

UNIVERSIDAD AUTÓNOMA DE CHIHUAHUA

FACULTAD DE INGENIERÍA

SECRETARÍA DE INVESTIGACIÓN Y POSGRADO



**ANÁLISIS DEL COMPORTAMIENTO SEGURO/AGRESIVO DEL CONDUCTOR
DESDE UNA PERSPECTIVA DE MACHINE LEARNING**

POR:

ORALDO JACINTO SIMON

**TESIS PRESENTADA COMO REQUISITO PARA OBTENER EL GRADO
DE MAESTRO EN INGENIERÍA EN COMPUTACIÓN**

CHIHUAHUA, CHIH., MÉXICO

ENERO DE 2020



Análisis del comportamiento seguro/agresivo del conductor desde una perspectiva de Machine Learning. Tesis presentada por Oraldo Jacinto Simón como requisito parcial para obtener el grado de Maestro en Ingeniería en Computación, ha sido aprobada y aceptada por:

M.I. Javier González Cantú
Director de la Facultad de Ingeniería

Dr. Alejandro Villalobos Aragón
Secretario de Investigación y Posgrado

M.S.I. Karina Rocío Requena Yáñez
Coordinador(a) Académico

Dr. Luis Carlos Gurrola González
Director(a) de Tesis

Enero de 2020

Fecha

Comité:

Dr. Luis Carlos Gurrola González
Dra. Graciela María de Jesús Ramírez Alonso
Dra. Alberto Aguilar González
Dra. Vania Carolina Álvarez Olivas

© Derechos Reservados

Oraldo Jacinto Simón
Circuito No. 1, Campus Universitario 2,
CP: 31125, Chihuahua, Chihuahua,
México.

ENERO 2020



ING. ORALDO JACINTO SIMÓN

Presente

En atención a su solicitud relativa al trabajo de tesis para obtener el grado de Maestro en Ingeniería en Computación, nos es grato transcribirle el tema aprobado por esta Dirección, propuesto y dirigido por el director **Dr. Luis Carlos Gurrola González** para que lo desarrolle como tesis, con el título: **ANÁLISIS DEL COMPORTAMIENTO SEGURO/AGRESIVO DEL CONDUCTOR DESDE UNA PERSPECTIVA DE MACHINE LEARNING.**

Índice

Dedicatoria

Agradecimientos

Índice de contenido

Índice de tablas

Índice de figuras

Índice de ilustraciones

Capítulo 1. Introducción

- 1.1 Revisión de la literatura
- 1.2 Preguntas de investigación
- 1.3 Objetivos
- 1.4 Hipótesis
- 1.5 Justificación

Capítulo 2. Marco teórico

- 2.1 Conducción segura/agresiva
- 2.2 Enfoques de evaluación de conducción
- 2.3 Sensores de teléfonos inteligentes
- 2.4 Preprocesamiento de datos de sensores
- 2.5 Medidas de evaluación
- 2.6 Conjuntos de datos
- 2.7 Métodos de aprendizaje de máquina

Capítulo 3. Metodología y técnicas de representación

- 3.1 Bag of words (bow)
- 3.2 Bow para la representación de maniobras de conducción
- 3.3 Representación de valores estadísticos mediante ventanas deslizantes
- 3.4 Modlem
- 3.5 Protocolo de evaluación
- 3.6 Conclusiones del capítulo



UNIVERSIDAD AUTÓNOMA DE
CHIHUAHUA

Capítulo 4. Evaluación de clasificadores en conjuntos de datos propuestos en la literatura

- 4.1 Selección de las pruebas estadísticas
- 4.2 Experimentación y organización de las pruebas estadísticas
- 4.3 Conclusiones del capítulo

Capítulo 5. Validación estadística de los resultados

- 5.1 Primera prueba: Prueba de friedman en ferreira et al. [10] usando su metodología
- 5.2 Segunda prueba: prueba de friedman en ferreira et al. [10] usando bow
- 5.3 Tercera prueba: prueba de friedman en m. Ricardo et al. [14] usando bow
- 5.4 Cuarta prueba: prueba de friedman en bejani y ghattee [13] usando la metodología propuesta por ferreira et al. [10]
- 5.5 Quinta prueba: prueba de friedman en bejani y ghattee [13] usando la metodología bow
- 5.6 Conclusiones del capítulo

Capítulo 6. Discusión y conclusiones generales

- 6.1 Discusión de la hipótesis
- 6.2 Trabajos futuros
- 6.3 Recomendaciones

Capítulo 7. Apéndices (anexos)

Bibliografía

Solicitamos a Usted tomar nota de que el título del trabajo se imprima en lugar visible de los ejemplares de las tesis.

ATENTAMENTE

"Naturam subiecit aliis"

EL DIRECTOR

M.I. JAVIER GONZÁLEZ CANTÚ

FACULTAD DE
INGENIERÍA
U.A.CH.



DIRECCIÓN

EL SECRETARIO DE INVESTIGACIÓN
Y POSGRADO

DR. ALEJANDRO VILLALOBOS ARAGÓN

FACULTAD DE INGENIERÍA
Circuito No.1, Campus Universitario 2
Chihuahua, Chih., México. C.P. 31125
Tel. (614) 442-95-00
www.fing.uach.mx

DEDICATORIA

A mis padres, hermanos, amigos y profesores que con su apoyo me permitieron cumplir este sueño.

AGRADECIMIENTOS

Un agradecimiento muy especial para M.S.I. Karina Rocío Requena Yáñez, coordinadora de la maestría, por brindarme la oportunidad de haber cursado este posgrado. Gracias por apoyarme durante toda mi estancia.

Un agradecimiento a la profesora M.S.I. Ana Lucia Corral Flores por sus enseñanzas y por haberme apoyado en la coordinación de mi rehabilitación.

Un sincero agradecimiento a mi tutor el Dr. Luis Carlos Gurrola González, quien se hizo responsable de mi plan de estudios y me supo guiar durante el trabajo de tesis. Gracias por enseñarme y brindarme sus conocimientos de manera incondicional.

Al Consejo de la Ciencia y Tecnología (CONACYT) de México por su apoyo económico, sin este no creo haber podido solventar mi estancia en este maravilloso país.

A la Secretaría de Investigación y Posgrado de la Facultad de Ingeniería de la Universidad Autónoma de Chihuahua, por aceptarme siendo extranjero.

RESUMEN

La conducción agresiva de automóviles es conocida como el comportamiento del conductor que intenta directa o indirectamente provocar, perjudicar o causar daño de cualquier tipo a otras personas que comparten el espacio común por el que discurre el tráfico en la vía. El análisis del comportamiento de conducción mediante el uso de tecnología móvil, ha demostrado que la seguridad de conducción se puede inferir de los datos recopilados por los sensores de los Teléfonos Inteligentes (TI). Los algoritmos de clasificación de aprendizaje de máquina sobre los datos de un TI son muy usados en la clasificación de maniobras para detectar agresividad y está demostrado que pueden detectar con precisión eventos agresivos como: frenado, aceleración, giros y cambio de carril. La eficacia de un algoritmo de clasificación de aprendizaje de máquina para identificar eventos agresivos de conducción, depende en gran medida de la calidad de los datos y el método de extracción de características para su representación. Existen varias formas de representación de los datos que en combinación con algoritmos de clasificación permiten identificar maniobras agresivas de manera eficiente. Los mejores trabajos reportados en la literatura usan diferentes formas de representación y no usan las mismas métricas para medir sus resultados. Estos resultados no son comparables y se puede apreciar que no se ha hecho suficiente para determinar la mejor solución ante este problema. El objetivo de esta investigación es evaluar el comportamiento seguro/agresivo del conductor de un vehículo a través de los datos de teléfonos inteligentes. Para ello se experimenta con los métodos de extracción y representación de características de mejores resultados reportados en la literatura: (representación en valores estadísticos, Bag of Word (BoW) y el algoritmo de reglas mínimas MODLEM). Se determina la mejor combinación de parámetros más clasificador de acuerdo con los resultados obtenidos en la métrica Area Under Curve (AUC) en 8 clasificadores propuestos. A continuación, se muestran los mejores resultados obtenidos en este trabajo de investigación:

- El algoritmo MODLEM junto al Análisis Discriminante Lineal Fisher (ADLF) obtuvo el 100% en la actividad de clasificación en un conjunto de datos de 69 eventos agresivos. Este no obtuvo buenos resultados en el resto de los experimentos con otros conjuntos de datos.
- El incremento del tamaño de las ventanas deslizantes n_f en la representación de valores estadísticos mejora la actividad de clasificación, siendo el clasificador MHLDForest el de mejores resultados con un AUC de 99.3%.
- Con el uso del ADLF sobre la representación de BoW se mejora la actividad de clasificación, siendo el clasificador MLP el de mejores resultados con un AUC de 99.9%.

Palabras clave: conducción agresiva, teléfonos inteligentes, sensores, aprendizaje de máquina



ÍNDICE DE CONTENIDO

| | |
|--|-----|
| Dedicatoria..... | vi |
| Agradecimientos | vii |
| Índice de Contenido..... | vi |
| Índice de Tablas..... | ix |
| Índice de Figuras..... | x |
| Índice de Ilustraciones | xi |
| CAPÍTULO 1: Introducción | 17 |
| 1.1. Revisión de la literatura | 19 |
| 1.2. Preguntas de investigación..... | 23 |
| 1.3. Objetivos..... | 24 |
| 1.4. Hipótesis | 24 |
| 1.5. Justificación | 24 |
| CAPÍTULO 2: Marco teórico..... | 17 |
| 2.1. Conducción segura/agresiva | 17 |
| 2.2. Enfoques de evaluación de conducción | 18 |
| 2.3. Sensores de teléfonos inteligentes | 18 |
| 2.4. Preprocesamiento de datos de sensores | 20 |
| 2.4.1. Calidad de los datos | 21 |
| 2.4.2. Visualización de datos | 22 |
| 2.4.3. Limpieza de datos | 22 |
| 2.4.4. Integración de datos | 23 |
| 2.4.5. Reducción de datos | 23 |
| 2.4.6 Transformación y discretización de datos | 23 |
| 2.4.7. Extracción de características..... | 24 |



| | |
|--|----|
| 2.5. Medidas de evaluación | 25 |
| 2.6. Conjuntos de datos..... | 26 |
| 2.7. Métodos de aprendizaje de máquina..... | 28 |
| CAPÍTULO 3: Metodología y técnicas de representación | 30 |
| 3.1. Bag of Words (BoW)..... | 31 |
| 3.1.1. Suavizado..... | 31 |
| 3.1.2. Segmentado..... | 31 |
| 3.1.3. Codeword y Codebook | 32 |
| 3.1.4. Histograma..... | 33 |
| 3.2. BoW para la representación de maniobras de conducción | 33 |
| 3.3. Representación de valores estadísticos mediante ventanas deslizantes..... | 35 |
| 3.4. MODLEM..... | 37 |
| 3.5. Protocolo de evaluación..... | 37 |
| 3.5.1. Selección de los clasificadores para la experimentación | 38 |
| 3.6. Conclusiones del capítulo | 39 |
| CAPÍTULO 4: Evaluación de clasificadores en conjuntos de datos propuestos en la literatura...40 | |
| 4.1. Selección de las pruebas estadísticas | 40 |
| 4.2. Experimentación y organización de las pruebas estadísticas | 42 |
| 4.2.1. Experimentos de la metodología de Ferreira et al. [10] en su dataset | 43 |
| 4.2.2. Experimentos de la metodología de Ferreira et al. [10] en el dataset de M. Ricardo et al. [14]..... | 48 |
| 4.2.3. Experimentos de BoW en el dataset propuesto por Ferreira et al. [10]..... | 48 |
| 4.2.4. Experimentos de BoW en el dataset propuesto por M. Ricardo et al. [14] | 54 |
| 4.2.5. Experimentos de la metodología de Ferreira et al. [10] en el dataset de Bejani y Ghatee [13] | 60 |
| 4.2.6. Experimentos de BoW en el dataset propuesto por Bejani y Ghatee [13]..... | 64 |



| | |
|---|-----|
| 4.3. Conclusiones del capítulo | 67 |
| CAPÍTULO 5: Validación estadística de los resultados..... | 68 |
| 5.1. Primera prueba: Prueba de Friedman en Ferreira et al. [10] usando su metodología.... | 69 |
| 5.2. Segunda prueba: Prueba de Friedman en Ferreira et al. [10] usando BoW..... | 72 |
| 5.3. Tercera prueba: Prueba de Friedman en M. Ricardo et al. [14] usando BoW..... | 74 |
| 5.4. Cuarta prueba: Prueba de Friedman en Bejani y Ghattee [13] usando la metodología propuesta por Ferreira et al. [10] | 77 |
| 5.5. Quinta prueba: Prueba de Friedman en Bejani y Ghattee [13] usando la metodología BoW | 79 |
| 5.6. Conclusiones del capítulo | 82 |
| CAPÍTULO 6: Discusión y conclusiones generales..... | 83 |
| 6.1. Discusión de la hipótesis | 84 |
| 6.2. Trabajos futuros | 85 |
| 6.3. Recomendaciones | 85 |
| CAPÍTULO 7: Apéndices (Anexos)..... | 86 |
| Anexo 1. Resultados de la metodología de Ferreira et al. [10] en su conjunto de datos. | 86 |
| Anexo 2. Resultado de la metodología de Ferreira et al. [10] en el dataset propuesto por M. Ricardo et al. [14]. | 99 |
| Anexo 3. Resultado de la experimentación de la metodología BoW en el dataset de Ferreira et al. [10]..... | 99 |
| Anexo 4. Resultado de la experimentación de la metodología BoW en el datataset de M. Ricardo et al. [14] | 109 |
| Anexo 5. Resultado de la experimentación con la metodología de Ferreira et al. [10] en el dataset propuesto por Bejani y Ghattee [13]..... | 117 |
| Anexo 6. Resultado de la experimentación de la metodología BoW en el datataset propuesto por Bejani y Ghattee [13]..... | 123 |
| Anexo 7. Experimentación con el algoritmo MODLEM | 131 |
| 8. BIBLIOGRAFÍA | 133 |



ÍNDICE DE TABLAS

| | |
|--|-----|
| Tabla 1. Mejores trabajos reportados en la clasificación de conducción agresiva | 20 |
| Tabla 2. Métricas propuestas para medir la conducción agresiva | 25 |
| Tabla 3. Combinaciones de pruebas estadísticas | 43 |
| Tabla 4. Rango promedio de los clasificadores al aplicar Friedman en Ferreira et al. [10] | 70 |
| Tabla 5. Rangos de signos de Wilcoxon calculados en el dataset de Ferreira et al. [10] | 71 |
| Tabla 6. Resumen de la prueba de rango de signo de Wilcoxon con simbología en [10] | 71 |
| Tabla 7. Resultados de la prueba de rango de signo de Wilcoxon en Ferreira et al. [10] | 71 |
| Tabla 8. Rango promedio de los clasificadores al aplicar Friedman en [10] usando BoW | 73 |
| Tabla 9. Resultados de la prueba de rango de signo de Wilcoxon en [10] usando BoW | 74 |
| Tabla 10. Rango promedio de los clasificadores al aplicar Friedman en [14] usando BoW | 75 |
| Tabla 11. Resultados de la prueba de rango de signo de Wilcoxon en [14] usando BoW | 76 |
| Tabla 12. Rango promedio de los clasificadores al aplicar Friedman en [36] | 77 |
| Tabla 13. Resultados de la prueba de rango de signo de Wilcoxon en [36] | 79 |
| Tabla 14. Rango promedio de los clasificadores al aplicar Friedman en [36] usando BoW | 80 |
| Tabla 15. Resultados de la prueba de rango de signo de Wilcoxon en [36] usando BoW | 81 |
| Tabla 16. Mejores resultados obtenidos en el AUC | 82 |
| Tabla 17. Parámetros evaluados para optimizar los clasificadores en Ferreira et al. [10] | 86 |
| Tabla 18. Resultados de clasificadores en el dataset de Ferreira et al. [10] con su metodología | 86 |
| Tabla 19. Resultados de clasificadores del AUC en el dataset de [10] usando su metodología | 97 |
| Tabla 20. Clasificadores de mejores resultados respecto a la métrica AUC en [10] | 98 |
| Tabla 21. Resultados de clasificadores en el dataset M. Ricardo et al. [14] | 99 |
| Tabla 22. Parámetros evaluados para optimizar los clasificadores propuestos para [10] | 99 |
| Tabla 23. Parámetros utilizados en la optimización de BoW | 100 |
| Tabla 24. Resultados de clasificadores en el dataset de Ferreira et al. [10] | 100 |
| Tabla 25. Resultados de clasificadores del AUC en el dataset de [10] usando BoW | 105 |
| Tabla 26. Resultados sobre la norma $\ X, Y\ _2$ y el eje X , al aplicar el filtro ADLF en [10] | 107 |
| Tabla 27. Resultados de clasificadores en el dataset de M. Ricardo et al. [14] usando BoW | 109 |
| Tabla 28. Resultados de clasificadores del AUC en el dataset [14] usando BoW | 114 |
| Tabla 29. Resultados sobre la norma $\ X, Y\ _2$ y el eje X , al aplicar el filtro ADLF en [14] | 115 |
| Tabla 30. Resultados de clasificadores en el dataset de Bejani y Ghatee [36] | 117 |
| Tabla 31. Resultados de clasificadores del AUC en el dataset de Ghatee [36] | 122 |
| Tabla 32. Resultados de clasificadores en el dataset de Bejani et Ghatee [36] usando BoW | 123 |
| Tabla 33. Resultados de clasificadores del AUC en el dataset de Ghatee [36] usando BoW | 127 |
| Tabla 34. Resultados la norma $\ X, Y\ _2$ y el eje X , al aplicar el filtro ADLF en Ghatee [36] | 129 |
| Tabla 35. Resultados del algoritmo Modlem en el dataset propuesto por Ferreira et al. [10] | 131 |
| Tabla 36. Resultados del algoritmo Modlem en el dataset de M. Ricardo et al. [14] | 131 |
| Tabla 37. Resultados del algoritmo Modlem en el dataset propuesto por Ghatee [36] | 132 |



ÍNDICE DE FIGURAS

| | |
|---|----|
| Figura 1. Dataset propuesto por Ferreira et al. [10]..... | 26 |
| Figura 2. Dataset propuesto por M. Ricardo et al. [14] | 27 |
| Figura 3. Dataset propuesto por Bejani & Ghatee [36] | 28 |
| Figura 4. Fase de entrenamiento de BoW [14] | 33 |
| Figura 5. Fase de codificación de BoW [14] | 34 |
| Figura 6. Ejemplo de ventana deslizante de Ferreira et al. [10] | 35 |
| Figura 7. Ejemplo de construcción del vector de características [10] | 36 |
| Figura 8. Protocolo de evaluación | 38 |
| Figura 9. Resultados de AUC en el sensor Acelerómetro, en [10]..... | 46 |
| Figura 10. Resultados de AUC en el sensor Acelerómetro-Lineal, en [10] | 46 |
| Figura 11. Resultados de AUC en los sensores Acelerómetro-Giroscopio-Combinado, en [10]..... | 47 |
| Figura 12. Resultados de AUC en el sensor Giroscopio, en [10] | 47 |
| Figura 13. Resultados de AUC en el sensor Acelerómetro, ejes (X,Y,Z), en [10]..... | 50 |
| Figura 14. Resultados de AUC en el sensor Acelerómetro, ejes (X, Y), en [10] | 51 |
| Figura 15. Resultados de AUC en el sensor Acelerómetro, ejes $\ X, Y\ _2$, en [10]..... | 51 |
| Figura 16. Resultados de AUC en el sensor Acelerómetro, eje X, en [10] | 52 |
| Figura 17. Resultados de AUC en el sensor Acelerómetro, eje Y, en [10] | 52 |
| Figura 18. Resultados de AUC en el sensor Acelerómetro, ejes $\ X, Y\ _2$, con ADLF, en [10] | 53 |
| Figura 19. Resultados de AUC en el sensor Acelerómetro, eje X, con ADLF, en [10] | 54 |
| Figura 20. Resultados de AUC en el sensor Acelerómetro, ejes (X, Y, Z), en [14]..... | 56 |
| Figura 21. Resultados de AUC en el sensor Acelerómetro, ejes (X, Y), en [14] | 57 |
| Figura 22. Resultados de AUC en el sensor Acelerómetro, ejes $\ X, Y\ _2$, en [14]..... | 57 |
| Figura 23. Resultados de AUC en el sensor Acelerómetro, eje X, en [14] | 58 |
| Figura 24. Resultados de AUC en el sensor Acelerómetro, eje Y, en [14] | 58 |
| Figura 25. Resultados de AUC en el sensor Acelerómetro, ejes $\ X, Y\ _2$, con ADLF en [14] | 59 |
| Figura 26. Resultados de AUC en el sensor Acelerómetro, eje X, en [14] | 60 |
| Figura 27. Resultados de AUC en el sensor Acelerómetro, en [36]..... | 63 |
| Figura 28. Resultados de AUC en el sensor Magnetómetro, en [36] | 63 |
| Figura 29. Resultados de AUC en el sensor Acelerómetro-Magnetómetro-Combinado, en [36]..... | 64 |
| Figura 30. Resultados de AUC en el sensor Magnetómetro, ejes $\ X, Y\ _2$, con ADLF, en [36]..... | 66 |
| Figura 31. Resultados de AUC en el sensor Magnetómetro, eje X, con ADLF en [36] | 66 |
| Figura 32. Diagrama de diferencias críticas del AUC en [10] usando su metodología..... | 70 |
| Figura 33. Diagrama de diferencias críticas del resultado del AUC en [10] usando BoW | 73 |
| Figura 34. Diagrama de diferencias críticas del AUC en el dataset de [14] usando BoW | 76 |
| Figura 35. Diagrama de diferencias críticas del AUC en [36] usando la metodología de [10]..... | 78 |
| Figura 36. Diagrama de diferencias críticas del AUC en [36] usando BoW | 81 |



ÍNDICE DE ILUSTRACIONES

| | |
|--|----|
| Ilustración 1. Seguridad vial | 17 |
| Ilustración 2. Representación gráfica de la metodología..... | 30 |



CAPÍTULO 1: INTRODUCCIÓN

El Internet de las Cosas (IC) es parte del futuro de Internet y comprenderá miles de millones de "cosas" u Objetos Conectados a Internet (OCI), donde las cosas pueden detectar, comunicarse, computar y potencialmente actuar, así como tener inteligencia, interfaces multimodales, identidades físicas/virtuales y atributos. La recopilación de datos de estos objetos es una tarea importante, ya que permite que los sistemas de software comprendan mejor el entorno. Muchos dispositivos con diferentes tipos de hardware pueden involucrarse en el proceso de recopilación y carga de datos de sensores a la nube, donde puede ocurrir un procesamiento complejo. Además, no se puede esperar que todos estos objetos estén conectados a las computadoras, debido a razones técnicas y económicas. Por lo tanto, deberíamos poder utilizar dispositivos con los recursos necesarios para recopilar datos de estos OCI. Por otro lado, recopilar y procesar los datos de los sensores embebidos en dichos OCI antes de enviarlos a la nube, ayuda a garantizar la sostenibilidad de la infraestructura de IC. Esto se debe a las limitaciones energéticas y de procesamiento [1].

Una porción significativa de estos OCI son Teléfonos Inteligentes (TI), los cuales son muy usados en nuestra sociedad. La cantidad de dispositivos móviles conectados a Internet está creciendo rápidamente. Sin embargo, se espera que miles de millones de diferentes tipos de dispositivos computacionales de recursos limitados se conecten a Internet en la próxima década. Por otro lado, la cantidad de sensores desplegados a nuestro alrededor se está incrementando. Es una tarea cada vez más importante recopilar datos de estos sensores para analizarlos y actuar sobre ellos [1]. Un TI posee una variedad de sensores y normalmente la mayoría de las personas lo llevan hacia todas partes por el simple hecho de estar localizables y poder comunicarse. Esto multiplica las posibilidades medir los datos que pueden ser capturados por dichos dispositivos.

El uso del TI como herramienta de censado y procesamiento se está haciendo muy popular en la comunidad científica. Un área muy interesante en la cual dicho dispositivo tiene una participación activa, es el análisis del comportamiento de conducción que comenzó originalmente en las ciencias sociales [2]. Recientemente, con el crecimiento de herramientas



de censado y plataformas de computación, se han identificado nuevas áreas de oportunidad [3][4], una de ellas es el análisis de conducción mediante el uso de teléfonos inteligentes (SDA, por sus siglas en inglés). El análisis de conducción ha atraído la atención de la comunidad científica, ya que se ha demostrado que la seguridad de conducción se puede inferir de los datos recopilados por los sensores de los teléfonos inteligentes [5]. Además, se podría dar retroalimentación positiva al conductor, ayudando así a mejorar la planificación y ejecución de maniobras seguras [3].

La conducción de automóviles de forma agresiva es una de las causas principales de muerte en la actualidad. Esta es definida como un comportamiento que "aumenta deliberadamente el riesgo de colisión y muchas veces es motivado por la impaciencia, la molestia, la hostilidad o un intento de ahorrar tiempo" [6]. Los estudios indican que los conductores tienden a ser relativamente más seguros cuando son monitoreados o cuando se les brinda retroalimentación sobre sus maniobras [5]. Los Sistemas de Transporte Inteligentes (STI) actuales siguen confiando en una infraestructura compuesta de sensores estáticos y cámaras instaladas en las carreteras, lo que dificulta la recopilación, generalización y análisis de datos en tiempo real. Debido al alto costo de instalación y mantenimiento, los STI a menudo están restringidos a determinados caminos o vecindarios [7]. Muchos autos de gama alta tienen estos STI instalados, pero no todas las personas pueden adquirir autos tan costosos. Esto convierte al TI como una opción viable para capturar datos para el análisis del comportamiento de conductor seguro/agresivo. Este análisis es también conocido como el problema de clasificación multiclase de eventos agresivos de conducción que es el tema central de este trabajo de investigación.

Un caso de particular interés es cuando el vehículo es censado con uno o algunos de los sensores inerciales (acelerómetro, giroscopio y magnetómetro) [3][5][8][9]. Estos estudios han revelado en gran medida que la usabilidad de las mediciones del acelerómetro, lo convierten en el sensor más importante de la inferencia de conducción basada en TI. No obstante, los investigadores no han hecho suficiente para identificar entre los trabajos existentes, las mejores formas de representar los datos de los sensores de teléfonos inteligentes para la identificación de eventos agresivos.



Los algoritmos de clasificación de aprendizaje de máquina sobre los datos captados de un TI son muy usados en la identificación de maniobras y está demostrado que pueden detectar con precisión cuatro patrones distintos (frenado, aceleración, giro a la izquierda y a la derecha) [10]. En comparación con los eventos detectados por un dispositivo de diagnóstico a bordo (OBD-II, por sus siglas en inglés), que se consideran altamente precisos, el enfoque de recopilación de datos a través de los sensores de un TI puede considerarse confiable, ya que identifica patrones de conducción específicos con alta precisión.

El presente trabajo de investigación recoge las mejores investigaciones sobre el problema de análisis-multiclase de conducción agresiva reportadas hasta el momento. Recrea las investigaciones en igualdad de condiciones para que sean comparables, las compara y expone sus resultados en un ranking de acuerdo a la métrica AUC. Propone mejoras para el procesamiento de la señal y logra los mejores resultados reportados en la comunidad científica hasta el momento.

1.1. REVISIÓN DE LA LITERATURA

En 2018 Stratis Kanarachos et al. [11] realizaron una investigación para analizar el uso de teléfonos inteligentes como plataforma integrada para monitorear el comportamiento del conductor en el contexto de Sistemas de Transporte Inteligente (STI). Ellos para validar su hipótesis, revisaron las contribuciones científicas del área y llegaron a la conclusión de que se observa una falta de coherencia entre los enfoques de las métricas utilizadas en la literatura, para comparar los resultados de las investigaciones [11]. Los autores realizan un resumen de las investigaciones que tratan sobre la clasificación agresiva de maniobras de conducción hasta 2018 y reportan los trabajos de mejores resultados según el rendimiento de sus métricas.

Entre los trabajos de mejores resultados reportados por [11], se encuentran los dos primeros de la siguiente tabla. El cuarto y quinto trabajo de la tabla fueron incluidos en esta investigación como parte de la revisión de la literatura, debido a su importancia en los resultados obtenidos en la actividad de clasificación de eventos de conducción.



Tabla 1. Mejores trabajos reportados en la clasificación de conducción agresiva

| Clasificación agresiva de conducción | Sensores | Método | Conjunto de datos | Rendimiento |
|---------------------------------------|--|--|--|--|
| Vlahogianni & Barmounakis [12] (2017) | Acelerómetro Giroscopio GPS | Conjunto mínimo de reglas de decisión (MODLEM) | No tienen público el conjunto de datos | Total Accuracy: 99.4% TPR: 88.1% FPR: 0.3% |
| Ferreira et al. [10] (2017) | Acelerómetro Giroscopio Magnetómetro | Ventanas de tiempo, valores estadísticos, RF, MLP | Creado por los autores, 69 eventos capturados en 4 viajes de 13 minutos | AUC: varían de 98% a 9.99% valores medios en ROC con diferentes sensores y tipos de maniobras |
| Bejani & Ghatec [13] (2018) | Acelerómetro Magnetómetro | Evalúan el ángulo de rotación del vehículo. Ensamble de clasificadores: (C4.5, RBFNetwork, K-NN, Naive Bayes) | Creado por los autores, 165 eventos capturados | Precision, Recall, Accuracy, F-measure varían de 95% a 100% |
| [M. Ricardo et al. [14] (2019) | Acelerómetro | Bag of Word, GNB, MLP | Creado por los autores, 478 eventos capturados por diversos conductores en diferentes días | F-measure: 97.02% G-means: 97.58% Accuracy: 96.88% |

AUC: Área bajo la curva (AUC por sus siglas en inglés)

C4.5: es un algoritmo usado para generar un árbol de decisión, es utilizado en actividades de clasificación

GNB: Clasificador de Bayes ingenuo gaussiano (GNB por sus siglas en inglés)

GPS: Sistema de Posicionamiento Global (GPS por sus siglas en inglés)

RBFNetwork: Clasificador que implementa una red de función de base radial gaussiana normalizada.

RF: Bosques aleatorios (RF por sus siglas en inglés)

ROC: Curva operativa de características del receptor

MLP: Red neuronal Multilayer Perceptron (MLP por sus siglas en inglés)

Naive Bayes: clasificador bayesiano ingenuo

K-NN: K vecinos más cercanos (K-NN por sus siglas en inglés)

TPR: Tasa de verdaderos positivos (TPR por sus siglas en inglés)

FPR: Tasa de falsos positivos (FPR por sus siglas en inglés)

F-measure: también conocido como F1-score, la media armónica de precisión y sensibilidad.

G-means: La raíz del producto de la sensibilidad de clase. Se usa cuando hay un desequilibrio en el número de ejemplos para cada clase, y analiza los resultados cuando el clasificador es mejor en ejemplos positivos sobre negativos, y viceversa (Kubat y Matwin, 1997).

Accuracy: la proporción de resultados correctos (tanto verdadero positivo como verdadero negativo) entre el total de casos examinados



Vlahogianni & Barmounakis [12], utilizan los sensores acelerómetro, giroscopio, Sistema de Posicionamiento Global (GPS, por sus siglas en inglés) del teléfono para recopilar los datos. Emplean un algoritmo llamado MODLEM, inspirado en la teoría de conjuntos aproximación para la generación de reglas mínimas en un enfoque de aprendizaje automático. Sus mejores resultados en la actividad de clasificación están en las métricas, *Exactitud Total*: 99.4%, *Tasa de Verdaderos Positivos*: 88.1%, *Tasa de Falsos Positivos*: 0.3%. La principal problemática de esta investigación es su trazabilidad debido a que los umbrales obtenidos dentro de las reglas de inducción resultantes representan la imagen general promedio del comportamiento de un conductor. Estos valores deben estimarse constantemente a medida que aumenta la muestra de conductores y las horas de conducción. Esto significa que los umbrales optimizados pueden no ser transferibles. Evidentemente, como en todos los enfoques de minería de datos, el marco metodológico es transferible y debe calibrarse cada vez que se disponga de diferentes datos o se realicen diferentes experimentos.

Ferreira et al. [10], usan los sensores acelerómetro, giroscopio y magnetómetro del teléfono inteligente para recopilar los datos. Emplean una metodología basada en ventanas deslizantes que se mueven en el tiempo para analizar la señal. En cada ventana se calculan valores estadísticos para extraer las características de la señal y construir el vector de características. Sus mejores resultados están basados en la métrica área bajo la curva (AUC, por sus siglas en inglés), dichos resultados varían de 0.98 a 0.999 valores medios en la curva conocida como característica operativa del receptor (ROC, por sus siglas en inglés) con diferentes sensores y tipos de maniobras agresivas. Los resultados en este trabajo están enfocados en buscar la mejor configuración, tamaño de la ventana deslizante, sensor y clasificador. Si miramos los resultados de esta investigación tal parece que es muy difícil mejorarlos, pero si analizamos bien a fondo, surgen preguntas como:

- ¿Existirán clasificadores mejores que Random Forest (RF) y la red neuronal Multilayer Perceptron (MLP) para esta forma de representación?
- ¿Existirán tamaños más adecuados para las ventanas deslizantes que procesan la señal?

Bejani & Ghatee [13], utilizan los sensores acelerómetro y magnetómetro para la clasificación de eventos agresivos. Estos eventos son dependientes del ángulo de rotación del



vehículo ya que la mayoría de eventos agresivos que analizan son giros. Realizan un ensamble de los clasificadores C4.5, RBFNetwork, K-NN y Naive Bayes. Combinan el clasificador C4.5 con RBFNetwork en un algoritmo híbrido donde alcanzan valores de 95% a 100% en las métricas de Precision, Recall, Accuracy y F-measure. Estos resultados se alcanzan en una validación cruzada con 5 folds en la fase de entrenamiento. Esta investigación tiene solo 3 tipos de eventos agresivos de conducción, si hubiesen empleado una variedad mayor de eventos agresivos como en Ferreira et al. [10] su resultados tendrían un mayor impacto.

M. Ricardo et al. [14], utilizan el sensor acelerómetro del TI, para recopilar los datos. Tomando como base la hipótesis de que existen mejores representaciones de datos, proponen una representación de segundo orden, basada en la estrategia Bag of Word (BoW) [15], para modelar las marcas de tiempo del acelerómetro asociadas con las maniobras de conducción agresiva. Sus mejores resultados en la actividad de clasificación están en las métricas F-measure: 0.9702, G-means: 0.9758 y Accuracy: 0.9688. En esta investigación muestran que el modelo BoW obtiene mejores resultados que las representaciones estadísticas propuestas por Ferreira et al. [10], y que las estrategias similares de segundo orden: Bag of Features (BoF) [16] y Symbolic Aggregate Approximation (SAX) [17]. Los mejores resultados que obtuvo Ferreira et al. [10] fueron con ventanas de tiempo de tamaño $nf = \{8, 7, 6\}$, sin embargo, el experimento es recreado desde nf igual 1 hasta 4. Solo la maniobra agresiva “Left turn” obtuvo entre sus mejores cinco resultados una configuración de nf igual 4. Si uno de sus objetivos era evaluar su propuesta BoW con respecto a la metodología propuesta por Ferreira et al. [6], debieron compararse con la métrica AUC en igualdad de condiciones para medir sus resultados y probar con ventanas de tamaño $nf = \{8, 7, 6, 4\}$.

Las métricas reportadas no son las mismas en ninguno de los tres trabajos, con excepción de la métrica *Exactitud* (Accuracy, traducido al inglés) que se repite en Vlahogianni and Barmounakis [12], Bejani & Ghatee [13] y M. Ricardo et al. [14]. Cada investigador usó solo su propio conjunto de datos a excepción del trabajo de M. Ricardo et al. [14], por lo que esta medida tampoco muestra trazabilidad para su comparación. Afirmar



cuál de las investigaciones obtuvo mejores resultados con diferencia significativa o no, no debe hacerse a la ligera.

Revisando las investigaciones citadas en Stratis Kanarachos et al. [11], se llega a la siguiente conclusión, la mayoría de las investigaciones usan diferentes TI que tienen diferentes sensores donde varía su calidad. Los estudios y pruebas se realizaron en varios países con diferentes condiciones de tráfico y carretera. Sus métodos para establecer *groundtruth* o verdad básica para las pruebas no son necesariamente los mismos y podían variar debido a la subjetividad. No todos los trabajos publican o ceden el conjunto de datos en los que se basaron para realizar sus experimentos. Partiendo de que las métricas que utilizan para reportar los resultados no son las mismas, se necesitan recrear los experimentos en igualdad de condiciones para que cada una de las soluciones sean comparables. Aún, no está claro que metodología es la más idónea para resolver el problema de clasificación multiclase de eventos agresivos, ni cuáles son los clasificadores candidatos para resolver el problema de clasificación multiclase de conducción agresiva. Estas metodologías y algoritmos [10][14][18] resuelven dicho problema, pero no hay consenso de cuál es la más adecuada o cuál es la que obtiene mejores resultados.

1.2. PREGUNTAS DE INVESTIGACIÓN

Teniendo en cuenta la situación problemática anteriormente expuesta se plantean las siguientes preguntas **de investigación**:

- ¿Cómo evaluar el comportamiento seguro/agresivo del conductor de un vehículo a través de los datos de teléfonos inteligentes?
- ¿Cuál de los trabajos de la literatura seleccionados obtiene mejores resultados cuando se evalúan a través de distintos conjuntos de datos?
- ¿Qué tipo de representación de los datos será más útil para obtener una mejor clasificación?



1.3. OBJETIVOS

Para darle solución al problema a resolver de este trabajo de investigación se traza el siguiente **objetivo general**: Evaluar el comportamiento seguro/agresivo del conductor de un vehículo a través de los datos de teléfonos inteligentes.

Para darle cumplimiento al objetivo general se desglosan los siguientes **objetivos específicos**:

- Definir la mejor solución entre las existentes para el problema de análisis-multiclase de conducción agresiva de un conductor.
- Describir las bases de datos públicas asociadas a este problema.
- Identificar cuáles son los sensores que han sido más exitosos para capturar datos para este problema.
- Definir cuáles han sido las transformaciones de los datos de estos sensores al incorporarlos en vectores de características con buena discriminación.
- Identificar el tipo de modelo de aprendizaje computacional más exitoso para resolver este problema.
- Determinar la mejor combinación de parámetros más clasificador en la actividad de clasificación para cada método de extracción de características.

1.4. HIPÓTESIS

La extracción de características mediante el cálculo de valores estadísticos con ventanas deslizantes de tamaño $2 < n_f < 15$ mejora los resultados de clasificación en eventos seguros/agresivos.

1.5. JUSTIFICACIÓN

La industria automotriz es un mercado de gran importancia para el desarrollo de la humanidad ya que facilita la distribución de suministros y hace más operativa la fuerza de trabajo de la actividad humana. En la actualidad la conducción es una tarea cotidiana que se



ha convertido en una necesidad para la sociedad moderna, principalmente en las grandes ciudades. Según un informe de seguridad global de la Organización Mundial de la Salud, cada año 1,25 millones de personas mueren en las carreteras del mundo; esta cifra se ha mantenido estable desde 2007 [19]. La comunidad internacional ha prestado especial atención al problema de la seguridad vial al establecer en el marco de los Objetivos de Desarrollo Sostenibles, una nueva meta a efectos de reducir a la mitad el número de muertes y traumatismos ocasionados por los accidentes de tránsito de aquí a 2020 [19]. Los algoritmos de clasificación de aprendizaje de máquina son muy usados en la clasificación de maniobras de conducción para detectar agresividad [12].

Debido al incremento de muertes por accidentes viales y aprovechando el auge de los OCI, se han realizado muchos estudios e investigaciones enfocados al diagnóstico de vehículos con TI. La evaluación y monitoreo del comportamiento de los conductores, son medidas preventivas para evitar accidentes y garantizar la seguridad vial. La extracción y el procesamiento de flujos de datos móviles facilitan el análisis de datos en tiempo real [20].

El presente trabajo de investigación recoge las mejores investigaciones sobre el problema de clasificación multiclase de conducción agresiva reportadas hasta el momento. Recrea las investigaciones en igualdad de condiciones para que sean comparables, las compara y expone sus resultados en un ranking de acuerdo a la métrica AUC. Propone mejoras para el procesamiento de la señal y busca lograr los mejores resultados reportados en la comunidad científica hasta el momento.



CAPÍTULO 2: MARCO TEÓRICO

Con el objetivo de facilitar la comprensión del marco conceptual de la investigación, en el presente capítulo se exponen los principales conceptos y enfoques de conducción, asociados al problema a resolver en este trabajo investigativo. Se describen los conjuntos de datos públicos con los que se realizará la experimentación y las medidas de evaluación para medir la actividad de clasificación de eventos seguros/agresivos.

2.1. CONDUCCIÓN SEGURA/AGRESIVA



Ilustración 1. Seguridad vial

Fuente: U. Córdoba [45]

Seguridad al conducir: Conjunto de Medios, Medidas, Normas y Actuaciones personales que tienen como fin que el viaje se desarrolle sin violar la base reglamentaria para circular en la vía [21].

Conducción segura: El cumplimiento de las normas de seguridad al conducir [21].

Conducción agresiva: Comportamiento de un conductor, con la que intenta directa o indirectamente provocar, perjudicar o causar daño de cualquier tipo a otras personas que comparten el espacio común por el que discurre el tráfico [22].

En este trabajo identificaremos la conducción agresiva a través de la clasificación de eventos agresivos sobre los datos de un TI. En el caso de la conducción segura, será a través de los eventos no agresivos, para que un conductor sea seguro deberá no realizar maniobras que reflejen agresividad. En el marco de esta investigación identificaremos los eventos agresivos clasificándolos por tipo en un enfoque supervisado. El término de conducción segura será clasificado por una única clase que se corresponde a todos los eventos que no son agresivos.



2.2. ENFOQUES DE EVALUACIÓN DE CONDUCCIÓN

Realizando un estudio de las soluciones relacionadas con este tema de investigación, podemos identificar tres enfoques al analizar el comportamiento del conductor.

Modelar el análisis de conducción: se intenta modelar el comportamiento de conducción y estimar algunos parámetros relacionados con la conducción en función del modelo producido. Como ejemplo, Álvarez et al. [23] presentaron un sistema que obtuvo algunos datos de TI, incluidos la velocidad, la aceleración y movimientos bruscos del volante, y estimó la carga restante de un vehículo eléctrico considerando el comportamiento de conducción.

Clasificación basada en el reconocimiento de patrones de conducción: Algunos investigadores tratan de clasificar los comportamientos de manejo en patrones agresivos y seguros [5]. Tenga en cuenta que una persona agresiva no es un conductor agresivo exactamente [24]. Se clasifican los estilos y maniobras de conducción de los conductores, no el carácter del conductor. Sin embargo, se necesita reconocer el conductor en muchas aplicaciones.

Clasificación asignando puntajes a los conductores: El tercer grupo de investigadores ha intentado evaluar la conducción asignando puntos a los conductores según sus maniobras [25] [26]. Estos puntajes pueden simplificarse a agresivos o seguros.

En el presente trabajo de investigación trataremos el segundo enfoque utilizando técnicas de aprendizaje de máquina. Dado al estado actual de las investigaciones reportadas, en este trabajo se considera más importante descubrir la mejor solución de las existentes y realizar propuestas de mejoras para mejorar la actividad de clasificación. Identificar con precisión los eventos agresivos es el aspecto más importante de esta investigación.

2.3. SENSORES DE TELÉFONOS INTELIGENTES

Cuando necesitamos procesar los datos de teléfonos inteligentes para la evaluación de manejo, se deben descuidar los intervalos de tiempo de llamadas, mensajes y juegos con



teléfonos inteligentes. Para seleccionar los sensores útiles para apoyar una tarea, necesitamos definir un plan para elegir los sensores necesarios y un plan para el muestreo de datos. Eftekhari y Ghatee mencionaron los siguientes criterios para elegir la mejor configuración de sensores para sus propósitos [26]:

- **Accesibilidad:** recopilar datos durante todo el proceso y soporte en cualquier sitio.
- **Cobertura:** usar los sensores en diferentes modelos de teléfonos inteligentes.
- **Precisión:** proporcionar una precisión razonable para detectar y reconocimiento.
- **Costo:** costos directos e indirectos, como la energía y cualquier pago.

Por otro lado, solo se pueden usar algunos sensores para analizar el comportamiento de conducción. Por ejemplo, el uso de una cámara para monitorear el comportamiento de manejo requiere un análisis y procesamiento de imágenes de mucho costo computacional. Debido a las limitaciones de procesamiento de los teléfonos celulares, las operaciones pesadas no funcionan. Podemos clasificar los sensores útiles para aplicaciones de STI de la siguiente manera:

- **Sistema de Posicionamiento Global:** (GPS por sus siglas en inglés), consume una gran cantidad de energía para determinar la ubicación geográfica y la velocidad. No es accesible en algunos lugares, como túneles. La precisión de este sensor es muy diferente con respecto al tipo de sensor y algunos otros problemas topológicos. La precisión de este sensor en diferentes TI es muy diferente. Por lo tanto, la aplicación de GPS para fines de seguridad es limitado. La distancia relativa entre dos TI calculada a partir del GPS varía con mucha frecuencia.
- **Radar y Lidar:** Algunos de los sensores como Lidar o Radar requieren una línea de visión entre el emisor y el receptor, por lo que su uso es limitado.
- **Bluetooth:** La red ad hoc Bluetooth suele ser adecuada para el posicionamiento en interiores. Además, Bluetooth en los TI puede comunicarse directamente con las computadoras de los vehículos para usar su información, como la apertura del airbag para detectar los choques con precisión.
- **Acelerómetro:** El acelerómetro consume poca energía y puede medir fuerzas en el teléfono celular en tres direcciones. Al usar la ley de Newton, la aceleración se puede



estimar a partir del vector de fuerza y la velocidad se puede estimar. Generalmente la precisión del acelerómetro para estimar la aceleración es aceptable, mientras que la estimación de la velocidad necesita un mayor procesamiento.

- **Giroscopio:** El giroscopio puede determinar la dirección del teléfono celular y esto es muy útil en diferentes contextos. Pero, la cobertura de este sensor es limitada. Hoy en día, el uso de este sensor se extiende y, por lo tanto, se pueden considerar los datos de este sensor en las aplicaciones de manera eficiente.
- **Magnetómetro** El magnetómetro se usa para medir el campo magnético que rodea el TI. Este sensor es útil para ajustar la dirección del TI con respecto a la dirección del movimiento.
- **Unidad de Medida Inercial (UMI):** La combinación de acelerómetro, giroscopio y magnetómetro se denomina unidad de medición inercial. Estos sensores tienen un gran papel para el seguimiento humano y de vehículos.

Por las capacidades de un TI pudiera pensarse que puede tomar el lugar de una caja negra en vehículos. Realmente, los TI no son de caja negra debido a la incertidumbre de la postura y otros factores inductores de ruido [27]. La privacidad es muy importante cuando necesitamos usar datos de TI para aplicaciones públicas. Eftekhari y Ghatee discutieron la privacidad de los sensores GPS y acelerómetro en detalles demostrando que el uso del acelerómetro tiene el menor peligro [26]. En este trabajo investigativo usaremos los sensores de la UMI para la clasificación de eventos agresivos, generando conjuntos de datos a partir de las combinaciones de sus ejes (x, y, z).

2.4. PREPROCESAMIENTO DE DATOS DE SENSORES

Los sensores de los TI producen una gran cantidad de datos por segundo, mientras que la capacidad de memoria es limitada. Esto significa que guardar los datos de los sensores representa un desafío en los teléfonos inteligentes. Además, los usuarios prefieren usar la memoria limitada para su propio uso, esto le resta posibilidades a usar dicha memoria para un sistema de evaluación de manejo. Por lo tanto, cualquier aplicación que necesite guardar los datos de los sensores debe diseñarse en condiciones de poco espacio. El procesamiento



en tiempo real es una buena respuesta a estas situaciones, lo que ahorra solo los resultados importantes, no todos los datos en la memoria [28].

Por otro lado, los usos de TI en algunas aplicaciones de transporte detectan si el TI está dentro de un vehículo o no. Por ejemplo, la detección de accidentes, el estado de la superficie de la carretera, el estado del tráfico e incluso la detección del modo motorizado (autobús, metro, automóvil, etc.) deben especificar la ubicación del TI con respecto al vehículo. Esta fase de preprocesamiento consume tiempo, procesamiento y memoria. Por lo tanto, disminuir la cantidad de datos almacenados, aumentar la precisión de la detección y minimizar el consumo de energía son cuestiones importantes [29]. Podemos usar el preprocesamiento en varios pasos:

- Aseguramiento de la calidad al verificar las medidas estadísticas necesarias, como promedio, desviación estándar, mediana, modo, diagrama de caja, etc.
- Visualización de datos: análisis de histograma, diagrama de dispersión, correlación positiva y negativa, proyección geométrica, apilamiento dimensional entre otros.
- Limpieza de datos para eliminar los efectos de ruido y completar los datos faltantes.
- Integración de datos para unificar los datos de las diferentes fuentes.
- Reducción de datos para disminuir la cantidad de características y muestras.
- Transformación de datos y discretización de datos.

2.4.1. CALIDAD DE LOS DATOS

La precisión de los sensores es muy importante, lo que indica que los datos resultantes de cualquier sensor son razonables o no. A veces, necesitamos fusionar los datos de los diferentes sensores para tomar una decisión. Para la evaluación de manejo, debemos velar por la integridad de los datos, porque si la recopilación de datos se ha interrumpido durante un período prolongado no podemos evaluar la conducción con todos los elementos necesarios. Vries y otros investigadores, desarrollaron algunas técnicas para validar y seleccionar automáticamente subconjuntos de datos para verificar la calidad de los datos recopilados de un TI [30]. Los datos utilizados en esta investigación son de fuente pública y ya han sido tratados por otros investigadores de la comunidad científica.



2.4.2. VISUALIZACIÓN DE DATOS

Es importante visualizar los datos para encontrar un enfoque o patrón para cada tipo de evento de conducción. Por lo general, al graficar los valores de los sensores, se pueden obtener buenos puntos de vista sobre lo que sucede en la realidad. A veces, la traza de datos de un TI tiene un significado importante, a través del GPS se puede recrear el trayecto de un vehículo [31]. A veces, los datos en brutos de los sensores no reflejan ningún significado, debido a su alta dimensionalidad, los datos no se pueden ilustrar. En estas situaciones, podemos transformar los datos a algunos formatos nuevos o reducir la dimensión a alguna versión visualizada.

2.4.3. LIMPIEZA DE DATOS

Antes de usar los datos del TI para sus aplicaciones, debemos limpiar la señal de ruido, valores atípicos, valores incompletos y faltantes. Entonces, la inconsistencia y los problemas intencionales deben resolverse. Para este objetivo, se pueden utilizar medidas estadísticas como modo, mediana, desviación estándar, etc., para eliminar los valores atípicos. Por ejemplo, los datos son ruidosos cuando los valores del sensor son menores que $\mu - 2\sigma$ o mayores que $\mu + 2\sigma$. En estas condiciones, μ es la media y σ es la desviación estándar. Estas técnicas estadísticas se pueden utilizar en sensores para mostrar los datos no válidos y fuera de rango [32].

Los impulsos ingresados en el vehículo se detectan principalmente mediante un sensor de magnetómetro con respecto a las coordenadas del dispositivo. Para este propósito, uno puede usar los cambios de ángulo en tres ejes. Entonces, la variable de aceleración, el tamaño angular y la velocidad angular se pueden medir cada vez. Estas variables se recopilan para los tres ejes x, y, z. Además, las medidas de covarianza y correlación pueden usarse para evaluar la relevancia entre las muestras de sensores y las etiquetas de clase [32]. Cuando el TI no está en la posición correcta dentro del vehículo es necesario reorientar la señal para analizar correctamente los ejes de la señal y no introducir ruido en su análisis. Si la señal del TI es reorientada, este se se puede colocar libremente dentro del vehículo en cualquier posición deseada [12].



2.4.4. INTEGRACIÓN DE DATOS

En esta parte, los datos de diferentes TI se integran en una base de datos única para un mayor análisis. Por ejemplo, los datos de la antena del Sistema Global para las Comunicaciones Móviles (GSM por sus siglas en inglés) para todos los residentes se pueden procesar en un solo centro para estimar la cantidad de viajeros que desean viajar entre cada par de orígenes y destinos. Cuando este esquema se repite para todas las muestras, se pueden estimar los viajeros entre todos los puntos [32]. La integración de datos, así, como el crecimiento del volumen de datos de maniobras de conducción agresiva, contribuyen a mejorar la aplicación de técnicas de aprendizaje de máquina.

2.4.5. REDUCCIÓN DE DATOS

Para reducir las características, se puede aplicar un filtro, un contenedor o un método híbrido. Una función de filtro elimina las características irrelevantes. Por ejemplo, al usar la ganancia de información entre cada característica y cada clase, podemos eliminar algunas características cuyos efectos en la clasificación son escasos. Realmente, la ganancia de información mide la correlación entre los valores y etiquetas de clase. A veces, quitando las características irrelevantes o redundantes se puede mejorar el rendimiento de la clasificación. Además, es posible ordenar las características por un filtro como la ganancia de información para mejorar el proceso de clasificación. Tal enfoque se utiliza en la exploración de árboles de decisión, ver [33] [34]. Como técnica de reducción de la dimensionalidad (reducir las características), el Análisis Discriminante Lineal de Fisher (ADLF), es utilizado en problemas de clasificación multiclase en enfoques de aprendizaje supervisado. Esta técnica puede ser entendida como encontrar un espacio que parezca contener toda la variabilidad de la clase.

2.4.6. TRANSFORMACIÓN Y DISCRETIZACIÓN DE DATOS

El proceso de transformación es muy importante, porque hay muchas transformaciones que pueden seleccionarse para aplicaciones reales. Cualquier función



matemática se puede utilizar para transformar los datos bajo planes lineales o no lineales. La discretización de datos generalmente se realiza en los datos del sensor del TI para simplificar los datos y revelar patrones en los datos [26].

Transformación con enfoques estadísticos: Cuando se estima la distribución de los datos, se pueden procesar los datos en una función de distribución en lugar de procesar los datos en bruto. Además, en muchos casos los datos se distribuyen uniformemente y estos se pueden modelar mediante una transformación normal.

Transformación en álgebra matricial: La rotación es una buena transformación que se puede utilizar en los datos del TI para ajustar su dirección con la del vehículo [32]. Además, algunas de las transformaciones se basan en la señal de preprocesamiento. Por ejemplo, la transformación de Fourier y la Discrete Wavelet Transformation (DWT) son métodos famosos para la transformación de datos [26].

2.4.7. EXTRACCIÓN DE CARACTERÍSTICAS

Para extraer las funciones de los datos del sensor del TI, se pueden definir dos tipos de funciones:

Características obvias: por lo general, se han realizado algunas operaciones en los datos sin procesar para proporcionar las características necesarias para fines de clasificación, agrupación o regresión. A veces, al obtener el promedio, la suma, el máximo, etc., los datos sin procesar se pueden transformar en algunas características útiles. Eftekhari y Ghatee definieron algunas características obvias para la detección del modo de transporte utilizando datos de giroscopio y de aceleración [32].

Características ocultas: estas características generalmente no son explicables. Por ejemplo, al obtener transformación integral de Fourier, transformación de wavelet o muestreo aleatorio, se pueden definir algunas características que son muy efectivas para la clasificación, pero no son explicables. Tales características se han estudiado con DWT en [35] y [32]. Realmente, DWT extrae las características necesarias para la evaluación del controlador de datos de TI, pero es difícil explicar por qué son importantes.



En este trabajo de investigación se experimenta con las técnicas de extracción de características más exitosas, según los trabajos de mejores resultados reportados. Entre las que se encuentran: metodología de Ferreira et al. [10], BoW propuestas por M. Ricardo et al. [14] y el algoritmo MODLEM [18].

2.5. MEDIDAS DE EVALUACIÓN

Para evaluar los resultados de las diferentes técnicas de extracción de características junto al clasificador más apto, necesitamos seleccionar las siguientes medidas:

Tabla 2. Métricas propuestas para medir la conducción agresiva

| Métrica | Descripción | Fórmula |
|----------------------------|---|---|
| TP Rate | Tasa de verdaderos positivos. (Instancias clasificadas correctamente en una clase). Es sinónimo de exhaustividad, Recall o sensibilidad. | $\frac{TP}{TP + FN}$ |
| FP Rate | Tasa de falsos positivos (Instancias clasificadas incorrectamente en una clase). | $\frac{FP}{FP + TN}$ |
| Presicion | La proporción de verdaderos positivos entre el número total de detecciones. | $Precision = \frac{TP}{TP + FP}$ |
| Recall or Sensitive | La métrica de exhaustividad muestra la cantidad de verdaderos positivos que el modelo ha clasificado en función del número total de valores positivos en una clase | $Recall = \frac{TP}{TP + FN}$ |
| F-measure | También conocido como puntaje F1, la media armónica de precisión y sensibilidad. | $F - measure = \frac{2TP}{2TP + FP + FN}$ |
| MCC | Si dos matrices de confusión no son comparables, el coeficiente de correlación de Matthew (MCC), nos ayuda a representar la matriz de confusión con un solo valor. Es una medida de la calidad de clasificación binaria que dado que tiene en cuenta los positivos y negativos (verdaderos y falsos), puede usarse incluso si las clases son de tamaños muy diferentes. | $MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$ |
| AUC | Área bajo la curva, característica de funcionamiento del receptor (ROC). Refleja el grado de separabilidad entre clases en todos los niveles de los umbrales de decisión, en términos de probabilidad. | $AUC = \frac{1}{2} \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right)$ |



| | | |
|-----------------|---|---|
| PRC Area | Es el área bajo la curva de la relación que existe entre Precisión y Recall. Las curvas de recuperación de precisión son mejores para resaltar las diferencias entre modelos para conjuntos de datos altamente desequilibrados. En este caso el área bajo la curva PRC mostrará más diferencia que ROC. | Una curva de recuperación de precisión muestra la relación entre Precisión (valor predictivo positivo) y recuperación (Recall) para cada posible corte. El PRC es un gráfico de dos dimensiones (x, y) que muestra: <i>Eje x:</i> muestra la recuperación (Recall or Sensitive) <i>Eje y:</i> muestra Precisión (valor predictivo positivo) |
|-----------------|---|---|

En el presente trabajo se calcularán cada una de estas métricas en una validación cruzada con 10 folds para evaluar la actividad de clasificación de conducción agresiva. Las métricas se calcularán para cada uno de los clasificadores utilizados en la experimentación.

