

# 31

Fecha de presentación: enero, 2023

Fecha de aceptación: marzo, 2023

Fecha de publicación: mayo, 2023

## EL ENFOQUE

### DE APRENDIZAJE CONJUNTO EN LA DETECCIÓN DE FALLAS EN CAJAS DE ENGRANAJES

#### THE JOINT LEARNING APPROACH TO GEARBOX FAILURE DETECTION

Félix Antonio García Mora<sup>1</sup>

E-mail: [felix.garcia@epoch.edu.ec](mailto:felix.garcia@epoch.edu.ec)

ORCID: <https://orcid.org/0000-0001-5814-3694>

José Luis Escobar Chávez<sup>1</sup>

E-mail: [josel.escobar@epoch.edu.ec](mailto:josel.escobar@epoch.edu.ec)

ORCID: <https://orcid.org/0000-0002-5304-9022>

César Marcelo Gallegos Londoño<sup>1</sup>

E-mail: [cesar.gallegos@epoch.edu.ec](mailto:cesar.gallegos@epoch.edu.ec)

ORCID: <https://orcid.org/0000-0002-8685-7501>

Eduardo Segundo Hernández Dávila<sup>1</sup>

E-mail: [edhernandez@epoch.edu.ec](mailto:edhernandez@epoch.edu.ec)

ORCID: <https://orcid.org/0000-0003-4899-2371>

<sup>1</sup>Escuela Superior Politécnica de Chimborazo, Ecuador.

#### Cita sugerida (APA, séptima edición)

García Mora, F. A., Escobar Chávez, J. L., Gallegos Londoño, C. M. & Hernández Dávila, E. S. (2023). El enfoque de aprendizaje conjunto en la detección de fallas en cajas de engranajes. *Universidad y Sociedad*, 15(3), 325-333.

#### RESUMEN

El aprendizaje conjunto es un término que se utiliza para referirse a los métodos que combinan varios algoritmos para tomar una decisión, generalmente en tareas de aprendizaje automático supervisado. En los últimos años la regresión logística, Support Vector Machine, la red neuronal y otros algoritmos de aprendizaje de máquinas se han utilizado en el diagnóstico inteligente de fallas en maquinaria rotativa, sin embargo, cada algoritmo tiene sus ventajas y desventajas y pueden funcionar de distinta manera para un determinado conjunto de datos. En el presente artículo científico se proponen tres modelos de detección de fallas en cajas de engranajes bajo el enfoque de aprendizaje conjunto utilizando las técnicas de clasificación por votación mayoritaria, votación suave y apilamiento a través de las cuales se combinan las predicciones de varios estimadores a fin de mejorar la generalización y la robustez sobre un solo estimador. Para el desarrollo de los modelos especificados se utiliza un conjunto de datos de señales de falla de cajas de engranajes y cuatro técnicas de clasificación las cuales son: regresión logística, support vector machine, XGBoost y random forest. Los resultados finales de precisión de los modelos de votación mayoritaria y votación suave son 99.86% y 99.82% respectivamente, mientras que la precisión del modelo bajo el enfoque de apilamiento es de 99.85%. En comparación con los resultados de los clasificadores individuales, se demuestra que los modelos de aprendizaje conjunto compensan los errores cometidos por los clasificadores individuales y mejoran efectivamente la precisión de clasificación.

**Palabras clave:** aprendizaje conjunto, votación mayoritaria, votación suave, apilamiento, caja de engranajes

#### ABSTRACT

Joint learning is a term used to refer to methods that combine several algorithms to make a decision, usually in supervised machine learning tasks. In recent years logistic regression, Support Vector Machine, neural network and other machine learning algorithms have been used in intelligent fault diagnosis in rotating machinery, however, each algorithm has its advantages and disadvantages and may perform differently for a given data set. In this scientific paper, three gearbox fault detection models are proposed under the ensemble learning approach using the majority voting, soft voting and stacking classification techniques through which the predictions of several estimators are combined in order to improve generalization and robustness over a single estimator. A data set of gearbox failure signals and four classification techniques are used to develop the specified models: logistic regression, support vector machine, XGBoost and random forest. The final accuracy results of the majority voting and soft voting models are 99.86% and 99.82% respectively, while the accuracy of the model under the stacking approach is 99.85%. Compared with the results of the individual classifiers, it is shown that the ensemble learning models compensate for the errors made by the individual classifiers and effectively improve the classification accuracy.

