



ESCUELA  
POLITÉCNICA  
NACIONAL

# MODELO DE APRENDIZAJE AUTOMÁTICO NO SUPERVISADO PARA EL ANÁLISIS DE PACIENTES CON DETERIORO COGNITIVO RELACIONADO CON LA PATOLOGÍA DE ALZHEIMER.



DESARROLLO DE MODELOS DE APRENDIZAJE AUTOMÁTICO APLICADOS A LA PATOLOGÍA DE ALZHEIMER

Kleber Andrés Albarracín Vinueza

DIRECTOR: Dr. Aguiar Pontes Josafá de Jesús

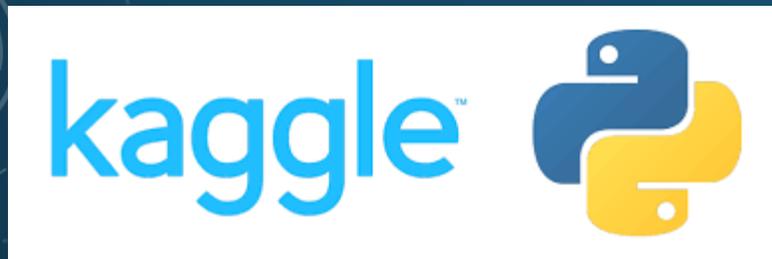
11 de septiembre de 2024





# INTRODUCCIÓN

El Alzheimer es una enfermedad prevalente en la tercera edad, lo que resalta la necesidad de analizar características que permitan mejorar su diagnóstico y seguimiento. Este trabajo desarrolla un modelo de aprendizaje no supervisado, basado en datos obtenidos de Kaggle y procesados en Google Colab.



# OBJETIVOS

## Objetivo General:

Desarrollar un modelo de aprendizaje automático no supervisado que permita analizar las características de los pacientes que presentan Alzheimer e identificar aquellas que se considera relevantes.

## Objetivos específicos:

- Aplicar técnicas de preprocesamiento de datos, como normalización y manejo de valores faltantes, para asegurar la coherencia y la calidad de los datos utilizados en el modelo.
- Desarrollar un modelo de aprendizaje no supervisado que pueda identificar patrones correlacionados con personas sanas y patrones correlacionados con personas enfermas.
- Seleccionar y probar algoritmos de aprendizaje no supervisado, ajustando sus parámetros hasta lograr resultados convenientes.



