

UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE TELECOMUNICACIÓN



GRADO EN INGENIERÍA DE TECNOLOGÍAS Y SERVICIOS DE
TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

DISEÑO E IMPLEMENTACIÓN DE UN
SISTEMA DE APRENDIZAJE AUTOMÁTICO
PARA LA PREDICCIÓN DEL NIVEL DE
HEMOGLOBINA EN SANGRE

ALEJANDRO NIETO REMACHA

29 de junio de 2023

GRADO EN INGENIERÍA DE TECNOLOGÍAS Y SERVICIOS DE TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

Título: Diseño e implementación de un sistema de aprendizaje automático para la predicción del nivel de hemoglobina en sangre

Autor: Alejandro Nieto Remacha

Tutor: Dr. Josué Pagán Ortiz

Cotutor: Dr. Jorge Pagán Ortiz

Departamento: Departamento de Ingeniería Electrónica

MIEMBROS DEL TRIBUNAL

Presidente: D.

Vocal: D.

Secretario: D.

Suplente: D.

Los miembros del tribunal arriba nombrados acuerdan otorgar la calificación de:
.....

Madrid, a de de 20...

UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE TELECOMUNICACIÓN



GRADO EN INGENIERÍA DE TECNOLOGÍAS Y SERVICIOS DE
TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

**Diseño e implementación de un
sistema de aprendizaje automático
para la predicción del nivel de
hemoglobina en sangre**

Alejandro Nieto Remacha

29 de junio de 2023

RESUMEN

La sangre es una sustancia vital e indispensable para la supervivencia del ser humano, puesto que juega un papel crucial en la distribución de nutrientes, oxígeno, hormonas y otros componentes importantes a lo largo de todo el organismo. Por lo tanto, es fundamental contar con un suministro adecuado de sangre para aquellos pacientes que puedan necesitarla.

A pesar de la importancia de la sangre, garantizar una cadena de transfusión sanguínea eficiente es una tarea compleja debido a la gran cantidad de variables que pueden afectar su disponibilidad. En este sentido, cabe destacar que alrededor de un 13,3 % de las donaciones de sangre son rechazadas, y de ellas, un 17,8 % se debe a niveles bajos de hemoglobina. Esta situación no solo implica la pérdida de una donación, sino también la posible pérdida de un donante primerizo.

Para solucionar este problema, el presente trabajo de fin de grado (TFG) tiene como objetivo principal colaborar con el Centro Regional de Hemodonación de Murcia en el diseño e implementación de un modelo de aprendizaje automático capaz de predecir el nivel de hemoglobina en sangre mediante el procesamiento de señales de fotopletismografía. Asimismo, se busca obtener otros parámetros sanguíneos a partir de estas señales.

Para ello, se utilizará un dispositivo no invasivo diseñado en un centro de investigación, que permitirá procesar las señales y obtener una serie de características que servirán como parámetros de entrada del modelo. Como resultado, se obtendrá el nivel de hemoglobina en sangre del paciente.

En definitiva, este proyecto busca ofrecer una alternativa menos invasiva y más eficiente para el cálculo del nivel de hemoglobina en sangre, mediante el uso de una red neuronal diseñada específicamente para procesar señales de fotopletismografía y obtener otros parámetros sanguíneos importantes. De este modo, se espera contribuir a mejorar la disponibilidad y calidad de la sangre destinada a pacientes que la necesiten.

SUMMARY

Blood is a vital and indispensable substance for human survival, as it plays a crucial role in the distribution of nutrients, oxygen, hormones and other important components throughout the body. Therefore, it is essential to have an adequate supply of blood for those patients who may need it.

Despite the importance of blood, ensuring an efficient blood transfusion chain is a complex task due to the many variables that can affect its availability. In this regard, it should be noted that around 13.3 % of blood donations are rejected, and of these, 17.8 % are due to low hemoglobin levels. This situation not only implies the loss of a donation, but also the possible loss of a first-time donor.

In order to solve this problem, the main objective of the present final thesis (TFG) is

to collaborate with the Centro Regional de Hemodonación de Murcia in the design and implementation of a neural network capable of predicting the hemoglobin level in blood by processing photoplethysmography signals. The aim is also to obtain other blood parameters from these signals.

For this purpose, a non-invasive device designed at a research center will be used to process the signals and obtain a series of characteristics that will serve as input parameters for the neural network. As a result, the hemoglobin level in the patient's blood will be obtained.

In short, this project seeks to offer a less invasive and more efficient alternative for the calculation of the hemoglobin level in blood, through the use of a neural network specifically designed to process photoplethysmography signals and obtain other important blood parameters. In this way, it is expected to contribute to improving the availability and quality of blood for patients in need.

PALABRAS CLAVE

Hemoglobina, hemodonación, modelo, aprendizaje automático, fotopletismografía, no invasivo, predicción

KEYWORDS

Hemoglobin, hemodonation, model, machine learning, photoplethysmography, noninvasive, prediction

Agradecimientos

A toda la gente que me ha acompañado durante esta bonita etapa universitaria. Espero que sigan acompañándome el resto de mi vida. A mis padres, quienes siempre me han apoyado incondicionalmente.

Agradecer especialmente a Josué Pagán, el mejor tutor que podría haber tenido, quien me ha guiado y enseñado durante este proyecto, permitiendo que disfrute del proceso.

Gracias a todos de corazón.

Índice

Resumen y Palabras Clave	II
Agradecimientos	IV
Lista de acrónimos	IX
1. Introducción y objetivos	1
1.1. Introducción	1
1.1.1. Papel crucial de la sangre en el cuerpo humano	1
1.1.2. La anemia y sus consecuencias en las donaciones	2
1.1.3. Métodos tradicionales de medición de hemoglobina	3
1.1.4. Métodos de prevención	4
1.1.5. Problemas y desafíos detectados	4
1.2. Objetivos	4
1.3. Procedimiento del estudio previo a la donación	5
2. Estado del arte	7
2.1. Introducción al estado del arte	7
2.2. Fundamentos teóricos	7
2.3. Métodos convencionales de medición de la hemoglobina	9
2.4. Investigaciones previas sobre la predicción de la hemoglobina	10
2.5. Brecha identificada y justificación del proyecto	13
3. Metodología	15
3.1. Obtención de señales PPG	15
3.2. Procesamiento de señales PPG	16
3.3. Diseño del modelo	22
3.3.1. Regresión Lineal	23
3.3.2. Random Forest	24
3.3.3. Gradient Boosting	25
3.3.4. Decision Tree Regression (DTR)	26
3.3.5. SVR	28
4. Implementación	29
4.1. Descripción de los datos utilizados	29
4.2. Descripción de las herramientas y recursos utilizados	29
5. Resultados	32
5.1. Regresión Lineal	32
5.2. Random Forest	33
5.3. Gradient Boosting	35
5.4. Decision Tree Regression	36
5.5. Support Vector Regression	39

5.6. Discusión de los resultados	41
6. Conclusiones y líneas futuras	42
6.1. Conclusiones	42
6.2. Líneas futuras	42
Bibliografía	44
Anexo A: Aspectos éticos, económicos, sociales y ambientales	47
Anexo B: Presupuesto económico	49

Índice de figuras

1.1. Relación entre donaciones y donantes 2017-2019 [8].	3
1.2. Objetivos del proyecto	5
2.1. Funcionamiento de un sistema de fotopletismografía basado en el sensor MAX30102	8
2.2. Principio básico de los sensores PPG funcionando en: (a) transmisión, (b) modos de reflexión [15]	8
2.3. Dispositivo de medida de hemoglobina capilar HemoCue	10
2.4. La señal PPG, las derivadas primera y segunda de la señal PPG y los puntos característicos	11
2.5. Rasgos característicos adquiridos a partir de la señal PPG	12
2.6. Rasgos característicos adquiridos a partir de la primera y la segunda derivada	12
3.1. Arquitectura del entorno completo, con los dispositivos hardware utilizados y los protocolos de comunicación establecidos entre ellos [11].	15
3.2. Proceso de obtención de señales PPG	16
3.3. Señales tras la eliminación de la componente continua	17
3.4. Filtrado paso bajo	18
3.5. Detección de picos y creación de envolventes	19
3.7. Fórmula para el cálculo del Sp02	19
3.6. Cálculo de la frecuencia respiratoria	20
3.8. Fórmula para el cálculo de R	20
3.9. Ejemplo de regresión lineal ²	24
3.10. Funcionamiento de un modelo Random Forest ³	24
3.11. Funcionamiento del train-test split ⁴	25
3.12. Funcionamiento de un modelo de Gradient Boosting ⁵	26
3.13. Ejemplo de Regresión de Árbol de Decisión ⁶	26
3.14. Ejemplo de modelo utilizando AdaBoost ⁷	27
3.15. Funcionamiento de un modelo de SVR ⁸	28
4.1. Prototipo del entorno de adquisición, donde se encuentra el nodo dentro de la carcasa, conectada por cables al sensor que está dentro del dedal y el móvil recibiendo las medidas [12].	30
4.2. Función de detección de picos de HeartPy ¹	30
5.1. Valores previstos frente a valores reales en el modelo de Regresión Lineal para la hemoglobina capilar	33
5.2. Valores previstos frente a valores reales en el modelo de Regresión Lineal para la hemoglobina venosa	34
5.4. Valores previstos frente a valores reales en el modelo de Random Forest para la hemoglobina venosa	34
5.3. Valores previstos frente a valores reales en el modelo de Random Forest para la hemoglobina capilar	36

5.5. Valores previstos frente a valores reales en el modelo de Gradient Boosting para la hemoglobina capilar	37
5.7. Valores previstos frente a valores reales en el modelo de Decision Tree Regression para la hemoglobina capilar	37
5.6. Valores previstos frente a valores reales en el modelo de Gradient Boosting para la hemoglobina venosa	38
5.8. Valores previstos frente a valores reales en el modelo de Decision Tree Regression para la hemoglobina venosa	38
5.9. Valores previstos frente a valores reales en el modelo de Support Vector Regression para la hemoglobina capilar	40
5.10. Valores previstos frente a valores reales en el modelo de Support Vector Regression para la hemoglobina venosa	40

Índice de tablas

3.1. Características basales	21
3.2. Características calculadas para la señal roja (idem para la señal infrarroja)	22
5.1. Características seleccionadas para el Random Forest	35
5.2. Características seleccionadas para el Gradient Boosting	35
5.3. Características seleccionadas para el Decision Tree Regression	39
5.4. Características seleccionadas para el Support Vector Regression	39
6.1. Presupuesto económico	49

Lista de acrónimos

PPG *Photoplethysmography*

Hb *Hemoglobin*

NIR-LED *Near-Infrared Light*

CRHM *Centro Regional de Hemodonación de Murcia*

IBI *Interbeat interval*

BR *Breath rate*

SW *Software*

HW *Hardware*

TFG *Trabajo de Fin de Grado*

SpO₂ *Oxygen saturation*

NO *Nitric oxide*

LED *Light-emitting diode*

IR *Infrared*

PHD *Photodetector*

MSE *Mean squared error*

MAE *Mean absolute error*

R₂ *Coeficiente de determinación*

DC *Direct current*

LPF *Low pass filter*

SNR *Signal-to-Noise Ratio*

HR *Heart rate*

1. Introducción y objetivos

1.1. Introducción

En este capítulo se presentarán los distintos componentes de la sangre y cómo esta juega un papel crucial en nuestras vidas. Se introducirá la enfermedad de la anemia y las consecuencias que produce a la hora de donar sangre, y por último, se presentarán los objetivos que pretende cumplir este trabajo, buscando una solución al problema de la ineficacia en la donación de sangre.

1.1.1. Papel crucial de la sangre en el cuerpo humano

La sangre es un fluido acuoso ligeramente alcalino (plasma) en el que están suspendidos un gran número de glóbulos rojos (eritrocitos), glóbulos blancos (leucocitos) y plaquetas que flotan libremente. La función principal de la sangre es transportar el oxígeno de los pulmones a todas las células del organismo y trasladar el dióxido de carbono en dirección de retorno tras ser producido por el metabolismo de las células. La sangre también transporta nutrientes a las células y lleva los productos de desecho al hígado y los riñones [1].

Los glóbulos rojos o eritrocitos regulan la función vascular mediante la modulación del suministro de oxígeno y la eliminación y generación de óxido nítrico (NO). En primer lugar, la hemoglobina del glóbulo rojo se une al oxígeno en los pulmones y lo transporta a los tejidos de todo el organismo en un proceso regulado alostéricamente, modulado por las concentraciones de oxígeno, dióxido de carbono y protones. La vasculatura responde a las bajas tensiones de oxígeno mediante la vasodilatación, lo que aumenta el flujo sanguíneo y el transporte de oxígeno por los eritrocitos [2].

Los glóbulos blancos o leucocitos, que circulan en la sangre, cumplen la mayor parte de sus funciones fuera de la circulación. Para ello disponen de sistemas que responden a estímulos específicos y que les permiten entrar y circular por el medio extravascular. Proporcionan principalmente una defensa de primera línea contra la invasión microbiana mediante la fagocitosis y la eliminación, además de respuestas inmunitarias celulares y humorales específicas [3].

Por último, las plaquetas son células sanguíneas especializadas que desempeñan un papel fundamental en los procesos fisiológicos y patológicos de hemostasia, inflamación, metástasis tumoral, cicatrización de heridas y defensa del huésped. La activación de las plaquetas es crucial para su función, que incluye una compleja interacción de moléculas de adhesión y señalización [4].

Podemos afirmar que la sangre es esencial para el cuerpo humano ya que realiza muchas

funciones vitales, como transportar nutrientes, oxígeno y hormonas por todo el cuerpo, eliminar desechos y toxinas, mantener el equilibrio ácido-base y regular la temperatura corporal. Además, la sangre es esencial para la defensa del cuerpo a través de la coagulación y las respuestas inmunitarias.

La sangre donada se utiliza en una amplia variedad de tratamientos médicos, desde transfusiones sanguíneas de emergencia hasta tratamientos de enfermedades crónicas. Cada componente de la sangre donada tiene una función específica y es esencial que se utilicen adecuadamente. Además, es importante someter la sangre donada a pruebas rigurosas para detectar enfermedades transmisibles por la sangre y garantizar la identificación adecuada del paciente y la compatibilidad de la sangre [5].

De acuerdo a la Federación Española de Donantes de Sangre¹, encontramos los siguientes datos:

- La sangre tiene una vida máxima de 7 días
- El 50 % de la población española necesitará alguna donación sanguínea
- La tasa de donación entre los posibles donantes es del 5 %

En conclusión, es crucial asegurar una buena cadena de transfusión sanguínea para garantizar unas reservas adecuadas de sangre para abordar emergencias médicas, cirugías y tratamientos que requieran transfusiones. Además, tener una buena disponibilidad de sangre también es importante para tratar pacientes con enfermedades crónicas como la anemia o el cáncer, que pueden necesitar transfusiones de forma regular.

1.1.2. La anemia y sus consecuencias en las donaciones

La anemia es una enfermedad de la sangre que se define como la disminución de la concentración de hemoglobina en el organismo, siendo generalmente los valores normales por encima a los 12 gramos por decilitro en la mujer, y a 13,5 en el hombre [6].

Según los datos de la Organización Mundial de la Salud², el 24 % de la población mundial presenta algún tipo de anemia. Esto provoca que un importante porcentaje de las donaciones sean rechazadas por esta causa, sin que el donante sepa que sufre de esta condición. Por ejemplo, en Estados Unidos aproximadamente el 10 % de las donaciones son rechazadas por este motivo. También cabe recalcar que los donantes rechazados por niveles bajos de hemoglobina donan un 30 % menos de lo que hubieran donado durante los siguientes 4-5 años [7].

Además, de acuerdo a las estadísticas de la Comunidad de Madrid aproximadamente el 75 % de estos donantes son primerizos, y al ser rechazados, la pérdida de la donación conlleva también en la mayoría de ocasiones la pérdida de un donante [8]. (ver Figura 1.1).

¹Federación Española de Donantes de Sangre:
<https://www.donarsangre.org/blog/7-datos-que-seguro-te-sorprenderan/>. Consultado el 18 de enero de 2023

²Organización Mundial de la Salud:
https://www.consalud.es/pacientes/anemia-problema-salud-publica-en-espana-en-mundo_118129_102.html. Consultado el 24 de enero de 2023

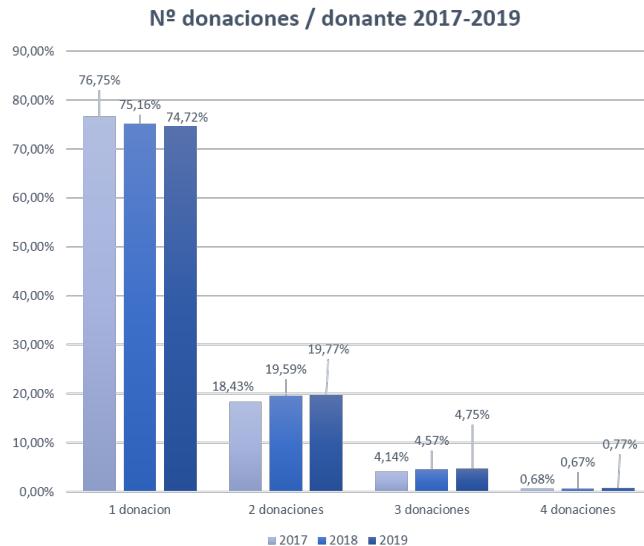


Figura 1.1: Relación entre donaciones y donantes 2017-2019 [8].

Por otra parte, esto supone unas pérdidas socio-económicas notables, ya que según Cruz Roja³, el importe de una bolsa de sangre es de 67,75€. A esto hay que añadirle los costes de los test previos a la donación, y el malgasto del material sanitario necesario para realizar el test.

1.1.3. Métodos tradicionales de medición de hemoglobina

La medición tradicional de los niveles de hemoglobina se realiza mediante métodos invasivos que requieren la extracción de una muestra de sangre. Dos de los métodos más comunes son la medición capilar y la medición venosa a través de un hemograma.