**Keywords:** joint learning, majority voting, soft voting, stacking, preventive maintenance

## INTRODUCCIÓN

Las cajas de engranajes cumplen un papel crucial en sistemas de transmisión mecánica y su fiabilidad y alto rendimiento son importantes para las diferentes aplicaciones industriales. Dado que la falla de un engranaje implica la incapacidad del sistema para operar de manera correcta las consecuencias de esta pueden ser graves y costosas. Por lo tanto, es esencial detectar con precisión las fallas que pueden ocurrir en una caja de engranajes de manera oportuna (Azamfar, 2021).

El estado de salud de las cajas de engranajes puede ser determinado mediante técnicas de monitoreo de la condición tales como, vibraciones, corriente eléctrica, térmicas, emisiones acústicas y tribológicas. El monitoreo de señales de vibración es la técnica más utilizada debido a que cualquier falla en la caja de engranajes modifica el espectro normal de la señal. Mediante el análisis de vibraciones se identifica con exactitud alrededor del 90% de fallas de tal manera que se pueden tomar acciones de mantenimiento de manera oportuna (Devendiran & Manivannan, 2016).

Tradicionalmente el proceso de diagnóstico de fallas se basa en establecer una relación entre los datos obtenidos mediante monitoreo y la condición de la máquina, este proceso se realiza manualmente lo que aumenta el trabajo y reduce la precisión del diagnóstico. Los métodos avanzados de preprocesamiento de señales de vibración permiten identificar fallas y el lugar de la máquina donde estas ocurren, sin embargo, estos métodos son demasiado complejos y requieren de la experiencia y conocimiento especializado de expertos de lo cual carecen los mantenedores y usuarios de máquinas (Lei et al., 2021).

Para resolver este problema, los investigadores han prestado mucha atención a la introducción de otras técnicas para el diagnóstico de fallas en las máquinas y una forma es utilizar técnicas de aprendizaje de máquinas. Los enfoques de aprendizaje de máquinas tienen la capacidad de identificar patrones a partir de datos históricos disponibles y crear modelos predictivos capaces de realizar un diagnóstico mucho más rápido y preciso acerca del estado de salud de una máquina. Sin embargo, el rendimiento de estos enfoques depende de la elección adecuada de la técnica de aprendizaje de máquinas.

No todos los algoritmos se desempeñan de la misma manera, algunos algoritmos como Support Vector Machine y árboles de decisión pueden manejar datos de gran dimensión mejor que otros (Wuest et al., 2016). Cada algoritmo de aprendizaje de máquinas tiene sus ventajas y desventajas por lo que puede funcionar bien para un determinado conjunto de datos, pero de manera deficiente

para otro. Para dar solución a este problema se han investigado otras metodologías y una de estas es el aprendizaje conjunto (Sagi & Rokach, 2018). Los métodos de aprendizaje por conjunto combinan múltiples algoritmos de aprendizaje de máquinas para generar predicciones individuales débiles y luego fusionar los resultados de predicción mediante mecanismos de votación para mejorar la precisión y obtener mejores rendimientos que los obtenidos por cualquier algoritmo individual del conjunto (Zhou, 2012).

Las técnicas de aprendizaje de máquinas individuales y los métodos de aprendizaje conjunto se han aplicado con éxito en la clasificación y detección de fallas en maquinaria rotativa. Por ejemplo, García et al. (2021), compararon cinco técnicas de aprendizaje de máquinas, entre estas *XGBoost* y *random forest*, para clasificar fallas en rodamientos. Los resultados mostraron que *XGBoost* es el mejor modelo como promedio para clasificar la mayoría de las fallas. Además, mencionan que, a pesar de que el análisis de la envolvente es uno de los métodos más empleados dentro del monitoreo basado en la condición este es claramente superado por las técnicas de aprendizaje de máquinas.

Han et al. (2018), compararon *random forest*, Support Vector Machine y redes neuronales artificiales para diagnosticar fallas en maquinaria rotativa, demostraron que *random forest* supera a los clasificadores comparativos en términos de precisión, estabilidad y solidez de las características, especialmente con pequeños conjuntos de entrenamiento. Patil et al. (2021), utilizaron cinco algoritmos, entre ellos Support Vector Machine y árbol de decisión, como clasificadores base para diagnosticar fallas de cojinetes y engranajes utilizando aprendizaje conjunto basado clasificación por voto mayoritario. Observaron que el clasificador de voto mayoritario logra un mejor rendimiento de predicción que cada clasificador individual.

Seleccionar un modelo de aprendizaje de máquinas que permita obtener un excelente rendimiento es un desafío y para hacer frente a ese desafío, en este documento se desarrollan tres modelos bajo el enfoque de aprendizaje conjunto con el objetivo de combinar las predicciones de varios estimadores bases a fin de mejorar el rendimiento en términos de precisión y compensar los errores de clasificación que se obtendrían sobre un solo estimador.

