



UNIVERSIDAD DE BURGOS
ESCUELA POLITÉCNICA SUPERIOR
Grado en Ingeniería Informática



TFG del Grado en Ingeniería
Informática

**Identificación de
Parkinson por visión
artificial**



Presentado por Catalin Andrei Cacuci
en Universidad de Burgos — 3 de junio
de 2023

Tutora: Alicia Olivares Gil
Cotutor: Álvar Arnaiz González



UNIVERSIDAD DE BURGOS
ESCUELA POLITÉCNICA SUPERIOR
Grado en Ingeniería Informática



Dña. Alicia Olivares Gil y D. Álvar Arnaiz González, profesores del departamento de Ingeniería Informática, área de Lenguajes y Sistemas Informáticos.

Expones:

Que el alumno D. Catalin Andrei Cacuci, con DNI X7451927L, ha realizado el Trabajo final de Grado en Ingeniería Informática titulado Identificación de Parkinson por visión artificial.

Y que dicho trabajo ha sido realizado por el alumno bajo la dirección del que suscribe, en virtud de lo cual se autoriza su presentación y defensa.

En Burgos, a 3 de junio de 2023

Vº. Bº. de la tutora:

Dña. Alicia Olivares Gil

Vº. Bº. del cotutor:

D. Álvar Arnaiz González

Resumen

La enfermedad de Parkinson afecta a cientos de miles de personas en España, cifra que algunas fuentes preveen que va a subir en el futuro. Aunque es una enfermedad incurable, una detección y tratamiento temprano pueden mejorar la calidad de vida de los pacientes considerablemente.

Este proyecto pretende utilizar técnicas de minería de datos y visión artificial para implementar un sistema capaz de detectar la presencia de bradicinesia, síntoma de la enfermedad de Párkinson, en un individuo de forma no intrusiva, mediante un simple vídeo.

Descriptores

enfermedad de Parkinson, bradicinesia, visión artificial, minería de datos, Docker, desarrollo web

Abstract

Parkinson's disease affects hundreds of thousands of people in Spain, this number is expected to increase in the future. Even though this illness has no cure, early detection and treatment can considerably improve the patient's quality of life.

This project attempts to use data mining and computer vision techniques in order to detect bradykinesia, a common symptom of Parkinson's disease, in a person in a non-intrusive way, with a simple video.

Keywords

Parkinson's disease, bradykinesia, computer vision, data mining, Docker, web development

Índice general

Índice general	III
Índice de figuras	V
Índice de tablas	VI
1. Introducción	1
2. Objetivos del proyecto	3
2.1. Objetivos generales	3
2.2. Objetivos técnicos	4
3. Conceptos teóricos	5
3.1. Minería de Datos	5
4. Técnicas y herramientas	13
4.1. Técnicas	13
4.2. Herramientas de desarrollo	15
5. Aspectos relevantes del desarrollo del proyecto	25
5.1. Fase de experimentación	25
5.2. Fase de desarrollo de la aplicación	36
6. Trabajos relacionados	41
6.1. A computer vision framework for finger-tapping evaluation	41
6.2. The discerning eye of computer vision	42
6.3. Supervised classification of bradykinesia	43

7. Conclusiones y Líneas de trabajo futuras **45**

Bibliografía **47**

Índice de figuras

1.1. Finger-tapping test	1
4.1. CRISP-DM	15
4.2. MediaPipe Hands	20
5.1. Mediapipe Hands «Landmarks»	27
5.2. Serie temporal de «landmarks»	28
5.3. Serie temporal de ángulos	29
5.4. Amplitudes	30
5.5. Amplitudes mediante desviación típica móvil	31
5.6. Amplitudes mediante desviación típica móvil con la enfermedad de Parkinson	31
5.7. Amplitudes mediante desviación típica móvil con movimiento errático	32
5.8. Amplitudes mediante desviación típica móvil con movimiento errático	32

Índice de tablas

3.1. Matriz de confusión	9
------------------------------------	---

1. Introducción

La enfermedad de Parkinson es una enfermedad neurodegenerativa irreversible que afecta a entre 120.000 y 150.000 personas en España [10] y algunas estimaciones preveen que estas cifras aumenten en el futuro [3]. Se caracteriza por la degradación de las células nerviosas que controlan el movimiento, lo que puede provocar temblores, rigidez muscular y pérdida del control postural [12]. Aunque no existe una cura para la enfermedad de Parkinson, la detección temprana y el tratamiento adecuado pueden mejorar significativamente la calidad de vida de los pacientes.

Un síntoma característico y de gran importancia durante el diagnóstico de la enfermedad de Parkinson es lo que se conoce como bradicinesia [9], que es la ralentización de un movimiento a medida que se realiza [1]. Para detectar la presencia de esta característica durante el diagnóstico de la enfermedad es común el uso de la prueba de «finger-tapping», en la que el paciente deberá juntar y separar los dedos índice y pulgar de una mano de forma rápida y rítmica. La figura 1.1 muestra en qué consiste el movimiento.



Figura 1.1: Finger-tapping test

En los últimos años, la inteligencia artificial se ha convertido en una herramienta de gran utilidad para el diagnóstico y tratamiento de enfermedades, incluida la enfermedad de Parkinson. En este proyecto, se ha desarrollado un sistema basado en inteligencia artificial para crear modelos de aprendizaje automático capaces de detectar bradicinesia en un vídeo de un individuo realizando el movimiento de «finger-tapping».

Interactuar con los modelos creados de forma directa puede resultar complicado si no se dispone de ciertos conocimientos previos sobre programación e inteligencia artificial. Con el objetivo de facilitar esta tarea para un usuario estándar se ha desarrollado una aplicación web que simplifica el uso de los modelos.

Los resultados obtenidos indican que los modelos creados son capaces de detectar la enfermedad de Parkinson con bastante precisión, lo que sugiere que podrían ser de utilidad para la realizar un diagnóstico inicial, y nunca sustituyente de una evaluación médica, sobre el estado de la enfermedad. Además, el enfoque basado en vídeo no es invasivo y fácil de usar, lo que lo hace potencialmente útil en entornos clínicos y de telemedicina.

Por último, me gustaría dar las gracias a todos aquellos que han dedicado su tiempo y esfuerzo a la obtención de los vídeos utilizados para entrenar los modelos, incluyendo la Asociación de Parkinson de Burgos, el personal del Hospital Universitario de Burgos y el equipo de la Dra. Esther Cubo.

2. Objetivos del proyecto

2.1. Objetivos generales

El objetivo principal de este proyecto es la creación de un sistema basado en la inteligencia artificial que permita la detección de bradicinesia de forma consistente utilizando como entrada un vídeo de la prueba de «finger-tapping» (además de otra información), además este sistema debería ser utilizable por la persona media. Para alcanzar este objetivo se plantean los siguientes objetivos específicos:

1. Revisar los trabajos previos relacionados para obtener unos conocimientos iniciales sobre el tema e identificar qué métodos funcionan mejor y peor.
2. Diseñar e implementar una secuencia de pasos o «pipeline» para extraer características relevantes de los vídeos para su uso en algoritmos de aprendizaje automático.
3. Buscar un modelo que se ajuste considerablemente bien a los datos disponibles mediante alguna de las técnicas existentes en el campo del aprendizaje automático para este propósito.
4. Crear una aplicación web que permita subir un vídeo de la prueba, extraer una serie de características del mismo y proporcionárselas a un modelo de aprendizaje automático que sea capaz de devolver una predicción relativa a la presencia o no de bradicinesia en el paciente.
5. Debido al público de una aplicación de este tipo se debe tener muy en cuenta la accesibilidad a la hora de diseñarla.

6. Incluir en la aplicación web la funcionalidad de gestión de usuarios así como la posibilidad de actualizar el modelo utilizado para realizar las predicciones por parte de un administrador.
7. Documentar el trabajo realizado.

2.2. Objetivos técnicos

Los siguientes objetivos han están relacionados con los aspectos técnicos de la aplicación.

- Utilizar Docker para facilitar el despliegue de la aplicación en cualquier equipo que tenga el motor de Docker instalado.
- Crear una API (Application Programming Interface) accesible mediante internet desde cualquier máquina para separar de forma clara el «backend» y el «frontend» y permitir la extensión futura de la aplicación.
- Internacionalizar la aplicación para permitir el uso de la misma en varios idiomas y realizar una detección automática del idioma preferido por el usuario.

3. Conceptos teóricos

Este capítulo define algunos de los conceptos teóricos que se mencionan en este documento.

3.1. Minería de Datos

La minería de datos es un área de la inteligencia artificial que consiste en el diseño de algoritmos y modelos que permitan a los ordenadores aprender la regla general que define un conjunto de datos sin ser explícitamente programados para ello.

La minería de datos incluye, además de algoritmos de aprendizaje, todas las metodologías y técnicas utilizadas para el tratamiento y filtrado del conjunto de datos disponible.

Preprocesado

La información con la que se entrene un modelo de aprendizaje automático determina en gran medida el rendimiento que podrá alcanzar, debido a esto es muy frecuente realizar un paso previo a la creación del modelo, denominado *preprocesado* [13].

El objetivo del preprocesado es transformar la entrada de datos iniciales con el fin de permitir y facilitar que un modelo se adapte a los mismos.

El preprocesado también tiene una gran influencia sobre el tiempo de computación necesario para entrenar un modelo y sobre la complejidad que necesitará para adaptarse a los datos.

El preprocesado suele componerse de algunos de los siguientes pasos.

Extracción de características

La extracción de características es el proceso de identificar, seleccionar y transformar atributos relevantes de los datos de entrada para su uso en un modelo. Por ejemplo, en este proyecto, se han extraído características como la velocidad o amplitud de movimiento de los dedos a partir de un vídeo de la prueba de «finger-tapping».

Existen diversas técnicas para la extracción de características, incluyendo la selección manual de características, o técnicas automatizadas como la reducción de dimensionalidad, la extracción de características basada en redes neuronales [4], entre otras. La elección de la técnica de extracción de características depende del conjunto de datos, del problema específico de aprendizaje automático que se está abordando y del tipo de modelo de aprendizaje automático que se está utilizando. En este proyecto se ha realizado una extracción manual para obtener algunas características como la frecuencia de toques y la velocidad del movimiento de «finger-tapping», además de una extracción de múltiples características genéricas de series temporales mediante la librería TSFresh [2].

Selección de características

Los datos de entrada pueden ser muy complejos y estar compuestos por una gran cantidad de información redundante o no relevante para el modelo. La selección de características se utiliza para identificar las características más relevantes y representativas de los datos, que pueden ser utilizadas para entrenar modelos de aprendizaje automático con mayor eficacia.

