

Comparación de modelos de inteligencia artificial para el diagnóstico de  
arritmia cardiaca mediante información de electrocardiogramas

**Presentado por:**

Cristian Camilo Gelves Higuera

**Código:**

U00115711

**Director:**

Andres Elias Mrad Ricaurte

Universidad Autónoma De Bucaramanga

Facultad De Ingeniería

Programa De ingeniería de Sistemas

Bucaramanga

2021

**Resumen**

El presente proyecto tiene como objetivo hacer un benchmarking de modelos de inteligencia artificial IA para determinar la presencia de arritmia cardiaca con base en la información de electrocardiogramas. Para esto se recopila registros provenientes de bases de datos medicas libres con el fin de entrenar y probar los modelos haciendo uso de metodologías de clasificación como lo son: reconocedor euclídeo y KNN (k vecinos más próximos). Adicionalmente se implementan métodos de reducción de dimensionalidad como lo son: PCA (análisis de componentes principales) y FSCNCA (Feature selection using neighborhood component analysis for classification). Para el desarrollo del presente proyecto se hace uso del software de cómputo numérico Matlab.

### **Abstract**

The objective of this project is to benchmark AI models to determine the presence of cardiac arrhythmia based on information from electrocardiograms. For this, records from medical databases are collected to train and test the models using classification methodologies such as: Euclidean recognizer and KNN (k-closest neighbors). In addition, dimensionality reduction methods are implemented such as: PCA (Principal Component Analysis) and FSCNCA (Selection of characteristics through neighborhood component analysis for classification). For development, the MATLAB programming language is used.

## Tabla de contenido

<b>1.</b>	<b>Problemática y justificación</b>	12
1.1.	Planteamiento de la problemática	12
1.1.1.	Árbol del problema	13
1.2.	Justificación	13
<b>2.</b>	<b>Objetivos</b>	14
2.1.	Objetivo General	14
2.2.	Objetivos Específicos	14
<b>3.</b>	<b>Antecedentes</b>	14
<b>4.</b>	<b>Estado del arte</b>	15
<b>5.</b>	<b>Marco Teórico</b>	20
5.1.	Enfermedad Cardiovascular	20
5.1.1.	Que es una enfermedad Cardiovascular	20
5.1.2.	¿Cómo se detecta una enfermedad cardiovascular?	21
5.2.	Arritmias	22

	5
<b>5.2.1. ¿Qué es una arritmia?</b>	22
<b>5.2.2. ¿Cómo se diagnostica una arritmia?</b>	23
<b>6. Diseño Metodológico</b>	25
<b>7. Presupuesto</b>	28
<b>8. Desarrollo</b>	30
<b>8.1. Herramientas de Desarrollo</b>	30
<b>8.2. Obtención de Datos</b>	30
<b>9. Metodología</b>	32
<b>9.1. Reconocedor Euclídeo</b>	32
<b>9.2. KNN</b>	33
<b>10. Reducción de Dimensionalidad</b>	34
<b>10.1. FSCNCA</b>	34
<b>10.2. PCA</b>	35
<b>11. Resultados</b>	36
<b>11.1. Reconocedor Euclídeo</b>	37
<b>11.2. Reconocedor Euclídeo FSCNCA</b>	37
<b>11.3. Reconocedor euclídeo PCA</b>	38
<b>11.4. KNN</b>	38
<b>11.5. KNN FSCNCA</b>	49

	6
<b>11.6. KNN PCA</b>	<b>60</b>
<b>12. Conclusiones</b>	<b>71</b>
<b>13. Recomendaciones</b>	<b>72</b>
<b>14. Referencias bibliográficas</b>	<b>74</b>

### **Lista de Figuras**

Figura 1 Árbol del problema	13
Figura 2 Variables Relevantes FSCNCA	37
Figura 3 variables relevantes PCA	38
Figura 4 KNN EUCLIDEAN	45
Figura 5 KNN CITYBLOCK	45
Figura 6 KNN MINKOWSKI	46
Figura 7 KNN CHEBYCHEV	47
Figura 8 KNN COSINE	48
Figura 9 KNN CORRELATION	49
Figura 10 KNN SPEARMAN	50
Figura 11 KNN FSCNCA EUCLIDEAN	55
Figura 12 KNN FSCNCA CITYBLOCK	56
Figura 13 KNN FSCNCA MINKOWSKI	57
Figura 14 KNN FSCNCA CHEBYCHEV	58
Figura 15 KNN FSCNCA COSINE	59

Figura 16 KNN FSCNCA CORRELATION	60
Figura 17 KNN FSCNCA SPEARMAN	61
Figura 18 KNN PCA EUCLIDEAN	67
Figura 19 KNN PCA CITYBLOCK	68
Figura 20 KNN PCA MINKOWSKI	69
Figura 21 KNN PCA CHEBYCHEV	70
Figura 22 KNN PCA COSINE	71
Figura 23 KNN PCA CORRELATION	72
Figura 24 KNN PCA SPEARMAN	73

### **Lista De Tablas**

Tabla 1 Criterios de Búsqueda del Estado del arte	13
Tabla 2 Estado del arte	13
Tabla 3 Cronograma de Actividades	26
Tabla 4 Presupuesto	28
Tabla 5 Herramientas de desarrollo	30
Tabla 6 Resultados Reconocedor Euclídeo	36
Tabla 7 Resultados Reconocedor Euclídeo FSCNCA	37
Tabla 8 Resultados reconocedor Euclídeo PCA	37
Tabla 9 Resultados KNN K=1	37
Tabla 10 Resultados KNN K=2	38



Tabla 11 Resultados KNN K=3	38
Tabla 12 Resultados KNN K=4	38
Tabla 13 Resultados KNN K=5	38
Tabla 14 Resultados KNN K=6	38
Tabla 15 Resultados KNN K=7	39
Tabla 16 Resultados KNN K=8	39
Tabla 17 Resultados KNN K=9	39
Tabla 18 Resultados KNN K=10	39
Tabla 19 Resultados KNN K=11	39
Tabla 20 Resultados KNN K=12	40
Tabla 21 Resultados KNN K=13	40
Tabla 22 Resultados KNN K=14	40
Tabla 23 Resultados KNN K=15	40
Tabla 24 Resultados KNN K=16	40
Tabla 25 Resultados KNN K=17	41
Tabla 26 Resultados KNN K=18	41

	10
Tabla 27 Resultados KNN K=19	41
Tabla 28 Resultados KNN K=20	41
Tabla 29 Resultados KNN FSCNCA K=1	46
Tabla 30 Resultados KNN FSCNCA K=2	46
Tabla 31 Resultados KNN FSCNCA K=3	47
Tabla 32 Resultados KNN FSCNCA K=4	47
Tabla 33 Resultados KNN FSCNCA K=5	47
Tabla 34 Resultados KNN FSCNCA K=6	47
Tabla 35 Resultados KNN FSCNCA K=7	47
Tabla 36 Resultados KNN FSCNCA K=8	48
Tabla 37 Resultados KNN FSCNCA K=9	48
Tabla 38 Resultados KNN FSCNCA K=10	48
Tabla 39 Resultados KNN FSCNCA K=11	48

	11
Tabla 40 Resultados KNN FSCNCA K=12	48
Tabla 41 Resultados KNN FSCNCA K=13	49
Tabla 42 Resultados KNN FSCNCA K=14	49
Tabla 43 Resultados KNN FSCNCA K=15	49
Tabla 44 Resultados KNN FSCNCA K=16	49
Tabla 45 Resultados KNN FSCNCA K=17	49
Tabla 46 Resultados KNN FSCNCA K=18	50
Tabla 47 Resultados KNN FSCNCA K=19	50
Tabla 48 Resultados KNN FSCNCA K=20	50
Tabla 49 Resultados KNN PCA K=1	54
Tabla 50 Resultados KNN PCA K=2	54
Tabla 51 Resultados KNN PCA K=3	55
Tabla 52 Resultados KNN PCA K=4	55

	12
Tabla 53 Resultados KNN PCA K=5	55
Tabla 54 Resultados KNN PCA K=6	55
Tabla 55 Resultados KNN PCA K=7	55
Tabla 56 Resultados KNN PCA K=8	56
Tabla 57 Resultados KNN PCA K=9	56
Tabla 58 Resultados KNN PCA K=10	56
Tabla 59 Resultados KNN PCA K=11	56
Tabla 60 Resultados KNN PCA K=12	56
Tabla 61 Resultados KNN PCA K=13	57
Tabla 62 Resultados KNN PCA K=14	57
Tabla 63 Resultados KNN PCA K=15	57
Tabla 64 Resultados KNN PCA K=16	57
Tabla 65 Resultados KNN PCA K=17	57

	13
Tabla 66 Resultados KNN PCA K=18	58
Tabla 67 Resultados KNN PCA K=19	58
Tabla 68 Resultados KNN PCA K=20	58
Tabla 69 Comparativa KNN RE	63

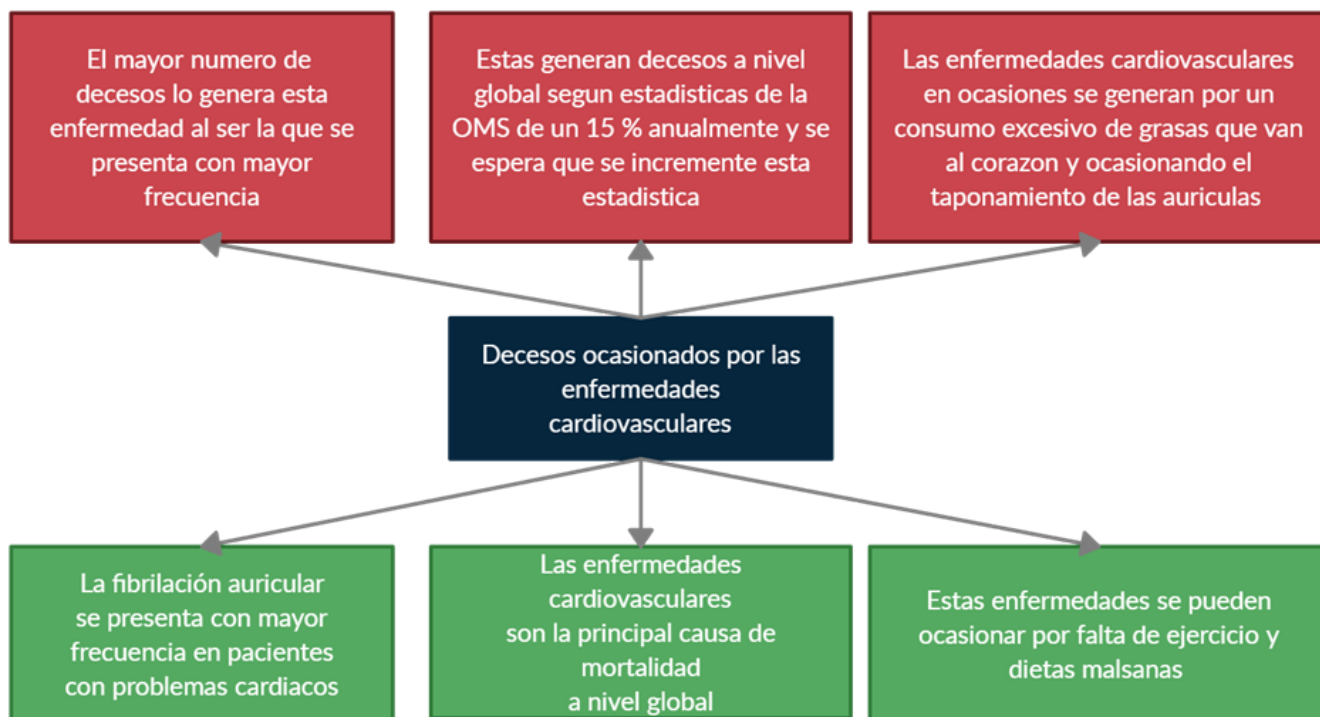
## 1. Problemática y justificación

### 1.1. Planteamiento de la problemática

Actualmente las enfermedades cardiovasculares son la causa más frecuente de muertes en el mundo anualmente; según la OMS se estima que para el año 2020 las muertes por enfermedades cardiovasculares incrementen de un 15% a un 20%, y también se pronostica que para el año 2030 las enfermedades cardiovasculares causan 23,6 millones de muertes siendo esta por consiguiente la principal causa de morbimortalidad a nivel global. ("Enfermedades cardiovasculares", 2021)

#### 1.1.1. Árbol del problema

Figura 1 Árbol del problema



## **1.2. Justificación**

Se plantea usar distintos modelos de IA para poder realizar una comparativa entre los resultados, y así poder escoger el modelo de mayor precisión para la predicción de la enfermedad en la cual nos centramos, la idea es facilitar por medio de la implementación del modelo más preciso un buen diagnóstico para ayudar a que las estadísticas de mortalidad vayan disminuyendo y no aumentando con el paso del tiempo.

## **2. Objetivos**

### **2.1. Objetivo General**

Implementar diferentes modelos de IA para el diagnóstico de arritmia cardiaca mediante información de electrocardiogramas.

### **2.2. Objetivos Específicos**

1. Determinar los métodos por medio de los cuales se realiza un diagnóstico de arritmia mediante información cardiaca de electrocardiogramas.
2. Recolectar el set de datos de diagnósticos médicos de la afección para la graduación del modelo.
3. Implementar distintos modelos de IA.
4. Comparar la precisión de los modelos mediante el set de datos recolectados.

### 3. Antecedentes

Arritmia cardiaca es una condición en la cual el ritmo cardiaco es irregular, el latido es demasiado lento o rápido. Ocurre por impulsos eléctricos inadecuados que coordinan los latidos del corazón.

Debido a la simplicidad y el bajo costo, realizar un Electrocardiograma (ECG) es la técnica más común para detectar arritmias cardiacas. La cantidad de ECG realizados cada día puede llegar a superar la capacidad de revisión por parte de operarios y técnicos. Variedad de métodos han sido propuestos para automatizar completamente la detección de arritmias cardiacas, métodos basados en machine learning y deep learning para monitoreo e interpretación de ECG con un gran aumento en el último.

Los modelos de deep learning se han vuelto un modo popular de clasificación de ECG, Los investigadores suelen usar variedad de técnicas de deep learning para esta aplicación, sin embargo, debido a que no existe una estandarización de los modelos de deep learning se pueden obtener modelos con mejores resultados que los hechos anteriormente.

### 4. Estado del arte

**Tabla 1.** Criterios de Búsqueda del Estado del arte

Bases de datos consultadas	ScienceDirect
	IEEE



Cantidad de referencias	ScienceDirect: 7
	IEEE: 3
Fechas de búsqueda	ScienceDirect octubre 2020
	IEEE septiembre 2020
Criterios de búsqueda	<p>Rango de la fecha de la búsqueda 2015 y 2020</p> <p>Palabras clave:</p> <p>ECG, Electrocardiogram, deep learning, Arrhythmia, machine learning, detection, classification.</p>

**Tabla 2.** Estado del arte

<b>Autor</b>	<b>Año</b>	<b>Palabras Clave</b>	<b>Título</b>	<b>Base de Datos</b>	<b>Aporte al Proyecto</b>
S. Sahoo, M. Dash, S. Behera, S. Sabut	2020	Cardiac arrhythmias, Computer-aided diagnosis, Electrocardiogram, QRS complex, Machine learning, Life threatening	Machine Learning Approach to Detect Cardiac Arrhythmias in ECG Signals: A Survey	ScienceDirect	Identificación del diagnóstico de las arritmias

<p>Fatma Murat, Ozal Yildirim, Muhammed Talo, Ulas Baran Baloglu, Yakup Demir, U. Rajendra Acharya</p>	<p>2020</p>	<p>Arrhythmia detection, Deep learning, ECG classification, CNN, LSTM</p>	<p>Application of deep learning techniques for heartbeats detection using ECG signals-anal ysis and review</p>	<p>ScienceDirect</p>	<p>Diferentes modelos de deep learning para detectar latidos del corazón</p>
<p>Rasmus S. Andersen, Abdolrahman Peimankar, Sadasivan Puthusserypady</p>	<p>2019</p>	<p>Electrocardiogram (ECG), Atrial fibrillation, Deep learning, Convolutional neural networks (CNNs), Recurrent neural networks (RNNs), Long short-term memory (LSTM)</p>	<p>A deep learning approach for real-time detection of atrial fibrillation</p>	<p>ScienceDirect</p>	<p>Un modelo el cual fue entrenado y validado en tres bases de datos diferentes para de esta manera poder tener un menor rango de error en cuanto a la detección de la fibrilación auricular.</p>