# METODOLOGÍA: preparación de datos

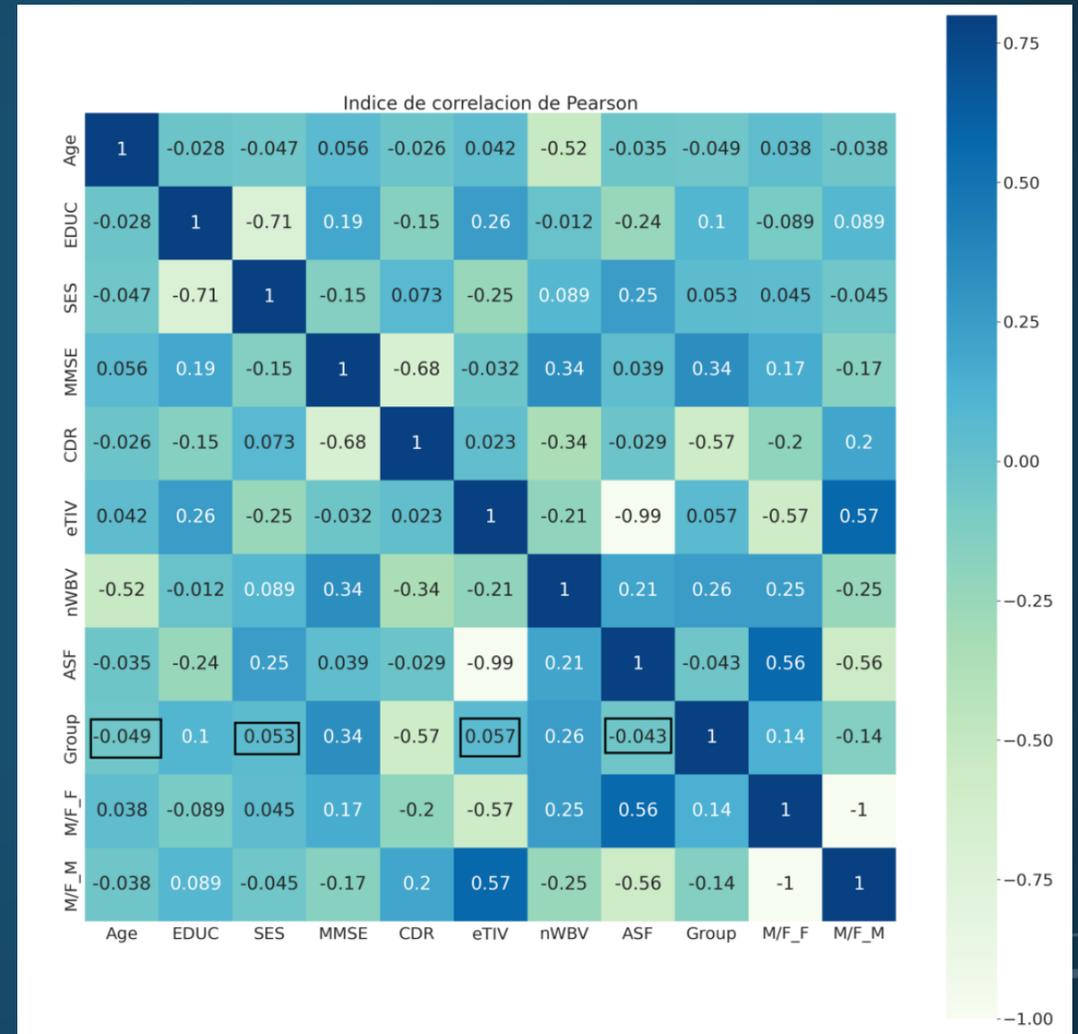
## Descripción del conjunto de datos

Dentro de la data o conjunto de datos encontramos una serie de características del Alzheimer las cuales se mencionan a continuación.

M/F, Age, Education, SES (Socioeconomic Status), MMSE (Mini-Mental State Examination), CDR (Clinical Dementia Rating), eTIV (Estimated Total Intracranial Volume), nWBV (Normalized Whole Brain Volume), ASF (Atlas Scaling Factor), Group (Demented, No Demented, Converted).

## Preprocesamiento de datos

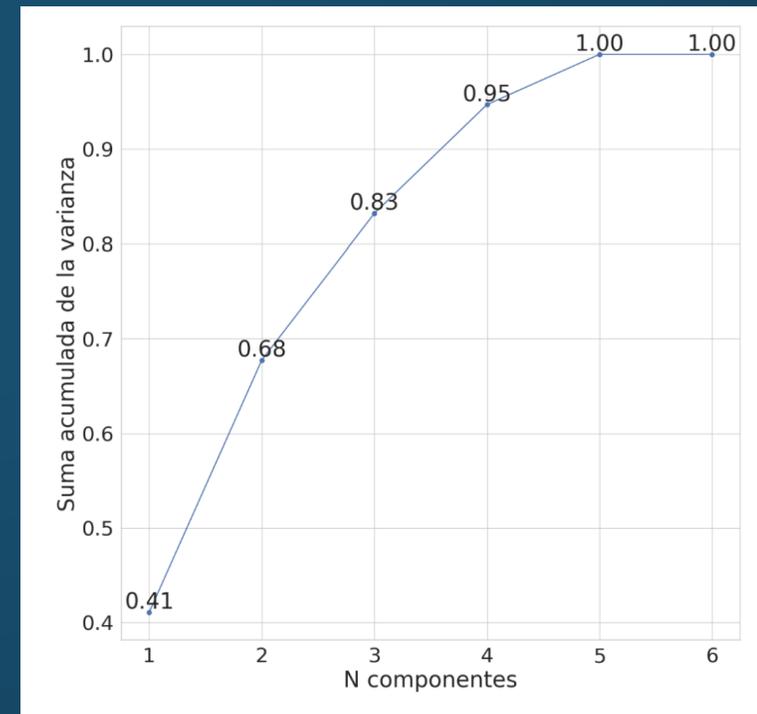
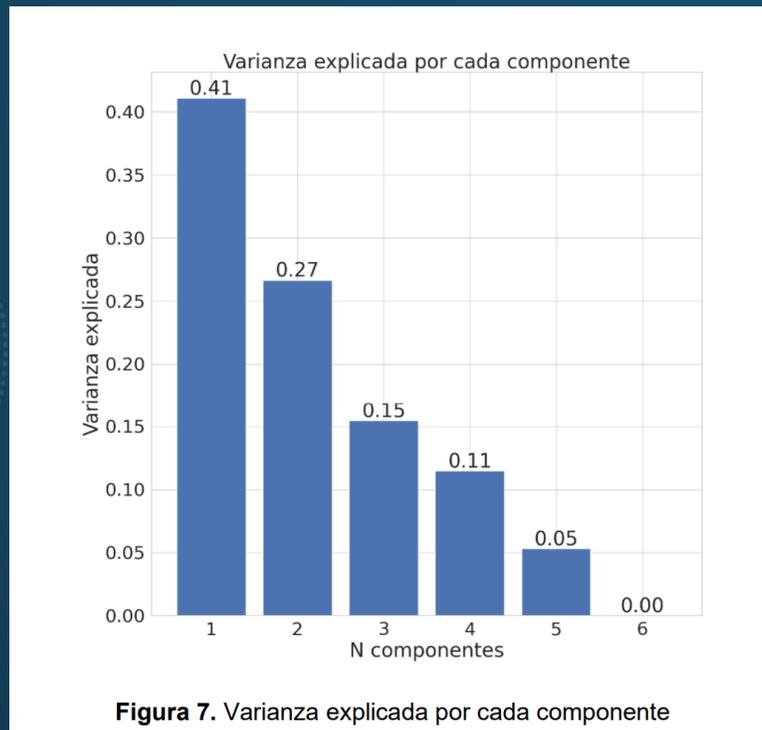
- Manejo de valores faltantes.
- Análisis de correlaciones
- Escalado de variables