En la selección de características, se pueden utilizar técnicas como el análisis de componentes principales (PCA) [7], el análisis discriminante lineal (LDA) [19] o pruebas de significancia, entre otras.

Normalización

Existen modelos muy sensibles a que existan diferencias en la escala de los distintos atributos, como, por ejemplo, *k-nearest neighbors*, por lo que es muy habitual que en la fase de preprocesado se realice una normalización de los datos, es decir, transformarlos de tal forma que utilicen la misma escala, en general se suelen tomar valores en los intervalos $[0, 1]$ o $[-1, 1]$.

Aunque lo más habitual es que la normalización se haga sin distorsionar las diferencias existentes entre los valores previos, existen situaciones en las que puede ser ventajoso utilizar un método de normalización que sí altere

estas diferencias, por ejemplo, la normalización por cuantiles [14], en la que se modifican los valores para que sigan una distribución normal, lo que consigue que, si existen valores muy lejanos a los valores más comunes (*outliers*), estos se acerquen al resto.

Aprendizaje supervisado

Dentro del aprendizaje automático existen tres ramas o variantes dependiendo del conjunto de datos del que se disponga, ya que estos datos son los que determinan las técnicas, algoritmos y metodologías que se podrán utilizar, estas variantes son:

- Aprendizaje no supervisado, en el que los datos no están etiquetados, es decir, el atributo que se desea predecir es desconocido (el ejemplo más característico es el *clustering*)
- Aprendizaje semisupervisado, en el que solo una parte de los datos están etiquetados.
- Aprendizaje supervisado, en el que el conjunto de datos está completamente etiquetado. Este es el tipo de aprendizaje utilizado realizado en el proyecto y del que trata este apartado.

El enfoque del aprendizaje supervisado consiste en la utilización de una muestra completamente etiquetada de un conjunto de datos para crear un modelo que «aprenda» de las instancias de la muestra la regla general que define el conjunto de datos completo [6].

Clasificación

En el ámbito del aprendizaje supervisado, clasificar instancias en categorías predeterminadas es uno de los principales problemas a resolver.

En general, la resolución de un problema de clasificación se caracteriza por delimitar qué zonas del espacio que contiene todas las instancias posibles pertenecen a cada categoría (espacio de datos), es decir, definir fronteras de decisión a partir de las cuales las instancias pasan de una categoría a otra.

Existe una gran cantidad de métodos y algoritmos de clasificación que se pueden utilizar para determinar estas fronteras, dependiendo del problema específico que se intente resolver unos algoritmos se comportarán mejor que otros. También cabe destacar que algunos algoritmos permiten obtener la

probabilidad de pertenencia a cada clase, como, por ejemplo, máquinas de vectores de soporte (SVM).

Además, dependiendo de la variable objetivo que se intenta predecir existen distintos tipos de clasificación:

- **Clasificación binaria:** Es el tipo de clasificación más simple, la variable objetivo tiene dos valores posibles, es decir, se va a intentar categorizar instancias en dos clases. Este es el tipo usado en este caso, ya que se está intentando determinar si un individuo tiene bradicinesia o no.
- **Clasificación multiclasa:** Este tipo de clasificación es similar al anterior, la diferencia está en que la variable objetivo puede tomar más de dos valores, es decir, se busca clasificar las instancias en más de dos clases.
- **Clasificación ordenada:** En este tipo de clasificación las clases tienen un orden, un ejemplo sería si en este proyecto se buscara determinar el grado de severidad de la bradicinesia detectada (por ejemplo inexistente, bajo, medio y algo). Además, este tipo de clasificación es muy similar a un problema de regresión, ya que se puede asignar valores numéricos a cada categoría determinando el orden

Sobreajuste

Normalmente solo se dispone de una pequeña muestra para entrenar un modelo que encuentre la regla general que define la población y se adapte lo mejor posible a esta.

El problema de esto es que, si se utilizan todos los datos disponibles durante el entrenamiento, no existe ninguna forma para verificar el comportamiento del modelo cuando se encuentre con instancias que no ha visto antes. Esto hace posible que el modelo “memorice” los datos con los que se ha entrenado pero no generalice bien al encontrarse con datos nuevos. Esto se conoce como sobreajuste.

Una solución muy común al sobreajuste es separar los datos disponibles en dos grupos, un conjunto de entrenamiento y un conjunto de test o prueba. Durante el entrenamiento el modelo solo tendrá acceso al conjunto de datos de entrenamiento, mientras que el conjunto de test es utilizado para determinar el rendimiento del modelo sobre datos nuevos mediante una de las métricas de evaluación que se verán a continuación.

Evaluación del modelo

Para determinar si un modelo es mejor que otro es necesario definir alguna métrica que asigne un valor numérico al rendimiento de cada modelo. Estas métricas utilizan las variables que contiene la matriz de confusión. Por ejemplo, la matriz de confusión para clasificación binaria es la siguiente.

		Predicción		
		p	n	total
Valor real	p'	True Positive	False Negative	P'
	n'	False Positive	True Negative	N'
total		P	N	

Tabla 3.1: Matriz de confusión

La métrica más intuitiva es lo que se conoce como precisión (*accuracy*), que es simplemente la proporción de predicciones (clasificaciones) acertadas del total de predicciones realizadas.

$$\text{precisión} = \frac{TP + TN}{TP + TN + FP + FN}$$

Al igual que ocurren con el algoritmo de clasificación empleado, la métrica que se use debe ser seleccionada en función del problema en cuestión. Una situación muy común en la que utilizar la precisión no es lo más ideal es cuando se trabaja con conjuntos de datos desbalanceados, es decir, cuando existen más instancias de una clase que de otra.

Lo anterior es muy común en el ámbito médico cuando se intenta crear un modelo que determine si un paciente padece una enfermedad concreta o

no. En estas situaciones se suele tener un grupo de control muy grande que no padece la enfermedad y un grupo relativamente pequeño que sí la padece. Si, por ejemplo, las proporciones de clases son 95 % y 5 % respectivamente, un modelo que siempre prediga que el paciente no padece la enfermedad en cuestión obtendrá una precisión del 95 %, lo que podría dar la falsa impresión de que se ajusta bien a los datos, cuando, en realidad, es un modelo completamente inútil.

Algunas alternativas a la precisión incluyen:

- **Sensibilidad:** Para una clase concreta, indica la capacidad del modelo de clasificar correctamente instancias de esa clase. En el caso médico esta métrica podría ser interesante para encontrar aquellos pacientes que sí padecen la enfermedad, aunque se clasifiquen mal algunos que no la padecen.

$$\text{sensibilidad} = \frac{TP}{TP + FP}$$

- **Especificidad:** Para una clase concreta, indica la capacidad del modelo para clasificar correctamente instancias que no pertenecen esa clase.

$$\text{especificidad} = \frac{TP}{TP + FN}$$

- **Valor-F1:** Es una medida que combina la sensibilidad y la especificidad mediante una media armónica.

$$F1 = \frac{\text{sensibilidad} \cdot \text{especificidad}}{\text{sensibilidad} + \text{especificidad}} = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

Fuga de información

En el campo del aprendizaje automático las fugas de información («information leaks») se producen cuando, de alguna forma, se utiliza información perteneciente al conjunto de datos de test para entrenar un modelo, esto no significa únicamente utilizar instancias de test durante el entrenamiento, las fugas de información se pueden producir de forma mucho más sutil y ser difíciles de detectar. Por ejemplo, si se utilizase el conjunto de test para seleccionar las características durante el preprocesado se estaría produciendo una fuga de información.

Las fugas de información no son algo que debería ser siempre evitado, existen situaciones en las que son necesarias dependiendo del caso concreto.

Pero sí es importante que sean detectadas y tenidas en cuenta a la hora de analizar los resultados obtenidos, ya que, en caso contrario, se podría llegar a conclusiones equívocas (demasiado optimistas por ejemplo).

Validación cruzada

En un apartado anterior se ha visto que para evitar el sobreajuste se puede dividir la muestra de datos disponible en un conjunto de datos de entrenamiento y uno de test. Al realizar esta división es posible que, en especial al trabajar con muestras pequeñas, se realice una división que no sea útil para validar el modelo. Por ejemplo, si por casualidad se seleccionase un conjunto de test “fácil de predecir” con el conjunto de entrenamiento dado, se medirá un rendimiento mayor al real.

La solución a esto es la validación cruzada, que consiste en utilizar múltiples pares de conjuntos de validación y test, de tal forma que se entrenen y validen tantos modelos como cantidad de estos pares y cada instancia sea utilizada para la validación al menos una vez, el resultado final de la validación es la media de la métrica elegida de cada iteración.

Existen varios métodos para determinar el número de iteraciones a realizar y el tamaño de los grupos de entrenamiento y test en cada iteración, por ejemplo:

- ***Leave-One-Out***: Para una muestra de tamaño N , consiste en realizar N iteraciones de entrenamiento y validación utilizando cada vez una única instancia para el conjunto de test. Es un método que se acerca bastante al rendimiento que se obtendría si se utilizase la muestra entera como conjunto de entrenamiento, pero tiene el problema de que se van a tener que realizar N iteraciones, lo que implica un tiempo de computación necesario muy grande.
- ***K-fold***: Consiste en dividir la muestra en conjuntos de entrenamiento y test de proporciones $(K - 1)/K$ y $1/K$ respectivamente, realizando un total de K iteraciones. Además se puede realizar varias repeticiones de este tipo de validación cruzada (con distintas particiones aleatorias) para reducir la variabilidad del error obtenido.

Optimización de hiperparámetros

Los algoritmos de clasificación suelen tener parámetros que determinan la forma en la que se van a ajustar a los datos, por ejemplo, la tasa de

aprendizaje o el número de iteraciones máximas, estos valores se denominan hiperparámetros.

El proceso de búsqueda dentro del espacio que contiene todas las combinaciones posibles de hiperparámetros para un algoritmo concreto se denomina optimización de hiperparámetros.