2.6. CONJUNTOS DE DATOS

Cuando buscamos conjuntos de datos públicos para evaluar la conducción agresiva, nos damos cuenta de que existen muy pocos de libre acceso. No todos los trabajos investigativos publican o ceden el conjunto de datos en los que se basaron para realizar sus experimentos. A continuación, se describen los tres conjuntos de datos encontrados en esta investigación:

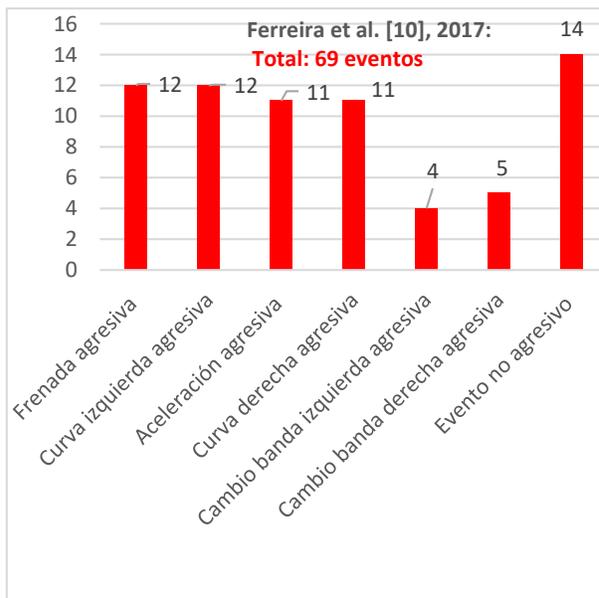


Figura 1. Dataset propuesto por Ferreira et al. [10]

Dataset de Ferreira et al. [10]: Los datos se recopilaron con un teléfono inteligente Motorola, utilizando el acelerómetro, el magnetómetro, el giroscopio y un acelerómetro lineal virtual. Los autores informan una frecuencia de muestreo entre 50 y 100 Hz, dependiendo del sensor, y de sus datos es posible determinar que emplearon una frecuencia de muestreo cercana a 50 Hz para los acelerómetros. Su experimento se realizó en cuatro viajes en automóvil de aproximadamente 13 minutos cada uno,



en el que dos conductores diferentes completaron estos viajes. El vehículo utilizado para recopilar las señales fue un Honda Civic 2011, y el teléfono inteligente se colocó en una posición fija, sin moverse ni funcionar mientras se realizaba la recopilación de datos. Los tipos de eventos de conducción y el número correspondiente de eventos recopilados son: frenado agresivo (12), aceleración agresiva (12), giro agresivo a la izquierda (11), giro agresivo a la derecha (11), cambio agresivo del carril izquierdo (4), cambio agresivo de carril derecho (5) y evento no agresivo (14). En total, hay 69 eventos de conducción clasificados en siete tipos diferentes. Aunque este conjunto de datos es pequeño con respecto a la cantidad de eventos de manejo agresivo, es uno de los más completos con respecto a los tipos de maniobras de manejo agresivo que se encuentran en la literatura.

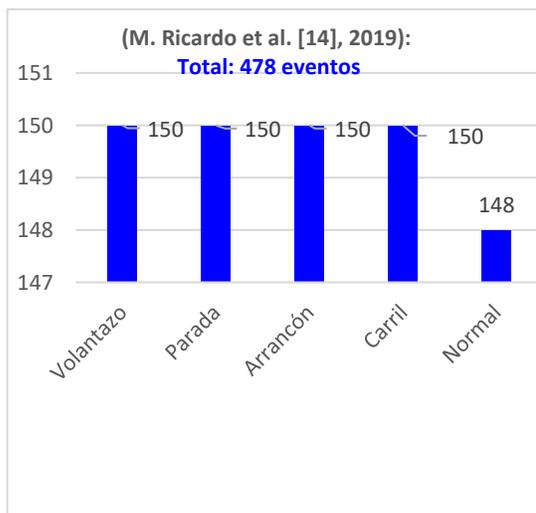


Figura 2. Dataset propuesto por M. Ricardo et al. [14]

Dataset de M. Ricardo et al. [14]:

Motivado por el trabajo de Ferreira et al. [10], presentaron un tercer conjunto de datos que está público, este se puede encontrar en este repositorio. El conjunto de datos consta de ejemplos de maniobras de conducción agresivas únicas, específicamente: desviarse a la izquierda, desviarse a la derecha, frenado repentino y aceleración repentina. Tenga en cuenta que las categorías nombradas como desviaciones podrían contener ejemplos de

cambio de carril agresivo, evitar obstáculos y esquinas. Estas maniobras fueron realizadas por sujetos entre 10 y 15 años de experiencia de conducción. En la sesión de recopilación de datos, los observadores externos registraron las etiquetas de verdad sobre el terreno para cada evento. Los datos se capturaron con los teléfonos inteligentes Motorola Moto G 2013 con sistema operativo Android versión 5.1, estos contienen un acelerómetro triaxial ST Micro LIS3DH. Los teléfonos inteligentes se colocaron libremente en el compartimiento inferior de la puerta del conductor y en el porta-vaso del vehículo. Tenga en cuenta que, dadas las maniobras de conducción que se estaban ejecutando, los teléfonos inteligentes podrían potencialmente cambiar su orientación en cualquier momento dado, esto está destinado a



replicar un escenario del mundo real. Se utilizaron dos vehículos para esta recopilación de datos: un Honda Accord y un Nissan Altima. La frecuencia de muestreo del acelerómetro se ajustó a 50 Hz, como en Ferreira et al. [10].

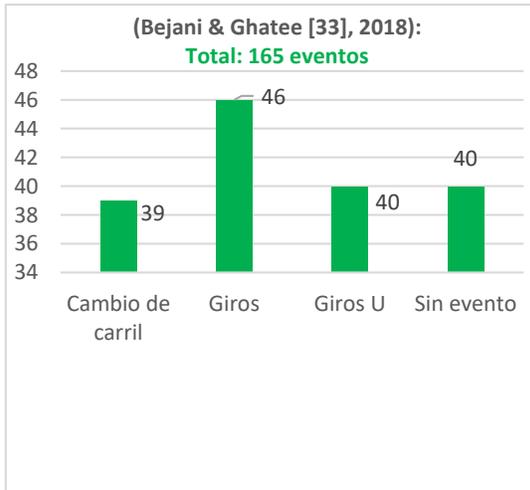


Figura 3. Dataset propuesto por Bejani & Ghatee [36]

Dataset de Bejani y Ghatee [13]:

Se considera para la recopilación de datos de los sensores una ventana rectangular con una longitud de 100 muestras, donde la frecuencia de muestreo es de 10 Hz. El conjunto de datos que se utiliza para el proceso de entrenamiento y prueba consta de 165 eventos, 40 de la clase sin eventos, 39 cambios de carril, 46 giros y 40 giros en U. Muestran que el árbol de decisión C4.5 es adecuado para la clasificación en las clases sin evento y cambio de carril. Además,

RBFNetwork es adecuado para la clasificación en las clases giro y giro en U. Por lo tanto, ambos algoritmos se combinan en un algoritmo híbrido para clasificar las maniobras. Mediante una validación cruzada con 5 folds, el resultado de las medidas de entrenamiento para el algoritmo de mejores resultados (algoritmo híbrido), muestran valores de 95% a 100% en las métricas de Precision, Recall, Accuracy y F-measure en los cuatro eventos antes mencionados.

2.7. MÉTODOS DE APRENDIZAJE DE MÁQUINA

Aprendizaje no supervisado: Cuando los datos no están etiquetados, en función de la similitud entre las muestras, se pueden dividir en algunos subconjuntos o grupos. La definición de las funciones de similitud es la base de los algoritmos de agrupamiento. Basado en la similitud, se puede definir una distancia entre las muestras y aplicando un algoritmo K-means las muestras se pueden asignar a los grupos de manera efectiva. La aplicación de clustering para analizar los datos del teléfono inteligente es limitada [32].



Aprendizaje supervisado: Cuando los datos tienen etiquetas, se implementa un método de aprendizaje supervisado para reconocer la relación entre las etiquetas y las características de las muestras. En estos problemas, se define una función de error para contar el número de muestras clasificadas erróneamente por el clasificador y la función de error se minimiza de forma iterativa. Hay muchos clasificadores como Naive Bayes, árbol de decisión, bosque aleatorio, entre otros. A veces, aplican puntos de vistas estadísticos y otros son definidos en base a los algoritmos de programación no lineal.

La aplicación de estos métodos en la evaluación de manejo es efectiva y se puede clasificar en dos partes. Algunos trabajos clasificaron el estilo de conducción en función de la calidad de las maniobras [26][13]. En estas investigaciones, los datos del sensor del TI se descompusieron en algunos intervalos que cubren las maniobras. Luego se detectó el tipo de cualquier maniobra y se evaluó el estilo de conducción en función de las similitudes entre todas las maniobras y las correspondientes maniobras seguras y peligrosas.

En la segunda categoría, el estilo de conducción se evalúa en general [36]. Esto significa que los datos del TI no se han descompuesto en los intervalos de maniobras. El clasificador asigna las etiquetas de acuerdo a las similitudes de la totalidad de las muestras de los conductores seguros o peligrosos. Ambos métodos producen las diferentes reglas para evaluar los estilos de conducción. La precisión de muchos clasificadores en datos complejos es baja. Sin embargo, podemos combinar varios clasificadores en un ensamble para aumentar la precisión de la clasificación [13].

El enfoque de mejores resultados según la literatura en esta área de investigación es el enfoque supervisado [10] [32][14]. En este trabajo no se les dejará todo el trabajo a los clasificadores, utilizaremos técnicas de extracción de características de vanguardia, para lograr mejores resultados en la clasificación de eventos agresivos.



CAPÍTULO 3: METODOLOGÍA Y TÉCNICAS DE REPRESENTACIÓN

La presente investigación presenta una metodología experimental, de tal manera que su desarrollo va orientado a experimentos que buscan la mejor combinación de parámetros sobre las técnicas de extracción de características de mejores resultados reportados. Con el uso de estas técnicas y algoritmos de clasificación se pueden identificar eventos de conducción seguros o agresivos.

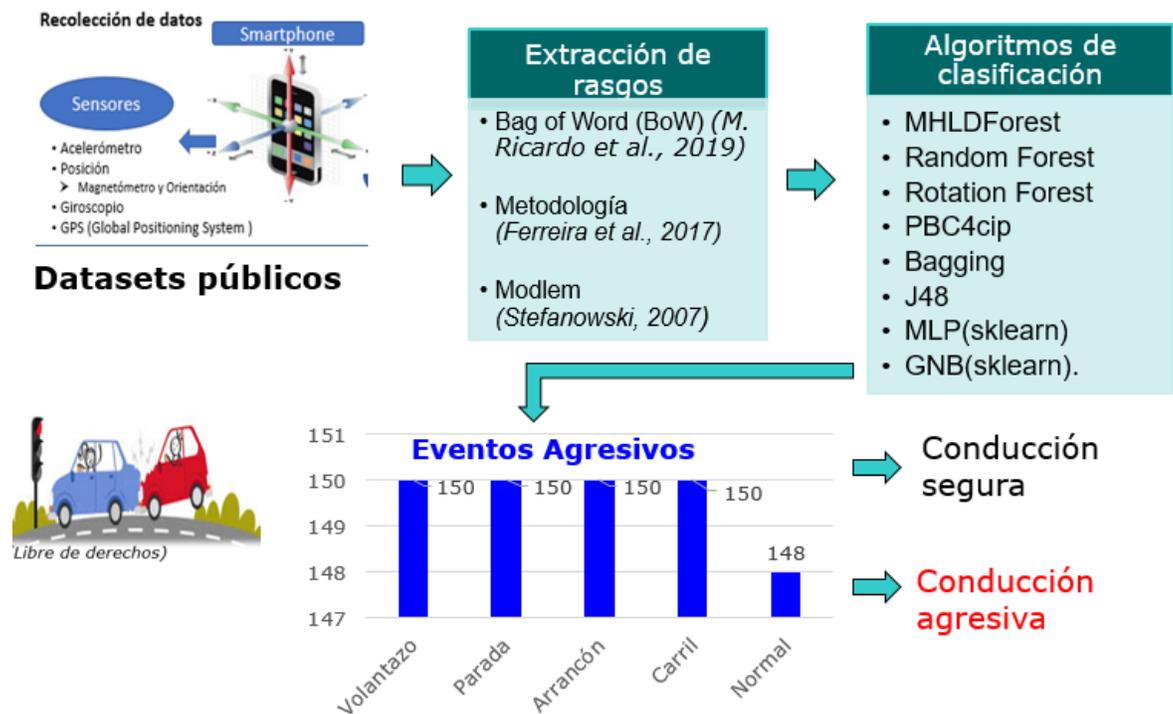


Ilustración 2. Representación gráfica de la metodología

Entre las de mejores resultados se encuentran: la representación de valores estadísticos a través de ventanas deslizantes presentado por Ferreira et al. [10], bolsa de palabras reportado por Ricardo et al. [14] y el algoritmo de reglas mínimas MODLEM [18]. El objetivo de experimentar con dichas técnicas es poder discernir la mejor técnica con su variante más exitosa para la clasificación de eventos seguros/agresivos. A continuación, se realiza una explicación de dichas técnicas y el algoritmo seleccionado.



3.1. BAG OF WORDS (BOW)

Bag of Words, en español bolsa de palabras, es una metodología muy usada en el procesamiento de textos para la clasificación de contenidos y se basa en guardar aquellas palabras más representativas. Dicha metodología facilita la clasificación de contenidos mediante el promedio de ocurrencias de palabras. La representación ignora el orden temporal de los segmentos locales dentro de una serie temporal y representa dicha serie como un histograma de palabras de código, es decir, segmentos locales. Luego se introducen varias medidas de distancia para la comparación de histogramas. En el ámbito de procesamiento de señales, el concepto de Bag of Words ha sido adaptado con el fin de extraer características de interés, desde un punto de vista distinto al que comúnmente es utilizado [37] [38] [39].

El procesamiento de señales temporales por medio de Bag of Words representa un caso de éxito para obtener de una manera simple las características deseadas con un costo de cómputo generalmente bajo. Primeramente, se debe obtener porciones de las señales o palabras que representan una clase. La metodología se puede resumir en 4 pasos: suavizado, segmentado, (codewords, codebook) y por último la creación de histogramas [37]. A continuación, se realiza breve descripción de que se hace en cada paso.

3.1.1.SUAVIZADO

El uso de un método de suavizado es una opción factible ya que hace posible eliminar las variaciones de la señal a causa de baches, bordos, la vibración del motor o cualquier otro elemento que haga interferencia en la recepción de la señal captada a través de los sensores del Teléfono inteligente (TI). Los filtros de media móvil o *Moving Average* se usan para suavizar las señales adquiridas. Al suavizar una señal se están eliminando los componentes de alta frecuencia de la misma.

3.1.2.SEGMENTADO

Una vez que la señal ha sido filtrada, cada porción de la señal con la que se formará la bolsa de palabras es segmentada en partes que puedan representar la forma que adquiere



una señal de cierta clase mediante la siguiente ecuación. Para cada señal S_i de una clase C_j , donde m es el total de clases, se define una ventana deslizante de tamaño r que recorre los diferentes puntos p , donde t es el tamaño de la señal con un barrido de elementos b para ser extraídos y procesados. X es el resultado de la segmentación de la señal y puede interpretarse como un listado que contiene los segmentos que representan cada clase. Cuando el número de elementos de la señal no empata con el largo de la ventana, los elementos de cola no son tomados en cuenta [37].

$$X_{i,j,p} = \sum_{p=0}^t \sum_{j=1}^m S_i C_j (r_p + b) \quad (1)$$

Con el objetivo de lograr una mejor representación de la señal, a cada segmento se le agregan algunos valores de representación extra a los puntos de la señal entre los que se encuentran: punto más alto, punto menos alto, promedio y desviación estándar. Las mejoras en el desempeño de estos valores serán probadas y evaluadas en los experimentos del presente trabajo.

3.1.3. CODEWORD Y CODEBOOK

Las señales o *codewords* que proveen las características principales de una clase, pueden definirse por todas aquellas señales con las cuales la metodología es entrenada, por lo tanto, una vez que se obtienen los segmentos de las señales, estas son sometidas al procesamiento mediante clustering; en este caso se hará uso del algoritmo k-means para obtener los promedios de señal con la siguiente ecuación, donde para cada punto p de los segmentos de la clase j se obtiene la media del número de *codewords* k .

Una vez que se han obtenido los *codewords* de las distintas clases a distinguir, es necesario crear un conjunto de todas estas palabras para conformar el llamado *codebook*, cuyo propósito es servir de referencia para crear los histogramas de las señales de prueba.

$$CB = \sum_{j=1}^m c_{k,j} \quad (2)$$



3.1.4. HISTOGRAMA

Para obtener los histogramas, las señales de prueba deben ser procesadas para obtener su representación en Bag of Words, por lo que, al igual que al inicio del proceso de la misma, cada una de las señales de prueba debe ser segmentada como se explica en 3.1.2. ver ecuación (1).

Una vez que las señales de prueba se segmentan, cada uno de ellos es comparado con cada uno de los *codewords* del *codebook*. De esta manera el segmento es asignado como aquel al cual más se parezca en el libro, el cual está conformado por las palabras de todas las clases, por lo que cabe la posibilidad de contener segmentos que sean más parecidos a palabras de una clase a la cual no pertenecen. Otro aspecto a resaltar de esta metodología es que su aplicación a señales de acelerómetro elimina por completo el concepto de temporalidad [37].

3.2. BOW PARA LA REPRESENTACIÓN DE MANIOBRAS DE CONDUCCIÓN

Recientemente esta metodología fue utilizada por M. Ricardo et al. [14] para la clasificación de maniobras agresivas a través del acelerómetro en un enfoque de aprendizaje de máquina supervisado. A continuación, se describen las etapas de entrenamiento y codificación.

Algoritmo 1: Etapa de entrenamiento

Entrada: S: lista de señales de entrenamiento de las mismas clases
K: cantidad de *codewords* a encontrar
L: longitud de la ventana deslizante
O: cantidad de marcas de tiempo que se superponen entre ventanas consecutivas

Resultado: K: *codewords* para representar la clase
W \leftarrow { } // Segmentos de la señal original
foreach s in S **do**
| W \leftarrow W + DivideEnVentanas(s, L, O)
end
codewords \leftarrow K – Media(W, K)
return K *codewords*

Figura 4. Etapa de entrenamiento de BoW [14]



Entrenamiento: El resultado de la etapa de entrenamiento es encontrar K secuencias que se asemejen a todas las sub-secuencias de las lecturas de aceleración del mismo grupo. Encontrar un vocabulario para representar uniformemente las series de tiempo en el problema de aprendizaje. El número de K segmentos a encontrar se llaman *codewords*. Al buscar estos *codewords*, se emplea una ventana deslizante de longitud L para atravesar cada señal de entrenamiento, y que existe una superposición de marcas de tiempo O entre ventanas consecutivas. Para una señal de acelerómetro dada s , la función *DivideEnVentanas* (s , L , O) extrae todos los segmentos de tamaño L que se obtienen de S con una ventana deslizante con superposición. Al final, con la lista W que contiene todos los segmentos de señal (de todas las señales en S), se aplica un algoritmo de agrupamiento para encontrar los K centroides (*codewords*). El conjunto de todas las palabras que representan los códigos se denomina *codebook*. El *codebook* puede verse como bloques de construcción del vector de características final [14]. Esta representación de BoW es la que se utilizará en el siguiente trabajo de investigación.

Algoritmo 2: Etapa de Codificación

Entrada: s : una muestra de la señal del acelerómetro
 L : longitud de la ventana deslizante
 O : cantidad de marcas de tiempo que se superponen entre ventanas consecutivas
 B : el *codebook*

Resultado: A : vector de característica de la señal s
 $F \leftarrow$ inicializar vector de tamaño $|B|$ // Salida del vector de características
 $W \leftarrow \{ \}$ // Segmentos extraídos desde s
 $W \leftarrow$ *DivideEnVentanas*(s , L , O)
foreach w in W **do**
 $i \leftarrow$ índice en B del *codeword* más cercano a w
 $F[i] \leftarrow F[i]++$ // Incrementar el recuento de *codeword* i
end
return F

Figura 5. Etapa de codificación de BoW [14]

Codificación: construye el vector de características de una muestra de señal de acelerómetro. Primero, extrae todos los segmentos de tamaño L de la misma manera que en la etapa de entrenamiento. Para cada uno de estos segmentos, encuentra el *codeword* más cercana dentro del *codebook* utilizando una distancia euclidiana en un espacio L -dimensional. Una vez que se encuentra el *codeword* más cercano, su recuento aumenta en 1.



El vector de características es un histograma donde cada compartimiento representa la frecuencia de cada *codeword* que se puede encontrar para la señal *s*. La racionalización detrás de esta representación es que una señal de muestra perteneciente a una clase *C_i* compartirá la distribución de palabras de código de las otras señales en la misma clase [14].

En este trabajo se utiliza la implementación hecha por los autores de este algoritmo. Se implementa, además, una interfaz para la generación de conjuntos de datos con diferentes configuraciones *K, L*.

3.3. REPRESENTACIÓN DE VALORES ESTADÍSTICOS MEDIANTE VENTANAS DESLIZANTES

Esta representación es tomada del trabajo original de Ferreira et al. [10] y a continuación se explica su funcionamiento básico. La señal se agrupa en marcos de *n* segundos de longitud para componer una ventana de tiempo deslizante. A medida que pasa el tiempo, la ventana se desliza en incrementos de 1 cuadro sobre la serie temporal hasta que recorra toda la señal [10]. En la Figura 6 se puede ver un ejemplo de cómo se mueve la ventana deslizante por la señal del teléfono inteligente.

f_0 : marco del segundo actual

f_{-1} : marco del segundo anterior y así sucesivamente $f_{-(nf-1)}$

nf : número de cuadros de la ventana deslizante

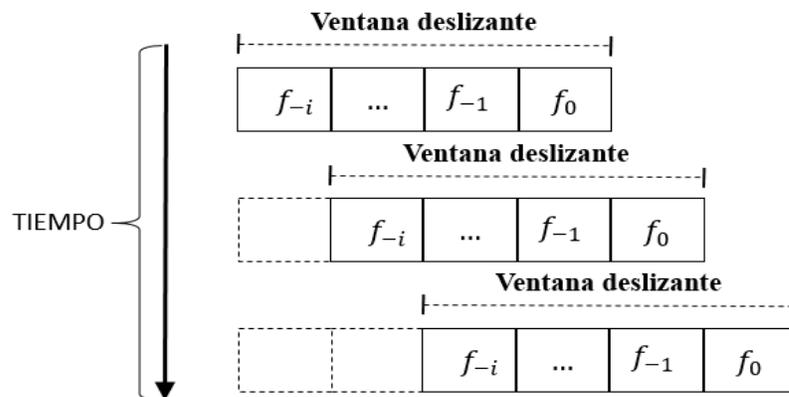


Figura 6. Ejemplo de ventana deslizante de Ferreira et al. [10]



Se define $nf = \{4, 5, 6, 7 \text{ y } 8\}$ donde cada número valores experimentalmente para que la ventana deslizante pueda acomodarse a la duración de los eventos de conducción recopilados que varían de 2 a 7 segundos.

Se genera una instancia del vector de atributos por cada ventana de tiempo que contiene un evento de conducción. El número de atributos del vector depende de la cantidad de cuadros de la ventana deslizante, lo que es equivalente al tamaño nf .

| Media | | | Mediana | | | Dev. Std | | | Tendencia | | | Etiqueta |
|-------|-----|-------|----------------------------------|-----|--------|----------|----------------------------------|--------|-----------|-----|-------|----------------------|
| M_0 | ... | M_i | MD_0 | ... | MD_i | DS_0 | ... | DS_i | T_0 | ... | T_i | Evento de Conducción |
| | | | $M_0 = M(f_0)$ | | | ... | $M_i = M(f_{-i}, f_0)$ | | | | | |
| | | | $MD_0 = MD(f_0)$ | | | ... | $MD_i = MD(f_{-i}, f_0)$ | | | | | |
| | | | $SD_0 = SD(f_0)$ | | | ... | $SD_i = SD(f_{-i}, f_0)$ | | | | | |
| | | | $T_1 = \frac{M(f_{-1})}{M(f_0)}$ | | | ... | $T_i = \frac{M(f_{-i})}{M(f_0)}$ | | | | | |

Figura 7. Ejemplo de construcción del vector de características [10]

Leyenda: (M) media, (MD) mediana, (DS) desviación estándar, (T) tendencia de aumento / disminución, esta última se calcula dividiendo la media del marco del segundo anterior representado por f_{-1} con la media del marco del segundo actual representada por f_0 .

La etiqueta o clase de cada vector de características provienen de los eventos de conducción y son tomados del registro de verdad básica o en inglés conocido como *groundtruth*. En este trabajo se implementa dicho algoritmo y se generan diferentes conjuntos de datos, a partir de los conjuntos de datos propuestos en la subsección 2.6 para comparar su efectividad.



3.4. MODLEM

Se basa en el esquema de una cobertura secuencial y produce heurísticamente un conjunto mínimo de reglas de decisión para cada concepto de decisión (por ejemplo, clase de decisión). El conjunto de reglas producido tiene como objetivo abordar todos (o los más significativos) ejemplos positivos de un concepto dado. En el procedimiento de inducción de reglas, la primera regla se construye eligiendo secuencialmente las condiciones elementales "mejores" según algunos criterios elegidos [18].

La primera condición candidata se elige formando una condición elemental y, si no cumple con el requisito de ser aceptado como una regla, entonces la segunda mejor condición elemental se agrega a la parte de la condición candidata, y así sucesivamente. Este procedimiento se repite hasta que se pueda aceptar una regla. Cuando se formula la regla, todos los ejemplos positivos de aprendizaje descritos por esta regla se eliminan de la muestra y el procedimiento de formación de reglas se repite para los datos de muestra restantes (ejemplos que permanecen descubiertos por el conjunto de reglas). Luego, el procedimiento se repite secuencialmente para cada conjunto de ejemplos de un concepto de decisión posterior. Las condiciones elementales se evalúan utilizando la entropía de clase [18].

Para identificar las potencialidades de MODLEM frente al problema de clasificación multiclase de maniobras agresivas de conducción, se obtuvo la versión original de dicho algoritmo de la plataforma Weka. Dicho algoritmo se adecuó a las necesidades de este trabajo de investigación para experimentar en los conjuntos de datos expuestos en la subsección 2.6.

3.5. PROTOCOLO DE EVALUACIÓN

En la siguiente figura se presenta el protocolo de evaluación propuesto en este trabajo de investigación:

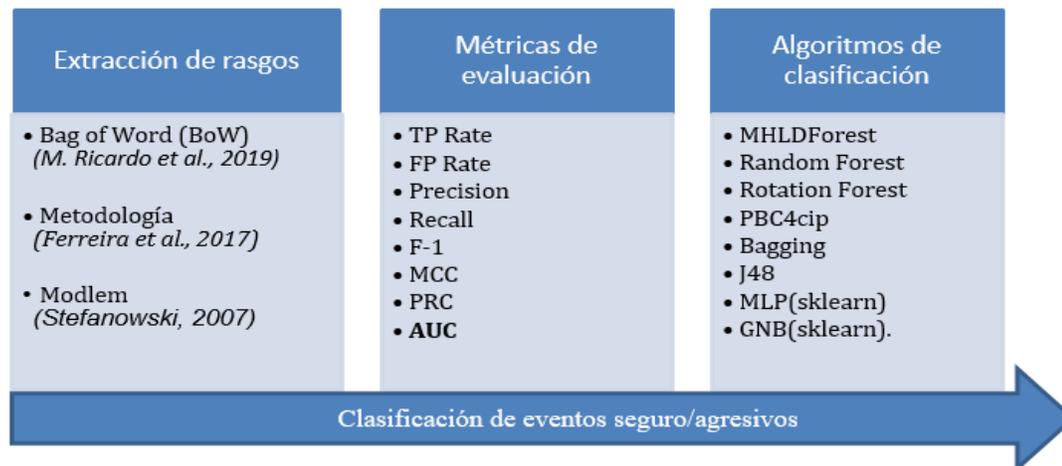


Figura 8. Protocolo de evaluación

Con el objetivo de identificar la técnica de extracción de características más idónea para la extracción y representación de los datos de un TI, se experimenta con cada una de ellas en los conjuntos de datos presentados en la subsección 2.6. Se busca la mejor combinación de parámetros más clasificador de acuerdo a los resultados de la métrica AUC en los métodos de extracción de características. Hay que resaltar que se calculan todas las métricas en una validación cruzada de 10 folds para cada uno de los clasificadores, en cada una de las técnicas de extracción de características. Para encontrar el clasificador más idóneo en cada técnica de extracción se experimenta con los 8 algoritmos de clasificación propuestos.

3.5.1. SELECCIÓN DE LOS CLASIFICADORES PARA LA EXPERIMENTACIÓN

Según los resultados de Ferreira et al. [10], Random Forest y MLP son los clasificadores que obtienen mejores resultados en su trabajo de investigación para la clasificación de eventos agresivos. De acuerdo con los resultados de M. Ricardo et al. [14], los mejores clasificadores son MLP(sklearn) y GNB(sklearn) en su propio conjunto de datos. En 2019 Cañete et al. [40], demostraron en 59 bases de datos que el clasificador PBC4cip multivariado y MHLDTForest superan a Random Forest en la actividad de clasificación.

PBC4cip (es un clasificador multivariado basado en patrones de contraste) [41].

MHLDTForest: es una variante de Random Forest que usa divisiones multivariadas en lugar de univariadas. Este clasificador fue introducido por Cañete et al. [40].



Ambos clasificadores se encuentran disponibles para Weka en la página¹. Partiendo de que Random Forest es el clasificador de mejores resultados para Ferreira et al. [10] y en la actualidad ya este tiene competidores, PBC4cip y MHLDTForest, serán utilizados en la actividad de experimentación del presente trabajo. Es importante preparar las condiciones adecuadas, en cuanto a recursos de hardware para correr los experimentos, teniendo en cuenta que estos clasificadores consumen muchos más recursos que Random Forest. Tomando el trabajo de Cañete et al. [40] como referencia, también se incluyen los clasificadores Rotation Forest, Bagging y J48 como parte de la etapa de experimentación del presente trabajo, estos están disponibles en los repositorios de Weka. En resumen, los algoritmos de clasificación seleccionados son: MHLDTForest, Random Forest, PBC4cip, Bagging, Rotation Forest, J48, MLP(sklearn) y GNB(sklearn).

Para realizar los experimentos con estos clasificadores es necesario invocar Weka en un ordenador con suficiente memoria de acceso aleatorio (RAM, por sus siglas en inglés), en este trabajo se reservaron 961600 megas.

3.6. CONCLUSIONES DEL CAPÍTULO

En este capítulo se describe el funcionamiento de las técnicas de extracción de características más exitosas en la actividad de clasificación de eventos seguros/agresivos. Estas se analizarán mediante un enfoque de aprendizaje de máquina supervisado. También se explica el protocolo de evaluación a seguir para encontrar la mejor variante de técnica más clasificador. Es importante realizar estas evaluaciones recreando los experimentos en igualdad de condiciones y parámetros.

¹ <https://sites.google.com/view/leocanetesifuentes/software/multivariate-pbc4cip?authuser=0>



CAPÍTULO 4: EVALUACIÓN DE CLASIFICADORES EN CONJUNTOS DE DATOS PROPUESTOS EN LA LITERATURA

Siguiendo el protocolo de evaluación de la subsección 3.5, se experimentó con los tres conjuntos de datos expuestos en la subsección 2.6. Utilizando el algoritmo MODLEM se obtuvo buenos resultados para el conjunto de datos propuesto por Ferreira et al. [10] en 69 eventos, no siendo así para el conjunto de datos propuesto por M. Ricardo et al. [14] en 478 eventos donde no superó un AUC del 91%. Este algoritmo fue descartado por no obtener buenos resultados en ambos datasets. Para más detalles ver el Anexo 7.

Al experimentar con las metodologías de Ferreira et al. [10] y BoW de los autores M. Ricardo et al. [14], los resultados superan un AUC del 98%. Con el objetivo de poder diferenciar cuál clasificador obtuvo mejores resultados en cada una de las metodologías y cuál de estas obtuvo mejores resultados es necesario realizar pruebas estadísticas. A continuación, se explica la selección y organización de dichas pruebas.

4.1. SELECCIÓN DE LAS PRUEBAS ESTADÍSTICAS

Los investigadores del área de Aprendizaje Automático (AA) se han vuelto cada vez más conscientes de la necesidad de la validación estadística de los resultados que se obtienen cuando se analizan grandes volúmenes de datos. Un artículo de AA típico, puede plantear la hipótesis implícita de que una mejora produce un rendimiento mejorado con respecto al algoritmo(s) existente(s). Se seleccionan una serie de conjuntos de datos para la prueba, se ejecutan los algoritmos y se evalúa la calidad de los modelos resultantes utilizando una medida adecuada, con mayor frecuencia la precisión de clasificación usando el AUC. Varios investigadores han abordado el problema de comparar dos clasificadores en un solo conjunto de datos usando varias soluciones.

Formalmente, supongamos que hemos probado k algoritmos de aprendizaje en N conjuntos de datos. Sea c_j^i la puntuación de rendimiento del algoritmo j en el conjunto de datos i . La tarea es decidir en función de los valores c_j^i , si los algoritmos de clasificación son estadísticamente significativos. Es importante verificar que los resultados medidos son



"confiables"; para ello, se requiere que se hayan realizado suficientes experimentos en cada conjunto de datos y preferiblemente que todos los algoritmos se evaluaron utilizando las mismas muestras aleatorias [42].

Si bien los procedimientos para la comparación de un par de clasificadores en un solo problema se han propuesto hace bastante tiempo, algunos estudios comparativos con más clasificadores y/o más conjuntos de datos todavía emplean soluciones parciales e insatisfactorias. Algunos trabajos no describen explícitamente los métodos de muestreo y prueba utilizados.

En la discusión de las pruebas para las comparaciones de dos clasificadores en múltiples conjuntos de datos, es importante resaltar que la prueba t-student ampliamente utilizada, no es bien utilizada y generalmente es conceptualmente inadecuada y estadísticamente insegura.

En el contexto de AA, la prueba t-student tiene tres puntos débiles:

- El primero es que esta prueba solo tiene sentido cuando las diferencias entre los conjuntos de datos son conmensuradas. En esta vista, usar la prueba t-student pareada para comparar un par de clasificadores tiene tan poco sentido como calcular los promedios sobre conjuntos de datos.
- El segundo problema con esta prueba es que, a menos que el tamaño de la muestra sea lo suficientemente grande (30 conjuntos de datos), la prueba t-student pareada requiere que las diferencias entre las dos variables aleatorias comparadas se distribuyan normalmente. La naturaleza de nuestros problemas no proporciona ninguna provisión para la normalidad y el número de conjuntos de datos suele ser mucho menor que 30.
- El tercer problema es que al igual que el promedio de los conjuntos de datos, esta prueba se ve afectada por valores atípicos que sesgan la estadística de la prueba y disminuyen la potencia de la prueba al aumentar el error estándar estimado.

La prueba de rangos con signo de Wilcoxon [43] es una alternativa no paramétrica a la prueba t-student pareada, que clasifica las diferencias en el rendimiento de dos algoritmos



para cada conjunto de datos, ignorando los signos, y compara los rangos de los positivos y las diferencias negativas. En muchos trabajos de aprendizaje automático terminan extrayendo conclusiones de una matriz y utilizan McNemar para comparar todos los pares de clasificadores, como si las pruebas para comparaciones múltiples, como ANOVA y la prueba de Friedman no existieran [42]. Debido a las ventajas teóricas y prácticas de la prueba de Friedman (facilidad de cálculo e interpretación, la capacidad de presentar el rendimiento general de clasificadores en forma de rangos en lugar de dudosos promedios), la prueba de Friedman debe preferirse a ANOVA [42].

En este trabajo se utilizará la prueba de Friedman para determinar que clasificador obtiene mejores resultados para cada conjunto de datos por separado, obteniendo el algoritmo de mejor rendimiento de la métrica AUC en el rango de 1, el segundo mejor rango 2 y así sucesivamente. Si el test de Friedman es significativo implica que al menos dos grupos de entre los comparados son significativamente diferentes, pero esta prueba no indica cuales. Para saberlo es necesario comparar todos los clasificadores entre sí. Una vez que se ha determinado que existen diferencias entre las medias, las pruebas de rango post hoc permiten determinar qué medias difieren, estas identifican subconjuntos homogéneos de medias que no se diferencian entre sí. En este caso utilizaremos como método de post hoc la prueba de Nemenyi, esta es similar a la prueba de Tukey para ANOVA y se usa cuando todos los clasificadores se comparan entre sí. El rendimiento de dos clasificadores es significativamente diferente si el promedio correspondiente de los rangos difieren al menos en la diferencia crítica [42].

4.2. EXPERIMENTACIÓN Y ORGANIZACIÓN DE LAS PRUEBAS ESTADÍSTICAS

Para explicar la generación de los conjuntos de datos utilizados en la experimentación y la selección de estos para realizar las pruebas estadísticas se presentan 5 subsecciones de la 4.2.1 a la 4.2.6, exceptuando la 4.2.2. En la subsección 4.2.2 se explica porque no se pudo aplicar la metodología de Ferreira et al [10] en el dataset propuesto por M. Ricardo et al [14]. En cada subsección se responden las siguientes preguntas:

- ¿Cómo se extraen las características del conjunto de datos propuesto?



- ¿Qué conjuntos de datos se generan?
- ¿Dónde se encuentran los resultados de la experimentación?
- ¿Qué datos se seleccionan para realizar la prueba estadística “x”?

El objetivo de las 5 pruebas estadísticas es establecer un ranking entre los clasificadores usados en la experimentación teniendo en cuenta la metodología seleccionada y determinar si existe diferencia estadísticamente significativa entre ellos, con un nivel de significancia de 0.5. A continuación se muestra la Tabla 3, donde se recoge la relación de las subsecciones de selección de datos y las pruebas estadísticas propuestas en los tres conjuntos de datos seleccionados:

Tabla 3. Combinaciones de pruebas estadísticas

| Prueba estadística | Metodología | Subsección de selección de datos | Dataset propuesto |
|----------------------|----------------------|----------------------------------|------------------------|
| Primera prueba (5.1) | Ferreira et al. [10] | 4.2.1 | Ferreira et al. [10] |
| Segunda prueba (5.2) | BoW[14] | 4.2.3 | Ferreira et al. [10] |
| Tercera prueba (5.3) | BoW[14] | 4.2.4 | M. Ricardo et al. [14] |
| Cuarta prueba (5.4) | Ferreira et al. [10] | 4.2.5 | Bejani y Ghattee [13] |
| Quinta prueba (5.5) | BoW[14] | 4.2.6 | Bejani y Ghattee [13] |

4.2.1. EXPERIMENTOS DE LA METODOLOGÍA DE FERREIRA ET AL. [10] EN SU DATASET

En esta subsección se explica la extracción de características y generación de conjuntos de datos a partir del dataset propuesto en [10]. Se muestra además una representación de los principales resultados en función de la métrica AUC (Weighted Avg.) por cada una de las variantes de los sensores con los que se experimenta.

¿Cómo se extraen las características?

Las características del dataset propuesto por Ferreira et al. [10] se extraen usando su metodología basada en valores estadísticos. Para determinar el tamaño de ventana más adecuado, en el trabajo original de Ferreira et al. [10] se experimenta con ventanas deslizantes de tamaño $nf = \{4, 5, 6, 7, 8\}$, donde nf representa la cantidad de marcos de 1 segundo de señal formando la ventana deslizante. En el trabajo de Ferreira et al. [10], los mejores resultados se obtuvieron con las ventanas de tamaño $nf = \{4, 6, 7, 8\}$. ¿Existirán tamaños nf



más adecuados para las ventanas deslizantes que procesan la señal, que los utilizados en la investigación de Ferreira et al. [10]?

Para responder esta pregunta se experimenta con diferentes tamaños nf de ventanas deslizantes. El rango de tiempo entre eventos agresivos de diferente clase oscila de 17.5 a 172.8 segundos. Específicamente, en este dataset pueden considerarse ventanas deslizantes de tamaño $17.5 > nf > 8$, sin que la ventana contenga dos eventos agresivos al mismo tiempo. Respetando este rango se decidió probar con los siguientes tamaños $nf = \{2, 4, 6, 7, 8, 9, 10, 11, 12, 13, 15\}$, de ventanas deslizantes. Tenga en cuenta que la clase de evento no agresivo, no es considerada como un evento agresivo para el cálculo del rango de tiempo entre eventos agresivos.

¿Qué conjuntos de datos se generan?

Siguiendo la metodología propuesta por Ferreira et al. [10] se generó un dataset por cada tamaño de ventana deslizante $nf = \{2, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15\}$, para cada una de las 4 variantes de sensores utilizadas. Las variantes de sensores utilizadas fueron: acelerómetro, acelerómetro lineal, giroscopio y una combinación por medio de concatenación de los sensores acelerómetro y giroscopio. En total se generaron $12 * 4 = 48$ datasets, donde 12 es la cantidad de tamaños de ventanas deslizantes nf y 4 las variantes de sensores. Se usaron todos los valores de los ejes x, y, z de cada uno de los sensores propuestos.

¿Dónde se encuentran los resultados de la experimentación?

En el Anexo 1, en la Tabla 18 están todos los resultados de los 8 clasificadores con los que se experimentó en los 48 datasets generados a partir del dataset de Ferreira et al. [10], utilizando su propia metodología. A cada clasificador se le realizó una validación cruzada con 10 folds.

8 clasificadores utilizados en la experimentación del dataset de Ferreira et al. [10]: MHLDForest, Random Forest, Rotation Forest, PBC4cip, Bagging, J48, MLP y GNB(sklearn).



¿Qué datos se seleccionan para realizar la prueba estadística 5.1?

Para evaluar los clasificadores en cada uno de los 48 datasets generados se seleccionó la métrica AUC (Weighted Avg.). Se construyó una matriz de 9 columnas X 48 filas. La primera columna se corresponde al identificador del dataset y el resto de las columnas son los 8 clasificadores utilizados en la experimentación. La intersección del identificador del dataset con cada uno de los 8 clasificadores corresponde al valor de AUC obtenido en ese dataset para ese clasificador. En la Tabla 18 del Anexo 1, se encuentran los valores de AUC utilizados para construir la matriz, los valores de la tabla 18 se extrajeron de la tabla 17 donde están los resultados de todas las métricas calculadas en los 48 datasets generados para los 8 clasificadores.

Representación de los resultados agrupados por sensor

Con el objetivo de mostrar los resultados que se obtuvieron en la métrica AUC (Weighted Avg.) en la intersección, clasificador, dataset, se muestra una gráfica para cada una de las 4 variantes de sensores utilizados. Esto permite analizar el comportamiento de la actividad de clasificación de eventos agresivos, a través de la métrica AUC, a medida que aumenta el tamaño de la ventana deslizante aumenta el valor de AUC en cada uno de los sensores. Es importante resaltar que cada tamaño de ventana deslizante, en cada uno de los sensores, se corresponde a un dataset generado a partir de ese valor nf .

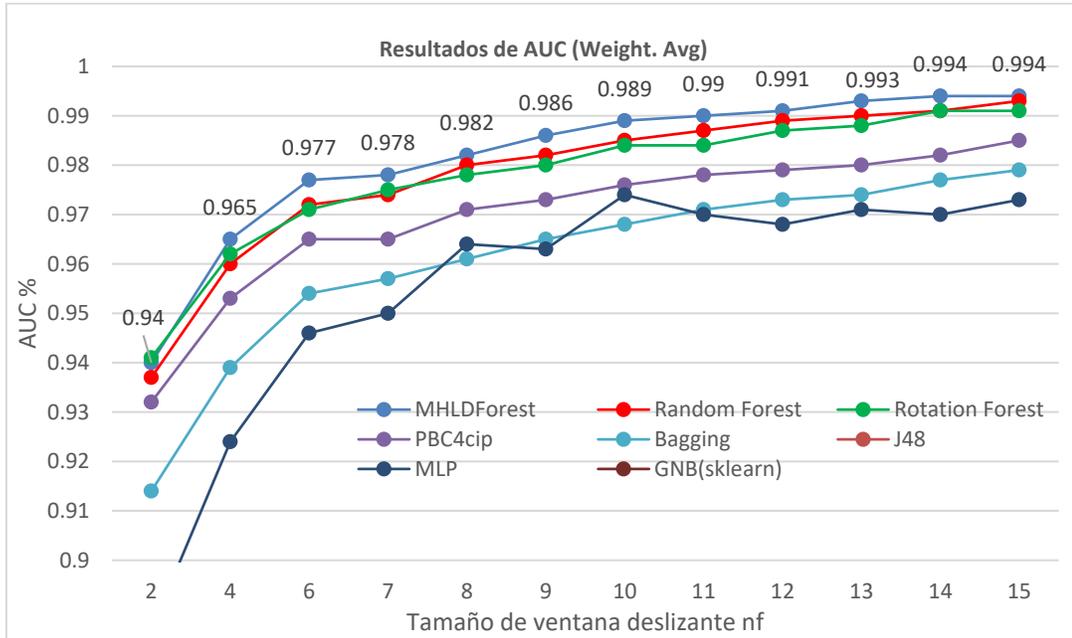


Figura 9. Resultados de AUC en el sensor Acelerómetro, en [10]

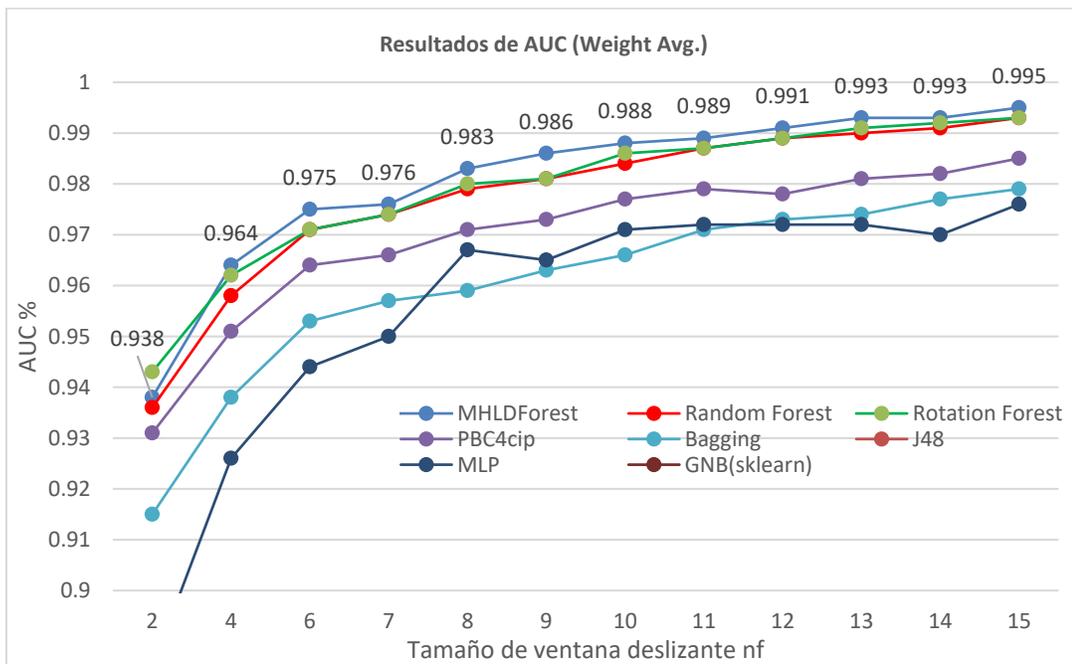


Figura 10. Resultados de AUC en el sensor Acelerómetro-Lineal, en [10]

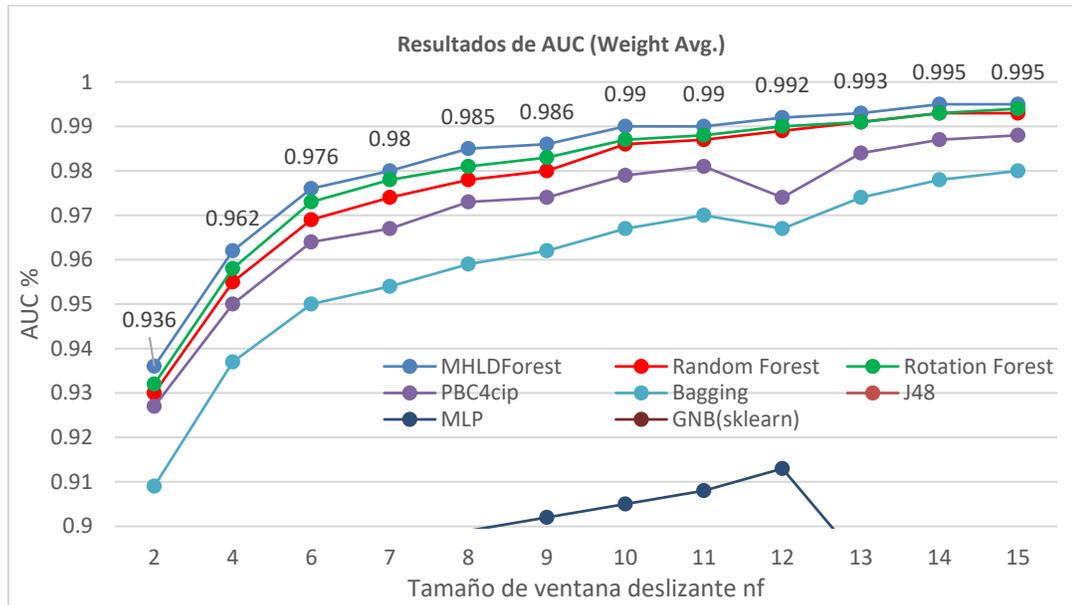


Figura 11. Resultados de AUC en los sensores Acelerómetro-Giroscopio-Combinado, en [10]

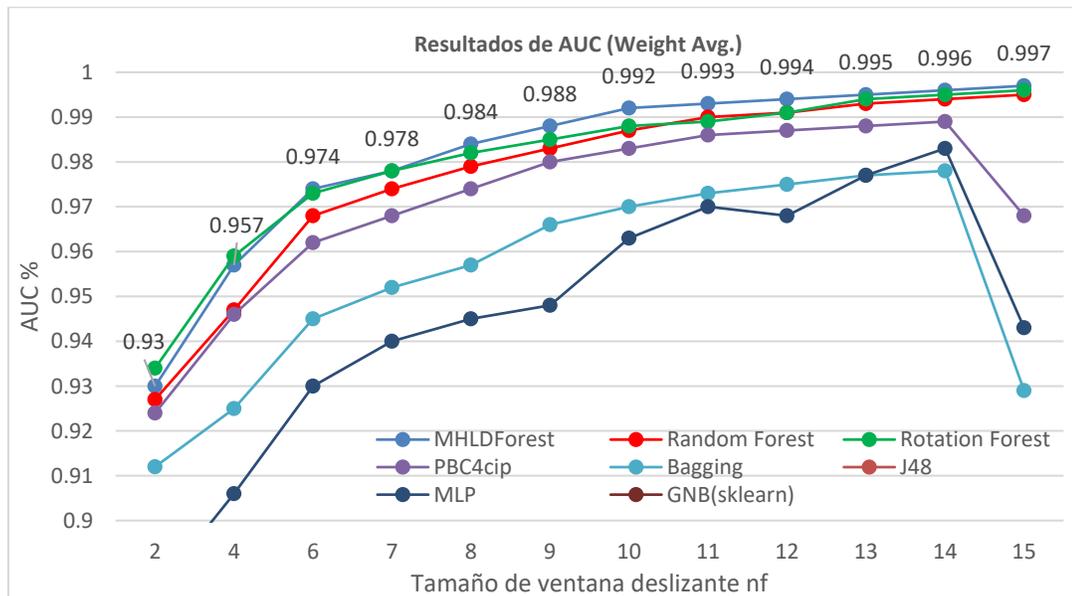


Figura 12. Resultados de AUC en el sensor Giroscopio, en [10]

Como se puede apreciar los mejores resultados respecto al AUC están por encima del 99%, para determinar cuáles son los mejores clasificadores y ver si existen diferencias significativas entre sus resultados es necesario realizar pruebas estadísticas.



4.2.2. EXPERIMENTOS DE LA METODOLOGÍA DE FERREIRA ET AL. [10] EN EL DATASET DE M. RICARDO ET AL. [14]

La metodología propuesta por Ferreira et al. [10], solo se puede aplicar en este dataset con un tamaño de ventana deslizante $nf = 2$. Esto se debe a que el menor de los eventos en este dataset dura 1.96 segundos, para poder aplicar un tamaño nf mayor necesitaríamos información de la señal antes y después de que ocurrió el evento para que la ventana deslizante pueda moverse sobre dicho evento en el tiempo. Esto sería posible si los eventos estuvieran representados en el tiempo con una secuencia de ocurrencia entre ellos. Los eventos en este dataset fueron capturados en diferentes días, por diferentes conductores sin un orden lógico. El dataset de M. Ricardo et al. [14], solo tiene datos del sensor acelerómetro. Debido a que solo se pudo generar un solo dataset, no se realizarán pruebas estadísticas, ya que contamos con pocos datos. Para ver más información ver el Anexo 2.

4.2.3. EXPERIMENTOS DE BOW EN EL DATASET PROPUESTO POR FERREIRA ET AL. [10]

En esta subsección se explica la extracción de características y generación de conjuntos de datos a partir del dataset propuesto en [10]. Se muestra además una representación de los principales resultados en función de la métrica AUC (Weighted Avg.) por cada una de las variantes de las combinaciones de ejes propuestas, en el sensor acelerómetro.

¿Cómo se extraen las características?

Las características de este dataset se extraen aplicando la metodología BoW propuesta en M. Ricardo et al. [14] a la señal del acelerómetro. Como resultado previo de la investigación de los autores de la metodología en este dataset, se tomaron las combinaciones $(K, L)^2$, en las cuales se obtuvieron mejores resultados para generar los vectores de

² (K, L) : donde K indica la longitud de las palabras con L lecturas.



características. Las mejores combinaciones (K, L) son las siguientes: $\{(5, 100), (10, 5), (10, 150), (50, 5), (50, 100)\}$.

¿Qué conjuntos de datos se generan?

Teniendo en cuenta las siguientes combinaciones de los ejes del sensor acelerómetro: $\{(X, Y, Z), (X, Y), \|X, Y\|_2, X, Y\}$, se generó por cada combinación de eje 5 datasets que se corresponden a $(K, L) = \{(5, 100), (10, 5), (10, 150), (50, 5), (50, 100)\}$. En total se generaron $5 * 5 = 25$ datasets, donde 5 es la cantidad de combinaciones de ejes del acelerómetro y 5 la cantidad de combinaciones (K, L) . De estos datasets se seleccionaron las dos combinaciones de ejes $\{\|X, Y\|_2, X\}$ de mejores resultados al aplicar el filtro de Análisis Discriminante Lineal de Fisher (ADLF). Se generaron $2 * 5 = 10$ datasets, donde 2 son las combinaciones de ejes y 5 la cantidad de combinaciones (K, L) . A estos 10 datasets se le volvió a repetir la actividad de clasificación.

¿Dónde se encuentran los resultados de la experimentación?

En el Anexo 3, en la Tabla 23 están todos los resultados de los 8 clasificadores con los que se experimentó en los 25 datasets generados a partir del dataset de Ferreira et al. [10] usando BoW. A cada clasificador se le realizó una validación cruzada con 10 folds.

8 clasificadores utilizados en la experimentación del dataset de Ferreira et al. [10]: MHLDForest, Random Forest, Rotation Forest, PBC4cip, Bagging, J48, MLP(sklearn) y GNB(sklearn).

¿Qué datos se seleccionan para realizar la prueba estadística 5.2?

Para realizar las pruebas estadísticas se seleccionan los 10 datasets donde se obtuvieron mejores resultados después de aplicar el filtro de ADLF. Estos fueron los generados con las combinaciones de ejes $\{\|X, Y\|_2, X\}$. Para evaluar los clasificadores en cada uno de los 10 datasets seleccionados se seleccionó la métrica AUC (Weighted Avg.). Se construyó una matriz de 9 columnas X 10 filas. La primera columna se corresponde al identificador del dataset y el resto de las columnas son los 8 clasificadores utilizados en la experimentación. La intersección del identificador del dataset con cada uno de los 8 clasificadores, corresponde al valor de AUC obtenido en ese dataset para ese clasificador. En



la Tabla 24 del Anexo 3, se encuentran los valores de AUC utilizados para construir la matriz, estos valores son el resultado de la segunda actividad de clasificación sobre los 10 datasets seleccionados.