En el método capilar, se obtiene una pequeña muestra de sangre generalmente de la yema del dedo mediante una punción. Esta muestra se coloca en un dispositivo de análisis, como un glucómetro, que mide los niveles de hemoglobina presente en la muestra[9].

Por otro lado, la medición venosa se realiza a través de un hemograma, que implica la extracción de una muestra de sangre de una vena, generalmente en el brazo. Esta muestra se envía a un laboratorio donde se lleva a cabo un análisis de sangre completo, incluyendo la medición de los niveles de hemoglobina⁴.

Ambos métodos tradicionales de medición de hemoglobina son efectivos, pero tienen limitaciones significativas. Requieren la intervención de personal médico capacitado para realizar la extracción de sangre, lo que aumenta los costos y la complejidad del proceso. Además, estas mediciones solo se pueden realizar en entornos clínicos o de laboratorio, lo que dificulta la monitorización ambulatoria y frecuente de los niveles de hemoglobina en los donantes de sangre.

³Cruz Roja:<https://acortar.link/JKH1by>. Consultado el 26 de enero de 2024

⁴Dra. Gemma Cardona:

<https://www.saludsavia.com/contenidos-salud/articulos-especializados/que-detecta-un-hemograma-y-como-interpretarlo>
Consultado el 4 de abril de 2023

1.1.4. Métodos de prevención

A partir de todas estas pérdidas e inconvenientes, se plantea la pregunta de si existe la posibilidad de detectar los niveles de hemoglobina en la sangre de una persona sin necesidad de utilizar métodos invasivos como los que se utilizan a día de hoy. En el estado del arte, se puede comprobar que existe la posibilidad de medir el nivel de hemoglobina a partir de señales fotopletismografía (PPG)[10].

Uno de los métodos que se ha propuesto consiste en recoger datos de la yema del dedo mediante un diodo emisor de luz infrarroja cercana (NIR-LED) con la cámara de un smartphone [11]. Procesando el vídeo, se genera la señal PPG aplicando varios métodos de filtrado. A continuación, se extraen características de la señal PPG y sus derivadas, se realiza un análisis de Fourier y se utiliza un algoritmo genético para seleccionar las características más relevantes y descartar las redundantes e irrelevantes para medir la hemoglobina o la glucosa.[11].

Otro método propuesto consiste en la obtención de señales PPG a partir de una tarjeta de adquisición de datos PPG basada en ARM STM32F103xC. Se extrajeron 40 características a partir de estas señales, se midió el nivel de hemoglobina utilizando el dispositivo "Hemocue Hb201TM", y se utilizaron diferentes técnicas de regresión de aprendizaje automático junto con métodos de selección de características para predecir el nivel de hemoglobina[10].

1.1.5. Problemas y desafíos detectados

La detección no invasiva de los niveles de hemoglobina en la sangre es un desafío importante para mejorar la eficacia y la comodidad de las donaciones. Actualmente, los métodos tradicionales requieren pinchazos y personal especializado, lo cual no permite una estimación ambulatoria y conlleva costos significativos. El control preciso de la hemoglobina en donantes anémicos entre donaciones es de vital importancia.

Para abordar este problema, se han propuesto métodos basados en señales PPG, que permiten medir la hemoglobina de forma no invasiva. Estos métodos involucran la extracción de características de la señal PPG, el uso de algoritmos de selección de características, y técnicas de regresión de aprendizaje automático para predecir los niveles de hemoglobina.

Estas soluciones podrían eliminar la necesidad de pinchazos y permitir un seguimiento más conveniente de los donantes anémicos, mejorando así la eficacia de las donaciones de sangre.

1.2. Objetivos

En este proyecto se pretende diseñar e implementar un sistema de aprendizaje automático el cual, a partir de señales PPG, sea capaz de predecir el nivel de hemoglobina del paciente. Estas serán señales obtenidas a partir de un dispositivo portátil de bajo coste diseñado por miembros del equipo de investigación en el que se desarrolla este trabajo, y el cual se usa en un proyecto de investigación del Centro Regional de Hemodonación de Murcia (CRHM)[12].

El dispositivo se encuentra dentro de un estudio clínico del proyecto, el cual está financiado por la convocatoria “Apoyo a la Investigación 2022, ETSIT (UPM)” (comité de ética

“DATOS-20220715-PJMM”).

Las metas principales a alcanzar con el desarrollo de este trabajo son los siguientes (ver Figura 1.2):

- Diseño de una metodología para el procesamiento y modelado predictivo de hemoglobina capilar y venosa a partir de señales PPG.
- Generación y selección de características para su posterior uso en el modelo de aprendizaje automático.
- Análisis de la metodología y selección de modelo de aprendizaje automático, en particular se probará la Regresión Lineal, Random Forest y Support Vector Regression
- Diseño e implementación del modelo de aprendizaje automático, utilizando las características previamente obtenidas como parámetro de entrada.
- Ajuste de los resultados obtenidos a partir del modelo de red neuronal, evaluando la efectividad de la predicción y la precisión de los resultados.

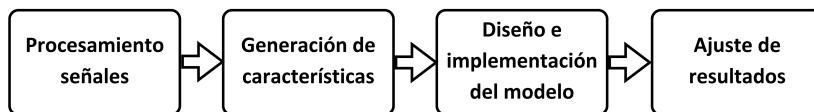


Figura 1.2: Objetivos del proyecto

1.3. Procedimiento del estudio previo a la donación

En el CRHM y en colaboración con la Universidad Politécnica de Madrid, se está llevando a cabo un estudio para desarrollar un modelo de monitorización y predicción de los niveles de hemoglobina en los donantes de sangre. El propósito principal es optimizar el sistema de la cadena transfusional y mejorar la seguridad y comodidad de los donantes. El estudio consta de dos fases: una fase inicial de dos meses y una segunda fase de cinco meses, separadas por un periodo de tres meses.

En la primera fase, los participantes acuden al CRHM en cuatro ocasiones. Durante la primera visita, se realizan las donaciones habituales y se recogen datos como los niveles de hemoglobina capilar y venosa, además de una medida de la señal de fotopletismografía y temperatura en el dedo. En las visitas posteriores, que tienen lugar cada dos semanas, se recopilan nuevamente los niveles de hemoglobina capilar y venosa, junto con las mediciones de la señal de fotopletismografía y temperatura.

En la segunda fase, a los participantes se les entrega un nodo de monitorización desarrollado por el equipo de investigación, junto con una aplicación móvil. Durante un periodo de diez semanas, realizan registros semanales de la señal de fotopletismografía y temperatura en el dedo desde la comodidad de sus hogares, generando un total de 20 registros. Además, se programan tres visitas al CRHM para medir los niveles de hemoglobina capilar y venosa.

Se espera que los resultados de este estudio proporcionen información valiosa para mejorar la calidad de la atención a los donantes de sangre, así como para ampliar el conocimiento

de los parámetros de calidad de vida en la población.

2. Estado del arte

2.1. Introducción al estado del arte

El propósito de esta sección es proporcionar una visión completa y actualizada del estado del arte en la predicción del nivel de hemoglobina y el procesamiento de señales de fotopletismografía. Esta sección juega un papel fundamental en el desarrollo de este trabajo, ya que proporciona un marco teórico sólido y una comprensión profunda de los avances científicos y tecnológicos en el campo.

A través de un análisis exhaustivo de los métodos convencionales de medición de hemoglobina, así como de estudios previos sobre la predicción de este parámetro, exploraremos las bases teóricas y técnicas utilizadas en el procesamiento de señales de fotopletismografía. Esto nos permitirá comprender cómo se han abordado hasta ahora los desafíos de la predicción de la hemoglobina y qué enfoques se han mostrado prometedores.

Además, investigaremos la aplicación de técnicas de aprendizaje automático a la predicción de hemoglobina, incluyendo un análisis de trabajos previos utilizando redes neuronales u otros algoritmos de aprendizaje automático. Esta revisión nos permitirá comprender el estado actual de la investigación y cómo podemos contribuir con propuestas innovadoras y efectivas.

2.2. Fundamentos teóricos

La hemoglobina generalmente se mide analizando una muestra de sangre extraída del cuerpo. Sin embargo, este método de medición causa inconvenientes y angustia a los pacientes, especialmente a aquellos que requieren mediciones continuas debido a condiciones tales como oligocitemia y anemia [10].

La hemoglobina juega un papel vital en la homeostasis del cuerpo, ya que los niveles adecuados son esenciales para garantizar el suministro adecuado de oxígeno a los tejidos. Los niveles bajos de hemoglobina, conocidos como anemia, pueden afectar negativamente la capacidad del cuerpo para funcionar correctamente, causando fatiga, debilidad y dificultad para respirar [13].

Además de su importancia clínica, el procesamiento de señales de fotopletismografía juega un papel crucial en la estimación no invasiva de los niveles de hemoglobina. La fotopletismografía se basa en la detección de cambios en la absorción de luz por parte de los tejidos que contienen hemoglobina, lo que proporciona información valiosa sobre la concentración de este componente sanguíneo [14].

Uno de los dispositivos más comunes para la medición del pulso cardíaco mediante señales

PPG es el MAX30102. Se trata de un pulsioxímetro y sensor de frecuencia cardíaca de alta sensibilidad que emite dos diodos emisores de luz (LED), uno infrarrojo (IR) con una longitud de onda máxima de 880 nm y otro rojo con una longitud de onda máxima de 660 nm (ver Figura 2.1)[15].

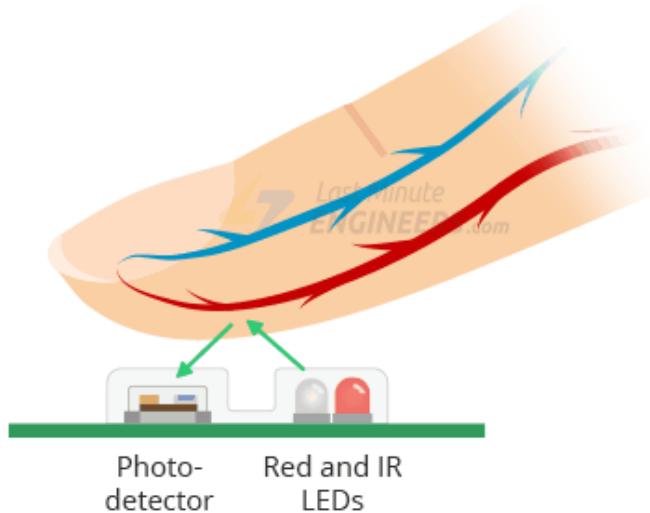


Figura 2.1: Funcionamiento de un sistema de fotopletismografía basado en el sensor MAX30102

Los sensores ópticos para la medición de la señal PPG pueden funcionar en modo de transmisión o de reflexión. Ambos tipos de sensores constan de dos partes: un transmisor (fuente de luz) y un receptor (fotodetector). La disposición de la(s) fuente(s) de luz, compuesta(s) por uno o varios diodos emisores de luz infrarroja (IF/RD LED) y un fotodetector (PHD), depende del modo de funcionamiento, tal y como se documenta en el esquema principal de la Figura 2.2 [16] .

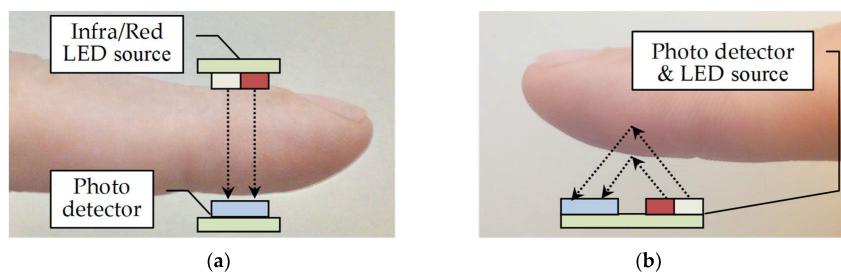


Figura 2.2: Principio básico de los sensores PPG funcionando en: (a) transmisión, (b) modos de reflexión [15]

Estas señales PPG son exactas y precisas, pero resulta incómodo para los usuarios medir la señal PPG en puntos como el dedo o el lóbulo de la oreja. Al medir la PPG en el dedo, hay muchos movimientos y actividades diarias. En el caso lóbulo de la oreja, el dispositivo para medir la PPG debe ser pequeño y ligero [17]. Esto provoca que las señales tengan mucho ruido y sea más complejo analizarlas y trabajar con ellas.

El procesamiento adecuado de estas señales, combinado con algoritmos de aprendizaje automático, puede predecir de forma precisa y no invasiva los niveles de hemoglobina. Comprender la base teórica asociada con el procesamiento de señales de fotopletismografía

es fundamental para optimizar algoritmos y técnicas para predecir la hemoglobina en la sangre.

2.3. Métodos convencionales de medición de la hemoglobina

Los métodos convencionales para medir los niveles de hemoglobina en la sangre incluyen pruebas de laboratorio y métodos invasivos. Las pruebas de laboratorio proporcionan mediciones precisas y confiables, pero son caras, no son en tiempo real, tienen un alto riesgo de infección y necesitan técnicas especiales[18].

Además del uso de los recursos de laboratorio tradicionales, el diagnóstico de la anemia también puede realizarse mediante la evaluación de la concentración de hemoglobina con dispositivos de análisis en el punto de atención, como los sistemas de análisis HemoCue (ver Figura 2.3). En varias situaciones estos dispositivos podrían sustituir a las pruebas de laboratorio tradicionales, por ejemplo en varios ámbitos de la atención sanitaria en los que puede ser necesaria una medición muy rápida de la Hb para tomar decisiones terapéuticas inmediatas[19].

Dado que el sistema HemoCue utiliza muestras de sangre sin diluir y puede ser realizado de forma fiable por trabajadores no cualificados tras una instrucción muy breve, inicialmente se pensó para su uso en departamentos ambulatorios, consultas de médicos generalistas, y también se ha descubierto que es eminentemente adecuado para la selección de donantes para transfusiones de sangre [20].

Sin embargo, el uso de estos dispositivos debe cumplir algunos criterios básicos, como cuestiones económicas, clínicas y normativas; conocimiento de los requisitos, el rendimiento, las limitaciones y las posibles interferencias de las pruebas; el uso de muestras venosas y arteriales, cuando sea posible; y una evaluación rigurosa de la calidad, que debe estar bajo la responsabilidad de profesionales de laboratorio[19].

Además, este dispositivo no tiene una fiabilidad absoluta, ya que utilizando muestras de sangre capilar para comparar HemoCue y el análisis automatizado de hemoglobina se obtuvieron unos resultados de discrepancias de más de 1 g/dl en el 21 % de los casos y de más de 2 g/dl en el 4 % de los casos [21]. Estos dispositivos de análisis en el punto de atención tienen un precio aproximado de entre 1000 y 2000 euros.¹

¹Fitness y Bienestar:
<https://www.fitnessybienestar.com/analizador-de-hemoglobina-hemo-control/>. Consultado el 24 de marzo de 2023



Figura 2.3: Dispositivo de medida de hemoglobina capilar HemoCue

2.4. Investigaciones previas sobre la predicción de la hemoglobina

La siguiente sección presentará y examinará estudios científicos e investigaciones anteriores que se han concentrado en pronosticar los niveles de hemoglobina en la sangre. Cada estudio se describirá de manera concisa, destacando el enfoque, la metodología y el tamaño de la muestra utilizada. Se enfatizarán los avances y descubrimientos notables de cada investigación.

En primer lugar, se encuentra el estudio "Non-invasive prediction of hemoglobin level using machine learning techniques with the PPG signal's characteristics features", el cual se centró en proponer un método no invasivo para la predicción de la hemoglobina utilizando los rasgos característicos de las señales PPG y diferentes algoritmos de aprendizaje automático[10].

Como datos de entrada se utilizaron las señales PPG de 33 personas en 10 períodos y se extrajeron 40 rasgos característicos de la señal PPG, de la primera derivada, y de la segunda derivada de cada señal (ver Figura 2.4); además de estos rasgos, también se añadió información sobre su sexo (masculino o femenino), altura (en cm), peso (en kg) y edad, y el número de rasgos final aumentó hasta 44[10].

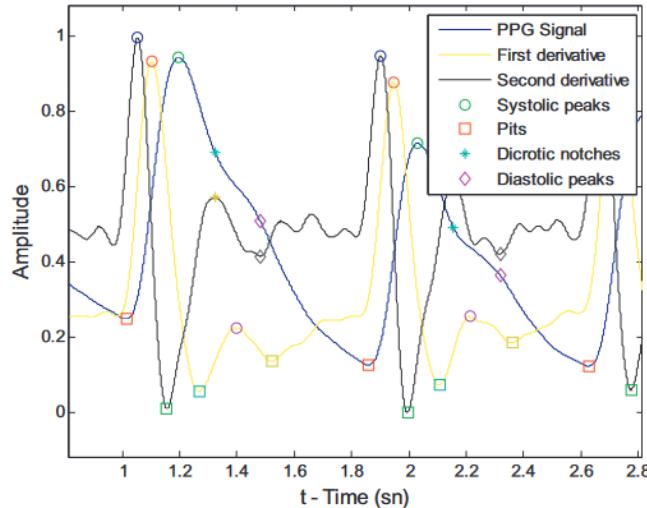


Figura 2.4: La señal PPG, las derivadas primera y segunda de la señal PPG y los puntos característicos

En este artículo se utilizaron dos métodos de selección de características para reducir el número de características en el conjunto de datos PPG: RELIEF feature selection (RFS) y correlation based feature selection (CFS). Tras aplicar los métodos RBS y CFS al conjunto de datos con 44 características, el número de características final fue de 10 y 11, respectivamente[10].

Los algoritmos de aprendizaje automático utilizados fueron los siguientes: CART (classification and regression trees), LSR (Least Square Regression), GLR (generalized linear regression), MVLR (multivariate linear regression), PLSR (partial least squares regression), GRNN (generalized regression neural network), MLP (multilayer perceptron), and SVR (support vector regression)[10].

Finalmente, se llegó a la conclusión de que SVR fue el mejor método de aprendizaje automático utilizando el criterio del error cuadrático medio en la predicción del nivel de hemoglobina.

