

# Informe final Taller Fourier - EDM

1<sup>st</sup> Joel Carabajal Fernández  
Facultad de Ingeniería  
Universidad de la República  
Montevideo, Uruguay  
joelcarabajal98@gmail.com

2<sup>nd</sup> Carlos Gruss  
Facultad de Ingeniería  
Universidad de la República  
Montevideo, Uruguay  
carlosgruss99@gmail.com

3<sup>rd</sup> Felipe López Folle  
Facultad de Ingeniería  
Universidad de la República  
Montevideo, Uruguay  
felipelopezfolle@gmail.com

**Abstract**—El problema a resolver se basa en identificar tipos de electrodomésticos a través de su voltaje y corriente, y a partir de eso caracterizarlo con el fin de poder determinar dicho electrodoméstico sin necesidad de un método invasivo. Donde una de las aplicaciones posibles donde podría ser muy útil sería para el diseño de tarifas inteligentes de consumo eléctrico, desde información extraída externamente, mejorando así el consumo energético general. Logrando así una mayor eficiencia en el suministro de energía eléctrica por parte de las empresas encargadas de esto. Los resultados obtenidos fueron positivos usando distintos métodos de aprendizaje automático, permitiendo esto, comparar ventajas y desventajas de cada uno de ellos que se detallan mas adelante.

**Index Terms**—harmonic analysis, signal processing, machine learning, IEEE power definitions

## I. INTRODUCCIÓN

La motivación de este trabajo es caracterizar los consumos eléctricos de distintos tipos de electrodomésticos. Esto se hace sobre un grupo acotado de tipos de electrodomésticos, que son los más usualmente encontrados en los hogares:

- 0) Aire acondicionado ("Air Conditioner"),
- 1) Lámpara fluorescente ("Compact Fluorescent Lamp"),
- 2) Ventilador ("Fan"),
- 3) Heladera ("Fridge"),
- 4) Secador de pelo ("Hairdryer"),
- 5) Estufa eléctrica ("Heater"),
- 6) Lámpara incandescente ("Incandescent Light Bulb"),
- 7) Laptop,
- 8) Microondas ("Microwave"),
- 9) Aspiradora ("Vacuum"),
- 10) Lava ropas ("Washing Machine").

Estos tipos de electrodomésticos se eligen en particular dado que son los tipos disponibles en la base "Plaid". Esta es una compilación de señales de corriente-voltaje de dos a cinco segundos, que se utilizan como referencia a lo largo de este trabajo. [1] Se debe tener en cuenta que estas señales se midieron con frecuencia de línea y frecuencia de muestreo respectivamente:

$$F_l = 60Hz \quad (1)$$

$$F_s = 30kHz \quad (2)$$

Para analizar las señales, se usan las definiciones de potencia del estándar IEEE-1459-2010 [2]. Asimismo, para evaluar el funcionamiento de la implementación de estas definiciones,

se utilizan los ejemplos disponibles en el artículo "The IEEE standard 1459: What and why?" [3].

Una vez implementada la adquisición y caracterización de señales, se adquieren señales de electrodomésticos propios, para así crear junto a otros compañeros del curso una base propia de señales de corriente-voltaje.

En la siguiente sección (II) se presenta la solución propuesta para la adquisición y clasificación de las señales. Luego, en la sección sección III se muestra brevemente el armado del hardware utilizado. Finalmente, en la sección IV se tienen los resultados del trabajo.

## II. SOLUCIÓN PROPUESTA

### A. Diseño del hardware de adquisición y adecuación de las señales

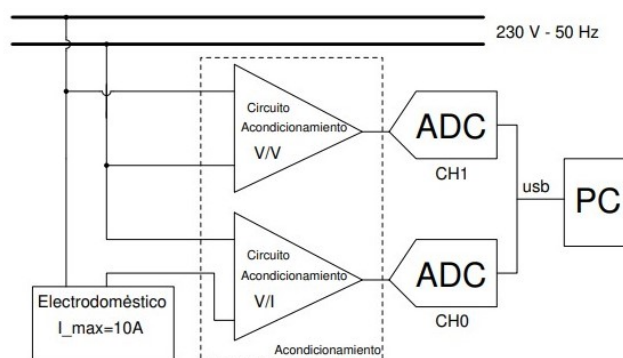


Fig. 1. Esquema de circuitos implicados para medir la tensión y corriente sobre un electrodoméstico

El objetivo fue diseñar los circuitos necesarios para medir adecuadamente la tensión y corriente consumidas por un electrodoméstico conectado a la red monofásica de baja tensión de UTE, tal como se muestra en la figura 1.

Una importante herramienta para el desarrollo de la solución fue el "laboratorio portátil Analog Discovery 2" ("AD2"), que se utilizó como herramienta para la verificación de los circuitos de acondicionamiento, como osciloscopio para la adquisición de señales e incluso como controlador de hardware.

También se contó con un "monitor" para los electrodomésticos, provista por el equipo docente del instituto. Esta proporciona los mecanismos necesarios para conectar un electrodoméstico a la red eléctrica, incorporando sensores

de las magnitudes de interés. La solución presentada en este informe trata a dicho monitor como "caja negra" de la cual se extraen cuatro señales, las cuales se quieren acondicionar para medir.

Con este fin, se diseñaron circuitos para la comunicación entre los dos dispositivos anteriormente nombrados. Por último, se hizo uso de una computadora para controlar la tarjeta AD2 y para encargarse del procesamiento y clasificación de las señales.

Los circuitos diseñados se pueden dividir en 2 módulos. En primer lugar, se tiene un circuito de acondicionamiento de voltaje a voltaje ("V/V").