Representación de los resultados agrupados por combinaciones de ejes

Seleccionamos de la Tabla 21 del Anexo 3 los resultados en la métrica AUC de los 8 clasificadores utilizados en la experimentación del sensor acelerómetro. Para ver el resultado de los clasificadores se agrupó por las combinaciones de ejes: $\{(X, Y, Z), (X, Y), \|X, Y\|_2, X, Y\}$. A continuación, se muestra una representación gráfica de dichos resultados:

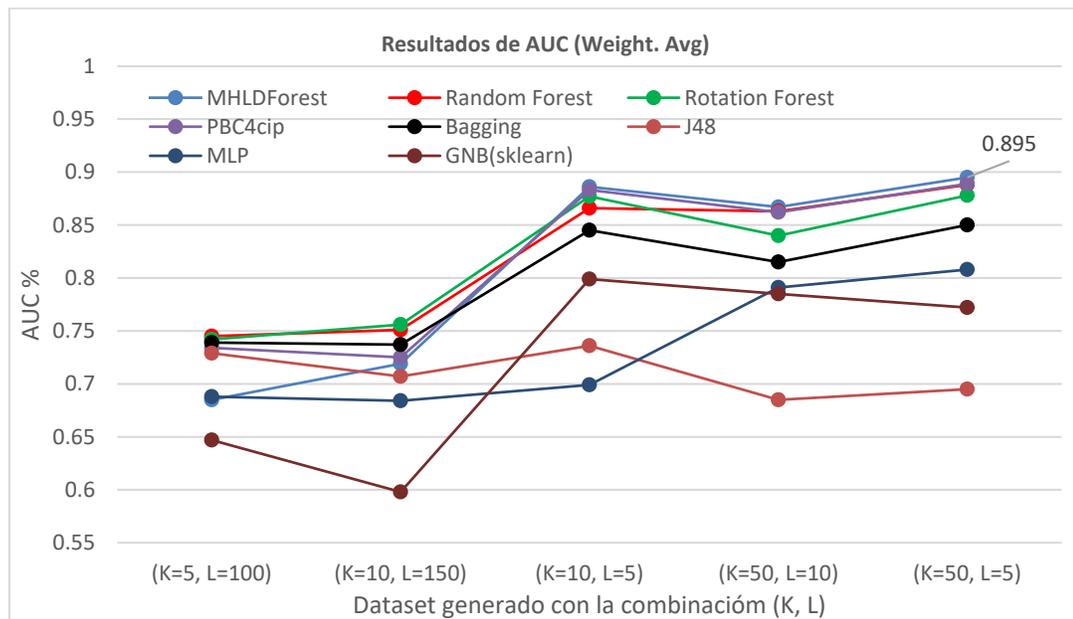


Figura 13. Resultados de AUC en el sensor Acelerómetro, ejes (X,Y,Z), en [10]

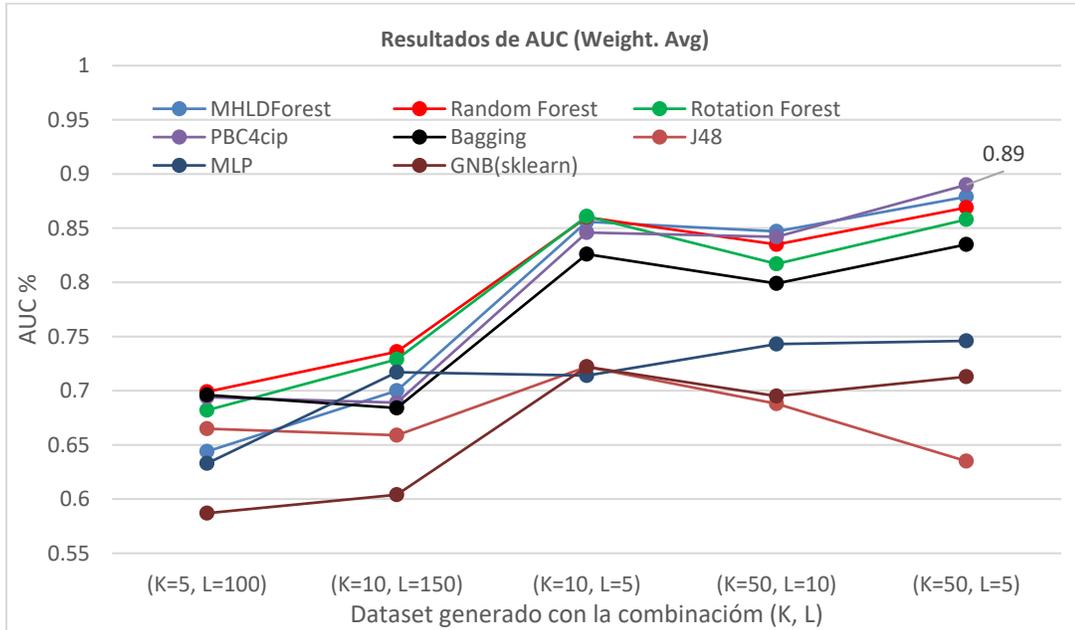


Figura 14. Resultados de AUC en el sensor Acelerómetro, ejes (X, Y), en [10]

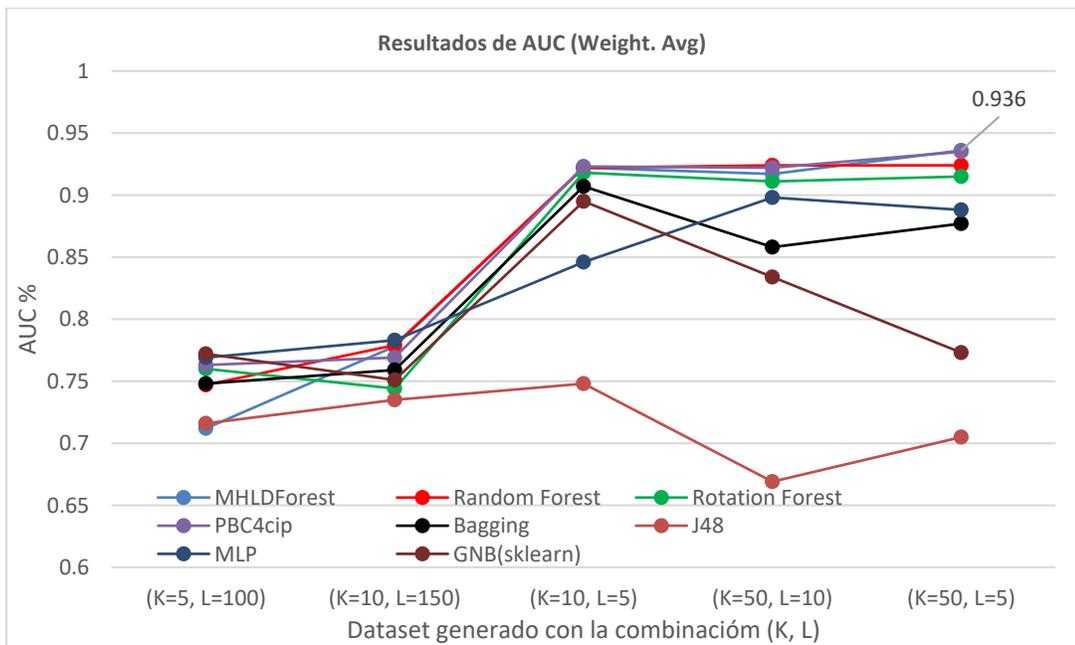


Figura 15. Resultados de AUC en el sensor Acelerómetro, ejes X_1, X_2 , en [10]

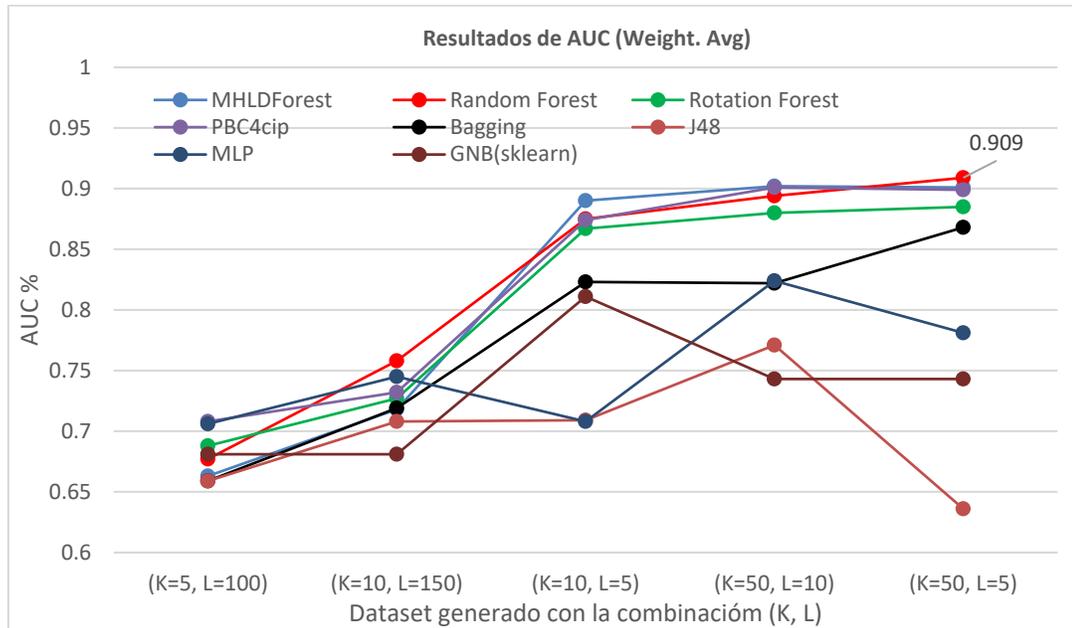


Figura 16. Resultados de AUC en el sensor Acelerómetro, eje X, en [10]

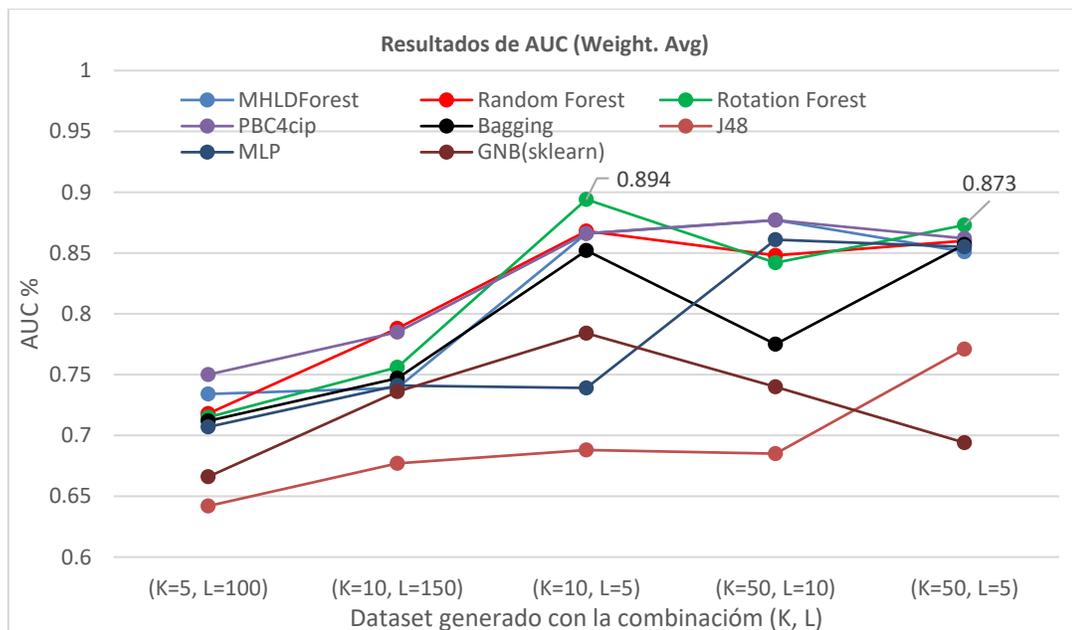


Figura 17. Resultados de AUC en el sensor Acelerómetro, eje Y, en [10]

Como en el trabajo de M. Ricardo et al. [14], los mejores resultados se obtuvieron en los conjuntos de datos generados a partir de la aplicación de BoW a los ejes $\{\|X, Y\|_2, X\}$, donde los resultados alcanzan en la métrica AUC valores por encima del 90%.



Con el objetivo de reducir la dimensionalidad, aplicamos el filtro de Análisis Discriminante Lineal de Fisher (ADLF) incluido en Weka para problemas de aprendizaje supervisado con clasificación multiclase. Al aplicar este filtro a todos los conjuntos de datos que se obtuvieron mediante la aplicación de la metodología BoW, obtuvimos nuevos conjuntos de datos a los cuales le repetimos la actividad de clasificación aplicando validación cruzada con 10 folds. A continuación se muestran los mejores resultados obtenidos en la métrica AUC en el dataset de Ferreira et al. [10] hasta el momento, 99.8% y 99.9% en los ejes $\{\|X, Y\|_2, X\}$. Para poder determinar cuáles son los mejores clasificadores y ver si existen diferencias significativas entre sus resultados es necesario realizar pruebas estadísticas.

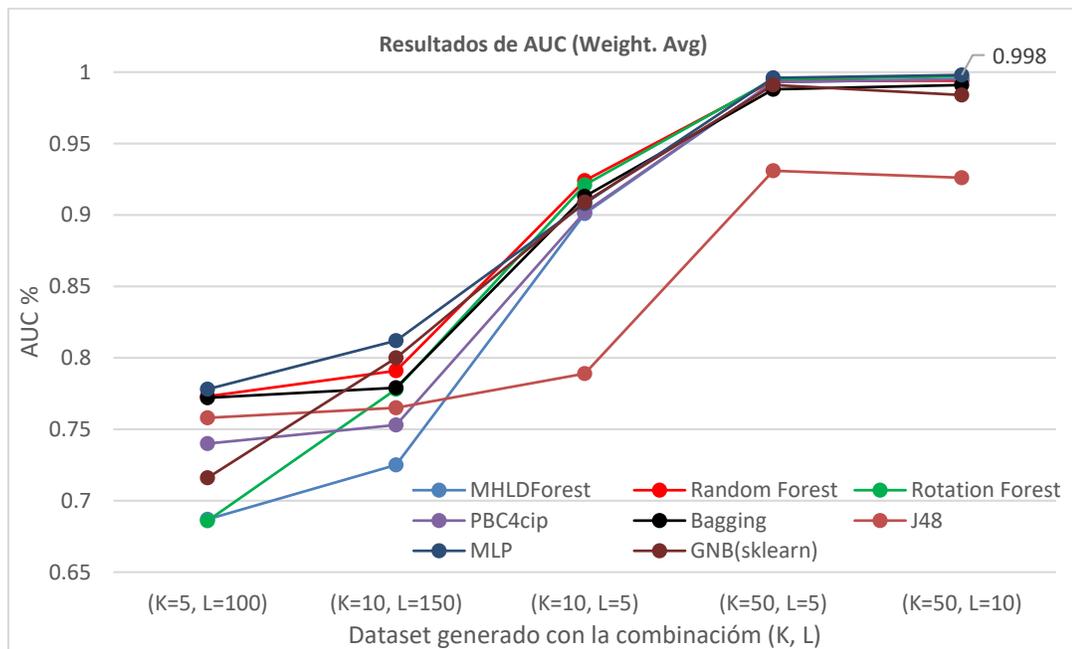


Figura 18. Resultados de AUC en el sensor Acelerómetro, ejes $\|X, Y\|_2$, con ADLF, en [10]

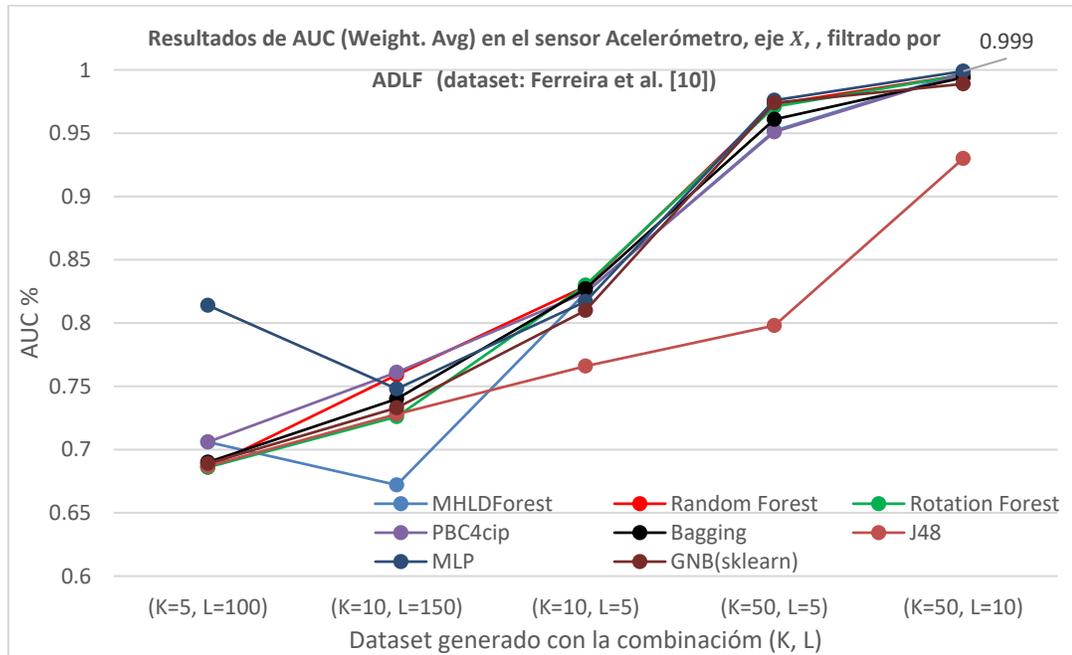


Figura 19. Resultados de AUC en el sensor Acelerómetro, eje X, con ADLF, en [10]

4.2.4. EXPERIMENTOS DE BOW EN EL DATASET PROPUESTO POR M. RICARDO ET AL. [14]

En esta subsección se explica la extracción de características y generación de conjuntos de datos a partir del dataset propuesto en [14]. Se muestra además una representación de los principales resultados en función de la métrica AUC (Weighted Avg.) por cada una de las combinaciones de ejes propuestas, en el sensor acelerómetro.

¿Cómo se extraen las características?

Las características de este dataset se extraen aplicando la metodología BoW propuesta en M. Ricardo et al. [14] a la señal del acelerómetro. Como resultado previo de la investigación de los autores de la metodología en este dataset, se tomaron las combinaciones (K, L) , en las cuales se obtuvieron mejores resultados para generar los vectores de características. Las mejores combinaciones (K, L) son las siguientes: $\{(50, 150), (100, 200), (100, 150), (150, 200)\}$.



¿Qué conjuntos de datos se generan?

Teniendo en cuenta las siguientes combinaciones de los ejes del sensor acelerómetro: $\{(X, Y, Z), (X, Y), \|X, Y\|_2, X, Y\}$, se generó por cada combinación de ejes 4 datasets que se corresponden a $(K, L) = \{(50, 150), (100, 200), (100, 150), (150, 200)\}$. En total se generaron $5 * 4 = 20$ datasets, donde 5 es la cantidad de combinaciones de ejes del acelerómetro y 4 la cantidad de combinaciones (K, L) . De estos datasets se seleccionaron las dos combinaciones de ejes de mejores resultados al aplicar el filtro de Análisis Discriminante Lineal de Fisher (ADLF), estas combinaciones de ejes fueron $\{\|X, Y\|_2, X\}$. Se generaron $2 * 4 = 8$ datasets, donde 2 son las combinaciones de ejes y 4 la cantidad de combinaciones (K, L) . A estos 8 datasets se le volvió a repetir la actividad de clasificación.

¿Dónde se encuentran los resultados de la experimentación?

En el Anexo 4, en la tabla 26 están todos los resultados de los 8 clasificadores con los que se experimentó en los 16 datasets generados a partir del dataset de M. Ricardo et al. [14] usando BoW. A cada clasificador se le realizó una validación cruzada con 10 folds.

8 clasificadores utilizados en la experimentación del dataset de Ferreira et al. [10]: MHLDForest, Random Forest, Rotation Forest, PBC4cip, Bagging, J48, MLP(sklearn) y GNB(sklearn).

¿Qué datos se seleccionan para realizar la prueba estadística 5.3?

Para realizar las pruebas estadísticas se seleccionan los 8 datasets donde se obtuvieron mejores resultados después de aplicar el filtro de ADLF. Estos fueron los generados con las combinaciones de ejes $\{\|X, Y\|_2, X\}$. Para evaluar los clasificadores en cada uno de los 8 datasets seleccionados se seleccionó la métrica AUC (Weighted Avg.). Se construyó una matriz de 9 columnas X 8 filas. La primera columna se corresponde al identificador del dataset y el resto de las columnas son los 8 clasificadores utilizados en la experimentación. La intersección del identificador del dataset con cada uno de los 8 clasificadores, corresponde al valor de AUC obtenido en ese dataset para ese clasificador. En la Tabla 27 del Anexo 4, se encuentran los valores de AUC utilizados para construir la matriz, estos son el resultado de la segunda actividad de clasificación sobre los 8 datasets seleccionados.



Representación de los resultados agrupados por combinaciones de ejes

Seleccionamos de la Tabla 27 del Anexo 4 los resultados en la métrica AUC de los 8 clasificadores utilizados en la experimentación del sensor acelerómetro. Para ver el resultado de los clasificadores se agrupó por las combinaciones de ejes: $\{(X, Y, Z), (X, Y), \|X, Y\|_2, X, Y\}$. A continuación, se muestra una representación gráfica de dichos resultados:

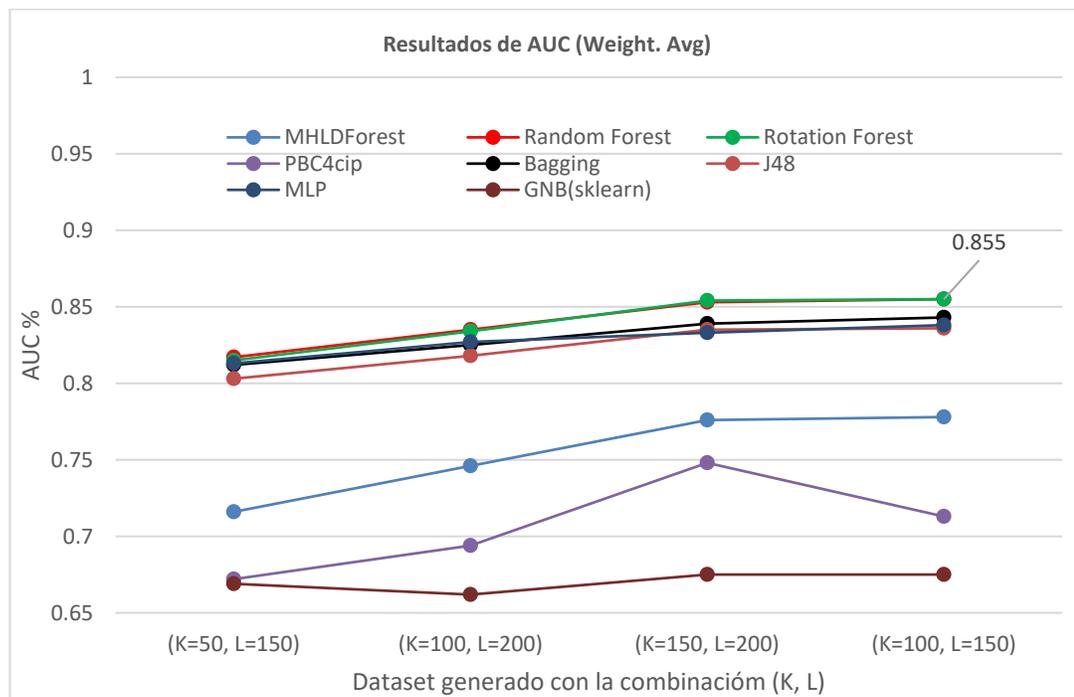


Figura 20. Resultados de AUC en el sensor Acelerómetro, ejes (X, Y, Z) , en [14]

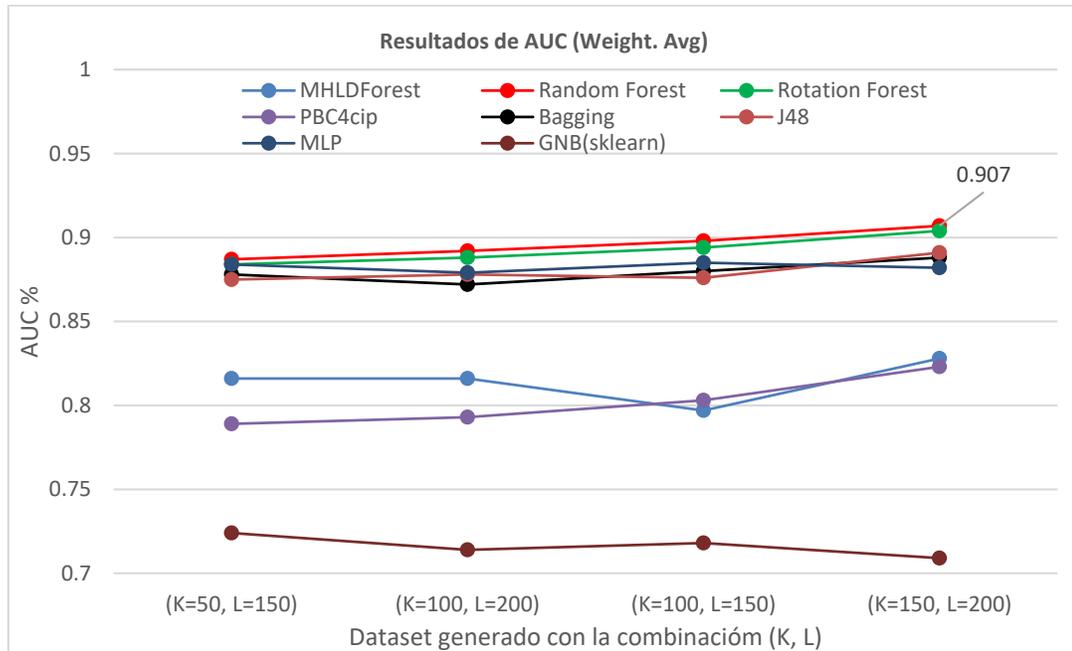


Figura 21. Resultados de AUC en el sensor Acelerómetro, ejes (X, Y), en [14]

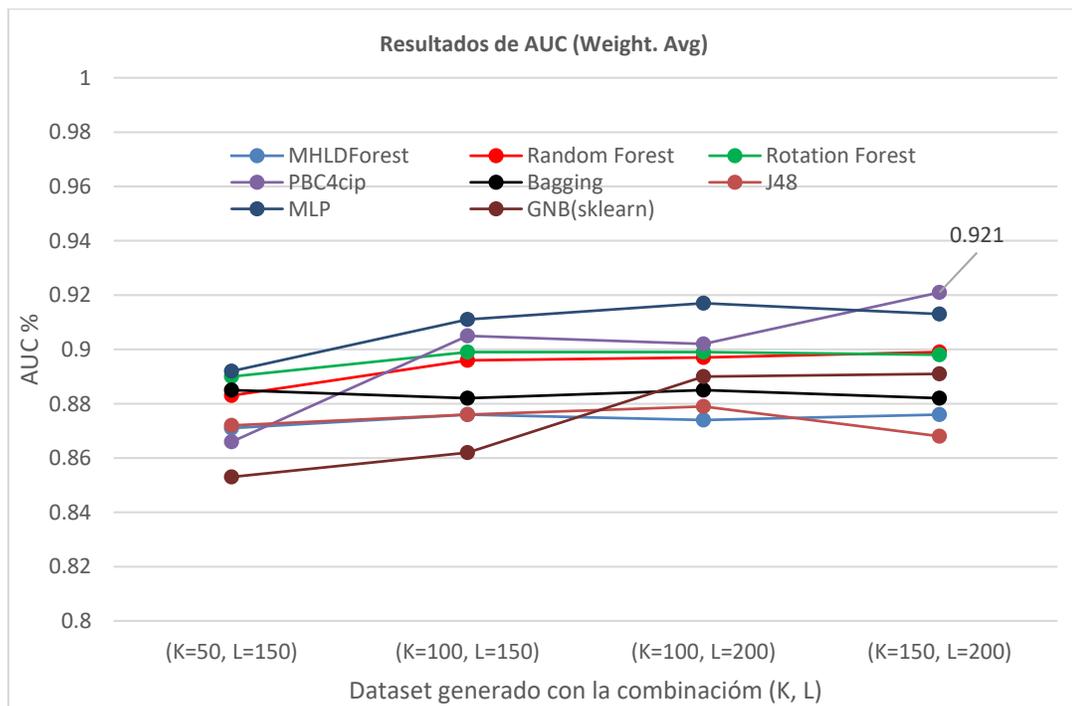


Figura 22. Resultados de AUC en el sensor Acelerómetro, ejes $|X, Y|_2$, en [14]

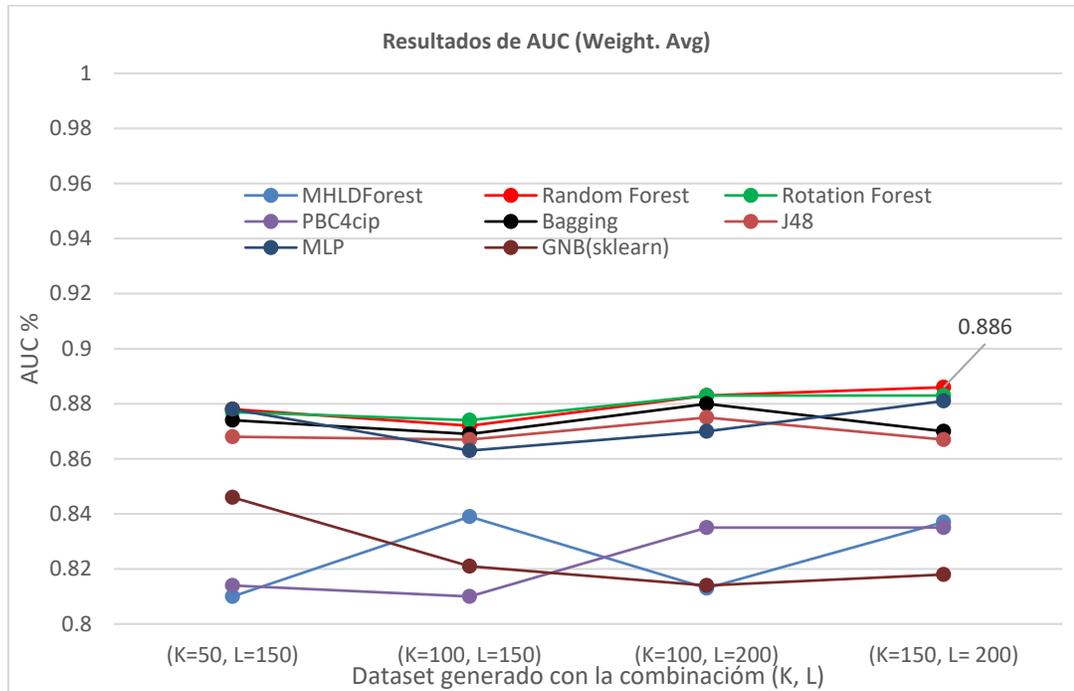


Figura 23. Resultados de AUC en el sensor Acelerómetro, eje X, en [14]

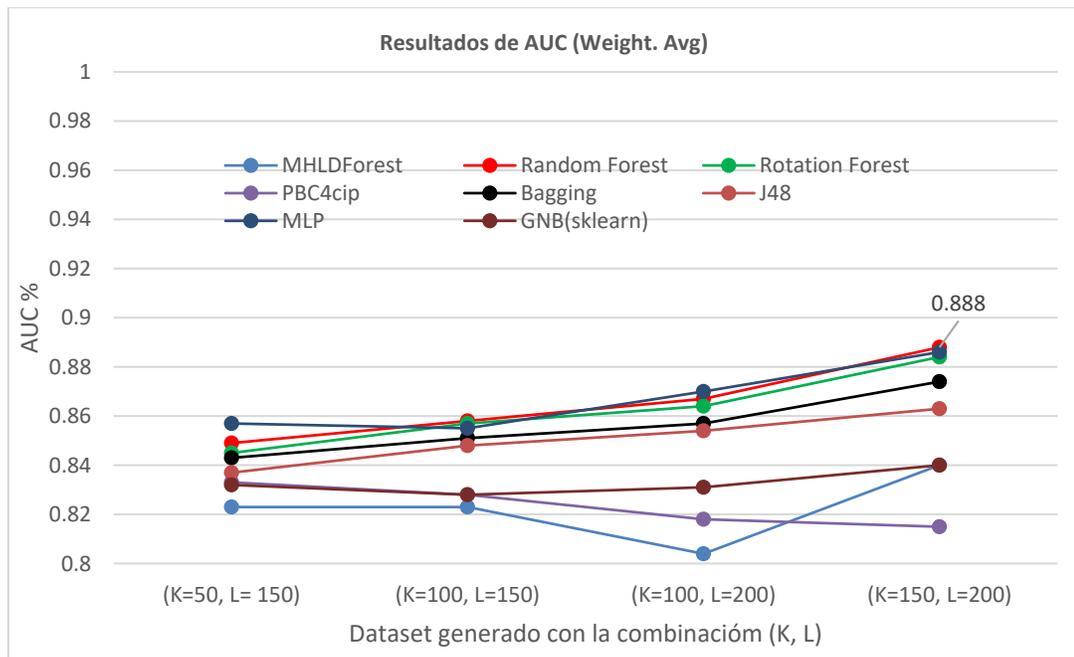


Figura 24. Resultados de AUC en el sensor Acelerómetro, eje Y, en [14]

Como se puede ver los mejores resultados se obtuvieron en la combinación $\{\|X, Y\|_2\}$. Al aplicar el filtro ADLF a todos los conjuntos de datos que se obtuvieron mediante



la aplicación de la metodología BoW, obtuvimos nuevos conjuntos de datos a los cuales le repetimos la actividad de clasificación, aplicando validación cruzada con 10 folds. Los mejores resultados se obtuvieron en los ejes: $\{\|X, Y\|_2, X\}$, a continuación se muestran los resultados:

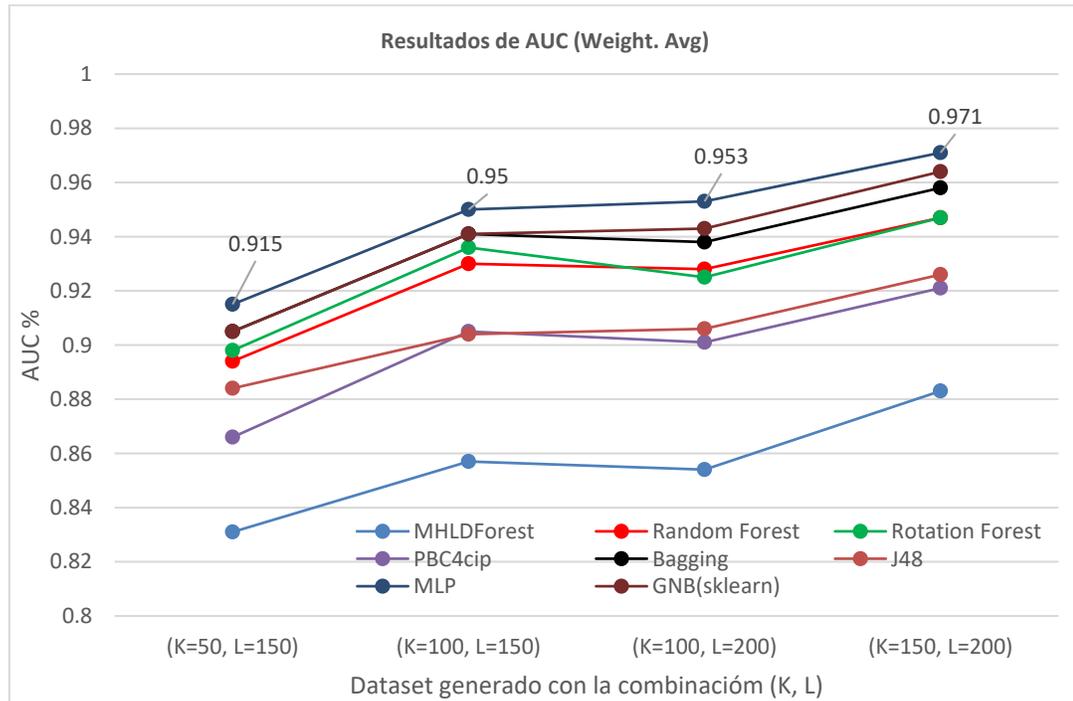


Figura 25. Resultados de AUC en el sensor Acelerómetro, ejes $\|X, Y\|_2$ con ADLF en [14]

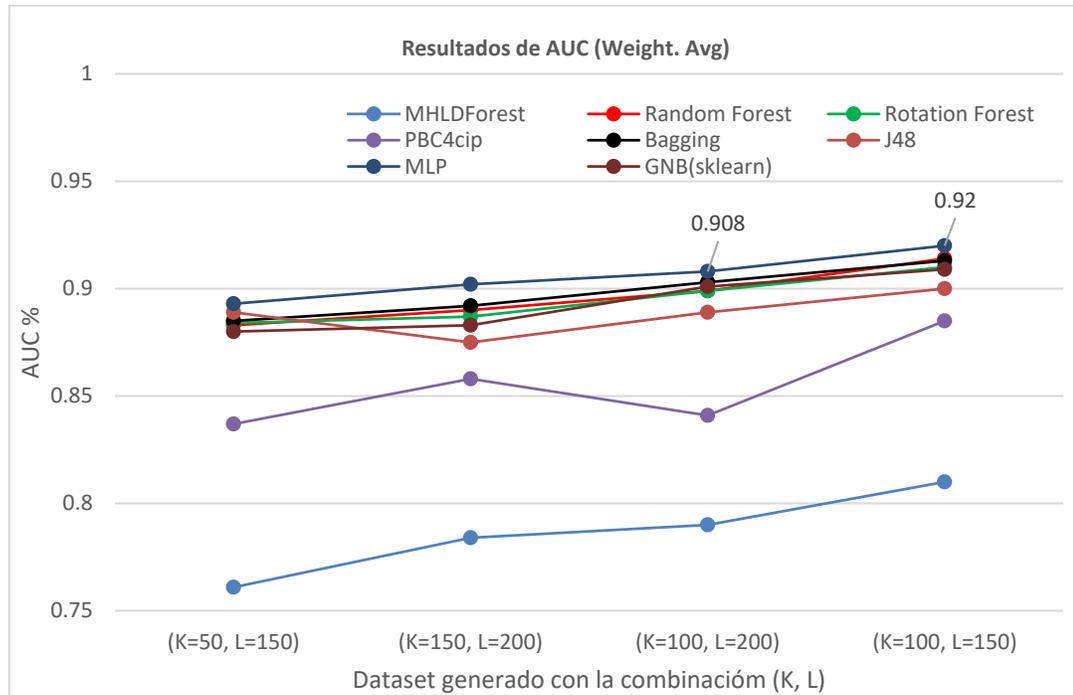


Figura 26. Resultados de AUC en el sensor Acelerómetro, eje X, en [14]

El valor de AUC más alto que se alcanzó con la metodología BoW en el dataset propuesto por M. Ricardo et al. [14] fue de 97.1%, esto fue usando como clasificador una red neuronal MLP en un dataset generado con la configuración, palabras de tamaño $K = 150$ con $L = 200$ lecturas. Este resultado no supera al alcanzado por la metodología de Ferreira et al. [10] en este mismo dataset.

El mejor resultado que hemos alcanzado fue usando la metodología Ferreira et al. [10], con una ventana deslizante de tamaño $nf = 2$, leyendo del acelerómetro y utilizando como clasificador MHLDTForest obtuvimos un AUC de 98.3%. Podemos concluir que en este dataset propuesto por M. Ricardo et al.[14] usando AUC como medida, el uso de la metodología de Ferreira et al. [10] obtuvo un resultado superior que la metodología BoW.

4.2.5. EXPERIMENTOS DE LA METODOLOGÍA DE FERREIRA ET AL. [10] EN EL DATASET DE BEJANI Y GHATEE [13]

En esta subsección se explica la extracción de características y generación de conjuntos de datos a partir del dataset propuesto en [13]. Se muestra además una



representación de los principales resultados en función de la métrica AUC (Weighted Avg.) por cada una de las variantes de los sensores con los que se experimenta.

¿Cómo se extraen las características?

Las características del dataset propuesto se extraen usando la metodología propuesta por Ferreira et al. [10]. Los eventos de dicho dataset están recolectados a una frecuencia de 10 Hz y todos los eventos del groundtruth tienen una duración de 10 segundos. Los tiempos entre eventos oscilan en un rango de 0 a 17 segundos y existen eventos agresivos de diferente clase donde la duración de tiempo entre ellos es 0. Si usamos ventanas deslizantes de tamaño $nf > 11$ con paso 1 segundo para recorrer la señal, habrá ocasiones donde la ventana contenga la ocurrencia de 3 eventos al mismo tiempo. En el dataset propuesto por Ferreira et al. [10], como el tiempo mínimo entre eventos agresivos era de 17.5 segundos nunca la ventana deslizante contenía información de dos eventos agresivos al mismo tiempo. En este experimento contemplamos el caso de que la ventana deslizante puede contener dos eventos agresivos al mismo tiempo, en este caso la clase del vector de características será la del primer cuadro $f_{-(nf-1)}$. Teniendo en cuenta esto se decidió probar con los siguientes tamaños $nf = \{2, 4, 6, 7, 8, 9, 10, 11\}$, de ventanas deslizantes.

¿Qué conjuntos de datos se generan?

Siguiendo la metodología propuesta por Ferreira et al. [10] se generó un dataset por cada tamaño de ventana deslizante $nf = \{2, 4, 6, 7, 8, 9, 10, 11\}$, para cada una de las 3 variantes de sensores utilizadas. Las variantes de sensores utilizadas fueron: acelerómetro, magnetómetro y una combinación por medio de concatenación de los sensores acelerómetro y magnetómetro. Es importante aclarar que en este conjunto de datos solo vienen grabaciones de los sensores: acelerómetro, magnetómetro y de localización. En total se generaron $8 * 3 = 24$ datasets, donde 8 es la cantidad de tamaños de ventanas deslizantes nf y 3 las variantes de sensores. Se usaron todos los valores de los ejes x, y, z de cada una de las variantes de sensores utilizadas.



¿Dónde se encuentran los resultados de la experimentación?

En el Anexo 5, en la Tabla 29 están todos los resultados de los 8 clasificadores con los que se experimentó en los 24 datasets generados a partir del dataset de Bejani y Ghatee [13], utilizando la metodología propuesta por Ferreira et al. [10]. A cada clasificador se le realizó una validación cruzada con 10 folds.

8 clasificadores utilizados en la experimentación del dataset de Bejani y Ghatee[13]: MHLDForest, Random Forest, Rotation Forest, PBC4cip, Bagging, J48, MLP y GNB(sklearn).

¿Qué datos se seleccionan para realizar la prueba estadística 5.4?

Para evaluar los clasificadores en cada uno de los 24 datasets generados se seleccionó la métrica AUC (Weighted Avg.). Se construyó una matriz de 9 columnas X 24 filas. La primera columna se corresponde al identificador del dataset y el resto de las columnas son los 8 clasificadores utilizados en la experimentación. La intersección del identificador del dataset con cada uno de los 8 clasificadores corresponde al valor de AUC obtenido en ese dataset para ese clasificador. En la tabla 30 del Anexo 5, se encuentran los valores de AUC utilizados para construir la matriz. Los valores de la Tabla 30 se extrajeron de la tabla 29 donde están los resultados de todas las métricas calculadas en los 24 datasets generados para los 8 clasificadores.

Representación de los resultados agrupados por sensor

Con el objetivo de mostrar los resultados que se obtuvieron en la métrica AUC (Weighted Avg.) en la intersección, clasificador, dataset, se muestra una gráfica para cada una de las 4 variantes de sensores utilizados. Esto permite analizar el comportamiento de la actividad de clasificación de eventos agresivos, a través de la métrica AUC, a medida que aumenta el tamaño de la ventana deslizante aumenta el valor de AUC en cada uno de los sensores. Es importante resaltar que cada tamaño de ventana deslizante, en cada uno de los sensores, se corresponde a un dataset generado a partir de ese valor nf .

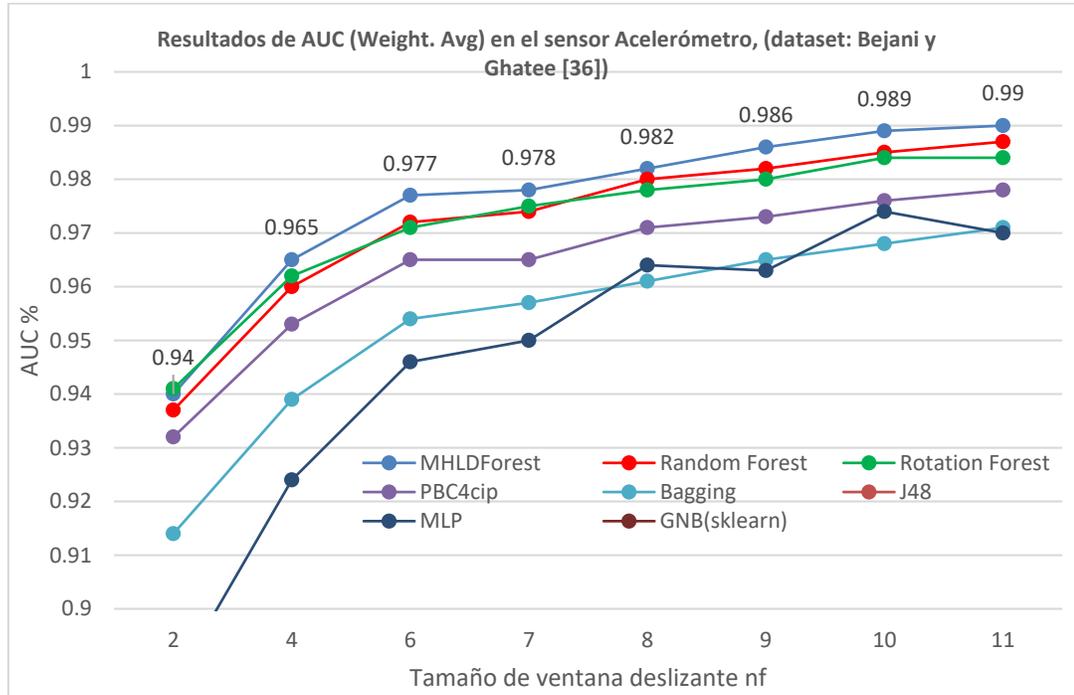


Figura 27. Resultados de AUC en el sensor Acelerómetro, en [36]

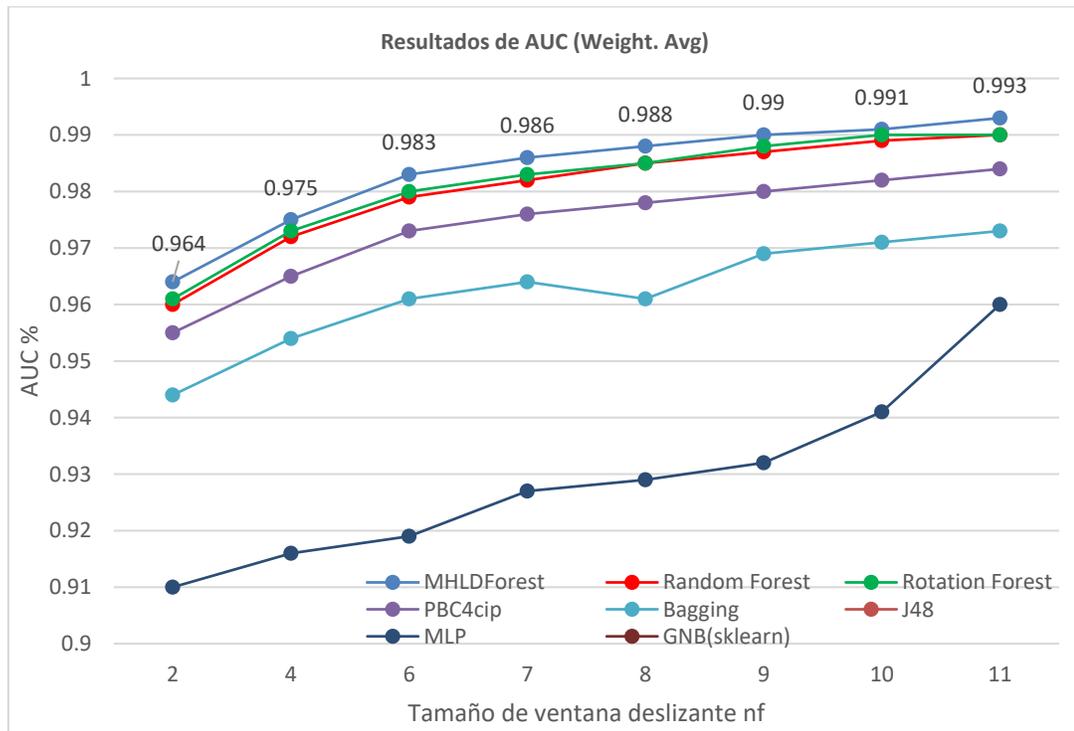


Figura 28. Resultados de AUC en el sensor Magnetómetro, en [36]

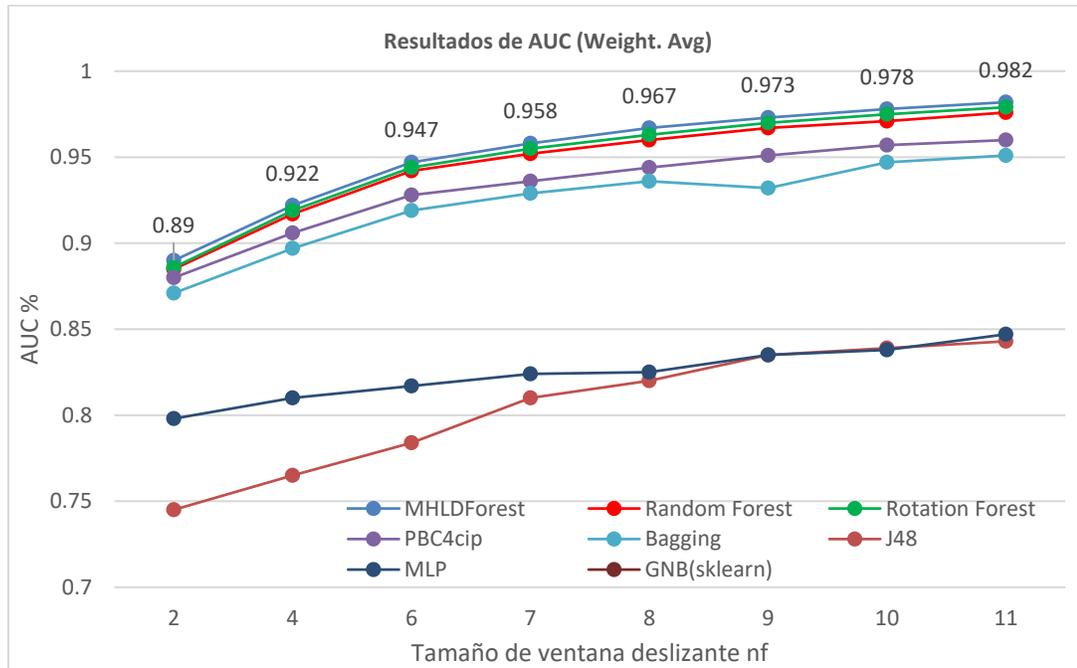


Figura 29. Resultados de AUC en el sensor Acelerómetro-Magnetómetro-Combinado, en [36]

4.2.6. EXPERIMENTOS DE BOW EN EL DATASET PROPUESTO POR BEJANI Y GHATEE [13]

En esta subsección se explica la extracción de características y generación de conjuntos de datos a partir del dataset propuesto en [13]. Se muestra además una representación de los principales resultados en función de la métrica AUC (Weighted Avg.) por cada una de las combinaciones de ejes del magnetómetro propuestas.

¿Cómo se extraen las características?

Las características de este dataset se extraen aplicando la metodología BoW propuesta en M. Ricardo et al. [14] a la señal del magnetómetro. Los eventos agresivos en este dataset entran en la clasificación de cambios de carril, giros y no evento. Se realizó una prueba exploratoria de la metodología BoW en los sensores acelerómetro y magnetómetro probando las combinaciones de ejes $\{\|X, Y\|_2, X\}$, como resultado el magnetómetro obtuvo mejores resultados. Las combinaciones (K, L) , en las cuales se obtuvieron mejores resultados para generar los vectores de características fueron las siguientes: $\{(150, 200), (150, 250), (150, 300), (200, 300)\}$.



¿Qué conjuntos de datos se generan?

Teniendo en cuenta las siguientes combinaciones de los ejes del sensor acelerómetro: $\{(X, Y, Z), (X, Y), \|X, Y\|_2, X, Y\}$, se generó por cada combinación de ejes 4 datasets que se corresponden a las $(K, L) = \{(150, 200), (150, 250), (150, 300), (200, 300)\}$. En total se generaron $5 * 4 = 20$ datasets, donde 5 es la cantidad de combinaciones de ejes del magnetómetro y 4 la cantidad de combinaciones (K, L) . De estos datasets se seleccionaron las dos combinaciones de ejes de mejores resultados al aplicar el filtro de Análisis Discriminante Lineal de Fisher (ADLF), estas combinaciones de ejes fueron $\{\|X, Y\|_2, X\}$. Se generaron $2 * 4 = 8$ datasets, donde 2 son las combinaciones de ejes y 4 la cantidad de combinaciones (K, L) . A estos 8 datasets se le volvió a repetir la actividad de clasificación.

¿Dónde se encuentran los resultados de la experimentación?

En el Anexo 6, en la Tabla 29 están todos los resultados de los 8 clasificadores con los que se experimentó en los 20 datasets generados a partir del dataset de Bejani y Ghatee [13] usando BoW. A cada clasificador se le realizó una validación cruzada con 10 folds.

8 clasificadores utilizados en la experimentación del dataset de Bejani y Ghatee [13]: MHLDForest, Random Forest, Rotation Forest, PBC4cip, Bagging, J48, MLP(sklearn) y GNB(sklearn)

¿Qué datos se seleccionan para realizar la prueba estadística 5.5?

Para realizar las pruebas estadísticas se seleccionan los 8 datasets donde se obtuvieron mejores resultados después de aplicar el filtro de ADLF. Estos fueron los generados con las combinaciones de ejes $\{\|X, Y\|_2, X\}$. Para evaluar los clasificadores en cada uno de los 8 datasets seleccionados se seleccionó la métrica AUC (Weighted Avg.). Se construyó una matriz de 9 columnas X 8 filas. La primera columna se corresponde al identificador del dataset y el resto de las columnas son los 8 clasificadores utilizados en la experimentación. La intersección del identificador del dataset con cada uno de los 8 clasificadores, corresponde al valor de AUC obtenido en ese dataset para ese clasificador. En la tabla 31 del Anexo 6, se encuentran los valores de AUC utilizados para construir la matriz, estos son el resultado de la segunda actividad de clasificación sobre los 8 datasets seleccionados.



A continuación se muestra una representación gráfica donde se muestran los mejores resultados en los ejes: $\{\|X, Y\|_2, X\}$ después de aplicar el filtro ADLF y repetir la actividad de clasificación.

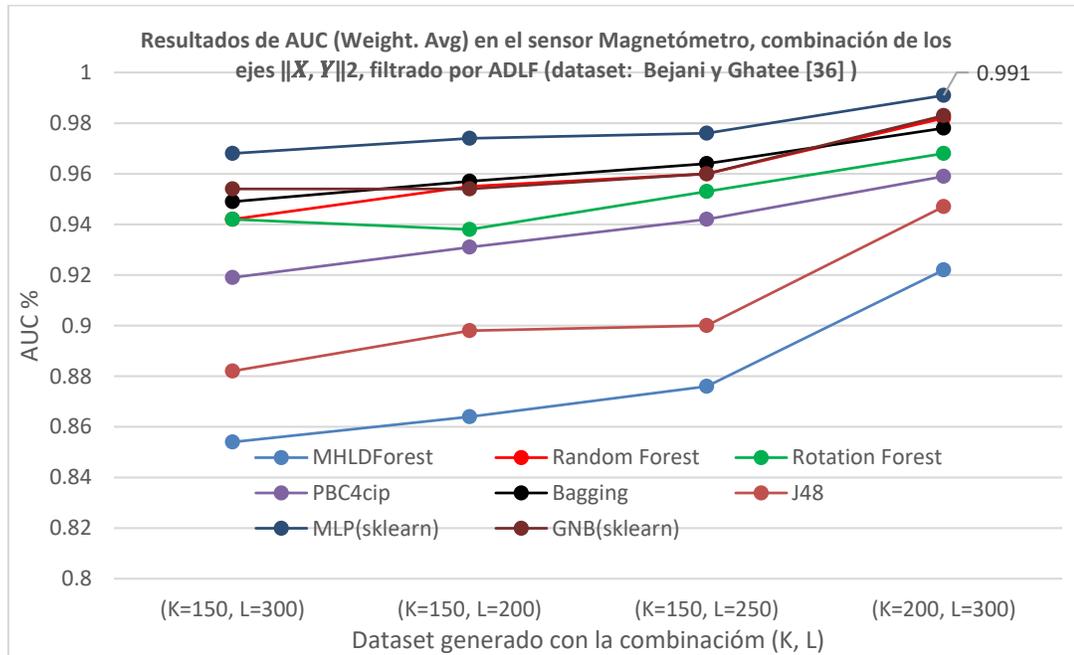


Figura 30. Resultados de AUC en el sensor Magnetómetro, ejes $\|X, Y\|_2$, con ADLF, en [36]

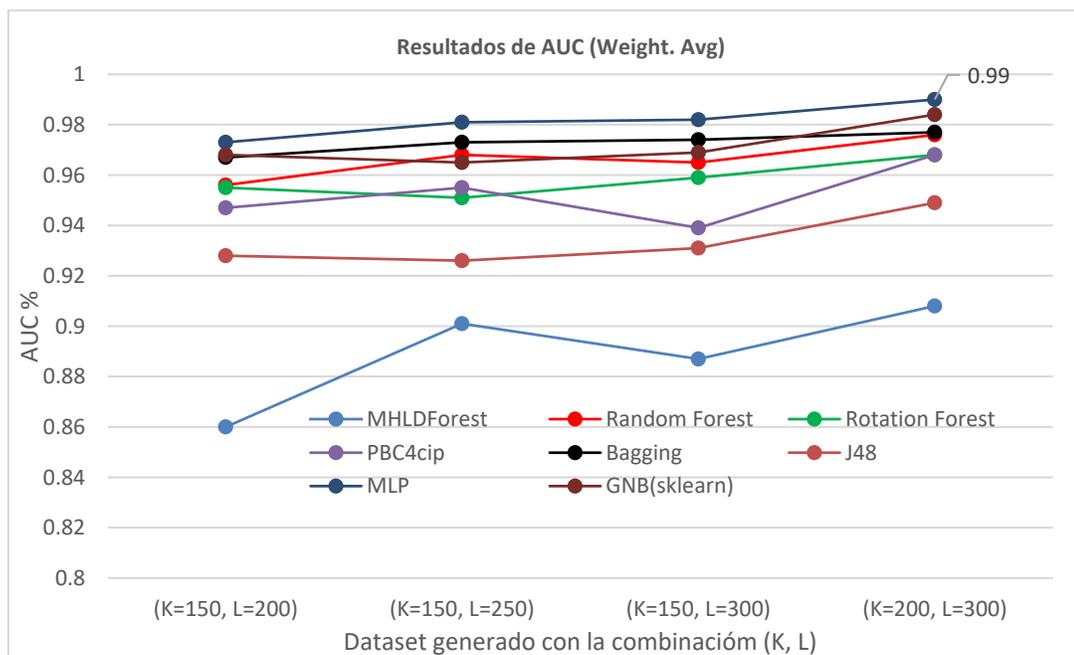


Figura 31. Resultados de AUC en el sensor Magnetómetro, eje X, con ADLF en [36]



Como en el trabajo de M. Ricardo et al. [14], los mejores resultados se obtuvieron en los conjuntos de datos generados a partir de la aplicación de BoW a los ejes $\{\|X, Y\|_2, X\}$. Después de aplicar el filtro ADLF, los resultados alcanzan en la métrica AUC valores hasta el 99.1%.

4.3. CONCLUSIONES DEL CAPÍTULO

En este capítulo se muestran los principales resultados de los algoritmos de clasificación sobre los conjuntos de datos propuestos en la subsección 2.6. Los resultados de la métrica AUC en los experimentos realizados al dataset propuesto por Ferreira et al. [10] usando su metodología están por encima del 99% en la actividad de clasificación de eventos agresivos de conducción. La aplicación del Análisis Discriminante Lineal de Fisher (ADLF) a los conjuntos de datos que se obtuvieron mediante la aplicación de la metodología BoW mejoró los resultados considerablemente. La aplicación de BoW sin el ADLF no supera a los resultados obtenidos con la metodología propuesta por Ferreira et al. [10] basada en la representación de valores estadísticos. Los mejores resultados obtenidos en la métrica AUC en el dataset de Ferreira et al. [10] fueron del 99.8% y 99.9% en los ejes $\{\|X, Y\|_2, X\}$ usando BoW en combinación con el ADLF. Como resultado de la experimentación en los conjuntos de datos propuestos por Bejani y Ghatee [13], M. Ricardo et al. [14], se llega a la conclusión que el uso de la metodología de representación de valores estadísticos obtuvo un resultado superior al de BoW.

Para determinar cuáles son los clasificadores de mejores resultados en cada conjunto de datos y ver si existen diferencias significativas entre sus resultados es necesario realizar pruebas estadísticas.



CAPÍTULO 5: VALIDACIÓN ESTADÍSTICA DE LOS RESULTADOS

Las siguientes pruebas estadísticas están orientadas a seleccionar los mejores clasificadores utilizados en la experimentación del problema de análisis multiclase de eventos agresivos de conducción en los datasets propuestos. Los clasificadores se medirán a través de la métrica AUC. Teniendo en cuenta que se usaron dos metodologías en la experimentación, es objetivo determinar los mejores clasificadores por cada una de ellas. Para esto se desarrollaron cinco pruebas estadísticas.

Primera prueba: Establecer un ranking entre los clasificadores utilizados en la experimentación del dataset de Ferreira et al. [10], de acuerdo a la métrica AUC usando su propia metodología. Determinar si existe diferencia significativa entre ellos con un 0.5% de nivel de significancia.

Segunda prueba: Establecer un ranking entre los clasificadores utilizados en la experimentación del dataset de Ferreira et al. [10], de acuerdo a la métrica AUC usando la metodología BoW. Determinar si existe diferencia significativa entre ellos con un 0.5% de nivel de significancia.

Tercera prueba: Establecer un ranking entre los clasificadores utilizados en la experimentación del dataset de M. Ricardo et al. [14], de acuerdo a la métrica AUC usando la metodología BoW. Determinar si existe diferencia significativa entre ellos con un 0.5% de nivel de significancia.

Cuarta prueba: Establecer un ranking entre los clasificadores utilizados en la experimentación del dataset propuesto por Bejani y Ghatee [13], de acuerdo a la métrica AUC usando la metodología propuesta por Ferreira et al. [10]. Determinar si existe diferencia significativa entre ellos con un 0.5% de nivel de significancia.

Quinta prueba: Establecer un ranking entre los clasificadores utilizados en la experimentación del dataset propuesto por Bejani y Ghatee [13], de acuerdo a la métrica AUC usando la metodología BoW. Determinar si existe diferencia significativa entre ellos con un 0.5% de nivel de significancia.



Para la ejecución de estas pruebas utilizaremos la prueba estadística no paramétrica desarrollada por Milton Friedman. Similar a las medidas repetidas paramétricas ANOVA, se usa para detectar diferencias en los tratamientos en múltiples intentos de prueba [44]. Para determinar si existe diferencia significativa entre los clasificadores usaremos el método *post hoc* propuesto por Nemenyi. En los casos donde no exista diferencia significativa usaremos la prueba de rangos de signos de Wilcoxon para realizar el ranking.

