

# **Predicción de resultados de ataque cerebrovascular mediante el análisis de series temporales fisiológicas**

**LUIS GARCÍA TERRIZA**

**MÁSTER EN INGENIERÍA INFORMÁTICA**

**FACULTAD DE INFORMÁTICA**

**DEPARTAMENTO DE ARQUITECTURA DE  
COMPUTADORES Y AUTOMÁTICA**

**UNIVERSIDAD COMPLUTENSE DE MADRID**



**TRABAJO FIN DE MÁSTER**

**Curso 2017-2018**

**Febrero 2018**

**Calificación: 8.5 - Notable**

**Directores:**

José Luis Risco Martín

José Luis Ayala Rodrigo



# Autorización de difusión y utilización

El abajo firmante, matriculado en el Máster de Ingeniería Informática de la Facultad de Informática, autoriza a la Universidad Complutense de Madrid (UCM) a difundir y utilizar con fines académicos, no comerciales y mencionando expresamente a su autor el presente Trabajo de Fin de Máster “*Predicción de resultados de ataque cerebrovascular mediante el análisis de series temporales fisiológicas*”, realizado durante el curso académico 2017-2018 bajo la dirección de José Luis Risco Martín y con la colaboración de José Luis Ayala Rodrigo en el Departamento de Arquitectura de Computadores y Automática, y a la Biblioteca de la UCM a depositarlo en el Archivo Institucional E-Prints Complutense con el objeto de incrementar la difusión, uso e impacto del trabajo en Internet y garantizar su preservación y acceso a largo plazo.

Luis García Terriza

Febrero de 2018, Madrid



# Agradecimientos

A mis directores, José Luis Risco Martín y José Luis Ayala Rodrigo, por darme la oportunidad de realizar el presente trabajo, por proporcionarme apoyo, dedicación continua y enseñarme tanto a lo largo de todo este tiempo. A ambos, muchas gracias.

A mi familia, que me ha apoyado desde que entré a la universidad y sin su ayuda no habría sido posible nada de esto.

Muchas gracias.



# Índice

<b>Índice de figuras</b>	<b>8</b>
<b>Índice de tablas</b>	<b>9</b>
<b>Resumen</b>	<b>10</b>
<b>Abstract</b>	<b>12</b>
<b>Introducción</b>	<b>14</b>
<b>Introduction</b>	<b>18</b>
<b>Visión general</b>	<b>21</b>
Regresión Logística	24
Support Vector Machines	25
Naive Bayes	26
K - Nearest Neighbors	26
Árboles de decisión	27
Redes Neuronales Artificiales	27
<b>Metodología</b>	<b>30</b>
Copia de respaldo y formateo de datos	30
Análisis de características	30
Preprocesamiento de datos y normalización	31
Entrenamiento y validación	32
<b>Resultados</b>	<b>36</b>
Predicción del tipo de ataque cerebrovascular	36
Pruebas realistas con nuevos pacientes	42
Predicción de pronóstico	46
Predicción de Resangrado	52
<b>Conclusiones</b>	<b>59</b>
<b>Conclusions</b>	<b>62</b>



# Índice de figuras

Figura 1. Ejemplo de mensaje HL7	22
Figura 2. Red neuronal MLP con una capa oculta	28
Figura 3. Esquema de transformación de datos con varias observaciones por instancia	33
Figura 4. Random Cross-Validation	34
Figura 5. Comparativa de algoritmos predicción de ataque cerebrovascular	40
Figura 6. Comparativa de mejores algoritmos predicción de ataque cerebrovascular en función de la ventana de monitorización	41
Figura 7. Comparativa de algoritmos predicción de pronóstico	49
Figura 8. Comparativa de mejores algoritmos predicción de pronóstico en función de la ventana de monitorización	51
Figura 9. Comparativa de algoritmos predicción de resangrado	56
Figura 10. Comparativa de mejores algoritmos predicción de resangrado en función de la ventana de monitorización	57

# Índice de tablas

Tabla 1. Resultados tipo de predicción de ataque cerebrovascular con 1 observación	37
Tabla 2. Resultados tipo de predicción de ataque cerebrovascular con 2 observaciones	37
Tabla 3. Resultados tipo de predicción de ataque cerebrovascular con 3 observaciones	38
Tabla 4. Resultados tipo de predicción de ataque cerebrovascular con 4 observaciones	38
Tabla 5. Resultados tipo de predicción de ataque cerebrovascular con 5 observaciones	39
Tabla 6. Comparativa mejores algoritmos con pacientes reales con ventana de tiempo y sin ventana	43
Tabla 7. Ataque cerebrovascular de tipo isquémico con pacientes reales y ventanas de minutos	44
Tabla 8. Ataque cerebrovascular de tipo hemorrágico con pacientes reales y ventanas de	45
Tabla 9. Resultados pronóstico tras ataque cerebrovascular con 1 observación	46
Tabla 10. Resultados pronóstico tras ataque cerebrovascular con 2 observaciones	47
Tabla 11. Resultados pronóstico tras ataque cerebrovascular con 3 observaciones	47
Tabla 12. Resultados pronóstico tras ataque cerebrovascular con 4 observaciones	48
Tabla 13. Resultados pronóstico tras ataque cerebrovascular con 5 observaciones	48
Tabla 14. Resultados de predicción de resangrado tras ataque cerebrovascular con 1 observación	52
Tabla 15. Resultados de predicción de resangrado tras ataque cerebrovascular con 2 observaciones	53
Tabla 16. Resultados de predicción de resangrado tras ataque cerebrovascular con 3 observaciones	54
Tabla 17. Resultados de predicción de resangrado tras ataque cerebrovascular con 4 observaciones	54
Tabla 18. Resultados de predicción de resangrado tras ataque cerebrovascular con 5 observaciones	55

# Resumen

En la actualidad, los ataques cerebrovasculares, también conocidos como ictus, son una de las principales causas de muerte en hombres y mujeres. El ictus tiene dos causas principales: ictus isquémico por obstrucción de un vaso sanguíneo, e ictus hemorrágico por la rotura de éstos.

En este Trabajo de Fin de Máster se utilizan técnicas conocidas de inteligencia artificial actuales como es el aprendizaje automático (Machine Learning) para realizar predicciones sobre pacientes que han sufrido un ataque cerebrovascular y poder tomar decisiones precozmente gracias a los modelos predictivos.

Los datos de los pacientes se obtienen de la Unidad de Ictus del Hospital Universitario de La Princesa, donde se monitoriza a los pacientes desde que ingresan hasta que son dados de alta o fallecen.

Se plantea la posibilidad de poder predecir distintas complicaciones o características por medio del procesamiento de los datos monitorizados de los pacientes como series temporales. Algunas de las predicciones planteadas son: si se va a producir un resangrado, el tipo de ataque cerebrovascular sufrido o predecir si el paciente va a sobrevivir al ataque cerebrovascular.

Además, también se tiene como objetivo encontrar los algoritmos de clasificación que mejor encajan en los problemas de predicción a tratar dadas las características del problema, por lo que se estudiará prácticamente cada tipo de algoritmo de aprendizaje supervisado viendo cómo se comporta tanto en resultados como en capacidad de cómputo.

**Palabras clave:** Aprendizaje automático, Ataque cerebrovascular, Ictus, Aprendizaje Supervisado, Inteligencia Artificial, Series temporales, Biomedicina



# Abstract

Nowadays, strokes are one of the main death reasons among men and women. A Stroke has two main causes: ischaemic stroke by blockage of blood vessel and hemorrhagic stroke by break of blood vessels.

In this Master's thesis some recognized technics from current artificial intelligence areas like machine learning have been used to perform predictions over patients that have suffered a stroke and be able to take precocious decisions thanks to predictive models.

Patients' data has been obtained from Stroke Unit of Hospital Universitario 'La Princesa' where patients are monitored from entering the hospital until they are discharged or they have passed away.

Predict different complications or characteristics are some of the goals of this work by processing monitored data from the patients as temporal series. Some of the proposed predictions to be carried out are predict if the patient is going to have a rebleeding, predict the sort of stroke suffered and predict if the patient is going to survive the stroke.

Furthermore, finding the best classifiers from supervised learning for these kind of predictions is another relevant goal. Each prediction may be different and each algorithm will be analyzed and studied in means of computing power and precision in order to see how well fits to the prediction problem.

**Keywords:** Machine Learning, Stroke attack, Ictus, Supervised learning, Artificial Intelligence, Temporal series, Biomedicine



# 1. Introducción

En España, el ictus es la primera causa de muerte en la mujer y la segunda en el hombre. En nuestro país, se producen entre 150-200 nuevos ictus por 100.000 habitantes por año, lo que supone alrededor de 100.000 nuevos casos al año. El ictus tiene dos causas principales: la obstrucción de un vaso sanguíneo (ictus isquémico que representa 85% de los casos) y la hemorragia (ictus hemorrágico, por rotura de vasos, 15% restante) [1].

Los síntomas del ictus (pérdida de fuerza, dificultad para hablar o entender, pérdida parcial de la visión, vértigo, etc.) aparecen de forma brusca. Sin embargo, ha de transcurrir un tiempo hasta que se produce la lesión cerebral por muerte neuronal. Este periodo, conocido como la “ventana terapéutica”, es normalmente las primeras 24 horas y hasta los primeros días tras un ictus. En este periodo es muy importante una correcta monitorización de las constantes vitales de los pacientes, las cuales sirven para detectar precozmente complicaciones y tomar decisiones terapéuticas. Los neurólogos deben evitar la muerte neuronal en la parte del cerebro dañada durante este periodo.

En este proyecto se propone la aplicación de técnicas de inteligencia artificial conocidas como el aprendizaje automático, en el cual se encuentran dos vertientes principales dentro del aprendizaje supervisado, que son el aprendizaje para clasificación y para regresión. Estos dos enfoques son parecidos, ambos tratan de predecir a partir de los datos de los que se “*aprende*”; sin embargo, en la clasificación se trata de “etiquetar” los ejemplos dados en un número conocido de etiquetas mientras que en la regresión se trata calcular un número, por ejemplo, predecir la temperatura.

Por medio de estas técnicas se pretende predecir diversos resultados de ataques cerebrovasculares en pacientes partiendo de datos biométricos de pacientes ingresados en la Unidad de Ictus del Hospital Universitario de La Princesa.

Existen algunos trabajos similares en los que se detallan las investigaciones y los resultados obtenidos.

Con el objetivo de detección temprana del ataque cerebrovascular, en [10], Villar *et al* desarrollaron un dispositivo de detección de ataques cerebrovasculares basada en reconocimiento de actividad, es decir, basado en sus movimientos. El dispositivo implementa dos algoritmos de

aprendizaje automático que, aplicando PCA (Principal Component Analysis), era capaz de detectar si la persona estaba experimentando una actividad distinta a la habitual. El dispositivo emite una alerta en el caso de detectar patrones de actividad inusuales para que sea evaluado y aplicado el tratamiento que sea necesario.

En otro estudio [9], Birkner *et al* utilizaron variables fisiológicas de pacientes como SBP (Systolic Blood Pressure), temperatura corporal junto con ciertos síntomas que padecían, como por ejemplo síntomas de incontinencia urinaria o respiración anormal, para predecir por medio de modelos predictivos si fallecería a los 30 días de haber sufrido el accidente.

También hay estudios de detección de secuelas basadas en aprendizaje automático aplicado a imágenes, como se realizó en [11], donde Chen *et al* analizaron TAC (Tomografía Axial Computarizada) por medio de algoritmos de aprendizaje automático para predecir las secuelas del ataque cerebrovascular.

En [2], Zhang *et al.* realizan un estudio en el que, a partir de datos de variables de ataques cerebrovasculares de tipo isquémico durante las 48 primeras horas de monitorización, se pretende predecir por medio de un algoritmo de clasificación (Regresión Logística) las secuelas del paciente a los 3 meses después del ataque. Además, realizan *feature selection* con el objetivo de encontrar las variables fisiológicas más predictivas.

En [5], Asadi *et al* utilizaron una base de datos con 107 pacientes sobre su circulación previa y posterior a un ataque cerebrovascular que además se sometieron a terapia intra-arterial. Procesando los datos con distintas herramientas y algoritmos como redes neuronales artificiales y *support vector machines* obtuvieron hasta un 70% de precisión prediciendo si las secuelas serían potencialmente buenas o malas.

Sin embargo, la solución propuesta en este trabajo se asimila más a [2]. No obstante, presentan varias diferencias: en [2] utilizan como variables fisiológicas predictivas únicamente cinco (nivel de azúcar en sangre, presión diastólica, presión sistólica, ritmo cardíaco y la temperatura corporal) mientras que en este trabajo se tienen en cuenta 7 variables fisiológicas, coincidiendo algunas de las variables fisiológicas utilizadas. Otra gran diferencia es que se tiene en cuenta no sólo las primeras 48 horas sino todo el tiempo en el que está ingresado el paciente. Además, en este trabajo el objetivo no es predecir las posibles secuelas que sufre el paciente, sino el tipo de ataque cerebrovascular sufrido, que vuelva a sufrir otro ataque (resangrado) y el pronóstico final (sobrevive o no).

El desarrollo realizado podrá ser de vital repercusión en el seguimiento de esta patología. Este trabajo se ha realizado en colaboración estrecha con el Servicio de Neurología del Hospital Universitario de La Princesa, donde se realiza una monitorización de los pacientes ingresados a través de un dispositivo de monitorización de constantes vitales (estimación del ritmo, frecuencia cardiaca, frecuencia respiratoria y saturación de oxígeno en sangre SpO2 entre otros parámetros biométricos) que automáticamente registra los datos medidos del paciente en una base de datos relacional. Con la información de sus constantes vitales y la afección sufrida se parte de la hipótesis de que se pueden predecir ciertas características de un ataque cerebrovascular de un paciente, como podrían ser el tipo de ataque que ha sufrido o el pronóstico del paciente.

Para realizar la predicción se utilizarán técnicas del campo de inteligencia artificial conocido como aprendizaje automático (Machine Learning) en el que se crean modelos de predicción con distintos algoritmos a partir de los datos recopilados como series temporales de los pacientes.

Como objetivos iniciales, se plantea la posibilidad de poder predecir el tipo de ataque que ha sufrido el paciente sin necesidad de recurrir a la prueba de imagen, predecir procesos de resangrado, en el cual el paciente tiende a empeorar debido a un nuevo ataque, y por último poder predecir un pronóstico final del paciente a modo informativo.

Estos objetivos dependen de la cantidad de información que se recoge del paciente y de si realmente existen indicadores reales para cada tipo de predicción en los datos monitorizados como series temporales, de forma que la posibilidad de conseguir resultados positivos se plantea como un trabajo de investigación con cierta incertidumbre. De igual forma es cierto que los médicos sugieren que hay indicadores y patrones subyacentes en los datos monitorizados que son claves en la predicción de los resultados del paciente, por lo que aunque a priori se plantea como una hipótesis, es probable obtener resultados determinantes que puedan servir en un futuro para una detección más rápida o un análisis del paciente más amplio.

Esta memoria se ha dividido en 4 capítulos aparte de la introducción ya expuesta (Capítulo 1):

- Capítulo 2: Visión general. Se aporta información actual del estado del arte explicando algunos trabajos ya realizados en este ámbito y se explican algunas de las técnicas existentes relacionadas con el trabajo realizado en este trabajo.
- Capítulo 3: Metodología. En este apartado se explica detalladamente en qué se ha basado este estudio, explicando de qué datos se parten, qué se hace con los datos que se tienen y qué pruebas se han realizado.

- Capítulo 4: Resultados. Se muestran todos los resultados de las pruebas realizadas explicando ampliamente por qué se han obtenido estos resultados.
- Capítulo 5: Conclusiones. Se detallan las conclusiones que se pueden obtener de este trabajo y posibles trabajos futuros a realizar.

