	0.0	1.0	4.0
thal	1025.0	2.323902	0.620660	0.0	2.0	2.0	3.0	3.0
target	1025.0	0.513171	0.500070	0.0	0.0	1.0	1.0	1.0

Figura 11. Factores asociados a Insuficiencia Cardíaca

Interpretabilidad del Modelo

En el presente trabajo, se abordó con profesionales de la salud cardiológica del Centro Médico REDIMA para evaluar la interpretabilidad del modelo, discutiendo cómo los profesionales de la salud pueden entender y confiar en las decisiones del modelo. La capacidad de explicar el razonamiento detrás de las predicciones es esencial para la aceptación y adopción del modelo en entornos clínicos.

Generalización del Modelo

Se examinaron los resultados de las pruebas en términos de la capacidad del modelo para generalizar a nuevas poblaciones. La robustez del modelo se evalúa a través de pruebas en conjuntos de datos externos y se discute la posible necesidad de ajustes para adaptarse a diferentes poblaciones o entornos clínicos, por lo que, se pretende que la base de datos sea cambiante y esto altere las decisiones del modelo para futuras investigaciones.

Comparación con Métodos Tradicionales

Se compara el rendimiento del modelo con métodos tradicionales de diagnóstico de insuficiencia cardíaca. Se destaca cómo el enfoque basado en razonamiento basado en casos con Machine Learning implementado en este trabajo, supera o complementa a las técnicas convencionales, proporcionando una justificación sólida para la adopción de este modelo en entornos clínicos.

6. CONCLUSION

Al finalizar este trabajo se puede concluir que:

- El Razonamiento Basado en Casos es perfecto para casos de aplicación médicos, debido a que, los doctores almacenan datos en base a sus historias clínicas, en casos anteriores, los cuales, pueden ser utilizados como referencia para ingresarlos a un sistema predictivo de enfermedades, en este caso, de insuficiencia cardiaca.
- Es imprescindible la recopilación de datos de historias clínicas, con referencia al proceso en cuestión, en este caso, enfermedades de insuficiencia cardiaca. Una vez obtenidos los datos necesarios, se debe realizar un pre-procesamiento de estos para la posterior manipulación. Cabe destacar que, en el caso de no contar con bases de datos amplias actualizadas, se pueden consultar en repositorios de datos en la web, siempre y cuando se verifique la confiabilidad de dichos datos.
- Utilizar Python para el procesamiento de los datos e información, resulta de mucha utilidad por ser de código abierto, de fácil acceso a librerías y por contar con soporte internacional y foros de ayuda.
- Al evaluar los resultados esperados se puede obtener más datos relevantes para seguir alimentando el repositorio de datos y mejorar el algoritmo con mayor precisión.