5.1. PRIMERA PRUEBA: PRUEBA DE FRIEDMAN EN FERREIRA ET AL. [10] USANDO SU METODOLOGÍA

Objetivo: Establecer un ranking entre los clasificadores utilizados en la experimentación del dataset de Ferreira et al. [10], de acuerdo a la métrica AUC usando la metodología BoW. Determinar si existe diferencia significativa entre ellos con un 0.5% de nivel de significancia.

Clasificadores utilizados en la experimentación del dataset de Ferreira et al. [10]: MHLDForest, Random Forest, Rotation Forest, PBC4cip, Bagging, J48, MLP y GNB(sklearn).

Nota: Para la realización de esta prueba se utilizaron los datos expuestos en la subsección 4.2.1, para tener claridad cómo se generan y seleccionan por favor lea las respuestas a las preguntas:

- ¿Qué conjuntos de datos se generan?
- ¿Qué datos se seleccionan para realizar pruebas estadísticas?

Hipótesis:

H_0 : No existe diferencia entre los clasificadores utilizados en la experimentación del dataset de Ferreira et al. [10] en la métrica AUC para eventos agresivos de conducción.

H_1 : Existe diferencia entre los clasificadores utilizados en la experimentación del dataset de Ferreira et al. [10] en la métrica AUC para eventos agresivos de conducción.

A continuación, se muestran los rangos promedios obtenidos al aplicar el procedimiento de Friedman:



Tabla 4. Rango promedio de los clasificadores al aplicar Friedman en Ferreira et al. [10]

| Clasificador | Ranking |
|-----------------|---------|
| MHLDForest | 1.0938 |
| Rotation Forest | 2.1979 |
| Random Forest | 2.7083 |
| PBC4cip 4 | 4 |
| Bagging | 5.1771 |
| MLP | 5.8438 |
| J48 | 7.1042 |
| GNB(sklearn) | 7.875 |

Estadística de Friedman considerando el rendimiento de reducción (distribuido según chi-cuadrado con 7 grados de libertad: 326.388889.

P-value calculado por la prueba de Friedman: 1.4062162545513957E-10.

Nivel de significancia: $\alpha = 0.05$.

Decisión: Como P-value < 0.05 se rechaza la hipótesis nula H_0 y aceptamos la hipótesis alternativa de que si existe diferencia significativa entre los clasificadores utilizados en la experimentación del dataset de Ferreira et al. [10] en la métrica AUC para eventos agresivos de conducción.

A continuación, se muestra el Diagrama de Diferencias Críticas (DDC) calculado, con una diferencia crítica de 1.5154395 con la prueba de Nemenyi.

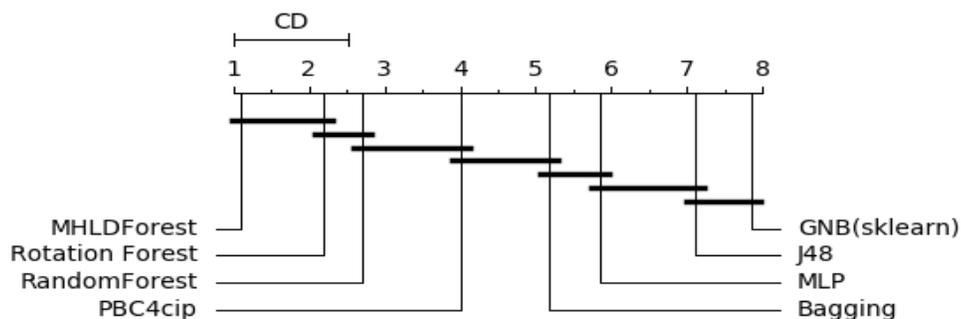


Figura 32. Diagrama de diferencias críticas del AUC en [10] usando su metodología

Como se puede observar hay presencia de subconjuntos homogéneos de medias en los clasificadores que no tienen diferencia significativa entre sí como, por ejemplo: MHLDForest vs Rotation Forest, Rotation Forest vs Random Forest, Random Forest vs



PBC4cip, PBC4cip vs Bagging, Bagging vs MLP, MLP vs J48 y J48 vs GNB(sklearn). Los datos no son suficientes para concluir si el rendimiento de la métrica AUC de eventos agresivos es mejor en el clasificador MHLDForest que en Rotation Forest, pasa lo mismo con Rotation Forest y Random Forest y así sucesivamente. Para clasificar las diferencias en el rendimiento de AUC de cada par de clasificadores mencionados anteriormente por cada conjunto de datos aplicamos la prueba de rango de signos de Wilcoxon[43]. A continuación, se muestra un resumen de los resultados obtenidos en dicha prueba:

Tabla 5. Rangos de signos de Wilcoxon calculados en el dataset de Ferreira et al. [10]

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---------------------|------|--------|--------|--------|--------|--------|--------|--------|
| MHLDForest (1) | - | 1176.0 | 1031.0 | 1176.0 | 1176.0 | 1176.0 | 1176.0 | 1176.0 |
| Random Forest (2) | 0.0 | - | 268.5 | 1176.0 | 1176.0 | 1176.0 | 1176.0 | 1176.0 |
| Rotation Forest (3) | 97.0 | 907.5 | - | 1176.0 | 1176.0 | 1176.0 | 1176.0 | 1176.0 |
| PBC4cip (4) | 0.0 | 0.0 | 0.0 | - | 1176.0 | 1176.0 | 1176.0 | 1176.0 |
| Bagging (5) | 0.0 | 0.0 | 0.0 | 0.0 | - | 1176.0 | 1023.5 | 1176.0 |
| J48 (6) | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | - | 1.0 | 1126.5 |
| MLP (7) | 0.0 | 0.0 | 0.0 | 0.0 | 104.5 | 1175.0 | - | 1176.0 |
| GNB(sklearn) (8) | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 49.5 | 0.0 | - |

● = el método en la fila mejora el método de la columna.

○ = el método en la columna mejora el método de la fila.

Diagonal hacia arriba el nivel de significancia $\alpha = 0.9$ y de la diagonal hacia abajo el nivel significancia $\alpha = 0.95$.

Tabla 6. Resumen de la prueba de rango de signo de Wilcoxon con simbología en [10]

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---------------------|-----|-----|-----|-----|-----|-----|-----|-----|
| MHLDForest (1) | - | ● | ● | ● | ● | ● | ● | ● |
| Random Forest (2) | ○ | - | ○ | ● | ● | ● | ● | ● |
| Rotation Forest (3) | ○ | ● | - | ● | ● | ● | ● | ● |
| PBC4cip (4) | ○ | ○ | ○ | - | ● | ● | ● | ● |
| Bagging (5) | ○ | ○ | ○ | ○ | - | ● | ● | ● |
| J48 (6) | ○ | ○ | ○ | ○ | ○ | - | ○ | ● |
| MLP (7) | ○ | ○ | ○ | ○ | ○ | ● | - | ● |
| GNB(sklearn) (8) | ○ | ○ | ○ | ○ | ○ | ○ | ○ | - |

Tabla 7. Resultados de la prueba de rango de signo de Wilcoxon en Ferreira et al. [10]

| Clasificador | $\alpha = 0.9$ | | $\alpha = 0.95$ | |
|--------------|----------------|---|-----------------|---|
| | + | ± | + | ± |
| MHLDForest | 7 | 7 | 7 | 7 |



| | | | | |
|-----------------|---|---|---|---|
| Rotation Forest | 6 | 6 | 6 | 6 |
| Random Forest | 5 | 5 | 5 | 5 |
| PBC4cip | 4 | 4 | 4 | 4 |
| Bagging | 3 | 3 | 3 | 3 |
| MLP | 2 | 2 | 2 | 2 |
| J48 | 1 | 1 | 1 | 1 |
| GNB(sklearn) | 0 | 0 | 0 | 0 |

Como se puede apreciar los mejores clasificadores según el rendimiento de la métrica AUC son MHLDForest y Rotation Forest en ese orden. Estos dos clasificadores superan a Random Forest que es el mejor clasificador reportado en Ferreira et al. [10], en el caso de MHLDForest lo supera con una diferencia estadísticamente significativa.

5.2. SEGUNDA PRUEBA: PRUEBA DE FRIEDMAN EN FERREIRA ET AL. [10] USANDO BOW

Objetivo: Establecer un ranking entre los clasificadores utilizados en la experimentación del dataset de Ferreira et al. [10], de acuerdo a la métrica AUC usando la metodología BoW. Determinar si existe diferencia significativa entre ellos con un 0.5% de nivel de significancia.

Clasificadores utilizados en la experimentación del dataset de Ferreira et al. [10]: MHLDForest, Random Forest, Rotation Forest, PBC4cip, Bagging, J48, MLP(sklearn) y GNB(sklearn).

Nota: Para la realización de esta prueba se utilizaron los datos expuestos en la subsección 4.2.3, para tener claridad cómo se generan y seleccionan por favor lea las respuestas a las preguntas:

- ¿Qué conjuntos de datos se generan?
- ¿Qué datos se seleccionan para realizar la prueba estadística?

Hipótesis:

H_0 : No existe diferencia entre los clasificadores utilizados en la experimentación del dataset de Ferreira et al. [10] en la métrica AUC para eventos agresivos de conducción.

H_1 : Existe diferencia entre los clasificadores utilizados en la experimentación del dataset de Ferreira et al. [10] en la métrica AUC para eventos agresivos de conducción.



A continuación, se muestran los rangos promedios obtenidos al aplicar el procedimiento de Friedman:

Tabla 8. Rango promedio de los clasificadores al aplicar Friedman en [10] usando BoW

| Clasificador | Ranking |
|-----------------|---------|
| MLP(sklearn) | 2.1 |
| Random Forest | 3.3 |
| Rotation Forest | 4.3 |
| PBC4cip | 4.4 |
| Bagging | 4.5 |
| GNB(sklearn) | 5.1 |
| MHLDForest | 5.3 |
| J48 | 7 |

Estadística de Friedman considerando el rendimiento de reducción (distribuido según chi-cuadrado con 7 grados de libertad: 24.166667.

P-value calculado por la prueba de Friedman: 0.0010649797338115885.

Nivel de significancia: $\alpha = 0.05$

Decisión: Como P-value < 0.05 se rechaza la hipótesis nula H_0 y aceptamos la hipótesis alternativa de que si existe diferencia significativa entre los clasificadores utilizados en la experimentación del dataset de Ferreira et al. [10] en la métrica AUC para eventos agresivos de conducción.

A continuación, se muestra el Diagrama de Diferencias Críticas (DDC) calculado, con una diferencia crítica de 3.320161594737401 con la prueba de Nemenyi.

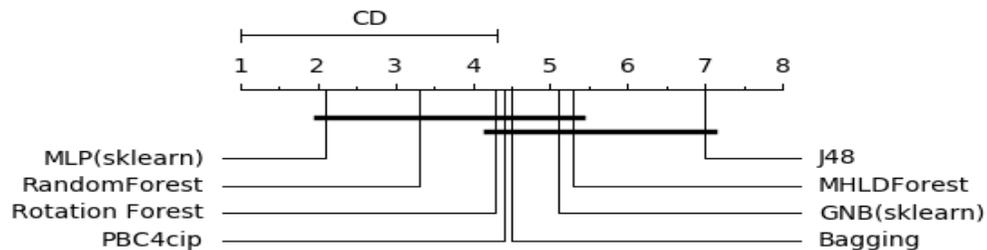


Figura 33. Diagrama de diferencias críticas del resultado del AUC en [10] usando BoW



Como se puede observar hay presencia de subconjuntos homogéneos de medias en los clasificadores que no tienen diferencia significativa entre sí. Solo existe diferencia significativa entre MLP(sklearn) vs J48 y Random Forest vs J48. Los datos no son suficientes para comparar el resto de los clasificadores de eventos agresivos entre sí, ya que no hay diferencias significativas. Para clasificar las diferencias en el rendimiento de AUC de cada par de clasificadores restantes aplicaremos la prueba de rango de signos de Wilcoxon[43]. A continuación, se muestra un resumen de los resultados obtenidos en dicha prueba:

Tabla 9. Resultados de la prueba de rango de signo de Wilcoxon en [10] usando BoW

| Clasificador | $\alpha = 0.9$ | | $\alpha = 0.95$ | |
|-----------------|----------------|----------|-----------------|----------|
| | + | \pm | + | \pm |
| MHLDForest | 0 | 5 | 0 | 5 |
| Random Forest | 4 | 7 | 3 | 7 |
| Rotation Forest | 0 | 7 | 0 | 7 |
| PBC4cip | 1 | 7 | 1 | 7 |
| Bagging | 1 | 5 | 1 | 6 |
| J48 | 0 | 2 | 0 | 2 |
| MLP(sklearn) | 4 | 7 | 3 | 7 |
| GNB(sklearn) | 1 | 5 | 1 | 5 |

Los clasificadores de mejor puntuación son MLP(sklearn) y Random Forest para la metodología BoW aplicada al dataset propuesto por Ferreira et al. [10].

5.3. TERCERA PRUEBA: PRUEBA DE FRIEDMAN EN M. RICARDO ET AL. [14] USANDO BOW

Objetivo: Establecer un ranking entre los clasificadores utilizados en la experimentación del dataset de M. Ricardo et al. [14], de acuerdo a la métrica AUC usando la metodología BoW. Determinar si existe diferencia significativa entre ellos con un 0.5% de nivel de significancia.

Clasificadores utilizados en la experimentación del dataset de M. Ricardo et al. [14]: MHLDForest, Random Forest, Rotation Forest, PBC4cip, Bagging, J48, MLP(sklearn) y GNB(sklearn)



Nota: Para la realización de esta prueba se utilizaron los datos expuestos en la subsección 4.2.4, para tener claridad cómo se generan y seleccionan por favor lea las respuestas a las preguntas:

- ¿Qué conjuntos de datos se generan?
- ¿Qué datos se seleccionan para realizar la prueba estadística?

Hipótesis:

H_0 : No existe diferencia entre los clasificadores utilizados en la experimentación del dataset de M. Ricardo et al. [14] en la métrica AUC para eventos agresivos de conducción.

H_1 : Existe diferencia entre los clasificadores utilizados en la experimentación del dataset de M. Ricardo et al. [14] en la métrica AUC para eventos agresivos de conducción.

A continuación, se muestra los rangos promedio obtenidos al aplicar el procedimiento de Friedman:

Tabla 10. Rango promedio de los clasificadores al aplicar Friedman en [14] usando BoW

| Clasificador | Ranking |
|---------------------|----------------|
| MLP(sklearn) | 1 |
| Bagging | 2.625 |
| GNB(sklearn) | 3.5 |
| Random Forest | 4.125 |
| Rotation Forest | 4.25 |
| J48 | 5.625 |
| PBC4cip | 6.875 |
| MHLDForest | 8 |

Estadística de Friedman considerando el rendimiento de reducción (distribuido según chi-cuadrado con 7 grados de libertad: 410.166667.

P-value calculado por la prueba de Friedman: 3.308299989512875E-8.

Nivel de significancia: $\alpha = 0.05$

Decisión: Como P-value < 0.05 se rechaza la hipótesis nula H_0 y aceptamos la hipótesis alternativa de que si existe diferencia significativa entre los clasificadores utilizados en la experimentación del dataset de Ferreira et al. [10] en el rendimiento de la métrica AUC para eventos agresivos de conducción.



A continuación, se muestra el Diagrama de Diferencias Críticas (DDC) calculado, con una diferencia crítica de 3.712053511058468 con la prueba de Nemenyi.

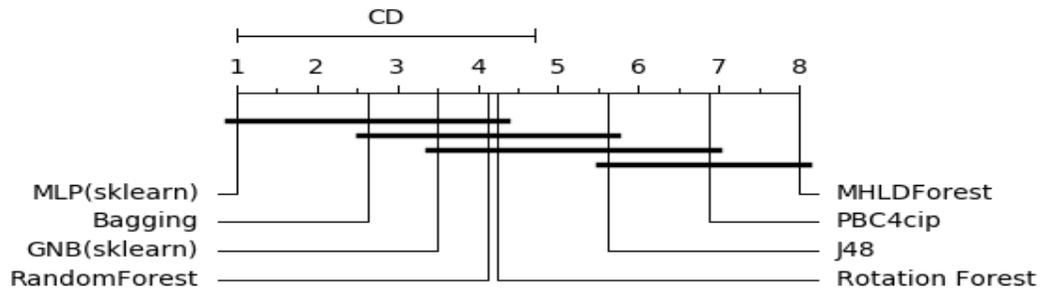


Figura 34. Diagrama de diferencias críticas del AUC en el dataset de [14] usando BoW

Como se puede observar existe diferencia significativa entre MHLDForest vs MLP(sklearn), PBC4cip vs MLP(sklearn), MHLDForest vs Bagging, J48 vs MLP(sklearn), MHLDForest vs GNB(sklearn), PBC4cip vs Bagging y MHLDForest vs Random Forest. Entre el resto de los clasificadores no existe diferencia significativa. Los datos no son suficientes para comparar el resto de los clasificadores de eventos agresivos entre sí, ya que no hay diferencias significativas. Para clasificar las diferencias en el rendimiento de AUC de cada par de clasificadores restantes aplicaremos la prueba de rango de signos de Wilcoxon[43].

Tabla 11. Resultados de la prueba de rango de signo de Wilcoxon en [14] usando BoW

| Clasificador | $\alpha = 0.9$ | | $\alpha = 0.95$ | |
|-----------------|----------------|-------|-----------------|-------|
| | + | \pm | + | \pm |
| MHLDForest | 0 | 0 | 0 | 0 |
| Random Forest | 3 | 5 | 3 | 5 |
| Rotation Forest | 3 | 5 | 3 | 5 |
| PBC4cip | 1 | 1 | 1 | 1 |
| Bagging | 5 | 6 | 5 | 6 |
| J48 | 2 | 2 | 2 | 2 |
| MLP(sklearn) | 7 | 7 | 7 | 7 |
| GNB(sklearn) | 3 | 6 | 3 | 6 |

Como se puede observar los clasificadores de mejor puntuación son MLP(sklearn) y Bagging para la metodología BoW aplicada al dataset propuesto por M. Ricardo et al.[14].



5.4. CUARTA PRUEBA: PRUEBA DE FRIEDMAN EN BEJANI Y GHATEE [13] USANDO LA METODOLOGÍA PROPUESTA POR FERREIRA ET AL. [10]

Objetivo: Establecer un ranking entre los clasificadores utilizados en la experimentación del dataset propuesto por Bejani y Ghattee [13], de acuerdo a la métrica AUC usando la metodología propuesta por Ferreira et al. [10]. Determinar si existe diferencia significativa entre ellos con un 0.5% de nivel de significancia.

Clasificadores utilizados en la experimentación del dataset de Bejani y Ghattee [13]: MHLDForest, Random Forest, Rotation Forest, PBC4cip, Bagging, J48, MLP y GNB(sklearn).

Nota: Para la realización de esta prueba se utilizaron los datos expuestos en la subsección 4.2.5, para tener claridad cómo se generan y seleccionan por favor lea las respuestas a las preguntas:

- ¿Qué conjuntos de datos se generan?
- ¿Qué datos se seleccionan para realizar la prueba estadística?

Hipótesis:

H_0 : No existe diferencia entre los clasificadores utilizados en la experimentación del dataset de Bejani y Ghattee [13] en la métrica AUC para eventos agresivos

H_1 : Existe diferencia entre los clasificadores utilizados en la experimentación del dataset de Bejani y Ghattee [13] en la métrica AUC para eventos agresivos.

A continuación, se muestran los rangos promedio obtenidos al aplicar el procedimiento de Friedman:

Tabla 12. Rango promedio de los clasificadores al aplicar Friedman en [36]

| Clasificador | Ranking |
|-----------------|---------|
| MHLDForest | 1.2292 |
| Rotation Forest | 1.8125 |
| Random Forest | 3 |
| PBC4cip 4 | 3.9583 |
| Bagging | 5 |
| MLP | 6.0625 |
| J48 | 6.9583 |



| | |
|--------------|--------|
| GNB(sklearn) | 7.9792 |
|--------------|--------|

Estadística de Friedman considerando el rendimiento de reducción (distribuido según chi-cuadrado con 7 grados de libertad: 165.215278.

P-value calculado por la prueba de Friedman: 8.145017993399506E-11.

Nivel de significancia: $\alpha = 0.05$

Decisión: Como P-value < 0.05 se rechaza la hipótesis nula H_0 y aceptamos la hipótesis alternativa de que si existe diferencia significativa entre los clasificadores utilizados en la experimentación del dataset de Bejani y Ghatte [13] en la métrica AUC para eventos agresivos de conducción.

A continuación, se muestra el Diagrama de Diferencias Críticas (DDC) calculado, con una diferencia crítica de 2.1431550938559023 con la prueba de Nemenyi.

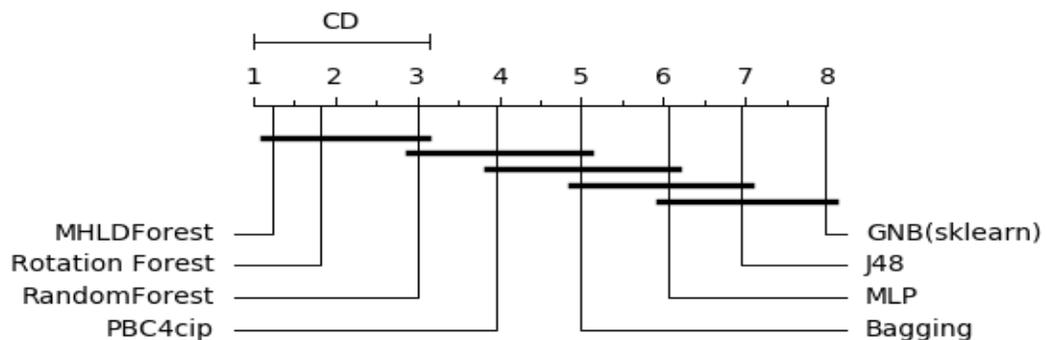


Figura 35. Diagrama de diferencias críticas del AUC en [36] usando la metodología de [10]

Como se puede observar hay presencia de subconjuntos homogéneos de medias en los clasificadores que no tienen diferencia significativa entre sí como, por ejemplo: MHLDForest vs Rotation Forest, MHLDForest vs Random Forest, Rotation Forest vs Random Forest, Rotation Forest vs PBC4cip, Random Forest vs PBC4cip, Random Forest vs Bagging, PBC4cip vs Bagging, PBC4cip vs MLP, Bagging vs MLP, Bagging vs J48, MLP vs GNB(sklearn), MLP vs J48 y J48 vs GNB(sklearn). Los datos no son suficientes para concluir si el rendimiento de la métrica AUC de eventos agresivos es mejor en el clasificador MHLDForest que en Rotation Forest, pasa lo mismo con Rotation Forest y Random Forest y así sucesivamente. Para clasificar las diferencias en el rendimiento de AUC de cada par de clasificadores mencionados anteriormente por cada conjunto de datos aplicamos la prueba



de rango de signos de Wilcoxon[43]. A continuación, se muestra un resumen de los resultados obtenidos en dicha prueba:

Tabla 13. Resultados de la prueba de rango de signo de Wilcoxon en [36]

| Clasificador | $\alpha = 0.9$ | | $\alpha = 0.95$ | |
|-----------------|----------------|-------|-----------------|-------|
| | + | \pm | + | \pm |
| MHLDForest | 7 | 7 | 7 | 7 |
| Rotation Forest | 6 | 6 | 6 | 6 |
| Random Forest | 5 | 5 | 5 | 5 |
| PBC4cip | 4 | 4 | 4 | 4 |
| Bagging | 3 | 3 | 3 | 3 |
| MLP | 2 | 2 | 2 | 2 |
| J48 | 1 | 1 | 1 | 1 |
| GNB(sklearn) | 0 | 0 | 0 | 0 |

Como se puede apreciar los mejores clasificadores según el rendimiento de la métrica AUC son MHLDForest y Rotation Forest en ese orden, como en la prueba estadística 5.1 realizada al dataset de Ferreira et al. [10]. En esta prueba los tres primeros clasificadores no tienen diferencia significativa entre ellos, pero si existe una diferencia estadísticamente significativa entre MHLDForest y PBC4cip que viene siendo el cuarto clasificador en el ranking.

5.5. QUINTA PRUEBA: PRUEBA DE FRIEDMAN EN BEJANI Y GHATEE [13] USANDO LA METODOLOGÍA BOW

Objetivo: Establecer un ranking entre los clasificadores utilizados en la experimentación del dataset propuesto por Bejani y Ghatee [13], de acuerdo a la métrica AUC usando la metodología BoW. Determinar si existe diferencia significativa entre ellos con un 0.5% de nivel de significancia.

Clasificadores utilizados en la experimentación del dataset de Bejani y Ghatee [13]: MHLDForest, Random Forest, Rotation Forest, PBC4cip, Bagging, J48, MLP(sklearn) y GNB(sklearn).

Nota: Para la realización de esta prueba se utilizaron los datos expuestos en la subsección 4.2.6, para tener claridad cómo se generan y seleccionan por favor lea las respuestas a las preguntas:



- ¿Qué conjuntos de datos se generan?
- ¿Qué datos se seleccionan para realizar la prueba estadística?

Hipótesis:

H_0 : No existe diferencia entre los clasificadores utilizados en la experimentación del dataset de Bejani y Ghatte [13] en la métrica AUC para eventos agresivos de conducción.

H_1 : Existe diferencia entre los clasificadores utilizados en la experimentación del dataset de Bejani y Ghatte [13] en la métrica AUC para eventos agresivos de conducción.

A continuación, se muestran los rangos promedio obtenidos al aplicar el procedimiento de Friedman:

Tabla 14. Rango promedio de los clasificadores al aplicar Friedman en [36] usando BoW

| Clasificador | Ranking |
|---------------------|----------------|
| MLP(sklearn) | 1 |
| Bagging | 2.625 |
| GNB(sklearn) | 2.8125 |
| Random | 3.625 |
| Rotation | 5.125 |
| PBC4cip | 5.8125 |
| J48 | 7 |
| MHLDForest | 8 |

Estadística de Friedman considerando el rendimiento de reducción (distribuido según chi-cuadrado con 7 grados de libertad: 53.322917.

P-value calculado por la prueba de Friedman: 3.265383741180017E-9.

Nivel de significancia: $\alpha = 0.05$

Decisión: Como P-value < 0.05 se rechaza la hipótesis nula H_0 y aceptamos la hipótesis alternativa de que si existe diferencia significativa entre los clasificadores utilizados en la experimentación del dataset de Ferreira et al. [10] en la métrica AUC para eventos agresivos.

A continuación, se muestra el Diagrama de Diferencias Críticas (DDC) calculado, con una diferencia crítica de 3.712053511058468 con la prueba de Nemenyi.

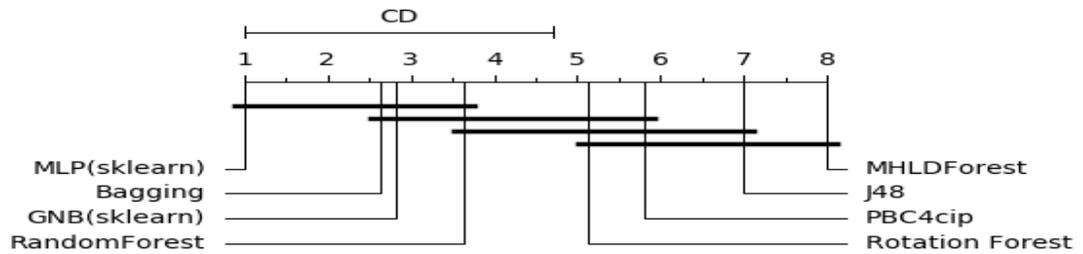


Figura 36. Diagrama de diferencias críticas del AUC en [13] usando BoW

Como se puede observar existe diferencia significativa entre MLP(sklearn) vs MHLDForest, MLP(sklearn) vs J48, MLP(sklearn) vs PBC4cip, MLP(sklearn) vs Rotation Forest, Bagging vs J48, Bagging vs MHLDForest, GNB(sklearn) vs J48, GNB(sklearn) vs MHLDForest y Random Forest vs MHLDForest. Entre el resto de los clasificadores no existe diferencia significativa. Los datos no son suficientes para comparar el resto de los clasificadores de eventos agresivos entre sí, ya que no hay diferencias significativas. Para clasificar las diferencias en el rendimiento de AUC de cada par de clasificadores restantes aplicaremos la prueba de rango de signos de Wilcoxon[43]. A continuación, se muestra un resumen de los resultados obtenidos en dicha prueba:

Tabla 15. Resultados de la prueba de rango de signo de Wilcoxon en [36] usando BoW

| Clasificador | $\alpha = 0.9$ | | $\alpha = 0.95$ | |
|-----------------|----------------|---|-----------------|---|
| | + | ± | + | ± |
| MLP | 7 | 7 | 7 | 7 |
| Bagging | 5 | 6 | 5 | 6 |
| GNB(sklearn) | 4 | 6 | 4 | 6 |
| Random Forest | 4 | 5 | 4 | 5 |
| Rotation Forest | 3 | 3 | 3 | 3 |
| PBC4cip | 2 | 2 | 2 | 2 |
| J48 | 1 | 1 | 1 | 1 |
| MHLDForest | 0 | 0 | 0 | 0 |

Como se puede observar los clasificadores de mejor puntuación son MLP(sklearn) y Bagging para la metodología BoW aplicada al dataset propuesto por Bejani y Ghatee [13]. El clasificador MHLDForest que es el primero en el ranking en la aplicación de la metodología propuesta por Ferreira et al. [10] en las pruebas realizadas en la subsección 5.1 y 5.4 , en esta prueba es el último en el ranking.



5.6. CONCLUSIONES DEL CAPÍTULO

Como resultado del capítulo obtuvimos las siguientes configuraciones por cada conjunto de datos, siendo los valores de la métrica AUC más altos que se alcanzaron en toda la actividad de experimentación.

Tabla 16. Mejores resultados obtenidos en el AUC en la actividad de experimentación

| Dataset | Extracción de características | Parametrización | Clasificador | AUC |
|--------------------------------|---|--|--------------|-------------|
| Ferreira et al. [10] (2017) | BoW <i>M. Ricardo et al. [14] (2019)</i> | K, L = 10, 5 Acelerómetro Filtro: ADLF | MLP(sklearn) | 99.9 |
| Bejani & Ghatee [13] (2018) | Valores estadísticos <i>Ferreira et al. [10] (2017)</i> | nf = 11 Magnetómetro | MHLDForest | 99.3 |
| (M. Ricardo et al. [14] (2019) | Valores estadísticos <i>Ferreira et al. [10] (2017)</i> | nf = 2 Acelerómetro | MHLDForest | 98.3 |

Como resultado de las pruebas estadísticas los clasificadores que quedaron mejor en el ranking fueron: MLP(sklearn) al aplicar BoW y MHLDForest en combinación con la representación basada en valores estadísticos.



CAPÍTULO 6: DISCUSIÓN Y CONCLUSIONES GENERALES

A modo de conclusión podemos decir que los mejores resultados según la métrica AUC en el dataset de Ferreira et al. [10], fueron obtenidos con la aplicación de la metodología BoW junto a la aplicación del filtro Análisis Discriminante Lineal de Fisher con un 99.8% y 99.9%. En este escenario de acuerdo a las pruebas estadísticas, los clasificadores de mejor puntuación son MLP(sklearn) y Random Forest. No se obtuvieron malos resultados en este dataset aplicando su propia metodología, donde también se obtuvieron resultados hasta un AUC de 99.7%. Los clasificadores de mejores resultados según las pruebas estadísticas fueron MHLDForest y Rotation Forest en ese orden. Estos dos clasificadores superan a Random Forest que es el mejor clasificador reportado en Ferreira et al. [10]. Un factor que ayudó a mejorar la puntuación de la metodología propuesta por Ferreira et al. [10] fue el aumento del tamaño de la ventana deslizante hasta $nf=15$. Con los resultados reportados en este trabajo, se supera el estado del arte en la métrica AUC (Weighted. Avg) con el uso de nuevos clasificadores y el aumento de tamaño de la ventana deslizante.

El uso del algoritmo MODLEM logró un 100% en la clasificación de todos los eventos agresivos el dataset de Ferreira et al. [10] en los 69 eventos, no siendo así para el dataset propuesto por M. Ricardo et al. [14] en 478 eventos donde no superó un AUC del 91%. En el dataset propuesto por Bejani y Ghatee [13] este algoritmo no superó un AUC del 83%. Para más información ver el Anexo 7.

Los mejores resultados según la métrica AUC en el dataset propuesto por M. Ricardo et al. [14], fueron alcanzados con el uso de la metodología Ferreira et al. [10] una ventana deslizante de tamaño $nf = 2$, leyendo del sensor acelerómetro y utilizando como clasificador MHLDTForest se obtuvo un AUC de 98.3%. El valor de AUC más alto que se alcanzó con la metodología BoW en este mismo dataset fue de 97.1%, esto fue usando como clasificador una red neuronal MLP(sklearn). Según las pruebas estadísticas los clasificadores de mayor puntuación con el uso de la metodología BoW en este dataset son MLP(sklearn) y Random Forest.

Los mejores resultados según la métrica AUC en el dataset propuesto por Bejani y Ghatee [13] fueron alcanzados con la metodología de Ferreira et al.[10], con una ventana



deslizante de tamaño $nf = 11$, leyendo del sensor magnetómetro y utilizando como clasificador MHLDTForest se obtuvo un AUC de 99.3%. De acuerdo a las pruebas estadísticas los mejores clasificadores en la metodología de Ferreira et al. [10], en este dataset son MHLDTForest y Rotation Forest . El valor de AUC más alto que se alcanzó con la metodología BoW en este mismo dataset fue de 99.1%, esto fue usando como clasificador una red neuronal MLP(sklearn). Según las pruebas estadísticas los clasificadores de mayor puntuación con el uso de la metodología BoW en este dataset son MLP(sklearn) y Bagging.

Al aplicar el ADLF a los datos en brutos se reduce la dimensionalidad de los datos. Cuando se aplicó el algoritmo MODLEM sobre esta transformación se obtuvieron mejores resultados en la métrica AUC. Se comportó de igual manera al aplicar dicha técnica ADLF sobre la representación BoW. Después de esto se repitió la actividad de clasificación nuevamente obteniendo resultados superiores a los alcanzados por BoW inicialmente. No pasó así al aplicar el ADLF sobre la representación de valores estadísticos propuesta por Ferreira et al. [10].

Con el desarrollo de este trabajo se ha logrado demostrar que ambas metodologías BoW y la propuesta por Ferreira et al. [10], obtienen resultados superiores al 98% de AUC frente al problema de clasificación multiclase de eventos agresivos en los tres datasets con los que se experimentó.

6.1. DISCUSIÓN DE LA HIPÓTESIS

“La extracción de características mediante el cálculo de valores estadísticos con ventanas deslizantes de tamaño $2 < nf < 15$ mejora los resultados de clasificación en eventos seguros/agresivos”.

Con los resultados alcanzados en este trabajo de investigación podemos afirmar que al aplicar metodología propuesta por Ferreira et al. [10] con tamaños de ventana deslizantes nf entre 2 y 14 se obtuvo un comportamiento creciente en la actividad de clasificación seguro/ agresiva de eventos de conducción. A medida que aumenta el tamaño nf mejora el valor de la métrica AUC para los clasificadores de vanguardia reportados en este trabajo.



6.2. TRABAJOS FUTUROS

En trabajos futuros sería interesante seguir validando los resultados alcanzados en este trabajo de investigación en un número mayor de conjuntos de datos.

Integrar los enfoques de análisis de conducción para diseñar una solución sistémica que incluya, modelado, identificación de patrones y sistemas de puntuación por conductores agresivos.

6.3. RECOMENDACIONES

Se recomienda aplicar los resultados obtenidos en este trabajo de investigación, en una aplicación móvil, almacenando los datos en la nube.

Investigar las soluciones que proponen modelos difusos para clasificar el comportamiento del conductor en términos de conducción agresiva y segura.



CAPÍTULO 7: APÉNDICES (ANEXOS)

Las siguientes tablas contienen la parametrización y los resultados de los clasificadores que se probaron en la experimentación de los tres conjuntos de datos propuestos en la subsección 2.6. A cada clasificador se le realizó una validación cruzada de 10 folds usando la herramienta Weka, los resultados que se reportan están expresados en Weighted Avg. Los clasificadores que tienen signo de interrogación en algunas de las métricas seleccionadas, es porque la métrica no se pudo calcular debido a que el clasificador identificó cero objetos en alguna de las clases del conjunto de datos.

ANEXO 1. Resultados de la metodología de Ferreira et al. [10] en su conjunto de datos.

Tabla 17. Parámetros evaluados para optimizar los clasificadores en Ferreira et al. [10]

| Clasificador | Parámetros | Valores |
|-----------------|-------------------------------|--|
| Bagging | # of iterations | 200, 100, 10 |
| GNB(sklearn) | Default parameters in sklearn | - |
| J48 | Default parameters in Weka | - |
| MLP | hidden_layer_sizes | (#attr. + #classes)/2, 40, 30, 20, 10 |
| MHLDTForest | # trees | 200, 100 |
| PBC4cip | # tress | 200, 100 |
| Random Forest | # of iterations | 200, 100 |
| | # of attributes to randomly | $\log_2(\# \text{ predictores} + 1, 10, 15)$ |
| Rotation Forest | # of iterations | 200, 100, 10 |

Tabla 18. Resultados de clasificadores en el dataset de Ferreira et al. [10] con su metodología

| Clasificadores | nf | TP Rate | FP Rate | Precision | Recall | F-1 | MCC | AUC | PRC |
|-----------------------------|----|---------|---------|-----------|--------|-------|-------|--------------|-------|
| Acelerómetro, nf = 2 | | | | | | | | | |
| MHLDForest | 2 | 0.756 | 0.053 | 0.756 | 0.756 | 0.749 | 0.705 | 0.940 | 0.807 |
| Random Forest | 2 | 0.732 | 0.058 | 0.731 | 0.732 | 0.723 | 0.674 | 0.937 | 0.800 |
| Rotation Forest | 2 | 0.735 | 0.057 | 0.733 | 0.735 | 0.726 | 0.677 | 0.941 | 0.814 |
| PBC4cip | 2 | 0.723 | 0.049 | 0.739 | 0.723 | 0.725 | 0.677 | 0.932 | 0.770 |
| Bagging | 2 | 0.672 | 0.072 | 0.666 | 0.672 | 0.655 | 0.596 | 0.914 | 0.741 |
| J48 | 2 | 0.596 | 0.079 | 0.595 | 0.596 | 0.595 | 0.517 | 0.776 | 0.464 |
| MLP | 2 | 0.690 | 0.066 | 0.696 | 0.690 | 0.688 | 0.628 | 0.886 | 0.670 |
| GNB(sklearn) | 2 | 0.491 | 0.098 | 0.482 | 0.491 | 0.479 | 0.387 | 0.836 | 0.496 |
| Acelerómetro, nf = 4 | | | | | | | | | |
| MHLDForest | 4 | 0.825 | 0.038 | 0.830 | 0.825 | 0.824 | 0.791 | 0.965 | 0.884 |



| | | | | | | | | | |
|-----------------------------|-----------|----------------|----------------|------------------|---------------|------------|------------|--------------|------------|
| Random Forest | 4 | 0.793 | 0.044 | 0.798 | 0.793 | 0.791 | 0.753 | 0.960 | 0.861 |
| Rotation Forest | 4 | 0.817 | 0.040 | 0.823 | 0.817 | 0.816 | 0.781 | 0.962 | 0.883 |
| PBC4cip | 4 | 0.742 | 0.042 | 0.767 | 0.742 | 0.747 | 0.704 | 0.953 | 0.843 |
| Bagging | 4 | 0.729 | 0.058 | 0.735 | 0.729 | 0.727 | 0.675 | 0.939 | 0.793 |
| J48 | 4 | 0.578 | 0.085 | 0.575 | 0.578 | 0.574 | 0.492 | 0.776 | 0.458 |
| MLP | 4 | 0.736 | 0.055 | 0.738 | 0.736 | 0.733 | 0.683 | 0.924 | 0.785 |
| GNB(sklearn) | 4 | 0.503 | 0.093 | 0.507 | 0.503 | 0.498 | 0.406 | 0.828 | 0.518 |
| Acelerómetro, nf = 6 | | | | | | | | | |
| MHLDForest | 6 | 0.851 | 0.033 | 0.857 | 0.851 | 0.851 | 0.823 | 0.977 | 0.922 |
| Random Forest | 6 | 0.846 | 0.035 | 0.855 | 0.846 | 0.847 | 0.817 | 0.972 | 0.907 |
| Clasificadores | nf | TP Rate | FP Rate | Precision | Recall | F-1 | MCC | AUC | PRC |
| Rotation Forest | 6 | 0.845 | 0.035 | 0.852 | 0.845 | 0.845 | 0.815 | 0.971 | 0.910 |
| PBC4cip | 6 | 0.777 | 0.037 | 0.799 | 0.777 | 0.780 | 0.744 | 0.965 | 0.880 |
| Bagging | 6 | 0.774 | 0.049 | 0.779 | 0.774 | 0.774 | 0.729 | 0.954 | 0.846 |
| J48 | 6 | 0.640 | 0.073 | 0.640 | 0.640 | 0.639 | 0.567 | 0.802 | 0.514 |
| MLP | 6 | 0.823 | 0.038 | 0.828 | 0.823 | 0.823 | 0.788 | 0.946 | 0.855 |
| GNB(sklearn) | 6 | 0.507 | 0.099 | 0.513 | 0.507 | 0.502 | 0.409 | 0.808 | 0.513 |
| Acelerómetro, nf = 7 | | | | | | | | | |
| MHLDForest | 7 | 0.860 | 0.031 | 0.868 | 0.860 | 0.861 | 0.834 | 0.978 | 0.928 |
| Random Forest | 7 | 0.860 | 0.031 | 0.868 | 0.860 | 0.861 | 0.834 | 0.974 | 0.912 |
| Rotation Forest | 7 | 0.872 | 0.029 | 0.878 | 0.872 | 0.873 | 0.848 | 0.975 | 0.925 |
| PBC4cip | 7 | 0.789 | 0.034 | 0.813 | 0.789 | 0.791 | 0.759 | 0.965 | 0.884 |
| Bagging | 7 | 0.776 | 0.048 | 0.783 | 0.776 | 0.777 | 0.732 | 0.957 | 0.859 |
| J48 | 7 | 0.686 | 0.063 | 0.690 | 0.686 | 0.687 | 0.625 | 0.836 | 0.581 |
| MLP | 7 | 0.832 | 0.035 | 0.834 | 0.832 | 0.832 | 0.798 | 0.950 | 0.882 |
| GNB(sklearn) | 7 | 0.522 | 0.096 | 0.527 | 0.522 | 0.517 | 0.426 | 0.805 | 0.507 |
| Acelerómetro, nf = 8 | | | | | | | | | |
| MHLDForest | 8 | 0.878 | 0.027 | 0.884 | 0.878 | 0.879 | 0.855 | 0.982 | 0.943 |
| Random Forest | 8 | 0.865 | 0.030 | 0.874 | 0.865 | 0.867 | 0.841 | 0.980 | 0.930 |
| Rotation Forest | 8 | 0.883 | 0.026 | 0.889 | 0.883 | 0.884 | 0.861 | 0.978 | 0.936 |
| PBC4cip | 8 | 0.789 | 0.032 | 0.816 | 0.789 | 0.794 | 0.762 | 0.971 | 0.904 |
| Bagging | 8 | 0.792 | 0.045 | 0.802 | 0.792 | 0.793 | 0.752 | 0.961 | 0.876 |
| J48 | 8 | 0.694 | 0.061 | 0.695 | 0.694 | 0.694 | 0.634 | 0.840 | 0.593 |
| MLP | 8 | 0.857 | 0.030 | 0.859 | 0.857 | 0.857 | 0.829 | 0.964 | 0.902 |
| GNB(sklearn) | 8 | 0.512 | 0.099 | 0.512 | 0.512 | 0.507 | 0.412 | 0.795 | 0.490 |
| Acelerómetro, nf = 9 | | | | | | | | | |
| MHLDForest | 9 | 0.900 | 0.022 | 0.904 | 0.900 | 0.900 | 0.881 | 0.986 | 0.953 |
| Random Forest | 9 | 0.873 | 0.028 | 0.879 | 0.873 | 0.874 | 0.849 | 0.982 | 0.936 |
| Rotation Forest | 9 | 0.881 | 0.026 | 0.885 | 0.881 | 0.881 | 0.857 | 0.980 | 0.943 |
| PBC4cip | 9 | 0.797 | 0.030 | 0.830 | 0.797 | 0.802 | 0.773 | 0.973 | 0.908 |
| Bagging | 9 | 0.804 | 0.042 | 0.811 | 0.804 | 0.805 | 0.766 | 0.965 | 0.886 |
| J48 | 9 | 0.739 | 0.052 | 0.741 | 0.739 | 0.739 | 0.687 | 0.860 | 0.637 |



| | | | | | | | | | |
|------------------------------|-----------|----------------|----------------|------------------|---------------|------------|------------|--------------|------------|
| MLP | 9 | 0.867 | 0.028 | 0.871 | 0.867 | 0.867 | 0.841 | 0.963 | 0.908 |
| GNB(sklearn) | 9 | 0.517 | 0.098 | 0.517 | 0.517 | 0.513 | 0.418 | 0.797 | 0.484 |
| Acelerómetro. nf = 10 | | | | | | | | | |
| MHLDForest | 10 | 0.899 | 0.022 | 0.903 | 0.899 | 0.900 | 0.880 | 0.989 | 0.961 |
| Random Forest | 10 | 0.885 | 0.025 | 0.890 | 0.885 | 0.886 | 0.863 | 0.985 | 0.947 |
| Rotation Forest | 10 | 0.892 | 0.024 | 0.896 | 0.892 | 0.892 | 0.871 | 0.984 | 0.952 |
| PBC4cip | 10 | 0.816 | 0.029 | 0.840 | 0.816 | 0.820 | 0.791 | 0.976 | 0.915 |
| Bagging | 10 | 0.803 | 0.041 | 0.806 | 0.803 | 0.803 | 0.764 | 0.968 | 0.893 |
| J48 | 10 | 0.716 | 0.057 | 0.715 | 0.716 | 0.715 | 0.658 | 0.848 | 0.608 |
| MLP | 10 | 0.885 | 0.025 | 0.887 | 0.885 | 0.885 | 0.862 | 0.974 | 0.933 |
| GNB(sklearn) | 10 | 0.523 | 0.097 | 0.521 | 0.523 | 0.519 | 0.424 | 0.793 | 0.490 |
| Clasificadores | nf | TP Rate | FP Rate | Precision | Recall | F-1 | MCC | AUC | PRC |
| Acelerómetro. nf = 11 | | | | | | | | | |
| MHLDForest | 11 | 0.919 | 0.018 | 0.922 | 0.919 | 0.919 | 0.903 | 0.990 | 0.965 |
| Random Forest | 11 | 0.892 | 0.024 | 0.897 | 0.892 | 0.893 | 0.871 | 0.987 | 0.952 |
| Rotation Forest | 11 | 0.902 | 0.022 | 0.905 | 0.902 | 0.902 | 0.883 | 0.984 | 0.954 |
| PBC4cip | 11 | 0.836 | 0.025 | 0.862 | 0.836 | 0.842 | 0.816 | 0.978 | 0.923 |
| Bagging | 11 | 0.815 | 0.038 | 0.820 | 0.815 | 0.815 | 0.779 | 0.971 | 0.899 |
| J48 | 11 | 0.759 | 0.050 | 0.760 | 0.759 | 0.759 | 0.710 | 0.871 | 0.667 |
| MLP | 11 | 0.889 | 0.023 | 0.892 | 0.889 | 0.889 | 0.867 | 0.970 | 0.934 |
| GNB(sklearn) | 11 | 0.526 | 0.097 | 0.523 | 0.526 | 0.522 | 0.427 | 0.795 | 0.486 |
| Acelerómetro. nf = 12 | | | | | | | | | |
| MHLDForest | 12 | 0.920 | 0.018 | 0.924 | 0.920 | 0.920 | 0.905 | 0.991 | 0.968 |
| Random Forest | 12 | 0.899 | 0.022 | 0.904 | 0.899 | 0.900 | 0.879 | 0.989 | 0.960 |
| Rotation Forest | 12 | 0.902 | 0.021 | 0.905 | 0.902 | 0.902 | 0.883 | 0.987 | 0.959 |
| PBC4cip | 12 | 0.831 | 0.026 | 0.857 | 0.831 | 0.836 | 0.810 | 0.979 | 0.923 |
| Bagging | 12 | 0.800 | 0.042 | 0.806 | 0.800 | 0.800 | 0.761 | 0.973 | 0.904 |
| J48 | 12 | 0.733 | 0.055 | 0.735 | 0.733 | 0.733 | 0.679 | 0.853 | 0.618 |
| MLP | 12 | 0.899 | 0.022 | 0.901 | 0.899 | 0.898 | 0.878 | 0.968 | 0.929 |
| GNB(sklearn) | 12 | 0.515 | 0.099 | 0.518 | 0.515 | 0.515 | 0.417 | 0.782 | 0.477 |
| Acelerómetro. nf = 13 | | | | | | | | | |
| MHLDForest | 13 | 0.923 | 0.017 | 0.926 | 0.923 | 0.923 | 0.908 | 0.993 | 0.974 |
| Random Forest | 13 | 0.906 | 0.021 | 0.909 | 0.906 | 0.906 | 0.887 | 0.990 | 0.965 |
| Rotation Forest | 13 | 0.917 | 0.018 | 0.918 | 0.917 | 0.916 | 0.9 | 0.988 | 0.965 |
| PBC4cip | 13 | 0.846 | 0.025 | 0.864 | 0.846 | 0.848 | 0.824 | 0.980 | 0.930 |
| Bagging | 13 | 0.821 | 0.038 | 0.825 | 0.821 | 0.821 | 0.786 | 0.974 | 0.911 |
| J48 | 13 | 0.768 | 0.049 | 0.770 | 0.768 | 0.768 | 0.721 | 0.873 | 0.680 |
| MLP | 13 | 0.901 | 0.021 | 0.903 | 0.901 | 0.900 | 0.881 | 0.971 | 0.936 |
| GNB(sklearn) | 13 | 0.512 | 0.100 | 0.515 | 0.512 | 0.511 | 0.413 | 0.790 | 0.488 |
| Acelerómetro. nf = 14 | | | | | | | | | |
| MHLDForest | 14 | 0.930 | 0.016 | 0.934 | 0.930 | 0.930 | 0.917 | 0.994 | 0.977 |
| Random Forest | 14 | 0.901 | 0.022 | 0.907 | 0.901 | 0.902 | 0.883 | 0.991 | 0.967 |



| | | | | | | | | | |
|------------------------------------|----|---------|---------|------------|--------|-------|-------|--------------|-------|
| Rotation Forest | 14 | 0.919 | 0.018 | 0.921 | 0.919 | 0.919 | 0.903 | 0.991 | 0.970 |
| PBC4cip | 14 | 0.849 | 0.025 | 0.864 | 0.849 | 0.850 | 0.826 | 0.982 | 0.935 |
| Bagging | 14 | 0.821 | 0.039 | 0.828 | 0.821 | 0.821 | 0.786 | 0.977 | 0.917 |
| J48 | 14 | 0.769 | 0.048 | 0.771 | 0.769 | 0.770 | 0.723 | 0.874 | 0.659 |
| MLP | 14 | 0.901 | 0.021 | 0.903 | 0.901 | 0.901 | 0.881 | 0.973 | 0.944 |
| GNB(sklearn) | 14 | 0.499 | 0.104 | 0.502 | 0.499 | 0.498 | 0.396 | 0.777 | 0.470 |
| Acelerómetro. nf = 15 | | | | | | | | | |
| MHLDForest | 15 | 0.935 | 0.014 | 0.937 | 0.935 | 0.935 | 0.923 | 0.994 | 0.979 |
| Random Forest | 15 | 0.911 | 0.020 | 0.914 | 0.911 | 0.911 | 0.893 | 0.993 | 0.971 |
| Rotation Forest | 15 | 0.922 | 0.017 | 0.923 | 0.922 | 0.921 | 0.906 | 0.991 | 0.971 |
| PBC4cip | 15 | 0.863 | 0.023 | 0.876 | 0.863 | 0.864 | 0.842 | 0.985 | 0.943 |
| Bagging | 15 | 0.825 | 0.038 | 0.831 | 0.825 | 0.825 | 0.791 | 0.979 | 0.925 |
| Clasificadores | nf | TP Rate | FP Rate | Preci-sion | Recall | F-1 | MCC | AUC | PRC |
| J48 | 15 | 0.767 | 0.049 | 0.768 | 0.767 | 0.767 | 0.719 | 0.872 | 0.660 |
| MLP | 15 | 0.901 | 0.021 | 0.901 | 0.901 | 0.900 | 0.88 | 0.973 | 0.945 |
| GNB(sklearn) | 15 | 0.51 | 0.102 | 0.512 | 0.510 | 0.508 | 0.409 | 0.786 | 0.494 |
| Acelerómetro-Lineal. nf = 2 | | | | | | | | | |
| MHLDForest | 2 | 0.753 | 0.054 | 0.753 | 0.753 | 0.744 | 0.700 | 0.938 | 0.800 |
| Random Forest | 2 | 0.747 | 0.057 | 0.761 | 0.747 | 0.739 | 0.696 | 0.936 | 0.799 |
| Rotation Forest | 2 | 0.750 | 0.055 | 0.754 | 0.750 | 0.740 | 0.696 | 0.943 | 0.824 |
| PBC4cip | 2 | 0.702 | 0.049 | 0.723 | 0.702 | 0.707 | 0.657 | 0.931 | 0.767 |
| Bagging | 2 | 0.675 | 0.071 | 0.668 | 0.675 | 0.658 | 0.599 | 0.915 | 0.743 |
| J48 | 2 | 0.614 | 0.075 | 0.613 | 0.614 | 0.613 | 0.539 | 0.778 | 0.469 |
| MLP | 2 | 0.687 | 0.065 | 0.687 | 0.687 | 0.684 | 0.623 | 0.885 | 0.664 |
| GNB(sklearn) | 2 | 0.500 | 0.097 | 0.490 | 0.500 | 0.484 | 0.395 | 0.820 | 0.480 |
| Acelerómetro-Lineal. nf = 4 | | | | | | | | | |
| MHLDForest | 4 | 0.817 | 0.040 | 0.824 | 0.817 | 0.815 | 0.782 | 0.964 | 0.879 |
| Random Forest | 4 | 0.789 | 0.045 | 0.793 | 0.789 | 0.787 | 0.747 | 0.958 | 0.858 |
| Rotation Forest | 4 | 0.817 | 0.040 | 0.823 | 0.817 | 0.816 | 0.781 | 0.962 | 0.881 |
| PBC4cip | 4 | 0.740 | 0.040 | 0.772 | 0.740 | 0.747 | 0.707 | 0.951 | 0.834 |
| Bagging | 4 | 0.731 | 0.057 | 0.737 | 0.731 | 0.729 | 0.678 | 0.938 | 0.790 |
| J48 | 4 | 0.578 | 0.083 | 0.577 | 0.578 | 0.576 | 0.494 | 0.761 | 0.435 |
| MLP | 4 | 0.742 | 0.053 | 0.746 | 0.742 | 0.740 | 0.691 | 0.926 | 0.784 |
| GNB(sklearn) | 4 | 0.450 | 0.102 | 0.444 | 0.450 | 0.435 | 0.338 | 0.793 | 0.460 |
| Acelerómetro-Lineal. nf = 6 | | | | | | | | | |
| MHLDForest | 6 | 0.845 | 0.034 | 0.850 | 0.845 | 0.844 | 0.815 | 0.975 | 0.915 |
| Random Forest | 6 | 0.846 | 0.034 | 0.853 | 0.846 | 0.847 | 0.817 | 0.971 | 0.903 |
| Rotation Forest | 6 | 0.861 | 0.031 | 0.868 | 0.861 | 0.861 | 0.835 | 0.971 | 0.911 |
| PBC4cip | 6 | 0.779 | 0.036 | 0.803 | 0.779 | 0.782 | 0.747 | 0.964 | 0.877 |
| Bagging | 6 | 0.772 | 0.049 | 0.778 | 0.772 | 0.772 | 0.727 | 0.953 | 0.843 |
| J48 | 6 | 0.664 | 0.066 | 0.665 | 0.664 | 0.664 | 0.598 | 0.809 | 0.529 |
| MLP | 6 | 0.812 | 0.039 | 0.815 | 0.812 | 0.811 | 0.775 | 0.944 | 0.851 |



| | | | | | | | | | |
|-------------------------------------|-----------|----------------|----------------|-------------------|---------------|------------|------------|--------------|------------|
| GNB(sklearn) | 6 | 0.473 | 0.101 | 0.477 | 0.473 | 0.464 | 0.368 | 0.783 | 0.471 |
| Acelerómetro-Lineal, nf = 7 | | | | | | | | | |
| MHLDForest | 7 | 0.863 | 0.031 | 0.871 | 0.863 | 0.864 | 0.838 | 0.976 | 0.922 |
| Random Forest | 7 | 0.850 | 0.033 | 0.857 | 0.850 | 0.850 | 0.821 | 0.974 | 0.91 |
| Rotation Forest | 7 | 0.860 | 0.031 | 0.867 | 0.860 | 0.861 | 0.834 | 0.974 | 0.923 |
| PBC4cip | 7 | 0.783 | 0.034 | 0.812 | 0.783 | 0.788 | 0.755 | 0.966 | 0.882 |
| Bagging | 7 | 0.776 | 0.048 | 0.783 | 0.776 | 0.777 | 0.732 | 0.957 | 0.859 |
| J48 | 7 | 0.688 | 0.062 | 0.692 | 0.688 | 0.689 | 0.628 | 0.829 | 0.568 |
| MLP | 7 | 0.841 | 0.033 | 0.844 | 0.841 | 0.840 | 0.810 | 0.950 | 0.864 |
| GNB(sklearn) | 7 | 0.486 | 0.101 | 0.484 | 0.486 | 0.475 | 0.380 | 0.782 | 0.469 |
| Acelerómetro-Lineal, nf = 8 | | | | | | | | | |
| MHLDForest | 8 | 0.872 | 0.029 | 0.879 | 0.872 | 0.874 | 0.848 | 0.983 | 0.942 |
| Random Forest | 8 | 0.867 | 0.030 | 0.873 | 0.867 | 0.868 | 0.841 | 0.979 | 0.926 |
| Clasificadores | nf | TP Rate | FP Rate | Preci-sion | Recall | F-1 | MCC | AUC | PRC |
| Rotation Forest | 8 | 0.883 | 0.026 | 0.889 | 0.883 | 0.884 | 0.861 | 0.980 | 0.935 |
| PBC4cip | 8 | 0.781 | 0.034 | 0.815 | 0.781 | 0.787 | 0.755 | 0.971 | 0.899 |
| Bagging | 8 | 0.793 | 0.045 | 0.802 | 0.793 | 0.794 | 0.754 | 0.959 | 0.871 |
| J48 | 8 | 0.686 | 0.063 | 0.688 | 0.686 | 0.686 | 0.624 | 0.829 | 0.563 |
| MLP | 8 | 0.863 | 0.029 | 0.866 | 0.863 | 0.863 | 0.836 | 0.967 | 0.903 |
| GNB(sklearn) | 8 | 0.497 | 0.099 | 0.499 | 0.497 | 0.489 | 0.396 | 0.780 | 0.470 |
| Acelerómetro-Lineal, nf = 9 | | | | | | | | | |
| MHLDForest | 9 | 0.892 | 0.024 | 0.898 | 0.892 | 0.893 | 0.872 | 0.986 | 0.951 |
| Random Forest | 9 | 0.863 | 0.030 | 0.868 | 0.863 | 0.864 | 0.837 | 0.981 | 0.934 |
| Rotation Forest | 9 | 0.885 | 0.026 | 0.889 | 0.885 | 0.885 | 0.862 | 0.981 | 0.943 |
| PBC4cip | 9 | 0.806 | 0.029 | 0.839 | 0.806 | 0.812 | 0.784 | 0.973 | 0.905 |
| Bagging | 9 | 0.807 | 0.041 | 0.814 | 0.807 | 0.808 | 0.77 | 0.963 | 0.883 |
| J48 | 9 | 0.708 | 0.059 | 0.709 | 0.708 | 0.708 | 0.649 | 0.840 | 0.588 |
| MLP | 9 | 0.871 | 0.027 | 0.875 | 0.871 | 0.871 | 0.846 | 0.965 | 0.916 |
| GNB(sklearn) | 9 | 0.502 | 0.098 | 0.499 | 0.502 | 0.493 | 0.400 | 0.780 | 0.459 |
| Acelerómetro-Lineal, nf = 10 | | | | | | | | | |
| MHLDForest | 10 | 0.906 | 0.021 | 0.910 | 0.906 | 0.907 | 0.888 | 0.988 | 0.957 |
| Random Forest | 10 | 0.878 | 0.027 | 0.882 | 0.878 | 0.879 | 0.854 | 0.984 | 0.945 |
| Rotation Forest | 10 | 0.897 | 0.023 | 0.900 | 0.897 | 0.897 | 0.876 | 0.986 | 0.953 |
| PBC4cip | 10 | 0.816 | 0.028 | 0.846 | 0.816 | 0.821 | 0.794 | 0.977 | 0.917 |
| Bagging | 10 | 0.790 | 0.044 | 0.796 | 0.79 | 0.791 | 0.749 | 0.966 | 0.888 |
| J48 | 10 | 0.726 | 0.056 | 0.727 | 0.726 | 0.726 | 0.670 | 0.847 | 0.607 |
| MLP | 10 | 0.873 | 0.027 | 0.876 | 0.873 | 0.873 | 0.848 | 0.971 | 0.929 |
| GNB(sklearn) | 10 | 0.510 | 0.097 | 0.509 | 0.510 | 0.504 | 0.410 | 0.773 | 0.456 |
| Acelerómetro-Lineal, nf = 11 | | | | | | | | | |
| MHLDForest | 11 | 0.919 | 0.018 | 0.923 | 0.919 | 0.919 | 0.903 | 0.989 | 0.964 |
| Random Forest | 11 | 0.896 | 0.023 | 0.901 | 0.896 | 0.897 | 0.876 | 0.987 | 0.952 |
| Rotation Forest | 11 | 0.903 | 0.022 | 0.907 | 0.903 | 0.903 | 0.884 | 0.987 | 0.958 |