Para el desarrollo de los tres modelos se utiliza un conjunto de datos de señales de falla de cajas de engranajes y cuatro algoritmos de clasificación comunes, los cuales son, regresión logística, *support vector machine*, *XGBoost* y *random forest*. Los algoritmos de conjunto se pueden

analizar en cuatro categorías: embolsado, impulso, apilamiento y votación.

## DESARROLLO

En este estudio se utiliza las técnicas de aprendizaje conjunto por votación mayoritaria, votación suave y la técnica de apilamiento. Se comparan los resultados de cada clasificador individual con los resultados de las técnicas de clasificación por votación mayoritaria, votación suave y la técnica de apilamiento y se discuten los resultados obtenidos.

### El aprendizaje conjunto como combinación de algoritmos

El aprendizaje conjunto es un término que se utiliza para referirse a los métodos que combinan varios algoritmos para tomar una decisión, generalmente en tareas de aprendizaje automático supervisado. La idea detrás del aprendizaje conjunto es que, al combinar varios modelos, los errores de un solo modelo probablemente serán compensados por otros modelos y como resultado, el rendimiento de predicción general del conjunto sería mejor que el de un solo modelo (Sagi & Rokach, 2018), (Dong et al., 2020).

La votación es una de las técnicas más conocidas de aprendizaje conjunto que se utiliza principalmente para aumentar la precisión de un modelo de predicción mediante la combinación de dos o más algoritmos. Primero, los conjuntos de entrenamiento se alimentan a diferentes algoritmos para construir modelos de clasificación individuales. Luego, un clasificador de votación promedia las predicciones de cada modelo para realizar predicciones para datos nuevos (Kumar et al., 2017).

La Figura 1 muestra la idea principal de un modelo clasificación por voto, el que consta de dos pasos: (1) se generan resultados de clasificación usando varios clasificadores y (2) se integran los resultados individuales para conseguir un resultado final mediante votación (Dong et al., 2020).

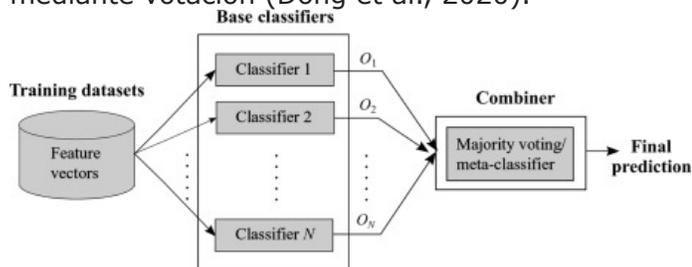


Figura 1. Clasificador de votación de conjunto

Fuente: elaboración propia

El clasificador de votación de conjunto puede realizar votaciones duras y suaves. Las votaciones duras

también se conocen como votaciones de voto mayoritario. La votación de conjunto duro es el caso más común de votación. Aquí, la etiqueta de clase  $Y$  se determina mediante la votación mayoritaria de cada clasificador  $C$  (Mahabub, 2020). Ecuación 1

(1)

El sistema de votación suave se emplea comúnmente para el clasificador base con salidas de probabilidad de clase (Cao et al., 2015). Un conjunto de votación suave clasifica nuevos puntos de datos utilizando un voto ponderado de probabilidades predichas, en lugar de clases predichas (Jones et al., 2021). En la votación de conjunto suave, se considera los nombres de clase sujetos a las probabilidades  $p$  previstas para el clasificador (Mahabub, 2020). Ecuación 2

(2)

El apilamiento construye modelos usando diferentes algoritmos de aprendizaje los cuales realizan predicciones individuales y posteriormente se entrena un algoritmo combinador que realiza predicciones finales en base a las predicciones generadas por los algoritmos base. El algoritmo combinador puede ser cualquier algoritmo base del conjunto (Sarajcev et al., 2021). El apilamiento tiene dos pasos: primero, recopila los resultados de predicción de cada aprendiz base y forma un nuevo conjunto de datos; segundo, los nuevos datos son utilizados como entrada para un nuevo algoritmo denominado alumno de segundo nivel o meta-clasificador que realiza una predicción final (Wang et al., 2011).