A. Rajkumar, M. Ganesan, R. Lavanya	2019	Electrocardiograph y, Feature extraction, Heart beat, Classification algorithms, Lead, Databases	Arrhythmia classificatio n on ECG using Deep Learning	IEEE	Maneras de clasificar los diferentes tipos de arritmias utilizando diferentes funciones y datos pre procesados
V. Sai Krishna, A. Nithya Kalyani	2019	Cardiac Arrhythmia, ECG, Deep Learning, Recurrent Neural Network	Prediction of Cardiac Arrhythmia using Artificial Neural Network	IEEE	Las Artificial Neural Network (ANN) nos proveen mayor exactitud comparado a otros modelos de machine learning
Saman Parvaneh, Jonathan Rubin, Saeed Babaeizadeh, Minnan Xu-Wilson	2019	Cardiac arrhythmias, Deep learning, ECG	Cardiac arrhythmia detection using deep learning: A review	ScienceDirect	Proveen diferentes dataset públicos seleccionados por la mayoría de

					los papers revisados
G. Sannino, G. De Pietro	2018	Deep learning, ECG classification, Heartbeat classification, Arrhythmia detection	A deep learning approach for ECG-based heartbeat classification for arrhythmia detection	ScienceDirect	No existe metodología estándar para la construcción de una red neuronal óptima con el número correcto de capas y número de neuronas en cada capa, el artículo nos proporciona una red neuronal desarrollada a prueba y error la cual obtiene resultados de hasta 99,09 de exactitud

Swapna G, Soman K P, Vinayakumar R	2018	ECG, cardiac arrhythmia, deep learning, CNN, RNN, LSTM, GRU	Automated detection of cardiac arrhythmia using deep learning techniques	ScienceDirect	proporciona comparaciones de rendimiento entre diferentes arquitecturas de deep learning
Ali Isin, Selen Ozdalili	2017	biomedical signal processing, deep learning, ECG classification, convolutional neural networks, cardiac arrhythmia, transfer learning	Cardiac arrhythmia detection using deep learning	ScienceDirect	Presenta un proceso para el reconocimiento y clasificación de ECG y demuestra la efectividad de utilizar métodos de detección basados en el feature extraction y transfer learning
Hilmy Assodiky, Iwan	2017	Machine learning, Biological neural	Deep learning	IEEE	Proporciona comparaciones

Syarif, Tessy Badriyah		networks, Neurons, Electrocardiograph y, Support vector machines, Sociology, Statistics	algorithm for arrhythmia detection		s entre diferentes algoritmos de machine learning y deep learning
---------------------------	--	--------------------------------------------------------------------------------------------------------	---------------------------------------------	--	----------------------------------------------------------------------------------

## 5. Marco Teórico

### 5.1. Enfermedad Cardiovascular

#### 5.1.1. Que es una enfermedad Cardiovascular

Las enfermedades cardiovasculares son desórdenes del corazón y de los vasos sanguíneos entre los que se incluyen:

- Cardiopatía coronaria: enfermedad de los vasos sanguíneos que irrigan el músculo cardiaco.
- Las enfermedades cerebrovasculares: enfermedades de los vasos sanguíneos que irrigan el cerebro.
- Las arteriopatías periféricas: enfermedades de los vasos sanguíneos que irrigan los miembros superiores e inferiores.

- La cardiopatía reumática: lesiones del músculo cardíaco y de las válvulas cardíacas debidas a la fiebre reumática, una enfermedad causada por bacterias denominadas estreptococos.
- Las cardiopatías congénitas: malformaciones del corazón presentes desde el nacimiento.
- Las trombosis venosas profundas y embolias pulmonares: coágulos de sangre (trombos) en las venas de las piernas, que pueden desprenderse (émbolos) y alojarse en los vasos del corazón y los pulmones.
- Arritmias: latidos anormales del corazón, ya sea irregulares, demasiado lentos(bradicardia) o demasiado rápidos(taquicardia).

#### **5.1.2. ¿Cómo se detecta una enfermedad cardiovascular?**

Después de haberle dado un vistazo a la historia médica del paciente y realizar una examinación física, para llegar al diagnóstico final el médico recurre a los siguientes exámenes:

- Electrocardiograma (ECG): representa la actividad eléctrica del corazón de manera gráfica en función del tiempo, que se obtiene, desde la superficie corporal, en el pecho, con un electrocardiógrafo en forma de cinta continua. Esta prueba puede mostrar latidos cardíacos anormales, daño al músculo cardíaco, problemas de flujo sanguíneo en las arterias y agrandamiento del corazón.

- Test de estrés: registra la actividad eléctrica del corazón durante el ejercicio, generalmente en una cinta de correr o bicicleta estática. Si tú no puedes hacer ejercicio, puedes tomar un medicamento en cambio eso muestra algún problema en la sangre fluir al corazón.

- Ecocardiografía: Convierte las ondas de sonido en imágenes que muestran el tamaño, la forma y movimiento. Las ondas sonoras también se pueden utilizar para ver cuánta sangre bombea el corazón cuando se contrae.

- Angiografía coronaria: Se generará una imagen de rayos X de sangre, problemas de flujo y bloqueos en la arteria coronaria. Se inserta un tubo delgado llamado catéter a través de una arteria de un brazo o pierna hasta el corazón. Se inyecta un tinte en el tubo, lo que permite el corazón y los vasos para ser filmados como el corazón zapatillas. Un ventriculograma: una imagen de la izquierda ventrículo del corazón, a veces se toma como parte del procedimiento de angiografía coronaria.

- Ecografía intracoronaria: usa un catéter para crear una imagen de las arterias coronarias. Muestra el grosor de las arterias y cualquier bloqueo en el flujo sanguíneo. (Sánchez-Arias et al., 1981)

## **5.2. Arritmias**

### **5.2.1. ¿Qué es una arritmia?**

Una arritmia es cualquier trastorno en los latidos o el ritmo del corazón. Significa que el corazón late demasiado rápido o demasiado lento o que tiene un patrón irregular. Cuando el corazón late más rápido de lo normal se denomina taquicardia. Cuando late demasiado lento se



llama bradicardia. El tipo más común de arritmia es la fibrilación auricular, que causa un latido irregular y rápido del corazón.

Existen muchos factores que pueden afectar el ritmo cardíaco, como haber tenido un ataque al corazón, fumar, defectos cardíacos congénitos y estrés. Algunas sustancias o medicinas también pueden causar arritmias.

Los síntomas de arritmia incluyen:

- Latidos cardíacos rápidos o lentos
- Latidos perdidos
- Mareos o vértigo
- Dolor en el pecho
- Falta de aire
- Sudoración

El médico puede realizar pruebas que detecten si padece de arritmia. El tratamiento para restablecer un ritmo cardíaco normal puede incluir medicamentos, un desfibrilador cardioversor implantable o un marcapasos y, a veces, cirugía.

### **5.2.2. ¿Cómo se diagnostica una arritmia?**

Para diagnosticar arritmias se usan diferentes pruebas como:

- Electrocardiograma (ECG): muestra qué tan rápido late el corazón y a qué ritmo (uniforme o irregular). También muestra la potencia y la sincronización de los impulsos eléctricos a medida que pasan por el corazón.
- Pruebas de sangre: Se pueden realizar pruebas de sangre para determinar la concentración de ciertas sustancias en la sangre, como el potasio y la hormona tiroidea. Las concentraciones anormales de estas sustancias pueden aumentar las probabilidades de tener una arritmia.
- Radiografía de tórax: La radiografía de tórax es una prueba indolora con la que se obtienen imágenes de las estructuras internas del tórax, como el corazón y los pulmones. En esta prueba se puede ver si el corazón está aumentado de tamaño.
- Prueba de esfuerzo: Algunos problemas del corazón son más fáciles de diagnosticar cuando el corazón está trabajando mucho más y latiendo rápido. Durante la prueba de esfuerzo, el paciente hace ejercicio para que el corazón trabaje mucho y lata rápido mientras se realizan pruebas cardíacas. Si no puede hacer ejercicio, es posible que le den medicinas para hacer que el corazón trabaje más intensamente y lata con rapidez.
- Estudio electrofisiológico: Esta prueba se usa para evaluar las arritmias graves. Durante el estudio electrofisiológico se introduce un alambre delgado y flexible en una vena del brazo o de la ingle (la parte superior del muslo) y se pasa hasta el corazón. El alambre registra los impulsos eléctricos del corazón. El médico puede usar el alambre para enviar estímulos eléctricos al corazón y causar una arritmia. Esto le permite ver si una medicina antiarrítmica puede solucionar

el problema. Durante el estudio electrofisiológico se puede realizar una ablación con catéter, que es un procedimiento con el que se tratan ciertos tipos de arritmia.

- Prueba de la mesa basculante: Esta prueba se usa a veces para tratar de hallar la causa de los desmayos. El paciente se acuesta en una mesa que pasa de posición horizontal a vertical. El cambio de posición puede hacer que el paciente se desmaye. El médico observa los síntomas, la frecuencia cardíaca, el trazado del electrocardiograma y la presión arterial durante la prueba. También puede darle al paciente una medicina y ver cómo responde a ésta.

- Angiografía coronaria: En la angiografía coronaria se usan un medio de contraste y rayos X especiales para mostrar el interior de las arterias coronarias. Para introducir el medio de contraste en las arterias coronarias, el médico emplea un procedimiento llamado cateterismo cardíaco. Un tubo delgado y flexible llamado catéter se inserta en un vaso sanguíneo del brazo, la ingle (la parte superior del muslo) o el cuello. El tubo se hace avanzar luego hasta las arterias coronarias y el medio de contraste se inyecta en la sangre. Mientras el medio de contraste fluye por las arterias coronarias se toman radiografías especiales. El medio de contraste le permite al médico estudiar la circulación de la sangre por el corazón y los vasos sanguíneos. De esta forma puede hallar bloqueos que puedan causar un ataque cardíaco.

- Registrador implantable subcutáneo: Este dispositivo detecta alteraciones del ritmo cardíaco. Se coloca bajo la piel del pecho en una intervención de cirugía menor. El registrador implantable subcutáneo les permite a los médicos determinar la razón de las palpitaciones o de los desmayos de una persona, especialmente si estos síntomas no suceden con mucha frecuencia. El dispositivo se puede usar por un tiempo de entre 12 y 24 meses. (GeoSalud, 2012)

## 6. Diseño Metodológico

La metodología seleccionada para realizar el proyecto es prototipado evolutivo, fue escogida porque es una metodología que permite mejorar a medida que se entrene el modelo que se construyó.

Esta metodología consta de las siguientes fases:

**Etapa 1 Investigación:** En esta fase se realiza la respectiva investigación necesaria para conocer sobre la enfermedad, cómo se detecta, los métodos por medio de los cuales se diagnostica y proyectos existentes. Para esto se definen unas actividades las cuales son:

- Búsqueda de literatura sobre la arritmia y su detección con ECG.
- Revisión de los documentos encontrados acordes al proyecto.
- Selección de la información más relevante para el proyecto.

Como resultado de esta etapa se genera el documento con la información de la investigación realizada.

**Etapa 2 Búsqueda:** En esta fase se realizará la búsqueda de los datos de electrocardiogramas en bases de datos públicas, así mismo seleccionaremos los modelos a usar para de esta manera

poder dar paso a la siguiente etapa en la cual se dará el uso de estos datos y modelos: Para ellos se definen estas actividades:

- Búsqueda de set de datos en bases de datos públicas.
- Realización de un filtro a los datos encontrados en la base de datos.
- Selección de información con mayor relevancia para el proyecto.

Como resultado de esta etapa se generará el set de datos identificando la afección que nos permitirá entrenar los modelos.

**Etapa 3 Desarrollo y pruebas:** En esta etapa se dará paso a desarrollar los modelos con el set de datos proporcionado el cual nos ayudará a entrenarlos y compararlos. Para esto se definen unas actividades las cuales son:

- Desarrollo de los modelos.
- Entrenamiento de los modelos con una parte de los datos filtrados.

Como resultado de esta etapa tendremos los modelos de IA que vamos a implementar.

**Etapa 4 Implementación:** En esta última fase se realizará la implementación de los modelos para de esta manera poder realizarle los respectivos ajustes con el fin de mejorar a medida que se vaya entrenando el modelo. Las actividades para esta etapa son las siguientes:

- Implementación de modelos con los datos restantes.
- Comparación de los resultados obtenidos de los modelos.
- Conclusiones

Como resultado de esta etapa se obtendrá una comparación de los resultados obtenidos de los modelos para su análisis.

## 7. Presupuesto

**Tabla 4.** Presupuesto

Categoría	Detalles	Descripción	Cantidad total	Valor(pesos)	Total(pesos)
-----------	----------	-------------	----------------	--------------	--------------


Personal	Ingeniero de sistemas	Cristian Camilo Gelves Higuera	4 horas de trabajo semanales para un total de 136 horas	30.000 COP	4.080.000 COP
	Director de proyecto de grado	Andres Elias Mrad Ricaurte	2 horas de asesorías semanales para un total de 68 horas	50.000 COP	3.400.000 COP
Equipos y software	Computador de escritorio	Computador utilizado para desarrollar el algoritmo	1	4.000.000 COP	4.000.000 COP
	Computador portátil	Computador empleado para desarrollar el algoritmo facilitando su transporte	1	2.500.000 COP	2.500.000 COP
	Licencia de software	Licencia de Windows 10 home	1	969.999 COP	969.999 COP

Otros	Servicio de internet	Precio mensual del internet	8	95.000 COP	760.000 COP
	Servicio de electricidad	Precio mensual del consumo de energía	8	50.000 COP	400.000 COP
Total					16.109.999 COP

## 8. Desarrollo

### 8.1. Herramientas de Desarrollo

**Tabla 5.** Herramientas de desarrollo

 <p>Fuente: <a href="https://la.mathworks.com/brand.html">https://la.mathworks.com/brand.html</a></p>	<p>MATLAB® es una plataforma de programación diseñada específicamente para ingenieros y científicos para analizar y diseñar sistemas. Matlab permite Analizar datos, Desarrollar algoritmos, Crear modelos y aplicaciones.</p>
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------



Con esta herramienta se obtiene una mayor facilidad al trabajar con matrices y graficar los datos gracias al lenguaje MATLAB, un lenguaje basado en matrices que permite la expresión más natural de las matemáticas computacionales

## **8.2. Obtención de Datos**

Con base en los diferentes artículos recolectados en el estado del arte se pudo acceder al origen de las bases de datos usadas en los mismos las cuales se alojaban en un repositorio llamado “physionet” el cual contiene una gran variedad de data bases médicos con la información necesaria para el entrenamiento y prueba de los modelos. Para el proceso de selección se realizaron dos filtros, el primero es contener datos de electrocardiogramas y el segundo es estar relacionados con las arritmias cardiacas. Finalizado el procesamiento de filtrado el resultado son ocho data bases con datos de electrocardiogramas (ECG) y relacionados a arritmias cardiacas:

- MIT-BIH Arrhythmia Data base ("MIT-BIH Arrhythmia Database v1.0.0", 2021)
- Intracardiac Atrial Fibrillation Data base ("Intracardiac Atrial Fibrillation Database v1.0.0", 2021)
- CU Ventricular Tachyarrhythmia Data base ("CU Ventricular Tachyarrhythmia Database v1.0.0", 2021)
- MIT-BIH Supraventricular Arrhythmia Data base ("MIT-BIH Supraventricular Arrhythmia Database v1.0.0", 2021)

- MIT-BIH Atrial Fibrillation Data base ("MIT-BIH Atrial Fibrillation Database v1.0.0", 2021)
- MIT-BIH Malignant Ventricular Ectopy Database ("MIT-BIH Malignant Ventricular Ectopy Database v1.0.0", 2021)
- ECG-ID Database ("ECG-ID Database v1.0.0", 2021)
- AF Termination Challenge Database ("AF Termination Challenge Database v1.0.0", 2021)

Seguido a esto, para procesar los datos fue necesario transformar las bases de datos a archivos “. mat”, los registros fueron descargados para tener un total de 541 y posteriormente organizados de la siguiente forma, el 70% en una matriz de training y el 30% restante en una matriz de validación, esto con la intención de poder entrenar teniendo un set de datos real y tener certeza de que el proceso era acertado teniendo como fuente de validación el 30% restante.

## **9. Metodología**

Una vez segmentados los datos, se eligieron dos métodos con los cuales generar los modelos de IA estos son, reconocedor euclídeo y KNN, adicionalmente se aplicaron dos métodos de reducción de dimensionalidad diferentes los cuales son FSRNCA y PCA.

### **9.1. Reconocedor Euclídeo**

Para este método se dividió la base de datos en ocho clases correspondientes a las diferentes opciones de clasificación, en cada clase se obtienen un centroide a partir de la matriz de training hallando la media de la misma, el centroide es el resultado de definir la media entre los patrones

presentes en las clases, para realizar este método de clasificación lo que se busca es obtener la distancia más cercana entre el patrón que se va a clasificar y el centroide de dicha clase por lo que para llevar a cabo este proceso se prueban 7 distancias como lo son:

- 'euclidean'
- 'cityblock'
- 'minkowski'
- 'chebychev'
- 'cosine'
- 'correlation'
- 'spearman'

Una vez se obtienen las distancias de cada patrón por cada uno de los centroides se hace una comparación entre las distancias buscando la más cercana, aquella que tenga la distancia más cercana entre el patrón y uno de los centroides da lugar a que ese patrón sea clasificado en la clase de ese centroide de esta forma lo que se hace es una clasificación de las ocho clases que se tienen planteadas.