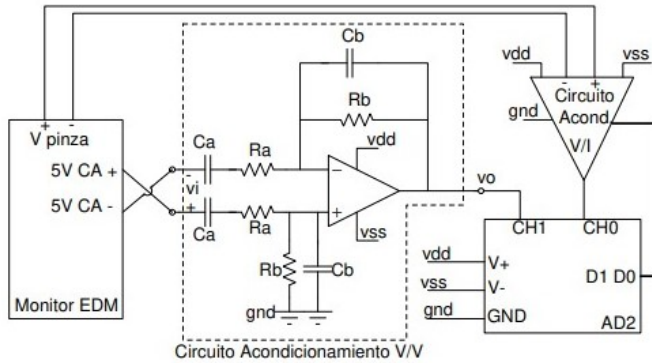


Fig. 2. Circuito de acondicionamiento V/V.

El circuito de la figura 2 muestra la interacción entre el monitor y la tarjeta AD2 a través de los circuitos de acondicionamiento. Como se puede ver, el monitor provee las cuatro señales (salida de transformador de voltaje y salida del sensor de corriente) y luego estos se condicionan por la electrónica desarrollada. Observar que la tarjeta AD2 provee los voltajes de alimentación  $V_{DD} = +5VDC$ ,  $V_{SS} = -5VDC$  y  $GND = 0VDC$  necesarios para los circuitos de acondicionamiento.

Las especificaciones de diseño para el circuito de acondicionamiento V/V exigen una respuesta pasa-banda con polo de baja frecuencia a una frecuencia inferior en  $5Hz$ , polo de alta frecuencia a una frecuencia de alrededor de  $2kHz$  y ganancia de  $0,45V/V$  en banda pasante.

Para el diseño se eligieron componentes resistivos de la serie E12 y capacitivos cerámicos de la E3. El amplificador operacional utilizado fue el "TLV274" para todos los circuitos de acondicionamiento. Los valores utilizados fueron:

- $C_a = 22nF$
- $C_b = 100pF$
- $R_a = 1,8M\Omega$
- $R_b = 820k\Omega$

Para estos valores, se tiene frecuencia inferior en  $4Hz$ , superior en  $2kHz$  y ganancia pasante de  $0,47V/V$ . Por elección de diseño, se fue conservador con algunos valores, prefiriendo arriesgar a estar un poco por encima del ancho de banda y ganancia requeridas.

En segundo lugar, se tiene el módulo de acondicionamiento voltaje a corriente ("V/I"). El monitor internamente utiliza un sensor de corriente no invasivo tipo pinza amperimétrica, entregando así un voltaje  $V_{pinza}$  que se quiere acondicionar.

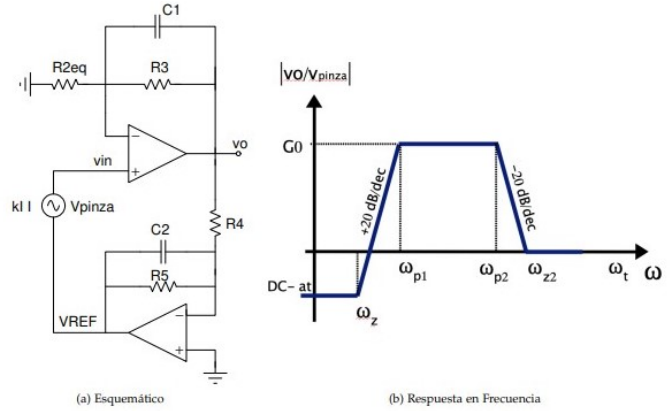


Fig. 3. Esquemático de circuito amplificador con ganancia programable ("PGA")

En la figura 3, se puede ver el esquemático del circuito diseñado para la comunicación entre la tarjeta AD2 y el monitor, junto con su respuesta en frecuencia.

La ganancia programable se debe a  $R_{2eq}$ , que fue diseñada de forma que, para las señales esperadas en  $v_{in}$ , siempre se puede seleccionar un valor para el que se cumpla  $V_o > 1V$ . Esto se implementó con resistencias en paralelo y el chip de llaves analógicas "Intersil DG412".

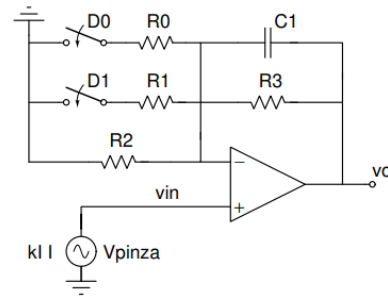


Fig. 4. Implementación de  $R_{2eq}$ . Las llaves se manejan con las salidas digitales del AD2.

Las especificaciones de diseño para este circuito fueron: polo de baja frecuencia a menos de  $5Hz$ , polo de alta frecuencia a mas de  $2kHz$ , ganancia variable programable en banda pasante de  $11V/V$ ,  $101V/V$  o  $1001V/V$  y atenuación en  $DC$  para anular el efecto del offset a la salida. Por ultimo se le añadió a la salida del circuito un filtro  $RC$  que atenúa las altas frecuencias (frecuencias mayores a  $20kHz$ ). Los valores de componentes utilizados para esto fueron:

- $R_0 = 6,8k\Omega$
- $R_1 = 0,68k\Omega$
- $R_2 = 68k\Omega$
- $R_3 = 820k\Omega$

- $R_4 = 1M\Omega$
- $R_5 = 3,3M\Omega$
- $C_1 = 100pF$
- $C_2 = 47\mu F$

Para estos valores, se cumplen los requerimientos de ancho de banda y offset, y se tienen las siguientes ganancias para las distintas configuraciones de  $[D_1, D_0]$ :

- $[OFF, OFF] : 13V/V$
- $[OFF, ON] : 126V/V$
- $[ON, OFF] : 1000V/V$

Se observa que el mayor problema de estas ganancias es que la corriente máxima que se podrá medir en electrodomésticos es menor que la especificada, pero la diferencia es pequeña y con las ganancias obtenidas se pudo medir una amplia gama de electrodomésticos.

