

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA INFORMÁTICA

MÁSTER UNIVERSITARIO EN INGENIERÍA DEL SOFTWARE E
INTELIGENCIA ARTIFICIAL

**Un clasificador estadístico para apoyar el
diagnóstico de demencias**

**A statistical classifier to support diagnose
dementias**

Realizado por

Estrella Lucena Sánchez

Tutorizado por

Eduardo Guzmán De los Riscos

M^a Victoria Belmonte Martínez

Departamento

Departamento de Lenguaje y Ciencias de la Computación

UNIVERSIDAD DE MÁLAGA

MÁLAGA, SEPTIEMBRE DE 2018

Resumen y palabras clave

Resumen

Se define demencia como una enfermedad cerebral que afecta a funciones cognitivas, conductuales y psicológicas. Los pacientes con demencias tienen una deficiente calidad de vida. A su vez, el rápido y progresivo deterioro del paciente una vez diagnosticada la enfermedad produce un fuerte impacto en numerosos sectores de la sociedad tanto económicos, como políticos y sociales. Hasta la actualidad no se ha encontrado un tratamiento que palie la enfermedad ya que los cambios cerebrales que la provocan se producen en años previos al surgimiento de los síntomas. Por lo tanto, se requieren de herramientas que ayuden a la detección precoz de las demencias y contribuyan a la búsqueda de la cura de estas.

En este trabajo se han construido modelos de decisión basados en datos para ayudar al sector sanitario en el diagnóstico de demencias. Para este propósito, hemos reunido un conjunto de datos de 67 registros de pacientes. Se han aplicado cinco técnicas de clasificación a partir de los datos de entrada de nueve síntomas y otra información sobre el paciente como la edad, el sexo, el déficit cognitivo y diferentes patologías, y se ha realizado una validación de 10 pliegues cruzados. Los resultados muestran que las técnicas aplicadas son adecuadas para predecir el déficit cognitivo de diferentes demencias.

Abstract

Dementia is defined as a brain disease which affects cognitive, conductual and psychologist functions. Life patients is deficient. Also, the progressive and rapid decline of patients has a strong impact on many sectors of society, both economic, political and social. At present, there is no treatment because the brain changes appear years before the emergence of the symptoms. So, tools are required to help early detection of dementias and contribute to the search for cure of these.

For this purpose, we have gathered together a dataset of records of patients form . Five classification techniques have been applied from input data of nine symptoms and other information about the patient such as age, sex, cognitive deficit and different pathologies, and a 10 cross-fold validation has been performed. Results show that the techniques applied are suitable for predicting the cognitive deficit of different dementias.

Palabras Clave

Diagnóstico de demencias, clasificador estadístico y toma de decisiones médicas.

Keywords

Dementia diagnosis, statistical classifier and medical decision making.

Índice general

| | |
|---|-----------|
| 1. Demencia | 9 |
| 1.1. Etiología de las demencias | 10 |
| 1.2. Tipos de demencias | 11 |
| 2. Técnicas y estudios previos | 15 |
| 2.1. Técnicas para construir los modelos de clasificación | 15 |
| 2.2. Estudios previos | 16 |
| 3. Análisis de datos | 19 |
| 3.1. Recogida de datos | 19 |
| 3.2. Preprocesamiento de los datos | 20 |
| 3.3. Análisis de los datos | 22 |
| 4. Resultados | 23 |
| 5. Conclusiones | 27 |

Introducción

El principal objetivo de este TFM es encontrar un mecanismo capaz de determinar si un paciente tiene un tipo u otro de demencia a partir de un conjunto de síntomas. Para ello se va a desarrollar un clasificador estadístico que, partiendo de diferentes atributos asociados a un grupo de pacientes, prediga el diagnóstico asociado a los mismos. Dicho diagnóstico consistirá en una demencia determinada por los síntomas que la caracteriza. Además, se emplearán los fármacos de los diferentes tratamientos para buscar una relación entre la hipertensión y la diabetes con las demencias.

Los pacientes con demencias tienen una deficiente calidad de vida debido a que dicha enfermedad cerebral afecta a funciones cognitivas, conductuales y psicológicas. Además, el rápido y progresivo deterioro del paciente una vez diagnosticada la enfermedad requiere una rápida adaptación por parte de las personas que forman parte de la vida de este. Esto produce un fuerte impacto en numerosos sectores de la sociedad tanto económicos, como políticos y sociales. Por lo que se considera este síndrome de gran interés y se ha desarrollado este trabajo tratando de aportar una herramienta capaz de facilitar su diagnóstico y ayudar a mejorar la calidad de vida de los pacientes.

Hasta la actualidad no se ha encontrado un tratamiento que palie la enfermedad ya que los cambios cerebrales que la provocan se producen en años previos al surgimiento de los síntomas. Por lo tanto, se requieren de herramientas que ayuden a la detección precoz de las demencias y contribuyan a la búsqueda de la cura de estas.

A su vez, se tratará de dilucidar la relación entre hipertensión y diabetes con las demencias, puesto que numerosos estudios revelan que las personas con hipertensión o diabetes durante sus vidas cuentan con mayor probabilidad de desarrollar demencias al llegar a la vejez.

Objetivos

Como se ha explicado en el primer capítulo hasta la actualidad no se ha encontrado la cura para las demencias debido a que los síntomas se manifiestan años más tarde de los cambios que se producen en el cerebro.

A su vez, las demencias suponen elevados costes en la sociedad afectando a sectores

políticos, sanitarios, sociales y económicos en todas las áreas del mundo. Además, estos costes continuarán aumentando conforme la población continúe envejeciendo y la longevidad sea mayor. Por lo que se requieren herramientas de diagnóstico que ayuden a la detección precoz de este síndrome.

En consecuencia, se ha desarrollado este trabajo para apoyar al diagnóstico de las demencias. El principal objetivo de este trabajo es encontrar un mecanismo capaz de determinar si un paciente tiene un tipo u otro de demencia a partir de un conjunto de atributos que incluyen síntomas, género, edad y fármacos asociados a otras enfermedades.

A su vez, como se ha comentado en la introducción, existe una relación entre enfermedades como la diabetes o la hipertensión y el desarrollo de demencias en las fases finales de la vida. Por lo que otro de los objetivos de este trabajo se centra en la búsqueda de la probable relación entre dichas enfermedades.

De esta forma, se pretende establecer la probabilidad de que un paciente que sufre diabetes o hipertensión durante su vida desarrolle una demencia en los años finales de la misma. Para ello se han empleado los fármacos de los tratamientos de los pacientes.