7. REFERENCIAS

- Aguilar, W. O., López, W. A. R., Salvador, D. D. I., Revelo, E. R., & Mendoza, Y. L. O. (2021). El análisis inteligente de datos mediante el uso de técnicas de Softcomputing. *Serie Científica de La Universidad de Las Ciencias Informáticas*, 14(7), 38–53.
- Alvarado Salazar, R. E. (2022). *Inteligencia artificial con enfoque a la discapacidad visual: un mapeo sistemático*. <http://dspace.ups.edu.ec/handle/123456789/23327>
- Barreto Vite, C. S. (2022). *Modelo de un sistema basado en casos orientados al diagnóstico médico de una enfermedad*.
- Berástegui Arbeloa, G. (2018). *Implementación y comparación del algoritmo de los k vecinos más cercanos (k-NN) con valores locales en k*.
- Bonilla-Palomas, J. L., Gámez-López, A. L., Moreno-Conde, M., López-Ibáñez, C., Ramiro-Ortega, E., Castellano-García, P., & Villar-Ráez, A. (2016). El colesterol total predice la mortalidad intrahospitalaria en pacientes de 70 años o mayores hospitalizados por insuficiencia cardíaca aguda. *Revista Española de Geriatria y Gerontología*, 51(5), 280–283. <https://doi.org/https://doi.org/10.1016/j.regg.2015.11.003>
- Cedeño Pinargote, K. E., & Chinga Quimis, E. A. (2020). *Métricas y pesos para el establecimiento de similitudes en rasgos para el sistema de razonamiento basado en casos, del proyecto FCI para pacientes con Dislalia*. Universidad de Guayaquil. Facultad de Ciencias Matemáticas y Físicas~....
- Flor Alvarado, A. S., & Rodríguez Hurtado, V. E. (2021). *Métodos de recuperación para sistema de razonamiento basado en casos de pacientes con dislalia*. Universidad de Guayaquil. Facultad de Ciencias Matemáticas y Físicas~....
- Grau, M. L., & Pérez, M. Á. M. (2022). Abordaje del final de vida en los pacientes con insuficiencia cardíaca avanzada. *FMC - Formación Médica Continuada En Atención Primaria*, 29(6, Supplement 1), 60–65. <https://doi.org/https://doi.org/10.1016/j.fmc.2022.05.008>
- Homem, T. P. D., Santos, P. E., Costa, A. H. R., da Costa Bianchi, R. A., & de Mantaras, R. L. (2020). Qualitative case-based reasoning and learning. *Artificial Intelligence*, 283, 103258.
- Ibrahim, N., Karim, R. A., Wahidah Arshad, N., Wahab, Y. A., & Harmoko Saputro, A. (2022). Hand Segmentation for Chest Pain Behaviour of A Car Driver in Vehicle by Using Fusion Watershed and Blob Analysis. *2022 IEEE 10th Conference on Systems, Process & Control (ICSPC)*, 197–202. <https://doi.org/10.1109/ICSPC55597.2022.10001809>
- Leake, D., Ye, X., & Crandall, D. J. (2021). Supporting Case-Based Reasoning with Neural Networks: An Illustration for Case Adaptation. *AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering*, 2.
- Macias Vera, G. G. (2023). *Estudio relacionado con nuevas tecnologías de machine Learning para determinar cuadros clínicos*. Babahoyo: UTB-FAFI. 2023.`
- Malepati, N., Fatima, R., Gupta, S., Ramsali, V., & K.R., S. (2020). Portable ECG Device for Remote Monitoring and Detection of Onset of Arrhythmia. *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, 1–5. <https://doi.org/10.1109/CONECCT50063.2020.9198658>
- Potluri, T., Shi, Y., Shahriar, H., Lo, D., Parizi, R., Chi, H., & Qian, K. (2023). Secure Software Development in Google Colab. *2023 IEEE World AI IoT Congress (AIIoT)*, 398–402. <https://doi.org/10.1109/AIIoT58121.2023.10174336>
- Rachmawati, D., Siregar, A. B., & others. (2022). Implementation Combination of Case-Based Reasoning and Rule-Based Reasoning for Diagnosis of Herpes Disease. *2022 6th International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM)*, 210–214. <https://doi.org/10.1109/ELTICOM58121.2022.9812121>
- Sanchez-Romero, J., & Llerena-Izquierdo, J. (2023). Revisión de la literatura sobre el uso del aprendizaje profundo enfocado en sistemas de inspección ópticos automatizados para la detección de defectos superficiales en el sector de la manufactura. *Revista InGenio*, 6(2), 1–19. <https://doi.org/10.18779/ingenio.v6i2.680>

- Scagliusi, S. F., Pérez-García, P., Oprescu, A. M., Fernández, D. M., Olmo, A., Sánchez, G. H., & Yúfera, A. (2023). Body Posture Determination for Heart Failure Patients From Ankle Orientation Measurements. *IEEE Access*, *11*, 48893–48900. <https://doi.org/10.1109/ACCESS.2023.3261554>
- Serrano Buenano, E. P. (2023). *Propuesta de un modelo predictivo con algoritmos de aprendizaje supervisado de machine learning aplicado al sobrepeso de las personas. P1*. Universidad de Guayaquil. Facultad de Ciencias Matemáticas y Físicas~....
- Torres Cordero, F. E., Torres Rodríguez, E. de la C., Chile Labrador, J. M., Reyes Valdés, D. de los, & González Benítez, N. (2019). Análisis neutrosófico para el diagnóstico de la hipertensión arterial a partir de un sistema experto basado en casos . *Revista Asociación Latinoamericana de Ciencias Neutrosóficas*. ISSN 2574-1101, 10(6 SE-), 28–38. <https://fs.unm.edu/NCML2/index.php/112/article/view/68>
- Varshavskii, P., Alekhin, R., Polyakov, S., Blashonkov, T., & Mukhacheva, I. (2020). Development of a Modular Case-Based Reasoning System for Data Analysis. *2020 International Youth Conference on Radio Electronics, Electrical and Power Engineering (REEPE)*, 1–4. <https://doi.org/10.1109/REEPE49198.2020.9059242>