## 9.2. KNN

Para este método se definió una variable  $K$  como número de vecinos más cercanos elegidos, luego por cada patrón se calculan diferentes distancias entre el resto del dataset las cuales, al igual que en reconocedor euclídeo son:

- 'euclidean'
- 'cityblock'
- 'minkowski'
- 'chebychev'
- 'cosine'
- 'correlation'
- 'spearman'

Terminado el cálculo de distancias se seleccionan  $K$  número de vecinos en base a las distancias más cortas, de los cuales se consulta la clase a la cual pertenecen para así por mayoría de clase calificar el patrón en la misma.

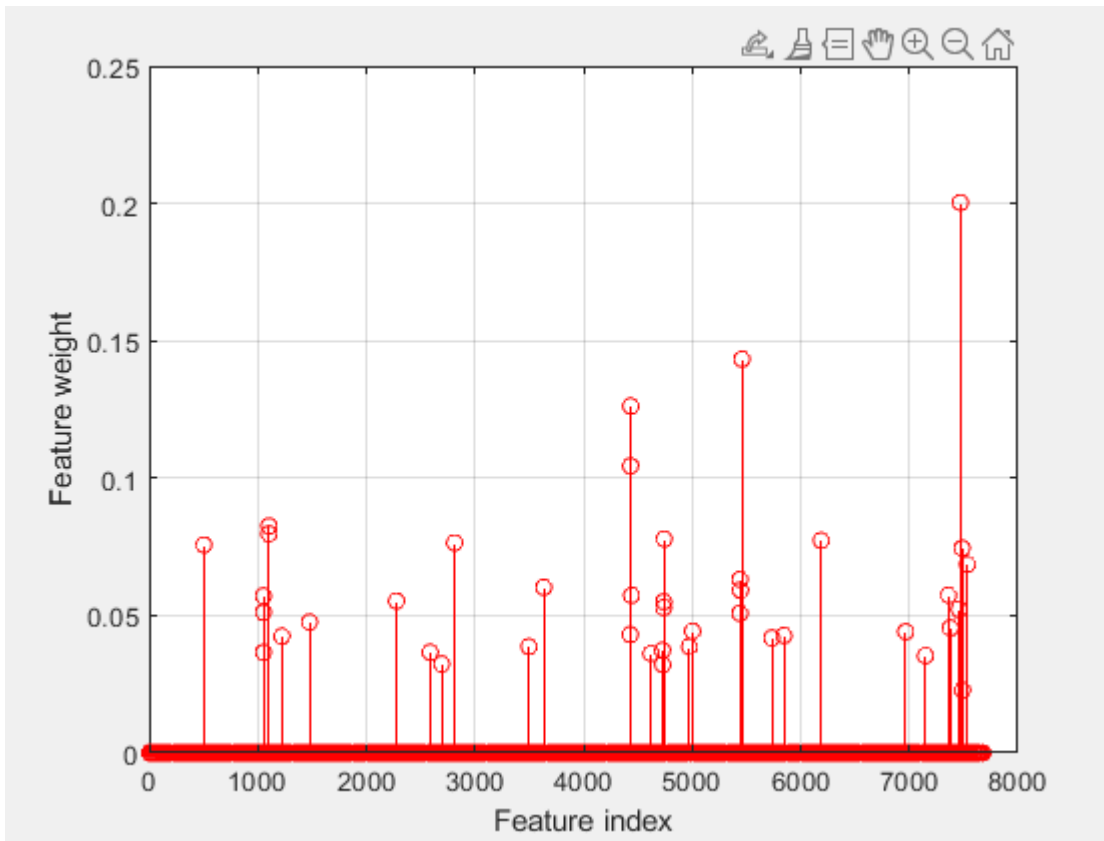
## 10. Reducción de Dimensionalidad

### 10.1. FSCNCA

El análisis de componentes de vecindad (NCA) es un método no paramétrico para seleccionar características con el objetivo de maximizar la precisión de predicción de los algoritmos de regresión y clasificación.

Fscnca aprende los pesos de las características mediante una adaptación diagonal de NCA y devuelve una instancia de un objeto FeatureSelectionNCAClassification. La función logra la selección de características al regularizar los pesos de las características y retorna un Modelo de análisis de componentes de vecindad para clasificación, devuelto como un objeto FeatureSelectionNCAClassification (Ver figura 2).

**Figura 2** Variables Relevantes FSCNCA



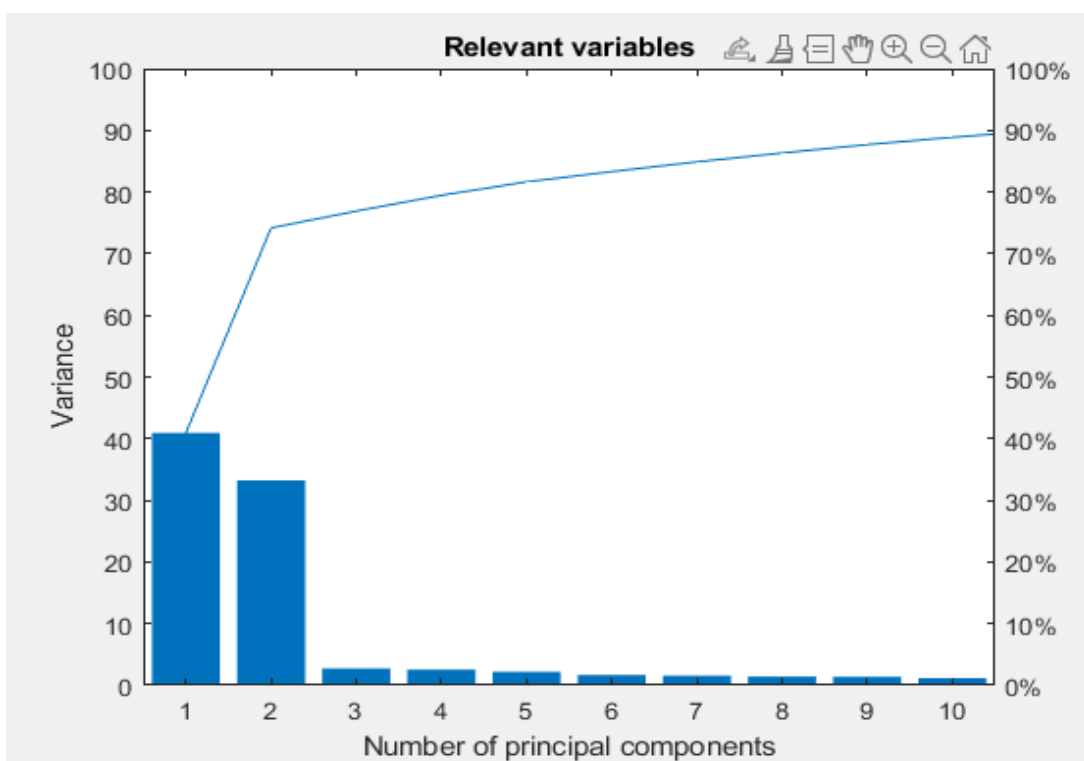
Se toman todos los datos por encima de 0.01 para su uso, el resto de ellos es descartado.

## 10.2. PCA

Se hizo uso de una función llamada “pac”, la cual requiere un único un argumento de entrada, el cual es la matriz de training. La función transforma ortogonalmente las  $n$  dimensiones numéricas originales de un conjunto de datos y da como resultado un nuevo conjunto de  $n$  dimensiones conocido como los coeficientes de componentes principales, el primer componente principal tiene la máxima varianza potencial. Cada componente principal sucesivo tiene la máxima varianza potencial bajo la limitación de que es ortogonal a los componentes principales precedentes, es decir, no está correlacionado con ellos. Mantener solo los primeros  $m < n$  componentes principales reduce la dimensionalidad de los datos, además también devuelve las

puntuaciones de componentes principales, devueltas como una matriz y las varianzas del componente principal, es decir, los valores propios de la matriz de covarianza de training, devueltos como vector de columna. Se halló la varianza acumulada de los componentes principales y se estableció como criterio de parada el 90% de la misma, es decir, cuando la varianza acumulada sea igual al 90%, los componentes principales que aún no han sido sumados serán eliminados (ver figura 3) (Directions, 2007).

**Figura 3** variables relevantes PCA



## 11. Resultados

En esta sección se presentan los resultados obtenidos luego de aplicar cada uno de los métodos nombrados anteriormente, así mismo, los resultados para cada uno de los métodos de reducción de dimensionalidad aplicados y por último el análisis de dichos resultados.

Para cada registro se decidió tomar una hora de duración de registro lo cual equivale a un total de 7680 variables por registro, luego de aplicarse los métodos de reducción de dimensionalidad dieron los siguientes resultados: 375 variables para PCA y 41 variables para FSCNCA. Después de obtenidos los resultados se aplicaron las diferentes metodologías.

### 11.1. Reconocedor Euclídeo

A continuación, se presentan los resultados obtenidos luego de aplicar el método reconocedor euclídeo y los dos métodos de reducción de dimensionalidad mencionados anteriormente, así mismo se muestra la media de porcentaje de clasificación resultante de cada modelo y la desviación estándar.

**Tabla 6.** Resultados Reconocedor Euclídeo

	Reconocedor Euclidean						
	Distancias						
	Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
Media	52,3004	52,6296	53,3004	57,5432	77,6996	85,3374	87,2798
Desv. Estándar	1,3435	3,0505	1,3435	2,726	2,7793	2,6835	2,4294

### 11.2. Reconocedor Euclídeo FSCNCA



**Tabla 7.** Resultados Reconocedor Euclídeo FSCNCA

Reconocedor Euclideo FSCNCA							
Distancias							
	Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
Media	79,3210	81,7901	79,3210	79,6296	85,4938	83,0247	82,7160
Desv. Estándar	± 6,5473	± 3,0554	± 6,5473	± 6,9838	± 2,1824	± 5,6743	± 6,9838

### 11.3. Reconocedor euclídeo PCA

**Tabla 8.** Resultados reconocedores Euclídeo PCA

Reconocedor Euclideo PCA							
Distancias							
	Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
Media	50,1811	51,7819	50,1811	52,9630	77,7613	77,,6584	79,5844
Desv. Estándar	4,4745	4,3434	4,4745	3,9332	4,0196	5,1276	4,4121

### 11.4. KNN

A continuación, se presentan los resultados obtenidos luego de aplicar el método KNN y los dos métodos de reducción de dimensionalidad mencionados anteriormente, así mismo se muestra la media de porcentaje de clasificación resultante de cada modelo, el valor de K correspondiente y la desviación estándar.

**Tabla 9.** Resultados KNN K=1

		KNN						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=1	Media	21,1640212	19,5767196	21,1640212	18,9594356	15,5202822	16,0493827	16,313933
	Desv. Estándar	1,45702925	1,36708007	1,45702925	2,48743951	1,86649122	3,16765521	1,84694628

**Tabla 10.** Resultados KNN K=2

		KNN						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=2	Media	24,6031746	22,6631393	24,6031746	18,6948854	15,6084656	21,6049383	20,2821869
	Desv. Estándar	1,83215074	2,03841722	1,83215074	3,55879564	3,3756325	5,07776242	4,1779465

**Tabla 11.** Resultados KNN K=3

		KNN						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=3	Media	27,9541446	27,1604938	27,9541446	17,5485009	16,9312169	23,9858907	22,7513228
	Desv. Estándar	2,18865496	3,32417581	2,18865496	3,88541132	3,561344	4,08572613	3,44480158

**Tabla 12.** Resultados KNN K=4

		KNN						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=4	Media	30,335097	29,9823633	30,335097	18,5185185	19,8412698	25,6613757	24,691358
	Desv. Estándar	3,65934637	3,48928624	3,65934637	3,73783994	3,51777068	4,10787115	3,58166568

**Tabla 13.** Resultados KNN K=5

		KNN						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=5	Media	31,7460317	32,4514991	31,7460317	20,0176367	21,2522046	27,6895944	27,2486772
	Desv. Estándar	3,02105986	2,97872059	3,02105986	2,8029754	3,32144548	4,25326965	5,89313384

**Tabla 14.** Resultados KNN K=6

		KNN						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=6	Media	32,010582	32,8042328	32,010582	20,3703704	21,957672	28,4832451	28,659612
	Desv. Estándar	2,65332959	3,72811863	2,65332959	3,45532009	3,561344	3,89474005	6,06759188

**Tabla 15.** Resultados KNN K=7

		KNN						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=7	Media	30,952381	33,9506173	30,952381	20,9876543	22,0458554	29,5414462	30,4232804
	Desv. Estándar	2,70077475	4,39385936	2,70077475	3,45532009	3,39439361	3,77887655	5,7212853

**Tabla 16.** Resultados KNN K=8

		KNN						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=8	Media	30,6878307	34,7442681	30,6878307	20,7231041	22,1340388	28,2186949	29,1005291
	Desv. Estándar	3,0390247	4,65649661	3,0390247	3,81947228	4,61146741	4,24259108	6,57562987

**Tabla 17.** Resultados KNN K=9

		KNN						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=9	Media	30,335097	34,2151675	30,335097	20,8994709	22,6631393	27,6014109	29,1887125
	Desv. Estándar	3,67666004	4,34819685	3,67666004	3,74511437	4,7776646	3,93183365	4,94744885

**Tabla 18.** Resultados KNN K=10

		KNN						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=10	Media	29,3650794	34,4797178	29,3650794	21,2522046	22,1340388	28,3068783	28,5714286
	Desv. Estándar	3,95024984	4,52808489	3,95024984	3,78607213	3,87840005	4,38559245	3,41305158

**Tabla 19.** Resultados KNN K=11

		KNN						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=11	Media	28,8359788	34,303351	28,8359788	21,6049383	22,4867725	27,4250441	27,9541446
	Desv. Estándar	3,76685337	4,67205721	3,76685337	4,24686573	3,81947228	3,7523747	3,39439361

**Tabla 20.** Resultados KNN K=12

		KNN						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=12	Media	29,6296296	33,3333333	29,6296296	21,2522046	23,1040564	28,042328	27,9541446
	Desv. Estándar	4,15617487	4,26179325	4,15617487	4,1844559	3,41570869	3,7523747	3,26079988

**Tabla 21.** Resultados KNN K=13

		KNN						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=13	Media	29,4532628	33,2451499	29,4532628	21,6049383	22,4867725	28,1305115	27,2486772
	Desv. Estándar	3,86668634	4,59767534	3,86668634	3,59935302	3,71837189	3,32144548	3,38904389

**Tabla 22.** Resultados KNN K=14

		KNN						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=14	Media	30,5114638	32,3633157	30,5114638	22,6631393	23,5449735	27,1604938	26,6313933
	Desv. Estándar	3,8360633	4,49187572	3,8360633	4,07460848	5,91464731	4,06346042	3,79564502

**Tabla 23.** Resultados KNN K=15

		KNN						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=15	Media	30,7760141	31,9223986	30,7760141	23,3686067	22,3104056	26,7195767	26,1904762
	Desv. Estándar	2,70077475	4,80417583	2,70077475	3,71104505	6,56596489	4,40417118	2,97872059

**Tabla 24.** Resultados KNN K=16

		KNN						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=16	Media	31,6578483	31,3051146	31,6578483	22,9276896	23,015873	26,0141093	26,5432099
	Desv. Estándar	2,27403826	3,45006484	2,27403826	3,51777068	6,21093477	3,72811863	2,13833433

**Tabla 25.** Resultados KNN K=17

		KNN						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=17	Media	32,7160494	31,0405644	32,7160494	22,6631393	22,1340388	26,2786596	25,1322751
	Desv. Estándar	2,36401755	2,82232874	2,36401755	3,16192189	6,25895255	4,37730993	3,08053525

**Tabla 26.** Resultados KNN K=18

		KNN						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=18	Media	32,4514991	29,8941799	32,4514991	22,8395062	22,0458554	26,2786596	25,308642
	Desv. Estándar	1,84694628	2,87014121	1,84694628	3,06577457	5,83122946	4,42062006	3,49188534

**Tabla 27.** Resultados KNN K=19

		KNN						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=19	Media	33,0687831	30,4232804	33,0687831	23,1040564	21,8694885	24,691358	24,691358
	Desv. Estándar	2,41339192	1,45702925	2,41339192	3,41570869	5,26029693	4,29149251	3,49188534

**Tabla 28.** Resultados KNN K=20

		KNN						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=20	Media	32,5396825	31,3932981	32,5396825	23,2804233	21,7813051	25,1322751	24,7795414
	Desv. Estándar	3,33779395	1,93334317	3,33779395	3,81709625	5,49480139	4,8436706	4,1779465

Las siguientes gráficas muestran el porcentaje de acierto según el número de K para las matrices de training y test.

