

Procesamiento y clasificación de señales

Machine Learning

28 de septiembre de 2024

Cervantes Martínez Rodrigo Fabián



Campus Irapuato-Salamanca | División de Ingenierías

Resumen

Este reporte describe el análisis de señales de corriente trifásica medidas desde un motor de inducción de jaula de ardilla. Los ensayos experimentales se realizaron para diferentes niveles de falla en cada fase de bobinado, recogiendo datos tanto en un estado saludable como en condiciones de fallo. El conjunto de datos consta de 13 categorías: 12 fallos de cortocircuito entre espiras por fase al 10 %, 20 %, 30 % y 40 %, y en buen estado. Se realizaron cinco repeticiones para cada condición experimental, y todas las mediciones se recogieron en un estado estacionario sin carga.

En este trabajo, se aplica una ventana deslizante sobre las señales de cada trifase, calculando imágenes de suma y diferencia con los valores extremos de la ventana. A partir de estas imágenes, se generan histogramas y se calculan 7 propiedades clave para construir una matriz de características de dimensión 252. Posteriormente, se clasifica en 13 categorías utilizando los clasificadores Random Forest y Extra Trees, evaluando el impacto de la reducción de dimensionalidad mediante PCA.

1. Introducción

El análisis de señales de corriente trifásica provenientes de un motor de inducción de jaula de ardilla resulta fundamental para detectar fallos en los bobinados. Con estas señales, se aplican ventanas deslizantes para calcular imágenes de suma y diferencia a partir de los valores extremos de cada ventana. Estas imágenes permiten extraer características locales que, posteriormente, se utilizan para clasificar las 13 categorías mediante clasificadores de árboles de decisión. Además, se evalúan técnicas de reducción de dimensionalidad como PCA con el fin de mejorar el rendimiento del clasificador.

Los objetivos principales de este ejercicio son:

- Aplicar ventanas a las señales de corriente para generar imágenes de suma y diferencia.
- Calcular 7 propiedades a partir de los histogramas de suma y diferencia.
- Realizar la clasificación de 13 categorías utilizando árboles de decisión.
- Evaluar el impacto de la reducción de dimensionalidad mediante PCA.

2. Metodología

2.1. Aplicación de ventanas

A cada fase de corriente (ver Figura 1) se le aplicó una ventana deslizante. El tamaño de la ventana fue de n valores y el desplazamiento se realizó de un valor a la vez. Para cada ventana, se calcularon las imágenes de suma y diferencia considerando el primer y el último valor de la ventana. Este proceso se repitió para las tres fases (A, B, C) y para un total de 12 ventanas por cada fase.