# 1. Introduction

In Spain, strokes are the first death cause in women and the second death cause in men. In our country, Spain, there are about 150-200 stroke attacks per 100.000 people per year, what is like 100.00 new cases every year. Strokes have two main reasons: the obstruction of a blood vessel (an ischaemic attack that represents approximately the 85% of the cases) and the haemorrhagic (haemorrhagic attack, by break of blood vessels, remaining 15%) [1].

The symptoms of strokes (loss of strength, difficulty to talk or understand, partial loss of the vision, vertigo, etc.) appear abruptly. However, it must pass a time until a brain lesion is caused by neuronal death. This period, known as the therapeutic window, is normally within the first 24 hours and until the first days after a stroke. In this period it is very important a correct monitoring of the vital constants of the patients, that are used to detect precociously possible complications and take therapeutic decisions. The neurologists must avoid the neural death in the part of the brain damaged in this period.

In this project it is proposed the application of artificial intelligence techniques known as machine learning, in which there are two main variants in supervised learning, that are the learning for classification and learning for regression. These two approaches are similar, because both try to predict “learning” from data, however, in the classification the objective is to “tag” the given samples in a known number of tags while the regression tries to calculate a number like, for example, predict the temperature of tomorrow in an area.

By using these techniques the aim is to predict some results from stroke attacks on patients from the biometrical data of patients admitted to the Stroke Unit of the Hospital Universitario de La Princesa.

There are some similar works that detail the researches and results obtained in this area.

With the aim of the early detection of the stroke attack, in [10] Villar *et al.* developed a device for detection of stroke attacks based in activity recognition, that is to say, based on the movements. The device implements two machine learning algorithms that, applying PCA (Principal Component Analysis), it was able to detect if the person was experiencing an activity different to the usual. The device emits an alert in case of detecting unusual activity patterns to be evaluated by medical professionals and be treated with the appropriate medication.

In another research [9], Birkner *et al.* used physiological variables of patients like SBP (Systolic Blood Pressure), body temperature along with certain symptoms that they suffered, like for example symptoms of urine incontinence or abnormal breathing, to predict by using predictive models if the patients would pass away 30 days after suffering the attack.

There are also studies of outcomes prediction based on machine learning applied to images, like [11] performed, where Chen *et al.* analysed CAT's (Computed Axial Tomography) by using machine learning algorithms to predict the outcomes of the stroke attack.

In [2], Zhang *et al.* perform a study where, from data of variables of ischaemic stroke attacks for the first 48 hours of monitoring, the aim is to predict by using a classification algorithm (Logistic Regression) the outcomes of the patients three months after the attack. Furthermore, they do a feature selection in order to find the physiological variables more predictive.

In [5], Asadi *et al.* used a database with 107 patients containing the information of their anterior and posterior circulation that furthermore were subjected to intra-arterial therapy. Processing the data with several tools and algorithms like artificial neural networks and support vector machines obtained up to a 70% of precision predicting if the outcomes would be potentially good or bad.

Although, the proposed solution in this work is more close to [2]. However, there are some presented differences: in [2] use as predictive physiological variables only five (sugar level in blood, diastolic pressure, systolic pressure, heart rate and body temperature) while in this work there are taken into account 7 hemodynamic variables, some of them being the same. Another great difference is that in this work is not only take into account the first 48 hours but the whole time the patient is admitted being monitored. Furthermore, in this work the aim is not to predict the possible outcomes or the grade of these outcomes, but the aim is diagnostic the stroke attack suffered by the patient along with others predictions like if the patient is going to suffer another attack (re-bleeding) and which is going to be the final prognostic (that is, if the patient is going to survive or not).

The development of this work may be of vital repercussion in the monitoring of this pathology. This work has been performed closely to the neurological service of the Hospital Universitario de la Princesa, where patients are monitored since they are admitted through a monitoring device of vital constants (heart rate estimation, heart rate, breathing rate and SpO<sub>2</sub> between other biometrical parameters) that automatically save the measured data from the patient in a relational database. By using the information of their vital constants and the suffered affection it started out the hypothesis that can be diagnosed some characteristics of a stroke attack suffered by a patient, like could be the

type of the suffered attack or the prognostic of the patient.

To do the predictions, technics from area of artificial intelligence known as machine learning (specifically supervised learning) will be used, in which different algorithms models are created from data collected as temporal series from the patients.

As initial objectives, the possibility of being able to diagnosis the type of stroke attack that the patient has suffered with needing a CAT test, predict re-bleeding processes where the patient tends to get worse due to a new attack, and finally to be able to predict the final prognosis of the patient informatively.

These objectives depends on the amount of information that is collected from the patient and if there are real indicators or predictors for each prediction in the monitored data as temporal series, so that the possibility of getting positive results is proposed as a research work with high uncertainty. Anyway it is true that medical professionals suggest that there are indicators and patterns under the monitored data that are really important in the prediction of the results of the patient, so that although a priori this is proposed as a hypothesis, it is likely getting determinant results that can be useful in the future for a faster diagnosis or a wider analysis of the patients.

This thesis has been divided in 4 chapters excluding the introduction here exposed (Chapter 1):

- Chapter 2: **General vision**. It is given an actual state of art explaining some of the already researches done in this area and some existing methods related to the work done in this work.
- Chapter 3: **Methodology**. In this chapter it is explained in detail how this work is based and done, explaining the achieved data, what is done with the data and what tests has been made.
- Chapter 4: **Results**. It is shown all the rests of the tests performed explaining widely why these results are obtained.
- Chapter 5: **Conclusions**. Finally, it is detailed the conclusions obtained in this work and the future works that could follow this research.

## 2. Visión general

En la actualidad, la forma más extendida y fiable para detectar el tipo de ataque cerebrovascular que ha sufrido el paciente (isquémico o hemorrágico), además de para saber la localización del ataque, se basa en realizar un TAC (Tomografía Axial Computarizada) en la que es el médico el que detecta el suceso analizando la imagen resultante. En cuanto a otros métodos que no dependan de imágenes cerebrales e involucren únicamente la monitorización de valores fisiológicos del paciente no existe ningún sistema de predicción de resultados de ataques cerebrovasculares fiable.

En [2], *Zhang et al.* realizaron análisis con algoritmos de regresión logística sobre una gran base de datos de pacientes que habían sufrido ataques cerebrovasculares. En él, tenían en cuenta los datos de las primeras 48 horas del paciente y 3 meses después del incidente se tenían en cuenta los resultados (pérdida de ciertas capacidades principalmente) del ataque cerebrovascular para realizar una correlación entre los valores fisiológicos del paciente en sus primeras 48 horas y el estado del paciente 3 meses después. Los resultados de predicción que obtuvieron se hallaron entre el 71 % de acierto en las secuelas basándose únicamente en variables estadísticas y alrededor del 90% si utilizaban más fuentes de datos alternativas.

En el estudio de *Zhang et al.* los métodos de medición de datos de los pacientes no es tan exhaustivo como la monitorización realizada en este trabajo. Ésta se ha realizado en la Unidad de Ictus del Hospital Universitario de la Princesa, donde se monitoriza a los pacientes desde el momento en que son hospitalizados y se extrae información de sus constantes vitales cada 30 segundos hasta que finalmente son dados de alta o no se recuperan, lo que aporta una gran cantidad de información sobre el estado de los pacientes a lo largo de todo el proceso de tratamiento y recuperación del ataque cerebrovascular. La toma de datos comenzó el abril de 2017 y continúa hasta la fecha, lo que ha permitido tener gran cantidad de pacientes y grandes cantidades de datos sobre los mismos, pudiéndose filtrar los datos basándonos en varias ventanas de tiempo y no sólo limitándose a las primeras 48 horas de hospitalización, como en el artículo anteriormente mencionado.

Este sistema de monitorización que se conoce internamente en el hospital como “*firewall*” fue instalado en febrero de 2017 junto con técnicos de Philips. Se trata de una central de monitorización de alta gama modelo Philips PIIC iX. Este dispositivo en términos de arquitectura y procesamiento es capaz de monitorizar desde 4 hasta 1024 pacientes, siendo posible monitorizar en tiempo real hasta 32 pacientes simultáneamente. Además, cuenta con los sistemas de procesamiento

de señales para realizar estimaciones de distintas variables monitorizadas del paciente como, por ejemplo, la estimación del ritmo cardíaco. Internamente, este dispositivo procesa mensajes HL7 (a continuación se puede ver un mensaje HL7 real) en los que se indican las variables leídas del paciente. Además de centralizar toda la información, procesarla, almacenarla internamente y permitir visualizarla tanto desde el panel del monitor como desde un portal web, este dispositivo permite el exportado de los datos, como es en este caso donde se encarga de insertar estructuradamente estos datos en una base de datos relacional MySQL 5.7.

```
MSH|^~\&|||||ORU^R01|HP0426104014759ICTUS189133|P|2.3|||||8859/1
PID|||123562^^^MR|||19640118|M
PV1||I|Mi unidad^^Cama 4&0&0
OBR|||||20170426103956
OBX||ST|0402-f8f5^sMode^MDIL|0|MONITORING|||||F||SETTING
OBX||NM|0002-4261^EV^MDIL|0|10|0004-0aa0^lpm^MDIL|<10||||F
OBX||NM|0002-0302^ST-II^MDIL|0|1.7|0004-0512^mm^MDIL|-2.0-2.0||||F
OBX||NM|0002-500a^FR^MDIL|0|17|0004-0ae0^rpm^MDIL|8-30||||F
OBX||NM|0002-4bb8^SpO2^MDIL|0|94|0004-0220^%MDIL|90-100||||F
OBX||NM|0002-4822^Pulso (SpO2)^MDIL|0|78|0004-0aa0^lpm^MDIL|50-120||||F
OBX||NM|0002-4bb0^Perf^MDIL|0|0.21|0004-0200^MDIL||||F
OBX||NM|0002-4182^FC^MDIL|0|83|0004-0aa0^lpm^MDIL|50-120||||F
OBX||ST|0002-d007^EstRit^MDIL|0|Ritmo Sinusal|||||F
OBX||ST|0002-d006^EstEct^MDIL|0|EVs Multiform. 10|||||F
```

Figura 1. Ejemplo de mensaje HL7

En la Figura 1 se puede visualizar un ejemplo real de un mensaje HL7 con los que funciona el dispositivo de monitorización centralizada de los pacientes. Estos ficheros tienen una estructura arbórea, siendo posible visualizar toda la información como si de un XML se tratara.

A continuación se explica qué significa cada campo del mensaje HL7:

- MSH: indica aspectos del resto del mensaje, como el parser a utilizar o la codificación.
- PID: Reporta información del paciente. Fecha de nacimiento y número de historial clínico.
- PV1: indica de dónde proviene el mensaje.
- OBR: Indica la fecha del mensaje con hora, minuto y segundo.
- OBX: Indica datos fisiológicos del paciente. Se almacenan todos en la base de datos. En cada línea se especifica un identificador, el valor que mide, lo que mide, en qué valor lo mide y un rango o un valor normal para ese parámetro.
  - sMode: MONITORING. Indica que se está haciendo la monitorización
  - EV: extrasístoles ventriculares, medido en *lpm*(latidos por minuto). Según HL7 debe ser < 10
  - ST-II: Medido en mm(milímetros). Systolic Time Interval Index, elevación del intervalo ST del ECG que diagnostica infarto agudo de miocardio y otras alteraciones del ritmo.
  - FR (se mide en rpm): frecuencia respiratoria
  - PNIs, PNId, PNIm: presión sistólica, diastólica y media
  - SP-O2: Saturación de oxígeno. Porcentaje.
  - Pulso SpO2: frecuencia cardíaca medida a partir de la curva de fotopleletismografía. Medido en latidos por minuto ( LPM).
  - Perf: distintos valores de perfusión.

- FC: frecuencia Cardíaca, medido en latidos por minuto.
- EstRit: un indicador del ritmo detectado.
- EstEct: indica los ritmos ectópicos, que son rachas intercaladas de un ritmo con un origen diferente en el corazón.

Gracias a la alta cantidad de datos extraídos en tiempo real por el dispositivo de monitorización es posible plantear la idea de identificar ciertos modelos de predicción en función de los datos fisiológicos obtenidos. Los modelos de predicción serán identificados aplicando distintas técnicas de aprendizaje automático (Machine Learning).

En cuanto a las distintas técnicas a aplicar, es importante conocer y entenderlas para determinar cuál utilizar en el contexto de este trabajo. Los tipos de aprendizaje máquina más utilizados son los siguientes:

- Aprendizaje no supervisado: se trata de un tipo de enfoque en Machine Learning en los que el proceso de aprendizaje por parte de los modelos se realiza automáticamente en base a los datos, es decir, no es necesario etiquetar los ejemplos de entrenamiento o pruebas en ningún momento y se le deja al algoritmo de aprendizaje no supervisado la tarea de encontrar patrones en los ejemplos que los distinga entre sí. Este tipo de modelos son muy útiles cuando se pretende descubrir patrones ocultos en los datos y no se sabe exactamente qué es lo que representan los datos.
- Aprendizaje por refuerzo: este tipo de aprendizaje trata de parecerse al aprendizaje que se realiza con los niños en el que se les da reportes (feedback) continuamente indicando si lo que han hecho está bien o mal, llegando a recompensar o penalizar en función de lo bien o mal que ha realizado la acción. Este tipo de modelos se utilizan principalmente en robótica, donde el robot realiza una serie de tareas y va aprendiendo de ellas en base a las recompensas e indicaciones que le dan los “educadores” del robot.
- Aprendizaje supervisado: en este tipo de enfoque se parte de ejemplos de entrenamiento que están “clasificados” por medio de etiquetas o valores y se le encomienda al algoritmo la tarea de encontrar patrones en las características de entrenamiento que determinen el valor o la etiqueta de cada ejemplo. Dentro de este tipo de modelos, existen otros tipos de clasificación:
  - Regresión: los ejemplos están etiquetados con un valor numérico y el algoritmo debe deducir (por regresión) el valor en función de las características de los ejemplos.
  - Clasificación: se le asigna a cada ejemplo una etiqueta, pudiendo haber un número de etiquetas mínimo de 2, lo que es una clasificación binaria, y un número máximo

de  $n$  etiquetas, lo que se conoce como una clasificación multiclase (o también conocido como *clasificación n-aria*).

En este trabajo las predicciones que se pretenden realizar son:

- Predicción del tipo de ataque cerebrovascular: se pretende determinar qué tipo de ataque cerebrovascular ha sufrido el paciente en base únicamente a los datos hemodinámicos monitorizados.
- Predicción del desenlace del paciente: se pretende predecir de forma temprana si el paciente va a fallecer a raíz del ataque cerebrovascular o si por el contrario sobrevivirá en función de los valores hemodinámicos monitorizados del paciente.
- Predicción de resangrado: se pretende predecir si el paciente va a sufrir un episodio de sangrado nuevo en el que, por lo general, se tiende a empeorar la situación del paciente.

Estas tres predicciones se pueden abordar en un contexto de clasificación binaria, por lo que el tipo de aprendizaje automático que se utiliza en este trabajo es aprendizaje supervisado. Dentro del aprendizaje supervisado se encuentra una gran cantidad de algoritmos que tratan los datos de entrada de distintas formas. Algunos ejemplos de estos algoritmos de aprendizaje supervisado que se van a utilizar son los que se explican a continuación.

## Regresión Logística

Es un algoritmo que generaliza la regresión lineal como una clasificación binaria realizando dos cambios: utiliza la distribución de Bernoulli en vez de la distribución Gaussian y después realiza una combinación lineal de las variables en un rango de 0 a 1 aplicando también la función sigmoide [3].

La ecuación lineal a ajustar (regresión lineal) es:

$$y(X) = W^T X + \varepsilon = \sum_{j=1}^D w_j x_j + \varepsilon$$

donde  $W^T X$  representa el producto escalar entre un vector de entrada  $X$  y el vector de pesos  $W$  e  $y \in \{0, 1\}$ .

La distribución de Bernoulli tiene la forma:

$$p(y|x, w) = \text{Ber}(y|\mu(x))$$

El resultado queda en el intervalo  $0 \leq \mu(x) \leq 1$  donde  $\mu(x) = \text{sigm}(w^T x)$  y  $\text{sigm}(\eta) \approx \frac{1}{1+\exp(-\eta)} = \frac{e^\eta}{e^\eta+1}$ .