Estructura de la memoria

En primer lugar se define el término demencia y se exponen los tipos y la etiología de las mismas.

A continuación, se explican las diferentes técnicas empleadas en el trabajo para construir el modelo de clasificación que permita la predicción de la demencia. Y se detalla el estado del arte considerando los estudios previos a este que emplean dichas técnicas.

Posteriormente se explica la recogida y el tratamiento de los datos así como en qué consiste la base de datos y la información que nos proporciona.

Por último se procede a mostrar y analizar el resultado de los diferentes experimentos realizados.

Capítulo 1

Demencia

Etimológicamente la palabra “demencia” está compuesta por el prefijo “de” (ausencia), “mens” (mente) y por el sufijo “ia” (condición o estado). La etimología nos acerca a las fases finales de la enfermedad, pero no refleja lo que ocurre en las primeras etapas de esta.

Se define demencia como un síndrome debido a una enfermedad del cerebro, generalmente de naturaleza crónica o progresiva, en la que hay una alteración de múltiples funciones corticales superiores, incluida la memoria, el pensamiento, la orientación, la comprensión, el cálculo, la capacidad de aprendizaje, el lenguaje y el juicio (Barrett and Burns, 2014).

Las manifestaciones clínicas o síntomas de la demencia son heterogéneas puesto que existen múltiples etiologías, expuestas en el apartado 1.1. No obstante, los principales síntomas observables son: alteración de la memoria episódica, posible manifestación de afasia, apraxia, agnosia, amnesia, alteración de la personalidad, de la conducta social y del lenguaje. Los síntomas quedan recogidos en la Tabla 1.1. Como se observa en la tabla, dependiendo de la demencia se distinguen unos síntomas u otros.

El reconocimiento temprano de los síntomas puede ayudar a los enfermos y sus familias reduciendo los niveles de ansiedad sobre los cambios en la memoria, el pensamiento, el estado de ánimo o el comportamiento y permitiendo la movilización de recursos que serán necesarios en el futuro. Sin embargo, existen múltiples obstáculos para el diagnóstico precoz, tanto en términos de conciencia pública como de comprensión profesional de la demencia (Yanguas, 2007).

Los estudios de prevalencia de pacientes con demencia en España han mostrado cifras que oscilan entre el 5 y el 15 % para mayores de 65 años y entre el 6 y el 17 % para mayores de 70 años según estudios estadísticos poblacionales realizados por el Centro Nacional de Epidemiología. Aunque para población menor de 65 años existen pocos estudios, se estimó que en un área de Londres la prevalencia era de 0.054 % en pacientes con edades comprendidas entre 30 y 64 años. En general la prevalencia es más alta en mujeres y aumenta con la edad.

Tabla 1.1: Síntomas

| Demencia | Síntomas |
|-------------------------|---|
| Alzheimer | Alteración de la memoria episódica Trastornos apráxicos, afásicos y agnósicos Liberación de los reflejos primitivos |
| Frontotemporal | Alteración de la personalidad y de la conducta social Alteración del lenguaje Conductas sociales desinhibidas y compulsiones Alteraciones de la conducta alimentaria o sexual Comportamientos estereotipados y apatía |
| Vascular | Lentitud del procesamiento mental (bradipsiquia) Reducción gradual de la atención sostenida y la fluidez verbal Alteración de funciones ejecutivas Risa o llanto espasmódicos Parkinsonismo |
| Supranuclear progresiva | Bradipsiquia Reducción gradual de la atención sostenida y la fluidez verbal Hipocinesia y síntomas pseudobulbares Risa o llanto espasmódicos Parkinsonismo |
| Cuerpos de Lewy | Alucinaciones visuales Trastorno de conducta asociado a la fase REM del sueño Trastorno de la autonomía Parkinsonismo |

1.1. Etiología de las demencias

Los factores de riesgo pueden considerarse genéticos y ambientales. Cabe destacar que los factores de riesgo genéticos modificarán la reacción de un individuo a los factores de riesgo ambientales a los que está expuesto (Barrett and Burns, 2014).

Dentro de los factores genéticos se distinguen:

- Los antecedentes familiares de enfermedad de Alzheimer o demencia vascular de aparición tardía, que se asocian con un mayor riesgo de desarrollar la afección en cualquier individuo.
- Las anormalidades cromosómicas discretas.
- El síndrome de Down. El riesgo de desarrollar demencias es significativamente mayor para las personas con síndrome de Down que para la población general.

Dentro de los riesgos ambientales, los más relevantes son la edad, problemas cardiovasculares, aterosclerosis y problemas psiquiátricos.

El riesgo de desarrollar demencia aumenta con el aumento de la edad. Luego, la edad puede ser un factor de riesgo en sí mismo o puede reflejar el efecto de aumentar el tiempo

durante el cual otros factores pueden ejercer su influencia.

Los factores de riesgo cardiovascular son fumar, presión arterial alta, diabetes e hiperlipidemia. Todos estos factores de riesgo predisponen al desarrollo de la aterosclerosis y al accidente cerebrovascular agudo, que se asocia con la demencia de todos los tipos clínicos.

Además, existen evidencias de la vinculación de un mayor riesgo de desarrollar Alzheimer con un historial de depresión a una edad temprana. Estudios han identificado un mayor riesgo de demencia con antecedentes de problemas psiquiátricos graves, siendo la esquizofrenia y la depresión los diagnósticos más comunes encontrados (Cooper y Holmes, 1998).

Estos riesgos han sido considerados como posibles agentes causantes de la enfermedad en este estudio. A los que se han añadido otras posibles causas, siguiendo un asesoramiento experto, como la carencia de determinadas sustancias y componentes químicos de la sangre o problemas de tiroides.

1.2. Tipos de demencias

Existen varias formas de clasificar las demencias atendiendo a la edad, la causa o etiología, las estructuras cerebrales afectadas o según la prevalencia en la población (Barrett and Burns, 2014) y (Bishara and Harwood, 2014).