### Los métodos de aprendizaje conjunto

Dentro de los métodos de aprendizaje conjunto se identifican la regresión logística, *Support Vector Machine*, *Random Forest* y *XGBoost*. La regresión logística es un algoritmo de aprendizaje de máquinas que se utiliza para resolver problemas de clasificación binaria. Es un modelo de regresión que predice la probabilidad de que un conjunto de datos de entrada pertenezca a una clase determinada en base al modelo de regresión. Para realizar el modelado de datos la regresión logística utiliza una función sigmoidea. Algunas ventajas de la regresión logística es su implementación simple, facilidad de regularización y no necesita que los datos estén escalados, sin embargo, es susceptible de sobreajuste y no es adecuada para resolver problemas no lineales (Osman et al., 2021).

*Support Vector Machine* es un algoritmo diseñado para realizar tareas de clasificación y regresión. El algoritmo primero proyecta las muestras etiquetadas en un espacio vectorial. Luego intenta sepa-

rar de manera óptima las muestras proyectadas de tal manera que exista una distancia máxima entre estas. Cuando se ingresa una nueva muestra al modelo entrenado, esta se proyecta en el espacio vectorial y se predice la clase a la que pertenece dependiendo del lado del hiperplano en el que se encuentre. El algoritmo emplea una superficie de decisión para separar las clases y maximizar el margen entre estas. Cuando las clases no son linealmente separables *Support Vector Machine* utiliza funciones *kernel* que mapean las muestras a un espacio de alta dimensión en el cual es posible la clasificación lineal.

En el algoritmo *Random Forest*, se ingresa una entrada en la parte superior y los datos se agrupan en conjuntos cada vez más pequeños a medida que atraviesan el árbol. El bosque aleatorio lleva esta noción al siguiente nivel al combinar árboles con una noción de conjunto. El beneficio de un clasificador de bosque aleatorio es que sus tiempos de ejecución son concisos, desequilibrados y se pueden tratar los datos faltantes. En el bosque aleatorio, el nuevo conjunto de datos o los datos de prueba se distribuyen a todos los subárboles creados. Cada subárbol de decisión en el bosque puede decidir la clase del conjunto de datos. Luego, el modelo seleccionará la clase más apropiada de la votación mayoritaria (Ibrahim & Abdulazeez, 2021).

*XGBoost* es una de las implementaciones de máquinas de aumento de gradiente, que se conoce como uno de los algoritmos de mejor rendimiento utilizados para el aprendizaje supervisado. Se puede utilizar tanto para problemas de regresión como de clasificación (Ibrahim & Abdulazeez, 2021). *XGBoost* es un modelo de optimización que combina un modelo lineal con un modelo de árbol de impulso. Utiliza no solo la primera derivada sino también la segunda derivada de la función de pérdida para la derivación de segundo orden. Esto permite que el algoritmo converja de manera óptima, rápida y mejore la eficiencia de la solución óptima del modelo (Yu et al., 2020).

### Metodología

La metodología para la detección de fallas en cajas de engranajes mediante la técnica de votación mayoritaria, votación suave y apilamiento se muestra en la Figura 2.

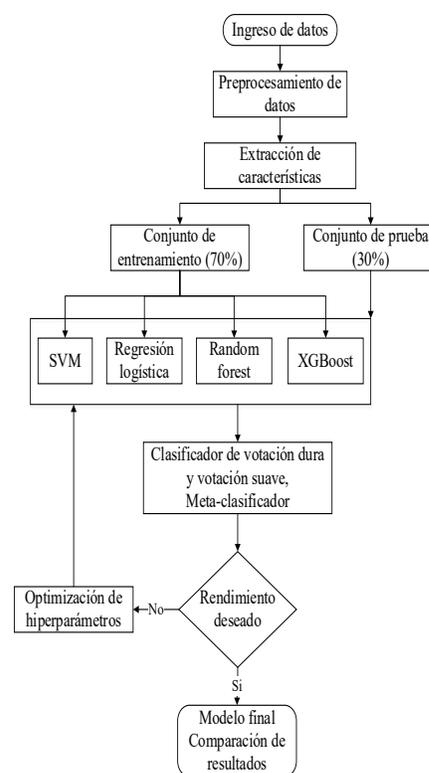


Figura 2. Metodología de clasificación por votación y apilamiento

Fuente: elaboración propia

La metodología consta de cinco fases. Primero, se realiza el pre-procesamiento del conjunto de datos en *Python*. Segundo, se extraen características en el dominio del tiempo y dominio de la frecuencia. Tercero, las características extraídas se dividen en 70% para entrenamiento y 30% para prueba. Cuarto, se utiliza cuatro clasificadores los cuales son: *Support Vector Machine*, *random forest*, *XGBoost* y regresión logística como clasificadores base para ser utilizados por las técnicas de aprendizaje de conjuntos de clasificación por votación y apilamiento. Quinto, se evalúan y comparan los resultados obtenidos.