De esta forma, en una clasificación binaria en la que hay únicamente dos clases  $\{0, 1\}$ , el resultado en la regresión logística será un valor continuo entre 0 y 1. Si el valor en este caso está por debajo de 0.5, la clase predicha es la clase 0 mientras que si está por encima será la clase 1. Esto es porque la Regresión Logística es un regresor, es decir, predice un valor en vez de una clase en función de las variables utilizadas.

Además, a la Regresión Logística es muy habitual aplicar factores correctores como la regularización o factores de aprendizaje para ajustar el modelo lo máximo posible.

## Support Vector Machines

Son un tipo de algoritmos para regresión o clasificación. Tratan los datos como vectores y construye un conjunto de hiperplanos de N dimensiones, donde si hay diferencias entre las clases a diferenciar obtendrá buenos resultados. Son muy costosos de entrenar, por lo que en muchos casos obtener buenos resultados no solo depende de la calidad de los datos sino también del software que se utiliza para entrenar los modelos.

La función de coste de las SVM es la siguiente:

$$J(w, \lambda) = \sum_{i=1}^N L(y_i, \hat{y}_i) + \lambda \|w\|^2$$

donde  $\hat{y}_i = w^T x_i + w_0$  y L, en el caso de las SVM para clasificación (*Support Vector Classifier, SVC*) es del tipo *Hinge*, que se calcula de la siguiente manera

$$L_{\text{hinge}}(y, \eta) = \max(0, 1 - y\eta) = (1 - \eta)_+$$

## Naive Bayes

Es un algoritmo de clasificación probabilístico que se basa en la probabilidad condicional y el teorema de Bayes.

La fórmula de la probabilidad condicional es la siguiente:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\left( \frac{\text{\#casos favorables } A \cap B}{\text{\#casos posibles}} \right)}{\left( \frac{\text{\#casos favorables } B}{\text{\#casos posibles}} \right)}$$

Desarrollando mínimamente la igualdad anterior se llega al teorema de Bayes:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Con este teorema la idea es simple: condicionar las salidas (clases posibles) en función de los atributos. En la fase de entrenamiento lo que se calculan son las probabilidades entre conjuntos de características y clases posibles:

$$p(x|y = c, \theta) = \prod_{j=1}^D p(x_j | y = c, \theta_{jc})$$

Finalmente, en la fase de pruebas en las que se predice la salida del ejemplo lo que se hace es calcular la probabilidad en base a lo que ya se conoce previamente en el entrenamiento, determinando la salida al porcentaje más alto de salida.

## K - Nearest Neighbors

Este algoritmo está basado en la clusterización por distancia y principalmente compara los ejemplos dados calculando la distancia entre ellos. Esta distancia es la métrica, que generalmente suele ser la distancia Euclídea (también se utiliza comúnmente la distancia Manhattan o la distancia Minkowski), que resulta fundamental porque va a ser la única forma de distanciar los ejemplos de

clases distintas de forma correcta o ingenua. Resulta especialmente útil cuando los ejemplos a clasificar no son linealmente separables y se pueden separar ejemplos de distintas clases aun estando mezclados.

## Árboles de decisión

Este tipo de algoritmos (CART, Classification and Regression Trees) se crean recursivamente particionando la entrada (los ejemplos de entrenamiento) en distintas ramas. Se presenta habitualmente como un árbol, aunque existen modelos más complejos y costosos que se componen de un conjunto de árboles (Random Forests). La idea del algoritmo consiste en ir creando un camino a lo largo del árbol desde la raíz hasta las hojas (clases posibles ó valores de regresión dependiendo del tipo de problema) donde en cada nodo se plantea una condición, de forma que si se cumple el camino de decisión sigue por una rama colgante del nodo o si por el contrario no se cumple se continua por otra rama.

La creación del modelo se puede representar de la siguiente forma:

$$f(x) = E[y|x] = \sum_{m=1}^M w_m I(x \in R_m) = \sum_{m=1}^M w_m \Phi(x; v_m)$$

donde  $R_m$  representa la región  $m$ ,  $W_m$  es respuesta media en esa región y  $V_m$  codifica la elección de la variable por la que dividir y el valor límite de la división.

## Redes Neuronales Artificiales

Las redes neuronales artificiales (ANN) son un tipo de estructura de computación que intenta emular la forma en que se conectan e interactúan las redes neuronales biológicas. Dentro de las redes neuronales, hay varios tipos: red neuronal multi-layer perceptron (MLP), redes neuronales convolucionales (CNN), redes neuronales con memoria a largo/corto plazo (LSTM) o redes neuronales recurrentes (RNN), entre otras. En este estudio el tipo de red neuronal escogido debido a las características del problema se trata de la Multi-Layer Perceptron Network, que sirve tanto para regresión como para clasificación. Este tipo de red neuronal se caracteriza por ser tener una serie de capas lógicas apiladas una detrás de otra, donde la función que emula la neurona es un perceptrón.

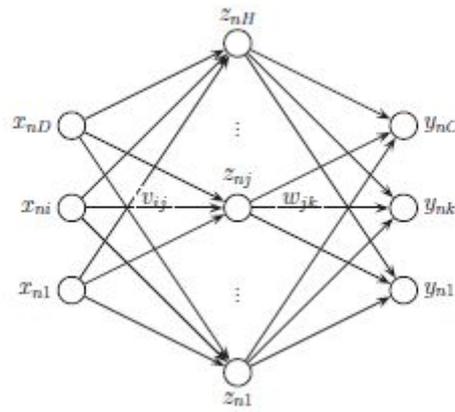


Figura 2. Red neuronal MLP con una capa oculta. [3]

En la Figura 2 se puede observar un modelo típico de MLP, donde en la primera capa de una MLP es la capa de entrada (input layer) y esta siempre está formada por N neuronas, siendo N el número de características que componen los ejemplos del problema a resolver. Después de la capa de entrada se suceden H capas ocultas (hidden layers) con una cantidad variable de neuronas en cada una de ellas. Se recomienda que el número de neuronas en cada capa intermedia sea el mismo aunque no es obligatorio que sea así. Además, no existe un número correcto o exacto de neuronas en la capa oculta ya que depende del problema, pero sí que se recomienda que sea mayor al número de unidades en la capa de entrada. Finalmente, se encuentra la capa de salida (output layer) que en el caso de un problema de tipo regresión hay una única unidad y devuelve el valor predicho mientras que si es un problema de clasificación se compondrá por el mismo número de clases a clasificar que hay. La codificación de la salida es de tipo one-hot.

Cada neurona se implementa como una función, la cual en función de los pesos que recibe toma un valor en un rango de 0 a 1 normalmente. Las funciones más conocidas y utilizadas ampliamente a lo largo de todos los tipos de redes neuronales artificiales son la función sigmoide y la función de activación ReLU (Rectified Linear Unit).

La función sigmoide es como sigue:

$$\text{sigm}(\eta) \approx \frac{1}{1+\exp(-\eta)} = \frac{e^\eta}{e^\eta+1}$$

Mientras que generalmente, ReLU ( Rectified Linear Unit) toma la siguiente forma:

$$f(x) = x^+ = \max(0, x)$$

donde  $x$  es la entrada de la neurona.

La principal ventaja de la ReLU respecto a la función sigmoide se produce en redes neuronales convolucionales y recurrentes (*Deep Learning*) donde genera mayor dispersión en la red y menor probabilidad de desvanecimiento del gradiente (*gradient vanishing*). Esto se debe a que, por lo general, la función sigmoide tiende a generar valores mayores a cero al contrario que la función ReLU, lo que genera una mayor densidad de cómputo entre capas y menos dispersión. Generalmente resultan más beneficiosos los modelos donde hay dispersión en vez de capas densas, por lo que puede ser más beneficioso utilizar unidades ReLU en caso de redes neuronales complejas como las recurrentes o las convolucionales.

Las redes neuronales son muy útiles en general, pero donde ofrecen resultados realmente excepcionales es en visión por computador (reconocimiento de imágenes) y reconocimiento del lenguaje (Natural Language Processing, NLP).

Todos estos modelos de aprendizaje automático se han usado para realizar las predicciones que más adelante se muestran. Cada tipo de algoritmo de predicción proporciona mejores o peores resultados en función del tipo de problema a resolver y del tipo de datos sobre los que se aplica el modelo.

Es por esto mismo que uno de los objetivos planteados en este proyecto es establecer una jerarquía de clasificadores en función de su efectividad y capacidad de predicción para los problemas planteados.

# 3. Metodología

La metodología llevada a cabo se compone de una serie de procedimientos variados y estructurados.

## Copia de respaldo y formateo de datos

Cada semana se descargaba un *backup* de la base de datos MySQL en la que se almacenaban todos los datos de los pacientes, incluyendo identificadores de pacientes y constantes medidas junto con los valores en cada instante de tiempo. Esta información había que extraerla, transformarla y almacenarla como ficheros de valores separados por comas (CSV's, Comma Separated Values) con procesos automatizados para que luego pueda ser utilizada en los modelos de predicción. La información de las variables recogidas de los pacientes se almacenaba en una única tabla de forma que cada fila de la tabla representa una variable medida con su valor en un instante de tiempo de un paciente concreto. Para realizar este trabajo con MySQL, se utilizó como apoyo un libro sobre MySQL [7] y como lenguaje de programación para el procesado de los datos Python 2.7. La transformación de los datos utilizando de Python es una parte clave en este proyecto, por lo que se utilizó como referencia el libro Python for Data Analysis [8], donde vienen expuestas y explicadas de forma tanto teórica como práctica varias metodologías relacionadas con la transformación de datos en Python.

## Análisis de características

Una vez los datos habían sido transformados y estructurados en un fichero de tipo CSV se realizó un análisis de las variables que se recogían y cuáles aportaban información relevante sobre el estado del paciente en relación al ataque cerebrovascular. Se detectó que algunas variables no se monitorizaban y almacenaban casi nunca, por lo que se decidió eliminarlas del conjunto de datos. Finalmente, después de descartar variables por la falta de relevancia o por la cantidad de ausencia en los reportes de los pacientes se escogió el siguiente conjunto de variables de los pacientes:

- Estimación del ritmo.
- EV.
- FC.
- Frecuencia Respiratoria.
- Perf.

- SpO2.
- ST-II.

## Preprocesamiento de datos y normalización

Con el conjunto de datos analizado y seleccionado, el siguiente paso consistía en realizar el preprocesamiento de los datos en busca de valores nulos o espurios (valores erróneos). Puesto que los datos son de pacientes concretos y en series temporales, se ha de tener en cuenta a la hora de preprocesar los datos. Cuando se daban casos de valores nulos o espurios puntuales por un fallo en la monitorización se realizaba un procedimiento de reemplazo, inserción o eliminación de la instancia monitorizada en función del tipo de variable. En el caso de los valores de la estimación del ritmo el procedimiento en caso de un valor nulo consistió en replicar el valor anterior medido correctamente en el tiempo a ese valor nulo de ese mismo paciente. En cambio, en variables numéricas como la frecuencia cardíaca o la perfusión se reemplazaron los datos nulos observando el conjunto total de la monitorización del paciente sustituyendo con la media en una ventana de tiempo determinada. En cuanto a la detección y reemplazo de espurios se realizaba únicamente en variables numéricas. Se consideró un valor como espurio cuando el valor medido era significativamente lejano de los valores típicos que se estaban leyendo en las monitorizaciones previas y posteriores.

Cuando se completaba el proceso explicado hasta aquí, los datos ya se consideraban prácticamente aptos y preparados para utilizarlos en modelos de aprendizaje automático. El único paso que quedaba de preprocesamiento consistía en normalizar los datos de forma que los valores se encuentren en un intervalo concreto. Para esta normalización se han utilizado dos técnicas de normalización ampliamente utilizadas:

- Normalización por Media y Desviación Estándar: consiste en aplicar la siguiente fórmula a cada característica del conjunto de datos:

$$\frac{X-\mu}{\sigma}$$

donde  $X$  es el dato actual de una característica concreta,  $\mu$  es la media de esta característica y  $\sigma$  es la desviación estándar de esa característica teniendo en cuenta todo el conjunto de datos.

- Normalización por Máximos y Mínimos: consiste en aplicar la siguiente fórmula a cada característica del conjunto de datos:

$$\frac{X - X_{min}}{X_{max} - X_{min}}$$

donde  $X$  es el dato actual,  $X_{max}$  es el máximo valor de la característica y  $X_{min}$  es el valor mínimo de la característica. Este tipo de normalización es conocida como *Scaling*, puesto que escala los valores a un rango definido  $[0,1]$ .

La normalización por máximos y mínimos tiende a obtener peores resultados en comparación con la normalización con respecto a la media y la desviación estándar puesto que la normalización por máximos y mínimos transforma los datos a un rango entre 0 y 1, no siendo representado un caso de alta desviación del dato, como sí se puede ver reflejado en la normalización por desviación estándar donde los datos no están en un rango estrictamente definido y es posible observar un valor anómalo o lejano a los valores “normales” cuando es un valor lejano a 0.

El tipo de normalización a utilizar depende principalmente del modelo algorítmico a utilizar. Por ejemplo, la normalización de los datos con respecto a la media y a la desviación estándar produce valores reales tanto positivos como negativos, lo que no es apto para modelos como Naive Bayes, que funciona únicamente con valores positivos. Sin embargo, el resto de modelos funcionan tanto con valores positivos como negativos, lo que permite elegir entre ambas normalizaciones y probar cuál funciona mejor con este conjunto de datos.

## Entrenamiento y validación

En todas las pruebas que se exponen en el apartado de Resultados se ha partido del mismo conjunto de datos inicial y dependiendo del tipo de experimento y finalidad se han preprocesado de distinta manera añadiendo nuevos parámetros en ciertos casos al conjunto de características de entrenamiento.

Puesto que los datos de los pacientes son series temporales, es importante tener en cuenta esa temporalidad y el orden en el que se recopiló cada medición de constantes del paciente. Para intentar descubrir cómo influye la evolución de los pacientes en este tipo de clasificaciones se han creado instancias de clasificación con varias series temporales de datos: desde instancias con 1 serie temporal (una medición de un paciente) hasta instancias con 2, 3, 4 y 5 series temporales (es decir,

instancias compuestas por N mediciones consecutivas y con los mismos atributos medidos). Esta idea se puede ver en la Figura 3, donde se parte de cuatro mediciones como ejemplo de un paciente que ha sufrido un ataque cerebrovascular de tipo hemorrágico y tiene esas mediciones en los primeros 2 minutos de monitorización. Dependiendo de si se utiliza una o dos series temporales, es decir, 1 o 2 observaciones por instancia de entrenamiento/testing, se puede ver cómo quedarían las instancias.