Sin embargo, es importante mencionar que este proyecto presenta limitaciones significativas que deben tenerse en cuenta. Por un lado, se tuvo un número reducido de donantes de datos, lo que podría afectar la representatividad y generalización de los resultados obtenidos. Además, la falta de información acerca de las características utilizadas limita la posibilidad de replicar el modelo propuesto. Además, los pacientes cuya información utilizaron como datos de entrada, no son pacientes reales, sino pacientes ficticios. Por lo tanto, se requiere una investigación adicional para mejorar la fiabilidad y aplicabilidad de este enfoque en la predicción no invasiva de los niveles de hemoglobina.

En segundo lugar, encontramos el estudio "Hemoglobin and glucose level estimation from PPG characteristics features of fingertip video using MGGP-based model", cuyo objetivo fue desarrollar un modelo no invasivo de estimación de los niveles de hemoglobina y glucosa basado en la programación genética multigénica (MGGP) utilizando los rasgos característicos del fotopletismograma extraídos del vídeo de la yema del dedo capturado por un smartphone[11].

Los vídeos generados por el smartphone se procesaron para generar la señal PPG. Analizando la señal PPG, su primera y segunda derivadas (ver Figuras 2.5 y 2.6) , y aplicando

el análisis de Fourier, se extrajeron un total de 46 características. Además, la edad y el sexo también se incluyeron en el conjunto de características[11].

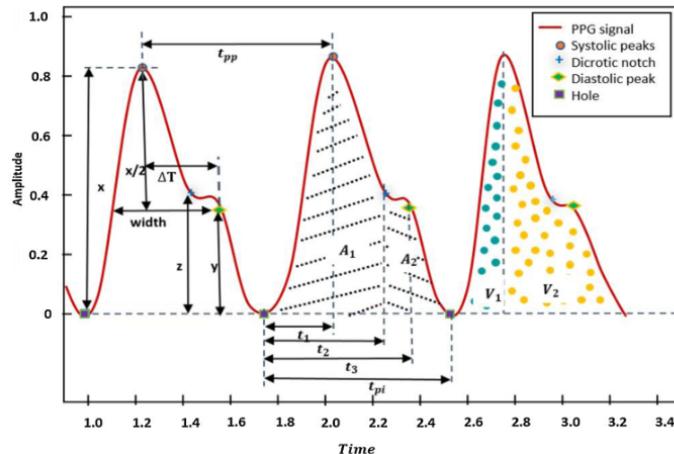


Figura 2.5: Rasgos característicos adquiridos a partir de la señal PPG

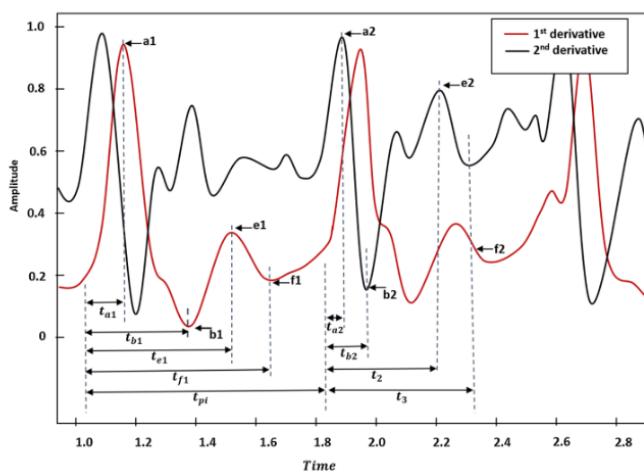


Figura 2.6: Rasgos característicos adquiridos a partir de la primera y la segunda derivada

En este trabajo, se aplica una selección de características basada en la correlación (CFS) utilizando un algoritmo genético[22]. Inicialmente, selecciona un subconjunto de características al azar como población inicial (un conjunto de posibles soluciones) y calcula la aptitud mediante una función objetivo que mide la idoneidad de las características. Según el algoritmo genético, se realizan una serie de operaciones genéticas sobre la población para generar una nueva población. Finalmente se obtuvieron 22 características de las 48 (46 + 2) iniciales[11].

Posteriormente se empleó la validación cruzada quíntuple para validar los modelos. El conjunto de datos de entrada se dividió aleatoriamente en cinco subconjuntos del mismo tamaño y en cada ronda se utilizó un único subconjunto como datos de validación para validar el modelo, y los cuatro subconjuntos restantes se utilizaron para entrenar el modelo. El proceso de validación cruzada se repitió 5 veces y cada subconjunto se utilizó exactamente una vez como conjunto de validación[11]. En cuanto a las conclusiones, para evaluar el rendimiento del enfoque basado en MGGP, los resultados se compararon con los modelos LR, SVR y RFR. A partir de los valores MSE, RMSE y MAE, se concluyó que el método de regresión simbólica basado en MGGP tuvo una mejor precisión de estimación ($\pm 0,304$ con R^2 , 0,807 para Hb y $\pm 0,324$ con R^2 , 0,881 para Gl) que los métodos de regresión convencionales[22].

Es importante tener en cuenta que este estudio no se ha replicado en otros entornos o poblaciones, lo que podría afectar la validez externa de los hallazgos. Además, este estudio no proporciona información acerca de las características utilizadas, por lo que se limita la posibilidad de replicar el modelo propuesto. Sin embargo, es relevante destacar que se utilizaron tanto la primera como la segunda derivada de la señal PPG para obtener los parámetros característicos, lo cual puede resultar útil en futuras investigaciones y es un aspecto a considerar al replicar el enfoque en otros contextos.

Por último, cabe destacar el estudio "Prediction of the hemoglobin level in hemodialysis patients using machine learning techniques". El objetivo de este estudio fue predecir el valor de hemoglobina (Hb) utilizando una base de datos de pacientes europeos en hemodiálisis proporcionada por Fresenius Medical Care (FMC) [23].

En Italia, se incluyó a 2,764 pacientes, mientras que en España se contó con 10,247 pacientes para este estudio. Con el objetivo de desarrollar modelos de predicción de hemoglobina, se utilizó la técnica de validación cruzada. Además, se aplicaron criterios adicionales como el establecimiento de un número mínimo de registros y la exclusión de valores atípicos [23].

Para abordar el problema de manera local y global, se emplearon diferentes modelos en este estudio. En el enfoque local, se utilizaron modelos de regresión lineal múltiple y árboles de regresión en cada uno de los clusters obtenidos mediante clustering jerárquico y ART. Estos modelos permitieron establecer relaciones entre las variables explicativas y la variable de respuesta en cada cluster específico. También se implementaron modelos de predicción globales, como el bagging, que combina múltiples modelos aplicados a muestras aleatorias de datos para mejorar la precisión de las predicciones. Otros modelos utilizados incluyeron el perceptrón multicapa (MLP) y las máquinas de vectores de soporte (SVM) [23].

En general, se observó que no había un modelo que superara claramente a los demás en términos de rendimiento. Los errores de predicción en los conjuntos de prueba rondaban los 0.6 g/dl. Aunque se esperaba que los modelos locales presentaran una mejora significativa en el error cometido, esta suposición fue refutada. En conclusión, se determinó que se había alcanzado el máximo rendimiento en el conjunto de datos utilizado y no se identificó un modelo sobresaliente en comparación con los demás [23].

Aunque se obtuvieron resultados prometedores, es importante destacar que se podrían encontrar dificultades al intentar replicar este estudio, especialmente en la obtención de una base de datos similar. Además, la falta de un modelo dominante en términos de rendimiento da pie a explorar nuevas técnicas o enfoques para mejorar las predicciones.

2.5. Brecha identificada y justificación del proyecto

En esta sección se hablará de las brechas y las limitaciones encontradas en los estudios y publicaciones mencionados anteriormente, para así ver de qué manera se pueden lograr avances y obtener mejores resultados en la predicción del nivel de hemoglobina.

Tras estudiar y revisar el estado del arte de la predicción del nivel de hemoglobina a partir del procesamiento de señales PPG, se pretenden replicar las técnicas y los modelos utilizados en los diferentes estudios para lograr unos resultados similares en el laboratorio del CRHM.

Una de las grandes limitaciones que se han encontrado a la hora de intentar replicar los proyectos de investigación es la falta de información acerca de los tipos de redes utilizados y de los datos usados.

Por otra parte, los estudios se realizaron con una pequeña cantidad de pacientes, por lo que los resultados podrían ser poco extrapolables a una situación de predicción con pacientes reales. Además, los pacientes cuya información utilizaron como datos de entrada, no son pacientes reales, sino pacientes ficticios.

Es por esto que en este proyecto se pretende superar estas limitaciones y brechas identificadas. En primer lugar, se buscará recopilar información detallada sobre los posibles tipos de redes a utilizar, a fin de asegurar una correcta replicación de las técnicas y modelos.

Además, se ampliará la muestra de pacientes involucrados en el estudio, utilizando los datos de los pacientes proporcionados por el CRHM. Esto permitirá obtener resultados más confiables y generalizables a la población en general.

3. Metodología

Este capítulo establece la importancia de desarrollar una metodología adecuada para predecir el nivel de hemoglobina en sangre y otros parámetros sanguíneos en el marco del proyecto. Esta metodología es crucial para optimizar la gestión de recursos sanguíneos y facilitar la detección temprana de enfermedades. A través de este capítulo, se sientan las bases para el desarrollo de la metodología en detalle.

3.1. Obtención de señales PPG

Para poder comenzar a desarrollar el sistema de aprendizaje automático, es necesario disponer de unos datos de entrada para poder entrenar el modelo que desarrollaremos. Estos datos de entrada son proporcionados por el CRHM, el cual, con el dispositivo desarrollado por el equipo de investigación durante el proyecto “Diseño e implementación de un sistema de monitorización ambulatoria y no intrusiva de hemoglobina” [12] (ver Figura 3.1), han obtenido 73 señales de 40 donantes distintos (señales tanto rojas como infrarrojas, como temperatura).

Además, se recopilaron otros datos como referencia, que incluyen mediciones de SpO₂ (saturación de oxígeno en sangre), presión sistólica y niveles de hemoglobina. La información fue obtenida mediante diferentes dispositivos, como pulsioxímetros y tensiómetro. Estos datos adicionales proporcionan una amplia gama de información relevante para el desarrollo del sistema de aprendizaje automático.

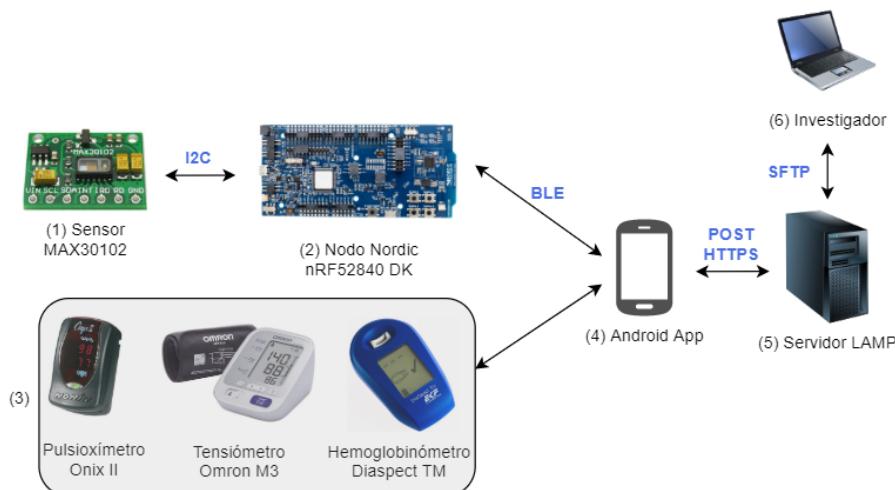


Figura 3.1: Arquitectura del entorno completo, con los dispositivos hardware utilizados y los protocolos de comunicación establecidos entre ellos [11].

Las señales se han obtenido a partir de un estudio en el que el Centro Regional de Hemodonación de Murcia (CRHM) y la Universidad Politécnica de Madrid buscan desarrollar

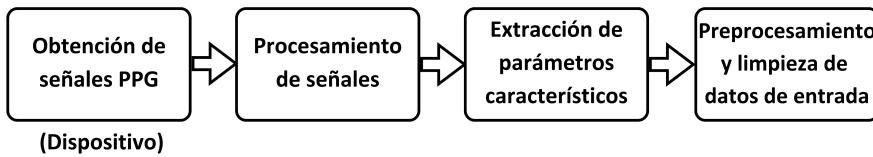


Figura 3.2: Proceso de obtención de señales PPG

un modelo de monitorización y predicción de los niveles de hemoglobina en los donantes de sangre. Posteriormente estas señales pasan por un flujo de procesamiento y extracción de características (ver Figura 3.2). El estudio consta de dos fases: una inicial de dos meses y una segunda de cinco meses, con recopilación de datos en visitas al centro y registros semanales desde el hogar. En este trabajo se centra en la primera fase.

Las señales de fotopletismografía se procesarán para obtener características de las mismas, que se utilizarán como entrada para el modelo de aprendizaje automático. El procesamiento de señales permitirá extraer información relevante y representativa de cada donante, que luego se utilizará para entrenar y ajustar modelos predictivos.

Este conjunto de parámetros característicos incluirá información relacionada con la señal de fotopletismografía, temperatura distal, así como datos adicionales como sexo, altura, peso y edad de cada paciente. Estos parámetros se utilizarán como variables independientes en el modelo y su combinación permitirá una predicción de los niveles de hemoglobina tanto capilar como venosa y tal vez otros parámetros sanguíneos en la sangre.

Es necesario enfatizar que la calidad de los datos de entrada juega un papel fundamental en la efectividad de los modelos de aprendizaje automático. Por ello, se realizará el preprocesamiento y limpieza de datos para eliminar datos anómalos, con ruido o incompletos que puedan afectar negativamente a la precisión del modelo.

3.2. Procesamiento de señales PPG

En este apartado se hablará del procesamiento de las señales para la obtención de los parámetros característicos. En primer lugar, se cargan las señales previamente mencionadas, de las cuales tenemos una señal infrarroja, una roja y la temperatura, y comienza el proceso de preprocesamiento.

Para comenzar con el procesamiento, las señales rojas e infrarrojas se invierten restando el valor máximo, lo que ayuda a normalizar las señales. A continuación, se calcula la media móvil para eliminar la componente continua (DC) utilizando una ventana de 1 segundo y un solapamiento del 50 %, como se puede observar en la Figura 3.3. Este paso permite eliminar la componente de continua no deseada de las señales. Por último, se resta la media móvil a las señales invertidas. Estos pasos son esenciales para preparar las señales PPG antes de realizar análisis y detección de características relevantes.

Posteriormente, se corrige un pequeño desplazamiento para lograr que las señales estén centradas alrededor de cero. Esto se logra restando el promedio de las señales originales 'red' e 'ir' a sus respectivas componentes de continua. Esta corrección facilita el posterior procesamiento de la señal para la detección de picos.

A continuación, se realiza un filtrado paso bajo en las señales procesadas 'red' e 'ir'.

3. Subtracting DC signal

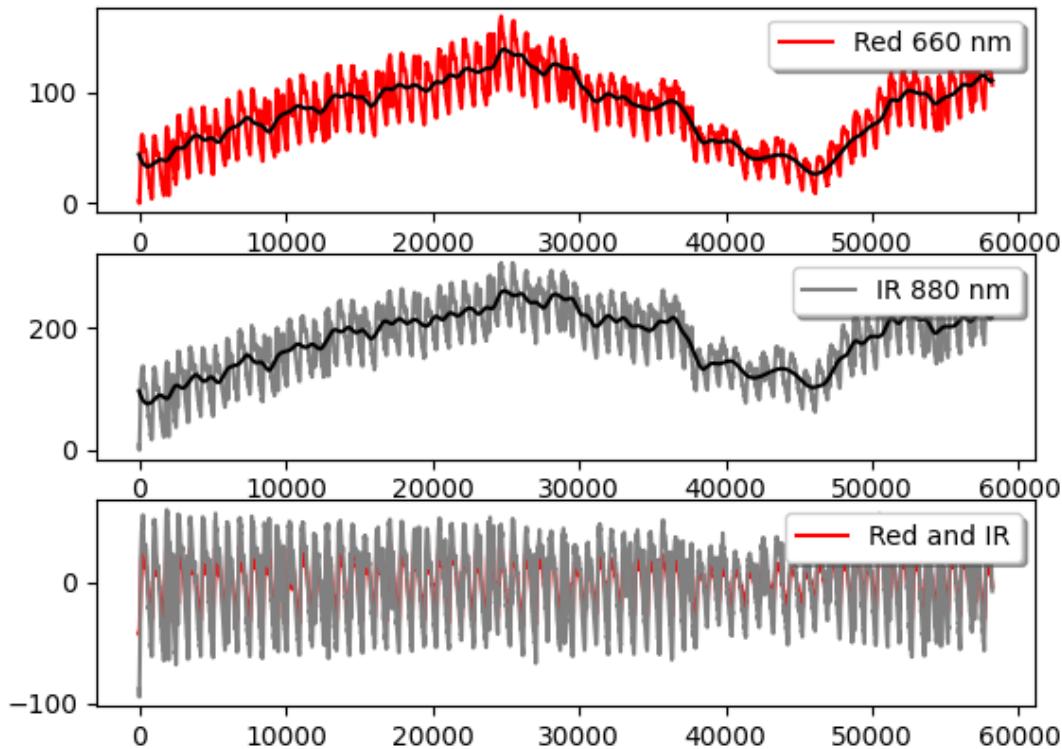


Figura 3.3: Señales tras la eliminación de la componente continua

Este filtro elimina las frecuencias altas no deseadas y retiene las componentes de baja frecuencia, lo que suaviza la señal y ayuda a resaltar las características relevantes para el análisis (ver Figura 3.4).

Finalmente, se aplica el algoritmo de spectrograma a las señales filtradas. El spectrograma proporciona una representación visual de la distribución de energía de la señal en función del tiempo y la frecuencia. Esto permite identificar patrones y cambios en la señal a lo largo del tiempo, lo cual puede ser útil para el análisis y la extracción de características.

En el siguiente paso, se procede a calcular los índices de calidad de las señales PPG procesadas. Para ello, se suma el componente de corriente continua (DC) previamente calculado a las señales procesadas. Estas señales, junto con las señales filtradas mediante un filtro paso bajo (LP) y también sumadas con el DC, son utilizadas para calcular diversas métricas de calidad, tales como Perfusión, Asimetría (Skewness), Curtosis (Kurtosis) y Relación Señal-Ruido (SNR) [24].