En general, la optimización de hiperparámetros consiste en probar diferentes combinaciones de valores y realizar el proceso de entrenamiento y test mediante validación cruzada con una métrica apropiado hasta dar con la combinación más adecuada. Existen varias formas para realizar esto, como, por ejemplo:

- ***Grid search***: Proceso mediante el cual se definen los valores posibles que puede tomar cada parámetro y se prueba cada combinación de los mismos hasta encontrar la mejor. Es una búsqueda exhaustiva que puede dar lugar a una gran cantidad de combinaciones, lo que puede implicar un tiempo de computación muy grande dependiendo del número de parámetros y sus posibles valores.
- ***Randomized search***: Proceso similar al anterior en el que se especifican los valores posibles para cada parámetro, pero en este caso se realiza la búsqueda de forma aleatoria de forma que se limita el tiempo de computación necesario. Se puede utilizar para obtener una vista general del espacio de combinaciones de parámetros para realizar posteriormente una búsqueda más exhaustiva en un subespacio más pequeño.
- ***Técnicas de optimización***: Existen diversos métodos que intentan determinar la forma que tiene el rendimiento del modelo con respecto al espacio de combinaciones de parámetros posibles utilizando iteraciones anteriores para buscar los valores óptimos sin tener que probar todas las combinaciones como ocurre con *grid search*. Un ejemplo de esto es la Optimización Bayesiana [18] en la que se intenta predecir la forma que toma la métrica de evaluación en función de los hiperparámetros elegidos, y así realizar una búsqueda más efectiva.

4. Técnicas y herramientas

En el área de la inteligencia artificial, y en general en la informática, la conocida expresión «Pararse a hombros de gigantes» tiene mucho sentido, ya que es muy difícil, y contraproducente, crear algo por completo desde cero, siempre se utilizan herramientas, librerías o frameworks (abstracciones en general) creados por otras personas para acelerar el trabajo y permitir que al programador centrarse en la lógica de negocio que quiere implementar.

Este capítulo enumera las técnicas y herramientas de otras personas que se han utilizado durante la realización de este proyecto.

4.1. Técnicas

Esta sección detalla las distintas técnicas que fueron utilizadas durante el desarrollo de este proyecto.

CRISP-DM

«Cross-industry standard process for data mining», conocido como (CRISP-DM), es un modelo estándar abierto del proceso que describe los enfoques comunes que utilizan los expertos en minería de datos [11]. Este estándar surgió a finales de los 90 a partir de la inexistencia y necesidad de un enfoque ampliamente aceptado para la minería de datos. Para conseguir esto, CRISP-DM es agnóstico a la industria y herramientas que se utilicen, lo que permite que sea utilizado en un amplio abanico de situaciones, lo que facilitó su adopción.

CRISP-DM divide el ciclo de vida de un proyecto de minería de datos en las siguientes 6 fases.

1. **Entendimiento del negocio:** La fase más importante de cualquier proyecto de minería de datos, consiste en obtener un entendimiento sobre los objetivos del proyecto desde el punto de vista del negocio y redefinir estos objetivos como un problema de minería de datos diseñado para cumplir con estos objetivos.
2. **Entendimiento de los datos:** La fase de entendimiento de datos comienza con una recolección inicial de datos. A continuación, se procede a analizar los datos para obtener cierta familiaridad con los mismos, identificar problemas y realizar hipótesis iniciales sobre la información oculta que pueda existir.
3. **Preparación de los datos:** La fase de preparación de los datos cubre todas las actividades realizadas para la construcción del conjunto de datos final a partir de la información sin procesar que se va a utilizar para entrenar los modelos de aprendizaje. Esta fase incluye la selección, limpieza, integración y formateado de los datos.
4. **Modelado:** En esta fase se seleccionan varios algoritmos de aprendizaje, cuyos parámetros deben ser calibrados para ajustarse al conjunto de datos de forma óptima. Diferentes algoritmos imponen diferentes requisitos sobre el conjunto de datos sobre el que puede trabajar, por lo que es posible que se deba volver a la fase anterior para reajustar el conjunto de datos final. Esta fase incluye la selección de los algoritmos a utilizar, la creación de los modelos y la evaluación del rendimiento de los mismos.
5. **Evaluación:** Antes del despliegue del modelo final es de gran importancia reevaluar si el modelo cumple realmente con los objetivos de negocio establecidos en la primera fase. Aquí se deberá determinar cómo se van a utilizar los resultados de la minería de datos. Los procesos clave de esta fase son la evaluación de los resultados, la revisión del proceso utilizado y la determinación de los pasos siguientes.

Las fases anteriores tienen una naturaleza cíclica, los conocimientos obtenidos en cada ciclo de CRISP-DM se utilizan para replantear los objetivos de negocio y repetir el proceso con los nuevos conocimientos. Como muestra la siguiente figura.

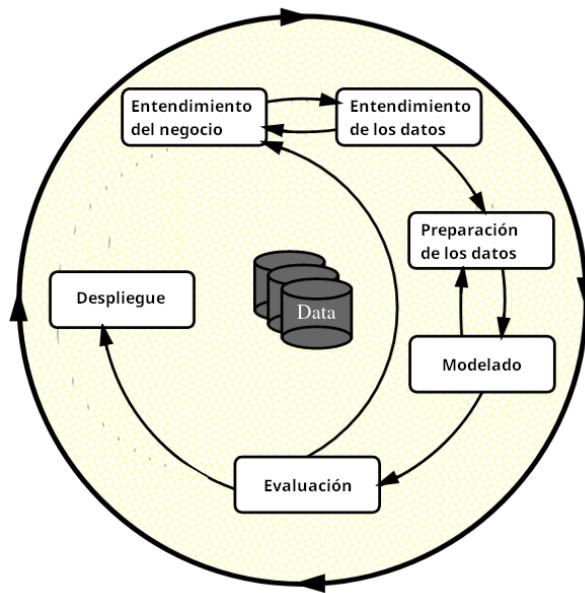


Figura 4.1: CRISP-DM [17]

4.2. Herramientas de desarrollo

Esta sección detalla las distintas herramientas, librerías y frameworks que fueron utilizadas durante el desarrollo de este proyecto.

Herramientas generales

Estas herramientas han sido usadas para realizar tareas de forma general dentro del proyecto.

Fedora Linux

Aunque al iniciar el proyecto se utilizó Windows 10 como sistema operativo principal, la mayor parte del desarrollo ha sido realizado en la distribución de Linux **Fedora**, en concreto versiones 37 y 38.

Fedora se caracteriza por que sus repositorios se actualizan regularmente para incluir las últimas versiones estables de la mayoría de programas, al contrario de lo que ocurriría en una distribución como Debian estable. Pese a esto, Fedora es considerada como una distribución bastante estable y, durante el desarrollo de este proyecto, no se han encontrado problemas fuera de lo común.

Han sido instalados de forma nativa Python 3.9 para la fase de investigación, [TeX Live](#) para la compilación de esta documentación y [Docker](#) para el desarrollo de la aplicación web.

Python

Python es un lenguaje de programación diseñado con el objetivo de ser fácil de escribir. Para lograr esto Python incluye de forma nativa una gran multitud de funciones de uso común (como funciones de manipulación de cadenas), mientras que en otros lenguajes debería encargarse el programador de implementarlas o utilizar librerías de terceros. Esto tiene muchas ventajas, pero también algunas desventajas, como, por ejemplo, que se pierde el control sobre la ejecución del programa que podría dar un lenguaje como C o C++, lo que significa que, en general, un mismo programa escrito completamente en Python va a ser más lento que su contrapartida escrita en un lenguaje de programación de más bajo nivel en el que se puedan especificar detalles como, por ejemplo, la gestión de memoria.

Pero Python tiene otra cualidad muy importante, y es su facilidad para interactuar con código escrito en C o C++ mediante lo que se conoce como *bindings* o *wrappers*. Esto permite al programador tener mayor control sobre las partes más críticas de un programa, escribiéndolas en C por ejemplo, mientras que las partes de menor importancia se pueden escribir rápidamente en Python.

Un ejemplo perfecto de lo anterior es la librería [NumPy](#), utilizada para la gestión de matrices, entre muchas otras cosas. Esta librería implementa su funcionalidad principal en C y expone una interfaz en Python para facilitar su uso.

Este patrón se repite una y otra vez en el campo del aprendizaje automático, donde los algoritmos de aprendizaje se implementan a bajo nivel para que su ejecución sea lo más rápida posible, mientras que se expone una interfaz a alto nivel en Python para interactuar con los mismos. Ejemplos de esto son las librerías [PyTorch](#) y [TensorFlow](#).

Git

Actualmente Git es el sistema de control de versiones más extendido en proyectos de código abierto [8], en este proyecto se ha usado para gestionar y guardar de forma segura los cambios que se han ido realizado con el tiempo en la plataforma GitHub.

Aunque existen varias herramientas con interfaces gráficas que envuelven el funcionamiento de Git (GitKraken, GitHub Desktop, etc.) para este proyecto se utilizó directamente el comando `git` desde una terminal de Linux.

Make

A través del proyecto hay comandos de cierta complejidad que se repiten con bastante frecuencia (*e.g.* formatear el código, lanzar contenedores, limpiar contenedores, etc.), para facilitar el uso de estos comandos se ha utilizado la utilidad GNU Make, que viene por defecto en la mayoría de sistemas operativos de tipo Linux y se puede instalar tanto en Windows como en MacOS.

El funcionamiento de Make consiste en la creación de un archivo denominado `Makefile` que contiene diferentes comandos y secuencias de comandos con las dependencias que existen entre los mismos.

Por ejemplo:

```
format:  
    autoflake src  
    isort src  
    black src
```

Un fichero `Makefile` con el contenido anterior permite ejecutar el comando `make format` (siempre desde el mismo directorio en el que se encuentra `Makefile`) para ejecutar la secuencia de comandos denominada `format`, que, en este caso ejecutará varios comandos para realizar un formato del código.

PyCharm

PyCharm es un Entorno de Desarrollo Integrado (IDE) creado y mantenido por JetBrains, una compañía que se dedica a crear software y cuyos productos más representativos son multitud de IDEs para diferentes lenguajes de programación (IntelliJ (Java), CLion (C y C++), GoLand (Go), etc.).

El objetivo de PyCharm es facilitar el desarrollo de aplicaciones que utilizan Python y da soporte amplio tanto para los habituales ficheros `.py` como para `.ipynb` (Jupyter Notebook).

En este proyecto PyCharm fue utilizado para esas desarrollar la librería PAD-DEL durante la fase de investigación, que contiene el código de Python común a los Notebooks de Jupyter utilizados durante la fase de investigación.

Visual Studio Code

Visual Studio Code (VS Code) es un editor de texto orientado al desarrollo creado por Microsoft. Gracias a su gran extensibilidad es posible utilizarlo cómodamente en casi cualquier escenario.

Fue utilizado en la creación de esta documentación en conjunto con la extensión [LATEX Workshop](#) que integra distintos compiladores de LATEX (pdfTeX, XeTeX, LuaTeX, etc.) dentro del programa y añade otras funcionalidades que hacen trabajar LATEX mucho más conveniente.