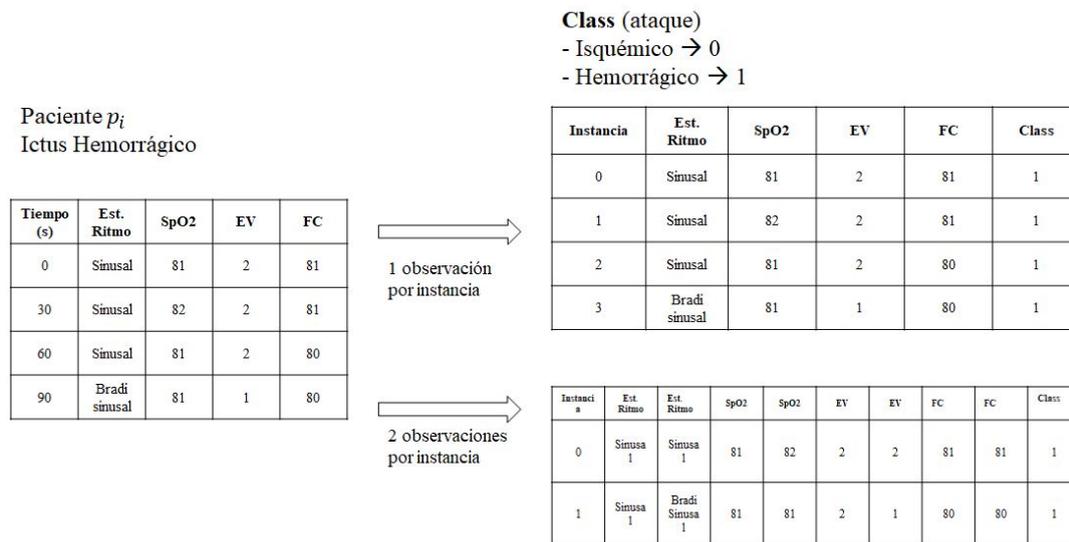


Figura 3. Esquema de transformación de datos con varias observaciones por instancia

De esta forma, se podrá observar si existe una variación en los resultados dependiendo del número de observaciones del paciente que se consideran para las clasificaciones y se podrá encontrar el conjunto mínimo de datos necesario que proporciona una predicción con una precisión aceptable. Es importante mencionar que, puesto que cada medición del paciente se realiza cada 30 segundos, aumentar el número de series temporales incluidas en los vectores de entrenamiento implica una monitorización durante más tiempo del paciente para realizar pruebas. Además, puede ser perjudicial al reducir el número de instancias para entrenamiento y validación, lo que a su vez posiblemente influya en una disminución de la tasa de aciertos debido a un *underfitting*.

Además de variar el número de observaciones por instancia, también se realiza un filtrado de datos por ventanas de tiempo. Esto es útil para saber si los datos son más determinantes al poco de sufrir el ataque cerebrovascular o si por el contrario son más ruidosos nada más ingresar al estar más reciente el suceso. Por tanto, se realizan clasificaciones utilizando instancias de la primera hora, de las primeras dos horas, de todos los datos omitiendo la primera hora, utilizando solo los datos de la

segunda hora, etc ... Es importante tener en cuenta que al realizar estos filtrados lo que se obtiene de forma resultante es también un conjunto de datos de menor tamaño, ya que se está reduciendo el conjunto a un rango de tiempo limitado, lo que puede implicar también un *underfitting*.

El conjunto de algoritmos utilizados ha sido el mismo para todos los experimentos aunque, como es previsible, algunos algoritmos han funcionado mejor en algunos experimentos y peor en otros. En cualquier caso, siempre se ha abordado la predicción como un problema de clasificación binaria. En las tablas que se muestran en el apartado de resultados se identifican los siguientes algoritmos con unas siglas que se muestran a continuación:

- Regresión Logística: LR
- Support Vector Machines (Support Vector Classifier): SVC
- K - Nearest neighbors: KNN
- Artificial Neural Network, MultiLayer Perceptron(MLP): NN
- Decision Tree: DT
- AdaBoost: AB
- Random Forest: RF
- Naive Bayes: NB
- Gradient Boosting: GB

La librería que se ha utilizado para la creación de modelos predictivos es Scikit-Learn, que es una librería para Python ampliamente utilizada en el ámbito de *machine learning* donde vienen implementados todos estos algoritmos con la posibilidad de configurar de la forma más óptima posible cada algoritmo.

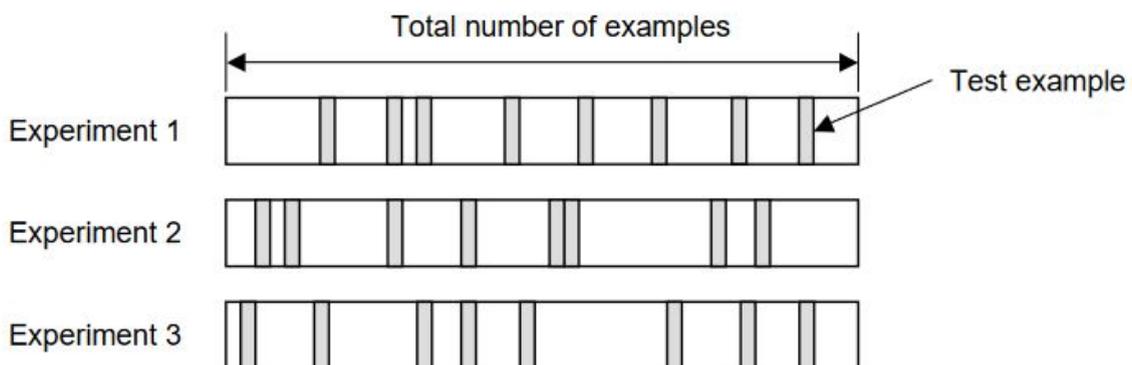


Figura 4. Random Cross-Validation [4].

El método de validación de resultados elegido ha sido *cross-validation* en su variante aleatoria, que consiste en entrenar el algoritmo con un porcentaje P del total de los datos existentes y utilizar el porcentaje 1-P restante para el testing de los algoritmos entrenados. Las distintas variantes de la *cross-validation* se diferencian en cómo se escogen los ejemplos de testing, siendo utilizada en este trabajo la variante aleatoria, tal y como se puede ver en la Figura 4, que escoge los ejemplos de *testing* de forma aleatoria en cada experimento.

En todos los experimentos se ha utilizado un porcentaje de entrenamiento del 70% y se ha utilizado el 30% restante de los ejemplos para comprobar si el modelo entrenado clasifica correctamente las instancias. Al utilizar números generados aleatoriamente en cada ejecución el conjunto de datos de entrenamiento y testing varía y en consecuencia los resultados también. Para evitar esto, se utiliza una semilla de números aleatorios de forma que sea un método pseudo-aleatorio y siempre se escoja el mismo conjunto de datos de *training* y *testing* para que todos los modelos se entrenen y validen en las mismas condiciones y con esta metodología se puedan comparar en igualdad de condiciones todos los modelos.

## 4. Resultados

En los siguientes apartados se exponen los resultados en cada tipo de predicción realizada y se explican con detalle el sentido y significado de los resultados obtenidos. Para todos los resultados que se presentan los datos vienen en porcentajes y la métrica utilizada para valorar los algoritmos es el F1 Score, que es un método de evaluación estadístico muy utilizado y fiable en clasificación binaria. Esta métrica es la media armónica de la precisión y el *recall* (también conocido como exhaustividad o sensibilidad).

En clasificaciones binarias, la precisión es la fracción de los ejemplos correctamente clasificados de una clase entre el total de clasificaciones predichas en esa misma clase, incluyendo los ejemplos predichos como esa clase de forma errónea. El *recall* es la fracción de ejemplos correctamente clasificados de una clase entre los ejemplos totales de esa clase.

$$F1_{score} = 2 \cdot \frac{1}{\frac{1}{recall} + \frac{1}{precision}} = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

### 4.1. Predicción del tipo de ataque cerebrovascular

En este tipo de experimentos lo que se pretende es predecir, con base en los datos medidos de los pacientes, si el tipo de ataque cerebrovascular que ha sufrido el paciente es de tipo isquémico o hemorrágico. Como se ha explicado anteriormente, se muestran a continuación distintas tablas en las que se considera como un ejemplo de entrenamiento o validación una o más observaciones sobre el paciente.

En la Tabla 1 se muestra el porcentaje de aciertos en función del algoritmo y la ventana de tiempo utilizando una observación para realizar la predicción. Una observación se trata de los valores medidos del paciente durante una monitorización de 30 segundos. Los resultados más altos se obtienen en las ventanas de tiempo más grandes de forma general, lo que parece indicar que no hay indicadores en algunos tramos del ingreso.

Ventana de tiempo	LR	SVC	KNN	NN	DT	AB	RF	NB	GB
Ventana 0-60 minutos	53,62	75,35	85,01	68,61	82,08	76,87	88,69	60,23	87,93
Ventana 0-120 minutos	56,47	84,07	84,74	75,32	86,6	81,75	89,91	60,12	88,06
Ventana 0-180 minutos	56,67	85,59	86,08	76,56	87,07	81,95	90,66	62,42	88,95
Ventana 0-240 minutos	55,09	66,59	86,48	78,56	88,58	82	92,53	60,01	90,78
Ventana 0-total minutos	57,13	57,73	88,14	88,03	90,61	85,49	92,96	60,64	91,86
Ventana 60-120 minutos	58,53	80,17	86,55	72,12	87,38	83,63	90,43	60,06	88,21
Ventana 60-180 minutos	56,84	88,06	87,58	78,07	89,59	85,08	93,2	60,44	93,13
Ventana 60-240 minutos	56,56	87,74	88,87	80	89,86	85	93,58	63,21	91,93
Ventana 60-final minutos	56,64	50,74	88,22	87,29	91,02	85,14	92,99	60,61	91,84

Tabla 1. Resultados tipo de predicción de ataque cerebrovascular con 1 observación

Ventana de tiempo	LR	SVC	KNN	NN	DT	AB	RF	NB	GB
Ventana 0-60 minutos	56,09	55,84	85,03	70,3	84,77	77,41	89,09	61,93	86,55
Ventana 0-120 minutos	59,55	63,4	84,08	68,97	85,28	81,3	91,91	60,48	90,19
Ventana 0-180 minutos	54,85	77,74	84,29	74,6	85,55	82,32	92,37	60,14	89,77
Ventana 0-240 minutos	57,19	84,38	87,13	73,92	87,89	83,07	93,12	62,22	91,81
Ventana 0-total minutos	56,93	57,63	87,01	86,82	90,51	86,26	94,25	61,02	92,38
Ventana 60-120 minutos	60,66	56,79	85,04	73,41	87,81	82,55	93,91	60,94	91,41
Ventana 60-180 minutos	58,81	67,55	85,58	72,26	89,04	84,33	93,9	60,33	92,93
Ventana 60-240 minutos	57,64	78,96	87,45	73,68	89,53	84,81	94,91	61,13	92,45
Ventana 60-final minutos	56,68	61,54	87,61	87,39	90,96	86,84	94,37	60,82	92,14

Tabla 2. Resultados tipo de predicción de ataque cerebrovascular con 2 observaciones

Ventana de tiempo	LR	SVC	KNN	NN	DT	AB	RF	NB	GB
Ventana 0-60 minutos	52,67	49,62	86,64	64,5	82,82	83,21	91,6	64,12	89,31
Ventana 0-120 minutos	53,19	58,57	86,25	71,91	87,05	78,09	90,24	59,56	87,65
Ventana 0-180 minutos	58,22	62,53	84,5	70,08	85,71	83,02	93,94	59,7	90,03
Ventana 0-240 minutos	57,85	65,6	84,71	74,28	86,67	84,19	93,8	60,95	92,15
Ventana 0-total minutos	56,78	56,35	85,73	85,69	89,91	85,94	93,69	60,87	91,84
Ventana 60-120 minutos	61,41	51,04	85,89	76,76	80,5	79,25	89,63	59,34	87,14
Ventana 60-180 minutos	60,5	54,05	84,62	75,05	87,94	83,16	92,93	61,95	91,68
Ventana 60-240 minutos	56,58	49,65	83,45	68,74	86,99	85,86	93,35	61,39	91,23
Ventana 60-final minutos	58,01	56,08	87,01	87,16	90,86	87,38	94,9	61,97	92,82

Tabla 3. Resultados tipo de predicción de ataque cerebrovascular con 3 observaciones

Ventana de tiempo	LR	SVC	KNN	NN	DT	AB	RF	NB	GB
Ventana 0-60 minutos	52,28	44,67	84,26	62,94	81,73	79,19	90,36	64,97	89,85
Ventana 0-120 minutos	53,05	48,01	84,88	68,7	83,82	76,66	89,66	60,74	87,27
Ventana 0-180 minutos	58,71	55,12	83,3	72,35	84,38	79,35	91,74	61,4	88,51
Ventana 0-240 minutos	55,85	47,46	83,49	73,59	84,46	81,16	91,61	60,11	88,58
Ventana 0-total minutos	57,8	66,68	85,54	85,59	89,77	87,11	94,15	60,28	91,99
Ventana 60-120 minutos	62,98	46,96	85,64	72,93	83,98	80,11	90,61	63,54	88,95
Ventana 60-180 minutos	54,02	51,8	83,66	67,04	84,21	78,95	93,35	57,62	88,92
Ventana 60-240 minutos	57,92	50,38	85,09	70,94	84,34	80,75	93,02	61,13	88,3
Ventana 60-final minutos	58,39	58,95	85,84	85,83	89,67	86,86	94,31	61,48	92,85

Tabla 4. Resultados tipo de predicción de ataque cerebrovascular con 4 observaciones

Ventana de tiempo	LR	SVC	KNN	NN	DT	AB	RF	NB	GB
Ventana 0-60 minutos	46,84	44,94	79,75	56,33	72,78	73,42	84,81	67,72	80,38
Ventana 0-120 minutos	49,01	47,68	81,46	67,22	75,83	74,83	88,41	59,6	85,43
Ventana 0-180 minutos	53,36	51,57	84,3	71,97	81,17	82,51	91,03	60,31	87,89
Ventana 0-240 minutos	55,42	48,54	84,51	71,77	83,65	80,21	91,91	58,52	88,98
Ventana 0-total minutos	57,75	59,52	85,72	84,68	88,34	86,39	93,32	62,29	91,75
Ventana 60-120 minutos	60,69	45,52	78,62	65,52	84,83	77,93	91,03	61,38	88,28
Ventana 60-180 minutos	50,87	45,33	88,24	70,24	83,39	79,93	94,12	63,67	92,39
Ventana 60-240 minutos	58,73	47,64	86,79	74,06	83,73	82,31	94,58	62,26	90,33
Ventana 60-final minutos	56,67	59,78	86,29	85,63	89,5	87,19	94,3	61,33	92,28

Tabla 5. Resultados tipo de predicción de ataque cerebrovascular con 5 observaciones

En los resultados que se muestran en las Tablas 2, 3, 4 y 5 se pueden ver valores para distintos algoritmos y en varias ventanas de tiempo. En cada tabla se muestran los experimentos utilizando 1 o más observaciones para realizar las predicciones. En la Tabla 1 se realizan experimentos con 1 observación por instancia de entrenamiento/testing, mientras que en la Tabla 2 se muestran con 2 observaciones por instancia, y así sucesivamente. El objetivo de añadir más observaciones de los pacientes por instancia de entrenamiento radica en ver si al añadir información de las variables en un margen de tiempo mayor se aportan más indicadores de la situación real del paciente. Los valores de las tablas son el F1-Score que muestra la precisión que ha tenido el algoritmo con observaciones monitorizadas de los pacientes en la ventana de tiempo indicada. Los valores F1-Score representan realmente porcentaje de aciertos. Observando las diferencias de resultados entre las Tablas 1 y 2, se puede ver que la precisión no mejora notablemente en ningún algoritmo en prácticamente ninguna ventana de tiempo, lo que ocurre de igual forma con el resto de Tablas 3, 4 y 5. Además, en este tipo de predicción los resultados tienden a disminuir ligeramente, lo que se debe a incluir más información de observaciones por instancia reduciendo el conjunto de entrenamiento y validación.

Una vez explicado esto, se continúa con un análisis de los resultados obtenidos. La Figura 5 muestra la precisión de todos los algoritmos utilizados sin ventana de tiempo, es decir, utilizando

todos los datos de monitorización de cada paciente. Lo que se puede ver rápidamente en la Figura 5 es que no todos los algoritmos obtienen de forma general resultados similares.