# METODOLOGÍA- Análisis De Componentes Principales (PCA)

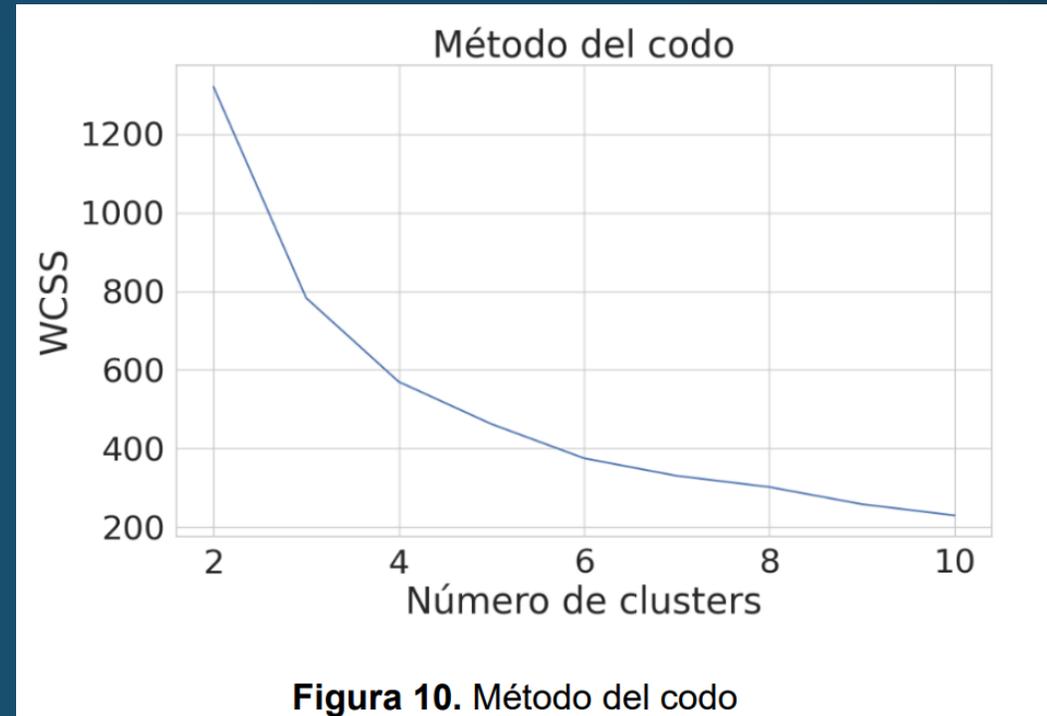
El PCA tiene como propósito la reducción de la dimensionalidad que transforma un conjunto de variables posiblemente correlacionadas en un conjunto más pequeño de variables no correlacionadas llamadas componentes principales, esto ayuda a simplificar el modelo, reducir el ruido y mejorar la interpretabilidad de los datos.