| | | | | | | | | | |
|-------------------------------------|-----------|----------------|----------------|-------------------|---------------|------------|------------|--------------|------------|
| PBC4cip | 11 | 0.826 | 0.027 | 0.851 | 0.826 | 0.831 | 0.804 | 0.979 | 0.923 |
| Bagging | 11 | 0.812 | 0.039 | 0.817 | 0.812 | 0.812 | 0.775 | 0.971 | 0.899 |
| J48 | 11 | 0.747 | 0.052 | 0.748 | 0.747 | 0.747 | 0.696 | 0.858 | 0.636 |
| MLP | 11 | 0.889 | 0.024 | 0.891 | 0.889 | 0.889 | 0.867 | 0.972 | 0.937 |
| GNB(sklearn) | 11 | 0.503 | 0.099 | 0.500 | 0.503 | 0.496 | 0.400 | 0.774 | 0.454 |
| Acelerómetro-Lineal, nf = 12 | | | | | | | | | |
| MHLDForest | 12 | 0.913 | 0.019 | 0.918 | 0.913 | 0.914 | 0.897 | 0.991 | 0.969 |
| Random Forest | 12 | 0.895 | 0.023 | 0.900 | 0.895 | 0.895 | 0.875 | 0.989 | 0.961 |
| Rotation Forest | 12 | 0.907 | 0.021 | 0.910 | 0.907 | 0.908 | 0.889 | 0.989 | 0.960 |
| PBC4cip | 12 | 0.831 | 0.026 | 0.859 | 0.831 | 0.837 | 0.811 | 0.978 | 0.921 |
| Bagging | 12 | 0.807 | 0.040 | 0.813 | 0.807 | 0.807 | 0.770 | 0.973 | 0.905 |
| J48 | 12 | 0.730 | 0.056 | 0.731 | 0.730 | 0.730 | 0.675 | 0.844 | 0.587 |
| MLP | 12 | 0.886 | 0.024 | 0.889 | 0.886 | 0.886 | 0.863 | 0.972 | 0.928 |
| GNB(sklearn) | 12 | 0.508 | 0.099 | 0.506 | 0.508 | 0.502 | 0.406 | 0.764 | 0.440 |
| Clasificadores | nf | TP Rate | FP Rate | Preci-sion | Recall | F-1 | MCC | AUC | PRC |
| Acelerómetro-Lineal, nf = 13 | | | | | | | | | |
| MHLDForest | 13 | 0.925 | 0.017 | 0.927 | 0.925 | 0.925 | 0.910 | 0.993 | 0.974 |
| Random Forest | 13 | 0.903 | 0.022 | 0.906 | 0.903 | 0.903 | 0.884 | 0.990 | 0.964 |
| Rotation Forest | 13 | 0.912 | 0.020 | 0.915 | 0.912 | 0.912 | 0.894 | 0.991 | 0.970 |
| PBC4cip | 13 | 0.843 | 0.026 | 0.859 | 0.843 | 0.845 | 0.820 | 0.981 | 0.933 |
| Bagging | 13 | 0.818 | 0.039 | 0.822 | 0.818 | 0.818 | 0.782 | 0.974 | 0.911 |
| J48 | 13 | 0.761 | 0.050 | 0.763 | 0.761 | 0.761 | 0.712 | 0.865 | 0.657 |
| MLP | 13 | 0.913 | 0.019 | 0.915 | 0.913 | 0.912 | 0.895 | 0.972 | 0.940 |
| GNB(sklearn) | 13 | 0.490 | 0.103 | 0.489 | 0.490 | 0.484 | 0.385 | 0.764 | 0.446 |
| Acelerómetro-Lineal, nf = 14 | | | | | | | | | |
| MHLDForest | 14 | 0.923 | 0.017 | 0.927 | 0.923 | 0.923 | 0.909 | 0.993 | 0.977 |
| Random Forest | 14 | 0.906 | 0.021 | 0.910 | 0.906 | 0.906 | 0.888 | 0.991 | 0.967 |
| Rotation Forest | 14 | 0.912 | 0.020 | 0.915 | 0.912 | 0.912 | 0.895 | 0.992 | 0.973 |
| PBC4cip | 14 | 0.840 | 0.027 | 0.853 | 0.840 | 0.841 | 0.814 | 0.982 | 0.937 |
| Bagging | 14 | 0.820 | 0.039 | 0.827 | 0.820 | 0.820 | 0.785 | 0.977 | 0.918 |
| J48 | 14 | 0.764 | 0.050 | 0.767 | 0.764 | 0.764 | 0.716 | 0.873 | 0.665 |
| MLP | 14 | 0.902 | 0.021 | 0.904 | 0.902 | 0.902 | 0.882 | 0.970 | 0.937 |
| GNB(sklearn) | 14 | 0.494 | 0.102 | 0.493 | 0.494 | 0.490 | 0.390 | 0.760 | 0.444 |
| Acelerómetro-Lineal, nf = 15 | | | | | | | | | |
| MHLDForest | 15 | 0.937 | 0.014 | 0.939 | 0.937 | 0.937 | 0.925 | 0.995 | 0.981 |
| Random Forest | 15 | 0.914 | 0.019 | 0.917 | 0.914 | 0.914 | 0.898 | 0.993 | 0.972 |
| Rotation Forest | 15 | 0.924 | 0.017 | 0.927 | 0.924 | 0.924 | 0.910 | 0.993 | 0.976 |
| PBC4cip | 15 | 0.863 | 0.024 | 0.874 | 0.863 | 0.864 | 0.841 | 0.985 | 0.945 |
| Bagging | 15 | 0.829 | 0.037 | 0.834 | 0.829 | 0.829 | 0.795 | 0.979 | 0.925 |
| J48 | 15 | 0.771 | 0.048 | 0.773 | 0.771 | 0.772 | 0.724 | 0.875 | 0.667 |
| MLP | 15 | 0.906 | 0.020 | 0.907 | 0.906 | 0.906 | 0.887 | 0.976 | 0.950 |
| GNB(sklearn) | 15 | 0.493 | 0.102 | 0.493 | 0.493 | 0.489 | 0.389 | 0.765 | 0.452 |



| Acelerómetro-Giroscopio-Combinado. nf = 2 | | | | | | | | | |
|--|-----------|----------------|----------------|-------------------|---------------|------------|------------|--------------|------------|
| MHLDForest | 2 | 0.729 | 0.059 | 0.725 | 0.729 | 0.719 | 0.67 | 0.936 | 0.787 |
| Random Forest | 2 | 0.705 | 0.065 | 0.705 | 0.705 | 0.692 | 0.640 | 0.930 | 0.768 |
| Rotation Forest | 2 | 0.709 | 0.065 | 0.712 | 0.709 | 0.701 | 0.646 | 0.932 | 0.792 |
| PBC4cip | 2 | 0.691 | 0.054 | 0.703 | 0.691 | 0.691 | 0.638 | 0.927 | 0.750 |
| Bagging | 2 | 0.64 | 0.079 | 0.635 | 0.64 | 0.626 | 0.559 | 0.909 | 0.708 |
| J48 | 2 | 0.566 | 0.089 | 0.566 | 0.566 | 0.565 | 0.479 | 0.779 | 0.483 |
| MLP | 2 | 0.538 | 0.101 | 0.526 | 0.538 | 0.526 | 0.434 | 0.829 | 0.56 |
| GNB(sklearn) | 2 | 0.291 | 0.167 | 0.362 | 0.291 | 0.283 | 0.154 | 0.596 | 0.261 |
| Acelerómetro-Giroscopio-Combinado. nf = 4 | | | | | | | | | |
| MHLDForest | 4 | 0.789 | 0.046 | 0.794 | 0.789 | 0.788 | 0.747 | 0.962 | 0.864 |
| Random Forest | 4 | 0.764 | 0.05 | 0.766 | 0.764 | 0.762 | 0.717 | 0.955 | 0.844 |
| Rotation Forest | 4 | 0.795 | 0.043 | 0.797 | 0.795 | 0.793 | 0.754 | 0.958 | 0.855 |
| PBC4cip | 4 | 0.732 | 0.044 | 0.749 | 0.732 | 0.734 | 0.689 | 0.950 | 0.823 |
| Bagging | 4 | 0.715 | 0.060 | 0.715 | 0.715 | 0.712 | 0.657 | 0.937 | 0.787 |
| Clasificadores | nf | TP Rate | FP Rate | Preci-sion | Recall | F-1 | MCC | AUC | PRC |
| J48 | 4 | 0.602 | 0.078 | 0.602 | 0.602 | 0.602 | 0.525 | 0.790 | 0.479 |
| MLP | 4 | 0.610 | 0.080 | 0.615 | 0.610 | 0.606 | 0.533 | 0.872 | 0.641 |
| GNB(sklearn) | 4 | 0.295 | 0.153 | 0.311 | 0.295 | 0.269 | 0.152 | 0.592 | 0.237 |
| Acelerómetro-Giroscopio-Combinado. nf = 6 | | | | | | | | | |
| MHLDForest | 6 | 0.856 | 0.031 | 0.859 | 0.856 | 0.856 | 0.828 | 0.976 | 0.914 |
| Random Forest | 6 | 0.826 | 0.037 | 0.829 | 0.826 | 0.826 | 0.792 | 0.969 | 0.894 |
| Rotation Forest | 6 | 0.838 | 0.035 | 0.841 | 0.838 | 0.838 | 0.806 | 0.973 | 0.903 |
| PBC4cip | 6 | 0.788 | 0.035 | 0.800 | 0.788 | 0.789 | 0.753 | 0.964 | 0.873 |
| Bagging | 6 | 0.750 | 0.053 | 0.755 | 0.750 | 0.749 | 0.700 | 0.950 | 0.833 |
| J48 | 6 | 0.661 | 0.069 | 0.661 | 0.661 | 0.661 | 0.593 | 0.812 | 0.536 |
| MLP | 6 | 0.626 | 0.079 | 0.637 | 0.626 | 0.628 | 0.554 | 0.885 | 0.68 |
| GNB(sklearn) | 6 | 0.288 | 0.154 | 0.283 | 0.288 | 0.252 | 0.137 | 0.603 | 0.239 |
| Acelerómetro-Giroscopio-Combinado. nf = 7 | | | | | | | | | |
| MHLDForest | 7 | 0.860 | 0.030 | 0.862 | 0.860 | 0.859 | 0.832 | 0.980 | 0.928 |
| Random Forest | 7 | 0.843 | 0.034 | 0.847 | 0.843 | 0.843 | 0.812 | 0.974 | 0.903 |
| Rotation Forest | 7 | 0.852 | 0.032 | 0.855 | 0.852 | 0.852 | 0.823 | 0.978 | 0.917 |
| PBC4cip | 7 | 0.790 | 0.035 | 0.801 | 0.790 | 0.790 | 0.754 | 0.967 | 0.887 |
| Bagging | 7 | 0.773 | 0.048 | 0.779 | 0.773 | 0.774 | 0.730 | 0.954 | 0.844 |
| J48 | 7 | 0.692 | 0.061 | 0.693 | 0.692 | 0.692 | 0.631 | 0.834 | 0.582 |
| MLP | 7 | 0.636 | 0.076 | 0.643 | 0.636 | 0.636 | 0.565 | 0.891 | 0.695 |
| GNB(sklearn) | 7 | 0.295 | 0.151 | 0.299 | 0.295 | 0.266 | 0.150 | 0.607 | 0.246 |
| Acelerómetro-Giroscopio-Combinado. nf = 8 | | | | | | | | | |
| MHLDForest | 8 | 0.889 | 0.024 | 0.892 | 0.889 | 0.889 | 0.867 | 0.985 | 0.941 |
| Random Forest | 8 | 0.857 | 0.031 | 0.859 | 0.857 | 0.857 | 0.828 | 0.978 | 0.919 |
| Rotation Forest | 8 | 0.862 | 0.030 | 0.865 | 0.862 | 0.862 | 0.835 | 0.981 | 0.930 |
| PBC4cip | 8 | 0.801 | 0.031 | 0.820 | 0.801 | 0.803 | 0.772 | 0.973 | 0.900 |



| | | | | | | | | | |
|---|-----------|----------------|----------------|-------------------|---------------|------------|------------|--------------|------------|
| Bagging | 8 | 0.783 | 0.045 | 0.788 | 0.783 | 0.784 | 0.741 | 0.959 | 0.859 |
| J48 | 8 | 0.685 | 0.062 | 0.685 | 0.685 | 0.685 | 0.623 | 0.832 | 0.577 |
| MLP | 8 | 0.649 | 0.071 | 0.662 | 0.649 | 0.650 | 0.584 | 0.899 | 0.720 |
| GNB(sklearn) | 8 | 0.298 | 0.151 | 0.322 | 0.298 | 0.283 | 0.161 | 0.614 | 0.255 |
| Acelerómetro-Giroscopio-Combinado, nf = 9 | | | | | | | | | |
| MHLDForest | 9 | 0.893 | 0.023 | 0.896 | 0.893 | 0.893 | 0.872 | 0.986 | 0.947 |
| Random Forest | 9 | 0.872 | 0.027 | 0.875 | 0.872 | 0.872 | 0.847 | 0.980 | 0.930 |
| Rotation Forest | 9 | 0.881 | 0.025 | 0.883 | 0.881 | 0.881 | 0.857 | 0.983 | 0.940 |
| PBC4cip | 9 | 0.814 | 0.029 | 0.830 | 0.814 | 0.816 | 0.785 | 0.974 | 0.905 |
| Bagging | 9 | 0.792 | 0.042 | 0.796 | 0.792 | 0.793 | 0.752 | 0.962 | 0.871 |
| J48 | 9 | 0.698 | 0.060 | 0.699 | 0.698 | 0.699 | 0.639 | 0.844 | 0.598 |
| MLP | 9 | 0.683 | 0.063 | 0.692 | 0.683 | 0.683 | 0.623 | 0.902 | 0.737 |
| GNB(sklearn) | 9 | 0.290 | 0.149 | 0.321 | 0.290 | 0.287 | 0.157 | 0.613 | 0.254 |
| Acelerómetro-Giroscopio-Combinado, nf = 10 | | | | | | | | | |
| MHLDForest | 10 | 0.909 | 0.02 | 0.911 | 0.909 | 0.91 | 0.891 | 0.990 | 0.961 |
| Random Forest | 10 | 0.886 | 0.025 | 0.889 | 0.886 | 0.886 | 0.863 | 0.986 | 0.945 |
| Clasificadores | nf | TP Rate | FP Rate | Preci-sion | Recall | F-1 | MCC | AUC | PRC |
| Rotation Forest | 10 | 0.891 | 0.024 | 0.893 | 0.891 | 0.891 | 0.869 | 0.987 | 0.951 |
| PBC4cip | 10 | 0.829 | 0.027 | 0.843 | 0.829 | 0.831 | 0.803 | 0.979 | 0.920 |
| Bagging | 10 | 0.799 | 0.042 | 0.802 | 0.799 | 0.799 | 0.759 | 0.967 | 0.885 |
| J48 | 10 | 0.701 | 0.060 | 0.702 | 0.701 | 0.701 | 0.641 | 0.835 | 0.583 |
| MLP | 10 | 0.671 | 0.065 | 0.676 | 0.671 | 0.67 | 0.607 | 0.905 | 0.740 |
| GNB(sklearn) | 10 | 0.307 | 0.145 | 0.327 | 0.307 | 0.298 | 0.172 | 0.615 | 0.257 |
| Acelerómetro-Giroscopio-Combinado, nf = 11 | | | | | | | | | |
| MHLDForest | 11 | 0.919 | 0.018 | 0.922 | 0.919 | 0.92 | 0.903 | 0.990 | 0.963 |
| Random Forest | 11 | 0.891 | 0.024 | 0.893 | 0.891 | 0.891 | 0.869 | 0.987 | 0.952 |
| Rotation Forest | 11 | 0.907 | 0.020 | 0.909 | 0.907 | 0.907 | 0.888 | 0.988 | 0.956 |
| PBC4cip | 11 | 0.831 | 0.027 | 0.846 | 0.831 | 0.833 | 0.805 | 0.981 | 0.925 |
| Bagging | 11 | 0.803 | 0.041 | 0.808 | 0.803 | 0.804 | 0.765 | 0.970 | 0.896 |
| J48 | 11 | 0.741 | 0.053 | 0.744 | 0.741 | 0.742 | 0.689 | 0.865 | 0.648 |
| MLP | 11 | 0.675 | 0.066 | 0.676 | 0.675 | 0.674 | 0.609 | 0.908 | 0.744 |
| GNB(sklearn) | 11 | 0.318 | 0.140 | 0.336 | 0.318 | 0.306 | 0.187 | 0.621 | 0.268 |
| Acelerómetro-Giroscopio-Combinado, nf = 12 | | | | | | | | | |
| MHLDForest | 12 | 0.922 | 0.017 | 0.923 | 0.922 | 0.922 | 0.906 | 0.992 | 0.971 |
| Random Forest | 12 | 0.898 | 0.022 | 0.900 | 0.898 | 0.899 | 0.878 | 0.989 | 0.957 |
| Rotation Forest | 12 | 0.898 | 0.022 | 0.900 | 0.898 | 0.899 | 0.878 | 0.990 | 0.962 |
| PBC4cip | 12 | 0.814 | 0.029 | 0.830 | 0.814 | 0.816 | 0.785 | 0.974 | 0.905 |
| Bagging | 12 | 0.799 | 0.042 | 0.802 | 0.799 | 0.799 | 0.759 | 0.967 | 0.885 |
| J48 | 12 | 0.750 | 0.051 | 0.750 | 0.750 | 0.750 | 0.699 | 0.863 | 0.648 |
| MLP | 12 | 0.678 | 0.062 | 0.683 | 0.678 | 0.679 | 0.616 | 0.913 | 0.752 |
| GNB(sklearn) | 12 | 0.308 | 0.140 | 0.322 | 0.308 | 0.291 | 0.174 | 0.617 | 0.263 |
| Acelerómetro-Giroscopio-Combinado, nf = 13 | | | | | | | | | |



| | | | | | | | | | |
|---|----|---------|---------|------------|--------|-------|-------|--------------|-------|
| MHLDForest | 13 | 0.930 | 0.015 | 0.932 | 0.930 | 0.931 | 0.917 | 0.993 | 0.973 |
| Random Forest | 13 | 0.904 | 0.021 | 0.906 | 0.904 | 0.905 | 0.885 | 0.991 | 0.965 |
| Rotation Forest | 13 | 0.911 | 0.019 | 0.913 | 0.911 | 0.911 | 0.893 | 0.991 | 0.967 |
| PBC4cip | 13 | 0.862 | 0.023 | 0.870 | 0.862 | 0.862 | 0.839 | 0.984 | 0.940 |
| Bagging | 13 | 0.825 | 0.036 | 0.827 | 0.825 | 0.825 | 0.790 | 0.974 | 0.906 |
| J48 | 13 | 0.749 | 0.053 | 0.751 | 0.749 | 0.749 | 0.698 | 0.867 | 0.658 |
| MLP | 13 | 0.660 | 0.073 | 0.670 | 0.660 | 0.662 | 0.594 | 0.892 | 0.730 |
| GNB(sklearn) | 13 | 0.314 | 0.135 | 0.324 | 0.314 | 0.293 | 0.181 | 0.621 | 0.271 |
| Acelerómetro-Giroscopio-Combinado, nf = 14 | | | | | | | | | |
| MHLDForest | 14 | 0.945 | 0.012 | 0.946 | 0.945 | 0.945 | 0.934 | 0.995 | 0.981 |
| Random Forest | 14 | 0.921 | 0.017 | 0.923 | 0.921 | 0.921 | 0.906 | 0.993 | 0.971 |
| Rotation Forest | 14 | 0.927 | 0.016 | 0.929 | 0.927 | 0.928 | 0.913 | 0.993 | 0.975 |
| PBC4cip | 14 | 0.877 | 0.020 | 0.885 | 0.877 | 0.878 | 0.857 | 0.987 | 0.952 |
| Bagging | 14 | 0.836 | 0.034 | 0.838 | 0.836 | 0.836 | 0.803 | 0.978 | 0.920 |
| J48 | 14 | 0.759 | 0.049 | 0.760 | 0.759 | 0.759 | 0.710 | 0.875 | 0.667 |
| MLP | 14 | 0.656 | 0.072 | 0.676 | 0.656 | 0.657 | 0.594 | 0.894 | 0.736 |
| GNB(sklearn) | 14 | 0.306 | 0.138 | 0.310 | 0.306 | 0.281 | 0.169 | 0.615 | 0.259 |
| Clasificadores | nf | TP Rate | FP Rate | Preci-sion | Recall | F-1 | MCC | AUC | PRC |
| Acelerómetro-Giroscopio-Combinado, nf = 15 | | | | | | | | | |
| MHLDForest | 15 | 0.942 | 0.012 | 0.943 | 0.942 | 0.942 | 0.931 | 0.995 | 0.982 |
| Random Forest | 15 | 0.920 | 0.017 | 0.922 | 0.920 | 0.920 | 0.905 | 0.993 | 0.975 |
| Rotation Forest | 15 | 0.932 | 0.015 | 0.933 | 0.932 | 0.932 | 0.919 | 0.994 | 0.977 |
| PBC4cip | 15 | 0.886 | 0.018 | 0.893 | 0.886 | 0.886 | 0.867 | 0.988 | 0.954 |
| Bagging | 15 | 0.841 | 0.034 | 0.843 | 0.841 | 0.842 | 0.809 | 0.980 | 0.927 |
| J48 | 15 | 0.775 | 0.047 | 0.776 | 0.775 | 0.775 | 0.729 | 0.878 | 0.676 |
| MLP | 15 | 0.636 | 0.074 | 0.643 | 0.636 | 0.633 | 0.564 | 0.877 | 0.706 |
| GNB(sklearn) | 15 | 0.301 | 0.139 | 0.305 | 0.301 | 0.275 | 0.163 | 0.617 | 0.260 |
| Giroscopio, nf = 2 | | | | | | | | | |
| MHLDForest | 2 | 0.732 | 0.060 | 0.735 | 0.732 | 0.725 | 0.675 | 0.930 | 0.789 |
| Random Forest | 2 | 0.720 | 0.061 | 0.718 | 0.720 | 0.710 | 0.660 | 0.927 | 0.775 |
| Rotation Forest | 2 | 0.708 | 0.067 | 0.732 | 0.708 | 0.702 | 0.651 | 0.934 | 0.803 |
| PBC4cip | 2 | 0.690 | 0.058 | 0.696 | 0.690 | 0.680 | 0.630 | 0.924 | 0.765 |
| Bagging | 2 | 0.654 | 0.077 | 0.648 | 0.654 | 0.635 | 0.574 | 0.912 | 0.730 |
| J48 | 2 | 0.587 | 0.080 | 0.602 | 0.587 | 0.592 | 0.512 | 0.794 | 0.516 |
| MLP | 2 | 0.657 | 0.070 | 0.652 | 0.657 | 0.651 | 0.585 | 0.884 | 0.689 |
| GNB(sklearn) | 2 | 0.542 | 0.080 | 0.591 | 0.542 | 0.550 | 0.476 | 0.833 | 0.540 |
| Giroscopio, nf = 4 | | | | | | | | | |
| MHLDForest | 4 | 0.776 | 0.048 | 0.779 | 0.776 | 0.774 | 0.731 | 0.957 | 0.851 |
| Random Forest | 4 | 0.751 | 0.054 | 0.754 | 0.751 | 0.747 | 0.700 | 0.947 | 0.824 |
| Rotation Forest | 4 | 0.761 | 0.050 | 0.764 | 0.761 | 0.758 | 0.713 | 0.959 | 0.849 |
| PBC4cip | 4 | 0.721 | 0.047 | 0.736 | 0.721 | 0.714 | 0.672 | 0.946 | 0.820 |
| Bagging | 4 | 0.695 | 0.064 | 0.696 | 0.695 | 0.692 | 0.633 | 0.925 | 0.762 |



| | | | | | | | | | |
|----------------------------|----|---------|---------|------------|--------|-------|-------|--------------|-------|
| J48 | 4 | 0.627 | 0.072 | 0.626 | 0.627 | 0.626 | 0.554 | 0.800 | 0.496 |
| MLP | 4 | 0.674 | 0.065 | 0.674 | 0.674 | 0.672 | 0.608 | 0.906 | 0.711 |
| GNB(sklearn) | 4 | 0.465 | 0.096 | 0.504 | 0.465 | 0.460 | 0.376 | 0.778 | 0.451 |
| Giroscopio, nf = 6 | | | | | | | | | |
| MHLDForest | 6 | 0.836 | 0.036 | 0.841 | 0.836 | 0.836 | 0.804 | 0.974 | 0.905 |
| Random Forest | 6 | 0.810 | 0.039 | 0.810 | 0.810 | 0.808 | 0.771 | 0.968 | 0.883 |
| Rotation Forest | 6 | 0.821 | 0.037 | 0.824 | 0.821 | 0.820 | 0.786 | 0.973 | 0.900 |
| PBC4cip | 6 | 0.765 | 0.041 | 0.773 | 0.765 | 0.757 | 0.721 | 0.962 | 0.868 |
| Bagging | 6 | 0.757 | 0.049 | 0.758 | 0.757 | 0.754 | 0.709 | 0.945 | 0.818 |
| J48 | 6 | 0.666 | 0.067 | 0.663 | 0.666 | 0.664 | 0.598 | 0.818 | 0.544 |
| MLP | 6 | 0.752 | 0.048 | 0.753 | 0.752 | 0.751 | 0.704 | 0.930 | 0.790 |
| GNB(sklearn) | 6 | 0.489 | 0.096 | 0.511 | 0.489 | 0.470 | 0.394 | 0.767 | 0.450 |
| Giroscopio, nf = 7 | | | | | | | | | |
| MHLDForest | 7 | 0.842 | 0.034 | 0.845 | 0.842 | 0.842 | 0.811 | 0.978 | 0.919 |
| Random Forest | 7 | 0.842 | 0.033 | 0.845 | 0.842 | 0.842 | 0.812 | 0.974 | 0.906 |
| Rotation Forest | 7 | 0.845 | 0.033 | 0.848 | 0.845 | 0.845 | 0.815 | 0.978 | 0.917 |
| PBC4cip | 7 | 0.782 | 0.038 | 0.794 | 0.782 | 0.777 | 0.743 | 0.968 | 0.885 |
| Bagging | 7 | 0.768 | 0.047 | 0.771 | 0.768 | 0.767 | 0.722 | 0.952 | 0.833 |
| Clasificadores | nf | TP Rate | FP Rate | Preci-sion | Recall | F-1 | MCC | AUC | PRC |
| J48 | 7 | 0.691 | 0.059 | 0.689 | 0.691 | 0.689 | 0.63 | 0.837 | 0.589 |
| MLP | 7 | 0.777 | 0.044 | 0.782 | 0.777 | 0.779 | 0.735 | 0.940 | 0.807 |
| GNB(sklearn) | 7 | 0.484 | 0.098 | 0.494 | 0.484 | 0.465 | 0.383 | 0.779 | 0.472 |
| Giroscopio, nf = 8 | | | | | | | | | |
| MHLDForest | 8 | 0.876 | 0.026 | 0.880 | 0.876 | 0.876 | 0.853 | 0.984 | 0.935 |
| Random Forest | 8 | 0.848 | 0.032 | 0.851 | 0.848 | 0.848 | 0.818 | 0.979 | 0.917 |
| Rotation Forest | 8 | 0.856 | 0.030 | 0.861 | 0.856 | 0.856 | 0.828 | 0.982 | 0.928 |
| PBC4cip | 8 | 0.789 | 0.034 | 0.808 | 0.789 | 0.786 | 0.756 | 0.974 | 0.898 |
| Bagging | 8 | 0.768 | 0.045 | 0.770 | 0.768 | 0.767 | 0.723 | 0.957 | 0.849 |
| J48 | 8 | 0.694 | 0.06 | 0.693 | 0.694 | 0.693 | 0.633 | 0.842 | 0.598 |
| MLP | 8 | 0.785 | 0.044 | 0.788 | 0.785 | 0.784 | 0.743 | 0.945 | 0.819 |
| GNB(sklearn) | 8 | 0.493 | 0.097 | 0.491 | 0.493 | 0.471 | 0.389 | 0.786 | 0.464 |
| Giroscopio, nf = 9 | | | | | | | | | |
| MHLDForest | 9 | 0.892 | 0.022 | 0.895 | 0.892 | 0.892 | 0.872 | 0.988 | 0.952 |
| Random Forest | 9 | 0.875 | 0.026 | 0.877 | 0.875 | 0.875 | 0.850 | 0.983 | 0.932 |
| Rotation Forest | 9 | 0.887 | 0.024 | 0.891 | 0.887 | 0.887 | 0.866 | 0.985 | 0.940 |
| PBC4cip | 9 | 0.809 | 0.032 | 0.824 | 0.809 | 0.806 | 0.778 | 0.980 | 0.916 |
| Bagging | 9 | 0.793 | 0.040 | 0.795 | 0.793 | 0.792 | 0.754 | 0.966 | 0.870 |
| J48 | 9 | 0.706 | 0.058 | 0.710 | 0.706 | 0.707 | 0.650 | 0.845 | 0.603 |
| MLP | 9 | 0.802 | 0.039 | 0.805 | 0.802 | 0.803 | 0.764 | 0.948 | 0.831 |
| GNB(sklearn) | 9 | 0.509 | 0.094 | 0.513 | 0.509 | 0.489 | 0.411 | 0.791 | 0.468 |
| Giroscopio, nf = 10 | | | | | | | | | |
| MHLDForest | 10 | 0.909 | 0.019 | 0.913 | 0.909 | 0.909 | 0.892 | 0.992 | 0.964 |



| | | | | | | | | | |
|----------------------------|----|---------|---------|------------|--------|-------|-------|--------------|-------|
| Random Forest | 10 | 0.882 | 0.024 | 0.884 | 0.882 | 0.882 | 0.859 | 0.987 | 0.947 |
| Rotation Forest | 10 | 0.897 | 0.021 | 0.900 | 0.897 | 0.897 | 0.877 | 0.988 | 0.956 |
| PBC4cip | 10 | 0.841 | 0.027 | 0.855 | 0.841 | 0.838 | 0.814 | 0.983 | 0.930 |
| Bagging | 10 | 0.807 | 0.038 | 0.808 | 0.807 | 0.806 | 0.769 | 0.970 | 0.890 |
| J48 | 10 | 0.740 | 0.053 | 0.74 | 0.74 | 0.739 | 0.687 | 0.864 | 0.64 |
| MLP | 10 | 0.838 | 0.033 | 0.844 | 0.838 | 0.840 | 0.808 | 0.963 | 0.880 |
| GNB(sklearn) | 10 | 0.495 | 0.096 | 0.492 | 0.495 | 0.475 | 0.39 | 0.788 | 0.465 |
| Giroscopio, nf = 11 | | | | | | | | | |
| MHLDForest | 11 | 0.927 | 0.015 | 0.931 | 0.927 | 0.928 | 0.914 | 0.993 | 0.970 |
| Random Forest | 11 | 0.898 | 0.022 | 0.901 | 0.898 | 0.898 | 0.878 | 0.990 | 0.958 |
| Rotation Forest | 11 | 0.902 | 0.021 | 0.906 | 0.902 | 0.903 | 0.883 | 0.989 | 0.960 |
| PBC4cip | 11 | 0.851 | 0.024 | 0.865 | 0.851 | 0.850 | 0.828 | 0.986 | 0.943 |
| Bagging | 11 | 0.827 | 0.034 | 0.829 | 0.827 | 0.827 | 0.793 | 0.973 | 0.901 |
| J48 | 11 | 0.744 | 0.053 | 0.746 | 0.744 | 0.745 | 0.692 | 0.862 | 0.647 |
| MLP | 11 | 0.867 | 0.027 | 0.869 | 0.867 | 0.867 | 0.840 | 0.970 | 0.903 |
| GNB(sklearn) | 11 | 0.521 | 0.093 | 0.52 | 0.521 | 0.501 | 0.422 | 0.806 | 0.492 |
| Giroscopio, nf = 12 | | | | | | | | | |
| MHLDForest | 12 | 0.925 | 0.016 | 0.928 | 0.925 | 0.925 | 0.911 | 0.994 | 0.976 |
| Random Forest | 12 | 0.898 | 0.021 | 0.900 | 0.898 | 0.898 | 0.878 | 0.991 | 0.966 |
| Clasificadores | nf | TP Rate | FP Rate | Preci-sion | Recall | F-1 | MCC | AUC | PRC |
| Rotation Forest | 12 | 0.898 | 0.021 | 0.900 | 0.898 | 0.898 | 0.878 | 0.991 | 0.964 |
| PBC4cip | 12 | 0.847 | 0.025 | 0.861 | 0.847 | 0.846 | 0.822 | 0.987 | 0.947 |
| Bagging | 12 | 0.827 | 0.034 | 0.828 | 0.827 | 0.827 | 0.793 | 0.975 | 0.907 |
| J48 | 12 | 0.758 | 0.05 | 0.761 | 0.758 | 0.759 | 0.710 | 0.870 | 0.649 |
| MLP | 12 | 0.856 | 0.030 | 0.860 | 0.856 | 0.857 | 0.828 | 0.968 | 0.906 |
| GNB(sklearn) | 12 | 0.516 | 0.093 | 0.511 | 0.516 | 0.496 | 0.414 | 0.800 | 0.478 |
| Giroscopio, nf = 13 | | | | | | | | | |
| MHLDForest | 13 | 0.933 | 0.014 | 0.935 | 0.933 | 0.934 | 0.921 | 0.995 | 0.980 |
| Random Forest | 13 | 0.924 | 0.016 | 0.926 | 0.924 | 0.924 | 0.909 | 0.993 | 0.974 |
| Rotation Forest | 13 | 0.906 | 0.020 | 0.909 | 0.906 | 0.906 | 0.887 | 0.994 | 0.975 |
| PBC4cip | 13 | 0.861 | 0.023 | 0.871 | 0.861 | 0.861 | 0.837 | 0.988 | 0.954 |
| Bagging | 13 | 0.827 | 0.035 | 0.829 | 0.827 | 0.827 | 0.793 | 0.977 | 0.915 |
| J48 | 13 | 0.751 | 0.048 | 0.75 | 0.751 | 0.75 | 0.702 | 0.864 | 0.632 |
| MLP | 13 | 0.886 | 0.023 | 0.888 | 0.886 | 0.886 | 0.864 | 0.977 | 0.928 |
| GNB(sklearn) | 13 | 0.518 | 0.092 | 0.519 | 0.518 | 0.5 | 0.42 | 0.801 | 0.469 |
| Giroscopio, nf = 14 | | | | | | | | | |
| MHLDForest | 14 | 0.944 | 0.011 | 0.945 | 0.944 | 0.944 | 0.934 | 0.996 | 0.983 |
| Random Forest | 14 | 0.927 | 0.015 | 0.929 | 0.927 | 0.927 | 0.913 | 0.994 | 0.975 |
| Rotation Forest | 14 | 0.924 | 0.015 | 0.926 | 0.924 | 0.924 | 0.910 | 0.995 | 0.979 |
| PBC4cip | 14 | 0.877 | 0.021 | 0.885 | 0.877 | 0.876 | 0.856 | 0.989 | 0.958 |
| Bagging | 14 | 0.831 | 0.034 | 0.833 | 0.831 | 0.831 | 0.798 | 0.978 | 0.916 |
| J48 | 14 | 0.791 | 0.044 | 0.793 | 0.791 | 0.791 | 0.749 | 0.894 | 0.707 |



| | | | | | | | | | |
|----------------------------|----|-------|-------|-------|-------|-------|-------|--------------|-------|
| MLP | 14 | 0.911 | 0.019 | 0.913 | 0.911 | 0.911 | 0.893 | 0.983 | 0.949 |
| GNB(sklearn) | 14 | 0.507 | 0.095 | 0.503 | 0.507 | 0.487 | 0.404 | 0.800 | 0.462 |
| Giroscopio. nf = 15 | | | | | | | | | |
| MHLDForest | 15 | 0.949 | 0.011 | 0.950 | 0.949 | 0.949 | 0.939 | 0.997 | 0.987 |
| Random Forest | 15 | 0.932 | 0.015 | 0.933 | 0.932 | 0.932 | 0.918 | 0.995 | 0.981 |
| Rotation Forest | 15 | 0.930 | 0.015 | 0.932 | 0.930 | 0.930 | 0.916 | 0.996 | 0.983 |
| PBC4cip | 15 | 0.891 | 0.018 | 0.898 | 0.891 | 0.890 | 0.872 | 0.992 | 0.968 |
| Bagging | 15 | 0.856 | 0.029 | 0.858 | 0.856 | 0.856 | 0.827 | 0.981 | 0.929 |
| J48 | 15 | 0.769 | 0.046 | 0.769 | 0.769 | 0.769 | 0.723 | 0.887 | 0.687 |
| MLP | 15 | 0.916 | 0.017 | 0.917 | 0.916 | 0.916 | 0.899 | 0.980 | 0.943 |
| GNB(sklearn) | 15 | 0.512 | 0.094 | 0.510 | 0.512 | 0.495 | 0.411 | 0.793 | 0.452 |

Resumen de resultados de la métrica AUC

Tabla 19. Resultados de clasificadores del AUC en el dataset de [10] usando su metodología

| AUC | D-1 | D-2 | D-3 | D-4 | D-5 | D-6 | D-7 | D-8 | D-9 | D-10 | D-11 | D-12 |
|--|------|------|------|------|------|------|-------|-------|-------|-------|-------|-------|
| Clasificador | nf=2 | nf=4 | nf=6 | nf=7 | nf=8 | nf=9 | nf=1 | nf=1 | nf=1 | nf=1 | nf=1 | nf=1 |
| Acelerómetro | | | | | | | | | | | | |
| MHLDForest | 0.94 | 0.96 | 0.97 | 0.97 | 0.98 | 0.98 | 0.989 | 0.990 | 0.991 | 0.993 | 0.994 | 0.994 |
| Random | 0.93 | 0.96 | 0.97 | 0.97 | 0.98 | 0.98 | 0.985 | 0.987 | 0.989 | 0.990 | 0.991 | 0.993 |
| Rotation | 0.94 | 0.96 | 0.97 | 0.97 | 0.97 | 0.98 | 0.984 | 0.984 | 0.987 | 0.988 | 0.991 | 0.991 |
| PBC4cip | 0.93 | 0.95 | 0.96 | 0.96 | 0.97 | 0.97 | 0.976 | 0.978 | 0.979 | 0.980 | 0.982 | 0.985 |
| Bagging | 0.91 | 0.93 | 0.95 | 0.95 | 0.96 | 0.96 | 0.968 | 0.971 | 0.973 | 0.974 | 0.977 | 0.979 |
| J48 | 0.77 | 0.77 | 0.80 | 0.83 | 0.84 | 0.86 | 0.848 | 0.871 | 0.853 | 0.873 | 0.873 | 0.872 |
| MLP | 0.88 | 0.92 | 0.94 | 0.95 | 0.96 | 0.96 | 0.974 | 0.970 | 0.968 | 0.971 | 0.970 | 0.973 |
| GNB(sklearn) | 0.83 | 0.82 | 0.80 | 0.80 | 0.79 | 0.79 | 0.793 | 0.795 | 0.782 | 0.790 | 0.760 | 0.786 |
| Acelerómetro-Lineal | | | | | | | | | | | | |
| MHLDForest | 0.93 | 0.96 | 0.97 | 0.97 | 0.98 | 0.98 | 0.988 | 0.989 | 0.991 | 0.993 | 0.993 | 0.995 |
| Random | 0.93 | 0.95 | 0.97 | 0.97 | 0.97 | 0.98 | 0.984 | 0.987 | 0.989 | 0.990 | 0.991 | 0.993 |
| Rotation | 0.94 | 0.96 | 0.97 | 0.97 | 0.98 | 0.98 | 0.986 | 0.987 | 0.989 | 0.991 | 0.992 | 0.993 |
| PBC4cip | 0.93 | 0.95 | 0.96 | 0.96 | 0.97 | 0.97 | 0.977 | 0.979 | 0.978 | 0.981 | 0.982 | 0.985 |
| Bagging | 0.91 | 0.93 | 0.95 | 0.95 | 0.95 | 0.96 | 0.966 | 0.971 | 0.973 | 0.974 | 0.977 | 0.979 |
| J48 | 0.77 | 0.76 | 0.80 | 0.82 | 0.82 | 0.84 | 0.847 | 0.858 | 0.844 | 0.865 | 0.873 | 0.875 |
| MLP | 0.88 | 0.92 | 0.94 | 0.95 | 0.96 | 0.96 | 0.971 | 0.972 | 0.972 | 0.972 | 0.970 | 0.976 |
| GNB(sklearn) | 0.82 | 0.79 | 0.78 | 0.78 | 0.78 | 0.78 | 0.773 | 0.774 | 0.764 | 0.764 | 0.760 | 0.765 |
| Acelerómetro-Giroscopio-Combinado | | | | | | | | | | | | |
| MHLDForest | 0.93 | 0.96 | 0.97 | 0.98 | 0.98 | 0.98 | 0.990 | 0.990 | 0.992 | 0.993 | 0.995 | 0.995 |
| Random | 0.93 | 0.95 | 0.96 | 0.97 | 0.97 | 0.98 | 0.986 | 0.987 | 0.989 | 0.991 | 0.993 | 0.993 |
| Rotation | 0.93 | 0.95 | 0.97 | 0.97 | 0.98 | 0.98 | 0.987 | 0.988 | 0.990 | 0.991 | 0.993 | 0.994 |
| PBC4cip | 0.92 | 0.95 | 0.96 | 0.96 | 0.97 | 0.97 | 0.979 | 0.981 | 0.974 | 0.984 | 0.987 | 0.988 |
| Bagging | 0.90 | 0.93 | 0.95 | 0.95 | 0.95 | 0.96 | 0.967 | 0.970 | 0.967 | 0.974 | 0.978 | 0.980 |
| J48 | 0.77 | 0.79 | 0.81 | 0.83 | 0.83 | 0.84 | 0.835 | 0.865 | 0.863 | 0.867 | 0.875 | 0.878 |
| MLP | 0.82 | 0.87 | 0.88 | 0.89 | 0.89 | 0.90 | 0.905 | 0.908 | 0.913 | 0.892 | 0.894 | 0.877 |



| | | | | | | | | | | | | |
|---------------------|------|------|------|------|------|------|-------|-------|-------|-------|-------|-------|
| GNB(sklearn) | 0.59 | 0.59 | 0.60 | 0.60 | 0.61 | 0.61 | 0.615 | 0.621 | 0.617 | 0.621 | 0.615 | 0.617 |
| Giroscopio | | | | | | | | | | | | |
| MHLDForest | 0.93 | 0.95 | 0.97 | 0.97 | 0.98 | 0.98 | 0.992 | 0.993 | 0.994 | 0.995 | 0.996 | 0.997 |
| Random | 0.92 | 0.94 | 0.96 | 0.97 | 0.97 | 0.98 | 0.987 | 0.990 | 0.991 | 0.993 | 0.994 | 0.995 |
| Rotation | 0.93 | 0.95 | 0.97 | 0.97 | 0.98 | 0.98 | 0.988 | 0.989 | 0.991 | 0.994 | 0.995 | 0.996 |
| PBC4cip | 0.92 | 0.94 | 0.96 | 0.96 | 0.97 | 0.98 | 0.983 | 0.986 | 0.987 | 0.988 | 0.989 | 0.968 |
| Bagging | 0.91 | 0.92 | 0.94 | 0.95 | 0.95 | 0.96 | 0.970 | 0.973 | 0.975 | 0.977 | 0.978 | 0.929 |
| J48 | 0.79 | 0.80 | 0.81 | 0.83 | 0.84 | 0.84 | 0.864 | 0.862 | 0.870 | 0.864 | 0.894 | 0.687 |
| MLP | 0.88 | 0.90 | 0.93 | 0.94 | 0.94 | 0.94 | 0.963 | 0.970 | 0.968 | 0.977 | 0.983 | 0.943 |
| GNB(sklearn) | 0.83 | 0.77 | 0.76 | 0.77 | 0.78 | 0.79 | 0.788 | 0.806 | 0.800 | 0.801 | 0.800 | 0.452 |

Resumen de mejores resultados

Tabla 20. Clasificadores de mejores resultados respecto a la métrica AUC en [10]

| AUC: Area Under Curve | D-1 | D-2 | D-3 | D-4 | D-5 | D-6 | D-7 | D-8 | D-9 | D-10 | D-11 | D-12 |
|--|------|------|------|------|------|------|-----------|-----------|-----------|-----------|-----------|-----------|
| Clasificador | nf=2 | nf=4 | nf=6 | nf=7 | nf=8 | nf=9 | nf=1 0 | nf=1 1 | nf=1 2 | nf=1 3 | nf=1 4 | nf=1 5 |
| Acelerómetro | | | | | | | | | | | | |
| MHLDFores | 0.94 | 0.96 | 0.97 | 0.97 | 0.98 | 0.98 | 0.989 | 0.990 | 0.991 | 0.993 | 0.994 | 0.994 |
| Random | 0.93 | 0.96 | 0.97 | 0.97 | 0.98 | 0.98 | 0.985 | 0.987 | 0.989 | 0.990 | 0.991 | 0.993 |
| Rotation | 0.94 | 0.96 | 0.97 | 0.97 | 0.97 | 0.98 | 0.984 | 0.984 | 0.987 | 0.988 | 0.991 | 0.991 |
| Acelerómetro-Lineal | | | | | | | | | | | | |
| MHLDFores | 0.93 | 0.96 | 0.97 | 0.97 | 0.98 | 0.98 | 0.988 | 0.989 | 0.991 | 0.993 | 0.993 | 0.995 |
| Random | 0.93 | 0.95 | 0.97 | 0.97 | 0.97 | 0.98 | 0.984 | 0.987 | 0.989 | 0.990 | 0.991 | 0.993 |
| Rotation | 0.94 | 0.96 | 0.97 | 0.97 | 0.98 | 0.98 | 0.986 | 0.987 | 0.989 | 0.991 | 0.992 | 0.993 |
| Acelerómetro-Giroscopio-Combinado | | | | | | | | | | | | |
| MHLDFores | 0.93 | 0.96 | 0.97 | 0.98 | 0.98 | 0.98 | 0.990 | 0.990 | 0.992 | 0.993 | 0.995 | 0.995 |
| Random | 0.93 | 0.95 | 0.96 | 0.97 | 0.97 | 0.98 | 0.986 | 0.987 | 0.989 | 0.991 | 0.993 | 0.993 |
| Rotation | 0.93 | 0.95 | 0.97 | 0.97 | 0.98 | 0.98 | 0.987 | 0.988 | 0.990 | 0.991 | 0.993 | 0.994 |
| Giroscopio | | | | | | | | | | | | |
| MHLDFores | 0.93 | 0.95 | 0.97 | 0.97 | 0.98 | 0.98 | 0.992 | 0.993 | 0.994 | 0.995 | 0.996 | 0.997 |
| Random | 0.92 | 0.94 | 0.96 | 0.97 | 0.97 | 0.98 | 0.987 | 0.990 | 0.991 | 0.993 | 0.994 | 0.995 |
| Rotation | 0.93 | 0.95 | 0.97 | 0.97 | 0.98 | 0.98 | 0.988 | 0.989 | 0.991 | 0.994 | 0.995 | 0.996 |



ANEXO 2. Resultado de la metodología de Ferreira et al. [10] en el dataset propuesto por M. Ricardo et al. [14].

Se utilizaron los mismos parámetros de optimización expuestos en el Anexo 1.

Tabla 21. Resultados de clasificadores en el dataset M. Ricardo et al. [14]

| Clasificador | nf | TP Rate | FP Rate | Precision | Recall | F-1 | MCC | AUC | PRC |
|------------------------|----|---------|---------|-----------|--------|-------|-------|--------------|-------|
| Acelerómetro, nf = 2 | | | | | | | | | |
| MHLDForest | 2 | 0.876 | 0.037 | 0.874 | 0.876 | 0.874 | 0.843 | 0.983 | 0.929 |
| Random Forest | 2 | 0.864 | 0.038 | 0.862 | 0.864 | 0.861 | 0.828 | 0.981 | 0.922 |
| Rotation Forest | 2 | 0.876 | 0.037 | 0.874 | 0.876 | 0.873 | 0.843 | 0.981 | 0.926 |
| PBC4cip | 2 | 0.860 | 0.039 | 0.857 | 0.860 | 0.857 | 0.822 | 0.979 | 0.916 |
| Bagging | 2 | 0.848 | 0.044 | 0.845 | 0.848 | 0.845 | 0.807 | 0.975 | 0.903 |
| J48 | 2 | 0.815 | 0.050 | 0.813 | 0.815 | 0.814 | 0.766 | 0.891 | 0.733 |
| MLP | 2 | 0.834 | 0.048 | 0.830 | 0.834 | 0.831 | 0.789 | 0.957 | 0.867 |
| GNB(sklearn) | 2 | 0.686 | 0.076 | 0.706 | 0.686 | 0.687 | 0.613 | 0.925 | 0.717 |

ANEXO 3. Resultado de la experimentación de la metodología BoW en el dataset de Ferreira et al. [10]

Tabla 22. Parámetros evaluados para optimizar los clasificadores propuestos para [10]

| Algoritmo | Parámetros | Valores |
|-----------------|---|--|
| Bagging | # of iterations | 200, 100, 10 |
| GNB(sklearn) | Default parameters in sklearn | - |
| J48 | Default parameters in Weka | - |
| MLP(sklearn) | hidden layer sizes | K/2, 100, 75 |
| | max iter | 2000 |
| MHLDTForest | # trees | 200, 100 |
| PBC4cip | # tress | 200, 100 |
| Random Forest | # of iterations | 200, 100 |
| | # of attributes to randomly investigate | $\log_2(\# \text{ predictores} + 1, 10, 15)$ |
| Rotation Forest | # of iterations | 200, 100, 10 |

K: Es el parámetro que se le pasa a BoW para construir el dataset, indica la longitud de las palabras con **L** lecturas.