Además, gracias a la extensión [Dev Containers](#), VS Code puede ser lanzado desde contenedores de Docker, permitiendo interactuar con las utilidades del contenedor. Gracias a esto VS Code ha sido la herramienta utilizada para desarrollar por completo la aplicación web.

Formateado de código Python

Se han utilizado las siguientes herramientas para realizar el formateado del código escrito en Python de forma automática:

- **Autoflake**: herramienta utilizada para la eliminación de sentencias `import` y variables innecesarias.
- **Isort**: herramienta para la reordenación de sentencias `import`.
- **Black**: formateador de código caracterizado por un enfoque en convenciones, por esto permite una configuración muy limitada por parte del usuario.

Herramientas para la fase de experimentación

Estas herramientas han sido de gran utilidad durante la fase de experimentación para realizar las distintas tareas como extracción de características y entrenamiento de modelos.

Jupyter Notebook

Jupyter Notebook es una herramienta que permite crear un archivo con extensión `.ipynb`, en el que el código Python se divide en “celdas” que

se pueden ejecutar individualmente, pero manteniendo y compartiendo las variables establecidas de forma global entre todas estas celdas.

Esta herramienta fue de especial utilidad durante la fase de experimentación, ya que permite observar detenidamente los cambios que se producen entre celdas y, así, encontrar problemas sin tener que ejecutar el código completo desde el principio cada vez que se realice un cambio, como ocurriría con un *debugger*.

Además de lo anterior Jupyter Notebook permite documentar de forma clara el código mediante otro tipo de celda, en el que se puede escribir código en Markdown, que es, en cierto modo, una alternativa menos verbosa a HTML.

OpenCV

OpenCV es una librería de visión artificial que contiene varias utilidades para interactuar con archivos de vídeo e imágenes de forma similar a como se trabajarían en un programa como MATLAB. OpenCV está escrito en C y C++ pero tiene *bindings* que permiten interactuar con sus interfaces desde Python.

En este proyecto se usa para leer y decodificar los archivos de vídeo utilizados durante la fase de investigación.

MediaPipe

MediaPipe es una librería de aprendizaje automático mantenida por Google que da acceso a varios modelos preentrenados que se centran en la detección de características corporales a partir de flujos de imágenes, en este caso se utilizó MediaPipe Hands [20], cuyo objetivo es extraer varios puntos que definen un esqueleto ligeramente simplificado de la mano humana.

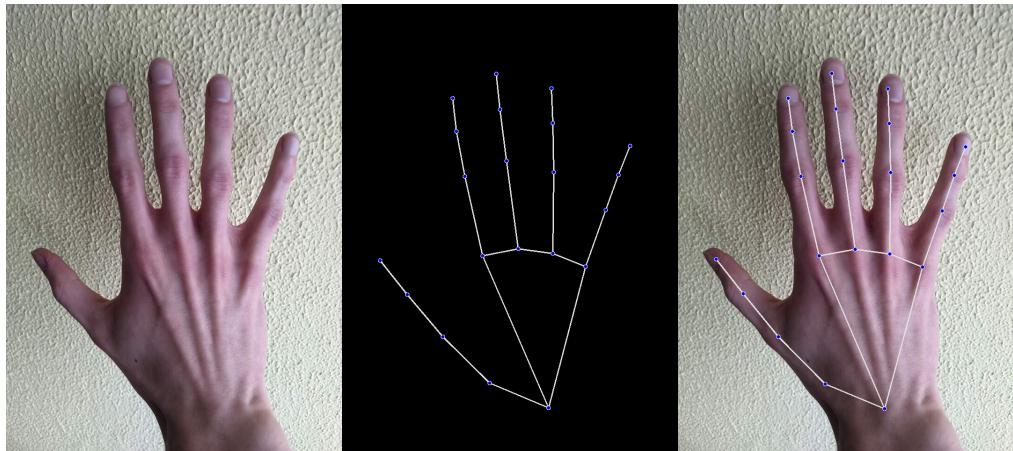


Figura 4.2: MediaPipe Hands

Como se puede ver el resultado es bastante bueno, además, al utilizar vídeos MediaPipe utiliza la información de los fotogramas anteriores para determinar de forma más exacta la pose en los fotogramas posteriores. Esto es muy útil para mejorar la precisión de las poses obtenidas, pero tiene la pega de que, al necesitar información de pasos previos, el proceso de extracción de poses no se puede paralelizar.

TSFresh

TSFresh [2] es una librería de Python que se encarga del preprocesado y la extracción de características de series temporales.

Una serie temporal es una secuencia ordenada de valores con una componente temporal asociada. Estas secuencias por sí solas no son de mucha utilidad para entrenar algunos modelos, por lo que en la fase de preprocesado se debe realizar una extracción de características sobre las mismas y así obtener características numéricas que sí se pueden utilizar para entrenar un modelo.

TSFresh establece una forma estándar para extraer estas características, permitiendo extraer características propias además de las, aproximadamente, 700 características que extrae por defecto.

Herramientas para el desarrollo de la aplicación

Estas utilidades fueron usadas para crear la aplicación web que se encuentra en [paddle.catalin.sh](#).

Docker

Docker permite la creación y gestión de contenedores, un contenedor se puede ver como un sistema operativo dentro del anfitrión, pero sin llegar a ser una máquina virtual, debido a que se utiliza el *kernel* de Linux del sistema anfitrión y los archivos del contenedor existen en el sistema de archivos del anfitrión. Esto hace que la diferencia de rendimiento entre utilizar un contenedor y ejecutar programas de forma nativa sea mínima.

Un contenedor de Docker se basa en lo que se conoce como una imagen, que es un sistema de archivos preconfigurado para el propósito del contenedor. Se puede encontrar una gran cantidad de imágenes en [Docker Hub](#), como, por ejemplo, postgres, node o python que ya incluyen los ejecutables para utilizar cada herramienta respectiva. Sin embargo, es posible que se necesite mayor control sobre la configuración de la imagen sobre la que se va a construir el contenedor, en estos casos se pueden utilizar lo que se conoce como *Dockerfile* para crear una imagen tomando como base otra imagen ya existente.

Un *Dockerfile* es un archivo que determina la imagen de base a utilizar (cualquier imagen de [Docker Hub](#) u otros sitios) y los comandos a ejecutar para crear la imagen que se desea obtener. Si se desease empezar de cero se puede utilizar una imagen de Ubuntu o Debian como base.

Además de *Dockerfile* existe otro tipo de archivo denominado *docker-compose.yml* utilizado para coordinar sistemas compuestos por multiples contenedores que interactúan entre sí. La aplicación web de este proyecto utiliza *docker-compose.yml* para determinar las interacciones entre los cuatro contenedores creados (web, API, proxy inverso y base de datos).

Caddy

En las fases iniciales de desarrollo se decidió que la API debería estar alojada en el subdominio `api.`, mientras el dominio principal debería llevar a la web (SvelteKit). Esto significa que van a existir dos servicios funcionando de forma paralela (Web y API), y que dependiendo del dominio al que se realice una petición, ésta debería ser redirigida a un servicio u otro.

Esta situación es uno de los casos de uso idóneos para un proxy inverso, que es un servicio que recibe peticiones HTTP, las altera si es necesario, y las redirige al servicio de destino (el cual se determina mediante ciertas reglas de redirección).

En un principio se optó por utilizar Nginx para implementar este proxy inverso con muy buenos resultados gracias a su fácil configuración, rendimiento y simplicidad. Pero surgieron problemas al añadir soporte para utilizar certificados SSL y realizar peticiones seguras mediante HTTPS, ya que era deseable una solución que gestionase de forma automática la petición y renovación de estos certificados y Nginx, debido a su simplicidad no dispone de esta funcionalidad.

Se probaron dos alternativas que solucionaban este problema, Traefik y Caddy, ambas escritas en GoLang. Traefik dispone de una gran cantidad de funcionalidades incluyendo gestión de certificados SSL, entre muchas otras, como un panel de control desde el que visualizar estadísticas sobre el nivel de uso del servicio.

La otra alternativa, Caddy, es mucho más simple, pero realiza la gestión de certificados de forma automática por defecto, e incluso permite el uso de certificados *self-signed* (firmados por uno mismo, en lugar de por una entidad certificadora). Esto es de gran utilidad durante el desarrollo para asegurar que el funcionamiento de la aplicación va a seguir siendo el mismo en el entorno final de producción (ya que hay ligeras diferencias dependiendo de si un navegador utiliza HTTP o HTTPS).

Es por todo lo anterior que, en última instancia, se decidió usar Caddy.

FastAPI

FastAPI es un *framework* de Python utilizado para crear APIs de tipo REST que se caracteriza por permitir al programador centrar sus esfuerzos en el código relacionado con la lógica de negocio al realizar la mayoría de acciones comunes entre diferentes APIs por defecto (como la creación de documentación, validación de tipos, serialización y deserialización de peticiones, etc.).

En este proyecto se ha creado una API REST para permitir la interacción con el código de Python utilizado para crear, entrenar y utilizar los modelos desde cualquier dispositivo y lenguaje de programación (siempre y cuando permita realizar peticiones HTTP). Con esto se consigue que en el futuro se puedan crear nuevas formas para interactuar con los modelos sin necesidad de alterar el código ya existente. Se podría, por ejemplo, crear una aplicación móvil para facilitar la toma y subida de vídeos. Además estas nuevas aplicaciones podrían estar creadas por personas ajenas al proyecto que necesiten una implementación diferente para su caso de uso específico.

SvelteKit

Para el frontend de la aplicación, es decir, todo lo que va a ver un usuario de la web (HTML, CSS y JavaScript), se ha utilizado **SvelteKit**, que es un framework de JavaScript basado en Svelte. Es similar a otros frameworks más populares como React o Angular, y se caracteriza por tener una sintaxis muy clara y simple además de reducir el tamaño de los archivos que se envían al navegador del cliente, haciendo que, en general, las páginas creadas con SvelteKit sean más rápidas que sus contrapartidas.

TypeScript

JavaScript es un lenguaje de programación que no dispone de tipado estático, lo cual hace muy difícil detectar problemas fuera del tiempo de ejecución.

TypeScript es un lenguaje de programación desarrollado por Microsoft que utiliza una sintaxis de base muy similar a la de JavaScript, pero añadiendo tipado estático. Los navegadores no “entienden” TypeScript, por lo que éste debe ser compilado a JavaScript antes de ser enviado al usuario. Durante esta compilación se pueden detectar multitud de errores gracias al tipado.

Typesafe i18n

Typesafe i18n es una librería escrita en TypeScript que se puede utilizar para internacionalizar aplicaciones web, es decir, añadir soporte para más idiomas, de forma simple y mediante unas convenciones predefinidas.