### Comparativa de algoritmos

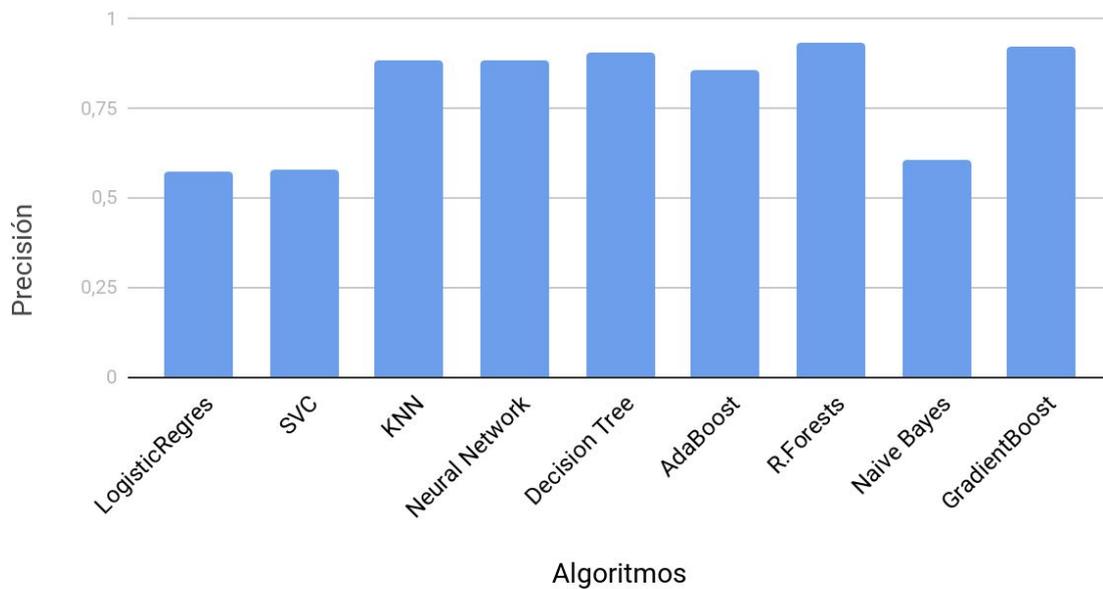


Figura 5. Comparativa de algoritmos predicción de ataque cerebrovascular

Tal como ilustra la Figura 5, algunos algoritmos de predicción obtienen resultados mucho peores en términos relativos a otros. Algoritmos como Regresión Logística ó Naive Bayes obtienen medias de porcentajes de aciertos entre el 50% y el 60%, lo que son porcentajes muy pobres y distantes con respecto a otros resultados como los picos de 93% de aciertos de Gradient Boosting y Random Forests. Entre estos dos grupos de algoritmos encontramos a SVM (SVC) y Redes Neuronales que obtienen resultados oscilantes entre el 50% y el 88% de aciertos. Por último se encuentran Nearest neighbors, AdaBoost y Decision Trees que se caracterizan por obtener resultados generalmente muy buenos en torno al 90% de aciertos, no obtienen en ningún momento resultados realmente cercanos al mejor porcentaje de aciertos, que se obtiene con el modelo de Random Forests.

Lo que se puede extraer de este pequeño análisis de la Figura 5 es que este tipo de problemas de clasificación con características en el dominio real es más adecuado modelarlos con ciertos algoritmos de aprendizaje automático. Algunos como la Regresión Logística o el clasificador Naive Bayes basado en estadística no encajan y no llegan a encontrar un patrón de distinción en las características para distinguir entre un ataque sufrido u otro. Por otra parte, se puede observar que algoritmos basados en similitud de ejemplos (Nearest neighbors, SVM) o algoritmos de árboles de decisión (Random Forests, Decision Trees, AdaBoost, Gradient Boosting) obtienen mejores resultados, consiguiendo encontrar un patrón de distinción en las características para decidir el tipo

de ictus sufrido.

Ahora bien, dentro de cada algoritmo se pueden observar distintos resultados en función de la ventana de tiempo. La Figura 6 muestra la precisión de los algoritmos que mejores resultados han obtenido en distintas ventanas de tiempo. Para elaborar la Figura 6 se han dejado de lado algoritmos como Regresión Logística o Naive Bayes que simplemente no parecen realizar una distinción real entre un tipo de ictus u otro se puede observar que hay una variación real en los resultados obtenidos para las distintas ventanas de tiempo.

### Mejores Algoritmos - Tiempo de monitorización

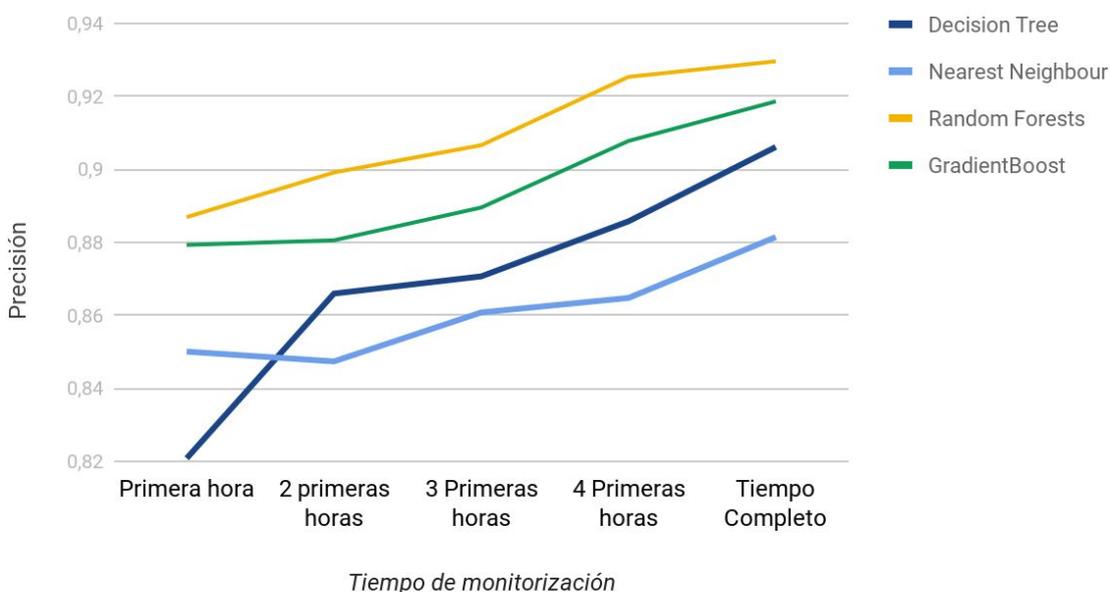


Figura 6. Comparativa de mejores algoritmos predicción de ataque cerebrovascular en función de la ventana de monitorización

Para empezar, en la Figura 6 se observa claramente que los porcentajes de acierto son ligeramente inferiores en la ventana de la primera hora de monitorización. Esto se repite de forma general cuando se comparan los resultados de las primeras dos horas de monitorización con las primeras tres horas de monitorización, lo que indica un posible caso de underfitting, ya que al reducir la ventana de tiempo a un periodo de tiempo corto (1,2,3... horas) el número de observaciones para entrenar y validar los algoritmos es menor. Este fenómeno se repite de forma general en las siguientes ventanas de tiempo, incluso cuando se ignoran las observaciones de la primera hora de monitorización. Generalmente se obtienen mejores resultados a medida que la ventana de tiempo se incrementa, lo que confirma que en este caso, con series de una observación del paciente, no hay un indicador claro de que las primeras horas de monitorización sean decisivas para determinar qué tipo de ictus, sino que a lo largo del tiempo se tiende a mantener un patrón habitual en las constantes

monitorizadas.

#### 4.1.1 Pruebas realistas con nuevos pacientes

Con el fin de descubrir si estos resultados obtenidos con el método de *cross-validation* eran realmente efectivos en casos reales se probó a predecir el tipo de ictus con observaciones de pacientes nuevos sobre los que no se tenía constancia inicialmente de qué tipo de ictus había sufrido, lo que suponía una prueba determinante para verificar la efectividad de los modelos.

La metodología llevada a cabo en estas pruebas es por lo tanto un poco distinta a la utilizada anteriormente y más cercana a lo que sería un caso realista en el caso de utilizar estos modelos algorítmicos de predicción como herramienta o método de detección del tipo de ictus de forma rápida y eficaz.

Para realizar estas pruebas han escogido los modelos entrenados con datos de pacientes ya existentes y analizados. Los algoritmos que mejores resultados habían obtenido en las pruebas expuestas en el apartado anterior son Random Forests, Gradient Boosting, Decision Trees y K-Nearest Neighbors. Para poder verificar si existía una ventana de tiempo en la que se decidiera con mayor acierto correctamente el tipo de ictus se utilizaron los modelos mencionados con dos versiones: una versión de los modelos entrenados con datos de pacientes con diagnóstico conocidos de la primera hora de monitorización tras el ingreso y la otra versión con todo el conjunto de observaciones de los pacientes con diagnóstico conocido desde el ingreso. De esta forma, se utilizarán modelos entrenados con pacientes con diagnósticos conocidos y se harán pruebas con pacientes nuevos de los que inicialmente se desconoce el diagnóstico y no formaron parte en ningún momento del entrenamiento de los modelos.

Un factor muy relevante a tener en cuenta cuando se crean estos modelos es la normalización de los datos. Todos los datos que se utilizan para entrenar los modelos están normalizados en base a esos mismos datos, es decir, en el caso de utilizar una normalización por la media y la desviación, la media y la desviación típica de cada característica de entrenamiento es de los datos de entrenamiento. Resulta imprescindible normalizar los datos de entrada de los pacientes nuevos para que las predicciones sobre los modelos entrenados funcionen, por lo que las normalizaciones se hicieron con respecto a los valores de los datos de entrenamiento. El criterio por el que se considera que se ha acertado en la predicción con un paciente consiste en si el 80% de las observaciones son predichas en la clase correcta, de forma que un paciente con un ataque cerebrovascular de tipo isquémico tendrá que salir como isquémico en al menos el 80% de las monitorizaciones.

Teniendo en cuenta todo esto, los primeros resultados generados se pueden observar en la Tabla 6

en la que se muestran los porcentajes de acierto de los mejores algoritmos diferenciando entre los resultados con modelos entrenados y testados con monitorizaciones de la primera hora o del tiempo completo que estuvo el paciente monitorizado. Como se puede ver claramente en la Tabla 6, los resultados son mejores limitando los datos a las monitorizaciones de la primera hora de monitorización. Los algoritmos que mejores resultados obtienen son los de tipo arbóreo, Random Forests, Gradient Boosting y Decision Trees.

	<b>Random Forests</b>	<b>Gradient Boost</b>	<b>Nearest Neighbors</b>	<b>Decision Trees</b>
<b>1ª hora de monitorización</b>	76,92	76,92	61,53	76,92
<b>Tiempo completo de monitorización</b>	76,92	69,23	53,84	69,23

Tabla 6. Comparativa mejores algoritmos con pacientes reales con ventana de tiempo y sin ventana

En el caso de las predicciones utilizando datos de la 1ª hora de monitorización, los porcentajes de acierto coinciden. Esto se debe a que coinciden en que han acertado y fallado en los mismos pacientes, lo que indica que, puesto que son tres tipos de modelos de clasificación similares, habiendo sido entrenados con los mismos datos y teniendo el mismo conjunto de datos de pruebas fallan en los mismos datos porque estos pacientes no parecen seguir los patrones de los pacientes utilizados para el entrenamiento de los modelos.

Además, no solo resulta relevante en este tipo de experimentos el porcentaje de aciertos sino que descubrir en qué fallaba también era clave en este estudio. Tras revisar los aciertos y los fallos en cada tipo de paciente se pudo concluir que el porcentaje de aciertos en ataques cerebrovasculares hemorrágicos estaba alrededor del 33% mientras que el porcentaje de aciertos en ataques cerebrovasculares isquémicos estaba en torno al 95%. Al tratarse de una distribución de pacientes descompensada en estas pruebas con aproximadamente el 80% de los pacientes con ataques de tipo isquémicos y 20% de tipo hemorrágicos, los resultados mezclando ambos tipos de ataques cercanos al 80% de aciertos no era totalmente fiable ya que en el caso de los hemorrágicos la predicción es realmente baja.

Estos resultados contrastan notablemente con los resultados obtenidos en el resto de las pruebas realizadas (explicadas en el apartado anterior) utilizando *cross-validation* ya que utilizando distribuciones de pacientes parejas (50-50 aproximadamente) se obtenían porcentajes de acierto prácticamente idénticos entre un tipo de paciente y otros, consiguiendo porcentajes superiores en

todos los casos al 90%.

Pese a los resultados negativos mostrados anteriormente aún quedaba la duda de cuánto tiempo de monitorización debería realizarse al paciente con el fin de predecir con cierta fiabilidad el tipo de ataque cerebrovascular.

Para comprobar las predicciones lo que se realizó es filtrar las monitorizaciones por ventana temporal de dos minutos y medio en dos minutos y medio, lo que equivale a cinco monitorizaciones por cada dos minutos y medio. Una vez filtrados los datos, lo que se realizó es una media de las predicciones. En el caso de un paciente con un ataque isquémico la predicción correcta sería clasificar cada instancia de monitorización como un cero, mientras que si es hemorrágico la predicción es uno. Al igual que en la predicción actual, se consideraría una predicción correcta si el 80% de las instancias predichas coinciden con la clase real de ataque cerebrovascular del paciente.

A continuación se pueden ver la Tabla 7 y la Tabla 8, que son dos reportes de resultados de dos pacientes de cada tipo de ataque cerebrovascular con monitorizaciones desde los primeros dos minutos y medio hasta los primeros veinte minutos.

	<b>Random Forests</b>	<b>Gradient Boosting</b>	<b>Decision Trees</b>	<b>Nearest Neighbors</b>
2.5 mins	100	100	100	60
5.0 mins	100	100	100	70
7.5 mins	100	100	100	80
10.0 mins	100	100	100	85
12.5 mins	100	100	100	88
15.0 mins	100	100	100	90
17.5 mins	100	100	100	91,43
20.0 mins	100	100	100	92,5

Tabla 7. Ataque cerebrovascular de tipo isquémico con pacientes reales y ventanas de minutos

Tal y como se puede ver en la Tabla 7 y la Tabla 8, que muestran los porcentajes de acierto de dos pacientes con resultados típicos en las pruebas realizadas, se puede observar cómo en el caso del paciente con ataque isquémico la precisión de los algoritmos arbóreos es muy alta (100% o muy cercana) mientras que las predicciones del clasificador K-Nearest Neighbors es ligeramente inferior

aunque tras siete minutos y medio de monitorización la predicción entra en los márgenes establecidos (80%) prediciendo correctamente la clase de ataque sufrido por el paciente. Estos resultados de pacientes con ataques isquémicos se produce prácticamente igual en todo el resto de pacientes de este tipo. Teniendo en cuenta esto, se puede establecer que los algoritmos de predicción de tipo arbóreo detectan con alta precisión a los pacientes con ataques de tipo isquémico.

	<b>Random Forests</b>	<b>Gradient Boosting</b>	<b>Decision Trees</b>	<b>Nearest Neighbors</b>
2.5 mins	60	60	60	80
5.0 mins	60	60	60	60
7.5 mins	40	40	40	40
10.0 mins	45	45	50	45
12.5 mins	56	56	60	56
15.0 mins	63,33	63,33	66,67	63,33
17.5 mins	68,57	68,57	71,42	68,57
20.0 mins	72,5	72,5	72,5	72,5

Tabla 8. Ataque cerebrovascular de tipo hemorrágico con pacientes reales y ventanas de minutos

Por otra parte, los resultados de este paciente con ataque de tipo hemorrágico son, como es de esperar tras observar los resultados generales con todos los pacientes, bastante bajos, especialmente al principio de la monitorización. Esto ocurre de forma especial con los algoritmos arbóreos, que no llegan en ningún momento a predecir correctamente el 80% de las instancias en ningún momento de los primeros veinte minutos de monitorización. En contraposición, K-Nearest Neighbor ha obtenido unos resultados al principio de la monitorización bastante superior al resto de clasificadores, mientras que según avanza el tiempo de monitorización tiende a igualarse con el resto. En el caso de los pacientes hemorrágicos sí que es importante mencionar que cada paciente reporta unos resultados distintos en cuanto a porcentajes de acierto, por lo que no se podría llevar a cabo una generalización en cuanto al tipo de resultados de este tipo de pacientes, a diferencia de los pacientes con ataque de tipo isquémico.