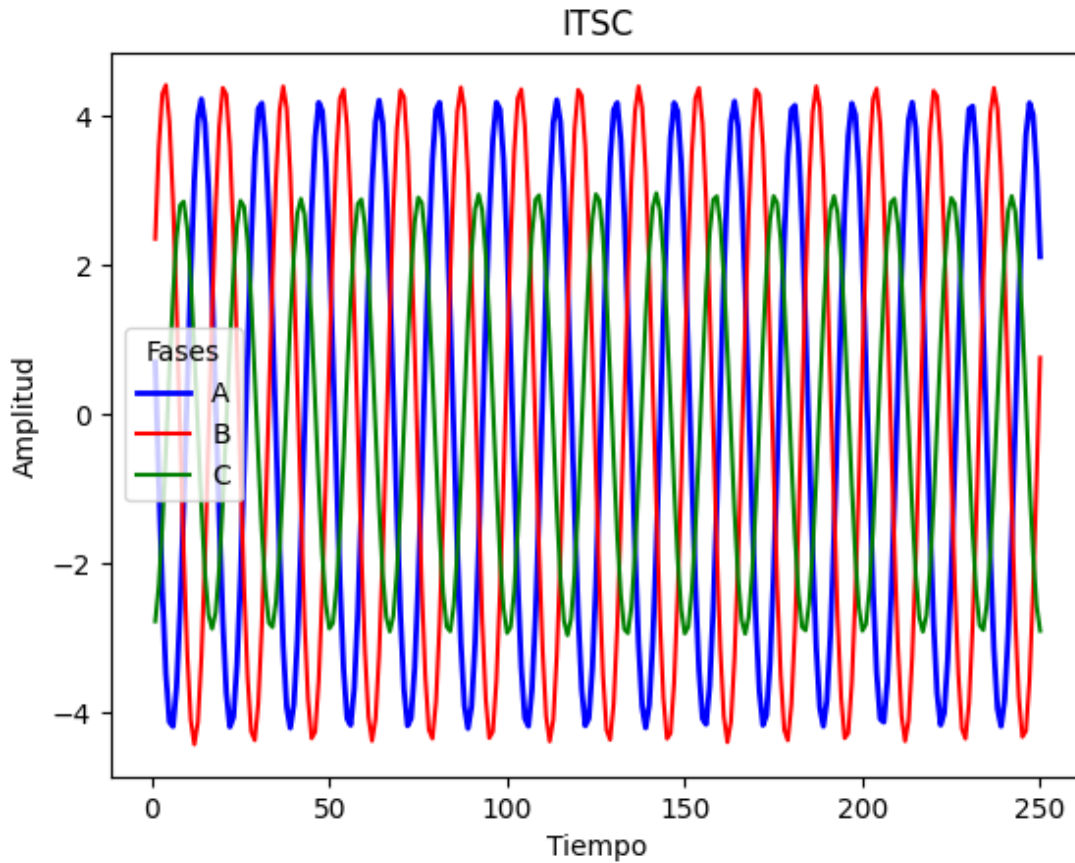


Figura 1: Visualización de cada fase de corriente.

2.2. Construcción de histogramas

Se generaron histogramas de las imágenes de suma y diferencia para cada ventana, considerando los valores máximos y mínimos en cada imagen. Esto permitió caracterizar las variaciones en los datos a lo largo de las señales.

2.3. Cálculo de las 7 propiedades

A partir de los histogramas de suma y diferencia, se calcularon las siguientes 7 propiedades para cada ventana:

1. Media.

2. Varianza.
3. Correlación.
4. Contraste.
5. Homogeneidad.
6. Sombra de clúster.
7. Prominencia de clúster.

Estas propiedades se calcularon para cada una de las 6 ventanas por fase, lo que resulta en 84 características para cada fase de la señal (A, B, C), sumando un total de 252 características.

2.4. Clasificación de las 13 categorías

Se utilizaron los clasificadores Random Forest y Extra Tress para realizar la clasificación de las 13 categorías a partir de las características calculadas. Se emplearon las siguientes métricas de evaluación:

- Accuracy.
- Precision.
- Recall.
- Matriz de confusión.

3. Resultados

3.1. Histogramas

Se obtuvieron los histogramas de las imágenes de suma y diferencia para cada fase de corriente. A continuación, se presentan algunos ejemplos representativos de estos histogramas (ver Figura 2).