ANEXO

Anexo 1

Algoritmo basado en RBC que permitió la predicción de enfermedades de Insuficiencia Cardíaca disponible en GitHub.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import matplotlib as mpl
#matplotlib inline
import seaborn as sns
pd.options.display.float_format = '{:.2f}'.format
import warnings
warnings.filterwarnings('ignore')

from google.colab import files
from google.colab import drive
drive.mount('/content/drive')

df=pd.read_csv('/content/drive/MyDrive/Insuficiencia
Cardiaca/heart.csv')
df.head()
def sano_enfermo(fila):
    result = 'enfermo'
    if int(fila['target']) == 0:
        result = 'sano'
    return result

df['estado'] = df.apply(sano_enfermo, axis=1)
df.head(20)

resultado = dict(zip(df['target'].unique(), df['estado'].unique()))
resultado

from sklearn.model_selection import train_test_split

X = df[['age', 'sex', 'cp', 'trestbps', 'chol', 'fbs', 'restecg',
'thalach', 'exang', 'oldpeak', 'slope', 'ca', 'thal']]
y = df['target']

X_train, X_test, y_train, y_test = train_test_split(X, y,
random_state=0, test_size=0.20)

from sklearn.neighbors import KNeighborsClassifier
```

```
knn = KNeighborsClassifier(n_neighbors=3)
knn.fit(X_train, y_train)

estimacion = knn.score(X_test, y_test)
print(f'La efectividad con el modelo de kNN es de {estimacion}')
```

```
estimacion_prueba = knn.score(X_test, y_test)
estimacion_entrenamiento = knn.score(X_train, y_train)
print(f'La efectividad con la data de prueba con el modelo de kNN es de {estimacion_prueba}')
```

Predicción de Enfermedades

```
prediccion = knn.predict([[30, 1, 0, 120, 180, 0, 1, 120, 0, 1.2, 1, 3, 3]])
resultado[prediccion[0]]
```

Anexo 2

Base de Datos que se tomó en cuenta para el desarrollo del algoritmo está disponible en GitHub

A	B	C	D	E	F	G	H	I	J	K	L	
1	Edad	Sexo	Tipo de dolor en el pecho	Presión Sanguínea	Colesterol	Nivel de Azúcar en la sangre	Resultados del Electrocardiograma	Frecuencia Cardíaca Máxima Alcanzada	Angina por ejercicio	Oldpeak	Pendiente ST	Insuficiencia Cardíaca
2	40	M	ATA	140	289	0	Normal	172	No	0	Up	0
3	49	F	NAP	160	180	0	Normal	156	No	1	Flat	1
4	37	M	ATA	130	283	0	ST	98	No	0	Up	0
5	48	F	ASY	138	214	0	Normal	108	Si	1.5	Flat	1
6	54	M	NAP	150	195	0	Normal	122	No	0	Up	0
7	39	M	NAP	120	339	0	Normal	170	No	0	Up	0
8	45	F	ATA	130	237	0	Normal	170	No	0	Up	0
9	54	M	ATA	110	208	0	Normal	142	No	0	Up	0
10	37	M	ASY	140	207	0	Normal	130	Si	1.5	Flat	1
11	48	F	ATA	120	284	0	Normal	120	No	0	Up	0
12	37	F	NAP	130	211	0	Normal	142	No	0	Up	0
13	58	M	ATA	136	164	1	ST	99	Si	2	Flat	1
14	39	M	ATA	120	204	0	Normal	145	No	0	Up	0
15	49	M	ASY	140	234	1	Normal	140	Si	1	Flat	1
16	42	F	NAP	115	211	0	ST	137	No	0	Up	0
17	54	F	ATA	120	273	0	Normal	150	No	0	Flat	1
18	38	M	ASY	110	196	0	Normal	166	Si	1	Up	0
19	43	F	ATA	120	201	0	Normal	165	No	3	Flat	1
20	60	M	ASY	100	248	0	Normal	125	No	0	Up	0
21	36	M	ATA	120	267	0	Normal	160	No	1	Up	0
22	43	F	TA	100	223	0	Normal	142	No	0	Flat	1
23	44	M	ATA	120	184	1	Normal	142	No	3	Flat	1
24	49	F	ATA	150	202	0	Normal	164	No	0	Up	0
25	44	M	ATA	130	288	0	Normal	150	No	0	Up	0
26	40	M	NAP	130	215	0	Normal	138	No	3	Flat	1
27	36	M	NAP	124	209	0	Normal	178	No	0	Flat	1
28	53	M	TA	120	260	0	ST	112	Si	0	Up	0
29	52	M	ATA	113	284	0	Normal	118	No	0	Flat	1
30	53	F	ATA	125	468	0	Normal	127	No	0	Up	0
31	51	M	ATA	145	188	0	Normal	145	Si	0	Flat	1
32	53	F	NAP	130	518	0	Normal	130	No	2	Up	0

Anexo 3

Enlace del algoritmo subido en la plataforma de GitHub para su libre acceso:

<https://github.com/KerlyRodriguez/ENFERMEDAD-CARDIACA/blob/main/Enf-cardiad.ipynb>