Tabla 23. Parámetros utilizados en la optimización de BoW

| Parámetros | Valores evaluados |
|--|--|
| Ejes del acelerómetro | X, Y, (X,Y), $\ X,Y\ _2$, (X, Y, Z) |
| Clasificadores | Bagging, GNB(sklearn), J48, MLP(sklearn), Random Forest MHLDTForest, Rotation Forest |
| Tamaños de codebook | K = {5, 10, 20, 50, 100, 150, 200, 250} |
| Duración del evento (marcas de tiempo) | L = {5, 10, 20, 50, 100, 150, 200, 250, 300} |

Tabla 24. Resultados de clasificadores en el dataset de Ferreira et al. [10]

| Ejes | Clasificador | (K, L) | TP Rate | FP Rate | Precision | Recall | F-1 | MCC | AUC | PRC |
|-----------|------------------------|-----------|---------|---------|-----------|--------|-------|-------|--------------|-------|
| (X, Y, Z) | MHLDTForest | (50, 5) | 0.643 | 0.078 | 0.616 | 0.643 | 0.605 | 0.553 | 0.895 | 0.693 |
| | PBC4cip | (50, 5) | 0.643 | 0.063 | 0.651 | 0.643 | 0.633 | 0.578 | 0.889 | 0.673 |
| | Random Forest | (50, 5) | 0.618 | 0.083 | 0.584 | 0.618 | 0.584 | 0.521 | 0.888 | 0.676 |
| | Rotation Forest | (50, 5) | 0.599 | 0.085 | 0.581 | 0.599 | 0.569 | 0.504 | 0.878 | 0.647 |
| | MLP(sklearn) | (50, 5) | 0.478 | 0.100 | 0.474 | 0.478 | 0.475 | 0.376 | 0.808 | 0.503 |
| | J48 | (50, 5) | 0.444 | 0.107 | 0.429 | 0.444 | 0.436 | 0.332 | 0.695 | 0.320 |
| | GNB(sklearn) | (50, 5) | 0.464 | 0.098 | 0.490 | 0.464 | 0.471 | 0.375 | 0.772 | 0.460 |
| | Bagging | (50, 5) | 0.522 | 0.105 | ? | 0.522 | ? | ? | 0.850 | 0.588 |
| | MHLDTForest | (10, 5) | 0.599 | 0.085 | 0.572 | 0.599 | 0.572 | 0.504 | 0.886 | 0.619 |
| | PBC4cip | (10, 5) | 0.556 | 0.077 | 0.557 | 0.556 | 0.550 | 0.475 | 0.883 | 0.610 |
| | Rotation Forest | (10, 5) | 0.546 | 0.096 | 0.523 | 0.546 | 0.520 | 0.441 | 0.877 | 0.591 |
| | Random Forest | (10, 5) | 0.556 | 0.093 | 0.508 | 0.556 | 0.524 | 0.445 | 0.866 | 0.585 |
| | GNB(sklearn) | (10, 5) | 0.372 | 0.093 | 0.370 | 0.372 | 0.323 | 0.256 | 0.799 | 0.416 |
| | J48 | (10, 5) | 0.454 | 0.109 | 0.432 | 0.454 | 0.440 | 0.338 | 0.736 | 0.370 |
| | MLP(sklearn) | (10, 5) | 0.333 | 0.147 | 0.316 | 0.333 | 0.304 | 0.188 | 0.699 | 0.328 |
| | Bagging | (10, 5) | 0.536 | 0.101 | ? | 0.536 | ? | ? | 0.845 | 0.536 |
| | Random Forest | (50, 10) | 0.570 | 0.093 | 0.531 | 0.570 | 0.534 | 0.463 | 0.863 | 0.619 |
| | PBC4cip | (50, 10) | 0.589 | 0.072 | 0.602 | 0.589 | 0.589 | 0.520 | 0.862 | 0.629 |
| | Rotation Forest | (50, 10) | 0.536 | 0.095 | 0.490 | 0.536 | 0.508 | 0.424 | 0.84 | 0.580 |
| | Bagging | (50, 10) | 0.464 | 0.117 | 0.405 | 0.464 | 0.419 | 0.329 | 0.815 | 0.550 |
| | MLP(sklearn) | (50, 10) | 0.464 | 0.098 | 0.473 | 0.464 | 0.464 | 0.368 | 0.791 | 0.488 |
| | GNB(sklearn) | (50, 10) | 0.459 | 0.085 | 0.515 | 0.459 | 0.461 | 0.385 | 0.785 | 0.464 |
| | J48 | (50, 10) | 0.425 | 0.110 | 0.412 | 0.425 | 0.417 | 0.310 | 0.685 | 0.359 |
| | MHLDTForest | (50, 10) | 0.565 | 0.095 | ? | 0.565 | ? | ? | 0.867 | 0.634 |
| | Rotation Forest | (10, 150) | 0.483 | 0.116 | 0.463 | 0.483 | 0.446 | 0.367 | 0.756 | 0.404 |



| | | | | | | | | | | |
|---------------------|------------------------|---------------------|---------------|----------------|----------------|------------------|---------------|--------------|--------------|------------|
| (X, Y, Z) | Random Forest | (10, 150) | 0.498 | 0.111 | 0.515 | 0.498 | 0.474 | 0.399 | 0.751 | 0.404 |
| | MLP(sklearn) | (10, 150) | 0.382 | 0.137 | 0.349 | 0.382 | 0.336 | 0.239 | 0.684 | 0.316 |
| | Bagging | (10, 150) | 0.459 | 0.121 | 0.418 | 0.459 | 0.417 | 0.330 | 0.737 | 0.391 |
| | PBC4cip | (10, 150) | 0.357 | 0.085 | 0.486 | 0.357 | 0.397 | 0.315 | 0.725 | 0.339 |
| | MHLDTForest | (10, 150) | 0.459 | 0.120 | 0.436 | 0.459 | 0.422 | 0.338 | 0.719 | 0.423 |
| | J48 | (10, 150) | 0.449 | 0.124 | 0.425 | 0.449 | 0.406 | 0.323 | 0.707 | 0.335 |
| | GNB(sklearn) | (10, 150) | 0.227 | 0.140 | ? | 0.227 | ? | ? | 0.598 | 0.241 |
| | Random Forest | (5, 100) | 0.401 | 0.131 | 0.436 | 0.401 | 0.37 | 0.278 | 0.745 | 0.353 |
| | PBC4cip | (5, 100) | 0.324 | 0.098 | 0.375 | 0.324 | 0.325 | 0.231 | 0.734 | 0.303 |
| | Rotation Forest | (5, 100) | 0.406 | 0.131 | ? | 0.406 | ? | ? | 0.742 | 0.326 |
| | Bagging | (5, 100) | 0.420 | 0.127 | ? | 0.420 | ? | ? | 0.739 | 0.306 |
| | J48 | (5, 100) | 0.382 | 0.135 | ? | 0.382 | ? | ? | 0.729 | 0.315 |
| | MLP(sklearn) | (5, 100) | 0.357 | 0.147 | ? | 0.357 | ? | ? | 0.688 | 0.296 |
| | MHLDTForest | (5, 100) | 0.430 | 0.126 | ? | 0.430 | ? | ? | 0.685 | 0.313 |
| | GNB(sklearn) | (5, 100) | 0.304 | 0.139 | ? | 0.304 | ? | ? | 0.647 | 0.272 |
| | Ejes | Clasificador | (K, L) | TP Rate | FP Rate | Precision | Recall | F-1 | MCC | AUC |
| (X, Y) | Rotation Forest | (10, 5) | 0.558 | 0.093 | 0.559 | 0.558 | 0.548 | 0.466 | 0.861 | 0.609 |
| | Random Forest | (10, 5) | 0.565 | 0.090 | 0.560 | 0.565 | 0.555 | 0.475 | 0.860 | 0.606 |
| | MHLDTForest | (10, 5) | 0.572 | 0.090 | 0.565 | 0.572 | 0.558 | 0.481 | 0.856 | 0.619 |
| | PBC4cip | (10, 5) | 0.522 | 0.085 | 0.515 | 0.522 | 0.515 | 0.430 | 0.846 | 0.572 |
| | J48 | (10, 5) | 0.449 | 0.109 | 0.454 | 0.449 | 0.443 | 0.343 | 0.722 | 0.376 |
| | GNB(sklearn) | (10, 5) | 0.384 | 0.105 | 0.426 | 0.384 | 0.380 | 0.287 | 0.722 | 0.398 |
| | MLP(sklearn) | (10, 5) | 0.348 | 0.136 | 0.333 | 0.348 | 0.331 | 0.208 | 0.714 | 0.402 |
| | Bagging | (10, 5) | 0.522 | 0.102 | ? | 0.522 | ? | ? | 0.826 | 0.519 |
| | MLP(sklearn) | (50, 5) | 0.442 | 0.106 | 0.450 | 0.442 | 0.443 | 0.340 | 0.746 | 0.447 |
| | PBC4cip | (50, 5) | 0.643 | 0.062 | 0.648 | 0.643 | 0.634 | 0.578 | 0.890 | 0.678 |
| | GNB(sklearn) | (50, 5) | 0.406 | 0.105 | 0.391 | 0.406 | 0.387 | 0.289 | 0.713 | 0.387 |
| | J48 | (50, 5) | 0.355 | 0.120 | 0.366 | 0.355 | 0.357 | 0.240 | 0.635 | 0.264 |
| | MHLDTForest | (50, 5) | 0.667 | 0.075 | ? | 0.667 | ? | ? | 0.879 | 0.661 |
| | Random Forest | (50, 5) | 0.623 | 0.082 | ? | 0.623 | ? | ? | 0.869 | 0.658 |
| | Rotation Forest | (50, 5) | 0.601 | 0.084 | ? | 0.601 | ? | ? | 0.858 | 0.615 |
| | Bagging | (50, 5) | 0.572 | 0.092 | ? | 0.572 | ? | ? | 0.835 | 0.556 |
| | PBC4cip | (50, 10) | 0.478 | 0.095 | 0.461 | 0.478 | 0.457 | 0.372 | 0.842 | 0.542 |
| | MLP(sklearn) | (50, 10) | 0.478 | 0.101 | 0.474 | 0.478 | 0.473 | 0.377 | 0.743 | 0.443 |
| GNB(sklearn) | (50, 10) | 0.406 | 0.115 | 0.401 | 0.406 | 0.400 | 0.289 | 0.695 | 0.346 | |
| J48 | (50, 10) | 0.384 | 0.117 | 0.369 | 0.384 | 0.375 | 0.262 | 0.688 | 0.315 | |
| MHLDTForest | (50, 10) | 0.536 | 0.105 | ? | 0.536 | ? | ? | 0.847 | 0.578 | |



| | | | | | | | | | | |
|---------------------|------------------------|---------------|----------------|----------------|------------------|---------------|------------|------------|--------------|------------|
| (X,Y) | Random Forest | (50, 10) | 0.507 | 0.110 | ? | 0.507 | ? | ? | 0.835 | 0.539 |
| | Rotation Forest | (50, 10) | 0.551 | 0.094 | ? | 0.551 | ? | ? | 0.817 | 0.531 |
| | Bagging | (50, 10) | 0.413 | 0.130 | ? | 0.413 | ? | ? | 0.799 | 0.449 |
| | Random Forest | (10, 150) | 0.457 | 0.115 | 0.482 | 0.457 | 0.449 | 0.355 | 0.736 | 0.421 |
| | Rotation Forest | (10, 150) | 0.406 | 0.129 | 0.390 | 0.406 | 0.382 | 0.276 | 0.729 | 0.408 |
| | MLP(sklearn) | (10, 150) | 0.406 | 0.130 | 0.387 | 0.406 | 0.373 | 0.274 | 0.717 | 0.378 |
| | MHLDTForest | (10, 150) | 0.449 | 0.120 | 0.441 | 0.449 | 0.421 | 0.332 | 0.700 | 0.393 |
| | PBC4cip | (10, 150) | 0.362 | 0.105 | 0.445 | 0.362 | 0.391 | 0.287 | 0.689 | 0.320 |
| | Bagging | (10, 150) | 0.377 | 0.135 | 0.359 | 0.377 | 0.345 | 0.237 | 0.684 | 0.350 |
| | J48 | (10, 150) | 0.406 | 0.136 | ? | 0.406 | ? | ? | 0.659 | 0.317 |
| | GNB(sklearn) | (10, 150) | 0.196 | 0.097 | ? | 0.196 | ? | ? | 0.604 | 0.263 |
| | | | | | | | | | | |
| | Random Forest | (5, 100) | 0.348 | 0.137 | 0.291 | 0.348 | 0.296 | 0.191 | 0.699 | 0.330 |
| | Bagging | (5, 100) | 0.341 | 0.141 | 0.234 | 0.341 | 0.271 | 0.165 | 0.696 | 0.296 |
| | PBC4cip | (5, 100) | 0.290 | 0.103 | 0.326 | 0.290 | 0.279 | 0.182 | 0.694 | 0.269 |
| | Rotation Forest | (5, 100) | 0.341 | 0.141 | 0.237 | 0.341 | 0.271 | 0.165 | 0.682 | 0.282 |
| | J48 | (5, 100) | 0.333 | 0.143 | 0.232 | 0.333 | 0.265 | 0.156 | 0.665 | 0.270 |
| | GNB(sklearn) | (5, 100) | 0.167 | 0.080 | 0.294 | 0.167 | 0.166 | 0.110 | 0.587 | 0.235 |
| | MLP(sklearn) | (5, 100) | 0.312 | 0.151 | ? | 0.312 | ? | ? | 0.633 | 0.278 |
| | MHLDTForest | (5, 100) | 0.348 | 0.140 | ? | 0.348 | ? | ? | 0.644 | 0.258 |
| | | | | | | | | | | |
| Ejes | Clasificador | (K, L) | TP Rate | FP Rate | Precision | Recall | F-1 | MCC | AUC | PRC |
| X, Y ² | PBC4cip | (10, 5) | 0.667 | 0.065 | 0.658 | 0.667 | 0.659 | 0.598 | 0.923 | 0.708 |
| | Random Forest | (10, 5) | 0.696 | 0.060 | 0.705 | 0.696 | 0.691 | 0.639 | 0.922 | 0.710 |
| | MHLDTForest | (10, 5) | 0.638 | 0.074 | 0.636 | 0.638 | 0.630 | 0.564 | 0.922 | 0.717 |
| | Rotation Forest | (10, 5) | 0.623 | 0.079 | 0.594 | 0.623 | 0.602 | 0.533 | 0.918 | 0.686 |
| | MLP(sklearn) | (10, 5) | 0.478 | 0.104 | 0.514 | 0.478 | 0.481 | 0.389 | 0.846 | 0.583 |
| | GNB(sklearn) | (10, 5) | 0.551 | 0.090 | 0.571 | 0.551 | 0.533 | 0.466 | 0.895 | 0.645 |
| | Bagging | (10, 5) | 0.623 | 0.079 | ? | 0.623 | ? | ? | 0.907 | 0.640 |
| | J48 | (10, 5) | 0.507 | 0.100 | ? | 0.507 | ? | ? | 0.748 | 0.398 |
| | | | | | | | | | | |
| | MLP(sklearn) | (5, 100) | 0.391 | 0.120 | 0.351 | 0.391 | 0.363 | 0.257 | 0.769 | 0.380 |
| | PBC4cip | (5, 100) | 0.391 | 0.096 | 0.385 | 0.391 | 0.382 | 0.289 | 0.763 | 0.361 |
| | Rotation Forest | (5, 100) | 0.536 | 0.092 | 0.498 | 0.536 | 0.486 | 0.427 | 0.760 | 0.386 |
| | Random Forest | (5, 100) | 0.435 | 0.112 | 0.395 | 0.435 | 0.410 | 0.310 | 0.747 | 0.368 |
| | GNB(sklearn) | (5, 100) | 0.420 | 0.114 | ? | 0.420 | ? | ? | 0.772 | 0.378 |
| | Bagging | (5, 100) | 0.449 | 0.112 | ? | 0.449 | ? | ? | 0.748 | 0.365 |
| | J48 | (5, 100) | 0.42 | 0.118 | ? | 0.42 | ? | ? | 0.716 | 0.336 |
| MHLDTForest | (5, 100) | 0.478 | 0.108 | ? | 0.478 | ? | ? | 0.712 | 0.350 | |



| | | | | | | | | | | | |
|------------------------|------------------------|---------------------|---------------|----------------|----------------|------------------|---------------|------------|--------------|------------|------------|
| X, Y 2 | MHLDTForest | (50, 5) | 0.623 | 0.077 | 0.581 | 0.623 | 0.597 | 0.531 | 0.936 | 0.738 | |
| | PBC4cip | (50, 5) | 0.667 | 0.063 | 0.639 | 0.667 | 0.649 | 0.591 | 0.935 | 0.728 | |
| | MLP(sklearn) | (50, 5) | 0.638 | 0.070 | 0.606 | 0.638 | 0.620 | 0.555 | 0.888 | 0.697 | |
| | J48 | (50, 5) | 0.464 | 0.1 | 0.460 | 0.464 | 0.456 | 0.363 | 0.705 | 0.359 | |
| | Random Forest | (50, 5) | 0.609 | 0.083 | ? | 0.609 | ? | ? | 0.924 | 0.714 | |
| | Rotation Forest | (50, 5) | 0.580 | 0.085 | ? | 0.580 | ? | ? | 0.915 | 0.685 | |
| | Bagging | (50, 5) | 0.536 | 0.101 | ? | 0.536 | ? | ? | 0.877 | 0.634 | |
| | GNB(sklearn) | (50, 5) | 0.522 | 0.100 | ? | 0.522 | ? | ? | 0.773 | 0.431 | |
| | | | | | | | | | | | |
| | MLP(sklearn) | (10, 150) | 0.464 | 0.113 | ? | 0.464 | ? | ? | 0.783 | 0.407 | |
| | Random Forest | (10, 150) | 0.478 | 0.107 | ? | 0.478 | ? | ? | 0.779 | 0.406 | |
| | MHLDTForest | (10, 150) | 0.478 | 0.106 | ? | 0.478 | ? | ? | 0.778 | 0.436 | |
| | PBC4cip | (10, 150) | 0.464 | 0.110 | 0.427 | 0.464 | 0.439 | 0.343 | 0.769 | 0.394 | |
| | Bagging | (10, 150) | 0.406 | 0.121 | ? | 0.406 | ? | ? | 0.759 | 0.366 | |
| | Rotation Forest | (10, 150) | 0.464 | 0.107 | ? | 0.464 | ? | ? | 0.744 | 0.375 | |
| | J48 | (10, 150) | 0.478 | 0.102 | ? | 0.478 | ? | ? | 0.735 | 0.360 | |
| | GNB(sklearn) | (10, 150) | 0.377 | 0.133 | ? | 0.377 | ? | ? | 0.751 | 0.355 | |
| | | | | | | | | | | | |
| | Random Forest | (50, 10) | 0.609 | 0.082 | 0.550 | 0.609 | 0.569 | 0.505 | 0.924 | 0.713 | |
| | PBC4cip | (50, 10) | 0.652 | 0.070 | 0.619 | 0.652 | 0.629 | 0.569 | 0.922 | 0.716 | |
| | Rotation Forest | (50, 10) | 0.609 | 0.082 | 0.543 | 0.609 | 0.572 | 0.502 | 0.911 | 0.696 | |
| | MHLDTForest | (50, 10) | 0.652 | 0.076 | ? | 0.652 | ? | ? | 0.917 | 0.723 | |
| | MLP(sklearn) | (50, 10) | 0.652 | 0.073 | ? | 0.652 | ? | ? | 0.898 | 0.722 | |
| | Bagging | (50, 10) | 0.522 | 0.104 | ? | 0.522 | ? | ? | 0.858 | 0.596 | |
| | GNB(sklearn) | (50, 10) | 0.667 | 0.071 | ? | 0.667 | ? | ? | 0.834 | 0.533 | |
| | J48 | (50, 10) | 0.391 | 0.125 | ? | 0.391 | ? | ? | 0.669 | 0.310 | |
| | Ejes | Clasificador | (K, L) | TP Rate | FP Rate | Precision | Recall | F-1 | MCC | AUC | PRC |
| | | | | | | | | | | | |
| (X) | MHLDTForest | (10, 5) | 0.536 | 0.095 | 0.528 | 0.536 | 0.517 | 0.439 | 0.890 | 0.706 | |
| | Random Forest | (10, 5) | 0.536 | 0.094 | 0.547 | 0.536 | 0.529 | 0.448 | 0.875 | 0.679 | |
| | PBC4cip | (10, 5) | 0.522 | 0.089 | 0.507 | 0.522 | 0.506 | 0.425 | 0.874 | 0.656 | |
| | Rotation Forest | (10, 5) | 0.551 | 0.095 | 0.554 | 0.551 | 0.544 | 0.460 | 0.867 | 0.671 | |
| | Bagging | (10, 5) | 0.493 | 0.107 | 0.424 | 0.493 | 0.448 | 0.359 | 0.823 | 0.592 | |
| | GNB(sklearn) | (10, 5) | 0.493 | 0.107 | 0.497 | 0.493 | 0.46 | 0.387 | 0.811 | 0.529 | |
| | J48 | (10, 5) | 0.478 | 0.099 | 0.465 | 0.478 | 0.468 | 0.373 | 0.709 | 0.394 | |
| | MLP(sklearn) | (10, 5) | 0.406 | 0.127 | 0.411 | 0.406 | 0.381 | 0.281 | 0.708 | 0.363 | |
| | | | | | | | | | | | |
| | PBC4cip | (5, 100) | 0.261 | 0.111 | 0.246 | 0.261 | 0.232 | 0.129 | 0.708 | 0.305 | |
| | GNB(sklearn) | (5, 100) | 0.333 | 0.104 | 0.351 | 0.333 | 0.309 | 0.225 | 0.681 | 0.353 | |
| | MLP(sklearn) | (5, 100) | 0.333 | 0.145 | ? | 0.333 | ? | ? | 0.706 | 0.342 | |
| Rotation Forest | (5, 100) | 0.391 | 0.134 | ? | 0.391 | ? | ? | 0.688 | 0.312 | | |
| Random Forest | (5, 100) | 0.333 | 0.146 | ? | 0.333 | ? | ? | 0.677 | 0.309 | | |



| | | | | | | | | | | | |
|----------------------|------------------------|------------------------|---------------|----------------|----------------|------------------|---------------|------------|--------------|--------------|------------|
| (X) | MHLDTForest | (5, 100) | 0.406 | 0.131 | ? | 0.406 | ? | ? | 0.663 | 0.316 | |
| | Bagging | (5, 100) | 0.333 | 0.144 | ? | 0.333 | ? | ? | 0.659 | 0.286 | |
| | J48 | (5, 100) | 0.333 | 0.148 | ? | 0.333 | ? | ? | 0.659 | 0.285 | |
| | MHLDTForest | (50, 5) | 0.652 | 0.077 | 0.619 | 0.652 | 0.621 | 0.562 | 0.901 | 0.726 | |
| | PBC4cip | (50, 5) | 0.638 | 0.066 | 0.598 | 0.638 | 0.604 | 0.548 | 0.899 | 0.710 | |
| | Rotation Forest | (50, 5) | 0.580 | 0.088 | 0.535 | 0.580 | 0.545 | 0.472 | 0.885 | 0.685 | |
| | MLP(sklearn) | (50, 5) | 0.580 | 0.082 | 0.555 | 0.580 | 0.563 | 0.487 | 0.781 | 0.578 | |
| | GNB(sklearn) | (50, 5) | 0.478 | 0.107 | 0.467 | 0.478 | 0.451 | 0.364 | 0.743 | 0.399 | |
| | J48 | (50, 5) | 0.377 | 0.122 | 0.397 | 0.377 | 0.378 | 0.265 | 0.636 | 0.317 | |
| | Random Forest | (50, 5) | 0.638 | 0.081 | ? | 0.638 | ? | ? | 0.909 | 0.746 | |
| | Bagging | (50, 5) | 0.580 | 0.093 | ? | 0.58 | ? | ? | 0.868 | 0.631 | |
| | Random Forest | (10, 150) | 0.420 | 0.122 | 0.376 | 0.42 | 0.385 | 0.283 | 0.758 | 0.464 | |
| | MLP(sklearn) | (10, 150) | 0.406 | 0.124 | 0.360 | 0.406 | 0.368 | 0.266 | 0.745 | 0.453 | |
| | PBC4cip | (10, 150) | 0.464 | 0.099 | 0.472 | 0.464 | 0.440 | 0.362 | 0.732 | 0.392 | |
| | MHLDTForest | (10, 150) | 0.449 | 0.116 | 0.432 | 0.449 | 0.422 | 0.328 | 0.718 | 0.406 | |
| | Rotation Forest | (10, 150) | 0.391 | 0.126 | ? | 0.391 | ? | ? | 0.727 | 0.397 | |
| | Bagging | (10, 150) | 0.435 | 0.119 | ? | 0.435 | ? | ? | 0.719 | 0.376 | |
| | J48 | (10, 150) | 0.362 | 0.131 | ? | 0.362 | ? | ? | 0.708 | 0.328 | |
| | GNB(sklearn) | (10, 150) | 0.362 | 0.125 | ? | 0.362 | ? | ? | 0.681 | 0.393 | |
| | MHLDTForest | (50, 10) | 0.638 | 0.083 | 0.708 | 0.638 | 0.612 | 0.579 | 0.902 | 0.741 | |
| | PBC4cip | (50, 10) | 0.609 | 0.077 | 0.590 | 0.609 | 0.594 | 0.523 | 0.901 | 0.724 | |
| | MLP(sklearn) | (50, 10) | 0.536 | 0.097 | 0.569 | 0.536 | 0.53 | 0.449 | 0.824 | 0.599 | |
| | J48 | (50, 10) | 0.478 | 0.104 | 0.487 | 0.478 | 0.475 | 0.379 | 0.771 | 0.475 | |
| | Random Forest | (50, 10) | 0.551 | 0.100 | ? | 0.551 | ? | ? | 0.894 | 0.725 | |
| | Rotation Forest | (50, 10) | 0.551 | 0.094 | ? | 0.551 | ? | ? | 0.880 | 0.661 | |
| | Bagging | (50, 10) | 0.478 | 0.117 | ? | 0.478 | ? | ? | 0.822 | 0.543 | |
| | GNB(sklearn) | (50, 10) | 0.478 | 0.106 | ? | 0.478 | ? | ? | 0.743 | 0.432 | |
| | Ejes | Clasificador | (K, L) | TP Rate | FP Rate | Precision | Recall | F-1 | MCC | AUC | PRC |
| | (Y) | Rotation Forest | (10, 5) | 0.681 | 0.069 | 0.699 | 0.681 | 0.681 | 0.621 | 0.894 | 0.701 |
| | | PBC4cip | (10, 5) | 0.638 | 0.066 | 0.624 | 0.638 | 0.621 | 0.559 | 0.866 | 0.666 |
| Random Forest | | (10, 5) | 0.652 | 0.077 | 0.665 | 0.652 | 0.644 | 0.582 | 0.868 | 0.658 | |
| MHLDTForest | | (10, 5) | 0.667 | 0.074 | 0.699 | 0.667 | 0.666 | 0.609 | 0.866 | 0.662 | |
| GNB(sklearn) | | (10, 5) | 0.623 | 0.086 | 0.654 | 0.623 | 0.59 | 0.545 | 0.784 | 0.549 | |
| MLP(sklearn) | | (10, 5) | 0.406 | 0.123 | 0.406 | 0.406 | 0.394 | 0.287 | 0.739 | 0.457 | |
| Bagging | | (10, 5) | 0.565 | 0.097 | ? | 0.565 | ? | ? | 0.852 | 0.602 | |
| J48 | | (10, 5) | 0.435 | 0.112 | ? | 0.435 | ? | ? | 0.688 | 0.356 | |
| PBC4cip | | | 0.449 | 0.077 | 0.529 | 0.449 | 0.443 | 0.383 | 0.750 | 0.401 | |



| | | | | | | | | | | | |
|------------|------------------------|-----------|-------|-------|-------|-------|-------|-------|--------------|-------|--|
| (Y) | GNB(sklearn) | (5, 100) | 0.290 | 0.120 | 0.314 | 0.290 | 0.229 | 0.155 | 0.666 | 0.339 | |
| | MHLDTForest | (5, 100) | 0.522 | 0.104 | ? | 0.522 | ? | ? | 0.734 | 0.426 | |
| | Random Forest | (5, 100) | 0.507 | 0.106 | ? | 0.507 | ? | ? | 0.718 | 0.397 | |
| | Rotation Forest | (5, 100) | 0.522 | 0.106 | ? | 0.522 | ? | ? | 0.715 | 0.397 | |
| | Bagging | (5, 100) | 0.522 | 0.106 | ? | 0.522 | ? | ? | 0.712 | 0.346 | |
| | MLP(sklearn) | (5, 100) | 0.435 | 0.127 | ? | 0.435 | ? | ? | 0.707 | 0.379 | |
| | J48 | (5, 100) | 0.493 | 0.112 | ? | 0.493 | ? | ? | 0.642 | 0.312 | |
| | | | | | | | | | | | |
| | Rotation Forest | (50, 5) | 0.652 | 0.077 | 0.685 | 0.652 | 0.642 | 0.586 | 0.873 | 0.691 | |
| | PBC4cip | (50, 5) | 0.623 | 0.073 | 0.602 | 0.623 | 0.603 | 0.539 | 0.862 | 0.695 | |
| | MLP(sklearn) | (50, 5) | 0.623 | 0.075 | 0.612 | 0.623 | 0.614 | 0.545 | 0.855 | 0.689 | |
| | J48 | (50, 5) | 0.580 | 0.081 | 0.560 | 0.580 | 0.561 | 0.489 | 0.771 | 0.513 | |
| | Random Forest | (50, 5) | 0.710 | 0.068 | ? | 0.710 | ? | ? | 0.860 | 0.687 | |
| | MHLDTForest | (50, 5) | 0.652 | 0.082 | ? | 0.652 | ? | ? | 0.851 | 0.688 | |
| | Bagging | (50, 5) | 0.594 | 0.095 | ? | 0.594 | ? | ? | 0.857 | 0.693 | |
| | GNB(sklearn) | (50, 5) | 0.449 | 0.109 | ? | 0.449 | ? | ? | 0.694 | 0.378 | |
| | | | | | | | | | | | |
| | PBC4cip | (10, 150) | 0.449 | 0.073 | 0.540 | 0.449 | 0.464 | 0.398 | 0.785 | 0.413 | |
| | GNB(sklearn) | (10, 150) | 0.391 | 0.086 | 0.476 | 0.391 | 0.395 | 0.322 | 0.736 | 0.394 | |
| | Random Forest | (10, 150) | 0.522 | 0.108 | ? | 0.522 | ? | ? | 0.788 | 0.468 | |
| | Rotation Forest | (10, 150) | 0.478 | 0.118 | ? | 0.478 | ? | ? | 0.756 | 0.411 | |
| | Bagging | (10, 150) | 0.406 | 0.133 | ? | 0.406 | ? | ? | 0.747 | 0.398 | |
| | MLP(sklearn) | (10, 150) | 0.435 | 0.118 | ? | 0.435 | ? | ? | 0.741 | 0.412 | |
| | MHLDTForest | (10, 150) | 0.478 | 0.12 | ? | 0.478 | ? | ? | 0.739 | 0.465 | |
| | J48 | (10, 150) | 0.377 | 0.14 | ? | 0.377 | ? | ? | 0.677 | 0.315 | |
| | | | | | | | | | | | |
| | PBC4cip | (50, 10) | 0.667 | 0.060 | 0.654 | 0.667 | 0.648 | 0.598 | 0.877 | 0.690 | |
| | MLP(sklearn) | (50, 10) | 0.580 | 0.081 | 0.601 | 0.580 | 0.577 | 0.507 | 0.861 | 0.653 | |
| | GNB(sklearn) | (50, 10) | 0.507 | 0.101 | 0.454 | 0.507 | 0.473 | 0.385 | 0.740 | 0.417 | |
| | MHLDTForest | (50, 10) | 0.609 | 0.092 | ? | 0.609 | ? | ? | 0.877 | 0.707 | |
| | Random Forest | (50, 10) | 0.551 | 0.103 | ? | 0.551 | ? | ? | 0.848 | 0.662 | |
| | Rotation Forest | (50, 10) | 0.565 | 0.096 | ? | 0.565 | ? | ? | 0.842 | 0.594 | |
| | Bagging | (50, 10) | 0.536 | 0.108 | ? | 0.536 | ? | ? | 0.775 | 0.529 | |
| J48 | (50, 10) | 0.377 | 0.126 | ? | 0.377 | ? | ? | 0.685 | 0.345 | | |

Tabla 25. Resultados de clasificadores del AUC en el dataset de [10] usando BoW

| AUC | D-1 | D-2 | D-3 | D-4 | D-5 |
|---|--------------|---------------|-------------|--------------|-------------|
| Clasificadores | (K=5, L=100) | (K=10, L=150) | (K=10, L=5) | (K=50, L=10) | (K=50, L=5) |
| Acelerómetro, combinación de los ejes X,Y,Z | | | | | |
| MHLDForest | 0.685 | 0.719 | 0.886 | 0.867 | 0.895 |



| | | | | | |
|--|-------|-------|-------|-------|-------|
| Random Forest | 0.745 | 0.751 | 0.866 | 0.863 | 0.888 |
| Rotation Forest | 0.742 | 0.756 | 0.877 | 0.84 | 0.878 |
| PBC4cip | 0.734 | 0.725 | 0.883 | 0.862 | 0.889 |
| Bagging | 0.739 | 0.737 | 0.845 | 0.815 | 0.850 |
| J48 | 0.729 | 0.707 | 0.736 | 0.685 | 0.695 |
| MLP(sklearn) | 0.688 | 0.684 | 0.699 | 0.791 | 0.808 |
| GNB(sklearn) | 0.647 | 0.598 | 0.799 | 0.785 | 0.772 |
| Acelerómetro, combinación de los ejes X,Y | | | | | |
| MHLDForest | 0.644 | 0.700 | 0.856 | 0.847 | 0.879 |
| Random Forest | 0.699 | 0.736 | 0.860 | 0.835 | 0.869 |
| Rotation Forest | 0.682 | 0.729 | 0.861 | 0.817 | 0.858 |
| PBC4cip | 0.694 | 0.689 | 0.846 | 0.842 | 0.890 |
| Bagging | 0.696 | 0.684 | 0.826 | 0.799 | 0.835 |
| J48 | 0.665 | 0.659 | 0.722 | 0.688 | 0.635 |
| MLP | 0.633 | 0.717 | 0.714 | 0.743 | 0.746 |
| GNB(sklearn) | 0.587 | 0.604 | 0.722 | 0.695 | 0.713 |
| Acelerómetro, norma de $\ X,Y\ _2$ | | | | | |
| MHLDForest | 0.712 | 0.778 | 0.922 | 0.917 | 0.936 |
| Random Forest | 0.747 | 0.779 | 0.922 | 0.924 | 0.924 |
| Rotation Forest | 0.760 | 0.744 | 0.918 | 0.911 | 0.915 |
| PBC4cip | 0.763 | 0.769 | 0.923 | 0.922 | 0.935 |
| Bagging | 0.748 | 0.759 | 0.907 | 0.858 | 0.877 |
| J48 | 0.716 | 0.735 | 0.748 | 0.669 | 0.705 |
| MLP | 0.769 | 0.783 | 0.846 | 0.898 | 0.888 |
| GNB(sklearn) | 0.772 | 0.751 | 0.895 | 0.834 | 0.773 |
| Acelerómetro eje X | | | | | |
| MHLDForest | 0.663 | 0.718 | 0.890 | 0.902 | 0.901 |
| Random Forest | 0.677 | 0.758 | 0.875 | 0.894 | 0.909 |
| Rotation Forest | 0.688 | 0.727 | 0.867 | 0.880 | 0.885 |
| PBC4cip | 0.708 | 0.732 | 0.874 | 0.901 | 0.899 |
| Bagging | 0.659 | 0.719 | 0.823 | 0.822 | 0.868 |
| J48 | 0.659 | 0.708 | 0.709 | 0.771 | 0.636 |
| MLP | 0.706 | 0.745 | 0.708 | 0.824 | 0.781 |
| GNB(sklearn) | 0.681 | 0.681 | 0.811 | 0.743 | 0.743 |
| Acelerómetro eje Y | | | | | |
| MHLDForest | 0.734 | 0.739 | 0.866 | 0.877 | 0.851 |
| Random Forest | 0.718 | 0.788 | 0.868 | 0.848 | 0.860 |
| Rotation Forest | 0.715 | 0.756 | 0.894 | 0.842 | 0.873 |
| PBC4cip | 0.750 | 0.785 | 0.866 | 0.877 | 0.862 |
| Bagging | 0.712 | 0.747 | 0.852 | 0.775 | 0.857 |
| J48 | 0.642 | 0.677 | 0.688 | 0.685 | 0.771 |
| MLP | 0.707 | 0.741 | 0.739 | 0.861 | 0.855 |
| GNB(sklearn) | 0.666 | 0.736 | 0.784 | 0.740 | 0.694 |



Resumen de mejores resultados:

Tabla 26. Resultados sobre la norma $\|X, Y\|_2$ y el eje X , al aplicar el filtro ADLF en [10]

| | Clasificador | (K, L) | TP Rate | FP Rate | Precision | Recall | F-1 | MCC | AUC | PRC | |
|--------------|-----------------|-----------|---------|---------|-----------|--------|-------|--------------|--------------|-------|--|
| $\ X, Y\ _2$ | Random Forest | (10, 5) | 0.594 | 0.081 | 0.564 | 0.594 | 0.575 | 0.502 | 0.924 | 0.693 | |
| | Rotation Forest | (10, 5) | 0.652 | 0.073 | 0.650 | 0.652 | 0.643 | 0.579 | 0.921 | 0.717 | |
| | Bagging | (10, 5) | 0.667 | 0.071 | 0.597 | 0.667 | 0.621 | 0.571 | 0.913 | 0.666 | |
| | GNB(sklearn) | (10, 5) | 0.681 | 0.064 | 0.676 | 0.681 | 0.673 | 0.616 | 0.909 | 0.686 | |
| | MLP(sklearn) | (10, 5) | 0.594 | 0.079 | 0.584 | 0.594 | 0.584 | 0.510 | 0.908 | 0.668 | |
| | PBC4cip | (10, 5) | 0.681 | 0.059 | 0.688 | 0.681 | 0.683 | 0.624 | 0.902 | 0.694 | |
| | MHLDTForest | (10, 5) | 0.652 | 0.066 | 0.651 | 0.652 | 0.651 | 0.586 | 0.901 | 0.684 | |
| | J48 | (10, 5) | 0.507 | 0.091 | 0.492 | 0.507 | 0.493 | 0.410 | 0.789 | 0.482 | |
| | | | | | | | | | | | |
| | MLP(sklearn) | (5, 100) | 0.435 | 0.114 | ? | 0.435 | ? | ? | 0.778 | 0.352 | |
| | Random Forest | (5, 100) | 0.449 | 0.11 | 0.401 | 0.449 | 0.418 | 0.323 | 0.773 | 0.378 | |
| | Bagging | (5, 100) | 0.522 | 0.099 | 0.455 | 0.522 | 0.468 | 0.399 | 0.772 | 0.392 | |
| | J48 | (5, 100) | 0.551 | 0.09 | ? | 0.551 | ? | ? | 0.758 | 0.378 | |
| | PBC4cip | (5, 100) | 0.406 | 0.095 | 0.416 | 0.406 | 0.397 | 0.310 | 0.740 | 0.347 | |
| | GNB(sklearn) | (5, 100) | 0.348 | 0.135 | ? | 0.348 | ? | ? | 0.716 | 0.335 | |
| | MHLDTForest | (5, 100) | 0.478 | 0.108 | 0.450 | 0.478 | 0.436 | 0.362 | 0.687 | 0.342 | |
| | Rotation Forest | (5, 100) | 0.362 | 0.140 | 0.272 | 0.362 | 0.296 | 0.195 | 0.686 | 0.322 | |
| | | | | | | | | | | | |
| | MLP(sklearn) | (50, 5) | 0.957 | 0.010 | 0.958 | 0.957 | 0.956 | 0.948 | 0.998 | 0.991 | |
| | Rotation Forest | (50, 5) | 0.942 | 0.012 | 0.942 | 0.942 | 0.942 | 0.930 | 0.997 | 0.988 | |
| | MHLDTForest | (50, 5) | 0.957 | 0.010 | 0.958 | 0.957 | 0.957 | 0.947 | 0.995 | 0.982 | |
| | PBC4cip | (50, 5) | 0.957 | 0.010 | 0.958 | 0.957 | 0.957 | 0.947 | 0.995 | 0.982 | |
| | Random Forest | (50, 5) | 0.942 | 0.013 | 0.944 | 0.942 | 0.942 | 0.930 | 0.994 | 0.978 | |
| | Bagging | (50, 5) | 0.913 | 0.017 | 0.918 | 0.913 | 0.914 | 0.897 | 0.991 | 0.965 | |
| | GNB(sklearn) | (50, 5) | 0.957 | 0.010 | 0.958 | 0.957 | 0.956 | 0.948 | 0.984 | 0.951 | |
| | J48 | (50, 5) | 0.870 | 0.022 | 0.875 | 0.870 | 0.868 | 0.848 | 0.926 | 0.777 | |
| | | | | | | | | | | | |
| | MLP(sklearn) | (10, 150) | 0.507 | 0.104 | 0.531 | 0.507 | 0.499 | 0.413 | 0.812 | 0.469 | |
| | GNB(sklearn) | (10, 150) | 0.536 | 0.098 | 0.565 | 0.536 | 0.523 | 0.450 | 0.800 | 0.454 | |
| | Random Forest | (10, 150) | 0.536 | 0.099 | ? | 0.536 | ? | ? | 0.791 | 0.448 | |
| | Bagging | (10, 150) | 0.507 | 0.102 | 0.454 | 0.507 | 0.474 | 0.386 | 0.779 | 0.402 | |
| | Rotation Forest | (10, 150) | 0.493 | 0.108 | 0.450 | 0.493 | 0.466 | 0.372 | 0.778 | 0.403 | |
| | J48 | (10, 150) | 0.493 | 0.106 | ? | 0.493 | ? | ? | 0.765 | 0.392 | |
| | PBC4cip | (10, 150) | 0.536 | 0.098 | 0.498 | 0.536 | 0.514 | 0.427 | 0.753 | 0.413 | |
| | MHLDTForest | (10, 150) | 0.522 | 0.100 | 0.488 | 0.522 | 0.502 | 0.413 | 0.725 | 0.414 | |
| | | | | | | | | | | | |
| MLP(sklearn) | (50, 10) | 0.928 | 0.017 | 0.929 | 0.928 | 0.927 | 0.913 | 0.996 | 0.986 | | |



| | | | | | | | | | | |
|------------------------|------------------------|---------------|----------------|----------------|------------------|---------------|------------|--------------|--------------|------------|
| | Rotation Forest | (50, 10) | 0.942 | 0.013 | 0.942 | 0.942 | 0.942 | 0.929 | 0.995 | 0.981 |
| | Random Forest | (50, 10) | 0.942 | 0.010 | 0.952 | 0.942 | 0.942 | 0.934 | 0.994 | 0.979 |
| | PBC4cip | (50, 10) | 0.957 | 0.010 | 0.958 | 0.957 | 0.957 | 0.947 | 0.993 | 0.976 |
| | MHLDTForest | (50, 10) | 0.971 | 0.007 | 0.971 | 0.971 | 0.971 | 0.964 | 0.993 | 0.976 |
| | GNB(sklearn) | (50, 10) | 0.957 | 0.010 | 0.959 | 0.957 | 0.957 | 0.948 | 0.991 | 0.972 |
| | Bagging | (50, 10) | 0.913 | 0.015 | 0.918 | 0.913 | 0.912 | 0.899 | 0.988 | 0.958 |
| | J48 | (50, 10) | 0.870 | 0.025 | 0.875 | 0.87 | 0.871 | 0.848 | 0.931 | 0.794 |
| Ejes | Clasificador | (K, L) | TP Rate | FP Rate | Precision | Recall | F-1 | MCC | AUC | PRC |
| (X) | Rotation Forest | (10, 5) | 0.536 | 0.095 | 0.516 | 0.536 | 0.524 | 0.434 | 0.830 | 0.589 |
| | Random Forest | (10, 5) | 0.507 | 0.097 | 0.488 | 0.507 | 0.494 | 0.401 | 0.829 | 0.559 |
| | Bagging | (10, 5) | 0.507 | 0.100 | 0.474 | 0.507 | 0.485 | 0.393 | 0.827 | 0.563 |
| | MHLDTForest | (10, 5) | 0.507 | 0.089 | 0.493 | 0.507 | 0.492 | 0.406 | 0.824 | 0.562 |
| | PBC4cip | (10, 5) | 0.507 | 0.089 | 0.493 | 0.507 | 0.492 | 0.406 | 0.824 | 0.562 |
| | MLP(sklearn) | (10, 5) | 0.536 | 0.096 | 0.538 | 0.536 | 0.534 | 0.443 | 0.817 | 0.571 |
| | GNB(sklearn) | (10, 5) | 0.464 | 0.109 | 0.491 | 0.464 | 0.461 | 0.364 | 0.810 | 0.562 |
| | J48 | (10, 5) | 0.478 | 0.098 | 0.505 | 0.478 | 0.482 | 0.390 | 0.766 | 0.437 |
| | | | | | | | | | | |
| | MLP(sklearn) | (5, 100) | 0.551 | 0.093 | 0.550 | 0.551 | 0.547 | 0.459 | 0.814 | 0.556 |
| | PBC4cip | (5, 100) | 0.261 | 0.108 | 0.323 | 0.261 | 0.241 | 0.152 | 0.706 | 0.315 |
| | MHLDTForest | (5, 100) | 0.261 | 0.108 | 0.323 | 0.261 | 0.241 | 0.152 | 0.706 | 0.315 |
| | Bagging | (5, 100) | 0.348 | 0.143 | 0.267 | 0.348 | 0.287 | 0.181 | 0.690 | 0.310 |
| | GNB(sklearn) | (5, 100) | 0.435 | 0.120 | 0.424 | 0.435 | 0.381 | 0.306 | 0.689 | 0.356 |
| | Rotation Forest | (5, 100) | 0.362 | 0.140 | 0.272 | 0.362 | 0.296 | 0.195 | 0.686 | 0.362 |
| | J48 | (5, 100) | 0.348 | 0.143 | ? | 0.348 | ? | ? | 0.687 | 0.295 |
| | Random Forest | (5, 100) | 0.348 | 0.143 | ? | 0.348 | ? | ? | 0.686 | 0.320 |
| | | | | | | | | | | |
| | MLP(sklearn) | (50, 5) | 0.986 | 0.004 | 0.986 | 0.986 | 0.985 | 0.982 | 0.999 | 0.995 |
| | MHLDTForest | (50, 5) | 0.957 | 0.010 | 0.958 | 0.957 | 0.957 | 0.947 | 0.998 | 0.993 |
| | PBC4cip | (50, 5) | 0.957 | 0.010 | 0.958 | 0.957 | 0.957 | 0.947 | 0.997 | 0.988 |
| | Random Forest | (50, 5) | 0.957 | 0.010 | 0.957 | 0.957 | 0.956 | 0.947 | 0.996 | 0.985 |
| | Rotation Forest | (50, 5) | 0.942 | 0.013 | 0.944 | 0.942 | 0.943 | 0.929 | 0.996 | 0.983 |
| | Bagging | (50, 5) | 0.928 | 0.017 | 0.928 | 0.928 | 0.927 | 0.911 | 0.994 | 0.977 |
| | GNB(sklearn) | (50, 5) | 0.957 | 0.01 | 0.961 | 0.957 | 0.957 | 0.949 | 0.989 | 0.952 |
| | J48 | (50, 5) | 0.855 | 0.031 | 0.876 | 0.855 | 0.858 | 0.833 | 0.930 | 0.792 |
| | | | | | | | | | | |
| PBC4cip | (10, 150) | 0.478 | 0.097 | 0.466 | 0.478 | 0.450 | 0.368 | 0.761 | 0.419 | |
| Random Forest | (10, 150) | 0.449 | 0.113 | 0.419 | 0.449 | 0.425 | 0.325 | 0.759 | 0.468 | |
| MLP(sklearn) | (10, 150) | 0.435 | 0.118 | 0.414 | 0.435 | 0.413 | 0.311 | 0.748 | 0.45 | |
| Bagging | (10, 150) | 0.449 | 0.111 | 0.387 | 0.449 | 0.403 | 0.314 | 0.740 | 0.430 | |
| GNB(sklearn) | (10, 150) | 0.449 | 0.114 | 0.399 | 0.449 | 0.407 | 0.321 | 0.733 | 0.428 | |
| Rotation Forest | (10, 150) | 0.464 | 0.111 | 0.394 | 0.464 | 0.417 | 0.326 | 0.726 | 0.420 | |



| | | | | | | | | | | |
|--|------------------------|-----------|-------|-------|-------|-------|-------|-------|--------------|-------|
| | J48 | (10, 150) | 0.478 | 0.105 | 0.416 | 0.478 | 0.433 | 0.35 | 0.728 | 0.421 |
| | MHLDTForest | (10, 150) | 0.464 | 0.111 | 0.427 | 0.464 | 0.435 | 0.339 | 0.672 | 0.343 |
| | | | | | | | | | | |
| | GNB(sklearn) | (50, 10) | 0.942 | 0.012 | 0.944 | 0.942 | 0.942 | 0.932 | 0.994 | 0.974 |
| | MLP(sklearn) | (50, 10) | 0.957 | 0.008 | 0.958 | 0.957 | 0.956 | 0.948 | 0.993 | 0.976 |
| | Rotation Forest | (50, 10) | 0.971 | 0.005 | 0.973 | 0.971 | 0.971 | 0.966 | 0.993 | 0.971 |
| | Random Forest | (50, 10) | 0.942 | 0.013 | 0.944 | 0.942 | 0.942 | 0.93 | 0.992 | 0.973 |
| | Bagging | (50, 10) | 0.899 | 0.024 | 0.907 | 0.899 | 0.892 | 0.878 | 0.991 | 0.961 |
| | PBC4cip | (50, 10) | 0.884 | 0.023 | 0.883 | 0.884 | 0.882 | 0.860 | 0.990 | 0.951 |
| | MHLDTForest | (50, 10) | 0.884 | 0.023 | 0.883 | 0.884 | 0.882 | 0.860 | 0.989 | 0.952 |
| | J48 | (50, 10) | 0.870 | 0.019 | 0.899 | 0.87 | 0.879 | 0.86 | 0.918 | 0.798 |
| | | | | | | | | | | |

ANEXO 4. Resultado de la experimentación de la metodología BoW en el dataset de M. Ricardo et al. [14]

Se utilizaron los mismos parámetros de optimización expuestos en el Anexo 2.

Tabla 27. Resultados de clasificadores en el dataset de M. Ricardo et al. [14] usando BoW

| Ejes | Clasificador | (K, L) | TP Rate | FP Rate | Precisio n | Recal l | F-1 | MCC | AUC | PRC | |
|---------------------|------------------------|------------|---------|---------|---------------|------------|-------|--------------|--------------|------|--|
| (X, Y, Z) | Rotation Forest | (100, 150) | 0.545 | 0.114 | 0.557 | 0.545 | 0.538 | 0.435 | 0.855 | 0.61 | |
| | Random Forest | (100, 150) | 0.538 | 0.115 | 0.549 | 0.538 | 0.533 | 0.427 | 0.855 | 0.61 | |
| | Bagging | (100, 150) | 0.520 | 0.120 | 0.542 | 0.52 | 0.506 | 0.407 | 0.843 | 0.57 | |
| | MLP(sklearn) | (100, 150) | 0.531 | 0.117 | 0.542 | 0.531 | 0.525 | 0.418 | 0.838 | 0.58 | |
| | J48 | (100, 150) | 0.532 | 0.117 | 0.545 | 0.532 | 0.525 | 0.419 | 0.836 | 0.56 | |
| | MHLDTForest | (100, 150) | 0.515 | 0.121 | 0.524 | 0.515 | 0.509 | 0.397 | 0.778 | 0.45 | |
| | PBC4cip | (100, 150) | 0.505 | 0.123 | 0.515 | 0.505 | 0.501 | 0.385 | 0.713 | 0.31 | |
| | GNB(sklearn) | (100, 150) | 0.321 | 0.168 | 0.428 | 0.321 | 0.262 | 0.194 | 0.675 | 0.33 | |
| | | | | | | | | | | | |
| | Rotation Forest | (150, 200) | 0.552 | 0.112 | 0.56 | 0.552 | 0.543 | 0.442 | 0.854 | 0.60 | |
| | Random Forest | (150, 200) | 0.553 | 0.112 | 0.558 | 0.553 | 0.545 | 0.442 | 0.853 | 0.60 | |
| | Bagging | (150, 200) | 0.509 | 0.123 | 0.523 | 0.509 | 0.49 | 0.389 | 0.839 | 0.54 | |
| | J48 | (150, 200) | 0.517 | 0.121 | 0.530 | 0.517 | 0.498 | 0.399 | 0.835 | 0.54 | |
| | MLP(sklearn) | (150, 200) | 0.539 | 0.115 | 0.545 | 0.539 | 0.532 | 0.426 | 0.833 | 0.57 | |
| | MHLDTForest | (150, 200) | 0.535 | 0.116 | 0.544 | 0.535 | 0.528 | 0.422 | 0.776 | 0.46 | |
| | PBC4cip | (150, 200) | 0.533 | 0.116 | 0.542 | 0.533 | 0.527 | 0.420 | 0.748 | 0.34 | |
| | GNB(sklearn) | (150, 200) | 0.285 | 0.177 | 0.358 | 0.285 | 0.222 | 0.135 | 0.675 | 0.31 | |
| | | | | | | | | | | | |
| | Random Forest | (100, 200) | 0.522 | 0.120 | 0.560 | 0.522 | 0.517 | 0.417 | 0.835 | 0.57 | |
| | Rotation Forest | (100, 200) | 0.517 | 0.121 | 0.569 | 0.517 | 0.513 | 0.416 | 0.834 | 0.57 | |
| MLP(sklearn) | (100, 200) | 0.513 | 0.122 | 0.523 | 0.513 | 0.511 | 0.395 | 0.827 | 0.57 | | |



| | | | | | | | | | | | |
|-------|------------------------|---------------------|---------------|----------------|----------------|-----------------------|--------------------|------------|--------------|-------------|------------|
| | Bagging | (100, 200) | 0.504 | 0.124 | 0.540 | 0.504 | 0.502 | 0.394 | 0.825 | 0.54 | |
| | J48 | (100, 200) | 0.511 | 0.123 | 0.570 | 0.511 | 0.504 | 0.41 | 0.818 | 0.53 | |
| | MHLDTForest | (100, 200) | 0.512 | 0.122 | 0.555 | 0.512 | 0.504 | 0.405 | 0.746 | 0.43 | |
| | PBC4cip | (100, 200) | 0.500 | 0.125 | 0.558 | 0.500 | 0.493 | 0.395 | 0.694 | 0.30 | |
| | GNB(sklearn) | (100, 200) | 0.310 | 0.171 | 0.383 | 0.310 | 0.253 | 0.171 | 0.662 | 0.31 | |
| | | | | | | | | | | | |
| | Random Forest | (50, 150) | 0.495 | 0.127 | 0.536 | 0.495 | 0.485 | 0.383 | 0.817 | 0.55 | |
| | Rotation Forest | (50, 150) | 0.495 | 0.127 | 0.540 | 0.495 | 0.484 | 0.384 | 0.815 | 0.54 | |
| | MLP(sklearn) | (50, 150) | 0.488 | 0.128 | 0.520 | 0.488 | 0.482 | 0.373 | 0.813 | 0.54 | |
| | Bagging | (50, 150) | 0.481 | 0.130 | 0.518 | 0.481 | 0.471 | 0.364 | 0.812 | 0.53 | |
| | J48 | (50, 150) | 0.488 | 0.128 | 0.527 | 0.488 | 0.477 | 0.373 | 0.803 | 0.51 | |
| | MHLDTForest | (50, 150) | 0.492 | 0.127 | 0.533 | 0.492 | 0.483 | 0.38 | 0.716 | 0.40 | |
| | PBC4cip | (50, 150) | 0.472 | 0.132 | 0.521 | 0.472 | 0.458 | 0.358 | 0.672 | 0.28 | |
| | GNB(sklearn) | (50, 150) | 0.316 | 0.169 | 0.479 | 0.316 | 0.265 | 0.206 | 0.669 | 0.34 | |
| | Ejes | Clasificador | (K, L) | TP Rate | FP Rate | Precisio n | Recal l | F-1 | MCC | AUC | PRC |
| (X,Y) | Random Forest | (150, 200) | 0.637 | 0.091 | 0.665 | 0.637 | 0.638 | 0.558 | 0.907 | 0.71 | |
| | Rotation Forest | (150, 200) | 0.630 | 0.093 | 0.656 | 0.630 | 0.631 | 0.548 | 0.904 | 0.70 | |
| | J48 | (150, 200) | 0.613 | 0.097 | 0.651 | 0.613 | 0.609 | 0.531 | 0.891 | 0.64 | |
| | Bagging | (150, 200) | 0.592 | 0.102 | 0.617 | 0.592 | 0.592 | 0.500 | 0.888 | 0.63 | |
| | MLP(sklearn) | (150, 200) | 0.594 | 0.102 | 0.615 | 0.594 | 0.595 | 0.501 | 0.882 | 0.66 | |
| | MHLDTForest | (150, 200) | 0.578 | 0.106 | 0.597 | 0.578 | 0.580 | 0.480 | 0.828 | 0.54 | |
| | PBC4cip | (150, 200) | 0.596 | 0.101 | 0.620 | 0.596 | 0.596 | 0.505 | 0.823 | 0.41 | |
| | GNB(sklearn) | (150, 200) | 0.345 | 0.162 | 0.390 | 0.345 | 0.296 | 0.205 | 0.709 | 0.34 | |
| | | | | | | | | | | | |
| | Random Forest | (100, 150) | 0.626 | 0.094 | 0.681 | 0.626 | 0.617 | 0.552 | 0.898 | 0.70 | |
| | Rotation Forest | (100, 150) | 0.610 | 0.098 | 0.662 | 0.610 | 0.597 | 0.531 | 0.894 | 0.68 | |
| | Bagging | (100, 150) | 0.583 | 0.104 | 0.636 | 0.583 | 0.567 | 0.497 | 0.880 | 0.63 | |
| | MLP(sklearn) | (100, 150) | 0.610 | 0.098 | 0.653 | 0.610 | 0.602 | 0.528 | 0.885 | 0.68 | |
| | J48 | (100, 150) | 0.600 | 0.100 | 0.655 | 0.600 | 0.584 | 0.520 | 0.876 | 0.62 | |
| | PBC4cip | (100, 150) | 0.602 | 0.100 | 0.648 | 0.602 | 0.591 | 0.519 | 0.803 | 0.38 | |
| | MHLDTForest | (100, 150) | 0.605 | 0.099 | 0.647 | 0.605 | 0.595 | 0.521 | 0.797 | 0.51 | |
| | GNB(sklearn) | (100, 150) | 0.371 | 0.156 | 0.456 | 0.371 | 0.334 | 0.254 | 0.718 | 0.37 | |
| | | | | | | | | | | | |
| | Random Forest | (100, 200) | 0.625 | 0.094 | 0.676 | 0.625 | 0.614 | 0.550 | 0.892 | 0.69 | |
| | Rotation Forest | (100, 200) | 0.614 | 0.097 | 0.667 | 0.614 | 0.600 | 0.537 | 0.888 | 0.67 | |
| | MLP(sklearn) | (100, 200) | 0.600 | 0.100 | 0.639 | 0.600 | 0.591 | 0.514 | 0.879 | 0.66 | |
| | J48 | (100, 200) | 0.608 | 0.098 | 0.664 | 0.608 | 0.593 | 0.530 | 0.878 | 0.63 | |
| | Bagging | (100, 200) | 0.600 | 0.100 | 0.650 | 0.600 | 0.584 | 0.518 | 0.872 | 0.62 | |
| | MHLDTForest | (100, 200) | 0.584 | 0.104 | 0.614 | 0.584 | 0.571 | 0.491 | 0.816 | 0.52 | |
| | PBC4cip | (100, 200) | 0.603 | 0.099 | 0.642 | 0.603 | 0.589 | 0.517 | 0.793 | 0.37 | |
| | GNB(sklearn) | (100, 200) | 0.372 | 0.156 | 0.441 | 0.372 | 0.324 | 0.245 | 0.714 | 0.36 | |



| | | | | | | | | | | | |
|-------------|------------------------|------------------------|----------------|----------------|-----------------------|--------------------|------------|------------|--------------|--------------|------|
| | | | | | | | | | | | |
| | Random Forest | (50, 150) | 0.596 | 0.101 | 0.645 | 0.596 | 0.584 | 0.512 | 0.887 | 0.68 | |
| | Rotation Forest | (50, 150) | 0.587 | 0.104 | 0.634 | 0.587 | 0.571 | 0.500 | 0.884 | 0.67 | |
| | MLP(sklearn) | (50, 150) | 0.589 | 0.103 | 0.631 | 0.589 | 0.576 | 0.500 | 0.884 | 0.67 | |
| | Bagging | (50, 150) | 0.569 | 0.108 | 0.613 | 0.569 | 0.556 | 0.477 | 0.878 | 0.65 | |
| | J48 | (50, 150) | 0.579 | 0.106 | 0.626 | 0.579 | 0.561 | 0.490 | 0.875 | 0.64 | |
| | MHLDTForest | (50, 150) | 0.584 | 0.104 | 0.614 | 0.584 | 0.571 | 0.491 | 0.816 | 0.52 | |
| | PBC4cip | (50, 150) | 0.577 | 0.106 | 0.610 | 0.577 | 0.561 | 0.483 | 0.789 | 0.37 | |
| | GNB(sklearn) | (50, 150) | 0.376 | 0.154 | 0.522 | 0.376 | 0.322 | 0.271 | 0.724 | 0.39 | |
| | | | | | | | | | | | |
| Ejes | Clasificador | (K, L) | TP Rate | FP Rate | Precisio n | Recal l | F-1 | MCC | AUC | PRC | |
| X, Y 2 | MLP(sklearn) | (100, 200) | 0.735 | 0.066 | 0.731 | 0.735 | 0.729 | 0.666 | 0.917 | 0.76 | |
| | PBC4cip | (100, 200) | 0.753 | 0.062 | 0.75 | 0.753 | 0.748 | 0.689 | 0.902 | 0.75 | |
| | Rotation Forest | (100, 200) | 0.709 | 0.073 | 0.707 | 0.709 | 0.701 | 0.634 | 0.899 | 0.70 | |
| | Random Forest | (100, 200) | 0.713 | 0.072 | 0.711 | 0.713 | 0.706 | 0.639 | 0.897 | 0.72 | |
| | GNB(sklearn) | (100, 200) | 0.595 | 0.101 | 0.664 | 0.595 | 0.575 | 0.518 | 0.890 | 0.65 | |
| | Bagging | (100, 200) | 0.674 | 0.082 | 0.673 | 0.674 | 0.665 | 0.591 | 0.885 | 0.66 | |
| | J48 | (100, 200) | 0.690 | 0.078 | 0.685 | 0.690 | 0.683 | 0.609 | 0.879 | 0.68 | |
| | MHLDTForest | (100, 200) | 0.671 | 0.082 | 0.672 | 0.671 | 0.662 | 0.588 | 0.874 | 0.67 | |
| | | | | | | | | | | | |
| | | PBC4cip | (150, 200) | 0.803 | 0.049 | 0.802 | 0.803 | 0.8 | 0.753 | 0.921 | 0.82 |
| | | MLP(sklearn) | (150, 200) | 0.723 | 0.069 | 0.719 | 0.723 | 0.720 | 0.652 | 0.913 | 0.76 |
| | | Random Forest | (150, 200) | 0.723 | 0.069 | 0.718 | 0.723 | 0.718 | 0.651 | 0.899 | 0.75 |
| | | Rotation Forest | (150, 200) | 0.715 | 0.071 | 0.710 | 0.715 | 0.710 | 0.641 | 0.898 | 0.73 |
| | | GNB(sklearn) | (150, 200) | 0.634 | 0.092 | 0.695 | 0.634 | 0.626 | 0.562 | 0.891 | 0.65 |
| | | Bagging | (150, 200) | 0.676 | 0.081 | 0.676 | 0.676 | 0.669 | 0.594 | 0.882 | 0.69 |
| | | MHLDTForest | (150, 200) | 0.635 | 0.091 | 0.628 | 0.635 | 0.622 | 0.539 | 0.876 | 0.67 |
| | | J48 | (150, 200) | 0.691 | 0.077 | 0.701 | 0.691 | 0.685 | 0.617 | 0.868 | 0.67 |
| | | | | | | | | | | | |
| | | MLP(sklearn) | (100, 150) | 0.731 | 0.067 | 0.727 | 0.731 | 0.727 | 0.662 | 0.911 | 0.76 |
| | | PBC4cip | (100, 150) | 0.747 | 0.063 | 0.746 | 0.747 | 0.744 | 0.683 | 0.905 | 0.77 |
| | | Rotation Forest | (100, 150) | 0.679 | 0.080 | 0.674 | 0.679 | 0.670 | 0.595 | 0.899 | 0.71 |
| | | Random Forest | (100, 150) | 0.690 | 0.078 | 0.684 | 0.690 | 0.685 | 0.609 | 0.896 | 0.72 |
| | | Bagging | (100, 150) | 0.659 | 0.085 | 0.658 | 0.659 | 0.650 | 0.572 | 0.882 | 0.68 |
| | | J48 | (100, 150) | 0.683 | 0.079 | 0.681 | 0.683 | 0.675 | 0.602 | 0.876 | 0.68 |
| | | MHLDTForest | (100, 150) | 0.635 | 0.091 | 0.628 | 0.635 | 0.622 | 0.539 | 0.876 | 0.67 |
| | | GNB(sklearn) | (100, 150) | 0.576 | 0.106 | 0.652 | 0.576 | 0.567 | 0.497 | 0.862 | 0.59 |
| | | | | | | | | | | | |
| | | MLP(sklearn) | (50, 150) | 0.726 | 0.069 | 0.722 | 0.726 | 0.722 | 0.655 | 0.892 | 0.70 |
| | | Rotation Forest | (50, 150) | 0.713 | 0.072 | 0.709 | 0.713 | 0.705 | 0.638 | 0.890 | 0.69 |
| | | Bagging | (50, 150) | 0.709 | 0.073 | 0.705 | 0.709 | 0.701 | 0.633 | 0.885 | 0.67 |