Conforme a la topología cerebral, las demencias se pueden clasificar en cuatro clases o grupos:

1. Demencias corticales. Son aquellas en las que el daño cerebral se produce principalmente en la corteza cerebral. Dentro de este grupo se distinguen las siguientes demencias:
 - Enfermedad de Alzheimer
 - Enfermedad de Pick
2. Demencias Subcorticales. Son aquellas que afectan a las estructuras que se encuentran por debajo del córtex cerebral. Se incluyen en este grupo las siguientes demencias:
 - Síndromes extrapiramidales
 - Enfermedad de Parkinson
 - Enfermedad de Huntington
 - Parálisis supranuclear progresiva
 - Enfermedad de Wilson
 - Degeneraciones espinocerebelosas

- Clasificación idopática de los ganglios basales
3. Demencias con disfunción cortical y subcortical. Son aquellas que afectan a la totalidad de las estructuras cerebrales sin predominio en ninguna de ellas. Caben destacar las siguientes demencias dentro de este tipo:
- Demencia con cuerpos de Lewy
 - Demencia frontotemporal
 - Postanóxica
 - Neoplásica
 - Demencia multiinfarto o vasculares
 - Infecciosas
 - Virus lentos
 - Parálisis general
4. Demencias axiales. Se producen por lesiones en las estructuras mediales del lóbulo temporal, hipocampo, cuerpos mamilares o hipotálamo. La demencia incluida en este grupo es la Demencia de Wernike-korsakoff. De este tipo no se contó con ningún paciente en la base de datos por lo que no se utilizó en las clasificaciones.

Atendiendo a la etiología de la demencia, se consideran tres categorías: demencias degenerativas primarias, demencias degenerativas secundarias y demencias combinadas o mixtas.

1. Demencias degenerativas primarias. Son aquellas causadas por un trastorno neurológico degenerativo con origen en el sistema nervioso. Dentro de este grupo se encuentra la enfermedad de Alzheimer, demencia frontotemporal, demencia por cuerpos de Lewy y enfermedad de Parkinson.
2. Demencias degenerativas secundarias. En este grupo se incluyen las demencias causadas por un trastorno neurológico no degenerativo o que no tiene su origen en el sistema nervioso. La causa puede ser de origen vascular, infecciosa, carencial, endocrino-metabólica, toxicológica o por otras patologías médicas que en su evolución clínica pueden acabar desarrollando demencia.
3. Demencias combinadas o mixtas. Los pacientes con demencias mixtas se caracterizan por la presencia de síntomas y signos patológicos de más de un tipo de demencia, por lo general, de demencia debida a la enfermedad de Alzheimer y de demencia vascular.

Según la prevalencia de las demencias en la población, se distinguen dos tipos de demencias: enfermedad de Alzheimer y demencia de tipo vascular. Ambas suponen conjuntamente más del 90 % de los casos.

En este trabajo se han elaborado cinco clasificaciones diferentes teniendo en cuenta los criterios expuestos anteriormente y los criterios de un experto.

Capítulo 2

Técnicas y estudios previos

2.1. Técnicas para construir los modelos de clasificación

La metodología utilizada para desarrollar este proyecto consiste en la aplicación de técnicas para construir modelos de clasificación para un conjunto de datos. Estas técnicas son provistas por el software Weka.

En primer lugar, hemos explorado diferentes técnicas de clasificación tomadas del campo del aprendizaje automático (Michalski et al., 2013), tales como Árboles de Decisión, Bosque Aleatorio, Máquina de Soporte Vectorial, Red Bayesiana y Naïve Bayes.

- Árboles de Decisión (DT). Se han utilizado para crear herramientas automatizadas que son útiles tanto para apoyar un diagnóstico como para hacer predicciones. Son modelos donde los algoritmos toman forma de árbol. Su objetivo es predecir el valor de una variable objetivo en función de varias variables de entrada. Se pueden usar varios algoritmos para generar un DT a partir de un conjunto de datos. Weka proporciona el algoritmo C4.5, un clasificador recursivo estadístico. C4.5 DT se construye a partir de un conjunto de datos de entrenamiento utilizando el concepto de entropía. J48 es una implementación de código abierto de C4.5 disponible en el software Weka.

Hay técnicas que construyen más de un DT, por ejemplo Random Forest (RF). Es un conjunto de árboles de regresión o clasificación sin podar creados mediante el uso de muestras de Bootstrap de los datos de entrenamiento y la selección de características aleatorias en la inducción de árboles. La predicción se obtiene a partir de los resultados de los árboles del predictor múltiple (promediación o elección de la mayoría).

- La Máquina de Soporte Vectorial (SVM) es una técnica eficaz de aprendizaje automático supervisado. La idea básica consiste en que dado un conjunto de puntos, subconjunto de un conjunto mayor (espacio), en el que cada uno de ellos pertenece

a una de dos posibles categorías, un algoritmo basado en SVM construye un modelo capaz de predecir si un punto nuevo (cuya categoría desconocemos) pertenece a una categoría o a la otra. Como en la mayoría de los métodos de clasificación supervisada, los datos de entrada (los puntos) son vistos como un vector p -dimensional (una lista ordenada de p números).

- Una Red Bayesiana (BN) es un modelo gráfico probabilístico que representa un conjunto de variables aleatorias y sus dependencias condicionales a través de un gráfico acíclico dirigido. En el grafo, los nodos representan variables aleatorias y los bordes representan dependencias condicionales. La dependencia entre dos nodos se describe por la presencia o ausencia de un arco entre ellos, y su influencia causal por la dirección del arco.
- Naïve Bayes (NB) es un caso especial de BN, donde se supone que los atributos son condicionalmente independientes, por esta razón se llama ingenuo. Este algoritmo es muy eficiente en ciertos dominios ya que es muy robusto con respecto a los atributos irrelevantes.

2.2. Estudios previos

En los últimos años, el aprendizaje automático y la inteligencia artificial en general se han utilizado como herramientas de diagnóstico para diferentes patologías. En la literatura, hay muchas contribuciones recientes de estudios de investigación que usan técnicas de aprendizaje automático (o diferentes modelos de predicción) para apoyar el diagnóstico de diferentes enfermedades mediante la predicción de estas. La neumonía (Jun et al., 2018), enfermedades de las arterias coronarias (Mokeddem, 2018), (Kim et al., 2015), (Kim et al., 2014) y (Babaoglu et al., 2009), las enfermedades cardiovasculares (Bashir et al., 2014), (Anooj, 2012) y (Eom et al., 2008), la leucemia (Mohammed et al., 2017), el cáncer (Takada et al., 2012), el asma (Farion et al., 2010), la diabetes (Temurtas et al., 2009) y la dermatosis (Übeyli, 2009) son solo algunos ejemplos. Sin embargo, no se ha encontrado ningún enfoque que utilice modelos predictivos para el diagnóstico de las demencias. En la Tabla 2.1 se muestra una selección de la investigación antes mencionada.