El estudio se realizó con un conjunto de datos de señales de vibración de cajas de engranajes del repositorio de Iniciativa de Datos de Energía Abierta del departamento de energía de EE. UU. Los datos han sido registrados con la ayuda de cuatro sensores de vibración colocados en cuatro posiciones diferentes del cuerpo, con una variación de cargas de 0 a 90 por ciento, frecuencia de muestreo de 30Hz y en dos escenarios diferentes: 1) condición saludable y 2) condición de diente roto. Los datos del análisis de vibraciones, tomados en un escenario de condición saludable tienen un tamaño de

1.015.808, mientras que los datos tomados para un escenario de condición de diente roto tienen un tamaño de 1.005.311 muestras.

En primera instancia se verifica el tipo de datos numéricos, no numéricos, enteros o decimales y se realiza la respectiva transformación. Luego, se realiza una limpieza al conjunto de datos, identificando valores faltantes, nulos y duplicados. Para el conjunto de datos utilizado no se encontraron valores faltantes ni duplicados.

La extracción de características se lleva a cabo utilizando la librería TSFEL (Barandas et al., 2020). Para este estudio las características extraídas en el dominio del tiempo son: la media, desviación estándar, asimetría, kurtosis, rango intercuartílico, máximo, mínimo, varianza, raíz cuadrática media, entropía y diferencia absoluta media, tal como se expone en la tabla 1. En el dominio de la frecuencia se selecciona la frecuencia media, espectro de potencia máxima, frecuencia media, entropía espectral, kurtosis espectral, asimetría espectral y variación espectral. Las características se seleccionan tomando en cuenta aquellas características que se reportan en la literatura como las de mayor uso para la detección de fallas en maquinaria rotativa. Tabla 1. Conjunto de características extraídas en el dominio del tiempo

Característica	Fórmula
Media	
Desviación estándar	
Varianza	
Asimetría	
Kurtosis	
Rango intercuartílico	
Máximo	
Mínimo	
Raíz media cuadrática	
Diferencia absoluta media	
Entropía	

Para la extracción de características mediante la librería TSFEL se aplicó un tamaño de ventana de 30 y para extraer las características en el dominio de la frecuencia se asignó una frecuencia de muestreo de 30Hz, dicho valor se tiene como dato. Para un tamaño de ventana de 30 el nuevo tamaño del conjunto de datos de señales de vibración en condición

saludable es de 33.860, mientras que el nuevo tamaño del conjunto de datos para un estado de condición de diente roto es de 33.510 muestras, como se expone en la tabla 2. Además, dado que los datos de señales de vibración fueron recopilados para un estado de condición saludable y un estado de condición de diente roto, se tiene dos categorías o clases a predecir; la clase 0 para las etiquetas de clase negativa o sin falla y la clase 1 para las etiquetas positivas o con falla de diente roto, convirtiéndose en un problema de clasificación binaria. Tabla 2. Conjunto de características extraídas en el dominio de la frecuencia

Característica	Fórmula
Frecuencia máxima	
Potencia total	
Frecuencia media	
Entropía espectral	
Kurtosis de frecuencia	
Asimetría de frecuencia	
Varianza de frecuencia	

Las características extraídas se dividen en 70% para entrenamiento y 30% para prueba, obteniéndose los resultados que se muestran en la figura 3.

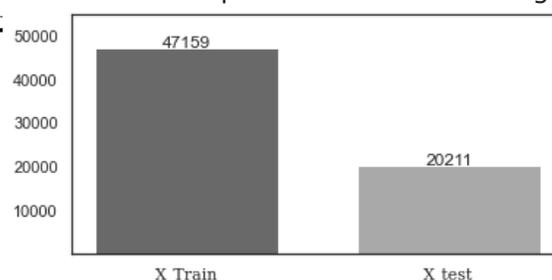


Figura 3. División de características para entrenamiento y prueba

Fuente: elaboración propia

Teniendo en cuenta la similitud entre características y sabiendo que el tiempo de cálculo y el costo computacional son directamente proporcionales a la cantidad de características disponibles, es importante y necesaria la reducción de características. Se aplicó un filtro de correlación con un umbral de 0.95 para reducir las características altamente correlacionadas. Posterior a la reducción de características se normalizó las características obtenidas, dado que las mismas se encuentran en diferentes escalas, para dicho propósito se utiliza el método *StandardScaler* mediante el cual las columnas de ca-

racterísticas se centran con una media de cero y una desviación estándar de uno.

La matriz de confusión es una herramienta muy útil en la evaluación del desempeño de un modelo de clasificación. En este estudio se utiliza los indicadores de desempeño descritos en la tabla 3 para evaluar y contrastar el desempeño entre los modelos propuestos y de esta manera definir e indicar el modelo óptimo. Además, se utiliza la curva ROC con su métrica de desempeño de área bajo la curva.