Estas métricas son de gran importancia, ya que proporcionan una evaluación objetiva de la calidad de las señales PPG procesadas. La métrica de Perfusión permite evaluar el grado de variabilidad en la amplitud de las señales, lo cual es crucial para detectar cambios en el flujo sanguíneo. La Asimetría y Curtosis son medidas estadísticas que describen la simetría y forma de la distribución de las señales, permitiendo detectar posibles distorsiones o anomalías. Por último, la Relación Señal-Ruido (SNR) indica la relación entre la potencia de la señal y el nivel de ruido presente, siendo un indicador clave para evaluar la calidad

4. Low pass filtered

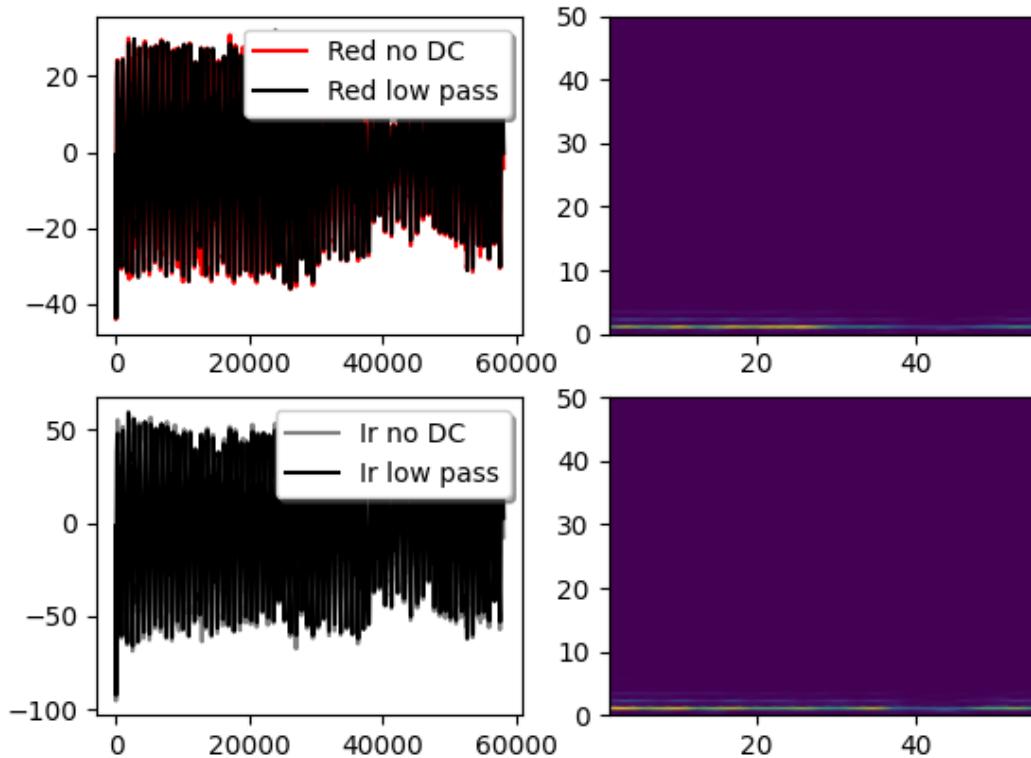


Figura 3.4: Filtrado paso bajo

general de la señal procesada.

Estos índices de calidad son utilizados para realizar una selección de las señales que cumplen con los estándares deseados. Los índices de calidad proporcionan una medida objetiva de la fiabilidad de las señales, lo que permite identificar aquellas que son más adecuadas para el análisis y la extracción de información relevante

Una vez hemos calculado los parámetros de calidad, detectamos los picos máximos y mínimos de las señales rojas e infrarrojas, como se muestra en la Figura 3.5. Ahora que hemos calculado los picos de las señales, podemos calcular los intervalos entre latidos (IBI) tanto para los picos máximos como los mínimos detectados en las señales rojas e infrarrojas. Estos intervalos permiten estimar el tiempo transcurrido entre cada latido cardíaco. A partir de los IBI, se calcula la frecuencia cardíaca (HR) en latidos por minuto, tanto para los picos máximos como los mínimos. Luego se calcula la media y la desviación estándar de las frecuencias cardíacas para las señales rojas e infrarrojas, y se obtiene un promedio de ambas frecuencias cardíacas.

Para obtener otra característica como es la frecuencia respiratoria (BR), se han utilizado técnicas similares, donde se sustrae la componente de corriente continua (DC) de la envolvente promedio y se detectan los picos correspondientes a la respiración (ver Figura 3.6).

Ahora que ya hemos obtenido todos estos datos, procederemos a calcular varias características (features) a partir de estos. En primer lugar, se calculan la media (mean) y la desviación estándar (std) de las componentes de corriente continua (DC) tanto para la

5. Peak detection

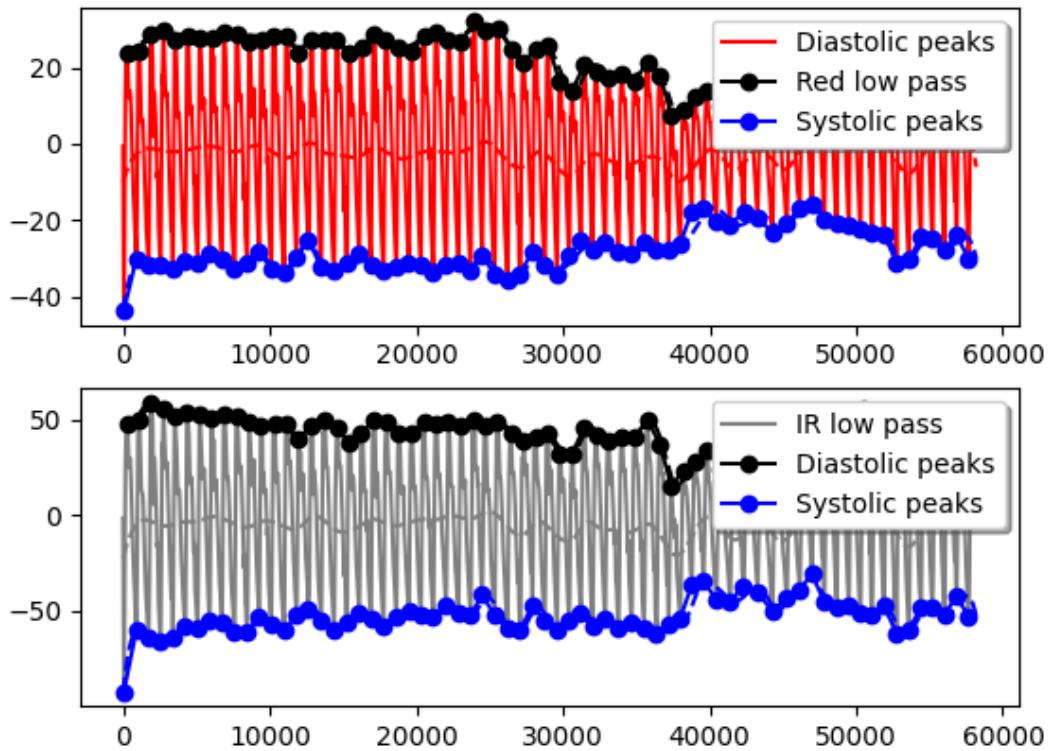


Figura 3.5: Detección de picos y creación de envolventes

señal roja como para la señal infrarroja. Estas medidas proporcionan información sobre los niveles medios y la variabilidad de la corriente continua en las señales PPG, lo cual es relevante para el análisis y la interpretación de las características cardíacas.

A continuación, se calcula la media y la desviación estándar de los intervalos entre latidos (IBI) tanto para la señal roja como para la señal infrarroja. Además, se calcula el promedio de los IBI de ambas señales. Estos valores son indicadores temporales importantes relacionados con la frecuencia cardíaca y el ritmo cardíaco.

Lo siguiente que haremos será obtener el dato del SpO₂ calculando la relación entre las señales PPG roja e infrarroja mediante la obtención del valor RMS (Root Mean Square) de cada señal. El SpO₂ se calcula mediante una fórmula empírica (ver Figuras 3.7 y 3.8), en la cual se utiliza la relación R multiplicada por un factor de escala y se resta de un valor de referencia. La relación R se define con la fórmula de la Figura 3.8 [25].

$$SpO_2 = 110 - 25 * R$$

Figura 3.7: Fórmula para el cálculo del SpO₂

12. Respiration rate

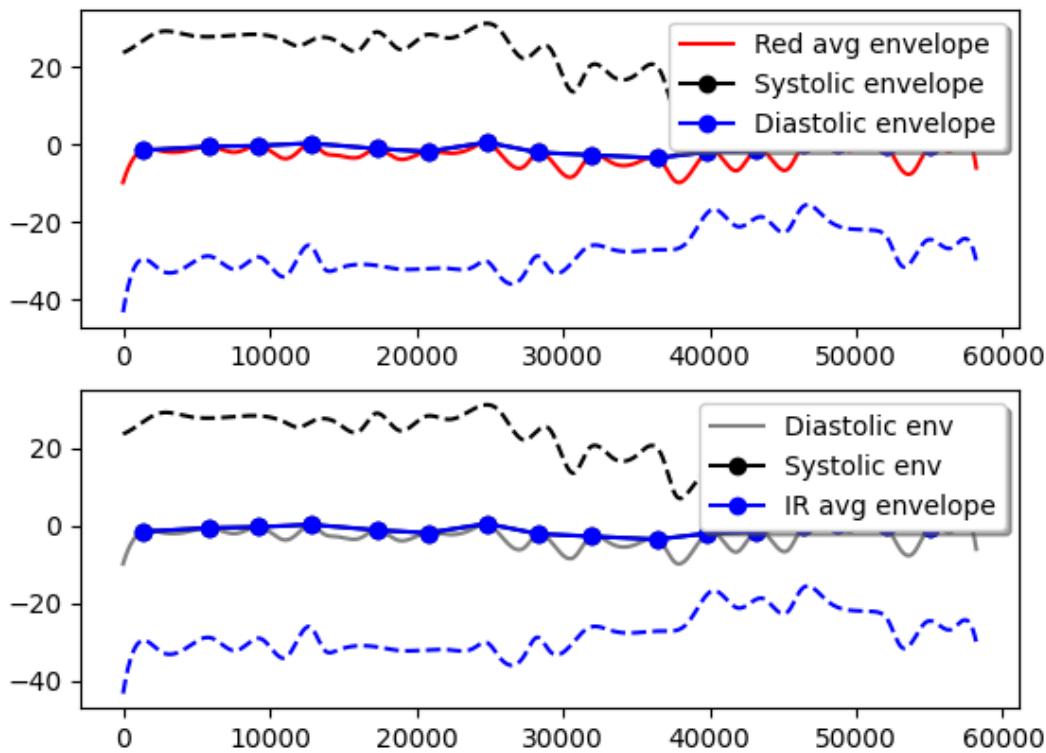


Figura 3.6: Cálculo de la frecuencia respiratoria

$$R = \frac{AC_{\text{RMS RED}}/DC_{\text{RED}}}{AC_{\text{RMS IR}}/DC_{\text{IR}}}$$

Figura 3.8: Fórmula para el cálculo de R

Posteriormente, se calcula la frecuencia respiratoria en pulsaciones por minuto (bpm) para las señales roja e infrarroja, así como el promedio de ambas frecuencias respiratorias. Luego, se calcula la amplitud promedio y la desviación estándar de las envolventes de las señales roja e infrarroja, lo cual proporciona información sobre la variación de amplitud de los pulsos cardíacos.

A continuación, se calcula el área promedio y la desviación estándar de los pulsos de las señales roja e infrarroja, lo cual refleja la información sobre la forma y duración de los pulsos. Finalmente, se calcula la variación promedio y la desviación estándar de los tiempos de subida y bajada de los pulsos para ambas señales.

El resto de parámetros característicos los obtendremos a partir de la biblioteca HeartPy¹. Se procesa la señal y se calculan varias medidas relacionadas con la frecuencia cardíaca, como las pulsaciones por minuto (bpm), la raíz cuadrada de las diferencias sucesivas (RMSSD), la desviación estándar de los intervalos RR, la proporción de diferencias su-

¹Documentación de Python Heart Rate Analysis Toolkit:
<https://python-heart-rate-analysis-toolkit.readthedocs.io/en/latest/>. Consultado el 16 de febrero de 2023

cesivas por encima de ciertos umbrales (pnn20 y pnn50), la desviación estándar de la proyección del gráfico de Poincaré (sd1 y sd2) y la frecuencia respiratoria.

Las características obtenidas a partir de la biblioteca HeartPy son de interés debido a que proporcionan información relevante sobre la frecuencia cardíaca y la variabilidad de los intervalos entre los latidos del corazón. A continuación, se enumeran algunas de las medidas calculadas:

- Pulsaciones por minuto (bpm): Indica la frecuencia cardíaca promedio en un minuto y puede ser utilizado como un indicador general del ritmo cardíaco.
- Raíz cuadrada de las diferencias sucesivas (RMSSD): Refleja la variabilidad de los intervalos RR, que representa la variabilidad de los latidos del corazón. Una mayor RMSSD sugiere una mayor variabilidad y se asocia con un sistema nervioso autónomo en equilibrio.
- Desviación estándar de los intervalos RR: Mide la dispersión de los intervalos RR y proporciona información sobre la variabilidad de la frecuencia cardíaca.
- Proporción de diferencias sucesivas por encima de ciertos umbrales (pnn20 y pnn50): Representan el porcentaje de diferencias sucesivas entre intervalos RR que exceden ciertos umbrales (20 ms y 50 ms, respectivamente). Estas medidas se utilizan para evaluar la presencia de variabilidad a corto plazo.
- Desviación estándar de la proyección del gráfico de Poincaré (sd1 y sd2): El gráfico de Poincaré es una representación gráfica de la relación entre los intervalos RR sucesivos. La desviación estándar de la proyección horizontal (sd1) y vertical (sd2) del gráfico proporciona información sobre la variabilidad de la frecuencia cardíaca a corto y largo plazo.
- Frecuencia respiratoria: Permite detectar la presencia de patrones respiratorios específicos en la señal de frecuencia cardíaca, lo cual puede ser útil para evaluar la función respiratoria y su relación con la actividad cardíaca.

Característica	Descripción
sex	Sexo
weight_kg	Peso (kg)
height_cm	Altura (cm)
spo2_ref	Referencia de SpO2
sys_mmHg	Presión sistólica (mmHg)
dia_mmHg	Presión diastólica (mmHg)
hgb_g_dL	Hemoglobina (g/dL)

Tabla 3.1: Características basales

Característica	Descripción
heart_bpm	Ritmo cardíaco en latidos por minuto (BPM)
breathing_rate	Frecuencia respiratoria en respiraciones por minuto
IBI	Intervalo entre latidos (IBI)
SPo2	Nivel de saturación de oxígeno en la sangre (SpO2)
RMSSD_red	Raíz cuadrada del promedio de las diferencias al cuadrado entre latidos consecutivos
SDNN_red	Desviación estándar de los intervalos entre latidos
SDSD_red	Desviación estándar de las diferencias entre latidos sucesivos
PNN20_red	Porcentaje de diferencias entre latidos consecutivos que superan los 20 ms
PNN50_red	Porcentaje de diferencias entre latidos consecutivos que superan los 50 ms
HR_MAD	Desviación media absoluta de los intervalos entre latidos
SD1_red	Desviación estándar del componente perpendicular de baja frecuencia del diagrama de puntos
SD2_red	Desviación estándar del componente paralelo de alta frecuencia del diagrama de puntos

Tabla 3.2: Características calculadas para la señal roja (idem para la señal infrarroja)

Con estos datos, creamos la tabla de datos, el cual almacena toda esta información, asociada a la donación correspondiente, la cual se identifica a partir de un id. Se cargan datos clínicos y basales de otros archivos CSV, asociados también a cada uno de los id, y se fusionan con los datos procesados previamente en una tabla final, el cual en un archivo CSV. Este archivo será la entrada del modelo.

3.3. Diseño del modelo

Para comenzar con el desarrollo del modelo, lo primero será importar las librerías necesarias para realizar el análisis de datos y construir los modelos de machine learning. Estas librerías incluyen herramientas para la manipulación de datos, selección de características, preprocesamiento, entrenamiento y evaluación de modelos, así como visualización de resultados.

Posteriormente, se definen una serie de métodos que serán utilizados en el código. Uno de los métodos destacados es `impute_missing_data`, el cual se encarga de imputar los valores ausentes en el conjunto de datos. Este método utiliza técnicas como `KNNImputer` e `IterativeImputer` para imputar los valores ausentes en las variables numéricas y categóricas respectivamente.

Después de definir los métodos necesarios, se carga el conjunto de datos que se utilizará en el estudio. Los datos se cargan desde un archivo CSV y se almacenan en una tabla de datos. Además, se realizan algunas transformaciones en los datos, como la sustitución de los valores erróneos (''-'') por NaN y la eliminación de columnas innecesarias.

A continuación, se realiza la imputación de los valores ausentes en el conjunto de datos

utilizando el método `impute_missing_data` previamente definido. Los valores ausentes se imputan utilizando diferentes estrategias dependiendo del tipo de variable. Los datos imputados se almacenan en nuevas tablas de datos.

El siguiente paso, será dividir el conjunto de datos en conjuntos de entrenamiento y test utilizando la función `train_test_split` de la librería scikit-learn. Esto permite evaluar el rendimiento de los modelos en datos no vistos durante el entrenamiento.

Durante el desarrollo de este proyecto, se utilizaron varios modelos para abordar el problema de predicción de valores numéricos continuos. Los modelos empleados fueron la Regresión Lineal, Random Forest, Gradient Boosting y SVR (Support Vector Regression).

La Regresión Lineal es un modelo estadístico que busca establecer una relación lineal entre una variable dependiente y una o más variables independientes. Se evaluó su rendimiento utilizando métricas como R², MAE y MSE, y se compararon las predicciones con los valores reales.