Los trabajos relacionados (Jun et al., 2018), (Mokeddem, 2018), (Kim et al., 2015), (Takada et al., 2012), (Farion et al., 2010) y (Eom et al., 2008), al igual que nuestro enfoque, utilizan diferentes clasificadores basados en DT con buenos resultados de rendimiento. En (Jun et al., 2018), Jun et al. generó un árbol de decisión como sistema de diagnóstico diferencial asistido por ordenador (CADD) que diferenciaba entre la neumonía intersticial habitual (UIP) y la neumonía intersticial no específica (NSIP) utilizando imágenes de tomografía computerizada de alta resolución.

Se aplicó un clasificador de máquina de soporte vectorial para entrenar y etiquetar las regiones de interés del parénquima pulmonar según las características de textura y forma

Tabla 2.1: Estudios previos

| Año de publicación/Referencia | Número de registros | Técnicas usadas | Objetivos |
|-------------------------------|---|--|---|
| 2018 (Jun et al., 2018) | 54 | SVM | Se desarrolló un CADD que distingue entre UIP y NSIP utilizando imágenes de tomografía computarizada de alta resolución. |
| 2018 (Mokeddem, 2018) | 834 | RF y C5.0 DT | En este trabajo se desarrolló un nuevo CDSS para diagnosticar CAD. El modelo se basa en el algoritmo RF, el algoritmo C5.0 DT y el modelado Fuzzy. |
| 2017 (Mohammed et al., 2017) | 6345 | SVM, Red Neuronal Artificial (ANN), DT, K-Nearest Neighborhood y Adaptive Boosting | Se demuestra el potencial de cinco algoritmos de aprendizaje automático para aprovechar el valor de las imágenes microscópicas de frotis de sangre para clasificar a los pacientes con leucemia linfocítica crónica. |
| 2015 (Kim et al., 2015) | 748 | CART DT | En este estudio se desarrolló un modelo de predicción de CHD impulsado por el árbol de decisiones y la lógica difusa CART. |
| 2014 (Kim et al., 2014) | 299 | C4.5 DT basado en el método Mamdani | Este artículo propone un modelo de soporte de predicción de enfermedad coronaria adaptable basado en la regla difusa FbACHD_PSM, que proporciona recomendaciones a los pacientes con enfermedad coronaria. |
| 2014 (Bashir et al., 2014) | 1316 | NB, DT basado en el índice Gini y SVM | En esta investigación se propone un clasificador de conjuntos que lleva a cabo la clasificación y predicción de datos de enfermedades del corazón. |
| 2012 (Takada et al., 2012) | 465 | ADTree | El objetivo de este estudio fue desarrollar un nuevo modelo para predecir la metástasis de los ganglios linfáticos axilares en el cáncer de mama primario. |
| 2012 (Anooj, 2012) | 398 | CDSS | Se presenta un sistema de apoyo a las decisiones clínicas basado en reglas difusas para el diagnóstico de la enfermedad cardíaca, obteniendo automáticamente el conocimiento de los datos clínicos del paciente. |
| 2010 (Farion et al., 2010) | 341 | C4.5 DT | En este artículo se desarrolla un modelo que ayuda a predecir la gravedad de las exacerbaciones de asma pediátricas en el departamento de emergencias a las 2 h después del triaje |
| 2009 (Babaoglu et al., 2009) | 330 | ANN | El objetivo de este estudio es comprobar cómo la red neuronal artificial determina la existencia de enfermedad de la arteria coronaria y la localización de la lesión a partir de los datos de la prueba de esfuerzo. |
| 2009 (Übeyli, 2009) | 111 Psoriasis 60 Dermatitis 71 Liquen plano 48 Pitiriasis rosada 48 Dermatitis crónica 20 Pitiriasis rubra pilaris | CNN | Este artículo ilustra el uso del modelo de redes neuronales combinadas (CNN) para guiar la selección del modelo para el diagnóstico de las enfermedades eritematoescamosas. |
| 2009 (Temurtas et al., 2009) | 768 | Red Neuronal Multicapa (MNN) y Red Neuronal Probabilística (PNN) | En este estudio se realizó un diagnóstico de la enfermedad de pima diabetes comparativo. |
| 2008 (Eom et al., 2008) | 66 Enfermedad cardiovascular 95 Problemas pulmonares 27 Tuberculosis 54 Cáncer | SVM, NN, DT y BN | En este artículo, se propone un método basado en un conjunto clasificador para apoyar el diagnóstico de la enfermedad cardiovascular |

(CADD) utilizando una base de datos con 54 imágenes de pacientes, y la mayor precisión obtenida fue de 0.91.

En (Mokeddem, 2018) se propuso un nuevo sistema de apoyo a la decisión clínica (CDSS) para diagnosticar las Enfermedades de las Arterias Coronarias (CAD). En este caso, el estudio utilizó 834 registros de pacientes, y el algoritmo de Random Forest se usa para clasificar las características y se desarrolló un enfoque basado en el árbol de decisión C5.0 para generar una regla nítida a partir de la que se generó el nuevo sistema. Se obtuvieron valores de precisión, sensibilidad y especificidad de 0.84, 0.87 y 0.81, respectivamente.

En (Kim et al., 2015) se propuso un nuevo modelo predictivo para las enfermedades coronarias en Korea. El modelo de predicción utilizó un modelo de inferencia basado en lógica difusa y las reglas se generaron utilizando un método de árbol de decisión y clasificación (CART). El estudio empleó 748 registros de sujetos, y le modelo obtuvo valores de precisión 0.69 y de sensibilidad de 0.93, elevados en comparación con otros cuatro modelos de predicción (ANN, SVM, LR y C5.0 DT).

En (Takada et al., 2012) se empleó un Árbol de Decisión Alternativo (ADTree) para desarrollar un nuevo modelo matemático para predecir la metástasis de los ganglios linfáticos axilares en pacientes con cáncer de mama primario. Este modelo usó información clínica-patológica preoperatoria de 465 pacientes, y el modelo obtenido mostró una buena precisión (ROC de 0.772) para predecir metástasis ganglionares.

En (Farion et al., 2010) se desarrolló un modelo DT basado en el algoritmo C4.5 para predecir la gravedad de las exacerbaciones de asma pediátricas en el departamento de emergencias a las 2 h después del triaje. Experimentaron con un conjunto de datos de 341 registros y 4 modelos de clasificación. Todos ellos fueron C4.5 DT y el mejor modelo mostró una precisión de 0.64 y una sensibilidad de 0.79.