Tabla 3. Métricas de evaluación del desempeño

Métrica	Fórmula
Exactitud	
Precisión	
Sensibilidad	
Puntuación F1	

Las técnicas de clasificación por votación dura y votación suave se componen de cuatro clasificadores bases, los cuales son: *Support Vector Machine*, *random forest*, *XGBoost* y regresión logística, mientras que la técnica de apilamiento toma como aprendices bases o de primer nivel los mismos cuatro algoritmos utilizados por el clasificador de votación y utiliza como meta-clasificador o alumno de segundo nivel el algoritmo de regresión logística. Además, se realiza la optimización de hiperparámetros de cada clasificador base individual mediante la técnica de búsqueda de cuadrícula con validación cruzada de  $k=5$ . A partir de los clasificadores bases ajustados se crea un clasificador de voto mayoritario, un clasificador de voto suave y un modelo de conjunto basado en apilamiento. Los resultados obtenidos por cada clasificador individual, así como de los modelos de conjunto se evalúan con las métricas de la tabla 3 y se comparan los resultados.

### Resultados de clasificación

Para evaluar la efectividad de los modelos de aprendizaje conjunto propuestos, primero se evaluó el desempeño individual de cada clasificador base y luego se comparó con los modelos de conjunto. Todos los clasificadores fueron entrenados y evaluados en la misma partición de datos de entrenamiento y en la misma partición de datos de prueba, utilizando parámetros optimizados. Los resultados de los cuatro clasificadores base, así como de los clasificadores de voto mayoritario y voto suave aplicados al conjunto de características extraídas se muestran en la tabla 4.

Tabla 4. Resultados de clasificación

Clasificador	Exactitud	Precisión
voto suave	0.9986	0.9982
voto mayoritario	0.9981	0.9986
XGBoost	0.9980	0.9975
Support Vector Machine	0.9980	0.9978
Random forest	0.9957	0.9944
Regresión logística	0.9936	0.9937

Fuente: Elaboración propia

Todos los clasificadores base lograron un excelente rendimiento de clasificación, sin embargo, de los cuatro clasificadores base el que tiene el mejor rendimiento es XGBoost el cual obtuvo una exactitud de 99.80%, precisión de 99.78% y puntaje f1 de 99.80%, mientras que el clasificador base que tiene el rendimiento más bajo, aunque con una mínima diferencia es la regresión logística que obtuvo una exactitud de 99.36%, precisión de 99.37% y puntaje f1 de 99.35%. El objetivo es mostrar que los modelos de aprendizaje por conjunto propuestos pueden lograr un mejor rendimiento de clasificación y compensar las debilidades de cada clasificador base individual y eso es justamente lo que se evidencia en la Tabla 4. Los modelos de conjunto de clasificación por voto suave y voto mayoritario obtuvieron un mejor rendimiento de clasificación de fallas, en el caso del clasificador por voto suave obtuvo una exactitud de 99.86%, precisión de 99.82% y puntaje f1 de 99.86% logrando una mejora en exactitud, precisión y puntaje f1 con respecto al mejor clasificador base de 0.006%, 0.007% y 0.006% respectivamente. Por otro lado, el modelo de conjunto basado en clasificación por votación mayoritaria obtuvo una exactitud de 99.81%, precisión de 99.86% y puntaje f1 de 99.81% logrando una mejora en exactitud, precisión y puntaje f1 con respecto al mejor clasificador base de 0.01%, 0.11% y 0.01% respectivamente. La figura 4 ilustra de mejor manera el margen de mejora obtenido por los clasificadores de votación.

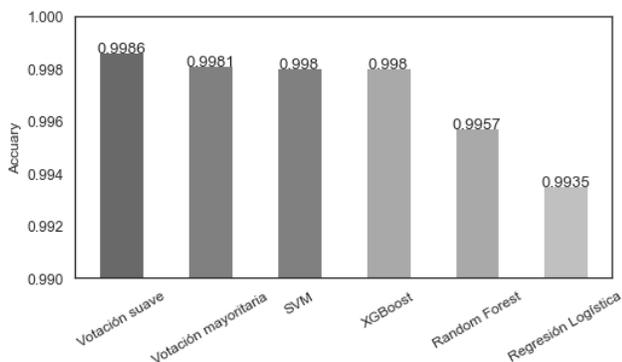


Figura 7. Puntuación de clasificación

Fuente: elaboración propia

Para construir el modelo mediante la técnica de apilamiento se utilizó *StackingCVClassifier* que emplea el concepto de validación cruzada y es una variante mejorada de la técnica de apilamiento estándar *StackingClassifier* en la cual puede existir sobreajuste. El modelo de apilamiento toma como clasificadores de primer nivel a, *Support Vector Machine*, *XGBoost*, *random forest* y regresión logística y como meta-clasificador o aprendiz de segundo nivel la regresión logística. Los resultados del modelo de aprendizaje de conjunto basado en apilamiento se muestran en la tabla 5, donde es comparado con los clasificadores de primer nivel y los clasificadores de aprendizaje de conjunto basado en votación suave y votación mayoritaria.