**Figura 4** KNN EUCLIDEAN

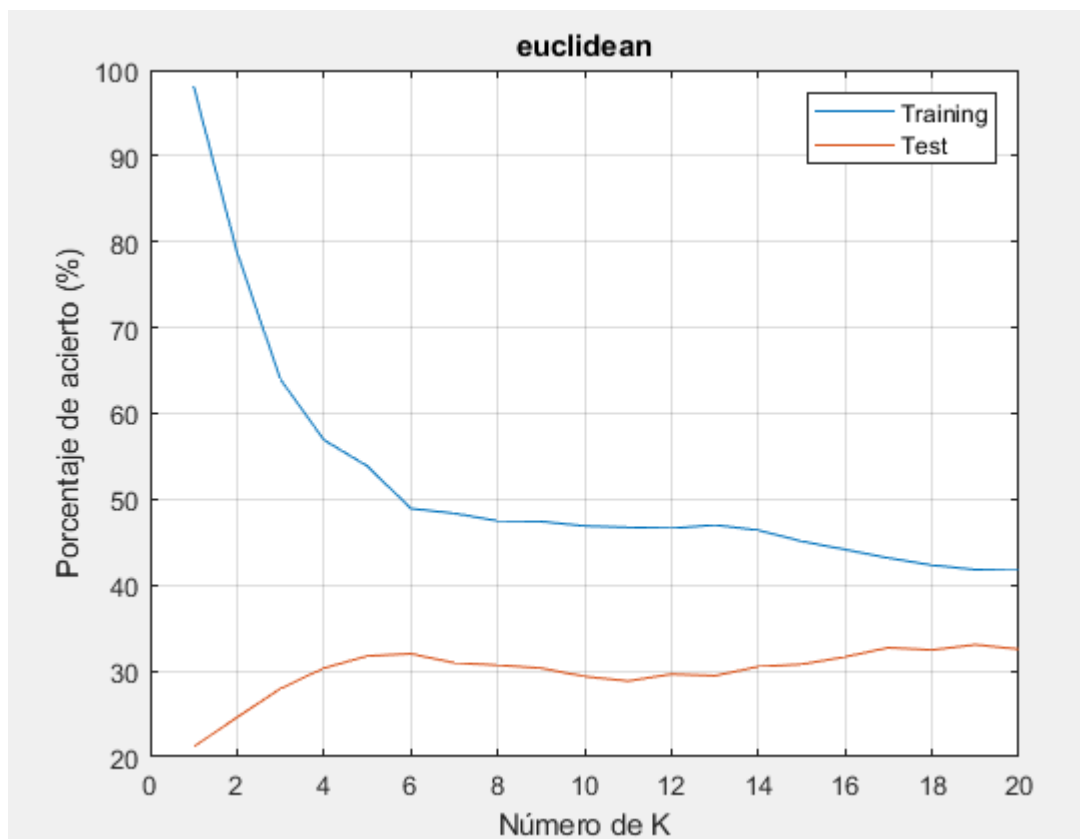


Figura 5 KNN CITYBLOCK

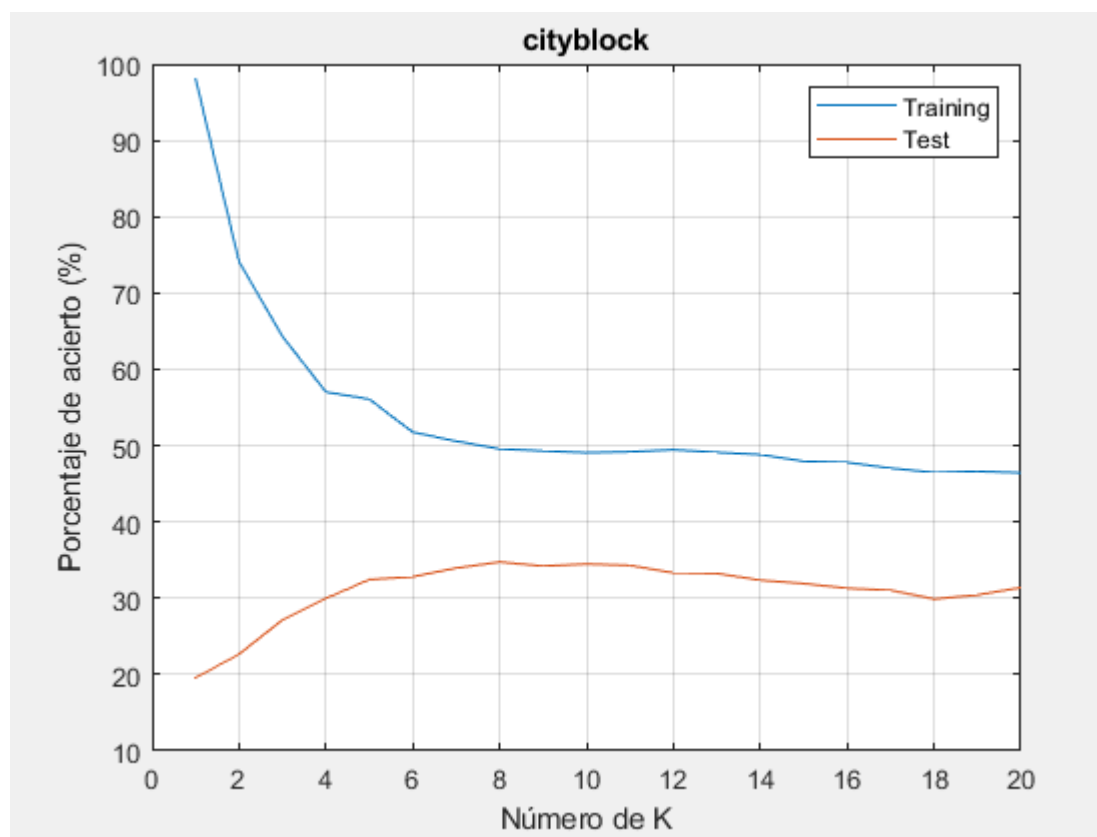
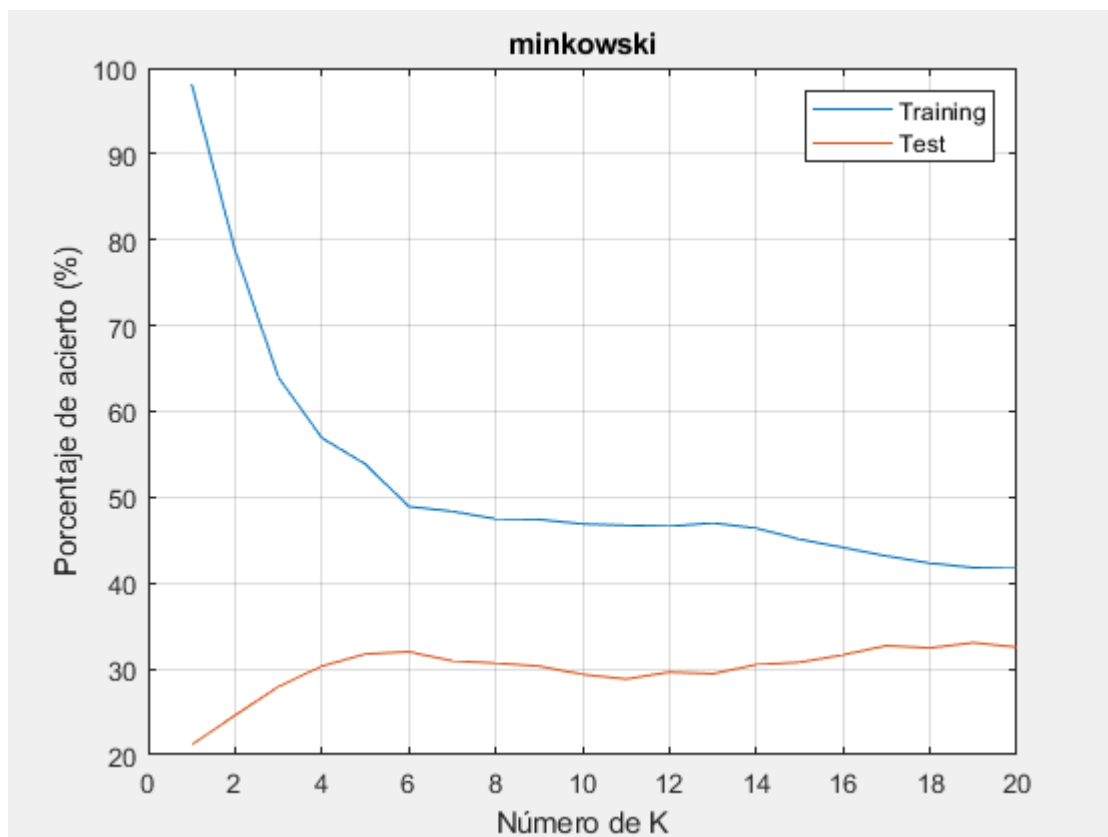