La varianza explicada y la varianza explicada acumulada, son vectores de mucha utilidad cuando se debe decidir el número de componentes principales que se van a utilizar.



# METODOLOGÍA- Clustering Con K-Means

La agrupación en clusters K-means es uno de los algoritmos de aprendizaje automático no supervisado más simple y popular, este algoritmo hace inferencias a partir de un conjunto de datos utilizando únicamente vectores de entrada sin hacer referencia a resultados conocidos, es decir, el objetivo de K-Means es agrupar puntos de datos similares y descubrir patrones subyacentes, es decir, características similares que permitan identificar a pacientes con demencia.

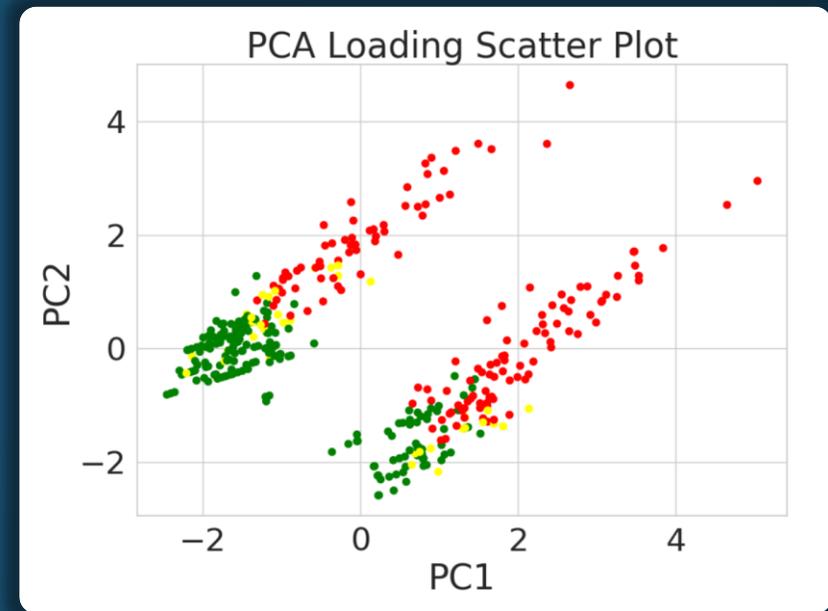


# RESULTADOS - PCA

El PCA se empleó para reducir la dimensionalidad de los datos y visualizar las principales fuentes de variación, al terminar de ejecutar el modelo de aprendizaje no supervisado se obtuvo tanto la varianza explicada como la varianza explicada acumulada por los primeros 4 componentes principales



```
Varianza explicada por las primeras componentes principales:  
[0.41092417 0.26661385 0.15458874 0.11482663]  
Varianza explicada acumulada: 0.9469533961357217
```