Para ello se crea un archivo para cada idioma al que se de soporte, en el que se definen las variables con las cadenas de texto utilizadas a través de la web en el idioma respectivo. Cada uno de estos archivos debe contener las mismas variables.

A continuación, en los lugares donde se usen cadenas de texto se debe importar la variable `LL` que es un objeto con todas las traducciones que se puede usar en la página respectiva para establecer las cadenas de texto dependiendo del idioma.

Axios

Axios es una librería de JavaScript que envuelve la función nativa `fetch` para hacerla más amigable y fácil de usar. Se caracteriza por soportar tipos de TypeScript y simplificar la realización de peticiones HTTP.

En este proyecto se ha utilizado esta librería para realizar peticiones REST a la API desde el navegador del usuario. Para ello se han creado multitud de funciones que reflejan los *endpoints* de la API y, así, permitir su uso desde el resto del sitio web.

Tailwind CSS

[Tailwind CSS](#) es una librería de clases de CSS, su funcionamiento es similar a [Bootstrap](#) en cuanto a que existen multitud de clases que realizan diferentes acciones. La diferencia está en que Tailwind CSS es mucho más modular que Bootstrap, es decir, las clases realizan cambios muy pequeños sobre los objetos a los que afectan. Esto hace que Tailwind sea más lento de escribir que su contrapartida, pero al mismo tiempo permite un nivel mucho mayor de configuración sobre el diseño de una página web.

Prettier

[Prettier](#) es un formateador de código, creado concretamente para dar un formato consistente a archivos de HTML, CSS, JavaScript, TypeScript, etc.

ChromeVox

Un aspecto que se tuvo muy en cuenta durante el desarrollo del sitio web es la accesibilidad, en especial para usuarios con visibilidad reducida que dependen de lectores de pantalla para descubrir el contenido de la web.

Para comprobar cómo interactuaría un lector de pantalla con la aplicación se ha utilizado la extensión de Google Chrome [ChromeVox](#).

5. Aspectos relevantes del desarrollo del proyecto

Este apartado contiene algunos aspectos relevantes sobre el desarrollo del proyecto, incluyendo las razones detrás de las decisiones tomadas durante este y el impacto que dichas decisiones han tenido sobre los resultados obtenidos.

5.1. Fase de experimentación

El conjunto de datos disponible durante la fase de experimentación consisten en 158 muestras con los siguientes «atributos»:

- Si la persona correspondiente a la instancia padece la enfermedad de Parkinson. Es lo que se busca predecir.
- Archivo de vídeo de una mano de la persona realizando la prueba de «finger-tapping».
- Mano que aparece en el vídeo (izquierda o derecha).
- Fecha en la que se ha tomado el vídeo (irrelevante).
- Edad de la persona.
- Sexo de la persona.
- Mano dominante de la persona.

Cabe destacar que se ha tomado una muestra de cada mano de cada persona. Por lo que este conjunto de datos ha sido obtenido de 79 personas. Aún así, se ha tomado cada mano como una instancia distinta con el objetivo de tener más instancias con las que trabajar.

Además, se dispone de 69 instancias de personas que padecen la enfermedad de Parkinson y 89 instancias de personas sanas, por lo que se está trabajando con un conjunto de datos ligeramente desbalanceado. Esto deberá tenerse en cuenta para que no influya en los resultados obtenidos, por ejemplo, seleccionando métricas adecuadas.

Procesado de vídeo

El objetivo de la fase de experimentación es crear un sistema que tenga como entrada los atributos antes mencionados (salvo el objetivo) y por salida si esos atributos corresponden con una persona que padece la enfermedad de Parkinson (posiblemente junto con el grado de confianza de la predicción).

Los algoritmos de aprendizaje automático típicamente están diseñados para trabajar con atributos numéricos, por lo que los atributos disponibles deben ser transformados durante el preprocesado. Esto se puede realizar en la mayoría de casos mediante simples métodos de minería de datos, como, por ejemplo, la codificación de variables categóricas. Pero hay un problema, el archivo de vídeo.

En este caso, una imagen es una matriz bidimensional donde cada celda contiene una 3-tupla representando los valores rojo, verde y azul del pixel con el que se corresponde. Un vídeo es una secuencia de imágenes junto con información adicional, es este caso sólamente es relevante la tasa de refresco, que es la frecuencia a la que ha sido capturada cada imagen. Esto es de gran importancia para dar una componente temporal al vídeo.

Con lo anterior se puede ver que un vídeo no se puede utilizar de forma directa, sino que deberá pasar por una fase multietapa de preprocesado para ser transformado a un conjunto de valores numéricos que lo describen.

Extracción de «landmarks»

Una «landmark» en el campo de la visión artificial es un punto de referencia que se corresponde con un objeto físico. El objetivo de esta fase es convertir el vídeo de la prueba de «finger-tapping» a una secuencia o serie temporal de conjuntos de «landmarks» los cuales representan la posición de la mano durante un instante concreto.

El primer paso para lograr esto es leer las imágenes del archivo de vídeo, esto ha sido logrado mediante la librería de «bindings» de OpenCV para Python. OpenCV es una librería de uso general para realizar tareas relacionadas con la visión artificial, y permite leer multitud de formatos de vídeo mediante un flujo de imágenes.

Se ha tomado especial precaución para no cargar en memoria todas las imágenes del vídeo, sino ir leyéndolas cuando son necesarias. Un archivo de vídeo ocupa relativamente poco espacio en disco debido a que está codificado mediante un «codec» determinado, lo que reduce el tamaño del archivo. Pero, para poder trabajar con las imágenes del vídeo, este debe ser decodificado.

La magnitud de lo anterior se ve claramente con el siguiente ejemplo. Un vídeo con una resolución de 1920x1080 a 30 fotogramas por segundo codificado mediante el codec H264 que ocupa 50,7MB, decodificado, ocuparía en torno a los 3,7GB.

A continuación se deben extraer las «landmarks» de las imágenes leidas. Esto ha sido logrado mediante la librería Mediapipe Hands [20], que permite extraer los siguientes puntos a partir de una imagen:

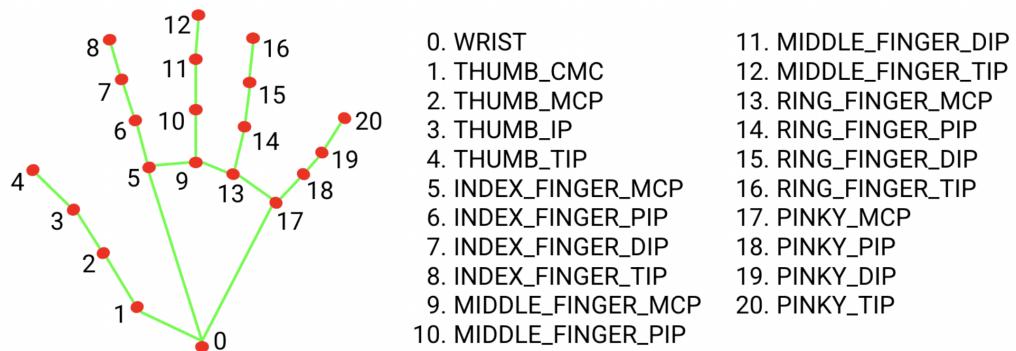


Figura 5.1: Mediapipe Hands «Landmarks»

Debido a la perspectiva desde la que se han tomados los vídeos de la prueba de «finger-tapping» hay algunas «landmarks» que quedan ocultas en varias ocasiones, lo que hace su posición detectada poco fiable. Por lo que se desechan las «landmarks» en el intervalo [9-17] para solo tomar en cuenta la parte «frontal» de la mano.

Mediapipe Hands ofrece la posibilidad de extraer «landmarks» de las imágenes como si estas no tuvieran relación entre sí o realizar la extracción teniendo en cuenta que se trata de un vídeo, así teniendo en cuenta resultados anteriores para determinar de forma más precisa las «landmarks» del

fotograma actual. Esto tiene la gran desventaja de que hace imposible la paralelización de este proceso (debido a que cada extracción depende de los resultados anteriores).

Lo anterior no tiene un gran efecto durante la fase de investigación, ya que sí se puede paralelizar la extracción de «landmarks» de cada vídeo. Donde sí tiene efecto es sobre la aplicación web, empeorando la experiencia de los usuarios.

En última instancia se tomó la decisión de no usar paralelización, y así obtener resultados más precisos. Una alternativa planteada fue realizar la extracción en paralelo solamente en la aplicación, pero esto se descartó ya que esto podría introducir inconsistencias y variables adicionales que deberían ser tomadas en cuenta.

Extracción de series temporales

Tras el proceso descrito en el apartado anterior se dispone una secuencia de datos similar a la siguiente:

WRIST		THUMB_CMC
1970-01-01 00:00:00.000000000	(0.7664212584495544, 0.5941680669784546, -1.0810111916725873e-06)	(0.6852301955223083, 0.59
1970-01-01 00:00:00.033289017	(0.75729900598526, 0.5936159491539001, -8.921984999687993e-07)	(0.685229480266571, 0.59,
1970-01-01 00:00:00.066578834	(0.7572722434997559, 0.5920841693878174, -9.016254125526757e-07)	(0.6872291564941406, 0.59
1970-01-01 00:00:00.099867051	(0.760816216468811, 0.5914657711982727, -4.4887269723403733e-07)	(0.6998681426048279, 0.59
1970-01-01 00:00:00.133156668	(0.7623580694198608, 0.5860837697982788, -2.6346916115471686e-07)	(0.7010113596916199, 0.59
1970-01-01 00:00:00.166445084	(0.7705487012863159, 0.5813321471214294, 5.089297587801411e-07)	(0.7006314992904663, 0.59
1970-01-01 00:00:00.199734103	(0.7820069193840027, 0.5774965286254883, 5.014093744648562e-07)	(0.7123239636421204, 0.59
1970-01-01 00:00:00.233023120	(0.7972635030746464, 0.5691201090812683, 6.951975706215308e-07)	(0.7247089147567749, 0.59
1970-01-01 00:00:00.266312137	(0.8012514114379883, 0.56582111120224, 7.189794359874213e-07)	(0.7269166707992554, 0.59
1970-01-01 00:00:00.29901155	(0.7900755405426025, 0.5591111779212952, 7.520919780290569e-07)	(0.722050666809082, 0.59

Figura 5.2: Serie temporal de «landmarks»

Cada fila representa la pose de la mano durante un fotograma concreto. Se puede observar que el índice es una fecha completa en lugar de otras opciones más convenientes como los segundos desde el inicio del vídeo o lo que se conoce como «timestamp». Esto es una imposición de TSFresh para poder realizar la extracción de características teniendo en cuenta esta componente temporal.