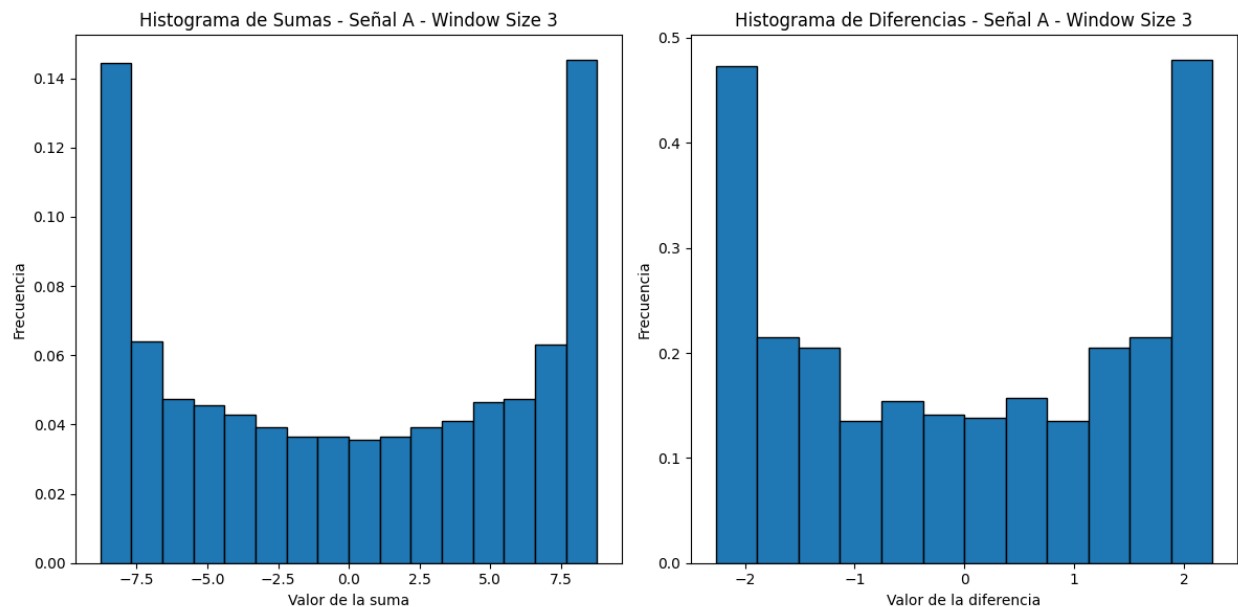


Figura 2: Histogramas de la imagen de suma y diferencia.

3.2. Matriz de características

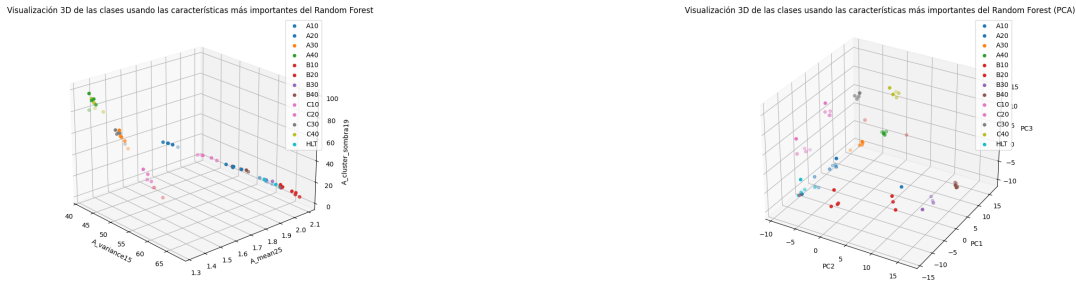
La matriz de características resultante tiene una dimensión de 252 (7 propiedades para cada una de las 12 ventanas aplicadas en cada fase). Además, se incluye una columna adicional para la clase correspondiente, lo que da un total de 253 columnas. A continuación, se muestra un fragmento de esta matriz.

| Propiedad | Fase A | Fase B | Fase C |
|--------------------------|--------|--------|--------|
| Contraste (ventana 3) | 0.23 | 0.35 | 0.27 |
| Correlación (ventana 3) | 0.89 | 0.78 | 0.91 |
| Contraste (ventana 25) | 0.21 | 0.33 | 0.25 |
| Correlación (ventana 25) | 0.88 | 0.79 | 0.90 |
| ... | ... | ... | ... |

Cuadro 1: Fragmento de la matriz de características con 126 columnas de propiedades más la columna de clase.

3.3. Distribución de Características en 3D

Las siguientes figuras muestran la distribución de las características calculadas en un espacio tridimensional, proporcionando una visualización de cómo se agrupan las clases en función de las características seleccionadas.



(a) Distribución sin PCA

(b) Distribución con PCA

Figura 3: Distribución tridimensional de las características extraídas de las señales con y sin PCA.