# RESULTADOS – K-MEANS

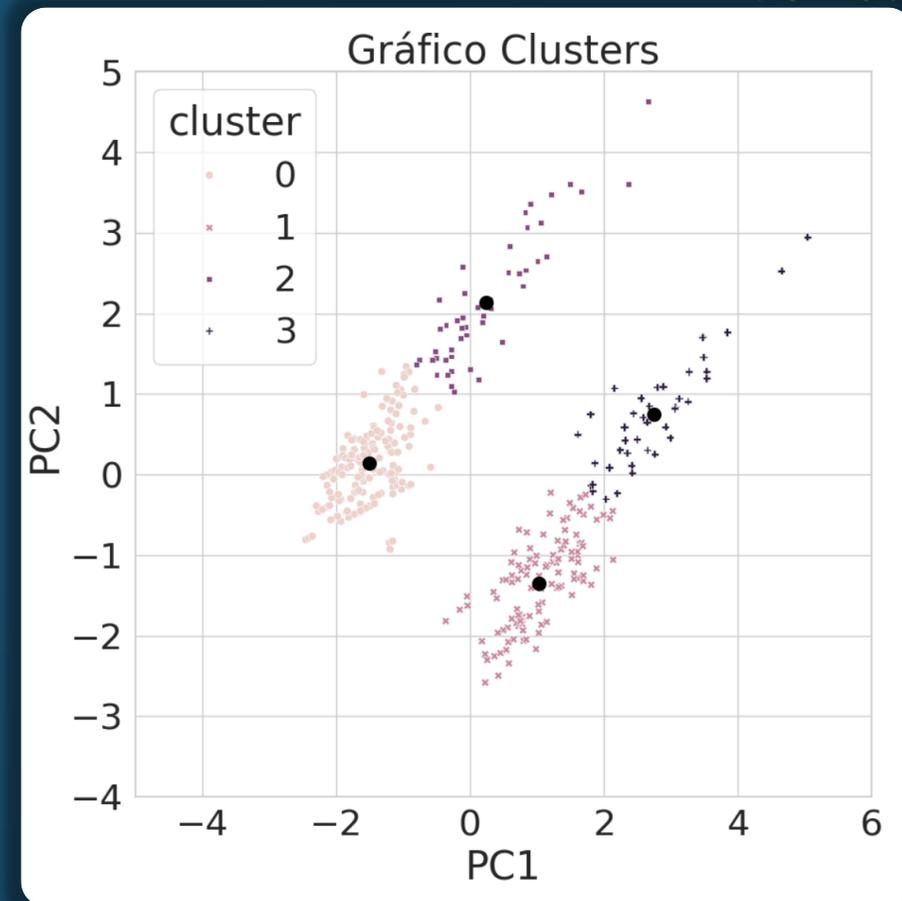
El algoritmo de K-Means se utilizó para agrupar los datos en clusters, por lo que se generaron varios clusters, cada uno representando a un grupo de individuos con características similares.

Al calcular la frecuencia y la frecuencia relativa se obtuvieron cuantos datos están en cada cluster y el porcentaje de total de los datos