Cada celda representa el punto en el espacio en el que se encuentra la «landmark» correspondiente en un instante de tiempo.

Aunque estas series temporales son una simplificación muy grande de lo vídeos iniciales, aún no son del todo útiles. Al ser información sobre la posición en el espacio, es dependiente de la forma en la que ha sido tomada

la grabación. En un entorno controlado esto no sería mucho problema, pero en este caso, como se desea hacer el uso del sistema accesible a cualquier persona sin necesitar equipamiento especial, se deben buscar alternativas.

En este caso se ha decidido tomar el ángulo entre el ángulo entre la punta del dedo pulgar, la muñeca y la punta del dedo índice. Esta medida es útil ya que no depende de cómo se ha tomado el vídeo ni del tamaño de la mano de la persona. Quedando como resultado una serie temporal como la siguiente:

	angle
1970-01-01 00:00:00.000000000	0.041790
1970-01-01 00:00:00.033289017	0.050749
1970-01-01 00:00:00.066578034	0.061877
1970-01-01 00:00:00.099867051	0.219051
1970-01-01 00:00:00.133156068	0.393220

Figura 5.3: Serie temporal de ángulos

La unidad usada en estos ángulos es el radián.

Extracción de características

El último paso para convertir un archivo de vídeo a atributos útiles para entrenar un modelo de aprendizaje es extraer características que describan la serie temporal obtenida. Algunos de estos atributos han sido extraídos de forma específica teniendo en cuenta este problema concreto, mientras que otros (la mayoría) son atributos genéricos de series temporales.

Detección de toques y amplitudes Para varios atributos específicos al problema han sido utilizados los instantes de tiempo en los que se producen toques (el pulgar y el índice tocan) y amplitudes máximas en el movimiento de «finger-tapping».

La detección de estos instantes ha sido realizada mediante una búsqueda de máximos gracias a la función `find_peaks` de la librería `scipy`. Pero esta función por si sola no da muy buenos resultados.

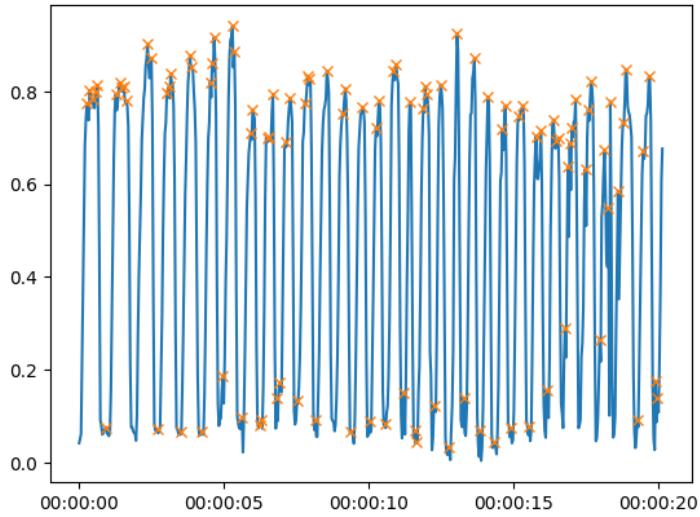


Figura 5.4: Amplitudes

El eje horizontal representa el tiempo que ha transcurrido desde el inicio del vídeo. El eje vertical representa el ángulo en radianes entre el dedo índice, la muñeca y el pulgar.

Se puede apreciar que obtener todos los máximos locales no es la mejor opción, debe incorporarse algún criterio para determinar si un máximo debería tenerse en cuenta o no. La función `find_peaks` tiene algunos parámetros para este propósito.

Tras probar varias opciones y comprobar manualmente los resultados obtenidos sobre diferentes vídeos, se determinó que la mejor opción es establecer la prominencia mínima (parámetro `prominence` de `find_peaks`) que debe tener un máximo local para ser considerado.

La prominencia se define como la medida que determina cuánto destaca un máximo con respecto a sus máximos más cercanos. Se probaron varios valores estáticos, pero como cada vídeo tiene características diferentes era necesario utilizar algo más dinámico. La desviación típica se comportaba bastante bien, pero cuando la amplitud variaba con el tiempo (síntoma de la enfermedad de Parkinson) los máximos de menor magnitud no eran detectados.

Lo anterior fue solucionado mediante el uso de una desviación típica móvil. Se establece una ventana que determina la «zona» a la que pertenece un valor y se utiliza la desviación típica de esa «zona» como prominencia mínima para que el punto pueda ser considerado un máximo.

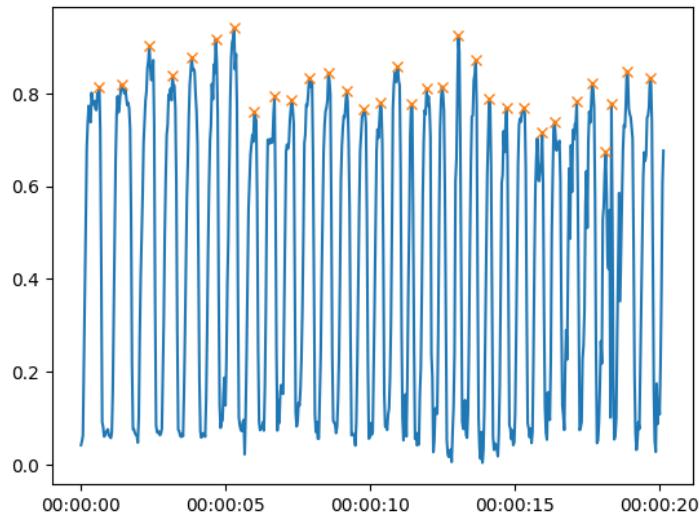


Figura 5.5: Amplitudes mediante desviación típica móvil

Se puede apreciar que los resultados son ampliamente mejores. Incluso cuando la amplitud decrece con el tiempo:

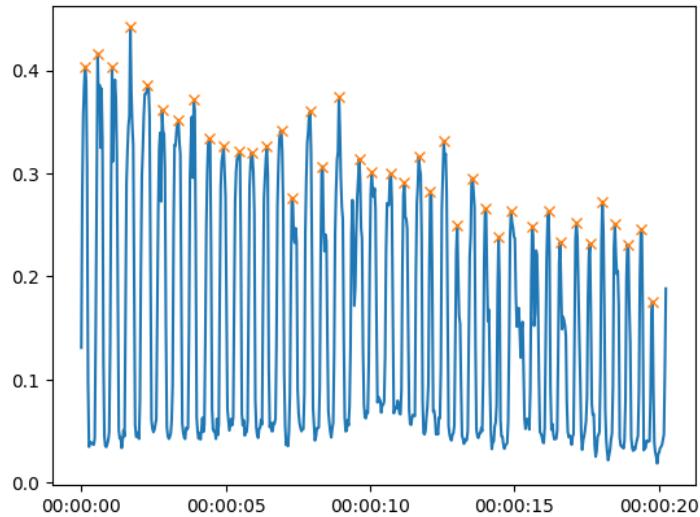


Figura 5.6: Amplitudes mediante desviación típica móvil con la enfermedad de Parkinson

Aunque si el movimiento es errático, los resultados también lo son, como es de esperar.

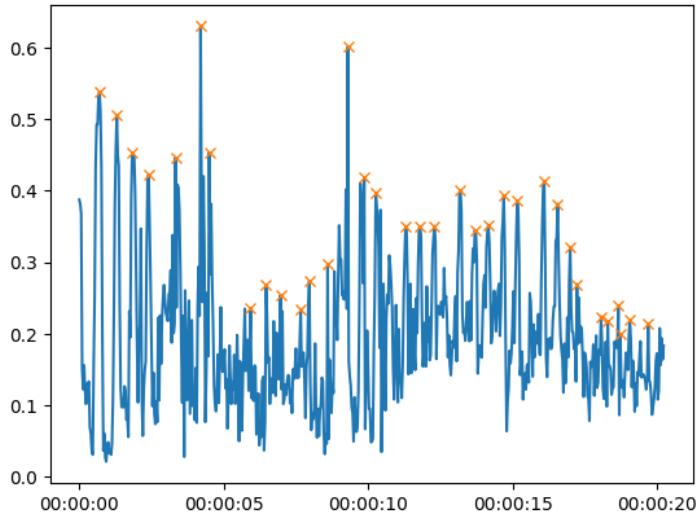


Figura 5.7: Amplitudes mediante desviación típica móvil con movimiento errático

Para detectar los toques (mínimos) es tan simple como utilizar la serie de ángulos negada para detectar los puntos donde se encuentran los máximos. Además se ha establecido un umbral para que solo se consideren toques si se encuentra entre 0 y 0,1 radianes.

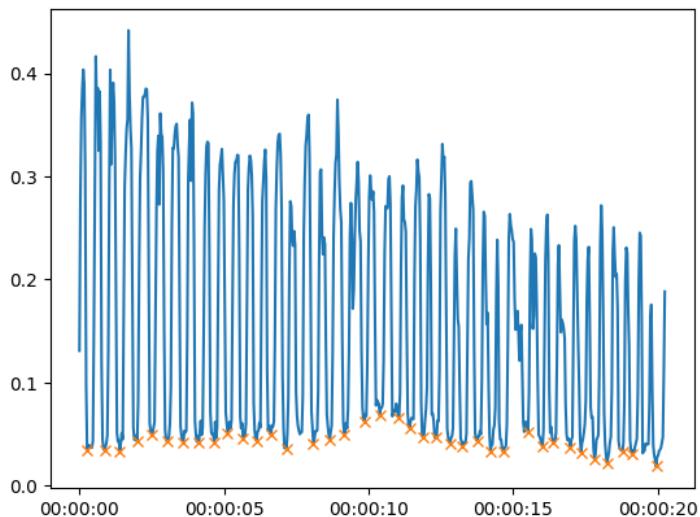


Figura 5.8: Amplitudes mediante desviación típica móvil con movimiento errático

Características detectadas A continuación se van a detallar algunas de las características extraídas de mayor interés. Estos valores ya son atributos que describen el vídeo, por lo que pueden ser utilizados para la generación de un modelo.

- **Velocidad media del movimiento:** Una característica de la bradicinesia es la ralentización del movimiento. Esto va a ser reflejado en la velocidad media (en radianes por segundo) que tiene el movimiento
- **Frecuencia de amplitudes y toques:** Dos magnitudes que miden cuantos toques o amplitudes se realizan por segundo de media durante el movimineto.
- **Amplitud media:** Otra característica de la enfermedad de Parkinson es la debilidad, lo que se puede ver reflejado sobre la amplitud media del movimiento. Es simplemente la media de las amplitudes antes detectadas.