3.4. Clasificación

Los resultados del clasificador de árboles de decisión se resumen en la Tabla 3. Se observa un rendimiento aceptable en la clasificación sin reducción de dimensionalidad.

| Clasificador | Parámetros |
|------------------------|--|
| RandomForestClassifier | n_estimators=100, max_depth=10, min_samples_split=5, criterion='entropy' |
| ExtraTreesClassifier | n_estimators=100, max_depth=None, min_samples_split=2, criterion='gini' |

Cuadro 2: Valores de los parámetros para cada clasificador

| Clasificador | Accuracy | Precision | Recall | Tiempo de Entrenamiento (s) |
|---------------|----------|-----------|--------|-----------------------------|
| Random Forest | 0.90 | 0.93 | 0.90 | 0.09 |
| Extra Trees | 0.90 | 0.94 | 0.90 | 0.04 |

Cuadro 3: Resultados de los clasificadores entrenados.

3.5. Reducción de dimensionalidad

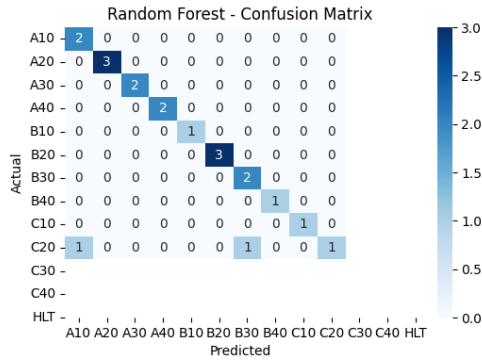
Los resultados con reducción de dimensionalidad PCA se comparan en la Tabla 4. Se observa un descenso en el rendimiento utilizando PCA.

| Clasificador | Accuracy | Precision | Recall | Tiempo de Entrenamiento (s) |
|---------------|----------|-----------|--------|-----------------------------|
| Random Forest | 0.85 | 0.91 | 0.85 | 0.07 |
| Extra Trees | 0.95 | 0.97 | 0.95 | 0.05 |

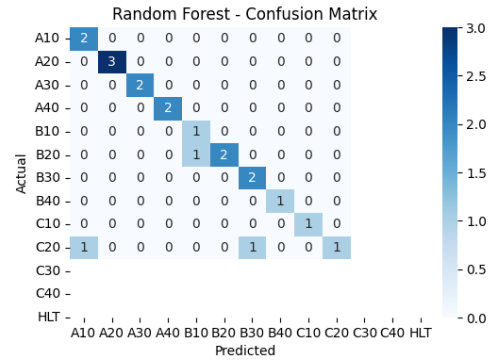
Cuadro 4: Resultados de los clasificadores entrandos usando PCA.

3.6. Matriz de Confusión

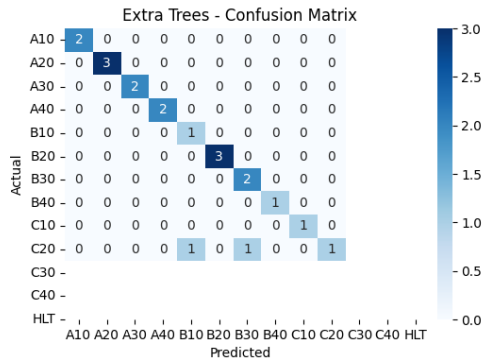
A continuación, se presentan las matrices de confusión para ambos clasificadores, tanto con como sin PCA. Es importante señalar que la ausencia de datos en las matrices se debe a la escasez de ejemplos disponibles tanto para el entrenamiento como para la predicción.