Figura 6 KNN MINKOWSKI





**Figura 7** KNN CHEBYCHEV

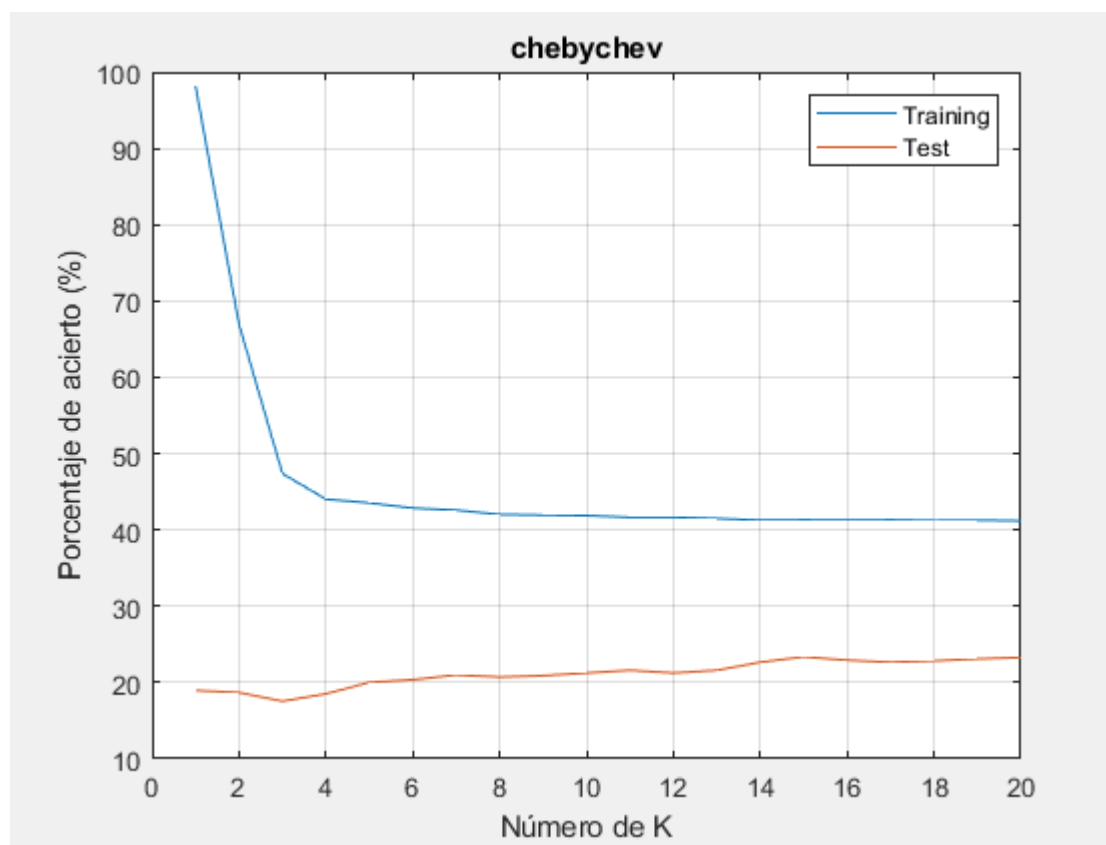


Figura 8 KNN COSINE

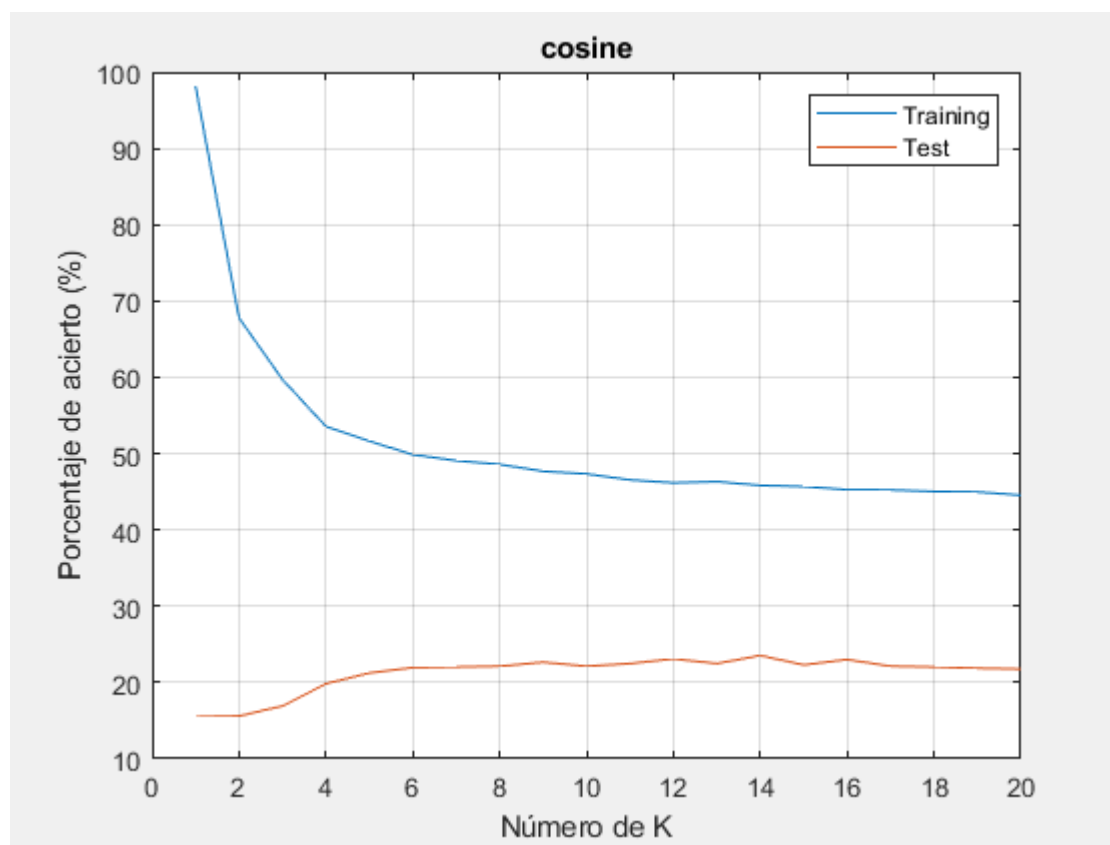


Figura 9 KNN CORRELATION

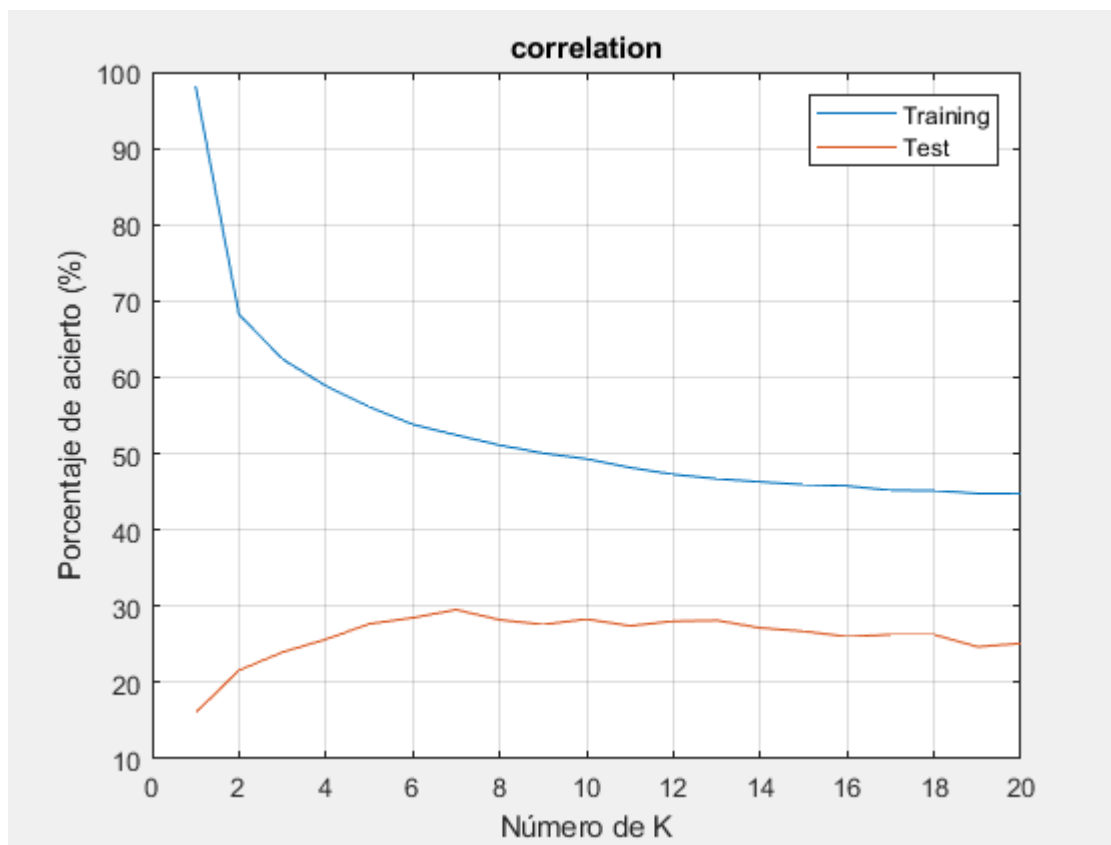
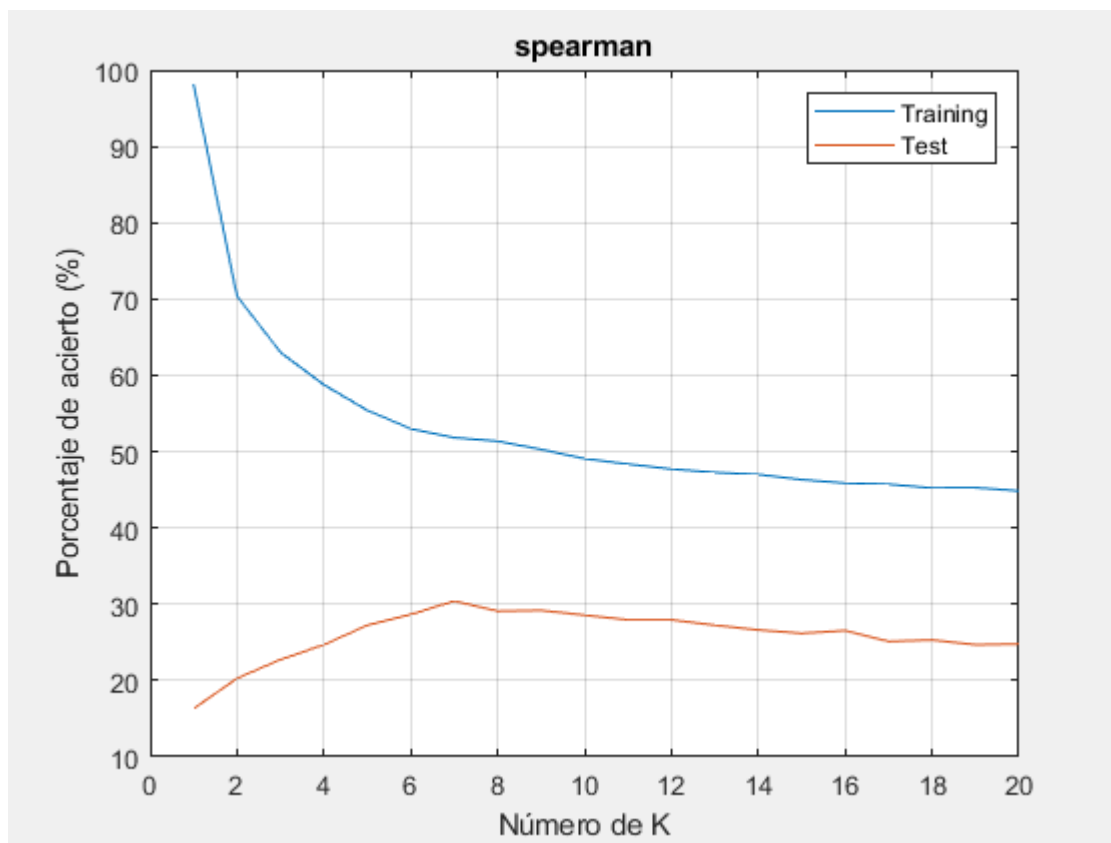


Figura 10 KNN SPEARMAN



## 11.5. KNN FSCNCA

**Tabla 29.** Resultados KNN FSCNCA K=1

		KNN FSCNCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=1	Media	11,1111111	13,5802469	11,1111111	24,0740741	14,1975309	11,7283951	12,345679
	Desv. Estándar	3,9275426	2,2295419	4,2254863	4,8294650	1,7018582	4,4067827	4,7615177

**Tabla 30.** Resultados KNN FSCNCA K=2

		KNN FSCNCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=2	Media	10,4938272	16,0493827	10,4938272	27,7777778	18,5185185	16,6666667	16,0493827
	Desv. Estándar	2,1202566	4,7162489	4,9925026	3,5504352	2,2558244	1,3444788	4,5568328

**Tabla 31.** Resultados KNN FSCNCA K=3

		KNN FSCNCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=3	Media	13,5802469	19,7530864	13,5802469	33,9506173	24,0740741	22,8395062	20,9876543
	Desv. Estándar	4,8808374	2,4858423	2,6871048	1,6318029	4,5792656	2,2204576	3,5322149

**Tabla 32.** Resultados KNN FSCNCA K=4

		KNN FSCNCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=4	Media	17,2839506	22,8395062	17,2839506	32,0987654	24,0740741	22,8395062	23,4567901
	Desv. Estándar	1,1824941	1,5798010	1,0503460	2,8063367	4,2371251	1,9076031	4,9850017

**Tabla 33.** Resultados KNN FSCNCA K=5

		KNN FSCNCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=5	Media	23,4567901	24,0740741	23,4567901	32,7160494	24,691358	22,2222222	25,308642
	Desv. Estándar	1,4327856	2,8643231	1,3033447	1,2955996	3,6598727	2,1056211	3,5793327

**Tabla 34.** Resultados KNN FSCNCA K=6

		KNN FSCNCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=6	Media	25,308642	25,9259259	25,308642	33,3333333	25,308642	24,0740741	25,9259259
	Desv. Estándar	3,9860963	1,6359769	3,2392943	1,3672151	1,1468974	3,9181251	2,3810744

**Tabla 35.** Resultados KNN FSCNCA K=7

		KNN FSCNCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=7	Media	25,308642	27,1604938	25,308642	32,7160494	25,308642	26,5432099	29,0123457
	Desv. Estándar	2,5070509	4,8363097	2,9237303	2,2529876	3,0740484	4,2066594	1,7468023

**Tabla 36.** Resultados KNN FSCNCA K=8

		KNN FSCNCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=8	Media	23,4567901	27,1604938	23,4567901	31,4814815	26,5432099	27,1604938	28,3950617
	Desv. Estándar	4,8166392	4,3201972	2,3108661	2,2743503	3,3002046	1,5314290	4,1978898

**Tabla 37.** Resultados KNN FSCNCA K=9

		KNN FSCNCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=9	Media	27,7777778	28,3950617	27,7777778	31,4814815	25,308642	25,9259259	27,1604938
	Desv. Estándar	1,8411082	1,9273688	2,4032820	4,5025633	1,3671910	3,9560465	1,4643337

**Tabla 38.** Resultados KNN FSCNCA K=10

		KNN FSCNCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=10	Media	30,8641975	30,8641975	30,8641975	32,0987654	25,9259259	25,308642	27,7777778
	Desv. Estándar	3,9584227	2,0945792	4,7481001	3,9495394	1,1732409	4,1928115	2,4939942

**Tabla 39.** Resultados KNN FSCNCA K=11

		KNN FSCNCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=11	Media	32,0987654	29,6296296	32,0987654	33,9506173	26,5432099	27,1604938	27,7777778
	Desv. Estándar	1,0301031	4,6089472	2,8506032	1,3131619	3,9580818	3,9968270	3,8530491

**Tabla 40.** Resultados KNN FSCNCA K=12

		KNN FSCNCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=12	Media	36,4197531	29,6296296	36,4197531	33,3333333	27,1604938	29,6296296	29,6296296
	Desv. Estándar	2,0904833	3,5042497	4,6578579	4,9421349	1,4115588	1,2100524	2,1625047

**Tabla 41.** Resultados KNN FSCNCA K=13

		KNN FSCNCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=13	Media	33,9506173	32,7160494	33,9506173	34,5679012	27,1604938	28,3950617	30,2469136
	Desv. Estándar	4,6263477	1,4842201	1,2766170	3,8910719	2,5272265	1,8122762	4,7884833

**Tabla 42.** Resultados KNN FSCNCA K=14

		KNN FSCNCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=14	Media	31,4814815	36,4197531	31,4814815	36,4197531	26,5432099	29,6296296	31,4814815
	Desv. Estándar	3,9972434	1,1601339	1,0347130	4,3945157	2,3797030	3,5842500	3,3939066

**Tabla 43.** Resultados KNN FSCNCA K=15

		KNN FSCNCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=15	Media	30,2469136	34,5679012	30,2469136	36,4197531	26,5432099	29,0123457	30,2469136
	Desv. Estándar	1,6812104	4,7394968	4,5097268	1,7440014	1,4042465	2,1804297	2,7763785

**Tabla 44.** Resultados KNN FSCNCA K=16



		KNN FSCNCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=16	Media	37,037037	32,7160494	37,037037	35,8024691	25,9259259	29,6296296	29,6296296
	Desv. Estándar	2,5510326	2,5190322	1,7736928	1,9123144	3,3451250	3,5380078	1,1108702

**Tabla 45.** Resultados KNN FSCNCA K=17

		KNN FSCNCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=17	Media	35,1851852	33,3333333	35,1851852	35,8024691	27,7777778	29,6296296	29,0123457
	Desv. Estándar	4,0005870	2,2273771	2,1722716	2,0752197	2,9906725	1,2697737	2,1049840

**Tabla 46.** Resultados KNN FSCNCA K=18

		KNN FSCNCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=18	Media	35,8024691	32,0987654	35,8024691	35,8024691	27,7777778	29,6296296	29,6296296
	Desv. Estándar	1,0750228	3,7941286	3,8840092	3,9593104	4,7575692	4,9863562	1,0076314

**Tabla 47.** Resultados KNN FSCNCA K=19

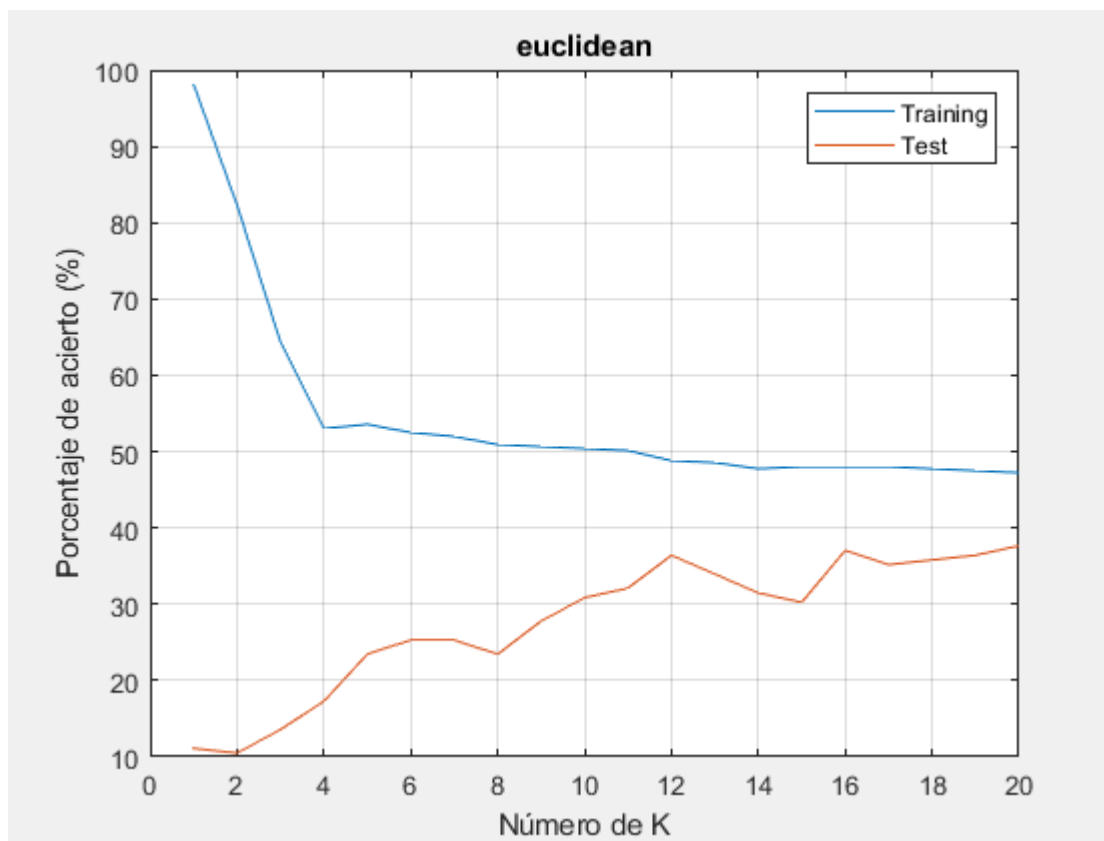
		KNN FSCNCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=19	Media	36,4197531	34,5679012	36,4197531	35,1851852	28,3950617	31,4814815	29,6296296
	Desv. Estándar	4,1222373	3,0281102	4,2993171	4,5735951	3,5003377	2,6279731	4,1712921

**Tabla 48.** Resultados KNN FSCNCA K=20

		KNN FSCNCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=20	Media	37,654321	36,4197531	37,654321	37,037037	27,1604938	32,0987654	30,2469136
	Desv. Estándar	4,6316866	2,2643081	1,5458980	4,9612823	2,9042849	1,2165024	1,7582594

Las siguientes gráficas muestran el porcentaje de acierto según el número de K para las matrices de training y test.