Para finalizar, es importante comentar que sí que se puede observar una tendencia general si se realiza un análisis detenido paciente a paciente. La predicción de los pacientes de tipo hemorrágico es inferior pero funciona mejor con algunos algoritmos como es el caso de K-Nearest Neighbor mientras que la predicción de pacientes de tipo isquémico se realiza con una precisión muy alta por

parte de algoritmos arbóreos. Como trabajo futuro se podría investigar la creación de un modelo mixto de dos niveles de predicción que esté compuesto por un predictor como Random Forests y otro como KNN y dependiendo de la salida de ambos, otro predictor como, por ejemplo, Regresión Logística o una máquina de soporte vectorial, calcule la salida final que sería la predicción real del paciente.

## 4.2. Predicción de pronóstico

En este apartado se muestran los resultados obtenidos en los experimentos realizados con el objetivo de realizar una predicción del pronóstico de los pacientes en base únicamente a las variables hemodinámicas. El pronóstico que se intenta predecir es simplemente si el paciente sobrevivirá a raíz del ataque cerebrovascular o fallecerá como consecuencia del mismo. Este tipo de predicción se plantea para realizarse de forma temprana y con fines informativos.

Al igual que en los resultados expuestos anteriormente, los experimentos consisten *cross-validation* con distintas series temporales y múltiples algoritmos de aprendizaje supervisado para ver qué algoritmos reportan mejores porcentajes de aciertos para este problema concreto.

Ventana de tiempo	LR	SVC	KNN	NN	DT	AB	RF	NB	GB
Ventana 0-60 minutos	62,74	89,08	91,85	49,57	93,76	87,18	96,19	59,27	96,36
Ventana 0-120 minutos	67,65	91,41	93,24	50,91	93,15	89,68	95,66	59,93	95,66
Ventana 0-180 minutos	68,65	92,77	94,22	50,14	93	88,43	96,01	59,22	95,08
Ventana 0-240 minutos	68,69	91,52	92,58	51,24	92,18	88,68	95,56	59,33	94,32
Ventana 0-final minutos	72,85	54,45	94,87	49,78	95,78	92,54	97,21	63,51	96,62
Ventana 60-120 minutos	70,71	91,68	95,32	50,43	94,28	90,12	96,88	59,62	97,05
Ventana 60-180 minutos	67,82	92,37	94,45	49,09	92,28	88,73	96,18	62,45	95,58
Ventana 60-240 minutos	68,2	91,47	94,09	50,36	92,78	89,68	96,3	59,55	91,23
Ventana 60-final minutos	72,87	59,49	95,23	50,1	96,01	93,37	97,42	63,88	96,93

Tabla 9. Resultados pronóstico tras ataque cerebrovascular con 1 observación

En la Tabla 9 se muestran los resultados F1-Score (porcentaje de aciertos) en predicciones del pronóstico con instancias de diferentes ventanas de tiempo para ver si las variables fisiológicas aportan más información del estado real del paciente en una ventana de tiempo concreta, como puede ser la primera hora de hospitalización. Analizando la Tabla 9 con detenimiento se puede ver

que los resultados tienden a mejorar según la ventana de monitorización es más grande, pero sí que se aprecia una ligera mejoría de precisión cuando se ignora la primera hora de monitorización. Esto puede deberse a que al poco de sufrir el ictus el estado del paciente es más inestable y las variables aportan ruido a la clasificación.

Ventana de tiempo	LR	SVC	KNN	NN	DT	AB	RF	NB	GB
Ventana 0-60 minutos	69,2	73,01	91	45,33	89,27	92,39	96,19	60,9	94,46
Ventana 0-120 minutos	72,44	82,32	91,85	50,43	89,77	88,39	97,23	58,75	95,32
Ventana 0-180 minutos	70,87	84,86	91,79	49,94	90,52	89,13	95,95	61,85	95,61
Ventana 0-240 minutos	70,43	85,44	93,43	50,27	92,18	89,17	96,09	60,92	95,29
Ventana 0-total minutos	76,23	75,54	95,18	49,96	95,41	93,82	97,95	63,75	97,37
Ventana 60-120 minutos	67,13	78,89	93,08	45,33	92,04	93,77	96,89	65,05	96,19
Ventana 60-180 minutos	72,27	84,4	92,72	50,43	87,18	88,21	96,01	59,79	95,67
Ventana 60-240 minutos	71,96	84,96	92,24	51,55	91,29	90,81	97,02	63,01	95,35
Ventana 60-final minutos	77,4	77,1	95,31	50,05	96,17	94,66	98	64,58	97,49

Tabla 10. Resultados pronóstico tras ataque cerebrovascular con 2 observaciones

Ventana de tiempo	LR	SVC	KNN	NN	DT	AB	RF	NB	GB
Ventana 0-60 minutos	70,47	51,3	90,16	55,44	82,38	87,05	92,75	67,36	91,71
Ventana 0-120 minutos	71,43	69,09	89,35	50,13	91,17	87,53	95,58	58,18	94,81
Ventana 0-180 minutos	72,44	69,32	89,95	50,43	87,18	88,56	94,28	62,74	93,59
Ventana 0-240 minutos	71,64	73,37	91,48	49,27	86,82	90,28	96,4	60,59	95,21
Ventana 0-total minutos	77,52	93,7	94,96	50,32	95,18	93,83	97,52	63,12	97
Ventana 60-120 minutos	74,09	55,44	90,67	44,56	85,49	90,16	96,89	61,14	96,37
Ventana 60-180 minutos	72,73	70,13	90,91	49,87	88,83	87,79	93,77	59,22	93,77
Ventana 60-240 minutos	74,78	63,86	90,52	50,45	86,94	88,73	94,45	62,08	93,56
Ventana 60-final minutos	77,84	94,37	95,06	50,17	95,09	94,3	97,7	62,89	97,19

Tabla 11. Resultados pronóstico tras ataque cerebrovascular con 3 observaciones

Ventana de tiempo	LR	SVC	KNN	NN	DT	AB	RF	NB	GB
Ventana 0-60 minutos	63,45	46,21	87,59	45,52	86,9	88,28	93,79	64,83	90,34
Ventana 0-120 minutos	75,43	51,56	87,89	45,33	84,08	88,93	93,08	65,74	90,66
Ventana 0-180 minutos	74,6	57,97	90,07	53,12	90,3	88,22	95,84	60,97	94,92
Ventana 0-240 minutos	72,82	59,5	90,59	50,98	86,86	88,45	96,09	61,63	94,32
Ventana 0-total minutos	77,79	94,26	95,21	50,34	94,74	94,3	97,79	63,23	97,77
Ventana 60-120 minutos	70,34	49,66	91,03	45,52	86,21	88,97	95,86	65,52	93,79
Ventana 60-180 minutos	76,12	50,87	89,97	45,33	91,35	91	95,5	60,55	95,16
Ventana 60-240 minutos	76,13	56,32	92,6	52,27	93,08	91,17	97,14	65,16	94,99
Ventana 60-final minutos	79,29	93,87	95,21	49,92	95,71	95,11	98,06	62,51	97,73

Tabla 12. Resultados pronóstico tras ataque cerebrovascular con 4 observaciones

Ventana de tiempo	LR	SVC	KNN	NN	DT	AB	RF	NB	GB
Ventana 0-60 minutos	74,14	48,28	89,66	48,28	89,66	87,93	97,41	59,48	95,69
Ventana 0-120 minutos	65,8	49,78	88,31	51,52	89,18	87,45	94,37	61,9	91,77
Ventana 0-180 minutos	71,1	51,45	93,64	47,4	90,17	87,57	96,82	60,4	94,51
Ventana 0-240 minutos	71,78	54,89	88,89	49,78	83,56	86,67	94,44	58,22	94,67
Ventana 0-total minutos	78,75	92,26	94,06	49,86	93,56	94,01	97,36	62,31	96,89
Ventana 60-120 minutos	75,00	51,72	83,62	48,28	81,9	87,93	98,28	62,07	93,97
Ventana 60-180 minutos	70,56	51,52	90,48	51,52	85,28	89,18	92,64	58,44	90,91
Ventana 60-240 minutos	68,36	50,15	89,85	48,96	84,48	87,46	94,93	57,61	92,24
Ventana 60-final minutos	77,96	92,39	95,22	49,37	94,67	94,37	97,58	62,05	97,33

Tabla 13. Resultados pronóstico tras ataque cerebrovascular con 5 observaciones

Ahora bien, en las Tablas 10, 11, 12 y 13, donde se pueden ver los resultados de todos los algoritmos en distintas ventanas y varias observaciones de pacientes por instancia de entrenamiento/testing. En la Tabla 10 se utilizan dos observaciones por instancia de entrenamiento/testing, mientras que en la Tabla 11 son 3 observaciones, en la Tabla 12 se predice con 4 observaciones y en la Tabla 13 con 5 observaciones de pacientes. El objetivo es ver si utilizando un mayor espectro de los pacientes se puede realizar una mejor predicción del pronóstico. En este caso, sí que se aprecia una mejoría al aumentar el número de observaciones por instancia, aunque si bien es cierto que la mejoría es mínima, alrededor de un 1%. En cuanto a los algoritmos, los algoritmos que obtienen mejores resultados siguen siendo, como en el apartado anterior, Random Forests, Nearest Neighbors, Decision Trees y Gradient Boosting. Resulta especialmente significativo que de estos cuatro clasificadores, tres son basados en árboles de clasificación, ya sean múltiples (*ensembles*) o simples, y el otro se basa en comparación de ejemplos comparando la cercanía basándose en la métrica de distancia de Minkowski.

Por otra parte, otros clasificadores obtienen resultados realmente bajos como es el caso de MLP(Neural Network) o Naive Bayes donde los resultados en algunos casos y en algunas ventanas de tiempo reportan resultados inferiores a la precisión mínima que se podría obtener, que es un 50% puesto que es una clasificación binaria y el conjunto de entrenamiento está equilibrado para que haya en promedio un 50% de datos de cada clase tanto en *testing* como en *training*. Esto indica claramente que estos algoritmos no consiguen llegar a obtener un patrón en los datos y no sirven realmente para este tipo de experimento.

### Comparativa Algoritmos

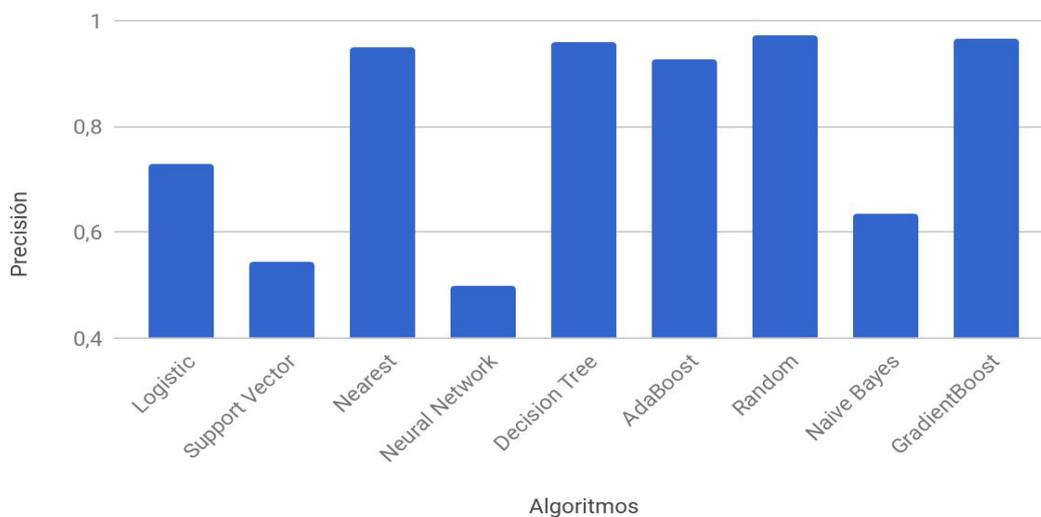


Figura 7. Comparativa de algoritmos predicción de pronóstico

En la Figura 7 se puede observar la precisión de todos los algoritmos utilizados para el experimento sin ventanas de tiempo. En la Figura 7 se ven con gran diferencia los algoritmos que obtienen mejores rendimientos en este tipo de predicción, que son KNN, Decision Trees, AdaBoost, Random Forests y Gradient Boosting mientras que el resto obtienen resultados de precisión notablemente inferiores.

En contraposición con los experimentos realizados para predecir el tipo de ataque cerebrovascular donde la Regresión Logística difícilmente llegaba a obtener en algunos casos un 60% de aciertos aquí se pueden llegar a ver resultados alrededor del 70% de aciertos e incluso hasta un 80% en algunas ventanas concretas y con ciertas series. Esto indica que la regresión logística, aun siendo un algoritmo que por lo general ha reportado resultados poco aceptables, puede llegar a ser una opción viable si se investiga en un futuro los parámetros del algoritmo buscando optimizar más estos resultados. Sin embargo, es bastante complicado que llegase a obtener resultados tan satisfactorios como los que se obtiene en ciertas ventanas de tiempo y con ciertas series temporales con algoritmos como Random Forests donde se llega a obtener resultados superiores al 98% de aciertos.

Otros algoritmos como AdaBoost o SVC obtienen resultados muy aceptables, alrededor del 90%, pero al ser resultados inferiores a los mejores obtenidos con otros algoritmos no se tienen excesivamente en cuenta a futuro como algoritmos que encajen especialmente bien en este tipo de predicción.

Explicado esto, a partir de aquí se van a tener en cuenta únicamente los resultados de los algoritmos KNN, Decision Trees, Gradient Boosting y Random Forests.

En este tipo de predicción, al igual que en la predicción del tipo de ataque cerebrovascular sufrido, es muy importante ver si los datos de monitorizaciones recogidas son más significantes utilizando únicamente datos de las primeras horas tras el ingreso hospitalario o si por el contrario no es un factor relevante. Para comprobar esto se realizan filtrados de datos por ventanas de tiempo, teniendo en cuenta las primeras horas únicamente o incluso desechando la primera hora de monitorización.

Cuando se utilizan únicamente datos de una ventana de datos concreta significa que se tienen únicamente los datos de esa ventana de tiempo tanto para entrenar los clasificadores como para testear, lo que implica que el conjunto de datos se reduce significativamente. Esto, al igual que en el resto de experimentos, es un handicap a la hora de realizar el entrenamiento ya que al disponer de menos instancias se pueden dar casos de *underfitting*, conllevando a problemas de convergencia. Además, no sólo al establecer ventanas de tiempo reducidas se está reduciendo el conjunto de datos, sino que cuando se aumentan las series temporales también se está reduciendo el conjunto de

datos puesto que se agrupan varias observaciones del paciente en ejemplos de *training* y *testing*.

En la Figura 8 se muestran los algoritmos que mejores resultados han obtenido en distintas ventanas de tiempo en los experimentos de predicción del pronóstico del paciente.

Observando la Figura 8, se puede ver con claridad que los porcentajes de aciertos tienden a aumentar a medida que aumenta el margen de la ventana de tiempo. Esto parece indicar que en las primeras horas de monitorización no hay mayores indicadores de pronóstico y además parece claro que al aumentar la ventana de muestreo de los pacientes se traduce en un mayor porcentaje de aciertos. Este hecho señala que al haber más observaciones de pacientes se realiza un mejor entrenamiento (evitando *overfitting*) y esta es la razón por la que aumenta la precisión de los resultados.