A partir de las ganancias obtenidas para los circuitos de acondicionamiento y la constante de transducción del sensor de corriente, se tiene una serie de constantes por las cuales se deben multiplicar las señales medidas por el AD2 a la hora de registrar un electrodoméstico:

- $K_V = 98$
- $K_{I00} = 2, 3$
- $K_{I01} = 0, 24$
- $K_{I10} = 0, 03$

### B. Análisis de la señal, propuesta de descriptores relevantes

El primer paso hacia el análisis de las señales es encontrar caracterizadores relevantes. Cómo se menciona en la sección I, en primera instancia se utilizaron las definiciones de potencia presentadas en el estándar IEEE-1459-2010 [2]:

- $S$ : Apparent power,
- $S_{11}$ : Fundamental apparent power,
- $S_H$ : Harmonic apparent power,
- $S_N$ : Non-fundamental apparent power,
- $P$ : Active power,
- $P_{11}$ : Fundamental active power,
- $P_H$ : Harmonics active power,
- $Q_{11}$ : Fundamental reactive power,
- $D_I$ : Current distortion power,
- $D_V$ : Voltage distortion power,
- $D_H$ : Harmonic distortion power,
- $N$ : Non-active apparent power,
- $THD_V$ : Total harmonic distortion for voltage,
- $THD_I$ : Total harmonic distortion for current.

Estas características se calculan utilizando la Transformada de Fourier de las señales de corriente-voltaje, y caracterizan las posibles relaciones entre armónicos de las señales. Una guía práctica para la implementación junto con ejemplos para ponerla a prueba se encuentra en "The IEEE standard 1459: What and why?" [3]. Consecuentemente, de estas características se desprenden de manera directa las siguientes:

- $FP$ : Power factor,
- $FP_{11}$ : Fundamental power factor.

Estas se calculan como:

$$FP = \frac{P}{S} \quad (3)$$

$$FP = \frac{P_{11}}{S_{11}} \quad (4)$$

Por otra parte, se observó durante experimentos preliminares que graficar la corriente en función del voltaje producía diagramas con formas características al tipo de electrodoméstico. Por lo tanto, se adapta un algoritmo presentado en "Plaid: a public dataset of high-resolution electrical appliance measurements for load identification research" [1] para, dada una señal de corriente-voltaje, calcular el diagrama  $V-I$  y devolverlo discretizado como una matriz. Este algoritmo admite argumentos para discretizar en valores reales o cuantificar estos en unos y ceros, y también permite modificar el tamaño de la matriz, afectando la resolución. Para la solución presentada, se decidió cuantizar en unos y ceros, y utilizar un tamaño de imagen de  $16 \times 16$ , dado que mediante experimentación se determinó que estos parámetros maximizaban el rendimiento sin subir demasiado el costo computacional.

Integrando todo lo anterior, se tiene para cada señal un arreglo de 17 características para calificarlas.

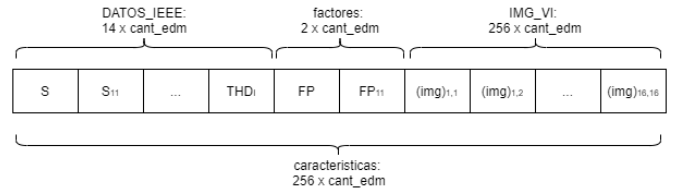


Fig. 5. Características utilizadas para clasificar

### C. Clasificadores utilizados

Para clasificar las señales a partir de sus características, se utilizan los clasificadores de la librería "scikit-learn" para Python [4]. En particular, las pruebas se realizaron con los métodos "K-Nearest Neighbors" y "Random Forest". A continuación se explica brevemente el funcionamiento de estos.

El método "K-Nearest Neighbors" [5] tiene la ventaja de ser muy intuitivo en su funcionamiento. Dadas  $n$  características para clasificar las señales, se ubican las señales de entrenamiento en  $\mathbb{R}^n$  según los valores de sus características. Luego, cada nueva señal a clasificar se ubica en este espacio vectorial. Dado un  $k \in \mathbb{N}^*$ , se encuentran los  $k$  puntos más cercanos (vecinos) al nuevo punto. Entonces, se hace un recuento de a que clase pertenecen estos puntos y se clasifica a la nueva señal a la clase a la que pertenezca la mayor cantidad de vecinos.

La principal ventaja de este método es su simplicidad y bajo costo computacional relativo, dado que ofrece resultados relativamente buenos. Lo principal a tener en cuenta en la implementación es que valores de  $k$  demasiado grandes o chicos no son representativos, y ciertos valores pueden tener el problema de generar casos de "empate" en la clasificación.

Para evitar esto último, se recurrió a utilizar un  $k$  mayor al número de clases que fuera número primo, para que siempre hubiera un vecino que cause el "desempate".