Finalmente, (Eom et al., 2008) propone un método basado en un conjunto clasificador para apoyar el diagnóstico de enfermedad cardiovascular (ECV) basado en chips de aptámero. El trabajo empleó cuatro bases de datos diferentes de 66, 95, 27 y 54 registros, respectivamente, y obtuvo una precisión de 0,95.

En comparación con nuestro trabajo, se observa que todos los enfoques anteriores tienen, de acuerdo a los resultados de rendimiento, buenos resultados, pero no mejores que los nuestros, con un área ROC y una precisión superiores a 0.94

Capítulo 3

Análisis de datos

3.1. Recogida de datos

Se han empleado un conjunto de datos compuesto por 67 enfermos de demencia adquiridos de la Asociación de Familiares de Enfermos de Alzheimer y otras Demencias Mijas-Fuengirola Costa (AFA Mijas-Fuengirola Costa).

Los datos fueron adquiridos de forma manual haciendo uso de las bases de datos en las que se encontraban registrados los pacientes y los diagnósticos y tratamientos almacenados en la institución de acuerdo a cada paciente.

Una vez recogidos los datos de todos los pacientes, se establecieron diferentes clasificaciones dependiendo de la demencia asociada y de la edad y el sexo de estos, estableciéndose finalmente una base de datos con 23 atributos que corresponden con la edad, el sexo, las causas, los síntomas y los fármacos de cada paciente. Y un total de 67 pacientes.

Como se ha comentado en el apartado anterior se llevaron a cabo diferentes tratamientos de los datos para tratar de obtener los mejores resultados al predecir el tipo de demencia. Para ello fue necesario la elaboración de cinco clasificaciones, a las que denominaremos C1, C2, C3, C4 y C5, teniendo en cuenta las diferentes tipologías expuestas en el apartado 1.3 y los criterios del experto. En la Tabla 3.1 se incluyen las 5 clasificaciones.

Tabla 3.1: Clasificaciones de las demencias empleadas en el estudio

| Clasificación | Tipos de demencias |
|---------------|---|
| C1 | Cortical, Subcortical o Disfunción cortical y subcortical |
| C2 | Degenerativa Primaria o Degenerativa Secundaria |
| C3 | Alzheimer, vascular, Parkinson, Cuerpos de Lewy, Frontotemporal o Mixta |
| C4 | Alzheimer, Vascular o Mixta |
| C5 | Alzheimer u otra |

A continuación, se llevó a cabo la reducción de la dimensionalidad. Este proceso reduce el número de atributos a considerar. Aunque en la literatura existen varias técnicas para este

propósito, en nuestro trabajo, la selección de atributos se ha realizado de acuerdo con los requisitos de nuestro enfoque, es decir, solo teniendo en cuenta aquellos síntomas y causas directamente observables y documentadas por los especialistas. Para esta tarea contamos con la experiencia del médico de la Asociación, responsable de generar los diagnósticos y establecer los tratamientos; y el enfermero, encargado de tratar a diario con los pacientes y responsable del cumplimiento de los tratamientos establecidos. En consecuencia, se seleccionaron del conjunto de datos 9 síntomas (ver Tabla 3.2)

Tabla 3.2: Valores de los atributos asociados a los síntomas y su ocurrencia en la base de datos

| Atributo | Si % | No % |
|----------------------------|-------|------|
| Ansiedad | 44,7 | 55,2 |
| Depresión | 50,7 | 49,2 |
| Trastornos del lenguaje | 100,0 | 0,0 |
| Memoria | 100,0 | 0,0 |
| Problemas de atención | 100,0 | 0,0 |
| Problemas de motricidad | 100,0 | 0,0 |
| Problemas comportamentales | 68,6 | 31,3 |
| Gritos | 29,8 | 70,1 |
| Conducta repetitiva | 29,8 | 70,1 |

Además, con el objetivo de encontrar una relación entre los síntomas precedentes al desarrollo de la demencia y de los cuales se pueda extraer una relación como posibles causas de su aparición en las últimas etapas de la vida, se incluyeron, siguiendo las sugerencias de los expertos, la edad, el sexo, el déficit cognitivo y diferentes patologías padecidas por los pacientes como colesterol, niveles elevados de azúcar, cardiopatía, carencia de hierro, calcio, ácido fólico o vitaminas, problemas de tiroides, diabetes o problemas psicotrópicos. Se subdividieron en grupos de acuerdo a las indicaciones de los expertos, como se muestra en la Tabla 3.3.

Los atributos restantes no utilizados contenían información sobre los diferentes tratamientos tanto farmacológico como no farmacológico. El tratamiento no farmacológico es actualmente el que mejores resultados proporciona a los pacientes mejorando su calidad de vida, sin embargo, puesto que todos los pacientes lo llevaban a cabo, no fue utilizado ya que no se podía obtener ningún dato de interés para este trabajo.

3.2. Preprocesamiento de los datos

Los modelos de clasificación requieren datos categóricos. Por lo tanto, todos los atributos seleccionados deben transformarse en categorías. Los atributos relacionados con los síntomas se transformaron fácilmente ya que originalmente tenían los siguientes valores: sufridos y no sufridos. La segunda y tercera columna de la Tabla 3.2 muestran el porcentaje de individuos del conjunto de datos con el valor correspondiente para cada síntoma.

Tabla 3.3: Otros atributos y su ocurrencia en la base de datos

| Atributo | Posible valor | Ocurrencia % |
|-------------------|--------------------------|--------------|
| Edad | >80 | 55,2 |
| | <80 | 38,8 |
| | <65 | 5,9 |
| Sexo | Hombre | 34,3 |
| | Mujer | 65,6 |
| Déficit cognitivo | Leve | 31,3 |
| | Moderado | 37,3 |
| | Grave | 31,3 |
| Causa | Colesterol | 26,8 |
| | Azúcar | 27,9 |
| | Carencia de hierro | 13,4 |
| | Carencia de calcio | 8,9 |
| | Carencia de Ácido fólico | 11,9 |
| | Carencia de vitamina | 13,4 |
| | Tiroides | 7,4 |
| | Problema cardiovascular | 67,1 |
| | Diabetes | 34,3 |
| | Problema psicotrópico | 70,1 |

El valor 'sí' indica que el paciente está sufriendo el síntoma correspondiente; si el paciente no lo sufre, el valor asignado es 'no'.

El atributo de edad, siguiendo las sugerencias de nuestro experto, se dividió en tres categorías: menor de ochenta años, mayor de ochenta años y menor de sesenta y cinco años. Esta división establece tres rangos de edad en los que la demencia se puede desarrollar siendo el rango de edad más avanzada, el más común. No obstante, como se puede observar, personas con edades inferiores a sesenta y cinco años, pueden desarrollar una demencia.