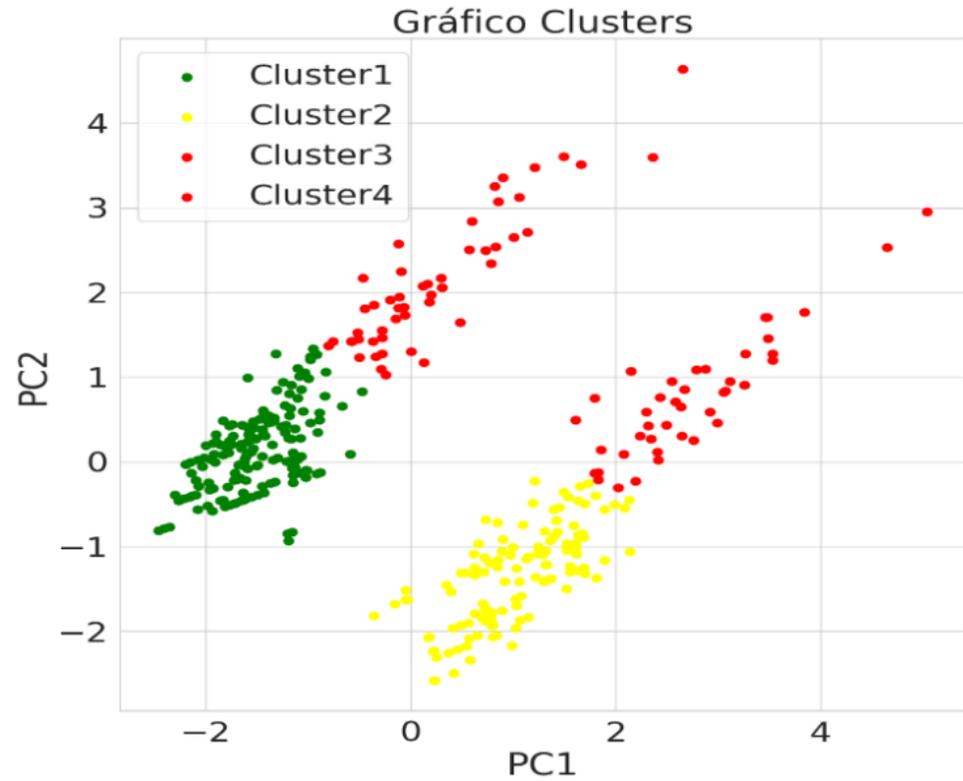
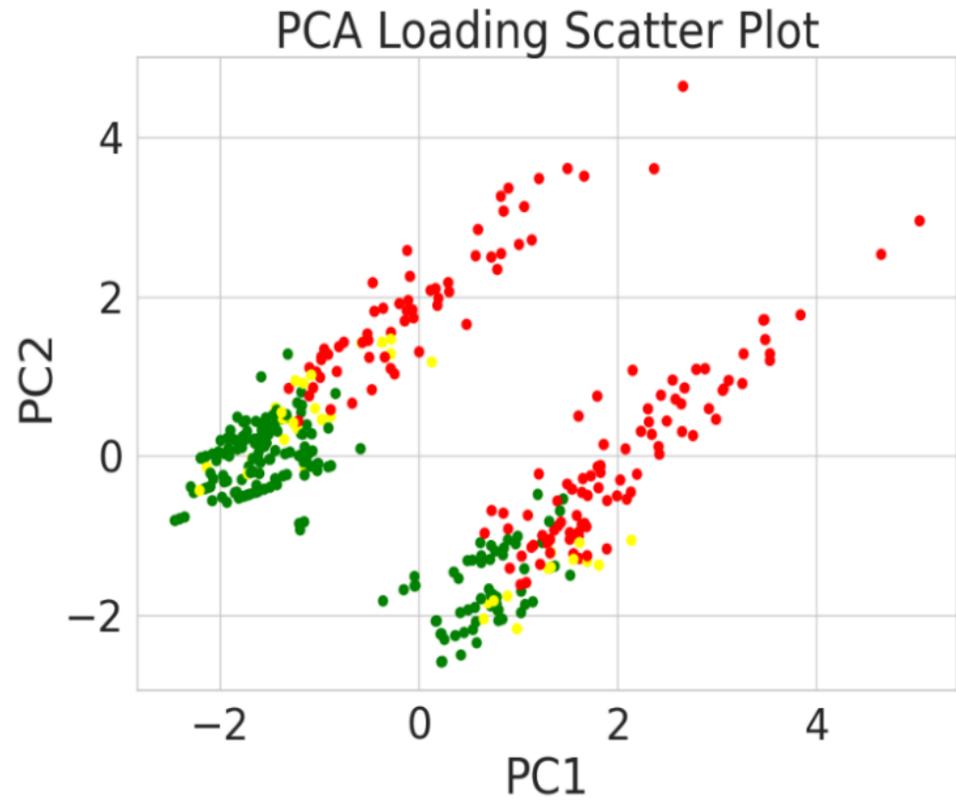


**Tabla 2.** Tabla de frecuencias

Frecuencia	Frecuencia Relativa
163	43.70%
119	31.90%
50	13.40%
41	10.99%



# RESULTADOS – comparativa entre los dos algoritmos



# CONCLUSIONES

- El desarrollo de un modelo de aprendizaje no supervisado utilizando específicamente los modelos PCA y K-Means, permitió identificar patrones significativos en los datos. Estos patrones distinguen claramente entre individuos no dementes, aquellos en transición y aquellos con demencia, proporcionando así una herramienta valiosa para la identificación temprana y el tratamiento de la enfermedad.
- Se analizaron los resultados obtenidos mostrando patrones relevantes que pueden ser de gran utilidad para los profesionales de la salud en el manejo del Alzheimer. Las conclusiones obtenidas indican que el modelo de aprendizaje no supervisado desarrollado es efectivo, no solo en la identificación de los estados de salud, sino que, también puede ofrecer una percepción para el desarrollo de estrategias de tratamiento y gestión de la enfermedad.

# TRABAJO FUTURO



En el futuro, se propone expandir el análisis del modelo mediante la incorporación de nuevos conjuntos de datos y variables que permitan un diagnóstico más preciso y temprano del Alzheimer. Además, se plantea la integración de modelos de aprendizaje supervisado, que aprovechen los resultados obtenidos del aprendizaje no supervisado, para mejorar la predicción de la progresión de la enfermedad.

Finalmente, se sugiere explorar la aplicación de técnicas de inteligencia artificial más avanzadas, para evaluar su efectividad en la identificación de patrones complejos en los datos clínicos permitiendo la integración de herramientas de diagnóstico y pronóstico de esta enfermedad.

# Preguntas y agradecimientos



Gracias