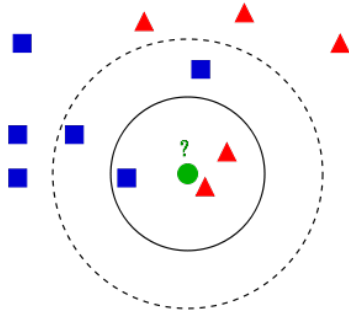


Fig. 6. Método "K-Nearest Neighbors" [6]

Por otro lado, el método "Random Forest" [7] toma como principal argumento un "número de árboles"  $a$ . Dado este valor, el método forma  $a$  subconjuntos aleatorios de la base de entrenamiento. Para cada subconjunto, se crea un árbol de decisión para clasificar futuras señales. Luego, cada nueva señal es pasada por todos los árboles de decisión, y se registra la clase que predice cada árbol. Finalmente, se clasifica como la clase que haya tenido mayor frecuencia de predicción.

Este método ofrece mejor rendimiento que "K-Nearest Neighbors", a cambio de mayor costo computacional. Cómo era también el caso antes, se debe elegir el valor de  $a$  de forma que las muestras no sean demasiado grandes o demasiado chicas, para evitar que el clasificador se sobre-ajuste a los datos de entrenamiento.

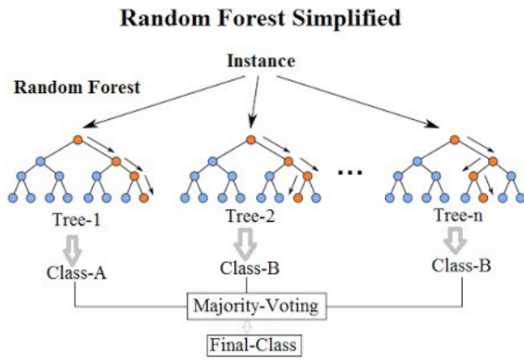


Fig. 7. Método "Random Forest" [8]

El entrenamiento de los clasificadores, se realizó siguiendo el esquema de "validación cruzada dejando uno afuera" [9]. Para esto, se separaron las señales de la base según la casa a la cual pertenecía el electrodoméstico medido. De esta forma, dadas  $N$  casas de dónde se tienen medidas, se entrena el clasificador seleccionando usando las señales pertenecientes a las  $N - 1$  casas restantes. Luego, se clasifican las  $N$  casas restantes y se evalúa que tan bien acertó el clasificador para la casa en cuestión.

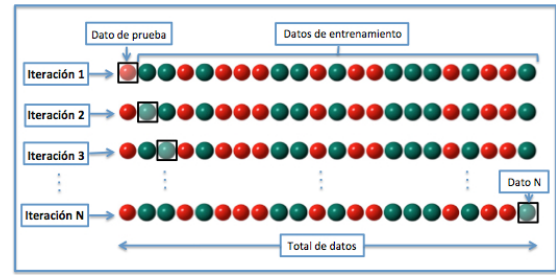


Fig. 8. Validación cruzada dejando uno afuera [10]

De esta manera, se termina entrenando al clasificador con todas las señales de la base y se evitan sesgos a la hora de aproximar el rendimiento real del clasificador.

### III. IMPLEMENTACIÓN

Para la implementación del hardware descrito en la sección II-A, luego de pruebas preliminares en proto-board, se realizó el desarrollo de un PCB utilizando la herramienta Autodesk EAGLE. En la figura 9 se puede observar el resultado final.

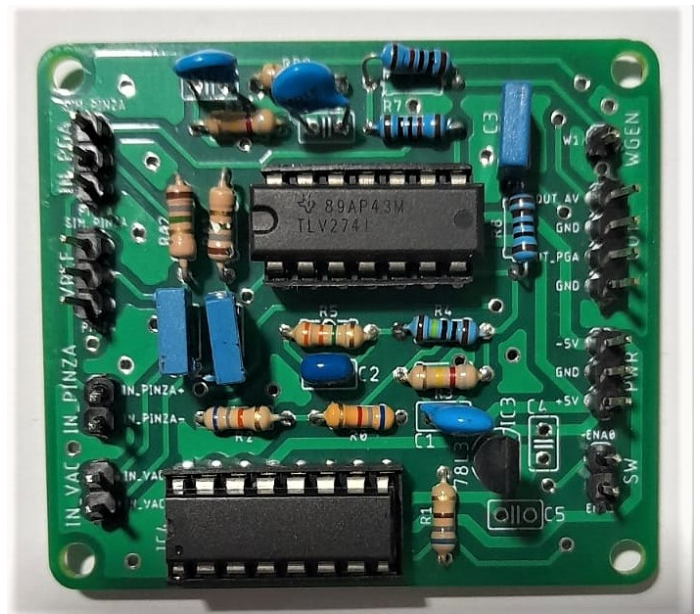


Fig. 9. Circuito de acondicionamiento implementado

Para la implementación de la solución propuesta se utilizó como se mencionó antes el Monitor, los circuitos de acondicionamiento de las señales sobre el PCB diseñado, la tarjeta AD2 y una computadora para controlar el sistema, como se muestra en la figura 10.

Se conectó las salidas del Monitor al circuito PCB para poder acondicionar las señales extraídas  $V_{ac}^+$ ,  $V_{ac}^-$  y  $V_{pinza}^+$ ,  $V_{pinza}^-$  proporcionales al voltaje y corriente consumidos por el electrodoméstico respectivamente.

Cabe resaltar que sobre el PCB se diseñó también un simulador de la pinza amperimétrica, que fue muy útil para verificar por medio del generador de señales del AD2 el



## IV. EXPERIMENTOS Y RESULTADOS

### A. Bases de Datos

Para poner a prueba el software desarrollado, se utilizaron dos bases de datos con señales de corriente-voltaje previamente adquiridos. Ambas bases consisten en archivos en formato ".csv", con la primera columna correspondiendo a la corriente y la segunda al voltaje. Ambos cuentan adicionalmente con un archivo de meta-data, en el cual se indica a que casa y tipo pertenece cada medición. Además, cada electrodoméstico tiene un identificador único dentro de cada base.