Para el atributo de sexo se hicieron dos categorías en función de si los pacientes eran mujeres u hombres.

El atributo de déficit cognitivo se extrajo de las bases de datos de la asociación donde se encontraba reflejado para cada individuo, el nivel de gravedad o déficit cognitivo de la demencia que padece.

Con respecto a los atributos correspondientes a las causas, se emplearon los fármacos del tratamiento de cada paciente. Se estudiaron las causas tratadas por los diferentes fármacos y se realizó una clasificación general de las causas más significativas. Para esto se siguieron las sugerencias de los expertos que permitieron reducir la lista de causas a 10 atributos (como se observa en la Tabla 3.3), siendo estos los más comunes entre todos los pacientes y los más relevantes respecto a las demencias. Cada causa de la primera columna se corresponde con un valor en porcentaje de la segunda columna de la tabla. Este valor indica el número de los pacientes totales que la padecen.

Inicialmente la base de datos contaba con 80 registros. Sin embargo, algunos atributos de varios pacientes no se encontraban especificados en la documentación y las bases de datos empleadas por lo que estos registros fueron eliminados quedando una base de datos de

67 individuos con 9 atributos de síntomas, 10 atributos de causas, 1 atributo de edad, 1 atributo de sexo y 1 atributo de déficit cognitivo.

3.3. Análisis de los datos

El análisis estadístico de los datos se llevó a cabo con la herramienta de software Weka (Michalski et al., 2013). Se usaron varios modelos de clasificación para entrenar el conjunto de datos. Concretamente, se usó J48 DT, SVM (SMO), RF, NB y BN. Todas estas técnicas intentan encontrar un modelo capaz de predecir la variable objetivo, que en nuestro caso es el tipo de demencia. Este modelo usa como variables de predicción los atributos relacionados con la ocurrencia de los nueve síntomas y las 10 causas, la edad, el déficit cognitivo y el sexo. El tipo demencia es la variable objetivo.

Para medir la calidad de los resultados se utilizaron diferentes indicadores, empleados comúnmente en el aprendizaje supervisado. La precisión global que, teniendo en cuenta la especificidad y sensibilidad, permite concluir si el diagnóstico obtenido es el adecuado. El análisis ROC permite evaluar el rendimiento de un clasificador independientemente de la distribución de clases o los costes de error. Los valores superiores a 0,8 se consideran buenos y los valores superiores a 0,9 se consideran excelentes.

El proceso de validación es necesario para determinar si el modelo construido puede o no proporcionar estimaciones precisas de las demencias. La validación cruzada es una de las técnicas de validación más aceptadas en los proyectos de inteligencia artificial y se utiliza para estimar la precisión de un modelo que se implementará en la práctica. Su objetivo es garantizar que los resultados de un modelo de predicción sean independientes de la partición entre el entrenamiento y los datos de prueba. La forma estándar de predecir la tasa de error de una técnica de aprendizaje dada una única muestra fija de datos es usar la validación cruzada K. Los datos se dividen en k subconjuntos en los que la clase se representa en aproximadamente las mismas proporciones que en el conjunto de datos completo de modo que cada vez, uno de los k subconjuntos se utiliza como el conjunto de prueba / conjunto de validación y los otros subconjuntos k-1 forman el conjunto de entrenamiento. La estimación del error se promedia en todos los k ensayos para obtener la efectividad total del modelo. De esta forma, cada dato forma parte del conjunto de validación exactamente una vez, y del conjunto de entrenamiento de k-1 veces. Esto reduce significativamente el sesgo ya que estamos usando la mayoría de los datos para la adaptación, y también reduce significativamente la varianza ya que la mayoría de los datos también se utilizan en el conjunto de validación. El intercambio de los conjuntos de entrenamiento y prueba también aumenta la efectividad de este método. La elección del número de iteraciones depende de la extensión del conjunto de datos. En general, se elige una validación cruzada de 10 veces.

Capítulo 4

Resultados

Las Tablas 4.1, 4.2, 4.3, 4.4 y 4.5 muestran los resultados de rendimiento de los algoritmos de clasificación que se han utilizado en este estudio para cada una de las cinco clasificaciones, respectivamente.

Tabla 4.1: Resultados de rendimiento para los algoritmos de clasificación para C1

| C1 | Cortical | | Subcortical | | Cortical y Subcortical | |
|--------|--------------------|--------------------|--------------------|--------------------|------------------------|--------------------|
| Modelo | Sensit. TP-Rate | Specif. Fp-Rate | Sensit. TP-Rate | Specif. Fp-Rate | Sensit. TP-Rate | Specif. Fp-Rate |
| J48 | 0,936 | 1,000 | 0,000 | 0,048 | 0,000 | 0,000 |
| RF | 0,894 | 1,000 | 0,000 | 0,000 | 0,000 | 0,000 |
| SMO | 0,936 | 1,000 | 0,000 | 0,016 | 0,000 | 0,039 |
| NB | 0,809 | 0,950 | 0,000 | 0,016 | 0,063 | 0,157 |
| BN | 0,766 | 0,900 | 0,000 | 0,032 | 0,125 | 0,176 |

Tabla 4.2: Resultados de rendimiento para los algoritmos de clasificación para C2

| C2 | Degenerativa Primaria | | Degenerativa Secundaria | |
|--------|-----------------------|--------------------|-------------------------|--------------------|
| Modelo | Sensit. TP-Rate | Specif. Fp-Rate | Sensit. TP-Rate | Specif. Fp-Rate |
| J48 | 1,000 | 1,000 | 0,000 | 0,000 |
| RF | 0,964 | 0,909 | 0,091 | 0,036 |
| SMO | 0,982 | 1,000 | 0,000 | 0,018 |
| NB | 0,857 | 0,909 | 0,091 | 0,143 |
| BN | 0,839 | 0,818 | 0,182 | 0,161 |

Tabla 4.3: Resultados de rendimiento para los algoritmos de clasificación para C3

| C3 | Alzheimer | | Vascular | | Parkinson | | Cuerpos de Lewy | | Frontotemporal | | Mixta | |
|--------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| Modelo | Sensit. TP-Rate | Specif. Fp-Rate | Sensit. TP-Rate | Specif. Fp-Rate | Sensit. TP-Rate | Specif. Fp-Rate | Sensit. TP-Rate | Specif. Fp-Rate | Sensit. TP-Rate | Specif. Fp-Rate | Sensit. TP-Rate | Specif. Fp-Rate |
| J48 | 0,978 | 0,909 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,048 | 0,000 | 0,000 |
| RF | 0,911 | 1,000 | 0,000 | 0,033 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,016 | 0,000 | 0,016 |
| SMO | 0,956 | 0,864 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,250 | 0,032 | 0,000 | 0,032 |
| NB | 0,844 | 0,909 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,250 | 0,048 | 0,000 | 0,081 |
| BN | 0,822 | 0,818 | 0,000 | 0,000 | 0,000 | 0,016 | 0,000 | 0,000 | 0,250 | 0,079 | 0,000 | 0,081 |

Como se observan en las cinco tablas, dependiendo de la técnica de clasificación, los valores de sensibilidad y especificidad oscilan entre 0 y 1, siendo en su mayoría valores bajos.