El Random Forest es un conjunto de árboles de decisión que promedia las predicciones de cada árbol. Se realizó una búsqueda de la mejor combinación de hiperparámetros mediante validación cruzada y se evaluó el rendimiento utilizando las mismas métricas.

El Gradient Boosting es un modelo de ensamblaje que entrena modelos secuenciales para corregir los errores del modelo anterior. También se ajustaron los hiperparámetros y se evaluó el rendimiento utilizando las métricas mencionadas.

El SVR, o Support Vector Regression, es un modelo de regresión basado en vectores de soporte que utiliza funciones de kernel para mapear los datos a un espacio de mayor dimensionalidad. Se ajustaron los hiperparámetros, como el tipo de kernel y la constante de regularización, y se evaluó el rendimiento utilizando las métricas mencionadas anteriormente.

3.3.1. Regresión Lineal

Una vez hemos dividido el conjunto de datos en conjunto de entrenamiento y test, se construye el modelo de regresión lineal. La regresión lineal es un modelo estadístico que busca establecer una relación lineal entre una variable dependiente y una o más variables independientes. Utiliza una función lineal para predecir valores numéricos continuos y se basa en el ajuste de los coeficientes de regresión para minimizar la diferencia entre los valores predichos y los valores reales en un conjunto de datos (ver Figura 3.9²).

²IArtificial.net:
<https://www.iartificial.net/regresion-lineal-con-ejemplos-en-python/>. Consultado el 5 de mayo de 2023

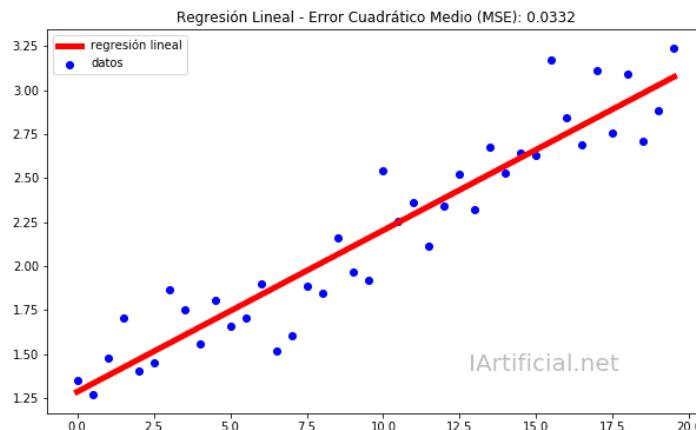


Figura 3.9: Ejemplo de regresión lineal²

Este modelo se ajusta a los datos de entrenamiento y se evalúa su rendimiento utilizando métricas como el coeficiente de determinación (R^2), el error absoluto medio (MAE) y el error cuadrático medio (MSE) en los conjuntos de entrenamiento y test. Además, se realizan visualizaciones para comparar los valores predichos con los valores reales en ambos conjuntos.

3.3.2. Random Forest

También probaremos a medir los resultados con un modelo de Random Forest. Un modelo de Random Forest es un conjunto de árboles de decisión en el que cada árbol se entrena con una muestra aleatoria del conjunto de datos y se promedia la predicción de cada árbol para obtener la predicción final (ver Figura 3.10³).

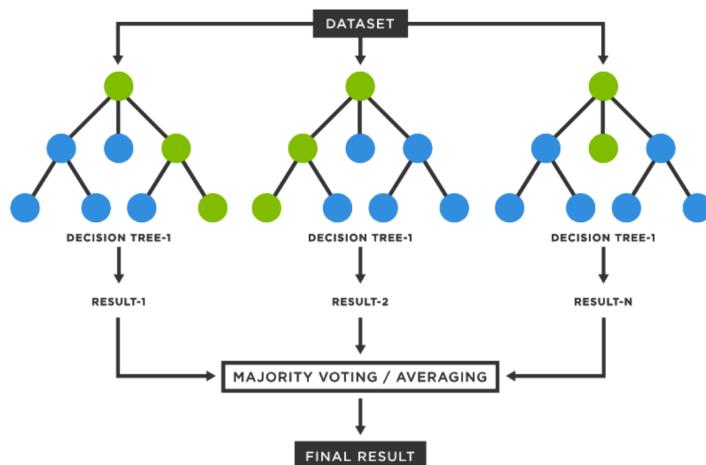


Figura 3.10: Funcionamiento de un modelo Random Forest³

En este modelo, se utilizó una selección de características basada en el MAE en el conjunto de test para identificar el mejor conjunto de características a utilizar en el modelo de Random Forest. Esta técnica ayuda a reducir la dimensionalidad y mejorar la eficiencia

³TIBCO:

<https://www.tibco.com/reference-center/what-is-a-random-forest>. Consultado el 5 de mayo de 2023

y precisión del modelo al seleccionar las características más relevantes para el problema en cuestión.

Para entrenar el modelo de Random Forest, se utilizó la técnica de división de datos en conjuntos de entrenamiento y test, conocida como train-test split. Esta técnica consiste en dividir el conjunto de datos en dos partes: una para el entrenamiento del modelo y otra para evaluar su rendimiento (ver Figura 3.11⁴).

El conjunto de entrenamiento se utilizó para ajustar los parámetros del modelo de Random Forest, mientras que el conjunto de test se reservó para evaluar su rendimiento final. Esta división garantiza que el modelo sea evaluado en datos no vistos previamente, lo que proporciona una medida más realista de su capacidad de generalización.

Cuando el modelo de Random Forest ya ha sido entrenado, se evalúa su rendimiento en el conjunto de test. Se calculan métricas como el coeficiente de determinación (R^2), el error absoluto medio (MAE) y el error cuadrático medio (MSE). En este modelo, al igual que en el de regresión lineal, se visualiza la relación entre los valores esperados y los valores predichos mediante un diagrama de dispersión.

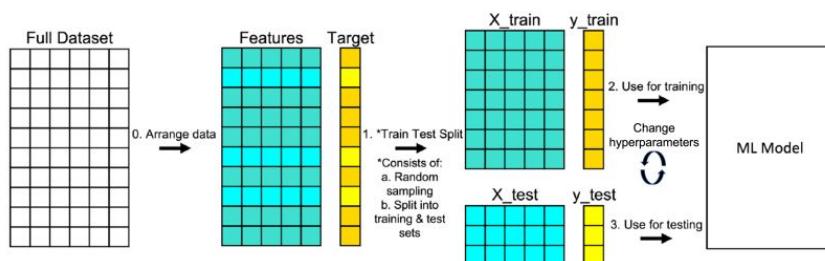


Figura 3.11: Funcionamiento del train-test split⁴

3.3.3. Gradient Boosting

Posteriormente, se construye un modelo de Gradient Boosting utilizando una metodología similar a la del modelo de Random Forest. El Gradient Boosting es una técnica de ensamblaje de modelos en la que se entrena múltiples modelos de forma secuencial, en los que cada modelo intenta corregir los errores del modelo anterior. En cada iteración, se ajusta un nuevo modelo utilizando un enfoque de “boosting” para enfocarse en los ejemplos que fueron mal clasificados por los modelos anteriores (ver Figura 3.12⁵).

⁴BuiltIn:

<https://builtin.com/data-science/train-test-split>. Consultado el 5 de mayo de 2023

⁵DataScience.eu:

<https://datascience.eu/machine-learning/gradient-boosting-what-you-need-to-know/>. Consultado el 5 de mayo de 2023

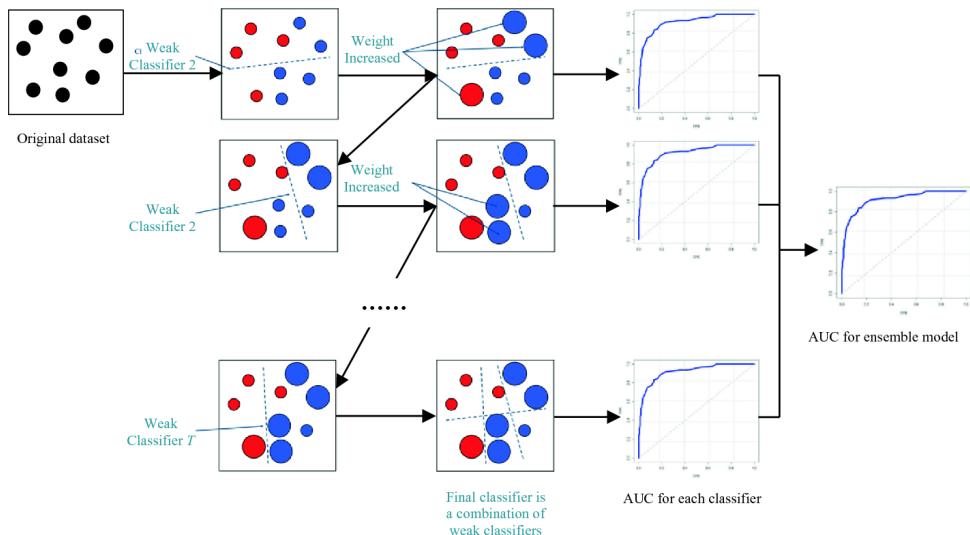


Figura 3.12: Funcionamiento de un modelo de Gradient Boosting⁵

Se llevó a cabo una selección de características basada en el MAE en el conjunto de test para identificar el mejor conjunto de características a utilizar en el modelo de Gradient Boosting.

Una vez entrenado el modelo, se evalúa su rendimiento. Al igual que con los otros modelos mencionados previamente, se calculan métricas como el coeficiente de determinación (R^2), el error absoluto medio (MAE) y el error cuadrático medio (MSE).

3.3.4. Decision Tree Regression (DTR)

En este estudio, también se utilizó el modelo de Regresión de Árbol de Decisión (DTR) para realizar las predicciones. El DTR es un algoritmo de aprendizaje supervisado utilizado para resolver problemas de regresión. A diferencia de los modelos anteriores, el DTR divide el espacio de características en regiones rectangulares y crea una estructura de árbol en la que cada nodo representa una división de características basada en una pregunta. Al llegar a las hojas del árbol, se toma el promedio de los valores de la variable objetivo en esa región como la predicción (ver Figura 3.13)⁶.

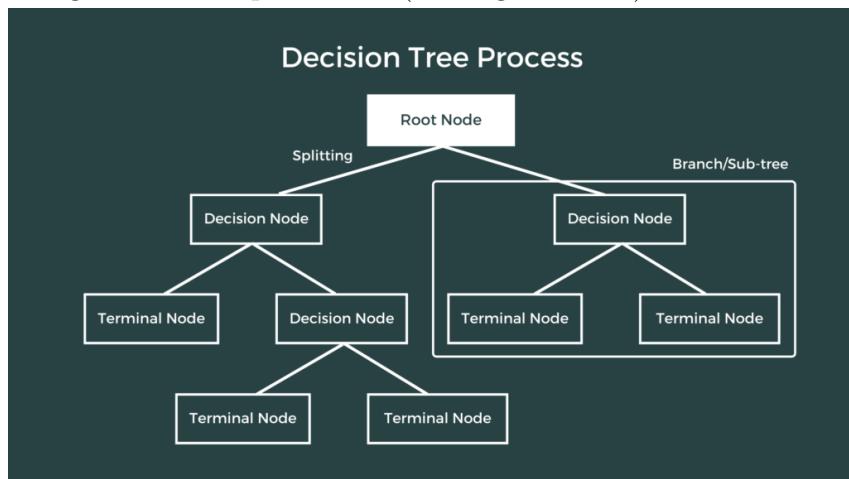


Figura 3.13: Ejemplo de Regresión de Árbol de Decisión⁶

Al igual que con los otros modelos mencionados anteriormente, se dividió el conjunto de

⁶TheClickReader: <https://www.theclickreader.com/decision-tree-regression/>

datos en conjuntos de entrenamiento y test. A continuación, se construyó y entrenó el modelo DTR utilizando los datos de entrenamiento. El DTR busca la mejor división de características en cada nodo del árbol para minimizar la suma de los cuadrados de las diferencias entre las predicciones y los valores reales. El proceso continúa hasta que se alcanza un criterio de parada predefinido.

Se llevó a cabo una selección de características basada en el MAE en el conjunto de test para identificar el mejor conjunto de características a utilizar en el modelo de Decision Tree Regression.

Una vez entrenado el modelo DTR, se evaluó su rendimiento utilizando métricas como el coeficiente de determinación (R^2), el error absoluto medio (MAE) y el error cuadrático medio (MSE) en los conjuntos de entrenamiento y test. Estas métricas proporcionan una medida de la calidad de las predicciones realizadas por el modelo.

Además, se utilizó una técnica de mejora conocida como “boosting” para mejorar el rendimiento del modelo DTR. El boosting es un enfoque de conjunto que combina múltiples modelos débiles para formar un modelo más fuerte y preciso. En este caso, se aplicó el algoritmo de boosting conocido como AdaBoost al DTR⁷.

El AdaBoost funciona entrenando iterativamente diferentes versiones del modelo DTR, dándole mayor importancia a los puntos de datos que fueron mal predichos en las iteraciones anteriores (ver Figura 3.14). De esta manera, el modelo se enfoca en aprender de los errores y mejorar progresivamente su capacidad de predicción. Al combinar los modelos DTR individuales, el AdaBoost genera un modelo final que es capaz de realizar predicciones más precisas y robustas.

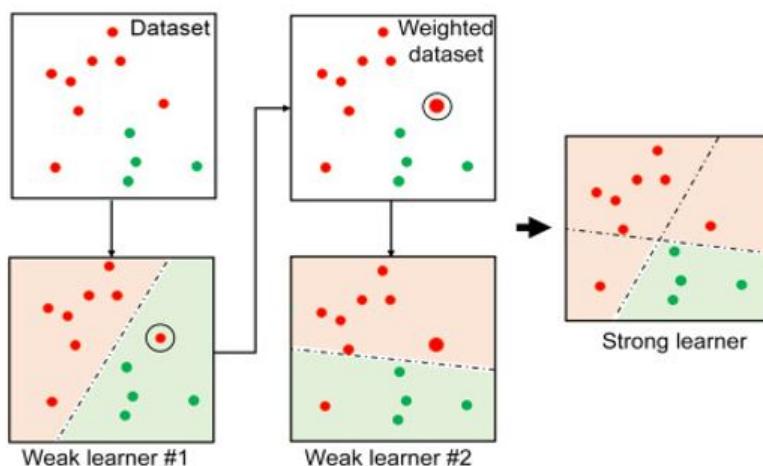


Figura 3.14: Ejemplo de modelo utilizando AdaBoost⁷

Se realizaron visualizaciones para comparar los valores predichos por el modelo DTR con los valores reales en ambos conjuntos de datos.

⁷Analytics Vidhya:
<https://www.analyticsvidhya.com/blog/2022/01/introduction-to-adaboost-for-absolute-beginners/>
 Consultado el 16 de mayo de 2023.

3.3.5. SVR

En este estudio, también se utilizó el modelo de regresión de vectores de soporte (SVR) para realizar las predicciones. El SVR es un algoritmo de aprendizaje supervisado utilizado para resolver problemas de regresión. A diferencia de la regresión lineal, el SVR busca encontrar una función no lineal que se ajuste a los datos y pueda realizar predicciones precisas (ver Figura 3.15⁸).

Al igual que con los otros modelos utilizados, se dividió el conjunto de datos en conjuntos de entrenamiento y test. A continuación, se construyó y entrenó el modelo SVR utilizando los datos de entrenamiento.

Se llevó a cabo una selección de características basada en el MAE en el conjunto de test para identificar el mejor conjunto de características a utilizar en el modelo de Support Vector Regression.

Una vez entrenado el modelo SVR, se evaluó su rendimiento utilizando métricas como el coeficiente de determinación (R^2), el error absoluto medio (MAE) y el error cuadrático medio (MSE) en los conjuntos de entrenamiento y test. Estas métricas proporcionan una medida de la calidad de las predicciones realizadas por el modelo.

Además, al igual que con los otros modelos, se realizaron visualizaciones para comparar los valores predichos por el modelo SVR con los valores reales en ambos conjuntos de datos (entrenamiento y test). Esto permitió evaluar la capacidad del modelo SVR para verificar la calidad de las predicciones realizadas.

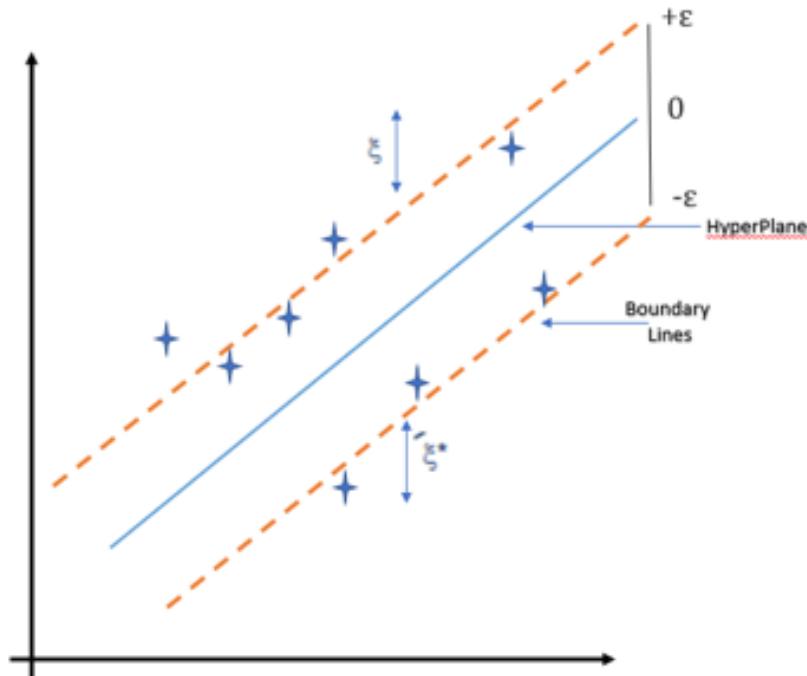


Figura 3.15: Funcionamiento de un modelo de SVR⁸

⁸EduCBA:
<https://www.educba.com/support-vector-regression/>. Consultado el 5 de mayo de 2023

4. Implementación

Este capítulo se centra en proporcionar detalles sobre el diseño y las herramientas utilizadas para abordar el problema planteado en la introducción. El objetivo es asegurar que el trabajo sea replicable y que otros investigadores puedan recrear la infraestructura descrita y obtener resultados similares utilizando las herramientas y los datos proporcionados. Este capítulo se divide en dos secciones principales: descripción de los datos utilizados y descripción de las herramientas y recursos empleados.