Tabla 5. Comparación de resultados de clasificación

Clasificador	Exactitud	Precisión	F1 Score
voto suave	0.9986	0.9982	0.9986
Meta-clasificador	0.9982	0.9985	0.9982
voto mayoritario	0.9981	0.9986	0.9981
XGBoost	0.9980	0.9975	0.9980
Support Vector Machine	0.9980	0.9978	0.9980
Random forest	0.9957	0.9944	0.9957
Regresión logística	0.9935	0.9937	0.9935

El modelo de conjunto basado en apilamiento obtuvo una exactitud de 99.82%, precisión de 99.85%, puntaje f1 de 99.82% y un AUC de 99.80% situándose en segundo lugar en términos de exactitud, mientras que en términos de precisión el clasificador de voto mayoritario es 0.01% más preciso, tal como se ilustra en la figura 8.

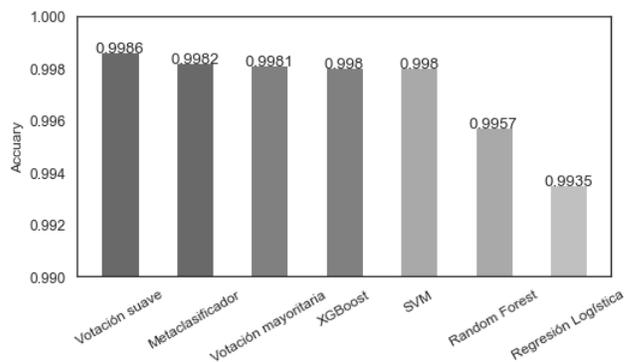


Figura 8. Puntuación de clasificación

Fuente: elaboración propia

De los tres modelos de conjunto desarrollados el clasificador de voto suave lidera y muestra un mejor rendimiento de clasificación de fallas, mostrando una mejora en exactitud de 0.04% con respecto al modelo de conjunto basado en apilamiento y 0.05% con respecto al clasificador de votación mayoritaria, mientras que en términos de precisión el clasificador de voto mayoritario es 0.04% más preciso que el clasificador de voto suave y 0.01% más preciso que el modelo de conjunto basado en apilamiento. Los tres métodos de conjunto producen mejores resultados de exactitud, precisión y puntuación F1 que cualquiera de los cuatro algoritmos de clasificación individuales, por lo tanto, se comprueba y se demuestra que los tres enfoques propuestos basados en aprendizaje conjunto pueden lograr un mejor rendimiento de clasificación y detección de fallas.

### CONCLUSIONES

En este trabajo se propuso tres modelos de aprendizaje conjunto, utilizando las técnicas de clasificación por votación mayoritaria, votación suave y un modelo bajo el enfoque de apilamiento. El objetivo fue utilizar el aprendizaje conjunto para combinar las ventajas de diferentes algoritmos y obtener un mejor rendimiento de clasificación y detección de fallas en cajas de engranajes.

En comparación con los valores de precisión de cada clasificador individual las precisiones de los modelos de conjunto fueron mejores. El clasificador de votación suave fue el modelo de aprendizaje conjunto de mejor rendimiento el cual obtuvo una exactitud de 99.86%, precisión de 99.82% y puntaje F1 de 99.86% seguido del modelo de aprendizaje conjunto basado en apilamiento que obtuvo una exactitud de 99.82%, precisión de 99.85% y puntaje f1 de 99.82% y por último del clasificador

por votación mayoritaria que obtuvo una exactitud de 99.81%, precisión de 99.86% y puntaje f1 de 99.91% los cuales en comparación con los clasificadores base obtuvieron un mejor rendimiento, mostrando ser métodos prácticos y eficientes para clasificar y detectar fallas con mayor precisión.