Tabla 4.4: Resultados de rendimiento para los algoritmos de clasificación para C4

| C4 | Alzheimer | | Vascular | | Mixta | |
|---------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| Modelo | Sensit. TP-Rate | Specif. Fp-Rate | Sensit. TP-Rate | Specif. Fp-Rate | Sensit. TP-Rate | Specif. Fp-Rate |
| J48 | 1,000 | 1,000 | 0,000 | 0,000 | 0,000 | 0,000 |
| RF | 0,909 | 1,000 | 0,000 | 0,051 | 0,000 | 0,032 |
| SMO | 0,982 | 1,000 | 0,000 | 0,017 | 0,000 | 0,000 |
| NB | 0,927 | 1,000 | 0,000 | 0,051 | 0,000 | 0,016 |
| BN | 0,855 | 0,750 | 0,125 | 0,102 | 0,000 | 0,063 |

Tabla 4.5: Resultados de rendimiento para los algoritmos de clasificación para C5

| C5 | Alzheimer | | Otra | |
|---------------|--------------------|--------------------|--------------------|--------------------|
| Modelo | Sensit. TP-Rate | Specif. Fp-Rate | Sensit. TP-Rate | Specif. Fp-Rate |
| J48 | 0,894 | 0,850 | 0,150 | 0,106 |
| RF | 0,872 | 0,900 | 1,000 | 0,128 |
| SMO | 0,872 | 0,950 | 0,050 | 0,128 |
| NB | 0,745 | 0,850 | 0,150 | 0,255 |
| BN | 0,702 | 0,800 | 0,200 | 0,298 |

Esto nos lleva a concluir que la clasificación no es buena y, por lo tanto, no se ha obtenido un clasificador que nos permita predecir exitosamente el tipo de demencia de un paciente.

A continuación, se desarrolló un clasificador atendiendo a la gravedad o déficit cognitivo de la demencia. De este modo, en lugar de predecir el tipo de demencia, sería capaz de predecir la gravedad de la demencia con el conjunto de datos inicial. Las Tablas 4.6 , 4.7 y 4.8 muestran los resultados de rendimiento de los algoritmos de clasificación que se han utilizado en este segundo estudio para la clasificación 5.

Tabla 4.6: Resultados de rendimiento para los algoritmos de clasificación según el déficit cognitivo para C5

| Déficit cognitivo | Leve | | Moderado | | Grave | |
|--------------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| Modelo | Sensit. TP-Rate | Specif. Fp-Rate | Sensit. TP-Rate | Specif. Fp-Rate | Sensit. TP-Rate | Specif. Fp-Rate |
| J48 | 1,000 | 0,022 | 0,960 | 0,024 | 0,950 | 0,000 |
| RF | 1,000 | 0,022 | 0,960 | 0,024 | 0,950 | 0,000 |
| SMO | 1,000 | 0,022 | 0,960 | 0,024 | 0,950 | 0,000 |
| NB | 0,950 | 0,043 | 0,920 | 0,048 | 0,950 | 0,000 |
| BN | 0,950 | 0,043 | 0,920 | 0,048 | 0,950 | 0,000 |

Como se puede ver, todas las técnicas de clasificación proporcionan buenos resultados. La sensibilidad para las tres categorías, es decir, Leve, Moderado y Grave exhiben valores alrededor de 0.96. Los valores de precisión y recuperación son superiores a 0,94 en todos los casos. Lo mismo se cumple para el área ROC cuyos valores son superiores a 0.90 y por lo tanto, los resultados pueden considerarse excelentes. Todas las técnicas de clasificación arrojaron una precisión de clasificación superiores al 90 %. Aunque todos los criterios

Tabla 4.7: Resultados promedio de rendimiento de los algoritmos de clasificación

| Modelo | Precision | Recall | F-Measure | ROC Area |
|--------|-----------|--------|-----------|----------|
| J48 | 0,971 | 0,970 | 0,970 | 0,963 |
| RF | 0,971 | 0,970 | 0,970 | 0,988 |
| SMO | 0,971 | 0,970 | 0,970 | 0,974 |
| NB | 0,942 | 0,940 | 0,941 | 0,978 |
| BN | 0,942 | 0,940 | 0,941 | 0,977 |

Tabla 4.8: Clasificación de instancias de algoritmos de clasificación

| Modelo | %Instancias clasificadas correctamente | %Instancias clasificadas incorrectamente |
|--------|---|---|
| J48 | 97,01 | 2,98 |
| RF | 97,01 | 2,98 |
| SMO | 97,01 | 2,98 |
| NB | 94,02 | 5,97 |
| BN | 94,02 | 5,97 |

apuntan al hecho de que cualquiera de las técnicas de clasificación utilizadas en este estudio sería adecuada para un diagnóstico de la gravedad de la demencia, se selecciona el J48 DT como la mejor opción debido a los resultados de rendimiento y las características que hacen que los DT sean los modelos más completos para usuarios sin un conocimiento experto en técnicas de clasificación basadas en el aprendizaje automático. La Figura 4.1 muestra el DT obtenido después de aplicar el algoritmo J48. Observamos que no todos los atributos de los síntomas seleccionados por el experto están presentes en el árbol. Se eliminan atributos menos discriminatorios, dejando un subconjunto de atributos que permiten la discriminación correcta entre las clases. El árbol muestra que los gritos es el atributo más discriminante. Para cualquier edad, se observa que, si el paciente experimenta fases de gritos, quiere decir que la demencia se diagnostica como grave. En caso contrario, otro síntoma como los problemas comportamentales, es el determinante para distinguir si la demencia se considera con déficit cognitivo leve o moderado.