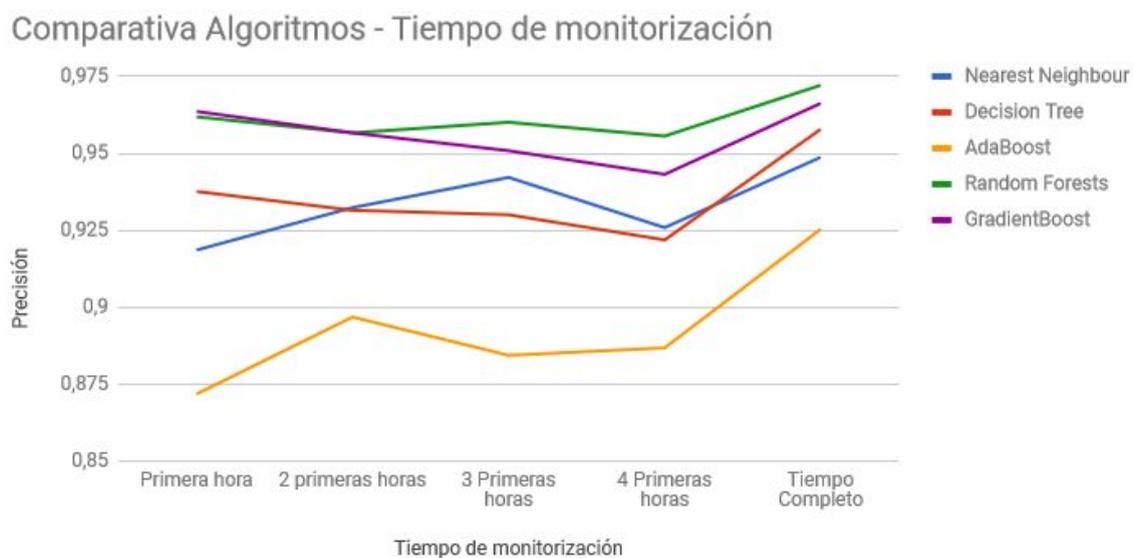


Figura 8. Comparativa de mejores algoritmos predicción de pronóstico en función de la ventana de monitorización

Una vez explicado esto, observando detenidamente las tablas (Tabla 9, 10, 11, 12 y 13) de resultados de los algoritmos más óptimos se pueden ver que sí que existe una correlación entre ventanas de tiempo reducidas y ventanas de tiempo más grandes, siendo ligeramente (alrededor de un 1% superior) mayor el porcentaje de aciertos cuando no se establecen ventanas de tiempo fijas.

En cuanto a los resultados en función del número de series temporales, no se aprecia una gran variación de los resultados. Por lo general, se obtiene mejores resultados de media con series temporales de 2 observaciones, y peores con series de 5 observaciones. Esto puede significar que la capacidad de predicción aumenta incrementando las observaciones por serie temporal pero al

disminuir el conjunto de datos de entrenamiento conlleva un pequeño descenso de la precisión de los clasificadores.

En conclusión, los resultados obtenidos en este experimento son realmente satisfactorios ya que de media se obtienen porcentajes alrededor del 97% con algoritmos como Random Forest, llegando con facilidad al 98% de aciertos. Estos resultados podrían ser fácilmente mejorables realizando una configuración más exhaustiva del algoritmo elegido para la clasificación, que por máximos de precisión y medias de precisión debería ser Random Forests.

### 4.3. Predicción de Resangrado

Finalmente, los últimos experimentos realizados tratan sobre la predicción de un posible caso de resangrado cerebral en el paciente. Para estos experimentos, a diferencia del resto de predicciones, se tiene en cuenta el tipo de ataque cerebrovascular como una característica del paciente para las predicciones. Esta característica se cree clave a la hora de poder predecir el paciente sufrirá una recaída en forma de resangrado por lo que se decidió tener en cuenta este factor durante la fase de *training* y de *testing*.

Ventana de tiempo	LR	SVC	KNN	NN	DT	AB	RF	NB	GB
Ventana 0-60 minutos	69,56	79,18	86,82	50,91	92,11	82,22	93,67	58,63	92,71
Ventana 0-120 minutos	67,46	82,52	86,98	49,5	90,93	81,17	95,18	61,04	93,88
Ventana 0-180 minutos	64,5	67,84	85,49	50,97	89,61	78,14	93,93	57,21	92,77
Ventana 0-240 minutos	66,96	61,79	85,55	50,25	90,36	78,98	93,66	59,51	92,04
Ventana 0-total minutos	59,67	52,09	83,74	49,79	90,16	77,67	92,9	56,81	91,6
Ventana 60-120 minutos	64,96	82,05	85,26	49,09	92,02	82,13	94,8	61,93	94,45
Ventana 60-180 minutos	64,5	83,91	85,3	49,85	89,36	80,11	93,81	61,71	93,15
Ventana 60-240 minutos	63,95	71,81	84,41	51,05	90,07	79,31	93,27	60,04	92,68
Ventana 60-final minutos	59,37	53,84	83,94	49,88	90,25	77,98	92,7	56,39	91,54

Tabla 14. Resultados de predicción de resangrado tras ataque cerebrovascular con 1 observación

En Tabla 14 se muestran los resultados (precisión F1-Score) de predicción de resangrado en varias ventanas de tiempo y algoritmos con una observación por instancia. Observando con detenimiento, se puede ver que en general los resultados de precisión son más bajos que en resto de experimentos, llegando en ventanas de tiempo concretas a máximos cercanos al 95%.

En este caso las ventanas de tiempo no parecen indicar una etapa de la monitorización del paciente en la que se indique con mayor precisión si va a sufrir un resangrado. Sin embargo, sí que se observa que durante las primeras 2 horas de monitorización se obtienen mejores resultados en algoritmos de tipo arbóreo.

Ventana de tiempo	LR	SVC	KNN	NN	DT	AB	RF	NB	GB
Ventana 0-60 minutos	71,23	58,58	85,44	50,43	90,12	85,62	95,15	70,19	93,07
Ventana 0-120 minutos	63,83	76,06	86,21	49,09	88,29	80,83	94,45	63,49	93,32
Ventana 0-180 minutos	66,11	83,52	85,26	49,97	89,26	79,69	94,72	62,68	92,98
Ventana 0-240 minutos	67,87	84,53	85,42	51,04	88,47	79,85	95,05	64,34	94,3
Ventana 0-total minutos	62,46	54,86	86,98	49,73	89,53	78,33	94,04	57,45	91,85
Ventana 60-120 minutos	65,34	55,98	88,56	49,57	89,77	85,1	96,19	63,6	94,8
Ventana 60-180 minutos	64,6	77,24	85,7	50,48	86,75	79,16	94,94	61,29	92,94
Ventana 60-240 minutos	64,2	82,81	84,94	49,97	88,92	80,2	94,13	62	92,95
Ventana 60-final minutos	61,88	50,46	87,07	49,82	89,35	78,88	93,98	57,78	92,14

Tabla 15. Resultados de predicción de resangrado tras ataque cerebrovascular con 2 observaciones

Ventana de tiempo	LR	SVC	KNN	NN	DT	AB	RF	NB	GB
Ventana 0-60 minutos	68,83	50,39	83,12	50,13	85,71	78,44	92,99	66,49	90,39
Ventana 0-120 minutos	65,67	50,98	84,53	48,89	87,39	80,88	94,8	64,5	91,94
Ventana 0-180 minutos	67,86	75,26	85,45	49,65	88,68	81,36	94,6	64,11	93,64
Ventana 0-240 minutos	68,17	70,89	85,94	50	88,46	82,29	95,49	64,59	94,5
Ventana 0-total minutos	62,53	52,74	87,14	49,87	88,71	78,38	93,89	58,28	91,81
Ventana 60-120 minutos	67,53	54,03	84,42	50,13	88,83	83,12	96,88	59,74	93,25
Ventana 60-180 minutos	66,88	59,69	87,7	50,79	88,35	83,77	95,03	64,53	93,46
Ventana 60-240 minutos	64,5	69,4	85,5	50,8	88,7	78,65	93,68	62,1	92,53
Ventana 60-final minutos	62,37	57,3	86,89	49,83	88,81	78,84	94,14	58,22	91,77

Tabla 16. Resultados de predicción de resangrado tras ataque cerebrovascular con 3 observaciones

Ventana de tiempo	LR	SVC	KNN	NN	DT	AB	RF	NB	GB
Ventana 0-60 minutos	66,09	45,33	80,62	54,67	82,35	78,2	93,43	68,86	88,24
Ventana 0-120 minutos	71,75	50,61	84,06	50,43	87,18	84,92	96,71	67,59	93,76
Ventana 0-180 minutos	63,34	50,35	81,79	49,3	84,8	77,73	93,39	62,65	92,23
Ventana 0-240 minutos	67,23	72,79	85,78	50,44	87,01	79,51	94,43	65,02	91,34
Ventana 0-total minutos	62,91	56,57	86,99	50,18	87,88	78,97	93,96	61,41	91,52
Ventana 60-120 minutos	63,32	45,33	82,35	54,67	84,08	76,12	92,39	61,59	89,27
Ventana 60-180 minutos	66,38	49,13	85,71	48,78	83,8	81,71	93,21	63,07	89,55
Ventana 60-240 minutos	60,66	49,05	84	51,78	85,78	79,62	92,77	60,31	91,59
Ventana 60-final minutos	62,62	56,01	86,42	49,71	88,15	78,74	93,97	61,84	91,55

Tabla 17 - Resultados de predicción de resangrado tras ataque cerebrovascular con 4 observaciones

Ventana de tiempo	LR	SVC	KNN	NN	DT	AB	RF	NB	GB
Ventana 0-60 minutos	64,07	48,48	80,09	48,48	78,79	75,32	90,04	61,47	84,42
Ventana 0-120 minutos	66,38	47,72	85,47	47,72	86,55	79,61	95,44	62,47	93,93
Ventana 0-180 minutos	69,96	49,06	84,33	49,06	85,49	79,97	95,36	61,97	93,47
Ventana 0-240 minutos	67,85	54,92	85,3	51,49	84,86	82,1	93,92	60,66	91,16
Ventana 0-total minutos	62,42	56,08	86,42	49,66	87,16	78,7	94,08	63,19	91,9
Ventana 60-120 minutos	57,58	48,48	87,88	48,48	83,55	81,82	90,48	55,84	85,71
Ventana 60-180 minutos	65,14	48,37	83,66	48,37	83,44	78,43	95,21	56,86	90,2
Ventana 60-240 minutos	67,41	49,63	83,7	50,37	87,7	80,3	94,07	63,56	91,7
Ventana 60-final minutos	63,31	57,91	86,41	49,98	87,14	79,2	93,71	63,07	91,26

Tabla 18. Resultados de predicción de resangrado tras ataque cerebrovascular con 5 observaciones

En las Tablas 15, 16, 17 y 18 se pueden ver los resultados de predicción de resangrado de los distintos modelos de predicción probados en función de distintas ventanas de tiempo y distintas observaciones por instancia de entrenamiento. En cada tabla se muestran los resultados utilizando desde 1 observación por instancia de entrenamiento, como se muestra en la Tabla 14, hasta 5 observaciones por instancia que son los resultados mostrados en la Tabla 18. En este tipo de predicción sí que parece aumentar de forma general el porcentaje de aciertos en algunos algoritmos, concretamente los de tipo arbóreo, obteniendo porcentajes de acierto de hasta un 97%, como ocurre con 2 y 3 observaciones por instancia de entrenamiento. Sin embargo, al aumentar a 5 observaciones por instancia los resultados no tienden a mejorar, sino a disminuir, lo que se debe a la disminución del conjunto de instancias para entrenamiento o testing, conllevando a un posible *underfitting*. Respecto a las ventanas de tiempo ocurre prácticamente lo mismo en general, aunque sí que es cierto que en algunas ventanas de tiempo como en las dos primeras horas de monitorización se obtienen mejores resultados, si bien es cierto que la diferencia es de apenas un 2%.

## Comparativa de algoritmos

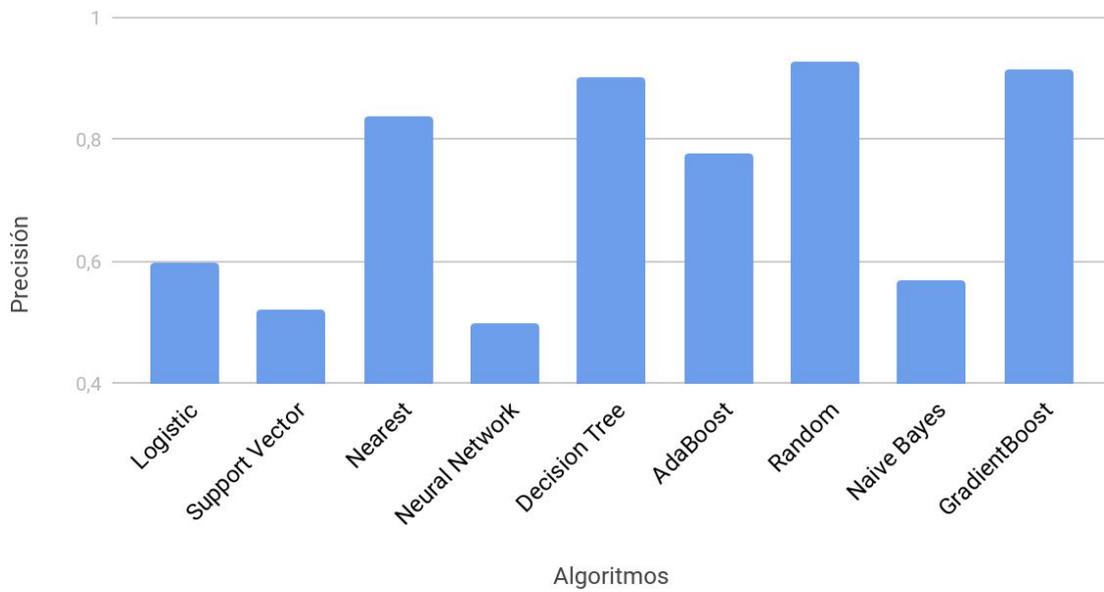


Figura 9. Comparativa de algoritmos predicción de resangrado

En la Figura 9 se pueden ver una comparativa de todos los algoritmos donde se muestran los resultados en términos de precisión de predicción de resangrado. Observando rápidamente la Figura 9, los algoritmos que han reportado mejores resultados en cuanto a precisión se refiere se puede observar claramente en las tablas que, al igual que todos los tipos de predicciones realizados y explicados en este trabajo, los mejores algoritmos son Random Forests, Decision Trees, Gradient Boost y Nearest Neighbors (KNN). La principal diferencia con respecto a los anteriores experimentos radica en que el porcentaje de aciertos de Nearest Neighbors es alrededor de un 10% inferior y que por lo general, la tasa de aciertos es ligeramente inferior (entre un 3-5%).

De forma similar a los resultados obtenidos en la predicción de pronóstico, los algoritmos que peor se han comportado son Neural Network y Naive Bayes. Definitivamente estos algoritmos, que apenas consiguen llegar al mínimo porcentaje de aciertos ( 50%) que se obtendría al clasificar todas las instancias de la misma clase, resultan totalmente inapropiados para este tipo de clasificación, por lo que seguir investigando en posibles configuraciones de estos clasificadores probablemente no sería un enfoque muy eficaz.

En un punto intermedio entre los mejores algoritmos y los peores se hayan Regresión Logística,

SVC y Ada Boost que, sin obtener malos resultados, no llegan a un porcentaje de acierto que se considere como aceptable. Los resultados que obtiene SVC son en muchas ocasiones similares a los que obtiene AdaBoost, que son bastante aceptables. Sin embargo, SVC obtiene resultados realmente bajos muy frecuentemente, lo que se debe a que es un tipo de algoritmo muy costoso en cómputo y conseguir que el algoritmo converja en un tiempo razonable es difícil en este tipo de problemas de decisión con multivariantes, lo que ha llevado a que el límite de iteraciones sea alcanzado con cierta facilidad y no haya conseguido terminar el entrenamiento, resultando en una clasificación ingenua con resultados cercanos al 60%. Las ocasiones en las que converge o deja de converger coinciden con ventanas de tiempo reducidas y con ventanas de tiempo amplias, respectivamente.

Puesto que los mejores resultados se obtienen con Random Forests, Gradient Boosting y Decision Trees, son los algoritmos que se van a tener en cuenta para los comentarios que se realizan a continuación, descartando el resto de algoritmos.