Además de las anteriores, se extrajeron 794 características genéricas mediante la librería TSFresh.

Limpieza de datos

En este punto se dispone de un conjunto de 158 instancias y unos 800 atributos, algunos de ellos categóricos, estos deberán ser codificados para que sean atributos numéricos, con los que se trabaja mejor.

Además, se determinó que todas las instancias deberán tener un tiempo de detección de la mano realizando la prueba de «finger-tapping» mínimo de 15 segundos para garantizar en cierto modo la precisión de los atributos extraídos y poder compara instancias con el mínimo número de influencias sobre las que no se tiene control.

Debido a la decisión anterior se pasó de disponer de 158 instancias a únicamente 156. Esto es desafortunado debido al ya pequeño número de muestras, pero se consideró una pérdida que merece la pena con el fin de obtener mejores resultados.

Por último se sustituyeron las característica «mano dominante» y «mano que aparece en el vídeo» por una única característica «es la mano que aparece en el vídeo la dominante» con el objetivo de reducir la conocida «maldición de la dimensionalidad» al mismo tiempo que se mantiene la información relevante.

Normalización

Algunos algoritmos se ven muy afectados por la escala en la que se presentan los atributos, uno de los casos más claros es «*k* vecinos más cercanos».

Debido a esto es importante normalizar los atributos. Se han probado varios métodos y encontrado que la normalización por cuantiles funciona ligeramente mejor que sus contrapartidas.

La normalización por cuantiles se caracteriza por hacer que los estadísticos de cada atributo sean idénticos a los de una distribución concreta (en este caso una distribución uniforme). Esto se consigue ignorando la magnitud de un atributo y teniendo en cuenta únicamente su posición en la secuencia de todas las instancias ordenadas según el atributo que se está normalizando.

Al no tener en cuenta la magnitud, la transformación anterior es de gran utilidad para tratar lo que se conoce como «outliers», que son instancias con atributos que se alejan mucho de la regla.

Búsqueda del mejor modelo

En este punto se dispone de 4 conjuntos de datos sobre 156 instancias:

- **Atributos iniciales:** Atributos de los que se disponía al inicio, como la edad o el sexo de la persona.
- **Atributos extraídos de vídeo propios:** Son los atributos cuya extracción ha sido descrita en los apartados anteriores, como la velocidad del movimiento o la frecuencia de toques.
- **Atributos extraídos de vídeo con TSFresh:** Son los 794 atributos del vídeo extraídos mediante TSFresh.
- **Variable objetivo:** Si la instancia corresponde con una persona que padece la enfermedad de Parkinson. Es lo que se intenta predecir.

A priori es muy difícil o incluso imposible saber cómo van a comportarse diferentes algoritmos al enfrentarse a distintos subconjuntos de la información disponible. Por lo que la única opción es probar diferentes configuraciones con la esperanza de encontrar una que se adapte considerablemente bien a los datos. En este sentido el aprendizaje automático es más un arte que una ciencia exacta.

En este caso se ha realizado una búsqueda exhaustiva (o *grid-search*) con las siguientes variables:

- **Conjunto de datos:** El conjunto de datos utilizado para el entrenamiento ha sido cambiado de dos formas distintas:
 - Distintas combinaciones del los conjuntos de datos anteriores.
 - Selección de un distinto número de los atributos según su relevancia estimada mediante TSFresh.
- **Algoritmo de aprendizaje:** Se han probado varios algoritmos de aprendizaje. (Máquinas de vectores de soporte, Naive Bayes, k vecinos más cercanos, Perceptrón multicapa, AdaBoost, etc.)
- **Parámetros de los algoritmos:** Se ha variado la configuración de cada uno de los algoritmos anteriores. Este proceso se denomina optimización de hiperparámetros.

Además, existen combinaciones de parámetros sin sentido, por ejemplo, variar el grado de la función para una máquina de vectores de soporte con «kernel» lineal. Esto se debe tener en cuenta para no aumentar el tiempo requerido de forma inútil.

Cabe destacar que la selección de características se realiza teniendo en cuenta la totalidad del conjunto de datos, no solo el de entrenamiento. Esto produce fugas de información durante la validación cruzada ya que las características seleccionadas son la consideradas mejores tanto para el conjunto de entrenamiento como el de test.

Una opción para solventar lo anterior es realizar la selección de características en cada «fold» de la validación cruzada, usando únicamente el conjunto de entrenamiento. Pero esta operación es muy costosa al trabajar con cientos de atributos. Este cambio incrementa la duración de la búsqueda «grid-search» en varios órdenes de magnitud. Se tomó la decisión de aceptar las fugas de información ya que lo que se pretende es comparar el rendimiento de distintos modelos, por lo que los resultados van a seguir siendo relevantes.

Resultados

Tras realizar la búsqueda se ha obtenido una tabla de dimensiones bastante grandes (con una fila por cada combinación de variables probada) con los resultados obtenidos.

En este apartado se exponen los resultados obtenidos además de algunos detalles interesantes.

5.2. Fase de desarrollo de la aplicación

Se ha creado una aplicación web que permite interactuar con un modelo de aprendizaje automático de forma simple. Debido a las características del programa una aplicación móvil podría ser una buena alternativa para facilitar la toma del vídeo. Por esto se quiso facilitar en la medida de lo posible la extensión futura del sistema con cambios mínimos de funcionalidad preexistente.

Para cumplir con lo anterior se decidió crear una API (Application Programming Interface) REST en la que se implemente toda la funcionalidad de la aplicación junto con una página web que sirva como «frontend» que se comunique con la API mediante peticiones HTTP. Este enfoque deja la puerta abierta a futuras extensiones (como una aplicación móvil) sobre el sistema sin la necesidad de realizar cambios sobre el código ya existente. Estas extensiones podrían ser realizadas por cualquier persona, incluso si no ha estado involucrada en este proyecto con antelación.

Docker

Para facilitar la creación del entorno de desarrollo y el despliegue de la aplicación se ha empleado la popular herramienta Docker. Esto permite la creación de «contenedores» de forma determinista, es decir, obteniendo siempre un sistema idéntico, solucionando el típico problema de «funciona en mi máquina».

Para esta aplicación se emplean cuatro contenedores:

- API (Basado en [FastAPI](#))
- Aplicación web (Basado en [SvelteKit](#))
- Proxy inverso (Basado en [Caddy](#))
- Base de datos (Basado en [PostgreSQL](#))

Proxy inverso

Un proxy inverso permite redireccionar peticiones de forma dinámica al destino adecuado. Esto es generalmente usado en lo que se conoce como balance de carga (*load balancing*).

Como tanto la API como la aplicación web están alojadas en la misma máquina se necesita alguna forma para diferenciar qué peticiones van a qué servicio. Se ha decidido que *dominio* vaya a la aplicación web y que *api.dominio* vaya a la api, en este caso *dominio* es [paddle.catalin.sh](#). Configuraciones como esta se deben realizar en el archivo `.env` en la raíz del proyecto de docker (directorio `/app` del repositorio).

Para lograr lo anterior se deben añadir registros DNS desde el panel de administración del servidor de nombres de dominio elegido (CloudFlare en este caso) para hacer que *dominio* y *api.dominio* apunten a la dirección ip del equipo en el que se aloja la aplicación. Con esto se consigue que las peticiones realizadas a los dominios vayan lleguen a la máquina en cuestión.

A continuación, se debe redirigir la petición dentro de la máquina al contenedor respectivo. Para esto se utiliza el servidor web Caddy, que permite la creación de un proxy inverso con el que se pueden redirigir peticiones en función del dominio al que vayan dirigidas.

Además, Caddy permite la obtención y renovación automática de certificados SSL de una entidad certificadora (en concreto [Let's Encrypt](#)) cuyos certificados son de confianza en la mayoría de navegadores.

Make

GNU Make es una herramienta usada para coordinar la ejecución de secuencias de comandos. Es muy usada en el proceso de compilación de programas escritos en C o C++ debido a lo complejos que pueden llegar a ser los comandos necesarios.

En este caso GNU Make ha sido usado para simplificar la ejecución de comandos comunes. Algunas acciones que se realizan son:

- Lanzar los contenedores de Docker para el desarrollo.
- Lanzar los contenedores de Docker para producción.
- Buscar posibles errores de tipado en código Python.
- Formatear el código Python.

- Compilar los archivos L^AT_EX para generar esta documentación.
- ...

Accesibilidad

El público objetivo de esta aplicación son aquellas personas que padeczan o puedan padecer la enfermedad de Parkinson. Los problemas de movilidad causados por esta enfermedad reducen la precisión con la que se puede usar una herramienta como un ratón y se debe utilizar métodos alternativos para los que se deberá implementar soporte explícitamente en el código de una página web.

Para asegurar el uso correcto de las diferentes herramientas de accesibilidad que los navegadores implementan se ha realizado el [curso sobre accesibilidad web de Udacity](#), que se puede ver en [esta lista de reproducción de YouTube](#).

El concepto más importante adquirido es lo que se conoce como «accessibility tree», que es una estructura de árbol que se genera automáticamente a partir del código HTML de una página web. En esta estructura se eliminan todos los elementos visuales, y quedan solo aquellos elementos con contenido (encabezados, párrafos, etc.) o funcionalidad (botones, links, etc.).

Cuando se utiliza una página web únicamente con el teclado (utilizando las teclas de tabulación y flechas) se está navegando a través de los elementos funcionales del árbol de accesibilidad.

Otros conceptos sobre accesibilidad de gran importancia son:

- Existen atributos ARIA (Accessible Rich Internet Applications) que deben ser utilizados para especificar aspectos relacionados con la accesibilidad como, por ejemplo, el orden de tabulado de los elementos.
- Es importante utilizar elementos HTML preexistentes cuando sea posible. Por ejemplo, utilizar el elemento `dialog` para crear ventanas «pop-up» o emergentes en lugar de `div` que es un contenedor genérico. Estos elementos más específicos añaden información semántica sobre el propósito y el uso de un componente, además de funcionalidad común adicional, por ejemplo, un elemento `dialog` se cierra automáticamente al pulsar la tecla `Esc` sin necesidad de añadir código JavaScript propio.
- Existe patrones y métodos estándar para implementar componentes y funcionalidades comunes en la web, como menús desplegables, listas,

etc. Estos se pueden encontrar en la [Guía de prácticas de autoría de ARIA](#), mantenida por el W3C (World Wide Web Consortium). Estas recomendaciones homogenizan la forma de uso de elementos comunes, lo que es especialmente importante para personas con accesibilidad limitada.