4.1. Descripción de los datos utilizados

En esta sección se presenta una descripción detallada de los datos utilizados en el estudio. Se mencionan las características de los datos, cómo se recopilaron y de dónde proceden.

Los datos utilizados fueron obtenidos del CRHM en colaboración con la Universidad Politécnica de Madrid. Estos datos se recopilaron como parte de un estudio para desarrollar un modelo de monitorización y predicción de los niveles de hemoglobina en donantes de sangre. Tras eliminar las señales que por errores de conexión o ruido no eran útiles, se obtuvieron un total de 73 mediciones de 40 donantes diferentes, incluyendo señales tanto rojas como infrarrojas, de temperatura y clínicas.

4.2. Descripción de las herramientas y recursos utilizados

En esta sección se describen las herramientas, software (SW), hardware (HW) y paquetes especiales de software utilizados en el desarrollo del estudio. Se proporcionan los nombres de las herramientas, sus versiones y cualquier información relevante que sea necesaria para replicar el trabajo.

Para la obtención de las señales PPG y el procesamiento de los datos, se utilizó un dispositivo desarrollado por el equipo de investigación durante el proyecto “Diseño e implementación de un sistema de monitorización ambulatoria y no intrusiva de hemoglobina” [12]. El dispositivo utiliza un nodo sensor formado por un sensor de pulsioximetría Maxim MAX30102 y se conecta mediante BLE a la aplicación móvil *Hemoglobulab* en Android. Este dispositivo permitió obtener las señales rojas e infrarrojas necesarias para el estudio (ver Figura 4.1).

En cuanto al software utilizado, se emplearon herramientas de procesamiento de señales y aprendizaje automático. Se utilizó un algoritmo de detección de picos desarrollado en

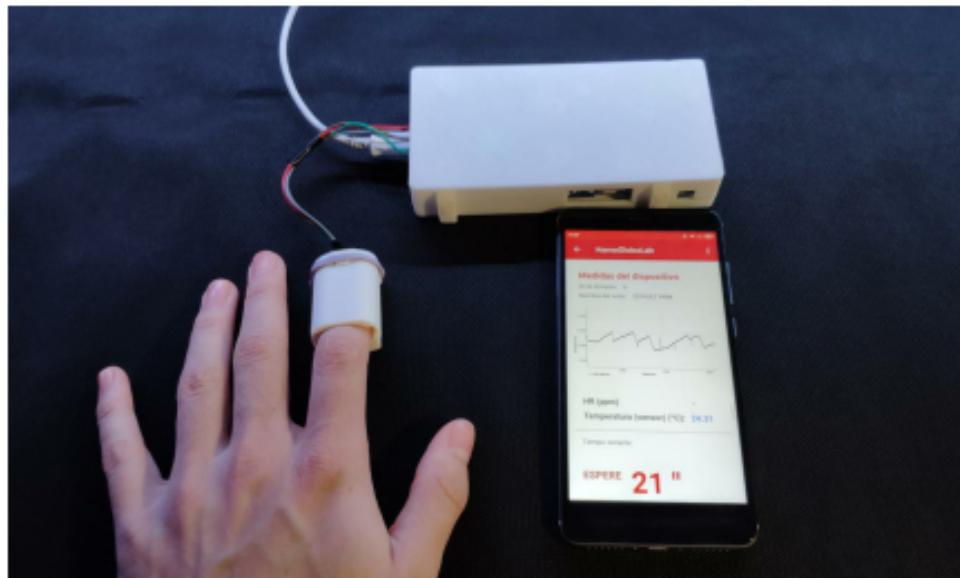


Figura 4.1: Prototipo del entorno de adquisición, donde se encuentra el nodo dentro de la carcasa, conectada por cables al sensor que está dentro del dedal y el móvil recibiendo las medidas [12].

el grupo de investigación en el que se ha realizado el TFG para detectar los picos en las señales PPG, lo cual resultó fundamental para el análisis de la actividad cardíaca.

Para el análisis y visualización de los datos, se utilizaron diferentes paquetes. El paquete “matplotlib.pyplot” permitió generar gráficos y visualizar los resultados de manera efectiva. Además, se utilizó el paquete “numpy” para el manejo eficiente de matrices y el paquete ”pandas” para el análisis de datos.

En el procesamiento de las señales de pulso, se hizo uso del paquete HeartPy¹, que proporcionó funcionalidades para el filtrado de señales, el cálculo de parámetros de calidad de señal y el análisis de la variabilidad de la frecuencia cardíaca (ver Figura 4.2).

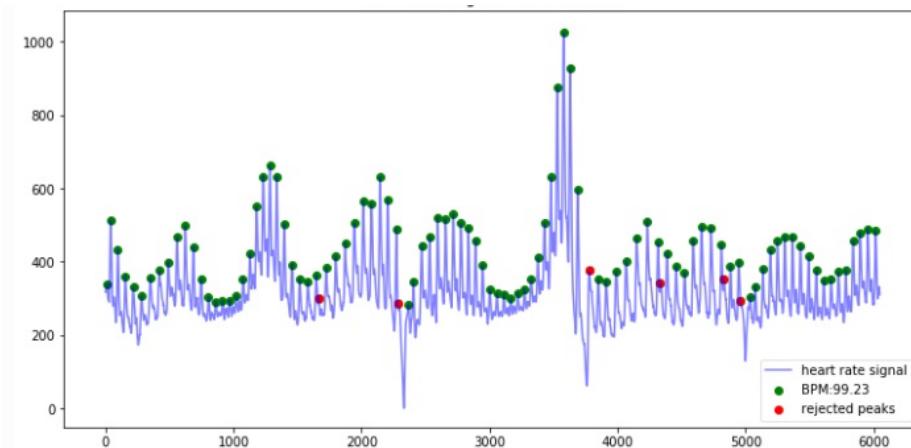


Figura 4.2: Función de detección de picos de HeartPy¹.

Además de estas herramientas y paquetes, se utilizó el paquete “scipy.interpolate” para

¹ Documentación de Python Heart Rate Analysis Toolkit:
<https://python-heart-rate-analysis-toolkit.readthedocs.io/en/latest/>. Consultado el 16 de febrero de 2023

realizar interpolaciones y el paquete “scipy.signal” para aplicar filtros adicionales en el procesamiento de las señales. La interpolación es útil para estimar valores entre puntos de datos conocidos y permite obtener una representación más suave y continua de una señal, mientras que los filtros son herramientas que permiten modificar una señal de entrada para resaltar o atenuar ciertas características, eliminar ruido no deseado o extraer información relevante.

Para el desarrollo de los modelos de aprendizaje automático, se utilizó la biblioteca scikit-learn, una de las bibliotecas más populares para el aprendizaje automático en Python. Scikit-learn ofrece una amplia gama de algoritmos y funciones que facilitan la construcción y evaluación de modelos de aprendizaje automático. En particular, se utilizaron varios módulos específicos de scikit-learn para esta investigación.

En primer lugar, se empleó el módulo `sklearn.model_selection`, que proporciona funciones y clases para la selección de modelos, la evaluación de rendimiento y la optimización de hiperparámetros. Se utilizó la función `train_test_split` para dividir los datos en conjuntos de entrenamiento y test, permitiendo una evaluación adecuada del rendimiento del modelo.

En cuanto a los algoritmos, se hizo uso del módulo `sklearn.ensemble`. Este módulo incluye implementaciones de algoritmos como el Random Forest y el Gradient Boosting. Estos algoritmos son conocidos por su capacidad para mejorar la precisión de los modelos de machine learning al combinar múltiples estimadores. Se utilizaron las clases `RandomForestRegressor` y `GradientBoostingRegressor` para construir modelos de regresión basados en estos algoritmos. Estas implementaciones fueron elegidas debido a su robustez y buen desempeño en problemas de regresión.

La evaluación del rendimiento de los modelos se realizó utilizando el módulo `sklearn.metrics`, el cual proporciona una variedad de funciones para calcular diferentes métricas de evaluación. Se utilizaron las funciones `r2_score`, `mean_absolute_error` y `mean_squared_error` para evaluar los modelos en términos de coeficiente de determinación (R2 score), error absoluto medio (mean absolute error) y error cuadrático medio (mean squared error), respectivamente. Estas métricas permiten medir la calidad de las predicciones realizadas por los modelos y comparar su rendimiento.

Además de las herramientas y paquetes mencionados anteriormente, se utilizó Google Colaboratory (Colab)². como entorno de desarrollo y ejecución de código. Google Colab es una plataforma en la nube que permite escribir y ejecutar código Python en un entorno de notebook de Jupyter. Proporciona acceso gratuito a recursos de computación, como unidades de procesamiento gráfico (GPU) y unidades de procesamiento tensorial (TPU), lo que resulta especialmente útil para tareas que requieren un alto rendimiento computacional, como el entrenamiento de modelos de aprendizaje automático.

²Google:
<https://shorturl.at/dmtVX>. Consultado el 24 de mayo de 2023

5. Resultados

En este capítulo de resultados, se presentan los hallazgos y conclusiones obtenidos a través de la realización del Trabajo de Fin de Grado (TFG). El objetivo principal del TFG fue desarrollar y evaluar un modelo predictivo para estimar la concentración de hemoglobina en gramos por decilitro en base a un conjunto de características seleccionadas. Este capítulo proporcionará un análisis detallado de los resultados obtenidos y discutirá su relevancia en el contexto del problema de investigación planteado.

Este capítulo proporcionará un análisis detallado de los resultados obtenidos durante el desarrollo y evaluación del modelo predictivo. Se discutirán los diferentes enfoques utilizados, como la regresión lineal, Random Forest, Gradient Boosting, Decision Tree Regression y Support Vector Regression, evaluando su rendimiento en términos de coeficiente de determinación (R^2), error absoluto medio (MAE) y otros indicadores relevantes.

Además, es importante destacar que los modelos se probaron tanto con la hemoglobina capilar como con la hemoglobina venosa, lo que permitió obtener una visión más completa y comparativa de su desempeño en diferentes tipos de muestras sanguíneas. Estos análisis comparativos permitirán comprender mejor la capacidad predictiva de los modelos y su aplicabilidad en diferentes contextos clínicos.

Podemos observar en los resultados de todos los modelos que hay menos datos en la predicción de la hemoglobina venosa que en la predicción de la hemoglobina capilar. Esto se debe a que en la hemoglobina venosa, solo se recogió una muestra por cada uno de los donantes.

Es relevante mencionar que se identificó una diferencia de 0.28 g/dL entre los valores de hemoglobina obtenidos mediante el hemoglobinómetro y los obtenidos en el laboratorio del CRHM. Es importante tener en cuenta este dato a la hora de analizar los resultados obtenidos a través de los modelos propuestos.

5.1. Regresión Lineal

En cuanto al modelo de regresión lineal, para la hemoglobina capilar se obtuvo un coeficiente de determinación (R^2 score) de aproximadamente 0.3 en el conjunto de entrenamiento y 0.23 en el conjunto de test. Estos valores indican que solo se puede explicar alrededor del 30% y 23% de la variabilidad de la concentración de hemoglobina en los conjuntos de entrenamiento y test, respectivamente.

Además, se calcularon los errores absolutos medios (MAE) para evaluar la precisión del modelo. En el conjunto de entrenamiento, se obtuvo un MAE de 0.851, mientras que en el conjunto de test, se obtuvo un MAE de 0.859, como se muestra en la Figura 5.1.

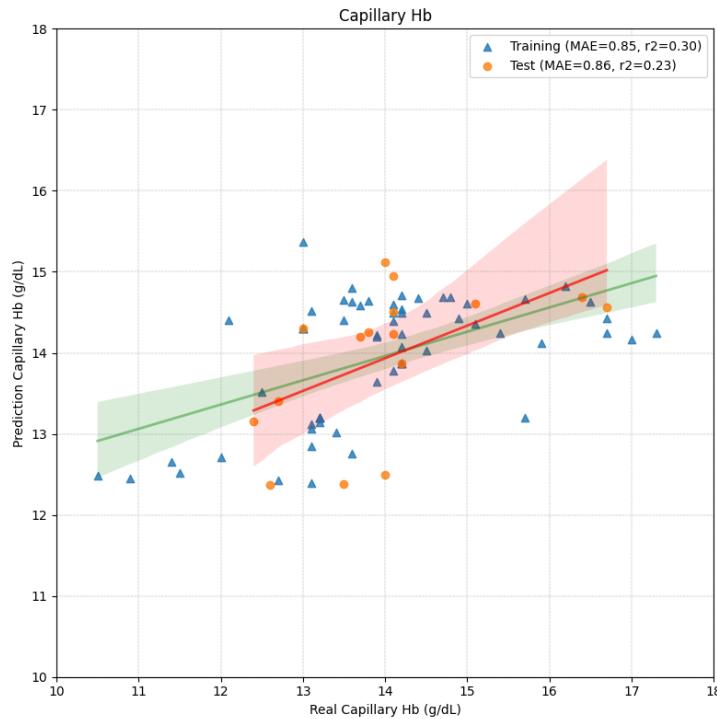


Figura 5.1: Valores previstos frente a valores reales en el modelo de Regresión Lineal para la hemoglobina capilar

Para la hemoglobina venosa, se obtuvo un coeficiente de determinación (R2 score) de aproximadamente 0.77 en el conjunto de entrenamiento y 0.57 en el conjunto de test. Estos valores indican que solo se puede explicar alrededor del 77 % y 57 % de la variabilidad de la concentración de hemoglobina en los conjuntos de entrenamiento y test, respectivamente.

Además, se calcularon los errores absolutos medios (MAE) para evaluar la precisión del modelo. En el conjunto de entrenamiento, se obtuvo un MAE de 0.37, mientras que en el conjunto de test, se obtuvo un MAE de 0.52, como se muestra en la Figura 5.2.

5.2. Random Forest

Durante el proceso de selección de características, se identificaron 8 características como las más relevantes para el modelo de Random Forest. Algunas de las características seleccionadas se muestran en la Tabla 5.1.

El modelo de Random Forest fue evaluado utilizando conjuntos de entrenamiento y test. En cuanto a la hemoglobina capilar, en el conjunto de entrenamiento, el modelo obtuvo un coeficiente de determinación (R2 Score) de 0.87, lo cual indica que explica aproximadamente el 87 % de la varianza en los datos. Además, el error absoluto medio (Mean Absolute Error) fue de 0.3892, lo que implica que las predicciones del modelo difieren en promedio en 0.3892 unidades del valor real.

En el conjunto de test, el modelo obtuvo un coeficiente de determinación de 0.63, lo que representa el 63 % de la varianza explicada en los datos. El error absoluto medio fue de 0.4916, indicando que las predicciones tienen una diferencia promedio de 0.4916 unidades respecto al valor real. Podemos observar estos resultados en la Figura 5.3.

En cuanto a la hemoglobina venosa, en el conjunto de entrenamiento se obtuvo un co-

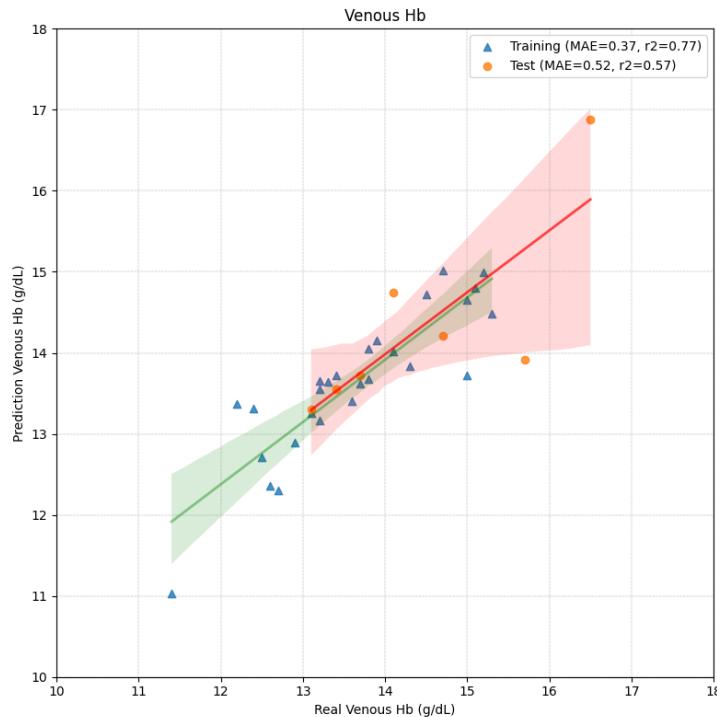


Figura 5.2: Valores previstos frente a valores reales en el modelo de Regresión Lineal para la hemoglobina venosa

eficiente de determinación de $R^2=0.82$, y un MAE=0.36, mientras que en el conjunto de test los resultados fueron un coeficiente de determinación de $R^2=0.54$ y un MAE=0.50. Podemos observar estos resultados en la Figura 5.4. Esto indica que la hemoglobina capilar tuvo mejores resultados para este modelo.