Por un lado, se utilizó la ya mencionada base PLAID [1]. Esta tiene la ventaja de tener un número de electrodomésticos alto (1074), y una cantidad aproximadamente uniforme de electrodomésticos de cada tipo. La principal desventaja de esta base, es que se realizó en Estados Unidos, con frecuencia de línea distinta a la de UTE, y con electrodomésticos distintos a los del mercado local.

Por otro lado, se utilizó la base de datos "EDM-III", desarrollada por estudiantes y docentes de la Facultad de Ingeniería UdelaR en los años 2019, 2020, y 2021. Ésta última base fue adquirida por alumnos del curso "Taller Fourier - EDM" mediante un acondicionamiento y procesamiento de señales análogo al descrito en este informe. En particular, se utilizó la del año 2019 para el entrenamiento, dado que esta contiene la mayor cantidad de electrodoméstico (699). Sin embargo, cabe resaltar que al ser menor cantidad que la base PLAID se esperan peores resultados. Además, es esperable que al no tratarse de un trabajo profesional y al no haber coordinación entre todos los actores tomando las medidas, los tipos de electrodomésticos medidos estén des-balanceados, lo cual pueda llevar a sesgos en los clasificadores. Por ejemplo, se tienen 61 aspiradoras y sólo 11 aires acondicionados.

Sin embargo, la principal ventaja de la base "EDM-III" es que se puede entrenar al clasificador con electrodomésticos del mismo mercado que los que se van a intentar clasificar. Por lo tanto, se esperan los mejores resultados con esta base.

### B. Extracción de características

La implementación antes mencionada referente a la extracción de características se realizó mediante programas Python sobre Jupyter Notebooks, donde se utilizaron fuertemente las librerías NumPy para trabajar con arreglos y Matplotlib para la realización de gráficos [11] [12].

Una primera manera de observar las diferencias entre los electrodomésticos es la realización de box-plots para ver la distribución del valor de cierta característica para los distintos electrodomésticos.

Por ejemplo en la figura 12 se puede ver que en general los secadores de pelo se podrían diferenciar bastante bien de los las CFL utilizando el factor de potencia total, mientras que estas últimas son difíciles de diferenciar de los microondas usando esta característica, pero podrían diferenciarse utilizando la potencia reactiva fundamental. Mirando solamente estos tres box-plots se pueden encontrar varios de estos casos donde la

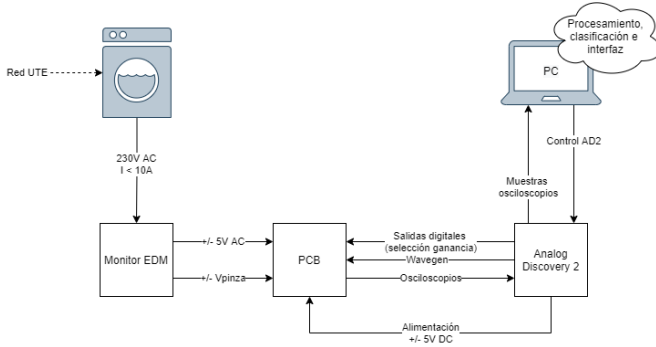


Fig. 10. Circuito de acondicionamiento implementado

correcto funcionamiento del diseño sin tener que conectar un electrodoméstico.

Para controlar la ganancia de  $R_{2eq}$  por medio del chip "Intersil DG412" se conectaron las salidas digitales 0 y 1 del AD2, encargadas de configurar la ganancia necesaria para que se cumpla  $V_o > 1V$ . Se alimentó el circuito PCB por medio del AD2 con  $V_{DD} = +5VDC$ ,  $V_{SS} = -5VDC$  y  $GND = 0VDC$ .

Se extrajeron de este circuito las señales acondicionadas mediante los osciloscopios 1 y 2 del AD2, ingresando estos datos a la PC por medio de una conexión USB entre el AD2 y esta.

Luego, se conecta el Monitor EDM a la red eléctrica, y por medio de un toma-corriente se conectan electrodomésticos. En la figura 11 se puede observar el sistema implementado en funcionamiento, realizando medidas sobre una lámpara incandescente.



Fig. 11. Sistema implementado evaluando las características de una luz fluorescente

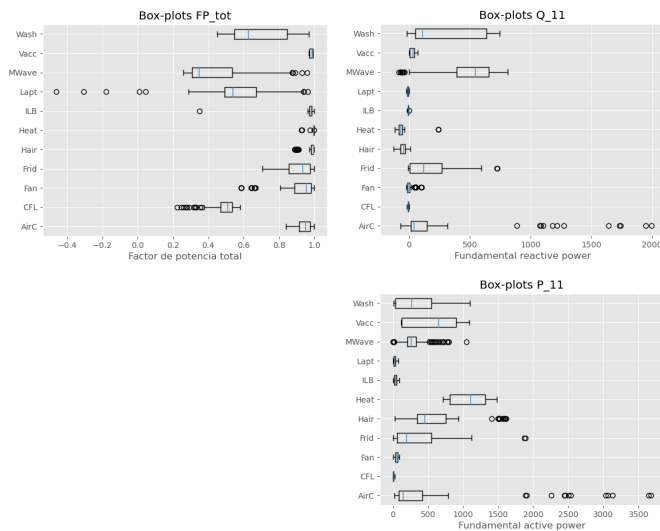


Fig. 12. Box-plots para tres distintas características en la base PLAID.

combinación de características podría servir para diferenciar entre electrodomésticos.