**Figura 11** KNN FSCNCA EUCLIDEAN



**Figura 12** KNN FSCNCA CITYBLOCK

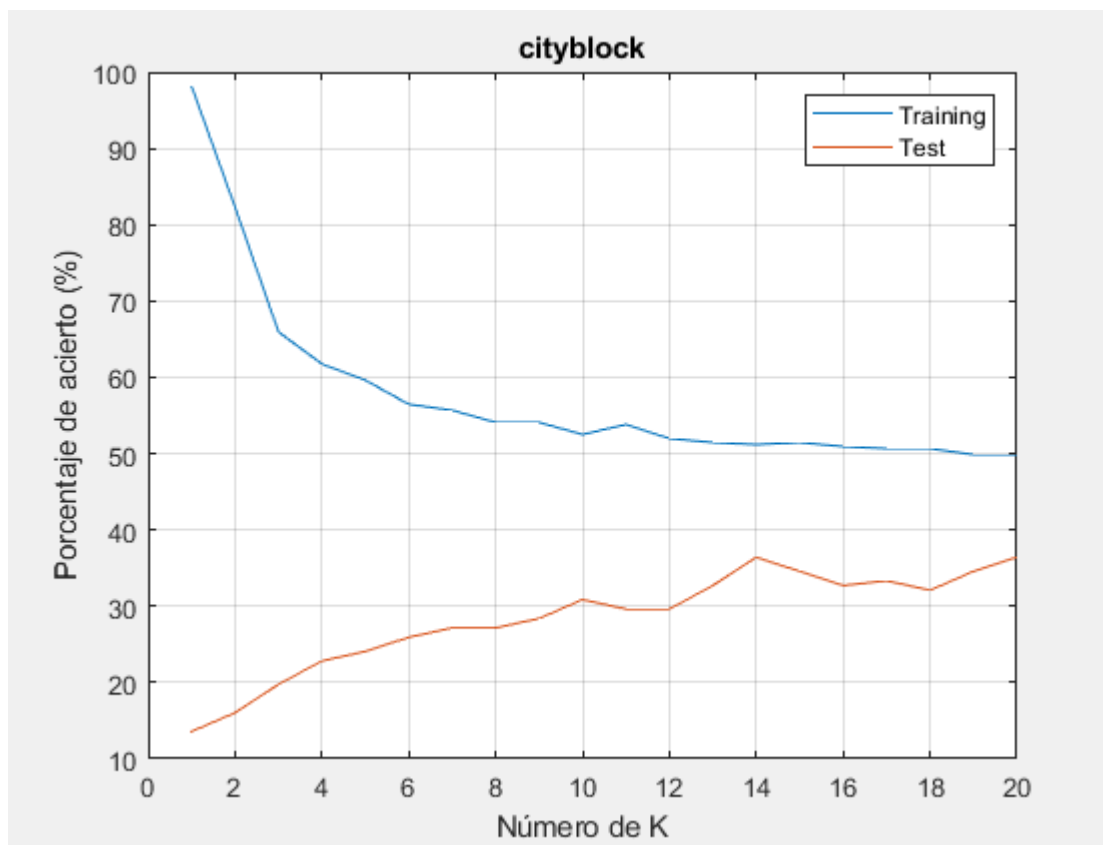


Figura 13 KNN FSCNCA MINKOWSKI

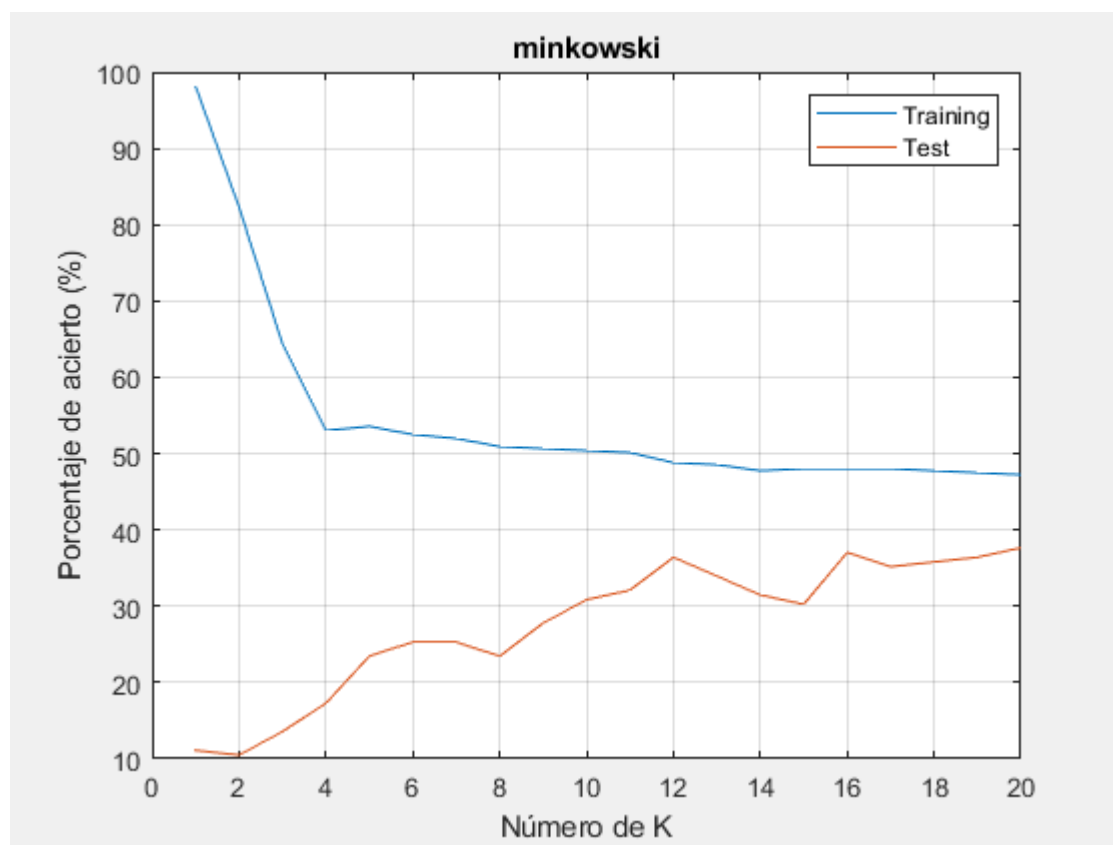


Figura 14 KNN FSCNCA CHEBYCHEV

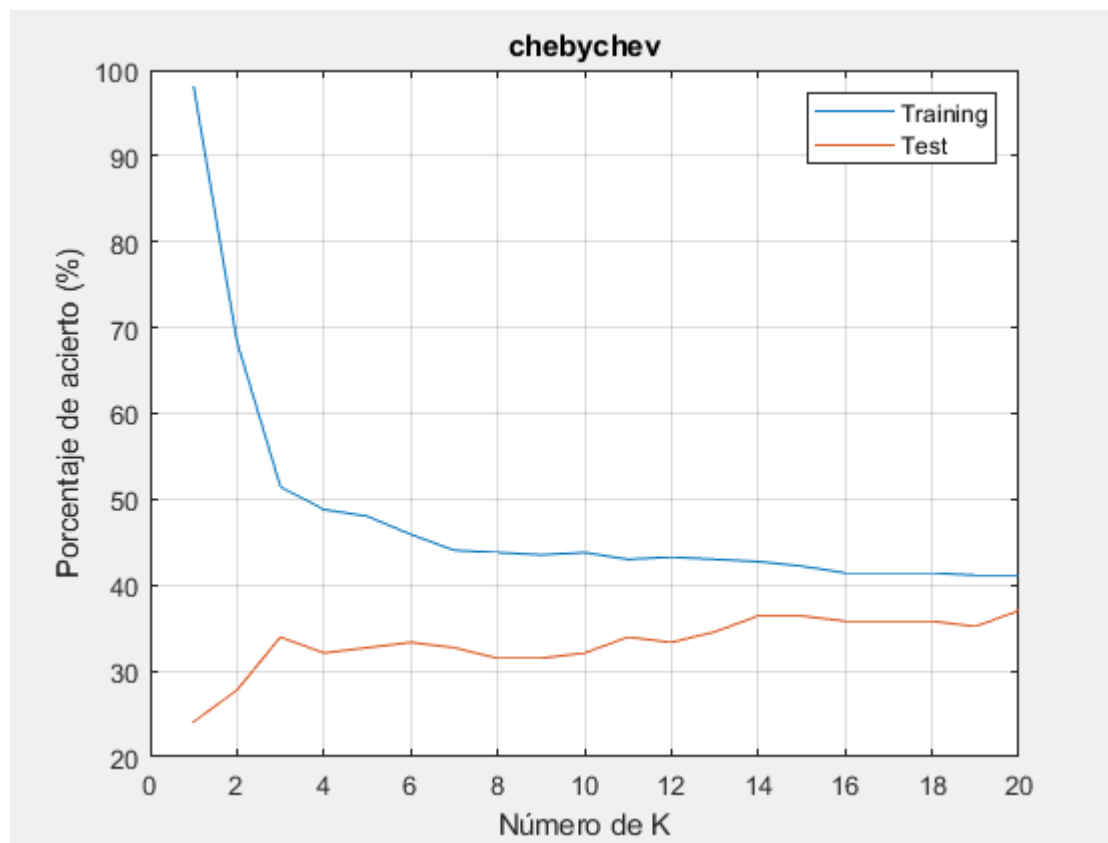
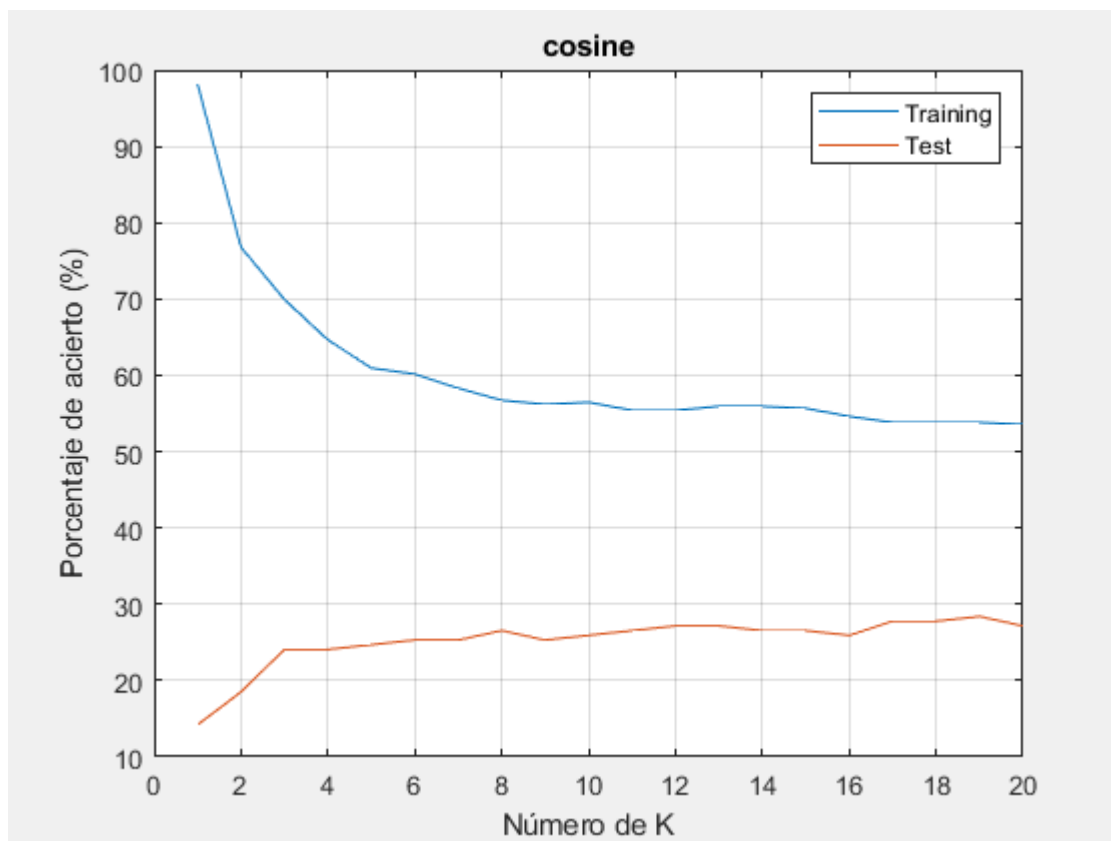
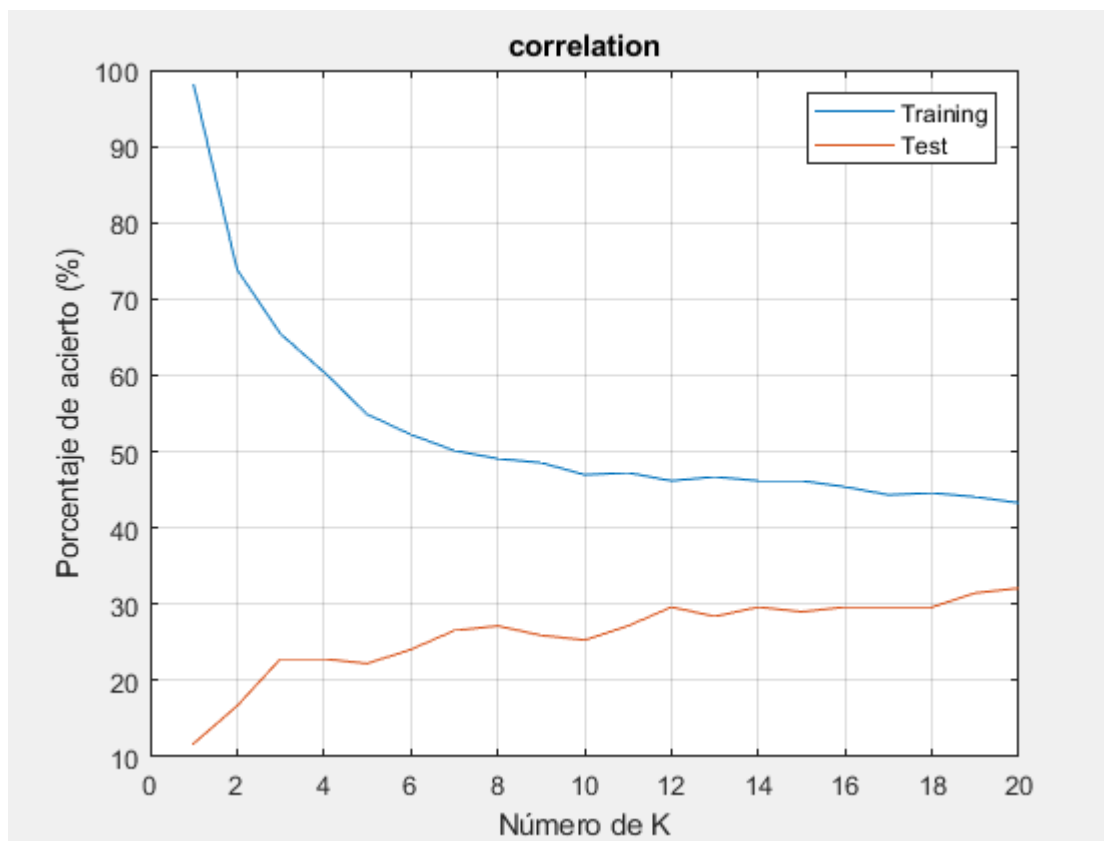


Figura 15 KNN FSCNCA COSINE

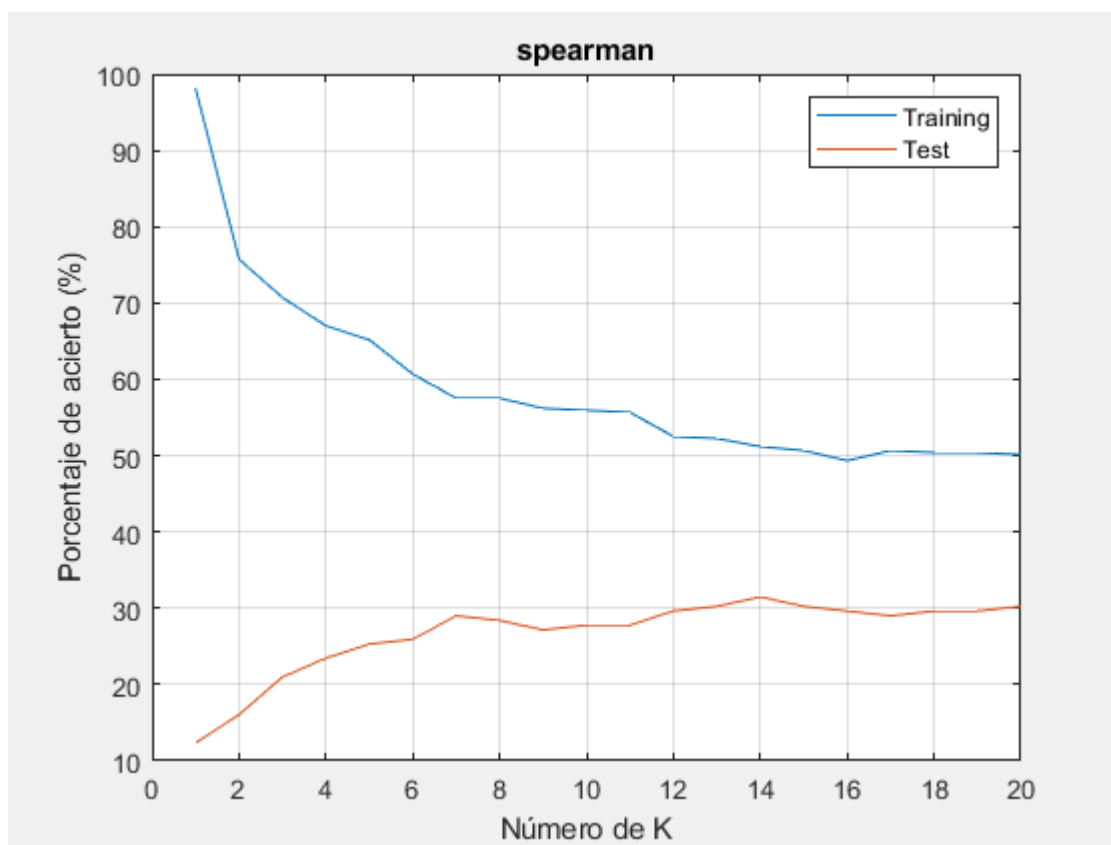


**Figura 16** KNN FSCNCA CORRELATION



**Figura 17** KNN FSCNCA SPEARMAN





## 11.6. KNN PCA

**Tabla 49.** Resultados KNN PCA K=1

		KNN PCA						
		Distancias						
K=1		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
		Media	15,3909465	15,1975309	15,3909465	15,6707819	18,0288066	18,4897119
	Desv. Estándar	2,77228692	2,53623914	2,77228692	2,61870568	2,64156473	2,56697437	6,2859537

**Tabla 50.** Resultados KNN PCA K=2

		KNN PCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=2	Media	21,4444444	21,1851852	21,4444444	21,526749	20,8353909	21,0288066	22,8353909
	Desv. Estándar	3,23206763	3,44790117	3,23206763	3,06880104	3,03413892	3,10555561	6,3074626

**Tabla 51.** Resultados KNN PCA K=3

		KNN PCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=3	Media	25,8395062	25,5144033	25,8395062	25,9465021	22,962963	22,9053498	25,0246914
	Desv. Estándar	3,57012549	3,53820495	3,57012549	3,55676761	3,41337269	3,58878433	6,17227191

**Tabla 52.** Resultados KNN PCA K=4

		KNN PCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=4	Media	28,654321	28,2222222	28,654321	29,1028807	24,8930041	24,7613169	25,8271605
	Desv. Estándar	3,89044379	3,67536873	3,89044379	3,66440517	3,4878526	3,88930425	6,2788478