| | | | | | | | | | | | |
|-------------|------------------------|------------------------|----------------|----------------|-----------------------|--------------------|------------|------------|--------------|--------------|------|
| | Random Forest | (50, 150) | 0.715 | 0.071 | 0.710 | 0.715 | 0.71 | 0.641 | 0.883 | 0.69 | |
| | J48 | (50, 150) | 0.694 | 0.077 | 0.692 | 0.694 | 0.687 | 0.615 | 0.872 | 0.66 | |
| | MHLDTForest | (50, 150) | 0.714 | 0.071 | 0.717 | 0.714 | 0.705 | 0.642 | 0.871 | 0.66 | |
| | PBC4cip | (50, 150) | 0.721 | 0.07 | 0.717 | 0.721 | 0.715 | 0.648 | 0.866 | 0.65 | |
| | GNB(sklearn) | (50, 150) | 0.594 | 0.102 | 0.604 | 0.594 | 0.564 | 0.490 | 0.853 | 0.59 | |
| | | | | | | | | | | | |
| Ejes | Clasificador | (K, L) | TP Rate | FP Rate | Precisio n | Recal l | F-1 | MCC | AUC | PRC | |
| (X) | Random Forest | (100, 150) | 0.623 | 0.094 | 0.639 | 0.623 | 0.588 | 0.531 | 0.872 | 0.60 | |
| | Bagging | (100, 150) | 0.596 | 0.101 | 0.611 | 0.596 | 0.555 | 0.495 | 0.869 | 0.58 | |
| | J48 | (100, 150) | 0.612 | 0.097 | 0.545 | 0.612 | 0.553 | 0.499 | 0.867 | 0.57 | |
| | MLP(sklearn) | (100, 150) | 0.590 | 0.103 | 0.621 | 0.59 | 0.593 | 0.501 | 0.863 | 0.60 | |
| | GNB(sklearn) | (100, 150) | 0.475 | 0.132 | 0.515 | 0.475 | 0.420 | 0.360 | 0.821 | 0.54 | |
| | PBC4cip | (100, 150) | 0.614 | 0.097 | 0.540 | 0.614 | 0.555 | 0.502 | 0.810 | 0.44 | |
| | Rotation Forest | (100, 150) | 0.611 | 0.097 | ? | 0.611 | ? | ? | 0.874 | 0.59 | |
| | MHLDTForest | (100, 150) | 0.616 | 0.096 | ? | 0.616 | ? | ? | 0.839 | 0.58 | |
| | | | | | | | | | | | |
| | | Random Forest | (150, 200) | 0.631 | 0.092 | 0.722 | 0.631 | 0.579 | 0.546 | 0.886 | 0.62 |
| | | Rotation Forest | (150, 200) | 0.62 | 0.095 | 0.747 | 0.62 | 0.565 | 0.539 | 0.883 | 0.62 |
| | | MLP(sklearn) | (150, 200) | 0.583 | 0.105 | 0.614 | 0.583 | 0.58 | 0.492 | 0.881 | 0.63 |
| | | J48 | (150, 200) | 0.603 | 0.099 | 0.749 | 0.603 | 0.539 | 0.516 | 0.867 | 0.57 |
| | | MHLDTForest | (150, 200) | 0.62 | 0.095 | 0.683 | 0.62 | 0.563 | 0.521 | 0.837 | 0.60 |
| | | PBC4cip | (150, 200) | 0.616 | 0.096 | 0.694 | 0.616 | 0.558 | 0.522 | 0.835 | 0.48 |
| | | GNB(sklearn) | (150, 200) | 0.447 | 0.139 | 0.576 | 0.447 | 0.392 | 0.351 | 0.818 | 0.53 |
| | | Bagging | (150, 200) | 0.595 | 0.101 | ? | 0.595 | ? | ? | 0.870 | 0.57 |
| | | | | | | | | | | | |
| | | Random Forest | (100, 200) | 0.638 | 0.091 | 0.758 | 0.638 | 0.586 | 0.558 | 0.883 | 0.62 |
| | | Rotation Forest | (100, 200) | 0.642 | 0.090 | 0.771 | 0.642 | 0.585 | 0.560 | 0.883 | 0.62 |
| | | J48 | (100, 200) | 0.639 | 0.090 | 0.767 | 0.639 | 0.582 | 0.556 | 0.875 | 0.59 |
| | | MLP(sklearn) | (100, 200) | 0.598 | 0.101 | 0.568 | 0.598 | 0.557 | 0.485 | 0.870 | 0.60 |
| | | PBC4cip | (100, 200) | 0.638 | 0.091 | 0.764 | 0.638 | 0.579 | 0.546 | 0.835 | 0.49 |
| | | GNB(sklearn) | (100, 200) | 0.483 | 0.130 | 0.524 | 0.483 | 0.427 | 0.366 | 0.814 | 0.51 |
| | | Bagging | (100, 200) | 0.631 | 0.092 | ? | 0.631 | ? | ? | 0.880 | 0.60 |
| | | MHLDTForest | (100, 200) | 0.626 | 0.094 | ? | 0.626 | ? | ? | 0.813 | 0.57 |
| | | | | | | | | | | | |
| | | Random Forest | (50, 150) | 0.594 | 0.102 | 0.585 | 0.594 | 0.566 | 0.486 | 0.878 | 0.63 |
| | | MLP(sklearn) | (50, 150) | 0.595 | 0.101 | 0.589 | 0.595 | 0.572 | 0.489 | 0.878 | 0.62 |
| | | Rotation Forest | (50, 150) | 0.602 | 0.100 | 0.556 | 0.602 | 0.553 | 0.483 | 0.877 | 0.61 |
| | | Bagging | (50, 150) | 0.594 | 0.102 | 0.581 | 0.594 | 0.559 | 0.483 | 0.874 | 0.61 |
| | | J48 | (50, 150) | 0.610 | 0.098 | 0.584 | 0.610 | 0.559 | 0.498 | 0.868 | 0.58 |
| | GNB(sklearn) | (50, 150) | 0.492 | 0.127 | 0.560 | 0.492 | 0.450 | 0.386 | 0.846 | 0.56 | |
| | PBC4cip | (50, 150) | 0.607 | 0.098 | 0.549 | 0.607 | 0.554 | 0.485 | 0.814 | 0.48 | |
| | MHLDTForest | (50, 150) | 0.604 | 0.099 | 0.553 | 0.604 | 0.553 | 0.485 | 0.810 | 0.57 | |



| Ejes | Clasificador | (K, L) | TP Rate | FP Rate | Precisio n | Recal l | F-1 | MCC | AUC | PRC | |
|---------------------|------------------------|------------|---------|---------|---------------|------------|-------|--------------|--------------|------|--|
| (Y) | Random Forest | (100, 150) | 0.591 | 0.103 | 0.652 | 0.591 | 0.581 | 0.508 | 0.858 | 0.61 | |
| | Rotation Forest | (100, 150) | 0.584 | 0.104 | 0.634 | 0.584 | 0.571 | 0.494 | 0.857 | 0.60 | |
| | MLP(sklearn) | (100, 150) | 0.578 | 0.106 | 0.609 | 0.578 | 0.574 | 0.484 | 0.855 | 0.61 | |
| | Bagging | (100, 150) | 0.544 | 0.114 | 0.568 | 0.544 | 0.536 | 0.438 | 0.851 | 0.59 | |
| | J48 | (100, 150) | 0.571 | 0.108 | 0.636 | 0.571 | 0.564 | 0.483 | 0.848 | 0.58 | |
| | GNB(sklearn) | (100, 150) | 0.544 | 0.114 | 0.637 | 0.544 | 0.549 | 0.467 | 0.828 | 0.57 | |
| | PBC4cip | (100, 150) | 0.564 | 0.109 | 0.589 | 0.564 | 0.568 | 0.466 | 0.828 | 0.49 | |
| | MHLDTForest | (100, 150) | 0.574 | 0.107 | 0.584 | 0.574 | 0.572 | 0.471 | 0.823 | 0.59 | |
| | | | | | | | | | | | |
| | Random Forest | (150, 200) | 0.632 | 0.092 | 0.683 | 0.632 | 0.63 | 0.561 | 0.888 | 0.68 | |
| | MLP(sklearn) | (150, 200) | 0.598 | 0.101 | 0.697 | 0.598 | 0.612 | 0.540 | 0.886 | 0.69 | |
| | Rotation Forest | (150, 200) | 0.62 | 0.095 | 0.672 | 0.620 | 0.616 | 0.546 | 0.884 | 0.67 | |
| | Bagging | (150, 200) | 0.592 | 0.102 | 0.644 | 0.592 | 0.587 | 0.511 | 0.874 | 0.63 | |
| | J48 | (150, 200) | 0.566 | 0.109 | 0.646 | 0.566 | 0.552 | 0.486 | 0.863 | 0.59 | |
| | MHLDTForest | (150, 200) | 0.622 | 0.095 | 0.671 | 0.622 | 0.62 | 0.547 | 0.840 | 0.64 | |
| | GNB(sklearn) | (150, 200) | 0.528 | 0.118 | 0.699 | 0.528 | 0.505 | 0.476 | 0.840 | 0.60 | |
| | PBC4cip | (150, 200) | 0.619 | 0.095 | 0.664 | 0.619 | 0.616 | 0.542 | 0.815 | 0.44 | |
| | | | | | | | | | | | |
| | MLP(sklearn) | (100, 200) | 0.636 | 0.091 | 0.729 | 0.636 | 0.626 | 0.576 | 0.870 | 0.65 | |
| | Random Forest | (100, 200) | 0.646 | 0.089 | 0.723 | 0.646 | 0.628 | 0.582 | 0.867 | 0.64 | |
| | Rotation Forest | (100, 200) | 0.644 | 0.089 | 0.709 | 0.644 | 0.627 | 0.576 | 0.864 | 0.62 | |
| | Bagging | (100, 200) | 0.607 | 0.099 | 0.680 | 0.607 | 0.592 | 0.534 | 0.857 | 0.62 | |
| | J48 | (100, 200) | 0.631 | 0.093 | 0.697 | 0.631 | 0.612 | 0.560 | 0.854 | 0.59 | |
| | GNB(sklearn) | (100, 200) | 0.548 | 0.113 | 0.662 | 0.548 | 0.541 | 0.48 | 0.831 | 0.58 | |
| | PBC4cip | (100, 200) | 0.624 | 0.094 | 0.694 | 0.624 | 0.609 | 0.554 | 0.818 | 0.47 | |
| | MHLDTForest | (100, 200) | 0.635 | 0.092 | 0.704 | 0.635 | 0.619 | 0.566 | 0.804 | 0.60 | |
| | | | | | | | | | | | |
| | MLP(sklearn) | (50, 150) | 0.634 | 0.092 | 0.678 | 0.634 | 0.598 | 0.550 | 0.857 | 0.61 | |
| | Random Forest | (50, 150) | 0.628 | 0.093 | 0.665 | 0.628 | 0.594 | 0.542 | 0.849 | 0.60 | |
| | Rotation Forest | (50, 150) | 0.619 | 0.096 | 0.644 | 0.619 | 0.581 | 0.526 | 0.845 | 0.59 | |
| Bagging | (50, 150) | 0.612 | 0.097 | 0.644 | 0.612 | 0.58 | 0.522 | 0.843 | 0.59 | | |
| J48 | (50, 150) | 0.622 | 0.095 | 0.665 | 0.622 | 0.589 | 0.536 | 0.837 | 0.57 | | |
| PBC4cip | (50, 150) | 0.620 | 0.095 | 0.693 | 0.620 | 0.582 | 0.537 | 0.833 | 0.57 | | |
| GNB(sklearn) | (50, 150) | 0.528 | 0.118 | 0.555 | 0.528 | 0.467 | 0.428 | 0.832 | 0.57 | | |
| MHLDTForest | (50, 150) | 0.626 | 0.094 | 0.696 | 0.626 | 0.581 | 0.538 | 0.823 | 0.59 | | |



Tabla 28. Resultados de clasificadores del AUC en el dataset [14] usando BoW

| AUC | D-1 | D-2 | D-3 | D-4 |
|---|----------------|----------------|----------------|---------------|
| Clasificadores | (K=100, L=150) | (K=150, L=200) | (K=100, L=200) | (K=50, L=150) |
| Acelerómetro, combinación de los ejes X,Y,Z | | | | |
| MHLDForest | 0.778 | 0.776 | 0.746 | 0.716 |
| Random Forest | 0.855 | 0.853 | 0.835 | 0.817 |
| Rotation Forest | 0.855 | 0.854 | 0.834 | 0.815 |
| PBC4cip | 0.713 | 0.748 | 0.694 | 0.672 |
| Bagging | 0.843 | 0.839 | 0.825 | 0.812 |
| J48 | 0.836 | 0.835 | 0.818 | 0.803 |
| MLP(sklearn) | 0.838 | 0.833 | 0.827 | 0.813 |
| GNB(sklearn) | 0.675 | 0.675 | 0.662 | 0.669 |
| Acelerómetro, combinación de los ejes X,Y | | | | |
| MHLDForest | 0.797 | 0.828 | 0.816 | 0.816 |
| Random Forest | 0.898 | 0.907 | 0.892 | 0.887 |
| Rotation Forest | 0.894 | 0.904 | 0.888 | 0.884 |
| PBC4cip | 0.803 | 0.823 | 0.793 | 0.789 |
| Bagging | 0.880 | 0.888 | 0.872 | 0.878 |
| J48 | 0.876 | 0.891 | 0.878 | 0.875 |
| MLP | 0.885 | 0.882 | 0.879 | 0.884 |
| GNB(sklearn) | 0.718 | 0.709 | 0.714 | 0.724 |
| Acelerómetro, norma de $\ X, Y\ _2$ | | | | |
| MHLDForest | 0.876 | 0.876 | 0.874 | 0.871 |
| Random Forest | 0.896 | 0.899 | 0.897 | 0.883 |
| Rotation Forest | 0.899 | 0.898 | 0.899 | 0.890 |
| PBC4cip | 0.905 | 0.921 | 0.902 | 0.846 |
| Bagging | 0.882 | 0.882 | 0.885 | 0.885 |
| J48 | 0.876 | 0.868 | 0.879 | 0.872 |
| MLP | 0.911 | 0.913 | 0.917 | 0.892 |
| GNB(sklearn) | 0.862 | 0.891 | 0.890 | 0.853 |
| Acelerómetro eje X | | | | |
| MHLDForest | 0.839 | 0.837 | 0.813 | 0.810 |
| Random Forest | 0.872 | 0.886 | 0.883 | 0.878 |
| Rotation Forest | 0.874 | 0.883 | 0.883 | 0.877 |
| PBC4cip | 0.810 | 0.835 | 0.835 | 0.814 |
| Bagging | 0.869 | 0.870 | 0.880 | 0.874 |
| J48 | 0.867 | 0.867 | 0.875 | 0.868 |
| MLP | 0.863 | 0.881 | 0.870 | 0.878 |
| GNB(sklearn) | 0.821 | 0.818 | 0.814 | 0.846 |
| Acelerómetro eje Y | | | | |
| MHLDForest | 0.823 | 0.840 | 0.804 | 0.823 |
| Random Forest | 0.858 | 0.888 | 0.867 | 0.849 |



| | | | | |
|-----------------|-------|-------|-------|-------|
| Rotation Forest | 0.857 | 0.884 | 0.864 | 0.845 |
| PBC4cip | 0.828 | 0.815 | 0.818 | 0.833 |
| Bagging | 0.851 | 0.874 | 0.857 | 0.843 |
| J48 | 0.848 | 0.863 | 0.854 | 0.837 |
| MLP | 0.855 | 0.886 | 0.870 | 0.857 |
| GNB(sklearn) | 0.828 | 0.840 | 0.831 | 0.832 |

Resumen de mejores resultados:

Tabla 29. Resultados sobre la norma $\|X, Y\|_2$ y el eje X , al aplicar el filtro ADLF en [14]

| | Clasificador | (K, L) | TP Rate | FP Rate | Precision | Recall | F-1 | MCC | AUC | PRC |
|--------------|--------------|------------|---------|---------|-----------|--------|-------|-------|--------------|-------|
| $\ X, Y\ _2$ | MLP(sklearn) | (100, 150) | 0.778 | 0.055 | 0.779 | 0.778 | 0.776 | 0.723 | 0.950 | 0.850 |
| | GNB(sklearn) | (100, 150) | 0.773 | 0.057 | 0.774 | 0.773 | 0.771 | 0.716 | 0.941 | 0.836 |
| | Bagging | (100, 150) | 0.758 | 0.061 | 0.756 | 0.758 | 0.754 | 0.696 | 0.941 | 0.823 |
| | Rotation | (100, 150) | 0.766 | 0.059 | 0.767 | 0.766 | 0.763 | 0.707 | 0.936 | 0.795 |
| | Random | (100, 150) | 0.747 | 0.063 | 0.745 | 0.747 | 0.744 | 0.683 | 0.930 | 0.808 |
| | PBC4cip | (100, 150) | 0.743 | 0.064 | 0.741 | 0.743 | 0.739 | 0.678 | 0.905 | 0.772 |
| | J48 | (100, 150) | 0.765 | 0.059 | 0.769 | 0.765 | 0.761 | 0.707 | 0.904 | 0.715 |
| | MHLDTFore | (100, 150) | 0.737 | 0.066 | 0.732 | 0.737 | 0.733 | 0.668 | 0.857 | 0.646 |
| | MLP(sklearn) | (150, 200) | 0.829 | 0.043 | 0.828 | 0.829 | 0.828 | 0.786 | 0.971 | 0.91 |
| | GNB(sklearn) | (150, 200) | 0.821 | 0.045 | 0.821 | 0.821 | 0.820 | 0.776 | 0.964 | 0.900 |
| | Bagging | (150, 200) | 0.811 | 0.047 | 0.810 | 0.811 | 0.809 | 0.763 | 0.958 | 0.872 |
| | Rotation | (150, 200) | 0.822 | 0.044 | 0.821 | 0.822 | 0.820 | 0.777 | 0.947 | 0.828 |
| | Random | (150, 200) | 0.807 | 0.048 | 0.805 | 0.807 | 0.805 | 0.758 | 0.947 | 0.861 |
| | J48 | (150, 200) | 0.813 | 0.047 | 0.811 | 0.813 | 0.81 | 0.765 | 0.926 | 0.768 |
| | PBC4cip | (150, 200) | 0.805 | 0.049 | 0.803 | 0.805 | 0.801 | 0.755 | 0.921 | 0.828 |
| | MHLDTFore | (150, 200) | 0.803 | 0.049 | 0.802 | 0.803 | 0.801 | 0.753 | 0.883 | 0.698 |
| | MLP(sklearn) | (100, 200) | 0.789 | 0.053 | 0.787 | 0.789 | 0.786 | 0.735 | 0.953 | 0.857 |
| | GNB(sklearn) | (100, 200) | 0.787 | 0.053 | 0.787 | 0.787 | 0.785 | 0.734 | 0.943 | 0.847 |
| | Bagging | (100, 200) | 0.781 | 0.055 | 0.781 | 0.781 | 0.776 | 0.725 | 0.938 | 0.814 |
| | Random | (100, 200) | 0.762 | 0.060 | 0.760 | 0.762 | 0.757 | 0.701 | 0.928 | 0.798 |
| | Rotation | (100, 200) | 0.785 | 0.054 | 0.787 | 0.785 | 0.781 | 0.731 | 0.925 | 0.769 |
| | J48 | (100, 200) | 0.787 | 0.053 | 0.791 | 0.787 | 0.782 | 0.735 | 0.906 | 0.719 |
| | PBC4cip | (100, 200) | 0.754 | 0.062 | 0.751 | 0.754 | 0.749 | 0.691 | 0.901 | 0.754 |
| | MHLDTFore | (100, 200) | 0.754 | 0.062 | 0.751 | 0.754 | 0.749 | 0.691 | 0.854 | 0.635 |
| | MLP(sklearn) | (50, 150) | 0.734 | 0.066 | 0.733 | 0.734 | 0.731 | 0.667 | 0.915 | 0.755 |



| | | | | | | | | | | |
|---------------------|---------------------|---------------|----------------|----------------|------------------|---------------|------------|--------------|--------------|------------|
| X, Y ₂ | GNB(sklearn) | (50, 150) | 0.725 | 0.069 | 0.724 | 0.725 | 0.722 | 0.655 | 0.905 | 0.744 |
| | Bagging | (50, 150) | 0.741 | 0.065 | 0.739 | 0.741 | 0.736 | 0.674 | 0.905 | 0.721 |
| | Rotation | (50, 150) | 0.747 | 0.063 | 0.746 | 0.747 | 0.743 | 0.683 | 0.898 | 0.713 |
| | Random | (50, 150) | 0.734 | 0.067 | 0.73 | 0.734 | 0.73 | 0.665 | 0.894 | 0.716 |
| | J48 | (50, 150) | 0.743 | 0.064 | 0.742 | 0.743 | 0.739 | 0.678 | 0.884 | 0.666 |
| | PBC4cip | (50, 150) | 0.721 | 0.070 | 0.717 | 0.721 | 0.715 | 0.648 | 0.866 | 0.659 |
| | MHLDTFore | (50, 150) | 0.722 | 0.070 | 0.716 | 0.722 | 0.716 | 0.649 | 0.831 | 0.591 |
| Ejes | Clasificador | (K, L) | TP Rate | FP Rate | Precision | Recall | F-1 | MCC | AUC | PRC |
| (X) | MLP(sklearn) | (100, 150) | 0.652 | 0.087 | 0.666 | 0.652 | 0.647 | 0.571 | 0.902 | 0.696 |
| | Bagging | (100, 150) | 0.655 | 0.086 | 0.677 | 0.655 | 0.621 | 0.572 | 0.892 | 0.668 |
| | Random | (100, 150) | 0.652 | 0.087 | 0.677 | 0.652 | 0.621 | 0.571 | 0.890 | 0.674 |
| | Rotation | (100, 150) | 0.655 | 0.086 | 0.718 | 0.655 | 0.602 | 0.569 | 0.887 | 0.649 |
| | GNB(sklearn) | (100, 150) | 0.635 | 0.091 | 0.614 | 0.635 | 0.585 | 0.531 | 0.883 | 0.641 |
| | J48 | (100, 150) | 0.651 | 0.087 | 0.665 | 0.651 | 0.615 | 0.564 | 0.875 | 0.587 |
| | PBC4cip | (100, 150) | 0.655 | 0.086 | 0.744 | 0.655 | 0.605 | 0.577 | 0.858 | 0.602 |
| | MHLDTFore | (100, 150) | 0.654 | 0.087 | 0.717 | 0.654 | 0.605 | 0.572 | 0.784 | 0.537 |
| | | | | | | | | | | |
| | MLP(sklearn) | (150, 200) | 0.701 | 0.075 | 0.801 | 0.701 | 0.655 | 0.636 | 0.920 | 0.74 |
| | Random | (150, 200) | 0.701 | 0.075 | 0.795 | 0.701 | 0.661 | 0.642 | 0.914 | 0.727 |
| | Bagging | (150, 200) | 0.701 | 0.075 | 0.791 | 0.701 | 0.656 | 0.637 | 0.913 | 0.718 |
| | Rotation | (150, 200) | 0.695 | 0.076 | 0.786 | 0.695 | 0.651 | 0.631 | 0.910 | 0.702 |
| | GNB(sklearn) | (150, 200) | 0.683 | 0.079 | 0.692 | 0.683 | 0.676 | 0.607 | 0.909 | 0.694 |
| | J48 | (150, 200) | 0.693 | 0.077 | 0.769 | 0.693 | 0.648 | 0.625 | 0.900 | 0.656 |
| | PBC4cip | (150, 200) | 0.701 | 0.075 | 0.808 | 0.701 | 0.660 | 0.644 | 0.885 | 0.654 |
| | MHLDTFore | (150, 200) | 0.698 | 0.076 | 0.810 | 0.698 | 0.659 | 0.643 | 0.810 | 0.587 |
| | | | | | | | | | | |
| | MLP(sklearn) | (100, 200) | 0.674 | 0.082 | 0.792 | 0.674 | 0.619 | 0.596 | 0.908 | 0.706 |
| | Bagging | (100, 200) | 0.666 | 0.084 | 0.784 | 0.666 | 0.613 | 0.589 | 0.903 | 0.691 |
| | GNB(sklearn) | (100, 200) | 0.674 | 0.082 | 0.757 | 0.674 | 0.62 | 0.597 | 0.901 | 0.689 |
| | Random | (100, 200) | 0.666 | 0.084 | 0.784 | 0.666 | 0.619 | 0.598 | 0.899 | 0.692 |
| | Rotation | (100, 200) | 0.670 | 0.083 | 0.762 | 0.670 | 0.620 | 0.597 | 0.899 | 0.672 |
| | J48 | (100, 200) | 0.668 | 0.083 | 0.753 | 0.668 | 0.616 | 0.591 | 0.889 | 0.628 |
| | PBC4cip | (100, 200) | 0.667 | 0.083 | 0.786 | 0.667 | 0.621 | 0.600 | 0.841 | 0.566 |
| | MHLDTFore | (100, 200) | 0.662 | 0.085 | 0.780 | 0.662 | 0.615 | 0.593 | 0.790 | 0.542 |
| | | | | | | | | | | |
| | MLP(sklearn) | (50, 150) | 0.632 | 0.092 | 0.622 | 0.632 | 0.597 | 0.531 | 0.893 | 0.664 |
| J48 | (50, 150) | 0.668 | 0.083 | 0.753 | 0.668 | 0.616 | 0.591 | 0.889 | 0.628 | |
| Bagging | (50, 150) | 0.615 | 0.096 | 0.607 | 0.615 | 0.585 | 0.513 | 0.885 | 0.653 | |
| Rotation | (50, 150) | 0.616 | 0.096 | 0.544 | 0.616 | 0.562 | 0.499 | 0.884 | 0.639 | |
| Random | (50, 150) | 0.604 | 0.099 | 0.596 | 0.604 | 0.576 | 0.5 | 0.883 | 0.646 | |
| GNB(sklearn) | (50, 150) | 0.624 | 0.094 | ? | 0.624 | ? | ? | 0.880 | 0.633 | |



| | | | | | | | | | | |
|--|------------------|-----------|-------|-------|-------|-------|-------|-------|--------------|-------|
| | PBC4cip | (50, 150) | 0.612 | 0.097 | 0.594 | 0.612 | 0.564 | 0.503 | 0.837 | 0.564 |
| | MHLDTFore | (50, 150) | 0.618 | 0.096 | 0.606 | 0.618 | 0.570 | 0.512 | 0.761 | 0.490 |

ANEXO 5. Resultado de la experimentación con la metodología de Ferreira et al. [10] en el dataset propuesto por Bejani y Ghatte [13].

Se utilizaron los mismos parámetros de optimización expuestos en el Anexo 2.

Tabla 30. Resultados de clasificadores en el dataset de Bejani y Ghatte [36]

| Clasificadores | nf | TP Rate | FP Rate | Precision | Recall | F-1 | MCC | AUC | PRC |
|-----------------------------|----|---------|---------|-----------|--------|-------|-------|--------------|-------|
| Acelerómetro, nf = 2 | | | | | | | | | |
| MHLDForest | 2 | 0.567 | 0.246 | 0.558 | 0.567 | 0.551 | 0.336 | 0.766 | 0.600 |
| Random Forest | 2 | 0.562 | 0.249 | 0.555 | 0.562 | 0.548 | 0.330 | 0.763 | 0.598 |
| Rotation Forest | 2 | 0.569 | 0.245 | 0.560 | 0.569 | 0.554 | 0.341 | 0.769 | 0.605 |
| PBC4cip | 2 | 0.497 | 0.163 | 0.563 | 0.497 | 0.508 | 0.329 | 0.764 | 0.594 |
| Bagging | 2 | 0.557 | 0.255 | 0.552 | 0.557 | 0.538 | 0.321 | 0.759 | 0.589 |
| J48 | 2 | 0.475 | 0.258 | 0.469 | 0.475 | 0.472 | 0.219 | 0.615 | 0.407 |
| MLP | 2 | 0.518 | 0.261 | 0.503 | 0.518 | 0.504 | 0.265 | 0.696 | 0.488 |
| GNB(sklearn) | 2 | 0.339 | 0.213 | 0.425 | 0.339 | 0.331 | 0.123 | 0.615 | 0.410 |
| Acelerómetro, nf = 4 | | | | | | | | | |
| MHLDForest | 4 | 0.628 | 0.214 | 0.623 | 0.628 | 0.614 | 0.432 | 0.836 | 0.702 |
| Random Forest | 4 | 0.630 | 0.212 | 0.624 | 0.630 | 0.618 | 0.435 | 0.828 | 0.692 |
| Rotation Forest | 4 | 0.641 | 0.201 | 0.635 | 0.641 | 0.631 | 0.455 | 0.838 | 0.704 |
| PBC4cip | 4 | 0.568 | 0.137 | 0.633 | 0.568 | 0.578 | 0.425 | 0.821 | 0.680 |
| Bagging | 4 | 0.612 | 0.218 | 0.605 | 0.612 | 0.600 | 0.410 | 0.814 | 0.667 |
| J48 | 4 | 0.524 | 0.229 | 0.524 | 0.524 | 0.524 | 0.296 | 0.657 | 0.443 |
| MLP | 4 | 0.527 | 0.249 | 0.515 | 0.527 | 0.518 | 0.284 | 0.704 | 0.504 |
| GNB(sklearn) | 4 | 0.355 | 0.197 | 0.456 | 0.355 | 0.350 | 0.158 | 0.632 | 0.431 |
| Acelerómetro, nf = 6 | | | | | | | | | |
| MHLDForest | 6 | 0.713 | 0.167 | 0.715 | 0.713 | 0.705 | 0.566 | 0.895 | 0.794 |
| Random Forest | 6 | 0.703 | 0.171 | 0.702 | 0.703 | 0.694 | 0.550 | 0.885 | 0.777 |
| Rotation Forest | 6 | 0.715 | 0.158 | 0.710 | 0.715 | 0.707 | 0.571 | 0.897 | 0.797 |
| PBC4cip | 6 | 0.627 | 0.115 | 0.688 | 0.627 | 0.637 | 0.505 | 0.867 | 0.627 |
| Bagging | 6 | 0.670 | 0.188 | 0.666 | 0.670 | 0.659 | 0.500 | 0.863 | 0.740 |
| J48 | 6 | 0.587 | 0.194 | 0.587 | 0.587 | 0.587 | 0.392 | 0.717 | 0.514 |
| MLP | 6 | 0.566 | 0.220 | 0.561 | 0.566 | 0.561 | 0.349 | 0.741 | 0.558 |
| GNB(sklearn) | 6 | 0.364 | 0.191 | 0.469 | 0.364 | 0.36 | 0.173 | 0.644 | 0.444 |
| Acelerómetro, nf = 7 | | | | | | | | | |
| MHLDForest | 7 | 0.754 | 0.143 | 0.755 | 0.754 | 0.746 | 0.630 | 0.918 | 0.836 |
| Random Forest | 7 | 0.731 | 0.155 | 0.731 | 0.731 | 0.724 | 0.595 | 0.907 | 0.814 |
| Rotation Forest | 7 | 0.750 | 0.140 | 0.748 | 0.750 | 0.743 | 0.625 | 0.920 | 0.836 |



| | | | | | | | | | |
|------------------------------|----|-------|-------|-------|-------|-------|-------|--------------|-------|
| PBC4cip | 7 | 0.655 | 0.101 | 0.720 | 0.655 | 0.664 | 0.546 | 0.887 | 0.782 |
| Bagging | 7 | 0.686 | 0.176 | 0.682 | 0.686 | 0.677 | 0.526 | 0.881 | 0.770 |
| J48 | 7 | 0.613 | 0.183 | 0.613 | 0.613 | 0.613 | 0.430 | 0.730 | 0.527 |
| MLP | 7 | 0.573 | 0.220 | 0.566 | 0.573 | 0.567 | 0.358 | 0.751 | 0.569 |
| GNB(sklearn) | 7 | 0.365 | 0.191 | 0.469 | 0.365 | 0.360 | 0.173 | 0.649 | 0.450 |
| Acelerómetro, nf = 8 | | | | | | | | | |
| MHLDForest | 8 | 0.786 | 0.125 | 0.787 | 0.786 | 0.780 | 0.679 | 0.937 | 0.869 |
| Random Forest | 8 | 0.760 | 0.135 | 0.759 | 0.760 | 0.754 | 0.641 | 0.925 | 0.845 |
| Rotation Forest | 8 | 0.779 | 0.125 | 0.779 | 0.779 | 0.774 | 0.670 | 0.938 | 0.870 |
| PBC4cip | 8 | 0.674 | 0.094 | 0.737 | 0.674 | 0.683 | 0.572 | 0.901 | 0.807 |
| Bagging | 8 | 0.710 | 0.167 | 0.708 | 0.710 | 0.700 | 0.561 | 0.895 | 0.793 |
| J48 | 8 | 0.619 | 0.178 | 0.619 | 0.619 | 0.619 | 0.441 | 0.736 | 0.534 |
| MLP | 8 | 0.607 | 0.199 | 0.601 | 0.607 | 0.602 | 0.413 | 0.781 | 0.608 |
| GNB(sklearn) | 8 | 0.367 | 0.191 | 0.469 | 0.367 | 0.363 | 0.175 | 0.654 | 0.456 |
| Acelerómetro, nf = 9 | | | | | | | | | |
| MHLDForest | 9 | 0.809 | 0.111 | 0.810 | 0.809 | 0.803 | 0.714 | 0.951 | 0.895 |
| Random Forest | 9 | 0.783 | 0.126 | 0.784 | 0.783 | 0.776 | 0.675 | 0.939 | 0.871 |
| Rotation Forest | 9 | 0.805 | 0.112 | 0.806 | 0.805 | 0.800 | 0.709 | 0.951 | 0.895 |
| PBC4cip | 9 | 0.698 | 0.085 | 0.759 | 0.698 | 0.707 | 0.604 | 0.914 | 0.830 |
| Bagging | 9 | 0.731 | 0.153 | 0.730 | 0.731 | 0.723 | 0.595 | 0.906 | 0.814 |
| J48 | 9 | 0.637 | 0.172 | 0.637 | 0.637 | 0.637 | 0.465 | 0.758 | 0.562 |
| MLP | 9 | 0.626 | 0.181 | 0.624 | 0.626 | 0.625 | 0.446 | 0.800 | 0.638 |
| GNB(sklearn) | 9 | 0.372 | 0.188 | 0.474 | 0.372 | 0.368 | 0.183 | 0.658 | 0.461 |
| Acelerómetro, nf = 10 | | | | | | | | | |
| MHLDForest | 10 | 0.837 | 0.096 | 0.839 | 0.837 | 0.833 | 0.757 | 0.962 | 0.918 |
| Random Forest | 10 | 0.806 | 0.112 | 0.807 | 0.806 | 0.801 | 0.711 | 0.950 | 0.891 |
| Rotation Forest | 10 | 0.832 | 0.094 | 0.832 | 0.832 | 0.828 | 0.750 | 0.960 | 0.912 |
| PBC4cip | 10 | 0.722 | 0.076 | 0.782 | 0.722 | 0.731 | 0.638 | 0.926 | 0.853 |
| Bagging | 10 | 0.746 | 0.146 | 0.746 | 0.746 | 0.739 | 0.619 | 0.916 | 0.830 |
| J48 | 10 | 0.651 | 0.165 | 0.651 | 0.651 | 0.651 | 0.486 | 0.755 | 0.562 |
| MLP | 10 | 0.660 | 0.169 | 0.657 | 0.660 | 0.657 | 0.494 | 0.828 | 0.681 |
| GNB(sklearn) | 10 | 0.373 | 0.188 | 0.471 | 0.373 | 0.367 | 0.181 | 0.662 | 0.466 |
| Acelerómetro, nf = 11 | | | | | | | | | |
| MHLDForest | 11 | 0.857 | 0.082 | 0.859 | 0.857 | 0.855 | 0.788 | 0.970 | 0.934 |
| Random Forest | 11 | 0.828 | 0.099 | 0.830 | 0.828 | 0.824 | 0.744 | 0.959 | 0.908 |
| Rotation Forest | 11 | 0.851 | 0.083 | 0.851 | 0.851 | 0.848 | 0.779 | 0.968 | 0.928 |
| PBC4cip | 11 | 0.741 | 0.067 | 0.801 | 0.741 | 0.751 | 0.665 | 0.936 | 0.871 |
| Bagging | 11 | 0.761 | 0.136 | 0.761 | 0.761 | 0.755 | 0.642 | 0.923 | 0.844 |
| J48 | 11 | 0.673 | 0.159 | 0.671 | 0.673 | 0.672 | 0.516 | 0.772 | 0.581 |
| MLP | 11 | 0.695 | 0.154 | 0.691 | 0.695 | 0.692 | 0.546 | 0.847 | 0.710 |
| GNB(sklearn) | 11 | 0.378 | 0.185 | 0.478 | 0.378 | 0.374 | 0.19 | 0.665 | 0.47 |
| Magnetómetro, nf = 2 | | | | | | | | | |
| MHLDForest | 2 | 0.817 | 0.086 | 0.817 | 0.817 | 0.813 | 0.738 | 0.964 | 0.902 |



| | | | | | | | | | |
|-----------------------------|---|-------|-------|-------|-------|-------|-------|--------------|-------|
| Random Forest | 2 | 0.806 | 0.093 | 0.805 | 0.806 | 0.800 | 0.721 | 0.960 | 0.893 |
| Rotation Forest | 2 | 0.820 | 0.085 | 0.820 | 0.820 | 0.815 | 0.743 | 0.961 | 0.897 |
| PBC4cip | 2 | 0.792 | 0.068 | 0.811 | 0.792 | 0.797 | 0.717 | 0.955 | 0.884 |
| Bagging | 2 | 0.772 | 0.111 | 0.775 | 0.772 | 0.762 | 0.673 | 0.944 | 0.857 |
| J48 | 2 | 0.736 | 0.115 | 0.733 | 0.736 | 0.734 | 0.623 | 0.851 | 0.657 |
| MLP | 2 | 0.740 | 0.114 | 0.736 | 0.740 | 0.734 | 0.630 | 0.910 | 0.788 |
| GNB(sklearn) | 2 | 0.602 | 0.124 | 0.661 | 0.602 | 0.616 | 0.476 | 0.848 | 0.687 |
| Magnetómetro, nf = 4 | | | | | | | | | |
| MHLDForest | 4 | 0.855 | 0.068 | 0.855 | 0.855 | 0.852 | 0.793 | 0.975 | 0.932 |
| Random Forest | 4 | 0.844 | 0.074 | 0.845 | 0.844 | 0.841 | 0.777 | 0.972 | 0.923 |
| Rotation Forest | 4 | 0.849 | 0.070 | 0.848 | 0.849 | 0.846 | 0.785 | 0.973 | 0.926 |
| PBC4cip | 4 | 0.818 | 0.060 | 0.832 | 0.818 | 0.822 | 0.751 | 0.965 | 0.908 |
| Bagging | 4 | 0.795 | 0.099 | 0.799 | 0.795 | 0.787 | 0.707 | 0.954 | 0.879 |
| J48 | 4 | 0.764 | 0.102 | 0.761 | 0.764 | 0.762 | 0.665 | 0.868 | 0.707 |
| MLP | 4 | 0.744 | 0.108 | 0.742 | 0.744 | 0.740 | 0.638 | 0.916 | 0.796 |
| GNB(sklearn) | 4 | 0.595 | 0.115 | 0.674 | 0.595 | 0.609 | 0.48 | 0.851 | 0.688 |
| Magnetómetro, nf = 6 | | | | | | | | | |
| MHLDForest | 6 | 0.887 | 0.054 | 0.887 | 0.887 | 0.885 | 0.839 | 0.983 | 0.952 |
| Random Forest | 6 | 0.863 | 0.064 | 0.863 | 0.863 | 0.861 | 0.804 | 0.979 | 0.943 |
| Rotation Forest | 6 | 0.872 | 0.060 | 0.871 | 0.872 | 0.870 | 0.817 | 0.980 | 0.944 |
| PBC4cip | 6 | 0.845 | 0.049 | 0.859 | 0.845 | 0.849 | 0.788 | 0.973 | 0.929 |
| Bagging | 6 | 0.807 | 0.093 | 0.811 | 0.807 | 0.801 | 0.724 | 0.961 | 0.897 |
| J48 | 6 | 0.784 | 0.091 | 0.784 | 0.784 | 0.784 | 0.693 | 0.876 | 0.727 |
| MLP | 6 | 0.759 | 0.104 | 0.756 | 0.759 | 0.755 | 0.659 | 0.919 | 0.804 |
| GNB(sklearn) | 6 | 0.592 | 0.114 | 0.676 | 0.592 | 0.606 | 0.479 | 0.853 | 0.684 |
| Magnetómetro, nf = 7 | | | | | | | | | |
| MHLDForest | 7 | 0.900 | 0.046 | 0.900 | 0.900 | 0.899 | 0.858 | 0.986 | 0.960 |
| Random Forest | 7 | 0.881 | 0.058 | 0.881 | 0.881 | 0.879 | 0.830 | 0.982 | 0.952 |
| Rotation Forest | 7 | 0.885 | 0.053 | 0.884 | 0.885 | 0.883 | 0.836 | 0.983 | 0.952 |
| PBC4cip | 7 | 0.858 | 0.045 | 0.869 | 0.858 | 0.861 | 0.805 | 0.976 | 0.937 |
| Bagging | 7 | 0.814 | 0.088 | 0.817 | 0.814 | 0.81 | 0.734 | 0.964 | 0.905 |
| J48 | 7 | 0.789 | 0.091 | 0.788 | 0.789 | 0.788 | 0.699 | 0.877 | 0.733 |
| MLP | 7 | 0.764 | 0.096 | 0.763 | 0.764 | 0.763 | 0.668 | 0.927 | 0.818 |
| GNB(sklearn) | 7 | 0.593 | 0.111 | 0.681 | 0.593 | 0.607 | 0.484 | 0.853 | 0.684 |
| Magnetómetro, nf = 8 | | | | | | | | | |
| MHLDForest | 8 | 0.912 | 0.043 | 0.912 | 0.912 | 0.910 | 0.874 | 0.988 | 0.966 |
| Random Forest | 8 | 0.894 | 0.051 | 0.893 | 0.894 | 0.892 | 0.848 | 0.985 | 0.959 |
| Rotation Forest | 8 | 0.888 | 0.052 | 0.888 | 0.888 | 0.886 | 0.840 | 0.985 | 0.959 |
| PBC4cip | 8 | 0.864 | 0.043 | 0.874 | 0.864 | 0.866 | 0.812 | 0.978 | 0.942 |
| Bagging | 8 | 0.807 | 0.093 | 0.811 | 0.807 | 0.801 | 0.724 | 0.961 | 0.897 |
| J48 | 8 | 0.800 | 0.083 | 0.800 | 0.800 | 0.800 | 0.718 | 0.884 | 0.742 |
| MLP | 8 | 0.781 | 0.089 | 0.781 | 0.781 | 0.781 | 0.694 | 0.929 | 0.823 |
| GNB(sklearn) | 8 | 0.590 | 0.109 | 0.683 | 0.590 | 0.604 | 0.482 | 0.853 | 0.684 |



| Magnetómetro, nf = 9 | | | | | | | | | |
|---|----|-------|-------|-------|-------|-------|-------|--------------|-------|
| MHLDForest | 9 | 0.916 | 0.040 | 0.915 | 0.916 | 0.915 | 0.880 | 0.990 | 0.972 |
| Random Forest | 9 | 0.901 | 0.048 | 0.900 | 0.901 | 0.899 | 0.858 | 0.987 | 0.964 |
| Rotation Forest | 9 | 0.899 | 0.047 | 0.899 | 0.899 | 0.897 | 0.856 | 0.988 | 0.966 |
| PBC4cip | 9 | 0.874 | 0.040 | 0.884 | 0.874 | 0.876 | 0.826 | 0.980 | 0.950 |
| Bagging | 9 | 0.828 | 0.083 | 0.832 | 0.828 | 0.824 | 0.754 | 0.969 | 0.919 |
| J48 | 9 | 0.813 | 0.08 | 0.813 | 0.813 | 0.813 | 0.733 | 0.887 | 0.752 |
| MLP | 9 | 0.795 | 0.085 | 0.794 | 0.795 | 0.794 | 0.712 | 0.932 | 0.831 |
| GNB(sklearn) | 9 | 0.586 | 0.107 | 0.686 | 0.586 | 0.601 | 0.481 | 0.853 | 0.681 |
| Magnetómetro, nf = 10 | | | | | | | | | |
| MHLDForest | 10 | 0.925 | 0.036 | 0.925 | 0.925 | 0.924 | 0.893 | 0.991 | 0.976 |
| Random Forest | 10 | 0.906 | 0.045 | 0.906 | 0.906 | 0.904 | 0.866 | 0.989 | 0.968 |
| Rotation Forest | 10 | 0.912 | 0.041 | 0.912 | 0.912 | 0.911 | 0.874 | 0.990 | 0.971 |
| PBC4cip | 10 | 0.882 | 0.037 | 0.891 | 0.882 | 0.884 | 0.837 | 0.982 | 0.954 |
| Bagging | 10 | 0.832 | 0.081 | 0.835 | 0.832 | 0.829 | 0.758 | 0.971 | 0.926 |
| J48 | 10 | 0.812 | 0.080 | 0.811 | 0.812 | 0.812 | 0.733 | 0.887 | 0.746 |
| MLP | 10 | 0.803 | 0.085 | 0.801 | 0.803 | 0.801 | 0.722 | 0.941 | 0.847 |
| GNB(sklearn) | 10 | 0.585 | 0.107 | 0.688 | 0.585 | 0.600 | 0.481 | 0.852 | 0.679 |
| Magnetómetro, nf = 11 | | | | | | | | | |
| MHLDForest | 11 | 0.928 | 0.034 | 0.928 | 0.928 | 0.927 | 0.898 | 0.993 | 0.980 |
| Random Forest | 11 | 0.917 | 0.040 | 0.917 | 0.917 | 0.916 | 0.881 | 0.990 | 0.971 |
| Rotation Forest | 11 | 0.913 | 0.041 | 0.912 | 0.913 | 0.912 | 0.875 | 0.990 | 0.973 |
| PBC4cip | 11 | 0.883 | 0.035 | 0.893 | 0.883 | 0.885 | 0.839 | 0.984 | 0.958 |
| Bagging | 11 | 0.839 | 0.079 | 0.841 | 0.839 | 0.835 | 0.768 | 0.973 | 0.930 |
| J48 | 11 | 0.818 | 0.079 | 0.816 | 0.818 | 0.817 | 0.741 | 0.892 | 0.760 |
| MLP | 11 | 0.814 | 0.051 | 0.845 | 0.814 | 0.820 | 0.752 | 0.960 | 0.913 |
| GNB(sklearn) | 11 | 0.584 | 0.106 | 0.690 | 0.584 | 0.600 | 0.482 | 0.852 | 0.676 |
| Acelerómetro- Magnetómetro-Combinado, nf = 2 | | | | | | | | | |
| MHLDForest | 2 | 0.697 | 0.163 | 0.694 | 0.697 | 0.689 | 0.547 | 0.890 | 0.788 |
| Random Forest | 2 | 0.687 | 0.169 | 0.683 | 0.687 | 0.678 | 0.530 | 0.885 | 0.778 |
| Rotation Forest | 2 | 0.693 | 0.165 | 0.690 | 0.693 | 0.686 | 0.541 | 0.886 | 0.781 |
| PBC4cip | 2 | 0.650 | 0.114 | 0.697 | 0.650 | 0.661 | 0.531 | 0.880 | 0.769 |
| Bagging | 2 | 0.667 | 0.182 | 0.664 | 0.667 | 0.654 | 0.500 | 0.871 | 0.749 |
| J48 | 2 | 0.604 | 0.181 | 0.602 | 0.604 | 0.603 | 0.424 | 0.745 | 0.527 |
| MLP | 2 | 0.595 | 0.250 | 0.586 | 0.595 | 0.561 | 0.371 | 0.798 | 0.639 |
| GNB(sklearn) | 2 | 0.406 | 0.153 | 0.574 | 0.406 | 0.432 | 0.280 | 0.663 | 0.520 |
| Acelerómetro- Magnetómetro-Combinado, nf = 4 | | | | | | | | | |
| MHLDForest | 4 | 0.746 | 0.140 | 0.744 | 0.746 | 0.739 | 0.621 | 0.922 | 0.842 |
| Random Forest | 4 | 0.735 | 0.145 | 0.734 | 0.735 | 0.728 | 0.604 | 0.917 | 0.832 |
| Rotation Forest | 4 | 0.740 | 0.138 | 0.737 | 0.740 | 0.734 | 0.613 | 0.919 | 0.835 |
| PBC4cip | 4 | 0.698 | 0.096 | 0.739 | 0.698 | 0.707 | 0.594 | 0.906 | 0.811 |
| Bagging | 4 | 0.703 | 0.161 | 0.700 | 0.703 | 0.694 | 0.556 | 0.897 | 0.792 |
| J48 | 4 | 0.645 | 0.163 | 0.644 | 0.645 | 0.644 | 0.483 | 0.765 | 0.557 |



| | | | | | | | | | |
|--|----|-------|-------|-------|-------|-------|-------|--------------|-------|
| MLP | 4 | 0.605 | 0.235 | 0.591 | 0.605 | 0.578 | 0.392 | 0.810 | 0.652 |
| GNB(sklearn) | 4 | 0.387 | 0.142 | 0.592 | 0.387 | 0.415 | 0.281 | 0.662 | 0.525 |
| Acelerómetro- Magnetómetro-Combinado, nf = 6 | | | | | | | | | |
| MHLDForest | 6 | 0.797 | 0.111 | 0.796 | 0.797 | 0.792 | 0.698 | 0.947 | 0.888 |
| Random Forest | 6 | 0.783 | 0.119 | 0.782 | 0.783 | 0.778 | 0.677 | 0.942 | 0.875 |
| Rotation Forest | 6 | 0.786 | 0.115 | 0.784 | 0.786 | 0.781 | 0.682 | 0.944 | 0.879 |
| PBC4cip | 6 | 0.737 | 0.081 | 0.775 | 0.737 | 0.744 | 0.647 | 0.928 | 0.849 |
| Bagging | 6 | 0.741 | 0.140 | 0.739 | 0.741 | 0.733 | 0.614 | 0.919 | 0.830 |
| J48 | 6 | 0.679 | 0.143 | 0.680 | 0.679 | 0.679 | 0.536 | 0.784 | 0.586 |
| MLP | 6 | 0.615 | 0.227 | 0.603 | 0.615 | 0.592 | 0.408 | 0.817 | 0.666 |
| GNB(sklearn) | 6 | 0.380 | 0.139 | 0.602 | 0.380 | 0.409 | 0.283 | 0.661 | 0.525 |
| Acelerómetro- Magnetómetro-Combinado, nf = 7 | | | | | | | | | |
| MHLDForest | 7 | 0.817 | 0.101 | 0.816 | 0.817 | 0.812 | 0.728 | 0.958 | 0.907 |
| Random Forest | 7 | 0.808 | 0.106 | 0.808 | 0.808 | 0.804 | 0.715 | 0.952 | 0.895 |
| Rotation Forest | 7 | 0.810 | 0.101 | 0.809 | 0.810 | 0.806 | 0.719 | 0.955 | 0.899 |
| PBC4cip | 7 | 0.754 | 0.075 | 0.789 | 0.754 | 0.761 | 0.669 | 0.936 | 0.867 |
| Bagging | 7 | 0.758 | 0.131 | 0.756 | 0.758 | 0.751 | 0.639 | 0.929 | 0.848 |
| J48 | 7 | 0.712 | 0.132 | 0.712 | 0.712 | 0.712 | 0.580 | 0.810 | 0.633 |
| MLP | 7 | 0.621 | 0.220 | 0.614 | 0.621 | 0.605 | 0.422 | 0.824 | 0.679 |
| GNB(sklearn) | 7 | 0.375 | 0.137 | 0.605 | 0.375 | 0.403 | 0.281 | 0.659 | 0.523 |
| Acelerómetro- Magnetómetro-Combinado, nf = 8 | | | | | | | | | |
| MHLDForest | 8 | 0.845 | 0.086 | 0.845 | 0.845 | 0.842 | 0.771 | 0.967 | 0.925 |
| Random Forest | 8 | 0.823 | 0.097 | 0.824 | 0.823 | 0.819 | 0.738 | 0.960 | 0.910 |
| Rotation Forest | 8 | 0.830 | 0.091 | 0.829 | 0.830 | 0.826 | 0.749 | 0.963 | 0.916 |
| PBC4cip | 8 | 0.773 | 0.068 | 0.807 | 0.773 | 0.779 | 0.695 | 0.944 | 0.882 |
| Bagging | 8 | 0.770 | 0.125 | 0.770 | 0.770 | 0.764 | 0.659 | 0.936 | 0.860 |
| J48 | 8 | 0.722 | 0.127 | 0.721 | 0.722 | 0.721 | 0.595 | 0.820 | 0.646 |
| MLP | 8 | 0.626 | 0.221 | 0.618 | 0.626 | 0.607 | 0.428 | 0.825 | 0.682 |
| GNB(sklearn) | 8 | 0.371 | 0.136 | 0.608 | 0.371 | 0.398 | 0.28 | 0.658 | 0.520 |
| Acelerómetro- Magnetómetro-Combinado, nf = 9 | | | | | | | | | |
| MHLDForest | 9 | 0.860 | 0.077 | 0.860 | 0.860 | 0.857 | 0.794 | 0.973 | 0.937 |
| Random Forest | 9 | 0.845 | 0.086 | 0.846 | 0.845 | 0.841 | 0.771 | 0.967 | 0.925 |
| Rotation Forest | 9 | 0.858 | 0.076 | 0.858 | 0.858 | 0.855 | 0.791 | 0.970 | 0.932 |
| PBC4cip | 9 | 0.786 | 0.062 | 0.820 | 0.786 | 0.792 | 0.713 | 0.951 | 0.894 |
| Bagging | 9 | 0.795 | 0.085 | 0.794 | 0.795 | 0.794 | 0.712 | 0.932 | 0.831 |
| J48 | 9 | 0.742 | 0.120 | 0.741 | 0.742 | 0.741 | 0.623 | 0.835 | 0.674 |
| MLP | 9 | 0.637 | 0.215 | 0.631 | 0.637 | 0.620 | 0.447 | 0.835 | 0.696 |
| GNB(sklearn) | 9 | 0.368 | 0.135 | 0.611 | 0.368 | 0.395 | 0.280 | 0.657 | 0.521 |
| Acelerómetro- Magnetómetro-Combinado, nf = 10 | | | | | | | | | |
| MHLDForest | 10 | 0.879 | 0.066 | 0.879 | 0.879 | 0.876 | 0.822 | 0.978 | 0.948 |
| Random Forest | 10 | 0.855 | 0.080 | 0.855 | 0.855 | 0.852 | 0.786 | 0.971 | 0.933 |
| Rotation Forest | 10 | 0.871 | 0.068 | 0.871 | 0.871 | 0.869 | 0.812 | 0.975 | 0.941 |
| PBC4cip | 10 | 0.800 | 0.056 | 0.834 | 0.800 | 0.806 | 0.734 | 0.957 | 0.906 |



| | | | | | | | | | |
|--|----|-------|-------|-------|-------|-------|-------|--------------|-------|
| Bagging | 10 | 0.797 | 0.109 | 0.797 | 0.797 | 0.792 | 0.700 | 0.947 | 0.882 |
| J48 | 10 | 0.749 | 0.115 | 0.749 | 0.749 | 0.749 | 0.635 | 0.839 | 0.680 |
| MLP | 10 | 0.636 | 0.213 | 0.630 | 0.636 | 0.620 | 0.445 | 0.838 | 0.705 |
| GNB(sklearn) | 10 | 0.365 | 0.134 | 0.615 | 0.365 | 0.391 | 0.279 | 0.656 | 0.519 |
| Acelerómetro- Magnetómetro-Combinado, nf = 11 | | | | | | | | | |
| MHLDForest | 11 | 0.891 | 0.058 | 0.891 | 0.891 | 0.890 | 0.841 | 0.982 | 0.956 |
| Random Forest | 11 | 0.876 | 0.067 | 0.877 | 0.876 | 0.874 | 0.818 | 0.976 | 0.944 |
| Rotation Forest | 11 | 0.880 | 0.063 | 0.879 | 0.880 | 0.878 | 0.825 | 0.979 | 0.949 |
| PBC4cip | 11 | 0.814 | 0.051 | 0.845 | 0.814 | 0.820 | 0.752 | 0.960 | 0.913 |
| Bagging | 11 | 0.806 | 0.105 | 0.807 | 0.806 | 0.801 | 0.714 | 0.951 | 0.889 |
| J48 | 11 | 0.758 | 0.109 | 0.757 | 0.758 | 0.757 | 0.649 | 0.843 | 0.681 |
| MLP | 11 | 0.649 | 0.209 | 0.647 | 0.649 | 0.634 | 0.465 | 0.847 | 0.723 |
| GNB(sklearn) | 11 | 0.363 | 0.133 | 0.617 | 0.363 | 0.390 | 0.279 | 0.655 | 0.518 |

Resumen de resultados de la métrica AUC

Tabla 31. Resultados de clasificadores del AUC en el dataset de Ghaate [36]

| AUC | D-1 | D-2 | D-3 | D-4 | D-5 | D-6 | D-7 | D-8 |
|---|-------|-------|-------|-------|-------|-------|-------|-------|
| Clasificadores | nf=2 | nf=4 | nf=6 | nf=7 | nf=8 | nf=9 | nf=10 | nf=11 |
| Acelerómetro | | | | | | | | |
| MHLDForest | 0.766 | 0.836 | 0.895 | 0.918 | 0.937 | 0.951 | 0.962 | 0.970 |
| Random Forest | 0.763 | 0.828 | 0.885 | 0.907 | 0.925 | 0.939 | 0.950 | 0.959 |
| Rotation Forest | 0.769 | 0.838 | 0.897 | 0.920 | 0.938 | 0.951 | 0.960 | 0.968 |
| PBC4cip | 0.764 | 0.821 | 0.867 | 0.887 | 0.901 | 0.914 | 0.926 | 0.936 |
| Bagging | 0.759 | 0.814 | 0.863 | 0.881 | 0.895 | 0.906 | 0.916 | 0.923 |
| J48 | 0.615 | 0.657 | 0.717 | 0.730 | 0.736 | 0.758 | 0.755 | 0.772 |
| MLP | 0.696 | 0.704 | 0.741 | 0.751 | 0.781 | 0.800 | 0.828 | 0.847 |
| GNB(sklearn) | 0.615 | 0.632 | 0.644 | 0.649 | 0.654 | 0.658 | 0.662 | 0.665 |
| Magnetómetro | | | | | | | | |
| MHLDForest | 0.964 | 0.975 | 0.983 | 0.986 | 0.988 | 0.990 | 0.991 | 0.993 |
| Random Forest | 0.960 | 0.972 | 0.979 | 0.982 | 0.985 | 0.987 | 0.989 | 0.990 |
| Rotation Forest | 0.961 | 0.973 | 0.980 | 0.983 | 0.985 | 0.988 | 0.990 | 0.990 |
| PBC4cip | 0.955 | 0.965 | 0.973 | 0.976 | 0.978 | 0.980 | 0.982 | 0.984 |
| Bagging | 0.944 | 0.954 | 0.961 | 0.964 | 0.961 | 0.969 | 0.971 | 0.973 |
| J48 | 0.851 | 0.868 | 0.876 | 0.877 | 0.884 | 0.887 | 0.887 | 0.892 |
| MLP | 0.910 | 0.916 | 0.919 | 0.927 | 0.929 | 0.932 | 0.941 | 0.960 |
| GNB(sklearn) | 0.848 | 0.851 | 0.853 | 0.853 | 0.853 | 0.853 | 0.852 | 0.852 |
| Acelerómetro- Magnetómetro-Combinado | | | | | | | | |
| MHLDForest | 0.890 | 0.922 | 0.947 | 0.958 | 0.967 | 0.973 | 0.978 | 0.982 |
| Random Forest | 0.885 | 0.917 | 0.942 | 0.952 | 0.960 | 0.967 | 0.971 | 0.976 |
| Rotation Forest | 0.886 | 0.919 | 0.944 | 0.955 | 0.963 | 0.970 | 0.975 | 0.979 |



| | | | | | | | | |
|---------------------|-------|-------|-------|-------|-------|-------|-------|-------|
| PBC4cip | 0.880 | 0.906 | 0.928 | 0.936 | 0.944 | 0.951 | 0.957 | 0.960 |
| Bagging | 0.871 | 0.897 | 0.919 | 0.929 | 0.936 | 0.932 | 0.947 | 0.951 |
| J48 | 0.745 | 0.765 | 0.784 | 0.810 | 0.820 | 0.835 | 0.839 | 0.843 |
| MLP | 0.798 | 0.810 | 0.817 | 0.824 | 0.825 | 0.835 | 0.838 | 0.847 |
| GNB(sklearn) | 0.663 | 0.662 | 0.661 | 0.659 | 0.658 | 0.657 | 0.656 | 0.655 |

ANEXO 6. Resultado de la experimentación de la metodología BoW en el dataset propuesto por Bejani y Ghatte [13]

Se utilizaron los mismos parámetros de optimización expuestos en el Anexo 3.