### Comparativa Algoritmos - Tiempo de monitorización

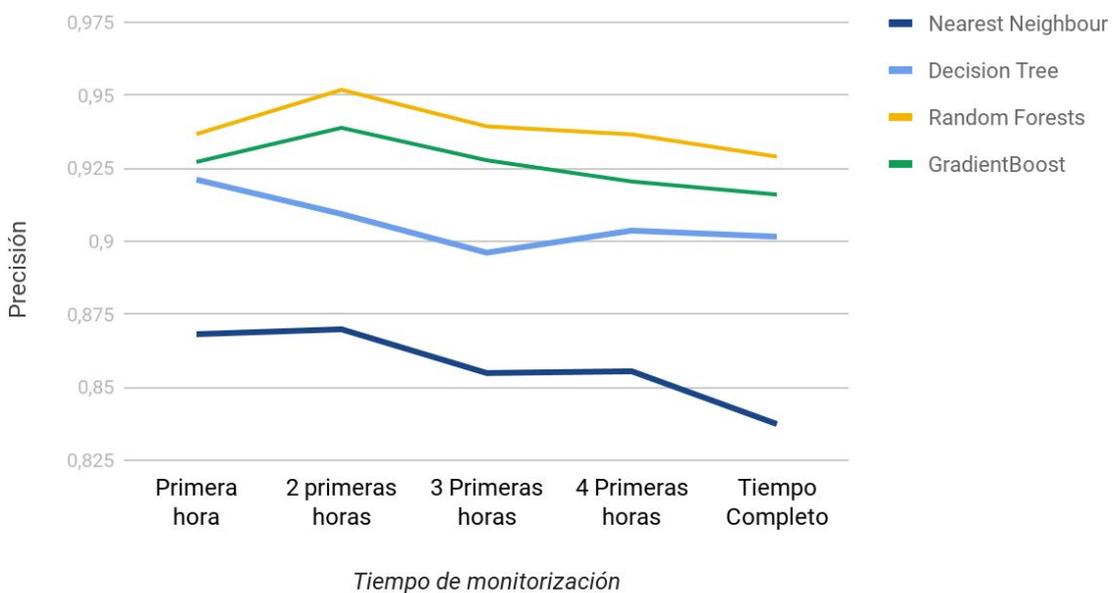


Figura 10. Comparativa de mejores algoritmos predicción de resangrado en función de la ventana de monitorización

En la Figura 10 se muestran la precisión de los algoritmos que mejores resultados han obtenido en función de distintas ventanas de monitorización de los pacientes. Tal y como se puede observar en la Figura 10, se produce un patrón distinto en cuanto a las ventanas de tiempo en comparación con el resto de predicciones. En el resto de predicciones los resultados tienden a mejorar ligeramente según aumenta el tiempo total de la ventana de tiempo, mientras que en este caso tiende a empeorar

ligeramente. Resulta especialmente llamativo que los mejores porcentajes de aciertos se obtienen en la ventana de tiempo de la segunda hora de monitorización tras el ingreso del paciente, obteniendo picos del 96% de aciertos en Random Forests.

Al igual que en el resto de experimentos, observando la Figura 9, el número de observaciones por serie temporal también parece tener cierta importancia, ya que se obtienen mejores resultados con series temporales de 2 y de 3 observaciones con respecto a observaciones de una instancia. Como es entendible debido a la importante disminución del conjunto de datos al aumentar el número de observaciones a 4 o 5 por ejemplo se obtienen peores resultados. Esto, al igual que en el resto de predicciones, tiene como origen del problema la falta de datos para *training* y *testing* aunque con un conjunto de datos con más pacientes es muy probable que se obtengan resultados aún mejores con series temporales de 4 o 5 observaciones.

Los resultados obtenidos son bastante positivos, con picos de acierto del 96%, lo que lleva a pensar que realizando una mayor optimización en la configuración de los algoritmos es factible obtener porcentajes de acierto aún mayores.

## 5. Conclusiones

En este trabajo se ha realizado un estudio sobre series temporales fisiológicas de pacientes que han sufrido un ataque cerebrovascular por medio de técnicas de inteligencia artificial. Se ha realizado una recopilación de datos de pacientes en colaboración con la Unidad de Ictus del Hospital Universitario de La Princesa. Los datos fueron unificados, procesados y transformados para poder ser utilizados para distintos objetivos predictivos por medio de múltiples algoritmos de aprendizaje supervisado con el fin de encontrar los modelos que mejor encajan para realizar cada tipo de predicción. Como método de evaluación de los distintos modelos se utilizó *random cross-validation*.

Para elaborar las conclusiones se discuten brevemente los resultados obtenidos. En este trabajo se han elaborado modelos predictivos de tres tipos:

- Detección de ataque cerebrovascular sufrido.
- Predicción de pronóstico del paciente.
- Predicción de resangrado.

Con respecto al primer tipo de predicción, se predice el tipo de ataque cerebrovascular que ha sufrido el paciente, que puede ser de dos tipos: isquémico y hemorrágico. En este tipo de predicción se parte de una proporción de ataques muy desequilibrada, siendo de aproximadamente un 90% los casos de ataques isquémicos y el 10% restante ataques hemorrágicos. Para evitar entrenamiento y predicciones sobre un único tipo debido a la cantidad de instancias de cada clase se han utilizado distribuciones de datos de pacientes equilibradas donde aproximadamente el 50% de los datos era de cada tipo. Los resultados han sido, en comparación con el resto de predicciones, los más bajos, siendo de igual forma bastante positivos, donde se obtienen con ciertos algoritmos porcentajes de acierto alrededor del 94%. Sin embargo, las pruebas realizadas con pacientes reales sobre los que inicialmente no se tenía constancia de su tipo de ataque cerebrovascular sufrido han resultado muy negativas. No solo los porcentajes de acierto no se acercaban a los obtenidos a los obtenidos en *cross-validation*, sino que tras analizar los resultados obtenidos se detectó un bajísimo porcentaje en detección de ataques de tipo hemorrágicos, lo que encubría el porcentaje final de aciertos al ser una distribución de pacientes descompensada (aproximadamente 90-10%). Es importante mencionar en este caso que, aunque los resultados en estas pruebas han resultado no satisfactorios,

es factible y probable mejorar las predicciones con modelos de predicción de varios niveles compuesto por algoritmos de predicción de distintos tipos.

Con respecto al segundo tipo de predicción, donde se intenta predecir si las variables monitorizadas del paciente indican que probablemente no supere el ataque y finalmente fallezca o no, los resultados son excelentes llegando a porcentajes de acierto cercanos al 98%. Este tipo de pronóstico se planteó desde el inicio como un posible mecanismo o herramienta adicional para el médico para informar a los pacientes del estado y de la previsión del paciente, siempre desde un enfoque informativo y nunca concluyente.

Finalmente, en el tercer tipo de predicción del resangrado donde se intenta predecir si, un paciente tras sufrir un tipo de ataque concreto y con el conjunto de variables monitorizadas, va a sufrir una recaída de forma que se produzca un resangrado en alguna zona del cerebro que en muchos casos empeora la situación del paciente. Los resultados en este experimento han resultado muy positivos, llegando en ciertas condiciones concretas a porcentajes de acierto del 95%.

Extrayendo información de todas las pruebas realizadas y todos los resultados en su conjunto, se puede afirmar que los algoritmos que mejor han funcionado para estos tipos de predicción (aprendizaje supervisado y clasificación binaria) son de tipo arbóreo, como lo son Random Trees, Gradient Boosting y Decision Trees. Otros algoritmos han obtenido para ciertos problemas de predicción resultados también muy positivos, como es el caso K-Nearest Neighbors o AdaBoost.

Existe un caso especial, el SVC (Support Vector Classifier), que por la propia complejidad en el proceso de entrenamiento no termina convergiendo en un tiempo razonable y alcanza finalmente el número de iteraciones máximas establecidas, llegando a no clasificar en ningún momento. Sin embargo, este algoritmo ha resultado en buenas predicciones cuando ha terminado convergiendo en algunos tipos de predicciones realizadas en este estudio. En un futuro, se podría tener en cuenta como un buen clasificador para este tipo de problemas realizando un estudio más exhaustivo de configuraciones posibles del algoritmo.

En cambio, se puede concluir que algunos algoritmos como las redes neuronales, regresión logística y naive bayes no obtienen resultados positivos al tratarse de una clasificación binaria y no terminan de ofrecer prácticamente ningún factor positivo (el tiempo de computación durante el entrenamiento es largo y los resultados son bajos).

Las conclusiones que se pueden extraer de este estudio son, en general, muy positivas y abren la posibilidad de continuar con la investigación en este enfoque sobre los ataques cerebrovasculares.

Como trabajo futuro se podría diseñar e implementar un dispositivo ergonómico y de bajo coste

que permita realizar las predicciones que se han experimentado en este trabajo.

Además, se podría plantear la creación de modelos predictivos como un modelo compuesto de varios niveles de predicción donde se implementen algoritmos predictivos de distinto tipo. Probablemente un modelo predictivo de varios niveles pueda llegar una fiabilidad real cercana al 100% en casos de predicción como en la predicción del tipo de ataque cerebrovascular, donde sí que se observa que algoritmos de un tipo funcionan bien para predecir un tipo de ataque mientras que otros algoritmos funcionan mejor para detectar otro tipo de ataque.

Otra posible mejora en cuanto a los modelos predictivos consistiría en intentar aplicar técnicas novedosas de procesamiento de series temporales con algoritmos de procesamientos de señales junto con modelos de detección de anomalías o modelos de markov ocultos (Hidden Markov Models, HMM) que podrían mejorar los resultados obtenidos en este trabajo.

## 5. Conclusions

In this work it has been performed a research about physiological temporal series from patients that have suffered a stroke attack by using techniques of artificial intelligence. It has been done a pickup of data from patients in collaboration with the Stroke Unit of the Hospital Universitario de La Princesa. The data was unified, processed and transformed to be able to use them for different predictive-related objectives by using multiple algorithms of supervised learning with the final purpose of finding the models that better fit to perform each kind of prediction. As evaluation method all the models were evaluated by cross-validation.

To elaborate the conclusions the obtained results are discussed briefly. In this work it has been developed predictive models of three types:

- Detection of the stroke attack suffered by the patient.
- Prediction of the prognosis of the patient.
- Prediction of a re-bleeding process of a patient.

Respecting to the first sort of prediction, it is predicted the type of stroke attack that the patient has suffered, that can be of two types: ischaemic and haemorrhagic. In this sort of prediction it has started out from a proportion of type attacks very unbalanced, being approximately about a 90% ischaemic attacks and the remaining 10% haemorrhagic attacks. To avoid trainings and predictions over only one type of attack due to the amount of instances of each class it has been used distributions of the data balanced where approximately the 50% of the data is of each stroke attack type. The results have been, in comparison to the rest of predictions, the lowest results, being equally very positive where on some algorithms there are precision percentages above 94%. However, the tests performed over real patients, who initially didn't have a diagnosis of the stroke attack suffered, have been clearly worse than the cross-validation test. Not only the precision percentages don't get close to the results obtained in the cross-validation test, but after analysing the results it was detected a really lower percentage of detection rate on haemorrhagic stroke attacks and this was finally covering the final results up due to being an unbalanced distribution of patients (approximately 90-10%). It is important to mention that in this case, although the results in these tests have not been completely successful, it is feasible and likely to improve the predictions with prediction models composed of several levels with different type algorithms in each level.

Continuing with the second prediction, where it is pretended to predict if the monitored variables from the patient indicate that probably the patient won't overcome the attack and finally will pass away or not, the results are excellent reaching precision percentages above 98%. This prognosis was proposed from the start as a possible mechanism or additional tool for the medical staff to know the real state of the patient and finally to inform the patient's family so that they know how their related is. Furthermore, this would be also useful to determine the critical window and then take out a contingency plan to help the patient.

Finally, in the third prediction the re-bleeding prediction, where it is proposed to predict if a patient that has already suffered an attack is going to suffer another attack in another area of the brain that in many cases aggravate the state of the patient. This is predicted from the physiological variables monitored from the patient. The results of this prediction have been really successful and positive with precision percentages above 95% in some temporal windows and temporal series. It has been verified that the results tends to be higher within the first two hours of the monitoring of the patient since the patient is admitted, so the first two hours seem to have a determinant window to predict if the patient is going to suffer a relapse or not.

Extracting information from all the tests done and all the results in the whole, it can be affirmed that the algorithms that have performed best for these kind of predictions (supervised learning, binary classification and real value variables) are the ensembles one, like it is the Random Forests algorithm, Gradient Boosting and tree-based Decision Trees. Others algorithms have obtained in some problems positive results, like the distance based algorithm K-Nearest Neighbors and the ensemble AdaBoost.

There is a special case, the SVC (Support Vector Classifier), due to the complexity and amount of the data the training process didn't converge in a reasonable time and reaches the maximum number of iterations allowed. However, this algorithm has given good results when it has ended converging in some tests with short temporal windows where there are fewer instances for training and testing. In a future, this kind of algorithm could be taken into account as a good classifier for this kind of problems performing a more exhaustive study in configurations of this algorithm.

Instead, it can be said that some algorithms like artificial neural networks, logistic regression and naïve Bayes don't get good results with this feature configuration. They didn't offer any positive factor that could be taken into account to keep using them for researches in these kind of problems because the computation time was long and the results were really low.

The conclusions that can be extracted from this work are, in general, very positive and open the possibility to continue the research in this approaches over stroke attacks.

For future work it could be designed and implemented an ergonomic, low cost and optimized device that allow making the predictions explained in this work.

Furthermore, it could be proposed the creation of new predictive models as a new model composed by several levels of prediction where in each level a different algorithm is implemented. This architecture for predictions could likely give a real confidence close to the 100% precision in cases like the diagnosis of the stroke type attack suffered where it can be observed that some algorithms perform really well in predicting one kind of attack and others algorithms perform better to predict the other type of attack.

Another possible improvement regarding the predictive models would consist on applying novelties techniques in temporal series processing with algorithms based on signal processing along with anomaly detections models or hidden Markov models (HMM) that could improve the results obtained in this work.



# Bibliografía

- [1] Díaz-Guzmán, J., Egido-Herrero, J. A., Gabriel-Sánchez, R., Barberà, G., & Fuentes, B. (2008). Bases metodológicas del estudio Iberictus. *Rev Neurol*, 47(12), 617-623.
- [2] Zhang, Q., Xie, Y., Ye, P., & Pang, C. (2013). Acute ischaemic stroke prediction from physiological time series patterns. *The Australasian medical journal*, 6(5), 280.
- [3] Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.
- [4] Ricardo Gutiérrez-Osuna, Leave-one-out Cross Validation .*Wright State University*
- [5] Asadi, H., Dowling, R., Yan, B., & Mitchell, P. (2014). Machine learning for outcome prediction of acute ischemic stroke post intra-arterial therapy. *PloS one*, 9(2), e88225.
- [6] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research*, 12(Oct), 2825-2830.
- [7] MySQL Bible, A. B. (2001). *MySQL*
- [8] McKinney, W. (2012). *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. " O'Reilly Media, Inc. "
- [9] Birkner, M. D., Kalantri, S. P., Solao, V., Badam, P., Joshi, R., Goel, A., ... & Hubbard, A. E. (2007). Creating diagnostic scores using data-adaptive regression: An application to prediction of 30-day mortality among stroke victims in a rural hospital in India. *Therapeutics and clinical risk management*, 3(3), 475.
- [10] Villar, J. R., González, S., Sedano, J., Chira, C., & Trejo-Gabriel-Galan, J. M. (2015). Improving human activity recognition and its application in early stroke diagnosis. *International journal of neural systems*, 25(04), 1450036.
- [11] Chen, Y., Dhar, R., Heitsch, L., Ford, A., Fernandez-Cadenas, I., Carrera, C., ... & Lee, J. M. (2016). Automated quantification of cerebral edema following hemispheric infarction: application of a machine-learning algorithm to evaluate CSF shifts on serial head CTs. *NeuroImage: Clinical*, 12, 673-680.