Para el conjunto de datos utilizado el rendimiento obtenido por cada uno de los clasificadores base es excelente cada uno obtuvo un puntaje de precisión superior a 99%, sin embargo, cuando se trata de clasificar y detectar fallas la exactitud y precisión son importantes por lo tanto se debe tratar de conseguir la mejor exactitud y precisión posible y para lograr dicho fin en este documento se demostró que mediante la combinación de las predicciones de varios clasificadores, se puede obtener un modelo con mejor rendimiento en términos de precisión.

La precisión del mejor clasificador base fue de 99.78%, mientras que la precisión del mejor modelo conjunto es de 99.86% mostrando una mejora no muy significativa, pero que a la hora de tomar una decisión en un entorno profesional marca diferencias en el orden práctico.

#### REFERENCIAS BIBLIOGRÁFICAS

- Azamfar, M., Singh, J., Bravo-Imaz, I., & Lee, J. (2021). Multisensor data fusion for gearbox fault diagnosis using 2-D convolutional neural network and motor current signature analysis. *Artificial Intelligence in Engineering*, 713–721.
- Barandas, M., Folgado, D., Fernandes, L., Santos, S., Abreu, M., Bota, P., ... & Gamboa, H. (2020). TSFEL: Time series feature extraction library. *SoftwareX*, 11, 100456.
- Cao, J., Kwong, S., Wang, R., Li, X., Li, K., & Kong, X. (2015). Class-specific soft voting based multiple extreme learning machines ensemble. *Neurocomputing*, 149, 275–284
- Devendiran, S., & Manivannan, K. (2016). Vibration based condition monitoring and fault diagnosis technologies for bearing and gear components—a review. *International Journal of Applied Engineering Research*, 11(6), 3966–3975.
- Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. *Frontiers of Computer Science*, 14, 241–258.
- García, Mora, F. G., Davila, E. H., & Villa, J. C. (2021). Clasificación de fallas en rodamientos utilizando aprendizaje de máquinas. *Dominio de las Ciencias*, 7(4), 70–89.
- Han, T., Jiang, D., Zhao, Q., Wang, L., & Yin, K. (2018). Comparison of random forest, artificial neural networks and support vector machine for intelligent diagnosis of rotating machinery. *Transactions of the Institute of Measurement and Control*, 40(8), 2681–2693.
- Osman, A. I. A., Ahmed, A. N., Chow, M. F., Huang, Y. F., & El-Shafie, A. (2021). Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia. *Ain Shams Engineering Journal*, 12(2), 1545–1556.
- Ibrahim, I., & Abdulzeez, A. (2021). The role of machine learning algorithms for diagnosing diseases. *Journal of Applied Science and Technology Trends*, 2(01), 10–19.
- Jones, I. A., Van Oyen, M. P., Lavieri, M. S., Andrews, C. A., & Stein, J. D. (2021). Predicting rapid progression phases in glaucoma using a soft voting ensemble classifier exploiting Kalman filtering. *Health Care Management Science*, 24(4), 686–701.
- Kumar, U. K., Nikhil, M. B. S., & Sumangali, K. (2017). 1-vote-Naive Bayes SMV J48. pdf.
- Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., & Nandi, A. K. (2021). Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical Systems and Signal Processing*, 138, 106587.
- Mahabub, A. (2020). A robust technique of fake news detection using Ensemble Voting Classifier and comparison with other classifiers. *SN Applied Sciences*, 2(4), 525.
- Patil, P. S., Patil, M. S., Tamhankar, S. G., Patil, S., & Kazi, F. (2021). Majority Voting Machine Learning Approach for Fault Diagnosis of Mechanical Components. In *Applications of Artificial Intelligence in Engineering: Proceedings of First Global Conference on Artificial Intelligence and Applications (GCAIA 2020)* (pp. 713–721). Springer Singapore.
- Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), e1249.
- Sarajcev, P., Kunac, A., Petrovic, G., & Despalatovic, M. (2021). Power system transient stability assessment using stacked autoencoder and voting ensemble. *Energies*, 14(11), 3148.
- Wang, G., Hao, J., Ma, J., & Jiang, H. (2011). A comparative assessment of ensemble learning for credit scoring. *Expert systems with applications*, 38(1), 223–230.
- Wuest, T., Weimer, D., Irgens, C., & Thoben, K. D. (2016). Machine learning in manufacturing:

advantages, challenges, and applications. *Production & Manufacturing Research*, 4(1), 23-45.

Yu, B., Qiu, W., Chen, C., Ma, A., Jiang, J., Zhou, H., & Ma, Q. (2020). SubMito-XGBoost: predicting protein submitochondrial localization by fusing multiple feature information and eXtreme gradient boosting. *Bioinformatics*, 36(4), 1074-1081.

Zhou, Z. H. (2012). *Ensemble methods: foundations and algorithms*. CRC press.