Tabla 32. Resultados de clasificadores en el dataset de Bejani et Ghatte [36] usando BoW

| Ejes | Clasificador | (K, L) | TP Rate | FP Rate | Precision | Recall | F-1 | MCC | AUC | PRC |
|-----------|------------------------|------------|---------|---------|-----------|--------|-------|-------|--------------|-------|
| (X, Y, Z) | MLP(sklearn) | (150,200) | 0.490 | 0.262 | 0.474 | 0.490 | 0.479 | 0.233 | 0.630 | 0.438 |
| | Random Forest | (150,200) | 0.506 | 0.275 | 0.483 | 0.506 | 0.483 | 0.242 | 0.676 | 0.489 |
| | MHLDTForest | (150,200) | 0.484 | 0.320 | 0.460 | 0.484 | 0.436 | 0.186 | 0.673 | 0.490 |
| | PBC4cip | (150,200) | 0.283 | 0.126 | 0.608 | 0.283 | 0.278 | 0.213 | 0.627 | 0.404 |
| | Bagging | (150,200) | 0.488 | 0.327 | 0.450 | 0.488 | 0.431 | 0.186 | 0.669 | 0.486 |
| | Rotation Forest | (150,200) | 0.479 | 0.330 | 0.420 | 0.479 | 0.418 | 0.164 | 0.672 | 0.476 |
| | J48 | (150,200) | 0.468 | 0.314 | 0.419 | 0.468 | 0.425 | 0.165 | 0.598 | 0.401 |
| | GNB(sklearn) | (150,200) | 0.260 | 0.169 | 0.414 | 0.260 | 0.245 | 0.095 | 0.568 | 0.367 |
| | MLP(sklearn) | (150, 250) | 0.468 | 0.264 | 0.451 | 0.468 | 0.456 | 0.206 | 0.661 | 0.454 |
| | Random Forest | (150, 250) | 0.470 | 0.276 | 0.438 | 0.470 | 0.449 | 0.195 | 0.690 | 0.499 |
| | MHLDTForest | (150, 250) | 0.495 | 0.302 | 0.450 | 0.495 | 0.453 | 0.202 | 0.674 | 0.489 |
| | PBC4cip | (150, 250) | 0.288 | 0.135 | 0.555 | 0.288 | 0.286 | 0.194 | 0.633 | 0.394 |
| | Bagging | (150, 250) | 0.488 | 0.331 | 0.444 | 0.488 | 0.431 | 0.177 | 0.680 | 0.495 |
| | Rotation Forest | (150, 250) | 0.490 | 0.325 | 0.424 | 0.490 | 0.432 | 0.178 | 0.684 | 0.485 |
| | J48 | (150, 250) | 0.463 | 0.356 | 0.406 | 0.463 | 0.393 | 0.131 | 0.578 | 0.376 |
| | GNB(sklearn) | (150, 250) | 0.250 | 0.164 | 0.418 | 0.250 | 0.232 | 0.092 | 0.585 | 0.378 |
| | MLP(sklearn) | (150, 300) | 0.447 | 0.272 | 0.428 | 0.447 | 0.436 | 0.176 | 0.631 | 0.423 |
| | Random Forest | (150, 300) | 0.484 | 0.273 | 0.457 | 0.484 | 0.466 | 0.216 | 0.678 | 0.484 |
| | MHLDTForest | (150, 300) | 0.480 | 0.301 | 0.411 | 0.480 | 0.433 | 0.181 | 0.663 | 0.461 |
| | PBC4cip | (150, 300) | 0.285 | 0.144 | 0.524 | 0.285 | 0.281 | 0.175 | 0.622 | 0.407 |
| | Bagging | (150, 300) | 0.483 | 0.337 | 0.422 | 0.483 | 0.415 | 0.172 | 0.665 | 0.466 |
| | Rotation Forest | (150, 300) | 0.505 | 0.316 | 0.453 | 0.505 | 0.445 | 0.209 | 0.675 | 0.478 |
| | J48 | (150, 300) | 0.472 | 0.292 | 0.430 | 0.472 | 0.436 | 0.181 | 0.633 | 0.426 |
| | GNB(sklearn) | (150, 300) | 0.230 | 0.172 | 0.382 | 0.230 | 0.210 | 0.060 | 0.569 | 0.369 |



| | | | | | | | | | | |
|-------------|------------------------|---------------|----------------|----------------|------------------|---------------|------------|------------|--------------|------------|
| (X, Y, Z) | MLP(sklearn) | (200, 300) | 0.445 | 0.278 | 0.429 | 0.445 | 0.435 | 0.169 | 0.646 | 0.446 |
| | Random Forest | (200, 300) | 0.451 | 0.287 | 0.418 | 0.451 | 0.43 | 0.164 | 0.669 | 0.490 |
| | MHLDTForest | (200, 300) | 0.459 | 0.313 | 0.410 | 0.459 | 0.421 | 0.152 | 0.653 | 0.472 |
| | PBC4cip | (200, 300) | 0.261 | 0.138 | 0.554 | 0.261 | 0.256 | 0.174 | 0.597 | 0.377 |
| | Bagging | (200, 300) | 0.473 | 0.351 | 0.426 | 0.473 | 0.402 | 0.152 | 0.647 | 0.467 |
| | Rotation Forest | (200, 300) | 0.474 | 0.342 | 0.411 | 0.474 | 0.409 | 0.150 | 0.651 | 0.471 |
| | J48 | (200, 300) | 0.436 | 0.436 | ? | 0.436 | ? | ? | 0.493 | 0.32 |
| | GNB(sklearn) | (200, 300) | 0.261 | 0.188 | 0.405 | 0.261 | 0.27 | 0.083 | 0.591 | 0.381 |
| Ejes | Clasificador | (K, L) | TP Rate | FP Rate | Precision | Recall | F-1 | MCC | AUC | PRC |
| (X, Y) | MLP(sklearn) | (150, 200) | 0.491 | 0.245 | 0.471 | 0.491 | 0.479 | 0.246 | 0.677 | 0.477 |
| | Random Forest | (150, 200) | 0.516 | 0.249 | 0.484 | 0.516 | 0.493 | 0.269 | 0.728 | 0.543 |
| | MHLDTForest | (150, 200) | 0.531 | 0.259 | 0.488 | 0.531 | 0.499 | 0.278 | 0.724 | 0.534 |
| | PBC4cip | (150, 200) | 0.307 | 0.151 | 0.534 | 0.307 | 0.318 | 0.194 | 0.680 | 0.468 |
| | Bagging | (150, 200) | 0.533 | 0.293 | 0.501 | 0.533 | 0.485 | 0.270 | 0.670 | 0.483 |
| | Rotation Forest | (150, 200) | 0.533 | 0.300 | 0.478 | 0.533 | 0.473 | 0.258 | 0.717 | 0.527 |
| | J48 | (150, 200) | 0.519 | 0.294 | 0.466 | 0.519 | 0.466 | 0.246 | 0.627 | 0.427 |
| | GNB(sklearn) | (150, 200) | 0.28 | 0.20 | 0.385 | 0.280 | 0.281 | 0.078 | 0.622 | 0.397 |
| | MLP(sklearn) | (150, 250) | 0.497 | 0.244 | 0.477 | 0.497 | 0.484 | 0.254 | 0.704 | 0.495 |
| | Random Forest | (150, 250) | 0.531 | 0.245 | 0.499 | 0.531 | 0.507 | 0.290 | 0.743 | 0.564 |
| | MHLDTForest | (150, 250) | 0.542 | 0.255 | 0.494 | 0.542 | 0.506 | 0.293 | 0.724 | 0.539 |
| | PBC4cip | (150, 250) | 0.321 | 0.135 | 0.612 | 0.321 | 0.342 | 0.246 | 0.682 | 0.448 |
| | Bagging | (150, 250) | 0.535 | 0.306 | 0.490 | 0.535 | 0.470 | 0.263 | 0.727 | 0.552 |
| | Rotation Forest | (150, 250) | 0.545 | 0.282 | 0.478 | 0.545 | 0.486 | 0.281 | 0.740 | 0.557 |
| | J48 | (150, 250) | 0.533 | 0.268 | 0.484 | 0.533 | 0.492 | 0.279 | 0.660 | 0.443 |
| | GNB(sklearn) | (150, 250) | 0.315 | 0.177 | 0.451 | 0.315 | 0.325 | 0.147 | 0.626 | 0.407 |
| | MLP(sklearn) | (150, 300) | 0.540 | 0.230 | 0.525 | 0.540 | 0.530 | 0.314 | 0.701 | 0.491 |
| | Random Forest | (150, 300) | 0.531 | 0.252 | 0.504 | 0.531 | 0.509 | 0.287 | 0.729 | 0.531 |
| | MHLDTForest | (150, 300) | 0.566 | 0.246 | 0.527 | 0.566 | 0.533 | 0.332 | 0.728 | 0.530 |
| | PBC4cip | (150, 300) | 0.331 | 0.142 | 0.604 | 0.331 | 0.357 | 0.245 | 0.679 | 0.451 |
| | Bagging | (150, 300) | 0.530 | 0.314 | 0.510 | 0.530 | 0.463 | 0.261 | 0.719 | 0.518 |
| | Rotation Forest | (150, 300) | 0.519 | 0.308 | 0.465 | 0.519 | 0.459 | 0.232 | 0.729 | 0.533 |
| | J48 | (150, 300) | 0.512 | 0.298 | 0.473 | 0.512 | 0.465 | 0.238 | 0.636 | 0.425 |
| | GNB(sklearn) | (150, 300) | 0.303 | 0.187 | 0.419 | 0.303 | 0.294 | 0.114 | 0.653 | 0.411 |
| (X, Y) | MLP(sklearn) | (200, 300) | 0.488 | 0.249 | 0.472 | 0.488 | 0.478 | 0.24 | 0.698 | 0.51 |
| | Random Forest | (200, 300) | 0.512 | 0.252 | 0.482 | 0.512 | 0.491 | 0.262 | 0.723 | 0.546 |
| | MHLDTForest | (200, 300) | 0.516 | 0.266 | 0.484 | 0.516 | 0.490 | 0.260 | 0.705 | 0.533 |
| | PBC4cip | (200, 300) | 0.310 | 0.144 | 0.577 | 0.310 | 0.330 | 0.220 | 0.619 | 0.384 |



| | | | | | | | | | | | |
|---------------------|------------------------|------------------------|----------------|----------------|------------------|---------------|------------|------------|--------------|--------------|-------|
| | Bagging | (200, 300) | 0.486 | 0.342 | 0.414 | 0.486 | 0.412 | 0.171 | 0.680 | 0.506 | |
| | Rotation Forest | (200, 300) | 0.488 | 0.327 | 0.418 | 0.488 | 0.426 | 0.179 | 0.713 | 0.530 | |
| | J48 | (200, 300) | 0.434 | 0.436 | ? | 0.434 | ? | ? | 0.496 | 0.321 | |
| | GNB(sklearn) | (200, 300) | 0.390 | 0.180 | 0.492 | 0.390 | 0.408 | 0.216 | 0.648 | 0.429 | |
| Ejes | Clasificador | (K, L) | TP Rate | FP Rate | Precision | Recall | F-1 | MCC | AUC | PRC | |
| X, Y ₂ | MLP(sklearn) | (150, 200) | 0.620 | 0.186 | 0.617 | 0.620 | 0.603 | 0.447 | 0.788 | 0.603 | |
| | Random Forest | (150, 200) | 0.624 | 0.192 | 0.618 | 0.624 | 0.610 | 0.447 | 0.788 | 0.593 | |
| | MHLDTForest | (150, 200) | 0.641 | 0.203 | 0.621 | 0.641 | 0.610 | 0.459 | 0.794 | 0.596 | |
| | PBC4cip | (150, 200) | 0.449 | 0.154 | 0.589 | 0.449 | 0.469 | 0.319 | 0.710 | 0.471 | |
| | Bagging | (150, 200) | 0.620 | 0.232 | 0.609 | 0.620 | 0.577 | 0.421 | 0.779 | 0.579 | |
| | Rotation Forest | (150, 200) | 0.596 | 0.234 | 0.563 | 0.596 | 0.557 | 0.383 | 0.775 | 0.570 | |
| | J48 | (150, 200) | 0.582 | 0.229 | 0.566 | 0.582 | 0.556 | 0.373 | 0.727 | 0.526 | |
| | GNB(sklearn) | (150, 200) | 0.321 | 0.157 | 0.485 | 0.321 | 0.309 | 0.170 | 0.720 | 0.479 | |
| | X, Y ₂ | MLP(sklearn) | (150, 250) | 0.610 | 0.183 | 0.606 | 0.610 | 0.603 | 0.434 | 0.788 | 0.579 |
| | | Random Forest | (150, 250) | 0.624 | 0.191 | 0.613 | 0.624 | 0.609 | 0.445 | 0.791 | 0.589 |
| | | MHLDTForest | (150, 250) | 0.638 | 0.196 | 0.611 | 0.638 | 0.609 | 0.455 | 0.789 | 0.587 |
| | | PBC4cip | (150, 250) | 0.422 | 0.162 | 0.564 | 0.422 | 0.443 | 0.282 | 0.703 | 0.485 |
| | | Bagging | (150, 250) | 0.613 | 0.225 | 0.581 | 0.613 | 0.571 | 0.410 | 0.772 | 0.568 |
| | | Rotation Forest | (150, 250) | 0.645 | 0.192 | 0.619 | 0.645 | 0.615 | 0.466 | 0.795 | 0.603 |
| | | J48 | (150, 250) | 0.575 | 0.224 | 0.542 | 0.575 | 0.542 | 0.355 | 0.751 | 0.530 |
| | | GNB(sklearn) | (150, 250) | 0.369 | 0.151 | 0.515 | 0.369 | 0.370 | 0.222 | 0.740 | 0.500 |
| X, Y ₂ | | MLP(sklearn) | (150, 300) | 0.592 | 0.199 | 0.59 | 0.592 | 0.574 | 0.407 | 0.777 | 0.574 |
| | | Random Forest | (150, 300) | 0.599 | 0.208 | 0.588 | 0.599 | 0.576 | 0.410 | 0.784 | 0.575 |
| | | MHLDTForest | (150, 300) | 0.652 | 0.196 | 0.607 | 0.652 | 0.609 | 0.469 | 0.780 | 0.581 |
| | | PBC4cip | (150, 300) | 0.373 | 0.152 | 0.539 | 0.373 | 0.376 | 0.245 | 0.687 | 0.447 |
| | | Bagging | (150, 300) | 0.627 | 0.228 | 0.563 | 0.627 | 0.568 | 0.415 | 0.770 | 0.556 |
| | | Rotation Forest | (150, 300) | 0.631 | 0.227 | 0.565 | 0.631 | 0.569 | 0.422 | 0.796 | 0.592 |
| | | J48 | (150, 300) | 0.596 | 0.219 | 0.571 | 0.596 | 0.557 | 0.387 | 0.739 | 0.527 |
| | | GNB(sklearn) | (150, 300) | 0.331 | 0.163 | 0.481 | 0.331 | 0.317 | 0.170 | 0.730 | 0.478 |
| | X, Y ₂ | MLP(sklearn) | (200, 300) | 0.596 | 0.182 | 0.607 | 0.596 | 0.595 | 0.424 | 0.801 | 0.600 |
| | | Random Forest | (200, 300) | 0.603 | 0.21 | 0.606 | 0.603 | 0.583 | 0.413 | 0.786 | 0.604 |
| | | MHLDTForest | (200, 300) | 0.585 | 0.247 | 0.567 | 0.585 | 0.547 | 0.369 | 0.773 | 0.579 |
| | | PBC4cip | (200, 300) | 0.449 | 0.174 | 0.558 | 0.449 | 0.465 | 0.291 | 0.649 | 0.423 |
| | | Bagging | (200, 300) | 0.564 | 0.274 | 0.480 | 0.564 | 0.500 | 0.314 | 0.759 | 0.553 |
| | | Rotation Forest | (200, 300) | 0.596 | 0.233 | 0.582 | 0.596 | 0.553 | 0.388 | 0.784 | 0.585 |
| | | J48 | (200, 300) | 0.589 | 0.229 | 0.586 | 0.589 | 0.552 | 0.385 | 0.708 | 0.485 |
| | | GNB(sklearn) | (200, 300) | 0.408 | 0.139 | 0.571 | 0.408 | 0.417 | 0.282 | 0.733 | 0.501 |



| Ejes | Clasificador | (K, L) | TP Rate | FP Rate | Precision | Recall | F-1 | MCC | AUC | PRC | |
|------|-----------------|--------------|------------|---------|-----------|-----------|--------|-------|--------------|--------------|-------|
| (X) | MLP(sklearn) | (150, 200) | 0.582 | 0.176 | 0.587 | 0.582 | 0.575 | 0.413 | 0.838 | 0.663 | |
| | Random Forest | (150, 200) | 0.599 | 0.186 | 0.587 | 0.599 | 0.578 | 0.418 | 0.832 | 0.676 | |
| | MHLDTForest | (150, 200) | 0.613 | 0.194 | 0.585 | 0.613 | 0.581 | 0.425 | 0.811 | 0.649 | |
| | PBC4cip | (150, 200) | 0.397 | 0.111 | 0.691 | 0.397 | 0.432 | 0.355 | 0.759 | 0.509 | |
| | Bagging | (150, 200) | 0.564 | 0.265 | 0.457 | 0.564 | 0.497 | 0.303 | 0.827 | 0.658 | |
| | Rotation Forest | (150, 200) | 0.582 | 0.237 | 0.481 | 0.582 | 0.522 | 0.341 | 0.836 | 0.667 | |
| | J48 | (150, 200) | 0.554 | 0.238 | 0.483 | 0.554 | 0.507 | 0.312 | 0.715 | 0.505 | |
| | GNB(sklearn) | (150, 200) | 0.498 | 0.115 | 0.623 | 0.498 | 0.518 | 0.397 | 0.798 | 0.589 | |
| | MLP(sklearn) | (150, 250) | 0.603 | 0.173 | 0.603 | 0.603 | 0.593 | 0.435 | 0.835 | 0.663 | |
| | Random Forest | (150, 250) | 0.592 | 0.196 | 0.554 | 0.592 | 0.562 | 0.394 | 0.820 | 0.658 | |
| | MHLDTForest | (150, 250) | 0.613 | 0.201 | 0.587 | 0.613 | 0.577 | 0.420 | 0.814 | 0.660 | |
| | PBC4cip | (150, 250) | 0.432 | 0.104 | 0.713 | 0.432 | 0.470 | 0.394 | 0.736 | 0.461 | |
| | Bagging | (150, 250) | 0.564 | 0.263 | 0.472 | 0.564 | 0.503 | 0.310 | 0.821 | 0.662 | |
| | Rotation Forest | (150, 250) | 0.613 | 0.220 | 0.542 | 0.613 | 0.554 | 0.397 | 0.842 | 0.673 | |
| | J48 | (150, 250) | 0.592 | 0.209 | 0.549 | 0.592 | 0.554 | 0.384 | 0.730 | 0.514 | |
| | GNB(sklearn) | (150, 250) | 0.470 | 0.127 | 0.605 | 0.470 | 0.496 | 0.362 | 0.776 | 0.562 | |
| | MLP(sklearn) | (150, 300) | 0.641 | 0.148 | 0.646 | 0.641 | 0.635 | 0.496 | 0.845 | 0.675 | |
| | Random Forest | (150, 300) | 0.617 | 0.181 | 0.585 | 0.617 | 0.591 | 0.433 | 0.841 | 0.688 | |
| | MHLDTForest | (150, 300) | 0.659 | 0.179 | 0.630 | 0.659 | 0.628 | 0.489 | 0.834 | 0.687 | |
| | PBC4cip | (150, 300) | 0.383 | 0.113 | 0.698 | 0.383 | 0.419 | 0.346 | 0.746 | 0.491 | |
| | Bagging | (150, 300) | 0.596 | 0.257 | 0.543 | 0.596 | 0.545 | 0.370 | 0.816 | 0.669 | |
| | Rotation Forest | (150, 300) | 0.648 | 0.204 | 0.581 | 0.648 | 0.603 | 0.452 | 0.846 | 0.699 | |
| | J48 | (150, 300) | 0.620 | 0.198 | 0.573 | 0.620 | 0.589 | 0.425 | 0.742 | 0.551 | |
| | GNB(sklearn) | (150, 300) | 0.519 | 0.110 | 0.657 | 0.519 | 0.550 | 0.430 | 0.794 | 0.594 | |
| | MLP(sklearn) | (200, 300) | 0.585 | 0.180 | 0.581 | 0.585 | 0.573 | 0.409 | 0.810 | 0.629 | |
| | Random Forest | (200, 300) | 0.596 | 0.196 | 0.599 | 0.596 | 0.571 | 0.411 | 0.797 | 0.63 | |
| | MHLDTForest | (200, 300) | 0.589 | 0.217 | 0.543 | 0.589 | 0.547 | 0.380 | 0.790 | 0.613 | |
| | PBC4cip | (200, 300) | 0.362 | 0.116 | 0.669 | 0.362 | 0.393 | 0.318 | 0.669 | 0.412 | |
| | Bagging | (200, 300) | 0.516 | 0.301 | 0.422 | 0.516 | 0.442 | 0.237 | 0.781 | 0.612 | |
| | Rotation Forest | (200, 300) | 0.571 | 0.250 | 0.476 | 0.571 | 0.512 | 0.327 | 0.803 | 0.622 | |
| | J48 | (200, 300) | 0.547 | 0.238 | 0.48 | 0.547 | 0.5 | 0.317 | 0.681 | 0.459 | |
| | GNB(sklearn) | (200, 300) | 0.484 | 0.135 | 0.597 | 0.484 | 0.515 | 0.366 | 0.760 | 0.546 | |
| | Ejes | Clasificador | (K, L) | TP Rate | FP Rate | Precision | Recall | F-1 | MCC | AUC | PRC |
| | | MLP(sklearn) | (150, 200) | 0.488 | 0.270 | 0.445 | 0.488 | 0.459 | 0.217 | 0.683 | 0.491 |



| | | | | | | | | | | |
|---------------------|------------------------|------------|-------|-------|-------|-------|-------|--------------|--------------|-------|
| (Y) | Random Forest | (150, 200) | 0.502 | 0.271 | 0.463 | 0.502 | 0.466 | 0.235 | 0.667 | 0.466 |
| | MHLDTForest | (150, 200) | 0.491 | 0.301 | 0.424 | 0.491 | 0.438 | 0.201 | 0.668 | 0.459 |
| | PBC4cip | (150, 200) | 0.216 | 0.162 | 0.472 | 0.216 | 0.205 | 0.095 | 0.580 | 0.373 |
| | Bagging | (150, 200) | 0.488 | 0.354 | 0.421 | 0.488 | 0.398 | 0.177 | 0.628 | 0.425 |
| | Rotation Forest | (150, 200) | 0.491 | 0.301 | 0.424 | 0.491 | 0.438 | 0.201 | 0.668 | 0.459 |
| | J48 | (150, 200) | 0.505 | 0.302 | 0.462 | 0.505 | 0.453 | 0.229 | 0.616 | 0.431 |
| | GNB(sklearn) | (150, 200) | 0.272 | 0.177 | 0.438 | 0.272 | 0.281 | 0.113 | 0.612 | 0.401 |
| | MLP(sklearn) | (150, 250) | 0.502 | 0.258 | 0.472 | 0.502 | 0.482 | 0.245 | 0.683 | 0.49 |
| | Random Forest | (150, 250) | 0.516 | 0.26 | 0.475 | 0.516 | 0.488 | 0.257 | 0.674 | 0.478 |
| | MHLDTForest | (150, 250) | 0.516 | 0.282 | 0.451 | 0.516 | 0.472 | 0.241 | 0.673 | 0.477 |
| | PBC4cip | (150, 250) | 0.223 | 0.168 | 0.471 | 0.223 | 0.230 | 0.097 | 0.587 | 0.378 |
| | Bagging | (150, 250) | 0.488 | 0.352 | 0.425 | 0.488 | 0.403 | 0.176 | 0.647 | 0.449 |
| | Rotation Forest | (150, 250) | 0.533 | 0.293 | 0.501 | 0.533 | 0.485 | 0.270 | 0.670 | 0.483 |
| | J48 | (150, 250) | 0.516 | 0.260 | 0.475 | 0.516 | 0.488 | 0.257 | 0.674 | 0.478 |
| GNB(sklearn) | (150, 250) | 0.275 | 0.190 | 0.414 | 0.275 | 0.282 | 0.095 | 0.613 | 0.390 | |
| (Y) | MLP(sklearn) | (150, 300) | 0.463 | 0.275 | 0.442 | 0.463 | 0.445 | 0.191 | 0.650 | 0.452 |
| | Random Forest | (150, 300) | 0.456 | 0.294 | 0.417 | 0.456 | 0.427 | 0.166 | 0.641 | 0.446 |
| | MHLDTForest | (150, 300) | 0.477 | 0.311 | 0.409 | 0.477 | 0.428 | 0.174 | 0.647 | 0.468 |
| | PBC4cip | (150, 300) | 0.240 | 0.167 | 0.473 | 0.240 | 0.235 | 0.111 | 0.568 | 0.351 |
| | Bagging | (150, 300) | 0.512 | 0.336 | 0.458 | 0.512 | 0.434 | 0.222 | 0.620 | 0.434 |
| | Rotation Forest | (150, 300) | 0.512 | 0.312 | 0.477 | 0.512 | 0.458 | 0.230 | 0.624 | 0.447 |
| | J48 | (150, 300) | 0.495 | 0.308 | 0.457 | 0.495 | 0.446 | 0.212 | 0.559 | 0.382 |
| | GNB(sklearn) | (150, 300) | 0.247 | 0.194 | 0.387 | 0.247 | 0.257 | 0.063 | 0.581 | 0.372 |
| | MLP(sklearn) | (200, 300) | 0.456 | 0.286 | 0.404 | 0.456 | 0.428 | 0.165 | 0.647 | 0.475 |
| | Random Forest | (200, 300) | 0.460 | 0.304 | 0.394 | 0.46 | 0.422 | 0.152 | 0.645 | 0.473 |
| | MHLDTForest | (200, 300) | 0.502 | 0.319 | 0.436 | 0.502 | 0.441 | 0.205 | 0.640 | 0.455 |
| | PBC4cip | (200, 300) | 0.293 | 0.167 | 0.510 | 0.293 | 0.300 | 0.165 | 0.547 | 0.343 |
| | Bagging | (200, 300) | 0.467 | 0.383 | 0.390 | 0.467 | 0.356 | 0.128 | 0.600 | 0.418 |
| | Rotation Forest | (200, 300) | 0.516 | 0.282 | 0.451 | 0.516 | 0.472 | 0.241 | 0.673 | 0.477 |
| J48 | (200, 300) | 0.467 | 0.323 | 0.444 | 0.467 | 0.431 | 0.162 | 0.567 | 0.396 | |
| GNB(sklearn) | (200, 300) | 0.279 | 0.189 | 0.441 | 0.279 | 0.303 | 0.113 | 0.578 | 0.385 | |

Tabla 33. Resultados de clasificadores del AUC en el dataset de Ghatee [36] usando BoW

| AUC | D-1 | D-2 | D-3 | D-4 |
|--|----------------|----------------|----------------|----------------|
| Clasificadores | (K=150, L=200) | (K=150, L=250) | (K=150, L=300) | (K=200, L=300) |
| Magnetómetro, combinación de los eies X.Y.Z | | | | |
| MLP(sklearn) | 0.630 | 0.661 | 0.631 | 0.646 |



| | | | | |
|---|-------|-------|-------|-------|
| Random Forest | 0.676 | 0.690 | 0.678 | 0.669 |
| MHLDTForest | 0.673 | 0.674 | 0.663 | 0.653 |
| PBC4cip | 0.627 | 0.633 | 0.622 | 0.597 |
| Bagging | 0.669 | 0.680 | 0.665 | 0.647 |
| Rotation Forest | 0.672 | 0.684 | 0.675 | 0.651 |
| J48 | 0.598 | 0.578 | 0.633 | 0.493 |
| GNB(sklearn) | 0.568 | 0.585 | 0.569 | 0.591 |
| Magnetómetro, combinación de los ejes X,Y | | | | |
| MLP(sklearn) | 0.677 | 0.704 | 0.701 | 0.698 |
| Random Forest | 0.728 | 0.743 | 0.729 | 0.723 |
| MHLDTForest | 0.724 | 0.724 | 0.728 | 0.705 |
| PBC4cip | 0.680 | 0.682 | 0.679 | 0.619 |
| Bagging | 0.670 | 0.727 | 0.719 | 0.680 |
| Rotation Forest | 0.717 | 0.740 | 0.729 | 0.713 |
| J48 | 0.627 | 0.660 | 0.636 | 0.496 |
| GNB(sklearn) | 0.622 | 0.626 | 0.653 | 0.648 |
| Magnetómetro, norma de $\ X, Y\ _2$ | | | | |
| MLP(sklearn) | 0.788 | 0.788 | 0.777 | 0.801 |
| Random Forest | 0.788 | 0.791 | 0.784 | 0.786 |
| MHLDTForest | 0.794 | 0.789 | 0.780 | 0.773 |
| PBC4cip | 0.710 | 0.703 | 0.687 | 0.649 |
| Bagging | 0.779 | 0.772 | 0.770 | 0.759 |
| Rotation Forest | 0.775 | 0.795 | 0.796 | 0.784 |
| J48 | 0.727 | 0.751 | 0.739 | 0.708 |
| GNB(sklearn) | 0.720 | 0.740 | 0.730 | 0.733 |
| Magnetómetro eje X | | | | |
| MLP(sklearn) | 0.838 | 0.835 | 0.845 | 0.810 |
| Random Forest | 0.832 | 0.820 | 0.841 | 0.797 |
| MHLDTForest | 0.811 | 0.814 | 0.834 | 0.790 |
| PBC4cip | 0.759 | 0.736 | 0.746 | 0.669 |
| Bagging | 0.827 | 0.821 | 0.816 | 0.781 |
| Rotation Forest | 0.836 | 0.842 | 0.846 | 0.803 |
| J48 | 0.715 | 0.730 | 0.742 | 0.681 |
| GNB(sklearn) | 0.798 | 0.776 | 0.794 | 0.760 |
| Magnetómetro eje Y | | | | |
| MLP(sklearn) | 0.683 | 0.683 | 0.650 | 0.647 |
| Random Forest | 0.667 | 0.674 | 0.641 | 0.645 |
| MHLDTForest | 0.668 | 0.673 | 0.647 | 0.640 |
| PBC4cip | 0.580 | 0.587 | 0.568 | 0.547 |
| Bagging | 0.628 | 0.647 | 0.620 | 0.600 |
| Rotation Forest | 0.668 | 0.670 | 0.624 | 0.673 |
| J48 | 0.616 | 0.674 | 0.559 | 0.567 |
| GNB(sklearn) | 0.612 | 0.613 | 0.581 | 0.578 |



Resumen de mejores resultados:

Tabla 34. Resultados la norma $\|X, Y\|_2$ y el eje X , al aplicar el filtro ADLF en Ghatee [36]

| | Clasificador | (K, L) | TP Rate | FP Rate | Precision | Recall | F-1 | MCC | AUC | PRC | |
|--------------|---------------|--------------|------------|---------|-----------|--------|-------|--------------|--------------|--------------|-------|
| $\ X, Y\ _2$ | MLP(sklearn) | (150, 200) | 0.864 | 0.064 | 0.861 | 0.864 | 0.862 | 0.806 | 0.974 | 0.937 | |
| | Random Forest | (150, 200) | 0.819 | 0.086 | 0.814 | 0.819 | 0.815 | 0.738 | 0.955 | 0.899 | |
| | MHLDTForest | (150, 200) | 0.808 | 0.080 | 0.810 | 0.808 | 0.807 | 0.734 | 0.864 | 0.727 | |
| | PBC4cip | (150, 200) | 0.794 | 0.071 | 0.820 | 0.794 | 0.797 | 0.726 | 0.931 | 0.862 | |
| | Bagging | (150, 200) | 0.843 | 0.080 | 0.838 | 0.843 | 0.838 | 0.772 | 0.957 | 0.897 | |
| | Rotation | (150, 200) | 0.836 | 0.082 | 0.829 | 0.836 | 0.824 | 0.763 | 0.938 | 0.863 | |
| | J48 | (150, 200) | 0.833 | 0.086 | 0.829 | 0.833 | 0.822 | 0.756 | 0.898 | 0.765 | |
| | GNB(sklearn) | (150, 200) | 0.833 | 0.060 | 0.849 | 0.833 | 0.832 | 0.773 | 0.954 | 0.878 | |
| | MLP(sklearn) | (150, 250) | 0.868 | 0.065 | 0.865 | 0.868 | 0.865 | 0.81 | 0.976 | 0.937 | |
| | Random Forest | (150, 250) | 0.822 | 0.082 | 0.821 | 0.822 | 0.822 | 0.743 | 0.960 | 0.905 | |
| | MHLDTForest | (150, 250) | 0.829 | 0.076 | 0.831 | 0.829 | 0.828 | 0.759 | 0.876 | 0.745 | |
| | PBC4cip | (150, 250) | 0.791 | 0.068 | 0.819 | 0.791 | 0.792 | 0.72 | 0.942 | 0.87 | |
| | Bagging | (150, 250) | 0.854 | 0.07 | 0.850 | 0.854 | 0.851 | 0.790 | 0.964 | 0.912 | |
| | Rotation | (150, 250) | 0.833 | 0.08 | 0.829 | 0.833 | 0.83 | 0.759 | 0.953 | 0.88 | |
| | J48 | (150, 250) | 0.833 | 0.078 | 0.829 | 0.833 | 0.83 | 0.76 | 0.900 | 0.772 | |
| | GNB(sklearn) | (150, 250) | 0.854 | 0.056 | 0.865 | 0.854 | 0.854 | 0.799 | 0.960 | 0.882 | |
| | MLP(sklearn) | (150, 300) | 0.857 | 0.07 | 0.853 | 0.857 | 0.854 | 0.794 | 0.968 | 0.919 | |
| | Random Forest | (150, 300) | 0.798 | 0.098 | 0.788 | 0.798 | 0.791 | 0.706 | 0.942 | 0.866 | |
| | MHLDTForest | (150, 300) | 0.794 | 0.086 | 0.794 | 0.794 | 0.793 | 0.714 | 0.854 | 0.711 | |
| | PBC4cip | (150, 300) | 0.76 | 0.084 | 0.791 | 0.76 | 0.759 | 0.676 | 0.919 | 0.826 | |
| | Bagging | (150, 300) | 0.826 | 0.087 | 0.817 | 0.826 | 0.818 | 0.747 | 0.949 | 0.863 | |
| | Rotation | (150, 300) | 0.798 | 0.098 | 0.788 | 0.798 | 0.791 | 0.706 | 0.942 | 0.866 | |
| | J48 | (150, 300) | 0.812 | 0.092 | 0.800 | 0.812 | 0.802 | 0.727 | 0.882 | 0.737 | |
| | GNB(sklearn) | (150, 300) | 0.826 | 0.066 | 0.844 | 0.826 | 0.824 | 0.765 | 0.954 | 0.856 | |
| | $\ X, Y\ _2$ | MLP(sklearn) | (200, 300) | 0.916 | 0.038 | 0.916 | 0.916 | 0.916 | 0.881 | 0.991 | 0.977 |
| | Random Forest | (200, 300) | 0.868 | 0.059 | 0.867 | 0.868 | 0.867 | 0.811 | 0.982 | 0.958 | |
| | MHLDTForest | (200, 300) | 0.892 | 0.048 | 0.891 | 0.892 | 0.891 | 0.848 | 0.922 | 0.829 | |
| | PBC4cip | (200, 300) | 0.854 | 0.053 | 0.862 | 0.854 | 0.854 | 0.794 | 0.959 | 0.916 | |
| Bagging | (200, 300) | 0.909 | 0.039 | 0.912 | 0.909 | 0.909 | 0.873 | 0.978 | 0.95 | | |
| Rotation | (200, 300) | 0.889 | 0.045 | 0.892 | 0.889 | 0.888 | 0.847 | 0.968 | 0.932 | | |
| J48 | (200, 300) | 0.885 | 0.048 | 0.886 | 0.885 | 0.884 | 0.841 | 0.947 | 0.880 | | |
| GNB(sklearn) | (200, 300) | 0.920 | 0.032 | 0.924 | 0.920 | 0.920 | 0.888 | 0.983 | 0.954 | | |



| Ejes | Clasificador | (K, L) | TP Rate | FP Rate | Precision | Recall | F-1 | MCC | AUC | PRC | |
|------|----------------------|----------------------|------------|---------|-----------|--------|-------|-------|--------------|--------------|-------|
| (X) | MLP(sklearn) | (150, 200) | 0.840 | 0.088 | 0.844 | 0.840 | 0.816 | 0.765 | 0.973 | 0.923 | |
| | Random Forest | (150, 200) | 0.787 | 0.079 | 0.786 | 0.787 | 0.787 | 0.708 | 0.956 | 0.885 | |
| | MHLDTForest | (150, 200) | 0.780 | 0.074 | 0.790 | 0.78 | 0.785 | 0.707 | 0.860 | 0.730 | |
| | PBC4cip | (150, 200) | 0.822 | 0.047 | 0.860 | 0.822 | 0.830 | 0.778 | 0.947 | 0.861 | |
| | Bagging | (150, 200) | 0.812 | 0.072 | 0.805 | 0.812 | 0.808 | 0.740 | 0.967 | 0.898 | |
| | Rotation | (150, 200) | 0.815 | 0.071 | 0.810 | 0.815 | 0.810 | 0.744 | 0.955 | 0.870 | |
| | J48 | (150, 200) | 0.822 | 0.067 | 0.818 | 0.822 | 0.819 | 0.755 | 0.928 | 0.817 | |
| | GNB(sklearn) | (150, 200) | 0.871 | 0.040 | 0.882 | 0.871 | 0.875 | 0.829 | 0.968 | 0.879 | |
| | | | | | | | | | | | |
| | | MLP(sklearn) | (150, 250) | 0.878 | 0.045 | 0.877 | 0.878 | 0.878 | 0.834 | 0.981 | 0.94 |
| | | Random Forest | (150, 250) | 0.840 | 0.067 | 0.830 | 0.84 | 0.831 | 0.775 | 0.968 | 0.907 |
| | | MHLDTForest | (150, 250) | 0.857 | 0.054 | 0.854 | 0.857 | 0.855 | 0.803 | 0.901 | 0.787 |
| | | PBC4cip | (150, 250) | 0.812 | 0.050 | 0.848 | 0.812 | 0.818 | 0.765 | 0.955 | 0.865 |
| | | Bagging | (150, 250) | 0.847 | 0.062 | 0.84 | 0.847 | 0.84 | 0.786 | 0.973 | 0.912 |
| | Rotation | (150, 250) | 0.826 | 0.072 | 0.809 | 0.826 | 0.809 | 0.753 | 0.951 | 0.860 | |
| | J48 | (150, 250) | 0.826 | 0.072 | 0.809 | 0.826 | 0.809 | 0.753 | 0.926 | 0.795 | |
| | GNB(sklearn) | (150, 250) | 0.864 | 0.047 | 0.868 | 0.864 | 0.866 | 0.817 | 0.965 | 0.877 | |
| (X) | | | | | | | | | | | |
| | | MLP(sklearn) | (150, 300) | 0.882 | 0.046 | 0.879 | 0.882 | 0.879 | 0.837 | 0.982 | 0.941 |
| | | Random Forest | (150, 300) | 0.836 | 0.061 | 0.836 | 0.836 | 0.836 | 0.775 | 0.965 | 0.892 |
| | | MHLDTForest | (150, 300) | 0.826 | 0.059 | 0.833 | 0.826 | 0.829 | 0.767 | 0.887 | 0.764 |
| | | PBC4cip | (150, 300) | 0.794 | 0.060 | 0.822 | 0.794 | 0.802 | 0.737 | 0.939 | 0.862 |
| | | Bagging | (150, 300) | 0.857 | 0.053 | 0.855 | 0.857 | 0.856 | 0.804 | 0.974 | 0.91 |
| | | Rotation | (150, 300) | 0.843 | 0.067 | 0.834 | 0.843 | 0.832 | 0.779 | 0.959 | 0.877 |
| | | J48 | (150, 300) | 0.836 | 0.066 | 0.828 | 0.836 | 0.830 | 0.771 | 0.931 | 0.808 |
| | | GNB(sklearn) | (150, 300) | 0.864 | 0.047 | 0.869 | 0.864 | 0.866 | 0.816 | 0.969 | 0.894 |
| | | | | | | | | | | | |
| | | MLP(sklearn) | (200, 300) | 0.902 | 0.047 | 0.924 | 0.902 | 0.885 | 0.872 | 0.990 | 0.966 |
| | | Random Forest | (200, 300) | 0.868 | 0.045 | 0.864 | 0.868 | 0.866 | 0.823 | 0.976 | 0.928 |
| | | MHLDTForest | (200, 300) | 0.843 | 0.048 | 0.848 | 0.843 | 0.845 | 0.795 | 0.908 | 0.796 |
| | | PBC4cip | (200, 300) | 0.857 | 0.038 | 0.874 | 0.857 | 0.862 | 0.820 | 0.968 | 0.911 |
| | Bagging | (200, 300) | 0.871 | 0.044 | 0.869 | 0.871 | 0.870 | 0.829 | 0.977 | 0.923 | |
| | Rotation | (200, 300) | 0.878 | 0.046 | 0.872 | 0.878 | 0.872 | 0.834 | 0.968 | 0.898 | |
| | J48 | (200, 300) | 0.875 | 0.048 | 0.868 | 0.875 | 0.869 | 0.828 | 0.949 | 0.845 | |
| | GNB(sklearn) | (200, 300) | 0.916 | 0.029 | 0.915 | 0.916 | 0.916 | 0.888 | 0.984 | 0.953 | |



ANEXO 7. Experimentación con el algoritmo MODLEM

Tabla 35. Resultados del algoritmo Modlem en el dataset propuesto por Ferreira et al.[10]

| | Algoritmo | Ejes | TP Rate | FP Rate | Precision | Recall | F-1 | MCC | AUC | PRC | |
|---------------|---|--------------|---------|---------|-----------|--------|-------|-------|--------------|--------------|-------|
| Weighted Avg. | Modlem | x | 0.449 | 0.113 | 0.391 | 0.449 | 0.403 | 0.313 | 0.668 | 0.314 | |
| | Modlem | $\ X, Y\ _2$ | 0.348 | 0.128 | 0.342 | 0.348 | 0.334 | 0.22 | 0.610 | 0.264 | |
| | Modlem | y | 0.348 | 0.143 | 0.319 | 0.348 | 0.315 | 0.201 | 0.602 | 0.238 | |
| | Modlem | xvz | 0.341 | 0.142 | 0.319 | 0.341 | 0.305 | 0.196 | 0.599 | 0.223 | |
| | Modlem | xy | 0.333 | 0.146 | 0.300 | 0.333 | 0.287 | 0.183 | 0.594 | 0.217 | |
| | Modlem | z | 0.377 | 0.129 | ? | 0.377 | ? | ? | 0.624 | 0.246 | |
| | Dataset filtrado por el filtro Análisis Discriminante Lineal de Fisher (ADLF) | | | | | | | | | | |
| | Modlem | x | 1.000 | 0 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | Modlem | y | 1.000 | 0 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | Modlem | z | 1.000 | 0 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | Modlem | xv | 1.000 | 0 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | Modlem | $\ X, Y\ _2$ | 1.000 | 0 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | Modlem | xvz | 1.000 | 0 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

Tabla 36. Resultados del algoritmo Modlem en el dataset de M. Ricardo et al. [14]

| | Algoritmo | Ejes | TP Rate | FP Rate | Precision | Recall | F-1 | MCC | AUC | PRC | |
|---------------|---|--------------|---------|---------|-----------|--------|-------|-------|--------------|-------|--|
| Weighted Avg. | Modlem | y | 0.715 | 0.071 | 0.712 | 0.715 | 0.712 | 0.642 | 0.822 | 0.609 | |
| | Modlem | x | 0.687 | 0.078 | 0.689 | 0.687 | 0.686 | 0.609 | 0.804 | 0.555 | |
| | Modlem | xvz | 0.612 | 0.097 | 0.605 | 0.612 | 0.607 | 0.511 | 0.757 | 0.467 | |
| | Modlem | $\ X, Y\ _2$ | 0.602 | 0.100 | 0.579 | 0.602 | 0.572 | 0.489 | 0.751 | 0.445 | |
| | Modlem | xv | 0.557 | 0.111 | 0.560 | 0.557 | 0.557 | 0.447 | 0.723 | 0.421 | |
| | Modlem | z | 0.519 | 0.121 | 0.533 | 0.519 | 0.522 | 0.404 | 0.699 | 0.398 | |
| | Dataset filtrado por el filtro Análisis Discriminante Lineal de Fisher (ADLF) | | | | | | | | | | |
| | Modlem | $\ X, Y\ _2$ | 0.853 | 0.037 | 0.854 | 0.853 | 0.853 | 0.816 | 0.908 | 0.761 | |
| | Modlem | y | 0.820 | 0.045 | 0.821 | 0.820 | 0.818 | 0.775 | 0.887 | 0.733 | |
| | Modlem | z | 0.767 | 0.058 | 0.772 | 0.767 | 0.768 | 0.711 | 0.855 | 0.651 | |
| | Modlem | x | 0.632 | 0.092 | 0.635 | 0.632 | 0.633 | 0.541 | 0.770 | 0.499 | |
| | Modlem | xv | 0.575 | 0.106 | 0.576 | 0.575 | 0.574 | 0.469 | 0.734 | 0.433 | |
| | Modlem | xvz | 0.521 | 0.120 | 0.523 | 0.521 | 0.521 | 0.402 | 0.701 | 0.380 | |



Tabla 37. Resultados del algoritmo Modlem en el dataset propuesto por Ghatee [36]

| | Algoritmo | Ejes | TP Rate | FP Rate | Precision | Recall | F-1 | MCC | AUC | PRC | |
|---------------|---|--------------|---------|---------|-----------|--------|-------|-------|--------------|-------|--|
| Weighted Avg. | Modlem | x | 0.659 | 0.167 | 0.645 | 0.659 | 0.638 | 0.501 | 0.746 | 0.545 | |
| | Modlem | $\ X, Y\ _2$ | 0.54 | 0.246 | 0.470 | 0.540 | 0.497 | 0.291 | 0.647 | 0.428 | |
| | Modlem | xvz | 0.493 | 0.286 | 0.457 | 0.493 | 0.462 | 0.215 | 0.604 | 0.388 | |
| | Modlem | v | 0.46 | 0.314 | 0.408 | 0.460 | 0.419 | 0.153 | 0.573 | 0.366 | |
| | Modlem | xv | 0.448 | 0.321 | 0.416 | 0.448 | 0.418 | 0.138 | 0.564 | 0.359 | |
| | Modlem | z | 0.425 | 0.351 | 0.364 | 0.425 | 0.376 | 0.080 | 0.537 | 0.343 | |
| | Dataset filtrado por el filtro Análisis Discriminante Lineal de Fisher (ADLF) | | | | | | | | | | |
| | Modlem | $\ X, Y\ _2$ | 0.777 | 0.119 | 0.778 | 0.777 | 0.77 | 0.675 | 0.829 | 0.661 | |
| | Modlem | x | 0.725 | 0.167 | 0.734 | 0.725 | 0.716 | 0.585 | 0.779 | 0.596 | |
| | Modlem | v | 0.704 | 0.177 | 0.708 | 0.704 | 0.695 | 0.55 | 0.763 | 0.575 | |
| | Modlem | z | 0.693 | 0.188 | 0.702 | 0.693 | 0.689 | 0.527 | 0.753 | 0.572 | |
| | Modlem | xvz | 0.528 | 0.277 | 0.515 | 0.528 | 0.507 | 0.271 | 0.626 | 0.408 | |
| | Modlem | xv | 0.479 | 0.302 | 0.459 | 0.479 | 0.457 | 0.19 | 0.588 | 0.377 | |



8. BIBLIOGRAFÍA

- [1] C. Perera, P. P. Jayaraman, A. Zaslavsky, D. Georgakopoulos, and P. Christen, “MOSDEN: An internet of things middleware for resource constrained mobile devices,” *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, pp. 1053–1062, 2014, doi: 10.1109/HICSS.2014.137.
- [2] J. Arnett, “Drunk driving, sensation seeking, and egocentrism among adolescents,” *Pers. Individ. Dif.*, vol. 11, no. 6, pp. 541–546, 1990, doi: 10.1016/0191-8869(90)90035-P.
- [3] P. Handel *et al.*, “Insurance telematics: Opportunities and challenges with the smartphone solution,” *IEEE Intell. Transp. Syst. Mag.*, vol. 6, no. 4, pp. 57–70, 2014, doi: 10.1109/MITS.2014.2343262.
- [4] G. A. M. Meiring and H. C. Myburgh, “A review of intelligent driving style analysis systems and related artificial intelligence algorithms,” *Sensors (Switzerland)*, vol. 15, no. 12, pp. 30653–30682, 2015, doi: 10.3390/s151229822.
- [5] D. A. Johnson and M. M. Trivedi, “Driving style recognition using a smartphone as a sensor platform,” *IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC*, pp. 1609–1615, 2011, doi: 10.1109/ITSC.2011.6083078.
- [6] L. Tasca, “A review of the literature on aggressive driving research. Ontario Advisory Group on Safe Driving Secretariat, Road User Safety Branch, Ontario Ministry of Transportation.,” pp. 1–25, 2000, doi: 10.1016/S0022-3182(80)80003-3.
- [7] H. L. Moosavi V, “Modeling urban traffic dynamics in coexistence with urban data streams,” *Proc. 2nd ACM SIGKDD Int. Work. Urban Comput.*, 2013.
- [8] M. Fazeen, B. Gozick, R. Dantu, M. Bhukhiya, and M. C. González, “Safe Driving Using Mobile Phones,” *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 3, pp. 1462–1468, 2012, doi: 10.1109/tits.2012.2187640.
- [9] J. Wahlström, I. Skog, P. Händel, and S. Member, “Smartphone-Based Vehicle Telematics : A Ten-Year Anniversary,” pp. 1–24, 2017.
- [10] J. F. Júnior *et al.*, “Driver behavior profiling: An investigation with different



- smartphone sensors and machine learning,” *PLoS One*, vol. 12, no. 4, pp. 1–16, 2017, doi: 10.1371/journal.pone.0174959.
- [11] S. Kanarachos, S. R. G. Christopoulos, and A. Chroneos, “Smartphones as an integrated platform for monitoring driver behaviour: The role of sensor fusion and connectivity,” *Transp. Res. Part C Emerg. Technol.*, vol. 95, no. April, pp. 867–882, 2018, doi: 10.1016/j.trc.2018.03.023.
- [12] E. I. Vlahogianni and E. N. Barmounakis, “Driving analytics using smartphones: Algorithms, comparisons and challenges,” *Transp. Res. Part C Emerg. Technol.*, vol. 79, pp. 196–206, 2017, doi: 10.1016/j.trc.2017.03.014.
- [13] M. M. Bejani and M. Ghatee, “A context aware system for driving style evaluation by an ensemble learning on smartphone sensors data,” *Transp. Res. Part C Emerg. Technol.*, vol. 89, no. February, pp. 303–320, 2018, doi: 10.1016/j.trc.2018.02.009.
- [14] M. R. Carlos, L. C. Gonzalez, J. Wahlstrom, G. Ramirez, F. Martinez, and G. Runger, “How Smartphone Accelerometers Reveal Aggressive Driving Behavior?--The Key Is the Representation,” *IEEE Trans. Intell. Transp. Syst.*, no. 1, pp. 1–11, 2019, doi: 10.1109/tits.2019.2926639.
- [15] L. C. Gonzalez, R. Moreno, H. J. Escalante, F. Martinez, and M. R. Carlos, “Learning Roadway Surface Disruption Patterns Using the Bag of Words Representation,” *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 11, pp. 2916–2928, 2017, doi: 10.1109/TITS.2017.2662483.
- [16] H. Gao *et al.*, “Papers Short Papers,” vol. 39, no. 3, pp. 1–8, 2000.
- [17] P. Senin and S. Malinchik, “SAX-VSM: Interpretable time series classification using sax and vector space model,” *Proc. - IEEE Int. Conf. Data Mining, ICDM*, pp. 1175–1180, 2013, doi: 10.1109/ICDM.2013.52.
- [18] J. Stefanowski, “On combined classifiers, rule induction and rough sets,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 4374 LNCS, no. PART 1, pp. 329–350, 2007.
- [19] Organización Mundial de la Salud, “Informe Sobre La Situación Mundial De La Seguridad Vial 2015,” *Ginebra, Suiza*, pp. 1–12, 2015.



- [20] G. M. Krishnaswamy S, Gama J, “Mobile Data Stream Mining: From Algorithms to Applications.,” *IEEE 13th Int. Conf. Mob. Data Manag.*, pp. 360–3, 2012.
- [21] FREMAP, “Conocimiento y acatamiento de normas y medidas. Respeto, cortesía e imagen,” 2019. [Online]. Available: <https://www.icv.csic.es/prevencion/Documentos/breves/FREMAP/vial.pdf>.
- [22] C. M. S. Martin, “Agresividad al volante,” 2019. [Online]. Available: <https://www.centromedicosanmartin.com/agresividad-al-volante/>.
- [23] A. Diaz Alvarez, F. Serradilla Garcia, J. E. Naranjo, J. J. Anaya, and F. Jimenez, “Modeling the driving behavior of electric vehicles using smartphones and neural networks,” *IEEE Intell. Transp. Syst. Mag.*, vol. 6, no. 3, pp. 44–53, 2014, doi: 10.1109/MITS.2014.2322651.
- [24] T. Lajunen and D. Parker, “Are aggressive people aggressive drivers? A study of the relationship between self-reported general aggressiveness, driver anger and aggressive driving,” *Accid. Anal. Prev.*, vol. 33, no. 2, pp. 243–255, 2001, doi: 10.1016/S0001-4575(00)00039-7.
- [25] C. S. Dula and M. E. Ballard, “Development and evaluation of a measure of dangerous, aggressive, negative emotional, and risky driving,” *J. Appl. Soc. Psychol.*, vol. 33, no. 2, pp. 263–282, 2003, doi: 10.1111/j.1559-1816.2003.tb01896.x.
- [26] H. R. Eftekhari and M. Ghatee, “A similarity-based neuro-fuzzy modeling for driving behavior recognition applying fusion of smartphone sensors,” *J. Intell. Transp. Syst. Technol. Planning, Oper.*, vol. 23, no. 1, pp. 72–83, 2019, doi: 10.1080/15472450.2018.1506338.
- [27] J. Paefgen, F. Kehr, Y. Zhai, and F. Michahelles, “Driving behavior analysis with smartphones,” *Proc. 11th Int. Conf. Mob. Ubiquitous Multimed. - MUM*, p. 1, 2012, doi: 10.1145/2406367.2406412.
- [28] H. Hromic *et al.*, “Real Time Analysis of Sensor Data for the Internet of Things by means of Clustering and Event Processing,” pp. 685–691, 2015.
- [29] H. Reza and M. Ghatee, “An inference engine for smartphones to preprocess data and detect stationary and transportation modes,” *Transp. Res. Part C*, vol. 69, pp. 313–



327, 2016, doi: 10.1016/j.trc.2016.06.005.

- [30] N. J. De Vries, M. H. Davel, J. Badenhorst, and W. D. Basson, “ScienceDirect A smartphone-based ASR data collection tool for under-resourced languages,” *Speech Commun.*, vol. 56, pp. 119–131, 2014, doi: 10.1016/j.specom.2013.07.001.
- [31] L. Eboli, G. Mazzulla, and G. Pungillo, “Combining speed and acceleration to define car users’ safe or unsafe driving behaviour,” *Transp. Res. Part C Emerg. Technol.*, vol. 68, pp. 113–125, 2016, doi: 10.1016/j.trc.2016.04.002.
- [32] M. Ghatee, *SMARTPHONES RECENT INNOVATIONS AND APPLICATIONS*, 2019 Nova. 2019.
- [33] R. Asadi and M. Ghatee, “A Rule-Based Decision Support System in Intelligent Hazmat Transportation System,” pp. 1–9, 2015.
- [34] S. Abpeykar and M. Ghatee, “An ensemble of RBF neural networks in decision tree structure with knowledge transferring to accelerate multi-classification,” *Neural Comput. Appl.*, vol. 9, 2018, doi: 10.1007/s00521-018-3543-9.
- [35] R. Lotfi and M. Ghatee, “Smartphone based Driving Style Classification Using Features Made by Discrete Wavelet Transformation.”
- [36] H. R. Eftekhari and M. Ghatee, “Hybrid of discrete wavelet transform and adaptive neuro fuzzy inference system for overall driving behavior recognition,” *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 58, pp. 782–796, 2018, doi: 10.1016/j.trf.2018.06.044.
- [37] R. M. Nieto, “La identificación de irregularidades del camino en base a patrones de acelerómetro,” Universidad Autónoma de Chihuahua, 2016.
- [38] J. Wang, P. Liu, S. Nahavandi, S. Member, and A. Kouzani, “Bag-of-words Representation for Biomedical Time Series Classification,” pp. 1–10, 2013.
- [39] J. Lin, R. Khade, and Y. Li, “Rotation-invariant similarity in time series using bag-of-patterns representation,” pp. 287–315, 2012, doi: 10.1007/s10844-012-0196-5.
- [40] L. Cañete-Sifuentes, R. Monroy, M. A. Medina-Pérez, O. Loyola-González, and F. Vera Voronisky, “Classification Based on Multivariate Contrast Patterns,” *IEEE Access*, vol. 7, pp. 55744–55762, 2019, doi: 10.1109/ACCESS.2019.2913649.



- [41] O. Loyola-González, M. A. Medina-Pérez, J. F. Martínez-Trinidad, J. A. Carrasco-Ochoa, R. Monroy, and M. García-Borroto, “PBC4cip: A new contrast pattern-based classifier for class imbalance problems,” *Knowledge-Based Syst.*, vol. 115, pp. 100–109, 2017, doi: 10.1016/j.knosys.2016.10.018.
- [42] J. Demš, “Statistical Comparisons of Classifiers over Multiple Data Sets,” vol. 7, pp. 1–30, 2006.
- [43] F. WILCOXON, “Individual comparisons of grouped data by ranking methods.,” *J. Econ. Entomol.*, vol. 39, no. 6, p. 269, 1946, doi: 10.1093/jee/39.2.269.
- [44] W. Myers, *Probabilidad estadística para Ingeniería y ciencia.* .
- [45] U. N. Córdoba, “Seguridad Vial,” 2019. [Online]. Available: <https://www.unc.edu.ar/campus-virtual/curso-seguridad-vial>.