**Tabla 53.** Resultados KNN PCA K=5

		KNN PCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=5	Media	30,9176955	30,4609053	30,9176955	31,2798354	26,1893004	26,2469136	26,9053498
	Desv. Estándar	4,07075635	3,86122523	4,07075635	3,51778557	3,71978905	4,01559172	6,52965995

**Tabla 54.** Resultados KNN PCA K=6

		KNN PCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=6	Media	32,9670782	32,2921811	32,9670782	33,0740741	27,308642	27,2510288	27,7407407
	Desv. Estándar	4,09215992	3,87256712	4,09215992	3,57959686	3,84642302	4,18074722	6,47461453

**Tabla 55.** Resultados KNN PCA K=7

		KNN PCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=7	Media	34,2139918	33,5020576	34,2139918	34,4279835	27,9753086	27,962963	28,691358
	Desv. Estándar	3,95826551	3,77340486	3,95826551	3,73543089	3,84285775	4,38225214	6,37129008

**Tabla 56.** Resultados KNN PCA K=8

		KNN PCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=8	Media	35,1111111	34,3333333	35,1111111	35,0658436	28,6625514	28,5514403	29,4897119
	Desv. Estándar	4,02970489	3,86101549	4,02970489	3,99194394	3,89842	4,1842935	6,70433176

**Tabla 57.** Resultados KNN PCA K=9

		KNN PCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=9	Media	35,6419753	35,1316872	35,6419753	35,8024691	29,3127572	29,0864198	30,1234568
	Desv. Estándar	3,92262511	3,97154579	3,92262511	3,76329555	3,98427478	4,18474167	6,68288077

**Tabla 58.** Resultados KNN PCA K=10

		KNN PCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=10	Media	36,4979424	35,8806584	36,4979424	36,5185185	29,9300412	29,5925926	30,9176955
	Desv. Estándar	3,9942495	4,18069217	3,9942495	3,69892254	3,99645135	4,27501197	6,77067571

**Tabla 59.** Resultados KNN PCA K=11

		KNN PCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=11	Media	36,8641975	36,4979424	36,8641975	37,1769547	30,4115226	30,0740741	31,5843621
	Desv. Estándar	3,94841578	4,51761734	3,94841578	3,67816941	4,09813614	4,16157783	6,79197493

**Tabla 60.** Resultados KNN PCA K=12

		KNN PCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=12	Media	37,5884774	37,1193416	37,5884774	37,6707819	31,0123457	30,5432099	32,2592593
	Desv. Estándar	4,00182925	4,51329612	4,00182925	3,67455224	4,15207903	4,43732686	7,11488801

**Tabla 61.** Resultados KNN PCA K=13

		KNN PCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=13	Media	37,9465021	37,5925926	37,9465021	38,037037	31,5596708	30,9670782	32,9465021
	Desv. Estándar	4,13357071	4,6385282	4,13357071	3,75218224	4,21420115	4,49384757	7,15793147

**Tabla 62.** Resultados KNN PCA K=14

		KNN PCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=14	Media	38,2757202	37,7695473	38,2757202	38,382716	31,909465	31,4279835	33,4650206
	Desv. Estándar	4,23067478	4,38204011	4,23067478	3,59727068	4,47361714	4,6253808	7,24177042

**Tabla 63.** Resultados KNN PCA K=15

		KNN PCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=15	Media	38,7078189	38,18107	38,7078189	38,8600823	32,0617284	31,8559671	34,0534979
	Desv. Estándar	4,24036541	4,54419565	4,24036541	3,83354095	4,5238022	4,62591154	7,45332547

**Tabla 64.** Resultados KNN PCA K=16

		KNN PCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=16	Media	39,1851852	38,3703704	39,1851852	39,2098765	32,473251	32,345679	34,7119342
	Desv. Estándar	4,14717348	4,30591693	4,14717348	3,75426041	4,61871925	4,62541582	7,09616817

**Tabla 65.** Resultados KNN PCA K=17

		KNN PCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=17	Media	39,2633745	38,7695473	39,2633745	39,617284	32,9176955	32,6666667	34,8312757
	Desv. Estándar	4,1577518	4,33646404	4,1577518	3,78029324	4,61018463	4,67683559	7,10118648

**Tabla 66.** Resultados KNN PCA K=18

		KNN PCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=18	Media	39,600823	38,9176955	39,600823	39,8765432	33,2633745	33,1604938	35,1646091
	Desv. Estándar	4,01305415	4,15125157	4,01305415	3,92234477	4,72850564	4,64974501	6,85605748

**Tabla 67.** Resultados KNN PCA K=19

		KNN PCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=19	Media	39,9382716	39,127572	39,9382716	40,0493827	33,5308642	33,399177	35,3415638
	Desv. Estándar	4,07063909	4,07505736	4,07063909	3,79234928	4,6619388	4,72504848	6,58270825

**Tabla 68.** Resultados KNN PCA K=20

		KNN PCA						
		Distancias						
		Euclidean	Cityblock	Minkowski	Chebychev	Cosine	Correlation	Spearman
K=20	Media	40,2757202	39,4485597	40,2757202	40,1687243	33,6790123	33,6995885	35,382716
	Desv. Estándar	3,97733675	4,07603622	3,97733675	3,70377045	4,73819998	4,7863237	6,44495903

Las siguientes gráficas muestran el porcentaje de acierto dependiendo del número de K en los casos de training y test.