Otra información que ponen en evidencia los box-plots es cuan parecidos entre sí son los electrodomésticos de un mismo tipo. Un electrodoméstico cuya caja está muy distribuida no será fácilmente reconocido mediante esa característica y viceversa. El caso ideal para caracterizar son cajas concentradas y en posiciones distintas a las demás. Otra manera de ver esta distribución de características es poniendo dos características en dos ejes y colocar a todos los electrodomésticos de la base en este plano:

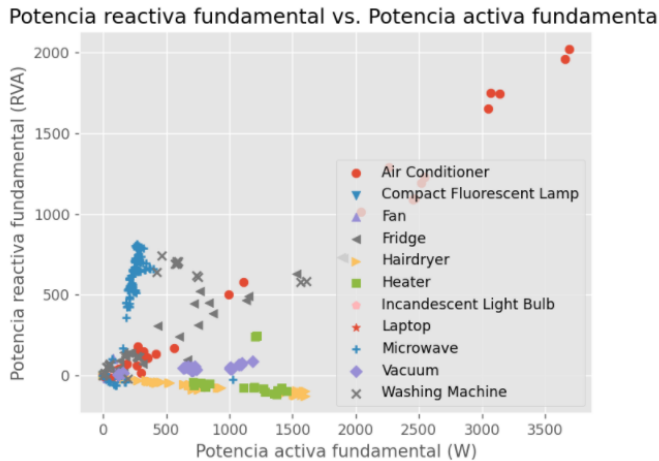


Fig. 13. Electrodomésticos de la base PLAID en el plano de potencia reactiva y activa.

En este plano, también puede apreciarse que electrodomésticos se encuentran más concentrados/distribuidos, y cuales se diferencian más claramente de los otros tipos mediante estas dos características. Observar que esta gráfica es una representación de  $n = 2$  de lo que se hace con las

características para clasificar con el método "KNN" antes descripto.

Como se mencionó anteriormente, también se utilizó una característica no estrictamente numérica con los diagramas  $V - I$ , los cuales también sirven para diferenciar ciertos electrodomésticos. Para extraer esta información se trabajó sin considerar los primeros ciclos de las medidas tomadas para eliminar los transitorios y luego tomando el diagrama promedio de todos los electrodomésticos de un mismo tipo, lo cual sirve para "limpiar" los diagramas y conseguir algo más característico al funcionamiento usual de un electrodoméstico.

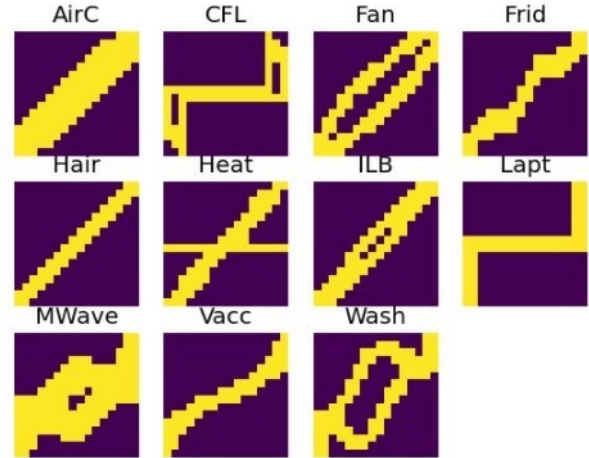


Fig. 14. Diagramas V-I por tipo para la base PLAID.

En la figura 14 se puede ver como todos los diagramas difieren entre si en mayor o menor medida. Por esto fue que se considero conveniente considerar esta información para la clasificación, aunque por ser imágenes, aumenten considerablemente el costo computacional.

Por último, otra característica que también se utilizó y cabe resaltar es la distorsión armónica total. Esto es un dato numérico pero indica directamente información visual cualitativa sobre las señales, ya que mientras mayor sea, mas difiere la señal en cuestión de una onda sinusoidal.

Tipo electrodoméstico	THD_V(%)	THD_I(%)
Air Conditioner	2.325	4.933
Compact Fluorescent Lamp	2.447	144.154
Fan	1.952	11.984
Fridge	1.622	6.777
Hairdryer	2.686	2.803
Heater	2.378	3.499
Incandescent Light Bulb	1.681	2.069
Laptop	1.520	86.696
Microwave	1.082	7.715
Vacuum	2.582	12.498
Washing Machine	2.379	2.237

Fig. 15. Tabla de distorsión armónica por tipo para la base PLAID.

En la figura 15 se puede ver la tabla de distorsiones promedio con las que se trabajó en la base PLAID. Entre ellas se puede ver como los valores para las señales de

voltaje no defieren mucho (lo cual es esperable dado que se trata del voltaje de línea), mientras las corrientes si lo hacen, debido a las distintas maneras en las que consumen los electrodomésticos.

### C. Evaluación de desempeño de la solución propuesta

Cómo se mencionó en la sección II-C, se estudió el uso de dos clasificadores en particular. Esto se hizo sobre la base PLAID primero, apuntando a construir un buen clasificador entrenado con esta base. Luego, se evaluó el desempeño de este clasificador sobre la base EDM-III. Finalmente, se creó un nuevo clasificador entrenado por la base EDM-III.

En las siguientes gráficas vemos el desempeño de los clasificadores considerados a medida que variamos sus parámetros y las características utilizadas.