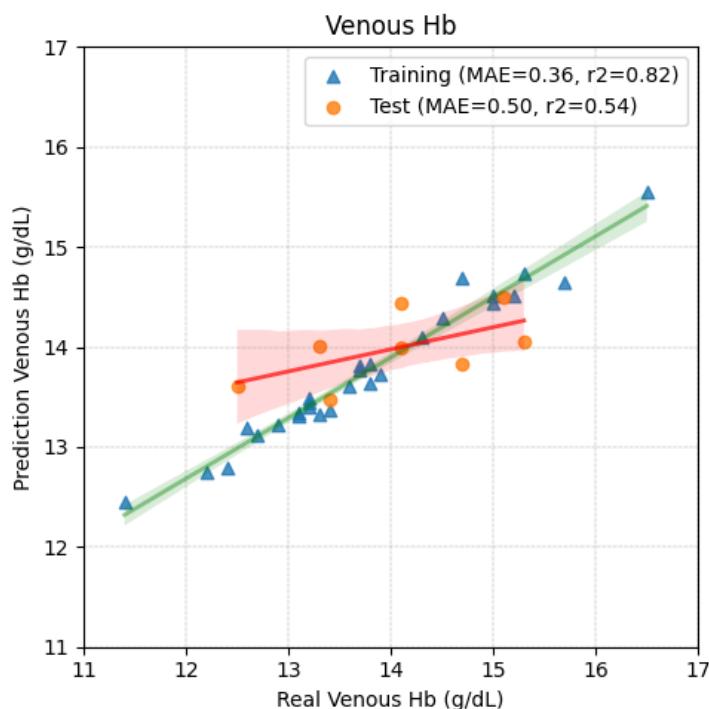


Figura 5.4: Valores previstos frente a valores reales en el modelo de Random Forest para la hemoglobina venosa

A pesar de que los valores de los coeficientes de determinación son bajos, siguen siendo

Número	Característica
1	sex
2	ir_breath
3	red_lf_hf
4	bmi
5	red_pi
6	ir_pi
7	ir_pnn20
8	temp_dc

Tabla 5.1: Características seleccionadas para el Random Forest

útiles, ya que el objetivo es realizar una evaluación general o de alto nivel, sin profundizar en detalles o matices más pequeños. No obstante, es necesario obtener más datos y mejores modelos para mejorar estos resultados.

5.3. Gradient Boosting

Tras la selección de características, se identificaron 9 características como las más importantes para el modelo de Gradient Boosting (ver Tabla 5.2).

Número	Característica
1	sex
2	ir_pi
3	ir_pnn20
4	red_lf_hf
5	temp_dc
6	ir_breath
7	red_pi
8	bmi
9	red_hf_perc

Tabla 5.2: Características seleccionadas para el Gradient Boosting

En el caso del modelo de Gradient Boosting, se han obtenido unos resultados prometedores. En el conjunto de entrenamiento, el coeficiente de determinación (R2 score) alcanzó un valor de 1, lo que indica que aproximadamente el 100 % de la variabilidad de la concentración de hemoglobina se puede explicar mediante este modelo. Además, el error absoluto medio (MAE) fue de 0, lo que implica que las predicciones del modelo difieren en promedio en 0 unidades del valor real.

Al evaluar el modelo para la hemoglobina capilar, en el conjunto de test, se obtuvo un coeficiente de determinación de 0.64, lo que significa que el modelo explica alrededor del 64 % de la varianza en los datos de test. El error absoluto medio en este conjunto fue de 0.5151, indicando una diferencia promedio de 0.5151 unidades entre las predicciones y los valores reales. Podemos observar estos resultados en la Figura 5.5.

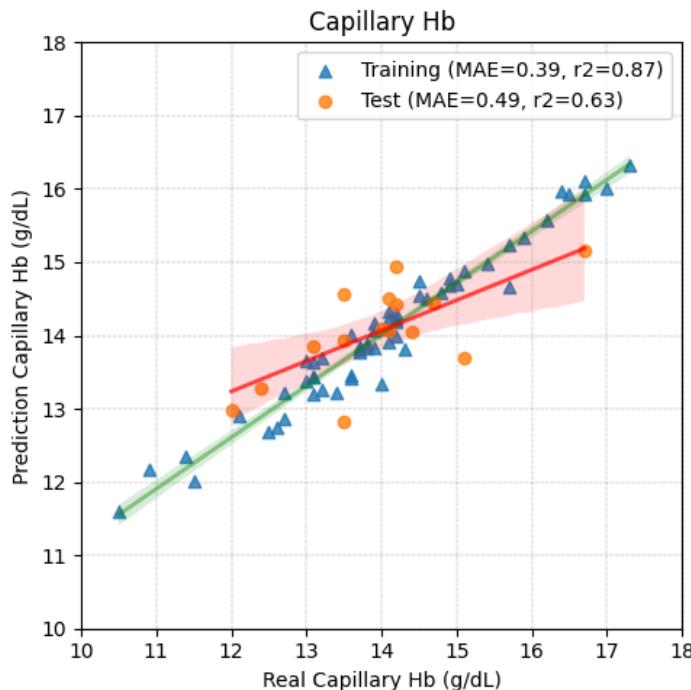


Figura 5.3: Valores previstos frente a valores reales en el modelo de Random Forest para la hemoglobina capilar

En el caso de la evaluación del modelo para la hemoglobina venosa, se observa un rendimiento excelente en el conjunto de entrenamiento, con un coeficiente de determinación (R^2) de 1 y un error absoluto medio (MAE) de 0.00. Estos valores, al igual que para la hemoglobina capilar, indican que el modelo es capaz de ajustarse perfectamente a los datos de entrenamiento y predecirlos con precisión absoluta.

Sin embargo, cuando evaluamos el conjunto de prueba, se observa una disminución significativa en el rendimiento, con un coeficiente de determinación de $R^2=0.64$ y un MAE=0.52. Estos valores indican que el modelo no puede generalizar adecuadamente a datos nuevos y desconocidos, lo que se traduce en una pérdida de precisión y un mayor error en las predicciones.

La discrepancia entre los resultados del conjunto de entrenamiento y del conjunto de prueba tanto para la hemoglobina capilar como para la venosa es un fuerte indicio de que el modelo ha capturado relaciones espurias o específicas de los datos de entrenamiento, en lugar de aprender patrones generales que se puedan aplicar a nuevos datos. Este comportamiento es característico del sobreajuste, ya que el modelo se ha ajustado demasiado a los datos de entrenamiento y no es capaz de generalizar de manera efectiva. Estos resultados se pueden observar en la Figura 5.6.

5.4. Decision Tree Regression

Durante la selección de características, se identificaron 10 características como las más importantes para el modelo de Decision Tree Regression (ver Tabla 5.3).

En cuanto al modelo de Decision Tree Regression, para la hemoglobina capilar, en el conjunto de entrenamiento, el modelo obtuvo un coeficiente de determinación (R^2 Score) de 1, lo cual indica que explica el 100 % de la varianza en los datos. Además, el error

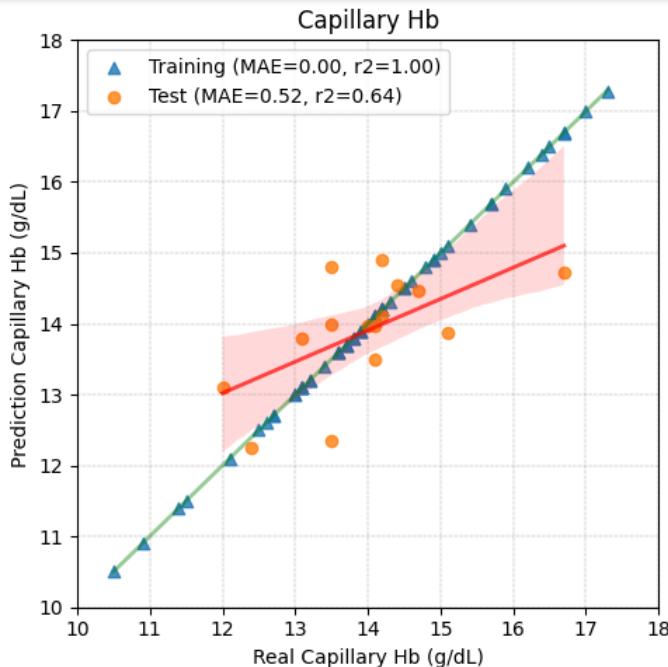


Figura 5.5: Valores previstos frente a valores reales en el modelo de Gradient Boosting para la hemoglobina capilar

absoluto medio (Mean Absolute Error) fue de 0.013, lo que implica que las predicciones del modelo difieren en promedio en 0.013 unidades del valor real.

En el conjunto de test, el modelo obtuvo un coeficiente de determinación de 0.69, lo que representa el 69 % de la varianza explicada en los datos. El error absoluto medio fue de 0.4166, indicando que las predicciones tienen una diferencia promedio de 0.4166 unidades respecto al valor real. Podemos observar estos resultados en la Figura 5.7.

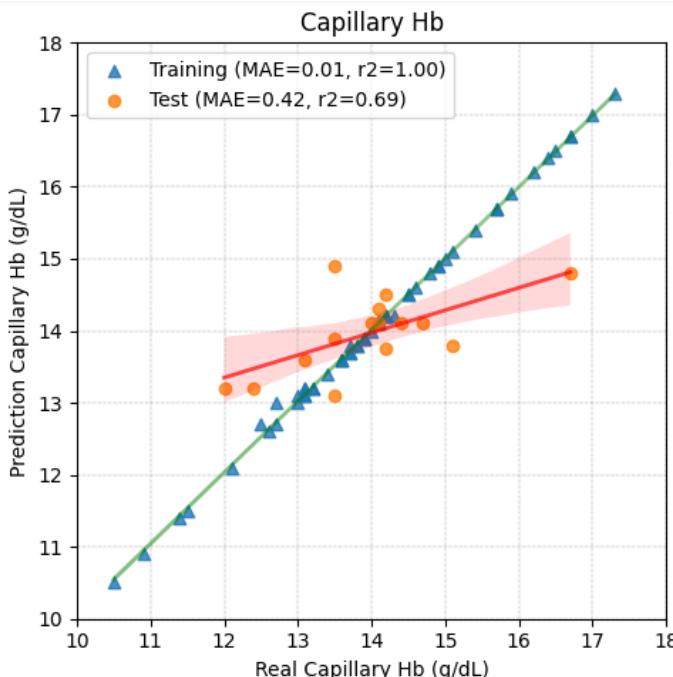


Figura 5.7: Valores previstos frente a valores reales en el modelo de Decision Tree Regression para la hemoglobina capilar

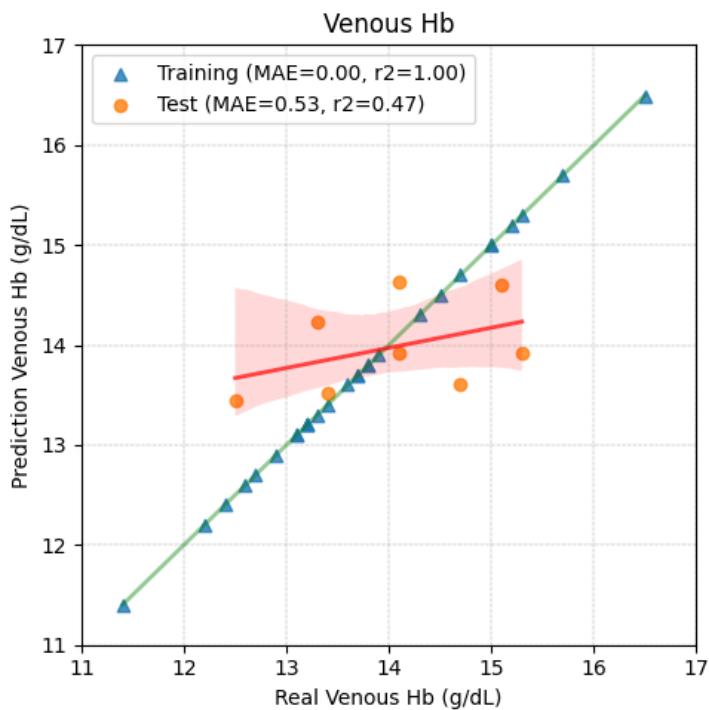


Figura 5.6: Valores previstos frente a valores reales en el modelo de Gradient Boosting para la hemoglobina venosa

Por otra parte, para la hemoglobina venosa, en el conjunto de entrenamiento se obtuvo un coeficiente de determinación de R²=1 y un MAE=0, y en el conjunto de test los resultados fueron de R²=0.47 y de MAE=0.46 , como se puede observar en la Figura 5.8.

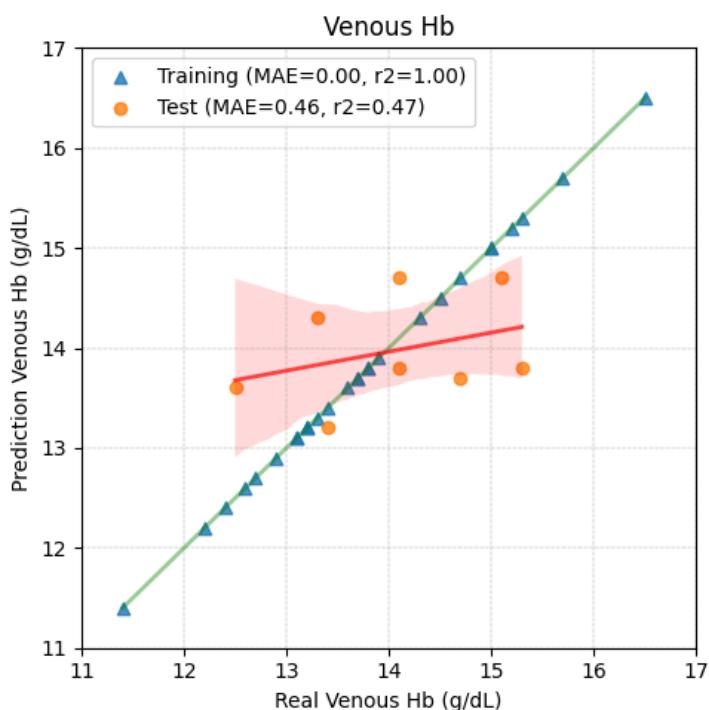


Figura 5.8: Valores previstos frente a valores reales en el modelo de Decision Tree Regression para la hemoglobina venosa

Número	Característica
1	sex
2	temp_dc
3	ir_breath
4	ir_pnn20
5	red_lf_hf
6	red_hf_perc
7	red_lf_perc
8	ir_pi
9	bmi
10	red_pi

Tabla 5.3: Características seleccionadas para el Decision Tree Regression

5.5. Support Vector Regression

Durante la selección de características, se identificaron 2 características como las más importantes para el modelo de Support Vector Regression (ver Tabla 5.4). Sin embargo, es importante tener en cuenta que este resultado puede indicar un posible sobreajuste en el modelo.

El sobreajuste ocurre cuando un modelo se ajusta demasiado a los datos de entrenamiento específicos y no generaliza bien a nuevos datos. En este caso, si el modelo se basa únicamente en las características *sex* y *ir_breath*, existe el riesgo de que esté capturando relaciones espurias o específicas de los datos de entrenamiento, lo que puede afectar negativamente su rendimiento en datos nuevos y desconocidos.

Número	Característica
1	sex
2	ir_breath

Tabla 5.4: Características seleccionadas para el Support Vector Regression

Por último, se repasarán los resultados obtenidos al utilizar el algoritmo Support Vector Regression. En el conjunto de entrenamiento, en la predicción de la hemoglobina capilar, se obtuvo un coeficiente de determinación (R^2) de 0.29, lo cual indica que se puede explicar aproximadamente el 29 % de la variabilidad en los datos de concentración de hemoglobina. Además, el error absoluto medio (MAE) fue de 0.8654, lo que implica que las predicciones del modelo tienen una diferencia promedio de 0.8654 unidades con respecto al valor real.

En el conjunto de test, para la hemoglobina capilar, el modelo logró un coeficiente de determinación (R^2) de 0.43, lo que representa el 43 % de la varianza explicada en los datos de concentración de hemoglobina. El error absoluto medio (MAE) fue de 0.5646, indicando que las predicciones tienen una diferencia promedio de 0.5646 unidades con respecto al valor real. Los resultados se pueden observar en la Figura 5.9.

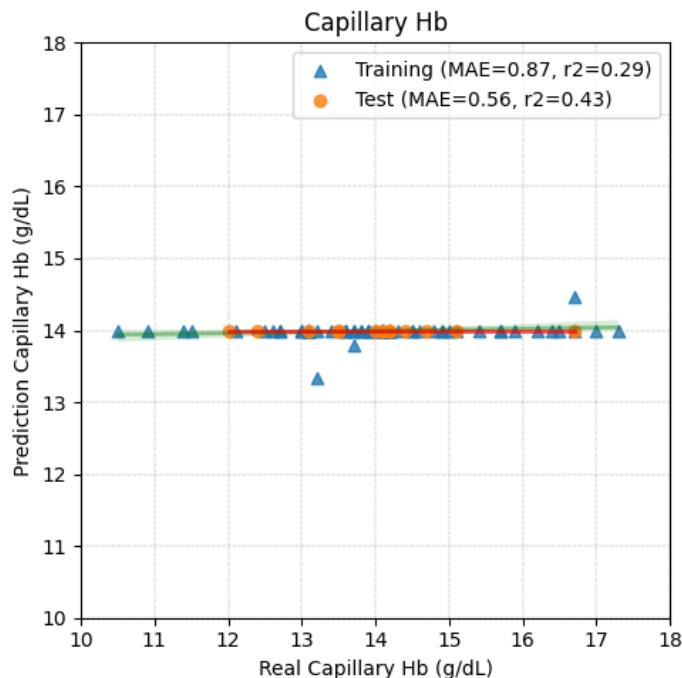


Figura 5.9: Valores previstos frente a valores reales en el modelo de Support Vector Regression para la hemoglobina capilar

Para la hemoglobina venosa, los resultados en el conjunto de entrenamiento fueron de un coeficiente de determinación de $R=0.06$ y un $MAE=0.78$, mientras que en el conjunto de prueba se obtuvo un $R=-0.12$ y un $MAE=0.76$, como podemos observar en la Figura 5.10.