Figura 18 KNN PCA EUCLIDEAN

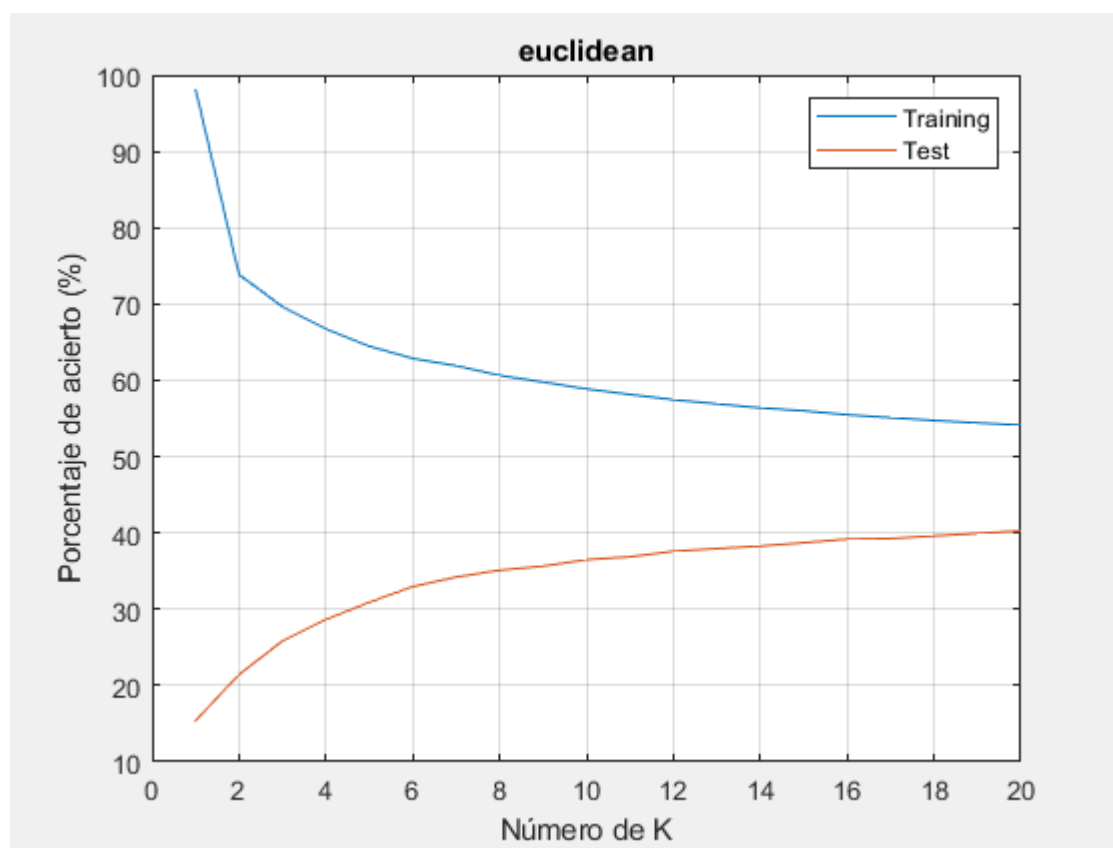


Figura 19 KNN PCA CITYBLOCK

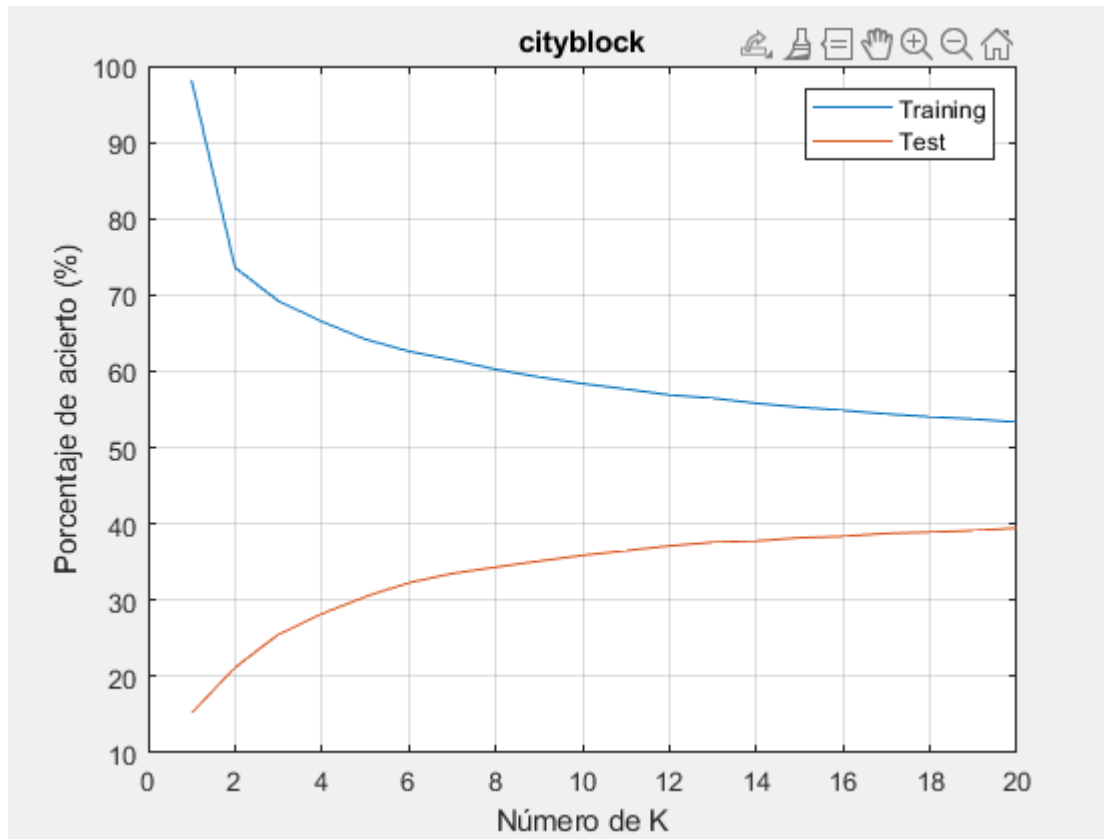




Figura 20 KNN PCA MINKOWSKI

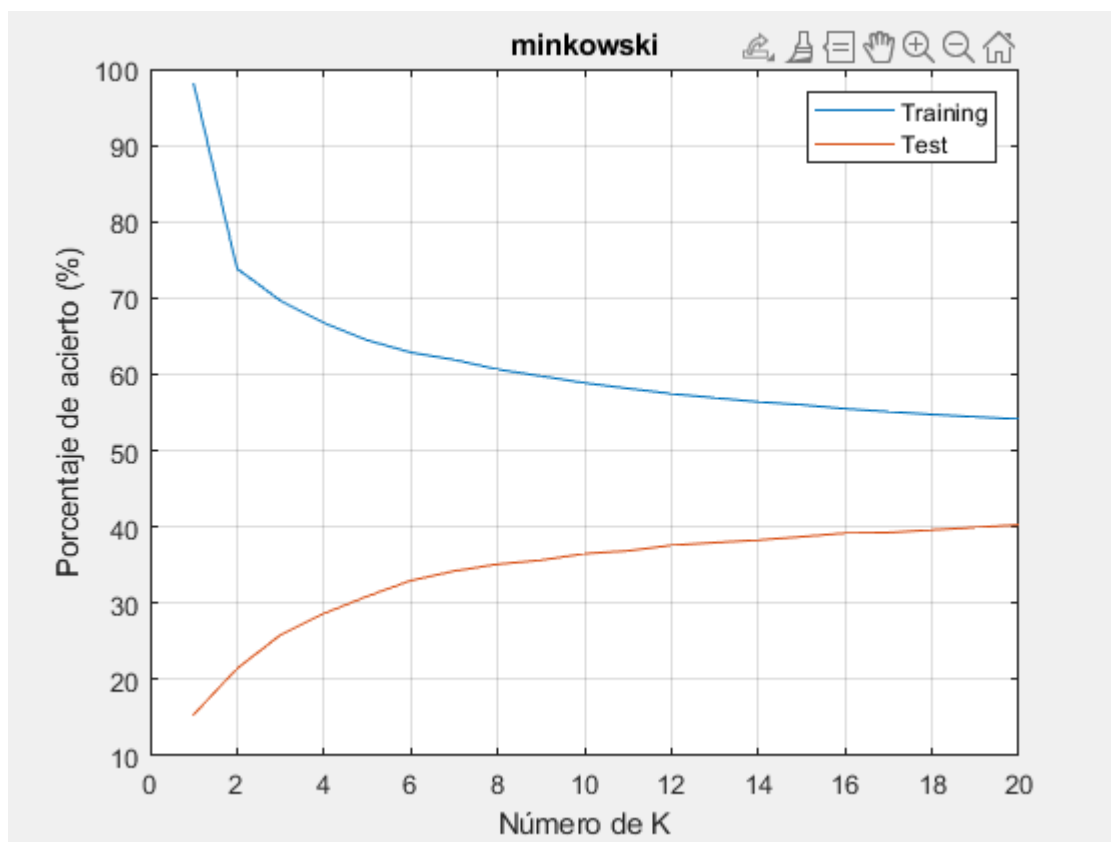


Figura 21 KNN PCA CHEBYCHEV

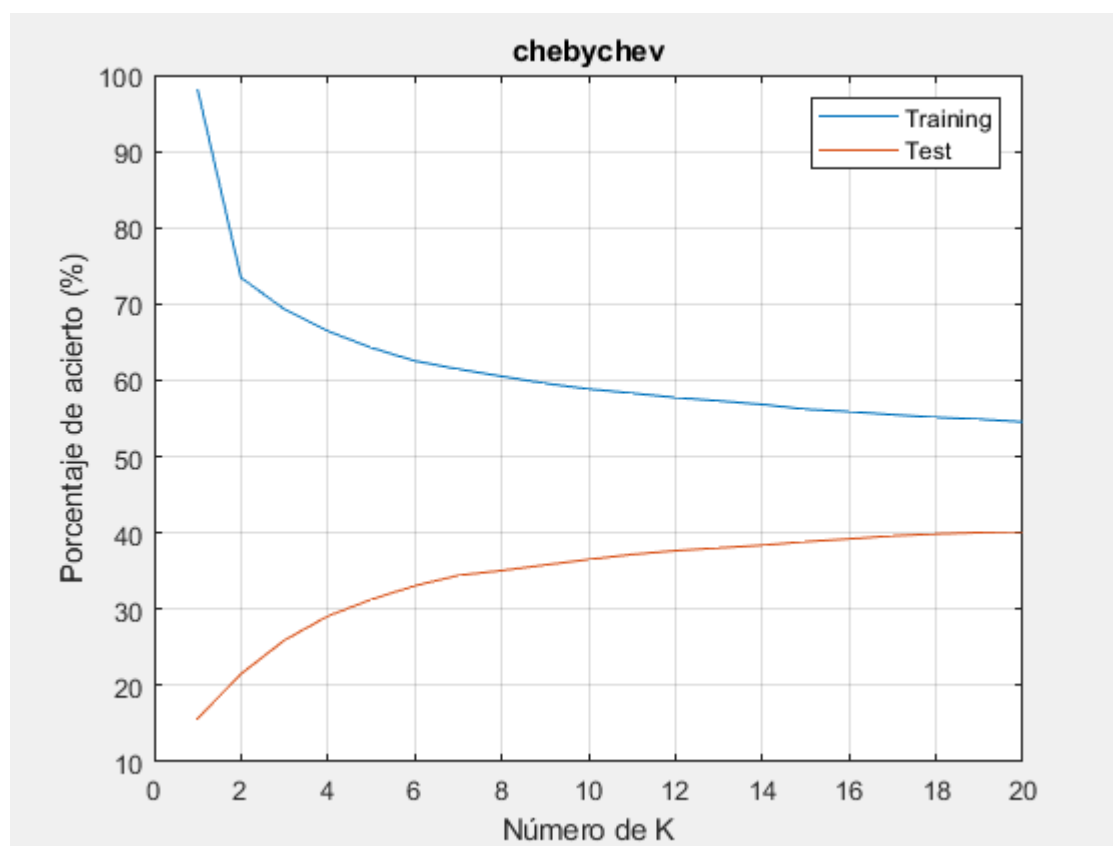


Figura 22 KNN PCA COSINE

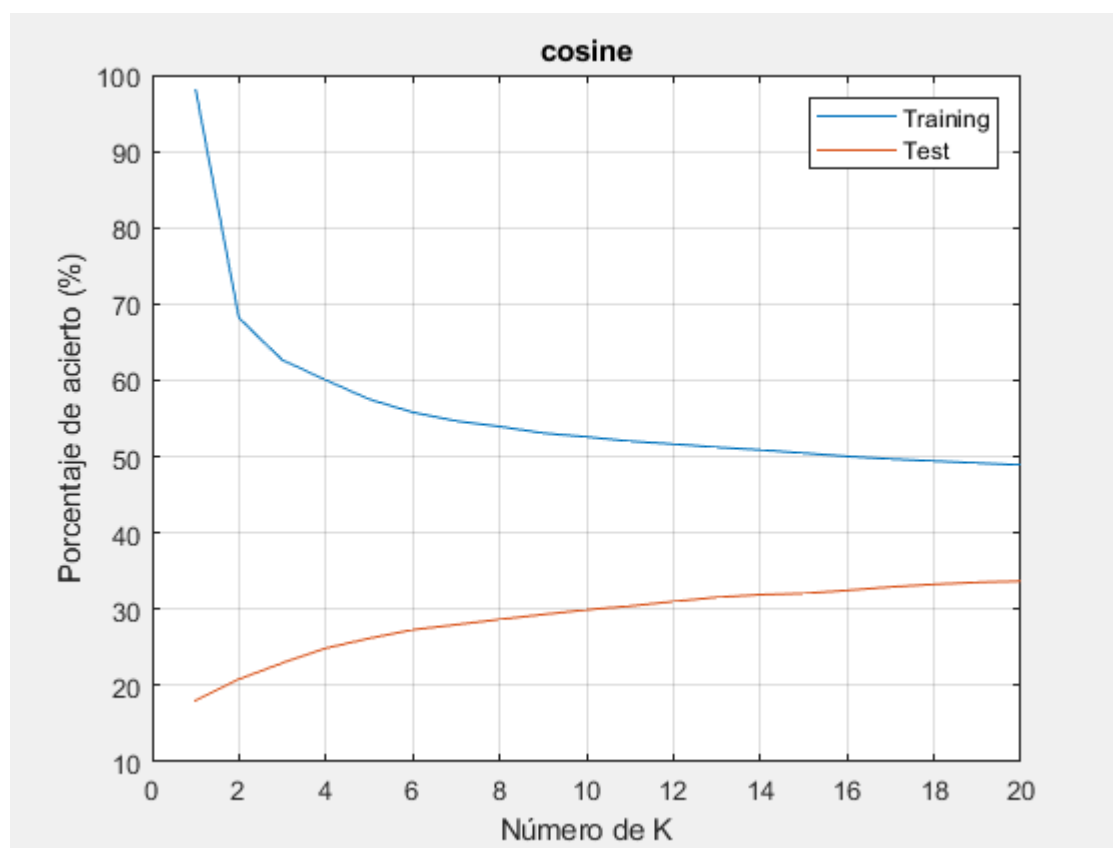
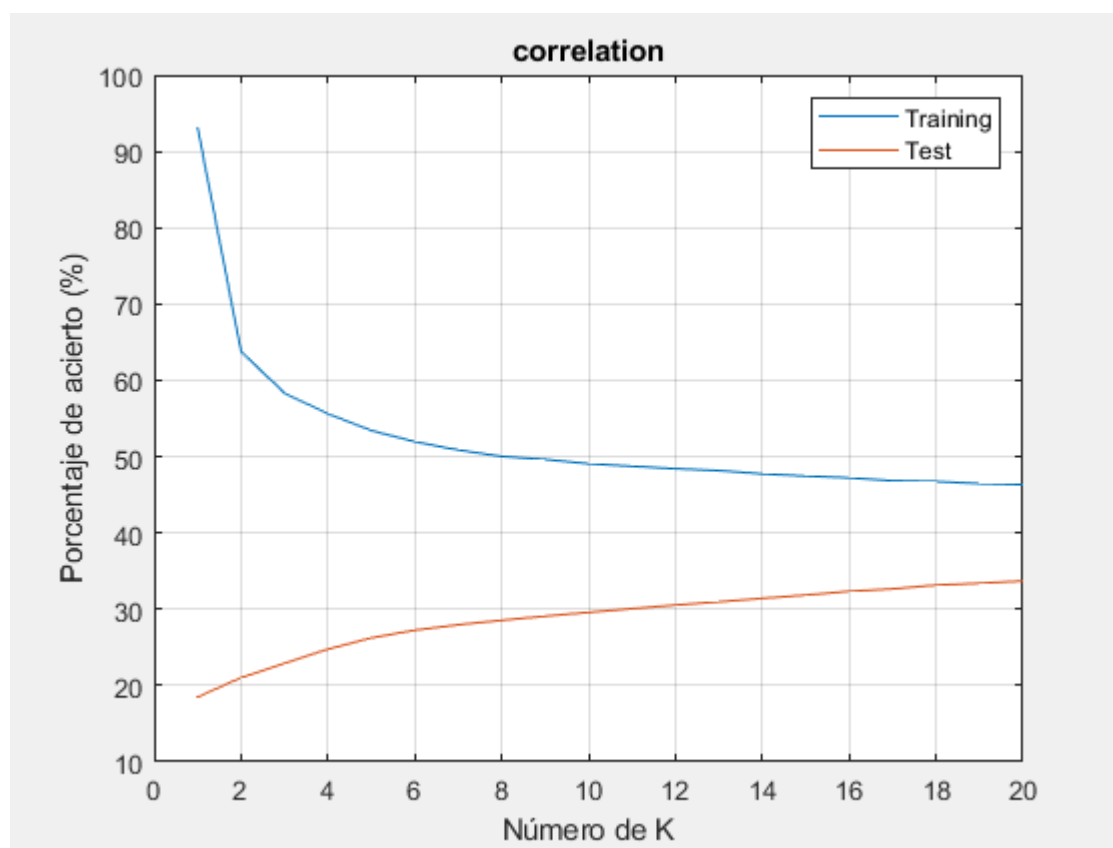
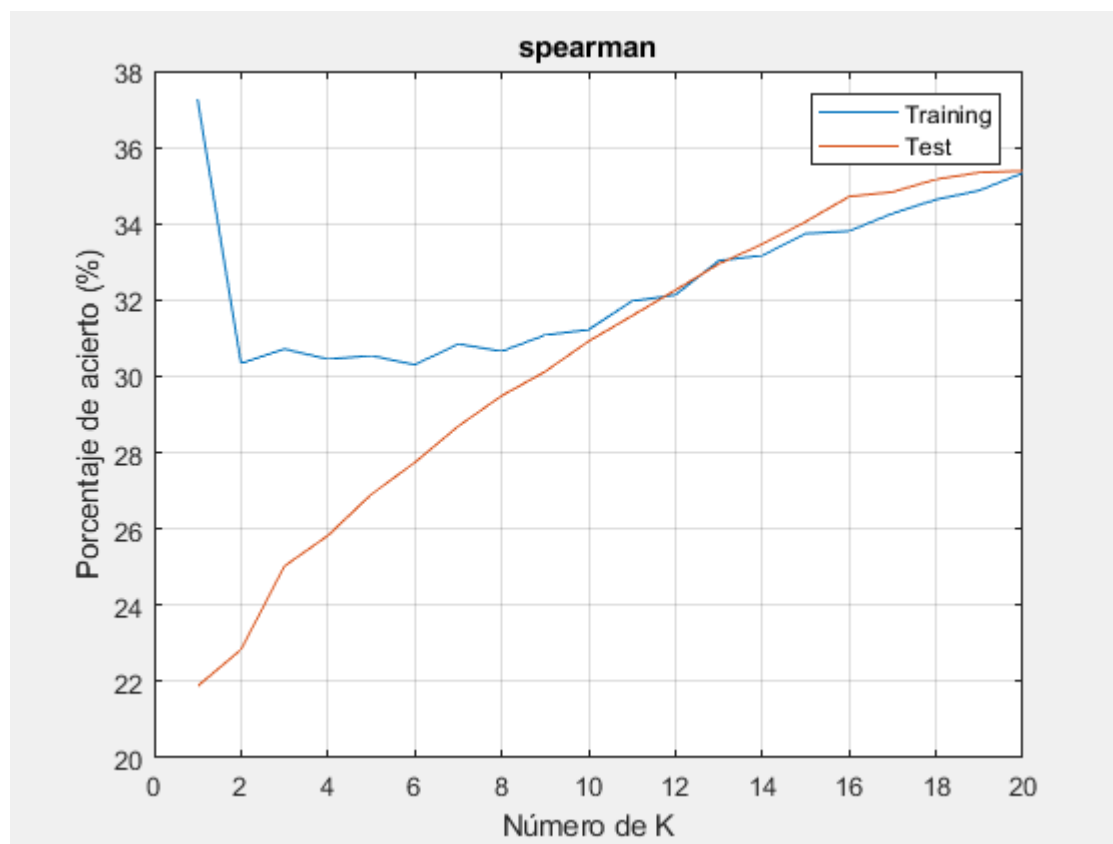


Figura 23 KNN PCA CORRELATION



**Figura 24** KNN PCA SPEARMAN



## 12. Conclusiones

Con un número inicial de 7680 se pudo observar un gran cambio en la cantidad de variables después de aplicar los respectivos métodos de reducción de dimensionalidad llegando a usarse hasta 41 variables por cada registro en el caso de FSCNCA y 375 con PCA, lo cual es esperado al aplicar dichos métodos, sin embargo al momento de comparar el rendimiento de los modelos con o sin métodos de reducción estos últimos tuvieron un mejor rendimiento, esto es debido a casos particulares donde la reducción de dimensionalidad por el contrario a su objetivo, no mejora el rendimiento, sino que lo disminuye.

Si observamos la media de clasificación máxima de reconocedor euclídeo sin métodos de reducción de dimensionalidad es de 87,2798% con una desviación estándar de 2,4294 en cambio aplicando FSCNCA en los datos se obtiene una media de clasificación máxima de 85,4938% con una desviación de 2,1824, en el caso de PCA la máxima media de clasificación fue de 79,5844% con una desviación estándar de 4,4121 lo cual confirma que en este caso la reducción de dimensionalidad no tuvo el resultado esperado con respecto al rendimiento, sin embargo al momento de comparar los valores máximos alcanzado en ejecuciones específicas los cuales son 93,3% para el reconocedor euclídeo sin métodos de reducción, 95,6% aplicando FSCNCA y PCA con 96,9% se puede observar cómo aplicando métodos de reducción de dimensionalidad se pudieron alcanzar valores máximos superiores al modelo sin métodos de reducción aplicados pero no llegando a una media superior.

En el caso de KNN la media alcanzada en los tres modelos fue mucho menor en comparación al método anterior llegando a alcanzar máximo una media de clasificación de 40% con PCA,

Metodo de Reduccion	KNN			Reconocedor Euclideo		
		PCA	FSCNCA		PCA	FSCNCA
Media de Clasificacion Maxima	34,479718	40,27572016	37,654321	87,2798	79,5844	85,4938
Desviacion Estandar	4,5280849	3,977336746	2,2661124	2,4294	4,4121	2,1824
K	10	20	20			

seguida de un 37% con FSCNCA y por último una media de 34% sin métodos de reducción de dimensionalidad.

**Tabla 69.** Comparativa KNN RE

### **13. Recomendaciones**

Se recomienda probar distintos métodos de reducción de dimensionalidad a los aplicados en este proyecto con el fin de lograr un rendimiento superior a los modelos sin métodos aplicados, con base en la comparativa de resultados se recomienda usar reconocedor euclídeo por encima de KNN para la clasificación de electrocardiogramas. Así mismo se recomienda probar otros métodos de clasificación como lo pueden ser redes neuronales debido a que ha mostrado menos porcentaje de error con respecto a método como KNN y arboles de decisión. (CARVAJAL & BEDOYA, 2005)

## 14. Referencias bibliográficas

CARVAJAL, J. L., & BEDOYA, J. W. (2005). *COMPARACIÓN DE MODELOS DE CLASIFICACIÓN*. Medellín: ResearchGate.

Burgos, F. J. (2003). Redes Neuronales con GNU/Linux. Obtenido de [https://www.ibiblio.org/pub/linux/docs/LuCaS/Presentaciones/200304curso-glisa/redes\\_neuronales/curso-glisa-redes\\_neuronales-html/x38.html](https://www.ibiblio.org/pub/linux/docs/LuCaS/Presentaciones/200304curso-glisa/redes_neuronales/curso-glisa-redes_neuronales-html/x38.html)

Calvo, D. (13 de Julio de 2017). Diego Calvo. Obtenido de <https://www.diegocalvo.es/clasificacion-de-redes-neuronales-artificiales/>

GeoSalud. (22 de septiembre de 2012). Obtenido de [https://www.geosalud.com/enfermedades%20cardiovasculares/arritmia/diagnostico.html#:~:text=Electrocardiograma%20\(ECG\)&text=Es%20la%20prueba%20m%C3%A1s%20com%C3%BAn,que%20pasan%20por%20el%20coraz%C3%B3n](https://www.geosalud.com/enfermedades%20cardiovasculares/arritmia/diagnostico.html#:~:text=Electrocardiograma%20(ECG)&text=Es%20la%20prueba%20m%C3%A1s%20com%C3%BAn,que%20pasan%20por%20el%20coraz%C3%B3n).



MayoClinic. (13 de septiembre de 2019). Obtenido de <https://www.mayoclinic.org/es-es/diseases-conditions/atrial-fibrillation/symptoms-causes/syc-20350624>

Bosogain Olabe, X. (n.d.). REDES NEURONALES ARTIFICIALES. [http://cvb.ehu.es/open\\_course\\_ware/castellano/tecnicas/redes\\_neuro/contenidos/pdf/libro-del-curso.pdf](http://cvb.ehu.es/open_course_ware/castellano/tecnicas/redes_neuro/contenidos/pdf/libro-del-curso.pdf)

Izaurieta, F., & Saavedra, C. (1999). Redes Neuronales Artificiales. Charlas de Fisica, 1–15. [https://doi.org/10.1016/S0210-5691\(05\)74198-X](https://doi.org/10.1016/S0210-5691(05)74198-X)

Pava-Molano, L. F., & Perafán-Bautista, P. E. (2016). Generalidades de la fibrilación auricular. *Revista Colombiana de Cardiología*, 23, 5–8. <https://doi.org/10.1016/j.rccar.2016.10.003>

Sánchez-Arias, A. G., Bobadilla-Serrano, M. E., Dimas-Altamirano, B., Gómez-Ortega, M., & González-González, G. (1981). Enfermedad cardiovascular: primera causa de morbilidad en un hospital de tercer nivel. *Proceedings - National Conference on Noise Control Engineering*, 27, 93–98.

Enfermedades cardiovasculares. (2021). Retrieved 9 March 2021, from [https://www.who.int/cardiovascular\\_diseases/es/](https://www.who.int/cardiovascular_diseases/es/)

AF Termination Challenge Database v1.0.0. (2021). Retrieved 19 May 2021, from <https://physionet.org/content/aftdb/1.0.0/>

CU Ventricular Tachyarrhythmia Database v1.0.0. (2021). Retrieved 19 May 2021, from <https://physionet.org/content/cudb/1.0.0/>

ECG-ID Database v1.0.0. (2021). Retrieved 19 May 2021, from <https://physionet.org/content/ecgidb/1.0.0/>

Intracardiac Atrial Fibrillation Database v1.0.0. (2021). Retrieved 19 May 2021, from <https://physionet.org/content/iafdb/1.0.0/>

MIT-BIH Arrhythmia Database v1.0.0. (2021). Retrieved 19 May 2021, from <https://physionet.org/content/mitdb/1.0.0/>

MIT-BIH Atrial Fibrillation Database v1.0.0. (2021). Retrieved 19 May 2021, from <https://physionet.org/content/afdb/1.0.0/>

MIT-BIH Malignant Ventricular Ectopy Database v1.0.0. (2021). Retrieved 19 May 2021, from <https://physionet.org/content/vfdb/1.0.0/>

MIT-BIH Supraventricular Arrhythmia Database v1.0.0. (2021). Retrieved 19 May 2021, from <https://physionet.org/content/svdb/1.0.0/>

Directions, A. (2007). Principal Component Analysis (PCA) Principal Component

Analysis (PCA). Statistics, June, 1–12.