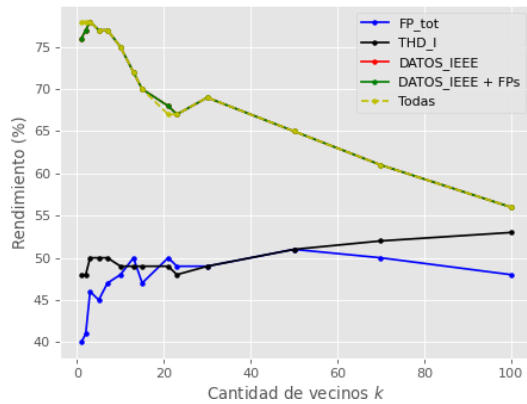


Fig. 16. Desempeño del clasificador K-Nearest Neighbors sobre base PLAID.

Por un lado, en la figura 16 se vé que el rendimiento mejora al aumentar características, lo cual era esperable. Lo que no esperábamos es que el rendimiento se maximizara con números tan bajos de  $k$ . Esto podría deberse al hecho de que, como se puede ver en los box-plot, hay tipos de electrodomésticos que para una cierta característica su distribución en muy concentrada, pero su media varía por muy poco con la de otro tipo. Es decir que al aumentar la cantidad de vecinos rápidamente el clasificador empieza a confundir con otros tipos.

Por otro lado, en la figura 17 se vé claramente cómo mejora la clasificación al aumentar la cantidad de características y la cantidad de árboles. Se determina que se obtiene un buen rendimiento en el rango 50-150 árboles, sin tener que manejar un tiempo de procesamiento demasiado grande.

El rendimiento de RF en este rango es alrededor de 7% mejor que el mejor rendimiento de KNN, por lo cual se decide usar este clasificador a costo de mayor tiempo de procesamiento, dado que este es sólo significativo a la hora de entrenar, y no hace ninguna diferencia de tiempo significativa a la hora de clasificar en tiempo real.

Entrenando un clasificador Random Forest con la base PLAID y las características planteadas en la figura 5, se obtiene el siguiente desempeño sobre esta misma base (usando el esquema de validación cruzada dejando uno afuera):

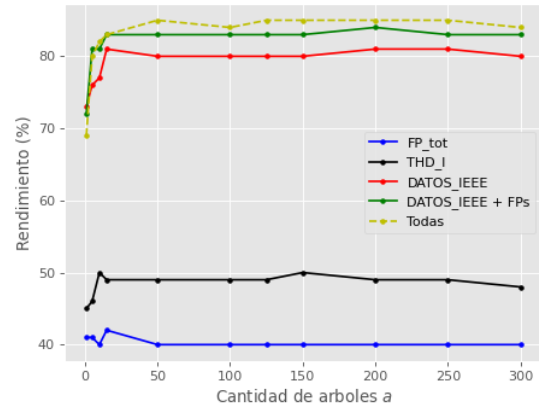


Fig. 17. Desempeño del clasificador Random Forest sobre base PLAID.

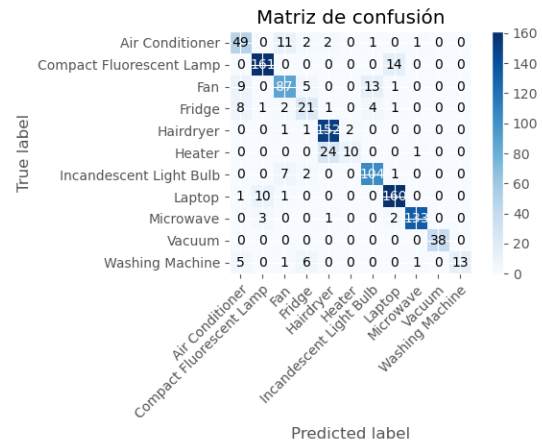


Fig. 18. Matriz de confusión PLAID sobre PLAID.

Para este caso, el acierto medio ronda en el 85%.

Luego, usando este mismo clasificador entrenados por la base PLAID se clasifican a todos elementos de la base EDM-III. En la siguiente figura, se puede ver su matriz de confusión:

El acierto medio en este caso es de 40%. Esta sustancial baja en el rendimiento era esperable como se explica en la sección IV-A: Los electrodomésticos con los que se realiza la base PLAID pertenecen a un mercado muy distinto al local uruguayo, e incluso la tensión de línea es distinta.

Finalmente, se entrena a un nuevo clasificador con la base EDM-III del año 2019 (dado que es la que mayor cantidad de electrodomésticos tiene entre estas), para así tener un clasificador que pueda funcionar relativamente de punta a punta, clasificando en el momento señales adquiridas por la placa. Cabe resaltar que se extraen de esta base los electrodomésticos del tipo "Otro" para poder hacer una comparación con la base PLAID, que no incluía este tipo. Los resultados nuevamente se presentan en una matriz de confusión:

Como era esperable, el acierto medio en este caso sube a un 60%. Posibles razones por las cuales esto sigue siendo menor al caso PLAID sobre PLAID se explican en detalle en la sección IV-A, e incluyen el hecho de que se tiene

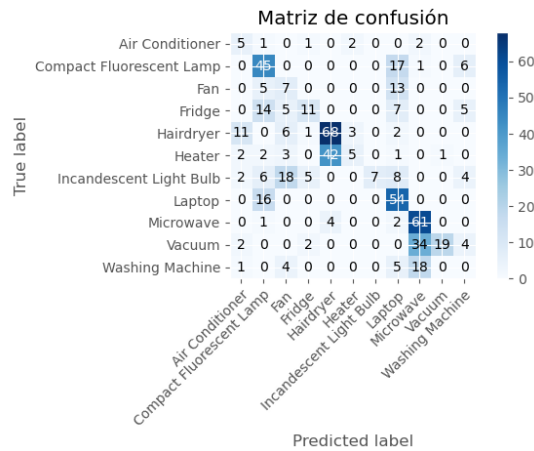


Fig. 19. Matriz de confusión EDM-IEE 2019 sobre PLAID.