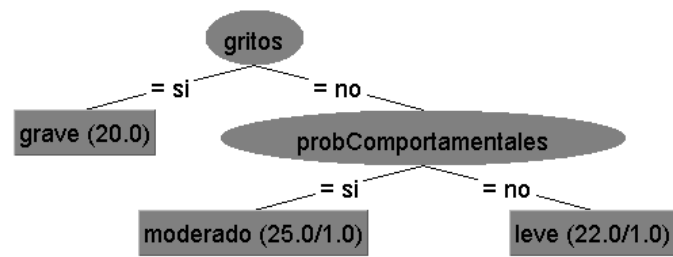


Figura 4.1: Árbol de decisión resultante después de aplicar el clasificador J48

Capítulo 5

Conclusiones

El objetivo de este estudio ha sido construir modelos de decisión basados en datos para ayudar al sector sanitario en el diagnóstico de demencias. Los resultados obtenidos permiten arrojar varias conclusiones. Por un lado, se ha comprobado la dificultad existente para determinar un tipo de demencia. Existen numerosas demencias dependiendo de los criterios adoptados para clasificarlas y en la sociedad existe dificultad para establecer un diagnóstico partiendo de los síntomas debido a que los enfermos comparten síntomas similares.

Como permiten concluir los resultados obtenidos en las Tablas 4.1, 4.2, 4.3, 4.4 y 4.5, no se ha conseguido obtener un clasificador que ayude a determinar el tipo de demencia de un paciente partiendo de los síntomas experimentados por este.

Por otro lado, se ha conseguido obtener un clasificador que ayude a la predicción del déficit cognitivo de demencias. Actualmente, el conocimiento deficiente de los déficits tiene implicaciones en diversos aspectos de la atención del paciente, como la autopresentación oportuna para el tratamiento, el cumplimiento del tratamiento, la seguridad personal, la autonomía financiera y la situación o comportamiento a adoptar por el cuidador. Nuestro estudio contribuye a la predicción de la gravedad de la demencia con un modelo de decisión ayudando de esta forma a reducir las dificultades encontradas en los aspectos comentados.

Respecto a la relación entre enfermedades como la diabetes o la hipertensión y el desarrollo de demencias en las fases finales de la vida se comprueba que un elevado porcentaje de los pacientes participantes en este trabajo posee alguna de las dos patologías. EL 67 % de los enfermos padecen hipertensión y el más del 30 % posee diabetes. Sin embargo, se observa que el resto de patologías como deficiencia de vitaminas, carencia de calcio o niveles elevados de azúcar son factores menos determinantes para el desarrollo de demencias.

Se ha examinado el rendimiento de cinco técnicas de clasificación aplicadas a nueve síntomas clínicos de un paciente, su edad y sexo. Todos esos modelos son capaces de predecir la gravedad de la demencia sufrida por el paciente. Por lo tanto, se ha conseguido aportar

una herramienta para el apoyo al diagnóstico de las demencias.

Bibliografía

- Anooj, P. (2012). Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules. *Journal of King Saud University-Computer and Information Sciences*, 24(1):27–40.
- Babaoglu, I., Baykan, O. K., Aygul, N., Ozdemir, K., and Bayrak, M. (2009). Assessment of exercise stress testing with artificial neural network in determining coronary artery disease and predicting lesion localization. *Expert Systems with Applications*, 36(2):2562–2566.
- Barrett, A. and Burns, A. (2014). Dementia revealed what primary care needs to know. *Hardwick CCG, NHS England, Department of Health, Royal College of General Practitioners*.
- Bashir, S., Qamar, U., Khan, F. H., and Javed, M. Y. (2014). Mv5: a clinical decision support framework for heart disease prediction using majority vote based classifier ensemble. *Arabian Journal for Science and Engineering*, 39(11):7771–7783.
- Bishara, D. and Harwood, D. (2014). Safe prescribing of physical health medication in patients with dementia. *International journal of geriatric psychiatry*, 29(12):1230–1241.
- Eom, J.-H., Kim, S.-C., and Zhang, B.-T. (2008). Aptacdss-e: A classifier ensemble-based clinical decision support system for cardiovascular disease level prediction. *Expert Systems with Applications*, 34(4):2465–2479.
- Farion, K., Michalowski, W., Wilk, S., O’Sullivan, D., and Matwin, S. (2010). A tree-based decision model to support prediction of the severity of asthma exacerbations in children. *Journal of medical systems*, 34(4):551–562.
- Jun, S., Park, B., Seo, J. B., Lee, S., and Kim, N. (2018). Development of a computer-aided differential diagnosis system to distinguish between usual interstitial pneumonia and non-specific interstitial pneumonia using texture-and shape-based hierarchical classifiers on hrcr images. *Journal of digital imaging*, 31(2):235–244.
- Kim, J., Lee, J., and Lee, Y. (2015). Data-mining-based coronary heart disease risk prediction model using fuzzy logic and decision tree. *Healthcare informatics research*, 21(3):167–174.

- Kim, J.-K., Lee, J.-S., Park, D.-K., Lim, Y.-S., Lee, Y.-H., and Jung, E.-Y. (2014). Adaptive mining prediction model for content recommendation to coronary heart disease patients. *Cluster computing*, 17(3):881–891.
- Michalski, R. S., Carbonell, J. G., and Mitchell, T. M. (2013). *Machine learning: An artificial intelligence approach*. Springer Science & Business Media.
- Mohammed, E. A., Mohamed, M. M., Naugler, C., and Far, B. H. (2017). Toward leveraging big value from data: chronic lymphocytic leukemia cell classification. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 6(1):6.
- Mokeddem, S. A. (2018). A fuzzy classification model for myocardial infarction risk assessment. *Applied Intelligence*, 48(5):1233–1250.
- Takada, M., Sugimoto, M., Naito, Y., Moon, H.-G., Han, W., Noh, D.-Y., Kondo, M., Kuroi, K., Sasano, H., Inamoto, T., et al. (2012). Prediction of axillary lymph node metastasis in primary breast cancer patients using a decision tree-based model. *BMC medical informatics and decision making*, 12(1):54.
- Temurtas, H., Yumusak, N., and Temurtas, F. (2009). A comparative study on diabetes disease diagnosis using neural networks. *Expert Systems with applications*, 36(4):8610–8615.
- Übeyli, E. D. (2009). Combined neural networks for diagnosis of erythematous-squamous diseases. *Expert Systems with Applications*, 36(3):5107–5112.
- Yanguas, J. J. (2007). *Modelo de atención a las personas con enfermedad de Alzheimer*. Number 21011. Imsero.