- Las personas con visión reducida dependen de un lector de pantalla para saber el contenido al que están accediendo. Esto se debe tener en cuenta para utilizar los atributos ARIA adecuados, por ejemplo:
 - Para comunicar sobre el estado de una barra de progreso se emplean los atributos `aria-valuenow`, `aria-valuemin` y `aria-valuemax`.
 - Para anunciar al usuario de un mensaje si, por ejemplo, le llega un correo mientras navega por su bandeja de entrada, se emplea el atributo `aria-live`. Que puede tomar los valores `assertive`, `off` y `default`, dependiendo del tipo de interrupción que se desee realizar.

Estos conocimientos han sido utilizados a través de la aplicación y, aunque no ha sido probada por una persona con accesibilidad reducida, se ha intentado navegar la página únicamente con un lector de pantalla para buscar posibles errores y mejoras.

6. Trabajos relacionados

Durante la última década se ha intentado utilizar la visión por computador para evaluar la Enfermedad de Parkinson en múltiples ocasiones con diferentes resultados. Este capítulo recopila de forma resumida algunos de los trabajos más relevantes.

6.1. A computer vision framework for finger-tapping evaluation

Este artículo [5] documenta el uso de visión por computador para determinar el nivel de severidad de la Enfermedad de Parkinson y distinguir entre individuos con esta enfermedad e individuos sin ella.

El método empleado se caracteriza por utilizar vídeos con la cara de la persona y ambas manos a los lados de la cabeza, apuntando las puntas de los dedos hacia la misma. Esto se utiliza para poder normalizar las distancias en base a características faciales.

El estudio se realizó sobre 13 pacientes con la Enfermedad de Parkinson, tomando 17 vídeos de cada paciente durante un día, además de un grupo de control de 6 individuos, tomando 2 vídeos al día durante una semana por cada uno. Aunque algunos de estos vídeos fueron descartados. En total se utilizaron 471 vídeos.

Metodología

1. Se detecta la cara del individuo para la normalización. Esto se basa en que la longitud de la mano de una persona adulta es aproximadamente

igual a la altura de su cara.

2. Se obtiene una serie temporal que representa la amplitud del movimiento de los dedos índice y pulgar de la mano dominante del individuo.
3. Se extraen un total de 15 características de esta serie temporal, por ejemplo:
 - Correlación cruzada media entre los máximos locales de dos intervalos distintos de tiempo de la serie temporal. Esto mismo se realiza también sobre los mínimos locales.
 - Número total de toques de dedos durante la grabación.
 - Velocidad media de la apertura de dedos.
 - Velocidad media del cierre de dedos.
 - ...
4. Se realiza una selección de características eliminando aquellas redundantes y usando el algoritmo chi-cuadrado.
5. Se entrena una máquina de vectores de soporte (SVM) mediante las características obtenidas para realizar la clasificación.

Resultados En cuanto a la distinción entre pacientes de la Enfermedad de Parkinson y el grupo de control se obtuvo una precisión del 95 %, que es una cifra que se debería tomar con precaución debido que, aunque se han utilizado 471 vídeos, estos provienen de únicamente 19 personas.

6.2. The discerning eye of computer vision

En este estudio [16] realizado sobre 39 pacientes con la Enfermedad de Parkinson y sobre un grupo de control de 30 individuos se tomaron vídeos de ambas manos de cada individuo mientras realizan toques de los dedos índice y pulgar (de forma similar a cómo se han tomado las muestras para este trabajo). Dando un total de 133 vídeos (se descartó uno).

De estos vídeos se han extraído diferentes características y comprobado la relación que existe entre éstas y diferentes escalas que clasifican el nivel de gravedad de la Enfermedad en un paciente, como la *Modified Bradykinesia Rating Scale* (MBRS) que categoriza el movimiento de los individuos en 5 niveles, del 0 al 4, siendo 0 un movimiento normal y 4 el nivel de mayor gravedad.

Metodología

1. Se utiliza una librería de visión por computador, en concreto DeepCutLab, para obtener una serie temporal de la amplitud entre las puntas de los dedos pulgar e índice.
2. Se normaliza esta serie temporal utilizando la amplitud máxima detectada, que va a convertirse en el valor 1, siendo todos los demás valores escalados proporcionalmente.
3. Se extraen las siguientes características:
 - Velocidad, calculada como la tasa media de cambio.
 - Variabilidad de la amplitud, calculada como el coeficiente de variación de la diferencia media entre máximos y mínimos de diferentes intervalos de 1 segundo de la serie temporal.
 - Regularidad del ritmo, calculada utilizando la Transformada Rápida de Fourier y, a continuación, midiendo la potencia de la frecuencia dominante más la potencia de las frecuencias en un intervalo de 0.4 Hz alrededor de ésta (un ritmo más regular concentra una mayor potencia en una única frecuencia).

Resultados Se observó una correlación bastante alta entre las características utilizadas y la categoría del individuo dentro de las escalas de medición de la Enfermedad de Parkinson utilizadas medida por un experto en el campo.

6.3. Supervised classification of bradykinesia

Este estudio [15] es muy similar al anteriormente explicado, y está realizado por un equipo compuesto por casi los mismos participantes. En este caso se utilizaron 70 vídeos, de ambas manos de 20 pacientes con la Enfermedad de Parkinson y de un grupo de control de 15 individuos.

Metodología La metodología es prácticamente igual que antes, la diferencia principal está en las características que se extraen de la serie temporal correspondiente con la amplitud, se ha obtenido:

- Frecuencia, medida como la frecuencia máxima de la Transformada Rápida de Fourier de la serie temporal.

- Amplitud, calculada como la densidad espectral, que se ha obtenido mediante la integral cuadrada del espectro de la Transformada Rápida de Fourier.

Con estas características se ha realizado clasificación binaria mediante clasificación bayesiana ingenua (naive bayes), regresión logística y máquina de vectores de soporte, tanto con función lineal como con función de base radial.

Resultados Los mejores resultados se obtuvieron con máquina de vectores de soporte con función de base radial, que coincide en un 73 % de los casos con la clasificación de expertos en el campo.

7. Conclusiones y Líneas de trabajo futuras

Bibliografía

- [1] Alfredo Berardelli, John C Rothwell, Philip D Thompson, and Mark Hallett. Pathophysiology of bradykinesia in parkinson's disease. *Brain*, 124(11):2131–2146, 2001.
- [2] Maximilian Christ, Nils Braun, Julius Neuffer, and Andreas W Kempa-Liehr. Time series feature extraction on basis of scalable hypothesis tests (tsfresh—a python package). *Neurocomputing*, 307:72–77, 2018.
- [3] Sociedad Española de Neurología (SEN). El número de afectados por la enfermedad de parkinson se duplicará en 20 años y se triplicará en 2050, 2021.
- [4] Nathan Intrator. Feature extraction using an unsupervised neural network. In *Connectionist Models*, pages 310–318. Elsevier, 1991.
- [5] Taha Khan, Dag Nyholm, Jerker Westin, and Mark Dougherty. A computer vision framework for finger-tapping evaluation in parkinson's disease. *Artificial intelligence in medicine*, 60(1):27–40, 2014.
- [6] Erik G Learned-Miller. Introduction to supervised learning. *I: Department of Computer Science, University of Massachusetts*, page 3, 2014.
- [7] Andrzej Maćkiewicz and Waldemar Ratajczak. Principal components analysis (pca). *Computers & Geosciences*, 19(3):303–342, 1993.
- [8] OpenHub. MS Windows NT kernel description. <https://www.openhub.net/repositories/compare>. Acceso: 2023-03-13.

- [9] Ronald B Postuma, Daniela Berg, Matthew Stern, Werner Poewe, C Warren Olanow, Wolfgang Oertel, José Obeso, Kenneth Marek, Irene Litvan, Anthony E Lang, et al. Mds clinical diagnostic criteria for parkinson's disease. *Movement disorders*, 30(12):1591–1601, 2015.
- [10] Diego Santos García, Marta Blázquez-Estrada, Matilde Calopa, Francisco Escamilla-Sevilla, Eric Freire, Pedro J García Ruiz, Francisco Grandas, Jaime Kulisevsky, Lydia López-Manzanares, Juan Carlos Martínez Castrillo, et al. Present and future of parkinson's disease in spain: Parkinson-2030 delphi project. *Brain Sciences*, 11(8):1027, 2021.
- [11] Colin Shearer. The crisp-dm model: the new blueprint for data mining. *Journal of data warehousing*, 5(4):13–22, 2000.
- [12] Wikipedia. Enfermedad de parkinson — wikipedia, la enciclopedia libre, 2023. [Internet; descargado 25-enero-2023].
- [13] Wikipedia contributors. Data pre-processing — Wikipedia, the free encyclopedia, 2023. [Online; accessed 15-May-2023].
- [14] Wikipedia contributors. Quantile normalization — Wikipedia, the free encyclopedia, 2023. [Online; accessed 1-May-2023].
- [15] Stefan Williams, Samuel D Relton, Hui Fang, Jane Alty, Rami Qahwaji, Christopher D Graham, and David C Wong. Supervised classification of bradykinesia in parkinson's disease from smartphone videos. *Artificial Intelligence in Medicine*, 110:101966, 2020.
- [16] Stefan Williams, Zhibin Zhao, Awais Hafeez, David C Wong, Samuel D Relton, Hui Fang, and Jane E Alty. The discerning eye of computer vision: Can it measure parkinson's finger tap bradykinesia? *Journal of the Neurological Sciences*, 416:117003, 2020.
- [17] Rüdiger Wirth and Jochen Hipp. Crisp-dm: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, volume 1, pages 29–39. Manchester, 2000.
- [18] Jia Wu, Xiu-Yun Chen, Hao Zhang, Li-Dong Xiong, Hang Lei, and Si-Hao Deng. Hyperparameter optimization for machine learning models based on bayesian optimization. *Journal of Electronic Science and Technology*, 17(1):26–40, 2019.

- [19] Petros Xanthopoulos, Panos M Pardalos, Theodore B Trafalis, Petros Xanthopoulos, Panos M Pardalos, and Theodore B Trafalis. Linear discriminant analysis. *Robust data mining*, pages 27–33, 2013.
- [20] Fan Zhang, Valentin Bazarevsky, Andrey Vakunov, Andrei Tkachenka, George Sung, Chuo-Ling Chang, and Matthias Grundmann. Media-pipe hands: On-device real-time hand tracking. *arXiv preprint arXiv:2006.10214*, 2020.