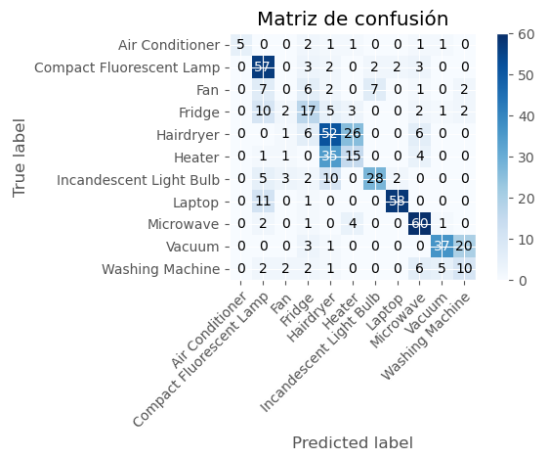


Fig. 20. Matriz de confusión EDM-IEE 2019 sobre EDM-IEE 2019.

aproximadamente 50% menos electrodomésticos en la base EDM-IEE, y sus cantidades por tipo son menos homogéneas.

## V. CONCLUSIONES

Sin dudas resultó satisfactorio haber conseguido un prototipo de sistema no invasivo de adquisición, procesamiento y clasificación "punta a punta", que puede identificar con 60% precisión y en pocos segundos el tipo de un electrodoméstico que se haya conectado. Cabe resaltar que el proceso de desarrollo e integración de las distintas partes del sistema fue muy enriquecedor, presentando oportunidades de trabajar en diversas etapas de desarrollo, desde diseño de PCB a programación de clasificadores.

Desde un punto de vista del hardware, creemos que el prototipo desarrollado cumple de buena forma con lo requerido. Creemos que el principal aspecto a mejorar en este frente en futuras iteraciones sería el tamaño físico del dispositivo, el cual aunque razonablemente portátil, está lejos de ser un dispositivo que se podría dejar instalado en sitio para medir el consumo de toma-corrientes, especialmente teniendo en cuenta que en

este trabajo todo el procesamiento se realizó con una PC y el laboratorio AD2.

Es claro también que los resultados de acierto medio se podrían mejorar a futuro. Una manera sería mejorando la base de entrenamiento. Esto podría lograrse compilando las bases EDM-IEE de distintos años para tener una mayor base de referencia de los electrodomésticos del mercado local. También se podría estudiar con mayor detalle que tipos de electrodomésticos tienen menos muestras en la base, y reforzar estos intencionalmente con más mediciones. Se debe tener en cuenta que el trabajo de recopilación de señales tampoco es realizado de manera profesional, por lo cual es esperable que hayan algunas mediciones erróneas en la base, las cuales son muy difíciles de detectar, y podrían estar empeorando el rendimiento frente a la base PLAID.

Otra manera de mejorar los resultados de clasificación sería identificar que tipos de electrodomésticos se confunden más seguido en con el clasificador actual. Luego, se podría estudiar si sacando algunas características de la base en las que estos tipos se parezcan el rendimiento del clasificador aumente.

## REFERENCES

- [1] Gao, Jingkun, et al. "Plaid: a public dataset of high-resolution electrical appliance measurements for load identification research: demo abstract." proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings. ACM, 2014.
- [2] IEEE Standard Definitions for the Measurement of Electric Power Quantities Under Sinusoidal, Nonsinusoidal, Balanced, or Unbalanced Conditions," in IEEE Std 1459-2010 (Revision of IEEE Std 1459-2000), vol. no., pp.1-50, 19 March 2010 doi: 10.1109/IEEESTD.2010.5439063.
- [3] J. L. Willems, "The IEEE standard 1459: What and why?," 2010 IEEE International Workshop on Applied Measurements for Power Systems, Aachen, 2010, pp. 41-46. doi: 10.1109/AMPS.2010.5609324.
- [4] Pedregosa, F., Varoquaux, Gael, Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... others. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12(Oct), 2825-2830.
- [5] Fix, E.; Hodges, J.L. (1989). "(1951): An Important Contribution to Nonparametric Discriminant Analysis and Density Estimation: Commentary on Fix and Hodges (1951)". International Statistical Review / Revue Internationale de Statistique 57 (3): 233-238. doi:10.2307/1403796.
- [6] Antti Ajanki AnAj (2007). "Example of k-nearest neighbour classification".
- [7] Breiman, Leo (2001). "Random Forests". Machine Learning 45 (1): 5-32. doi:10.1023/A:1010933404324.
- [8] Venkata Jagannath (2017). "Random Forest Template for TIBCO Spotfire®".
- [9] Charles Elkan, Evaluating Classifiers University of California, San Diego, 18 de enero de 2011.
- [10] Joan Domenech (2011) "Validación cruzada dejando uno fuera".
- [11] Harris, C.R., Millman, K.J., van der Walt, S.J. et al. Array programming with NumPy. Nature 585, 357-362 (2020). DOI: 10.1038/s41586-020-2649-2.
- [12] J. D. Hunter, "Matplotlib: A 2D Graphics Environment", Computing in Science & Engineering, vol. 9, no. 3, pp. 90-95, 2007.