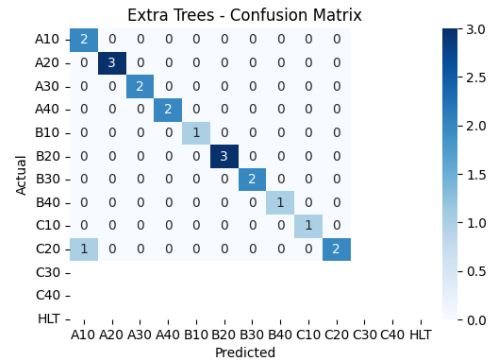
(a) Random Forest sin PCA



(b) Random Forest con PCA



(c) Extra Trees sin PCA



(d) Extra Trees con PCA

Figura 4: Matrices de confusión para los clasificadores Random Forest y Extra Trees, con y sin PCA.

4. Discusión

Los resultados obtenidos con los clasificadores *Random Forest* y *Extra Trees* muestran un desempeño considerable en la clasificación de las 13 categorías de fallas en el motor de inducción. El *Extra Trees* obtuvo la mejor precisión y recall tanto con como sin la reducción de dimensionalidad utilizando PCA, alcanzando

un valor máximo de **95 %** de exactitud sin PCA. Esto indica que, a pesar de la alta dimensionalidad del conjunto de características, el *Extra Trees* fue capaz de generalizar mejor que el *Random Forest*, manteniendo la capacidad de distinguir con precisión las diferentes clases.

Sin embargo, se observó una disminución en el rendimiento de ambos clasificadores cuando se aplicó PCA. La reducción de características a través de PCA, aunque útil para evitar el sobreajuste en algunos casos, pareció eliminar características relevantes para la identificación de las fallas. En particular, el *Random Forest* se vio más afectado, disminuyendo su precisión de un 90 % a un 85 % con la aplicación de PCA, lo cual sugiere que este modelo depende más de la información completa del conjunto de datos para alcanzar su mejor rendimiento.

Otro punto relevante es el tiempo de entrenamiento. Mientras que el tiempo de entrenamiento de *Random Forest* disminuyó marginalmente al usar PCA, el clasificador *Extra Trees* mantuvo tiempos de entrenamiento rápidos, independientemente de la reducción de características. Esto demuestra la capacidad de *Extra Trees* para procesar datos complejos de manera eficiente y refuerza su idoneidad para la clasificación de señales con gran cantidad de características.

Las matrices de confusión también revelaron que algunas categorías fueron confundidas con mayor frecuencia, particularmente aquellas con fallos de cortocircuito en niveles bajos (10 %). Estas categorías presentaron características similares a las del estado saludable, dificultando la separación de clases. Para mejorar este aspecto, se sugiere explorar técnicas adicionales de selección de características y métodos de clasificación más avanzados que consideren la similitud entre clases.

5. Conclusiones

El presente trabajo demostró la efectividad de los clasificadores basados en árboles de decisión para la identificación de fallos en motores de inducción. De manera particular, el clasificador *Extra Trees* presentó un rendimiento superior al *Random Forest*, con una precisión del 95 % sin aplicar reducción de dimensionalidad. La aplicación de PCA, aunque disminuyó la dimensionalidad de las características, resultó en una ligera pérdida de rendimiento en ambos clasificadores, lo cual resalta la importancia de preservar la mayor cantidad de información posible en estos casos.

En términos computacionales, *Extra Trees* también destacó por su eficiencia de tiempo, lo cual lo convierte en una opción robusta para aplicaciones en tiempo real. Sin embargo, el análisis de las matrices de confusión indica que ciertas categorías de fallas, como los cortocircuitos de bajo nivel, siguen siendo un desafío. Esto abre la puerta a futuras investigaciones que incluyan la integración de clasificadores híbridos, la incorporación de características de dominio temporal y espectral, así como la evaluación de técnicas de *ensemble learning* más sofisticadas.

Finalmente, el estudio sugiere que, si bien la reducción de dimensionalidad puede ser útil en algunos casos, es fundamental evaluar su impacto en la clasificación, especialmente cuando se trabaja con señales con características intrincadas y multivariadas. Se recomienda para trabajos futuros analizar la sensibilidad de las características calculadas e investigar la incorporación de métodos de selección de características basados en importancia y relevancia para cada clase.