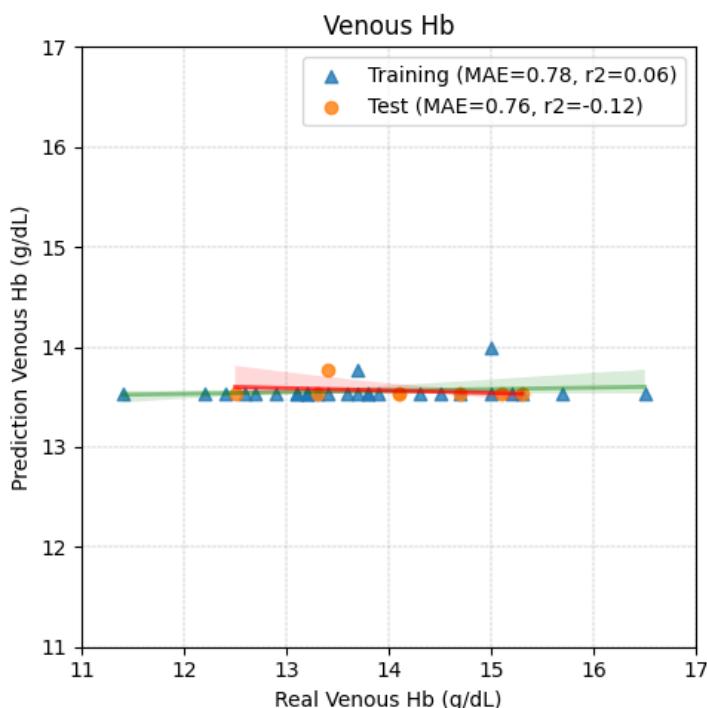


Figura 5.10: Valores previstos frente a valores reales en el modelo de Support Vector Regression para la hemoglobina venosa

Estos resultados sugieren que el modelo de Support Vector Regression no es capaz de capturar adecuadamente la relación entre las características de entrada y la concentración de hemoglobina. En lugar de proporcionar predicciones significativas, el modelo parece

estar produciendo resultados consistentes pero incorrectos. Esto podría indicar una falta de capacidad para capturar la complejidad de los datos o una limitación inherente del algoritmo utilizado.

5.6. Discusión de los resultados

Al analizar estos resultados, podemos observar que los modelos de Gradient Boosting y Decision Tree Regression mostraron el MAE más bajo en el conjunto de entrenamiento para ambas hemoglobinas. El modelo de Decision Tree Regression fue el modelo con el resultado más optimista, presentando un MAE=0.42 sobre el conjunto de prueba.

Sin embargo, es importante tener en cuenta que tanto el modelo de Gradient Boosting como el de Decision Tree Regression sufrieron de sobreajuste, ya que presentaron un mayor MAE en el conjunto de test en comparación con el conjunto de entrenamiento. Por otro lado, los modelos de regresión lineal y Random Forest también tuvieron buenos resultados en términos de MAE, aunque ligeramente más altos que los modelos anteriores.

En todos los modelos se han obtenido mejores resultados al realizar la predicción sobre la hemoglobina capilar que sobre la hemoglobina venosa, por lo que se propone trabajar sobre este dato en futuras investigaciones.

6. Conclusiones y líneas futuras

En este capítulo de conclusiones, retomaremos el problema detectado y los objetivos planteados inicialmente, evaluaremos los resultados y su validez y propondremos posibles direcciones para la investigación futura.

6.1. Conclusiones

Este proyecto tiene como objetivos fundamentales (i) procesar y analizar las señales PPG proporcionadas por el CRHM (ii) generar y seleccionar características para su posterior uso en el modelo de aprendizaje automático, (iii) diseñar e implementar un modelo de aprendizaje automático para la predicción del nivel de hemoglobina, .

El resultado del desarrollo del proyecto ha sido la creación de un sistema completo de extracción de características a partir de 73 señales PPG de 40 pacientes distintos, y de un modelo de aprendizaje automático para la predicción del nivel de hemoglobina a partir de estas señales PPG. Para lograr este resultado se han utilizado 26 parámetros característicos, y se han probado varios algoritmos de aprendizaje automático tanto para la medida de hemoglobina capilar como para la de hemoglobina venosa, como la regresión lineal, Random Forest, Gradient Boosting, Decision Tree Regression o Support Vector Regression. El mejor resultado ha sido obtenido para la predicción de la hemoglobina capilar con el modelo de Decision Tree Regression, con un error absoluto medio de 0.4166g/dL sobre el conjunto de prueba.

Este resultado es muy prometedor, teniendo en cuenta que se identificó una diferencia de 0.28 g/dL entre los valores de hemoglobina obtenidos mediante el hemoglobinómetro y los valores obtenidos en el laboratorio del CRHM. Este dato muestra la viabilidad y la utilidad del sistema, especialmente considerando el alto costo del hemoglobinómetro, además de los requerimientos y condiciones que implica su uso.

Este proyecto ha logrado avanzar en la predicción del nivel de hemoglobina utilizando señales PPG, lo que representa una alternativa prometedora y de bajo costo en comparación con los métodos tradicionales. Los resultados obtenidos abren nuevas posibilidades en la monitorización y seguimiento de los niveles de hemoglobina, lo que puede contribuir significativamente al cuidado de la salud y a la optimización del sistema de donación de sangre.

6.2. Líneas futuras

El presente proyecto ha sentado las bases para el desarrollo de futuras investigaciones y mejoras en el campo de la predicción del nivel de hemoglobina utilizando señales PPG y

modelos de aprendizaje automático.

Una posible línea futura consiste en la incorporación de nuevas características al modelo. Aunque se han generado y seleccionado características relevantes, es importante seguir investigando y evaluando la inclusión de otras que puedan mejorar la precisión de la predicción.

Otra dirección interesante sería explorar diferentes algoritmos de aprendizaje automático. Aunque se han probado varios en este proyecto, existen numerosas técnicas disponibles en el campo. Sería valioso investigar y comparar el rendimiento de otros algoritmos, como redes neuronales profundas, algoritmos de aprendizaje profundo y ensamblados de modelos, para determinar si alguno de ellos ofrece mejoras significativas en la precisión de la predicción.

Además, sería necesario validar el sistema en diferentes poblaciones y condiciones clínicas para evaluar su aplicabilidad y generalización. Esto implicaría probar el sistema en grupos demográficos diversos y en pacientes con condiciones médicas específicas. Esta validación adicional permitiría evaluar su desempeño en diferentes contextos, revelando patrones de predicción útiles y adaptaciones necesarias.

La integración del sistema desarrollado con tecnologías de monitoreo continuo sería otra línea futura interesante. El uso de señales PPG para la predicción de la hemoglobina ofrece la posibilidad de monitoreo no invasivo y en tiempo real. Por lo tanto, explorar la integración del sistema con dispositivos portátiles o sensores implantables permitiría un seguimiento constante del nivel de hemoglobina, detectando cambios y tendencias a lo largo del tiempo para así optimizar el sistema de donación de sangre.

Por último, se podría considerar la extensión del enfoque utilizado en este proyecto a otras variables clínicas de interés. Además de la predicción del nivel de hemoglobina, se podrían investigar otras variables sanguíneas relevantes, como el nivel de glucosa o la presencia de enfermedades cardiovasculares. Esto ampliaría la utilidad y el alcance del sistema desarrollado, permitiendo su aplicación en diferentes áreas de la medicina.

Bibliografía

- [1] C. G. Atkins, K. Buckley, M. W. Blades, and R. F. Turner, “Raman spectroscopy of blood and blood components,” *Appl. Spectrosc.*, vol. 71, no. 5, pp. 767–793, May 2017. [Online]. Available: <https://opg.optica.org/as/abstract.cfm?URI=as-71-5-767>
- [2] C. C. Helms, M. T. Gladwin, and D. B. Kim-Shapiro, “Erythrocytes and vascular function: oxygen and nitric oxide,” *Frontiers in physiology*, vol. 9, p. 125, 2018.
- [3] T. Gordon-Smith, “Structure and function of red and white blood cells,” *Medicine*, vol. 37, no. 3, pp. 119–124, 2009, haematology Part 1 of 2. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1357303909000255>
- [4] K. Jurk and B. E. Kehrel, “Platelets: physiology and biochemistry,” in *Seminars in thrombosis and hemostasis*, vol. 31, no. 04. Copyright © 2005 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New . . . , 2005, pp. 381–392.
- [5] J. Emmanuel and W. H. O. B. T. Safety, *The Clinical Use of Blood in Medicine, Obstetrics, Paediatrics, Surgery & Anaesthesia, Trauma & Burns*, ser. Blood transfusion safety. World Health Organization, Blood Transfusion Safety, 2001. [Online]. Available: <https://books.google.es/books?id=PAYoYPzwRv0C>
- [6] M. Bastos Oreiro, “Anemia ferropénica: Tratamiento,” *Revista Española de Enfermedades Digestivas*, vol. 101, no. 1, pp. 70–70, 2009.
- [7] A. E. Mast, K. S. Schlumpf, D. J. Wright, B. Custer, B. Spencer, E. L. Murphy, T. L. Simon, and for the NHLBI Retrovirus Epidemiology Donor Study-II, “Demographic correlates of low hemoglobin deferral among prospective whole blood donors,” *Transfusion*, vol. 50, no. 8, pp. 1794–1802, 2010. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1537-2995.2010.02649.x>
- [8] Estadísticas de la Comunidad de Madrid. (Accedido: 24-01-2023) Relación donaciones de sangre y donantes. [Online]. Available: <https://www.comunidad.madrid/servicios/salud/estadisticas-donacion-sangre>
- [9] L. Neufeld and et al., “Hemoglobin measured by hemocue and a reference method in venous and capillary blood: a validation study,” *Salud pública Méx*, vol. 44, no. 3, pp. 219–227, 2002. [Online]. Available: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0036-36342002000300005&lng=es&nrm=iso
- [10] A. R. Kavsaoglu, K. Polat, and M. Hariharan, “Non-invasive prediction of hemoglobin level using machine learning techniques with the ppg signal’s characteristics features,” *Applied Soft Computing*, vol. 37, pp. 983–991, 2015. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1568494615002227>
- [11] M. K. Hasan, M. H. Aziz, M. I. I. Zarif, M. Hasan, M. Hashem, S. Guha, R. R. Love, and S. Ahamed, “Noninvasive hemoglobin level prediction in a mobile phone environment: State of the art review and recommendations,”

- JMIR Mhealth Uhealth*, vol. 9, no. 4, p. e16806, Apr 2021. [Online]. Available: <https://mhealth.jmir.org/2021/4/e16806>
- [12] A. M. Miguel, “Diseño e implementación de un sistema de monitorización ambulatoria y no intrusiva de hemoglobina,” Bachelor’s Thesis, ETSI Telecomunicación, 2021.
 - [13] L. Lanser, D. Fuchs, K. Kurz, and G. Weiss, “Physiology and inflammation driven pathophysiology of iron homeostasis—mechanistic insights into anemia of inflammation and its treatment,” *Nutrients*, vol. 13, no. 11, 2021. [Online]. Available: <https://www.mdpi.com/2072-6643/13/11/3732>
 - [14] M. A. uddowla Golap, S. M. T. U. Raju, M. R. Haque, and M. Hashem, “Hemoglobin and glucose level estimation from ppg characteristics features of fingertip video using mggp-based model,” *Biomedical Signal Processing and Control*, vol. 67, p. 102478, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1746809421000756>
 - [15] H. Robertus, C. L. Hendiarta, and A. A. S. Gunawan, “A comparison of noisy photoplethysmography (PPG) signals using heartpy with MAX30102 and pulse sensor,” *AIP Conference Proceedings*, vol. 2594, no. 1, 04 2023, 060001. [Online]. Available: <https://doi.org/10.1063/5.0109091>
 - [16] J. Přibil, A. Přibilová, and I. Frollo, “Comparative measurement of the ppg signal on different human body positions by sensors working in reflection and transmission modes,” *Engineering Proceedings*, vol. 2, no. 1, 2020. [Online]. Available: <https://www.mdpi.com/2673-4591/2/1/69>
 - [17] S. H. Kim, D. W. Ryoo, and C. Bae, “Adaptive noise cancellation using accelerometers for the ppg signal from forehead,” in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2007, pp. 2564–2567.
 - [18] R. A. Rochmanto, H. Zakaria, R. D. Alviana, and N. Shahib, “Non-invasive hemoglobin measurement for anemia diagnosis,” in *2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, 2017, pp. 1–5.
 - [19] F. Sanchis-Gomar, J. Cortell-Ballester, H. Pareja-Galeano, G. Banfi, and G. Lippi, “Hemoglobin point-of-care testing: the hemocue system,” *Journal of laboratory automation*, vol. 18, no. 3, pp. 198–205, 2013.
 - [20] L. D. Morris, A. Osei-Bimpong, D. McKeown, D. Roper, and S. M. Lewis, “Evaluation of the utility of the hemocue 301 haemoglobinometer for blood donor screening,” *Vox Sanguinis*, vol. 93, no. 1, pp. 64–69, 2007. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1423-0410.2007.00919.x>
 - [21] L. Lamhaut, R. Apriotesei, X. Combes, M. Lejay, P. Carli, and B. Vivien, “Comparison of the Accuracy of Noninvasive Hemoglobin Monitoring by Spectrophotometry (SpHb) and HemoCue® with Automated Laboratory Hemoglobin Measurement,” *Anesthesiology*, vol. 115, no. 3, pp. 548–554, 09 2011. [Online]. Available: <https://doi.org/10.1097/ALN.0b013e3182270c22>
 - [22] R. Tiwari and M. P. Singh, “Correlation-based attribute selection using genetic algorithm,” *International Journal of Computer Applications*, vol. 4, no. 8, pp. 28–34, 2010.

- [23] J. M. Martínez-Martínez, P. Escandell-Montero, C. Barbieri, E. Soria-Olivas, F. Mari, M. Martínez-Sober, C. Amato, A. J. Serrano López, M. Bassi, R. Magdalena-Benedito, A. Stopper, J. D. Martín-Guerrero, and E. Gatti, “Prediction of the hemoglobin level in hemodialysis patients using machine learning techniques,” *Computer Methods and Programs in Biomedicine*, vol. 117, no. 2, pp. 208–217, 2014. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0169260714002569>
- [24] M. Elgendi, “Optimal signal quality index for photoplethysmogram signals,” *Bioengineering*, vol. 3, no. 4, 2016. [Online]. Available: <https://www.mdpi.com/2306-5354/3/4/21>
- [25] R. Strogonovs. (Accessed 2023) Implementing pulse oximeter using max30100. [Online]. Available: <https://morf.lv/implementing-pulse-oximeter-using-max30100>

Anexo A: Aspectos éticos, económicos, sociales y ambientales

A1. INTRODUCCIÓN

En el desarrollo de cualquier proyecto, es fundamental considerar no solo los aspectos técnicos y operativos, sino también los impactos que este pueda tener en diferentes ámbitos. Los aspectos éticos, económicos, sociales y ambientales son elementos cruciales que deben ser abordados de manera integral, garantizando una gestión responsable y sostenible.

En este anexo, se abordarán en detalle los aspectos éticos, económicos, sociales y ambientales relacionados con el proyecto en cuestión. Estos aspectos están intrínsecamente ligados y su consideración adecuada permite evaluar y gestionar de manera efectiva los impactos positivos y negativos que el proyecto puede tener en la sociedad, la economía, el entorno ambiental y los valores éticos.

A2. DESCRIPCIÓN DE IMPACTOS RELEVANTES RELACIONADOS CON EL PROYECTO

En este apartado se evalúan los aspectos éticos, económicos sociales y medioambientales que provoca el desarrollo de este proyecto. Es fundamental considerar estos aspectos para garantizar que el proyecto sea sostenible y beneficioso para todas las partes involucradas.

En el ámbito económico, este trabajo reporta beneficios, puesto que da pie a tener un sistema de hemovigilancia actualizado con medidas reales. Esto permite prever con mayor exactitud la cantidad de donantes aptos el día de la recogida de las bolsas, lo que se traduce en ahorro de mucho dinero en material sanitario y personal.

El párrafo anterior viene relacionado con el ámbito medioambiental, ya que al reducir la fabricación y el número de desplazamientos de material y de personal sanitario, se reduce la emisión de gases perjudiciales para el aire. De la misma manera, se reduce el número de desplazamientos de donantes en vehículo personal, beneficiando de la misma manera al medio ambiente.

En cuanto al ámbito ético, este trabajo no tendrá ningún impacto negativo a la hora de trabajar con los datos de los donantes, notificando previamente a los donantes y mediante la creación y uso de un código anónimo desvinculado del donante, con el que se trabajará a lo largo del proceso.

Por último, se verán grandes beneficios en el ámbito social, ya que los pacientes podrán

saber antes de desplazarse al punto de donación si van a ser aceptados o no, ahorrando así tiempo y evitando la pérdida de donantes que no hubieran vuelto a donar al ser rechazados

A3. ANÁLISIS DETALLADO DE ALGUNO DE LOS IMPACTOS

El impacto más importante de este trabajo será en el ámbito social, ya que cómo se explicó en el Capítulo 1, la sangre es esencial para el cuerpo humano ya que realiza muchas funciones vitales, además de ser esencial para la defensa del organismo. La sangre donada se utiliza en una amplia variedad de tratamientos médicos, desde transfusiones sanguíneas de emergencia hasta tratamientos de enfermedades crónicas. Por estos motivos, es crucial asegurar una buena cadena de transfusión sanguínea para garantizar unas reservas adecuadas de sangre.

El desarrollo de este trabajo busca optimizar la cadena de transfusión sanguínea para tener siempre unos niveles adecuados en las reservas de sangre de los centros sanitarios.

A4. CONCLUSIONES

El desarrollo de este proyecto no solo ofrece avances científicos en la predicción de los niveles de hemoglobina, sino que también proporciona mejoras sociales, económicas y medioambientales. Contribuye a la salud y el bienestar de los pacientes, genera ahorros económicos y promueve prácticas sostenibles.

Anexo B: Presupuesto económico

COSTE DE MANO DE OBRA (coste directo)

Horas	Precio/hora	Total
300	15 €	4.500 €

COSTE DE RECURSOS MATERIALES (coste directo)

	Precio de compra	Uso en meses	Amortización (en años)	Total
Ordenador personal (Software incluido).....	1.500,00 €	6	5	150,00 €
Otro equipamiento				

COSTE TOTAL DE RECURSOS MATERIALES

150,00 €

GASTOS GENERALES (costes indirectos)

15 %

sobre CD

697,50 €

BENEFICIO INDUSTRIAL

6 %

sobre CD + CI

279,00 €

SUBTOTAL PRESUPUESTO

5.626,50 €

IVA APPLICABLE

21 %

1.181,57 €

TOTAL PRESUPUESTO

6.808,07 €

Tabla 6.1: Presupuesto económico