

Interpretación de señales para la detección de obstrucciones en inspecciones instrumentadas de tuberías

Juan Camilo Bohórquez Reyes
Facultad de Fisicomecánicas
UNAB
Bucaramanga, Colombia
jbohorquez3@unab.edu.co

Álvaro Javier Flores Martínez
CIC
Bucaramanga
aflorez@corrosion.uis.edu.co

PhD. Carlos Julio Arizmendi
Facultad de Fisicomecánicas
UNAB
Bucaramanga, Colombia
carizmendi@unab.edu.co

Abstract— this study explains the methodology for conducting the classification and treatment of data obtained by a pipeline inspection device used to detect obstructions. The proposal involves the implementation of signal processing techniques and artificial intelligence applied to signals acquired by caliper type sensors in a inspection tool in a pipe used to transport hydrocarbons and detect events in the signals corresponding to the change of diameter on the pipe.

I. INTRODUCCIÓN

En la industria de los hidrocarburos se implementa un método de inspección de tuberías, para detectar imperfecciones tales como fisuras, abolladuras u obstrucciones entre otras. Este método es conocido como *Pigging*, que basa en una herramienta llamada *Piping Inspection Gauge* (PIG) [1, 2, 3, 4, 5, 6].

En conjunto con la Corporación para la Investigación de la Corrosión (CIC) en su proyecto ITION, se propuso un dispositivo PIG que viaje por las tuberías junto con el material a transportar, realizando una inspección del mismo mediante sistemas de medición tipo Caliper [1, 2, 3, 4, 5, 6] para adquirir los datos del diámetro de la tubería, los datos adquiridos serán tratados mediante técnicas provenientes del procesamiento de señales basado en la Transformada Wavelet [7]. Los datos procesados serán adaptados a las entradas de un clasificador basado en Redes Neuronales [8].

Mediante la implementación de métodos inteligentes es posible detectar las imperfecciones de una forma más eficiente, agilizando los procedimientos de mantenimiento y prevención en las tuberías de transporte, sin detener el funcionamiento de la tubería, permitiendo no afectar la producción mientras se realiza la inspección.

II. CONTENIDO

A. Adquisición de Datos

Cada señal se adquiere en un recorrido de 23 kilómetros en una tubería funcional, utilizando el dispositivo ITION. Las señales se registran utilizando dos brazos y una rueda en cada

uno de ellos, adyacente a la tubería y un sensor que detecta las imperfecciones al flexionar cada brazo. Ver Fig. 1.

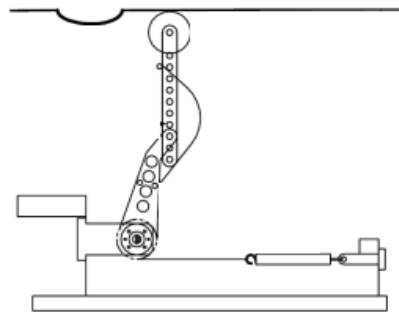


Fig. 1 Brazo Caliper. Fuente: CIC.

B. Señales

Se registraron dos señales cada una obtenida por un sensor tipo Caliper. En Fig. 2 se observa el comportamiento de la señal.

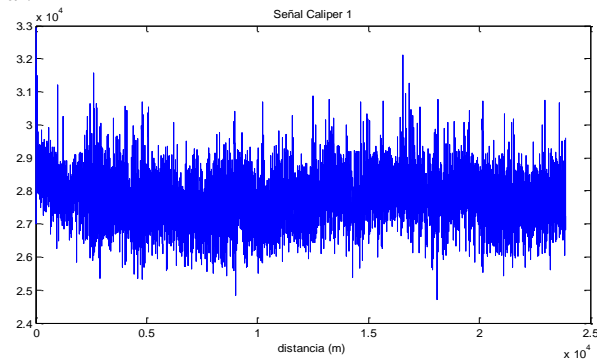


Fig. 2. Señal del sensor caliper 1. Fuente: CIC.

En Fig. 2 se observa la señal completa registrada por el primer sensor Caliper en su recorrido de 23 metros, la señal caliper dos posee el mismo comportamiento que la señal mostrada en la figura.

La toma de datos se realizó en condiciones reales donde los eventos que se quieren encontrar suceden a una alta frecuencia debido al tamaño que poseen.

Se observa una oscilación que sucede de manera recurrente en las señales. En la Fig. 3 Se observan estas oscilaciones que se repiten a una distancia entre 9 a 13 metros a lo largo de toda la señal. Se puede determinar si corresponden a un evento que sucede recurrentemente en la tubería como para luego analizar si existe algún evento aislado como válvulas o defectos como abolladuras que tengan una distancia diferente entre oscilaciones.

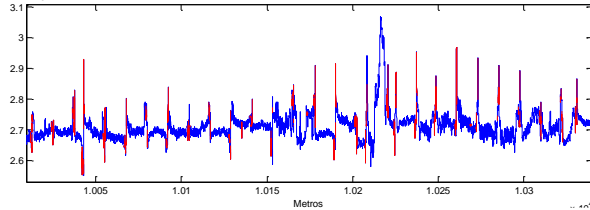


Fig. 3 Oscilaciones periódicas en una de las señales Caliper. Fuente: Autor.

Se busca la clasificación de estos eventos con el fin de hallar

C. Procesamiento de la Señal

Se realiza una descomposición multinivel Wavelet a 9 niveles de profundidad, encontrado como óptimo utilizando el criterio de la entropía de Shannon y una Wavelet madre Haar debido a que es la mejor para encontrar cambios en la señal. Este procedimiento se realiza utilizando la *Wavelet Packet* del software Matlab®.

La *toolbox Wavelet Packet* de Matlab® halla el árbol de descomposición de una señal donde encuentra todas las posibles derivaciones de esta. En la Fig. 4 se observa una señal de ejemplo descompuesta utilizando esta herramienta.

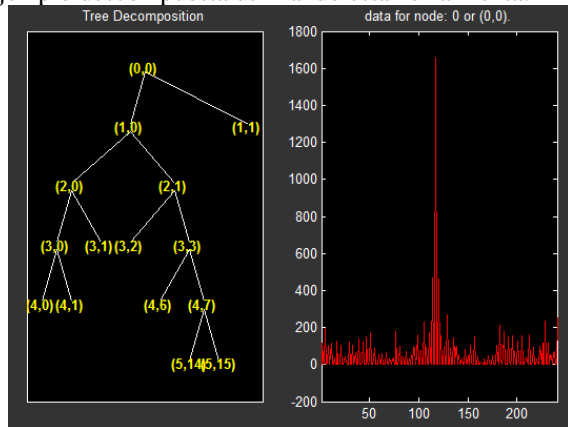


Fig. 4 Árbol de descomposición Wavelet Packet. Fuente: Autor.

La señal que se observa a la derecha en la Fig. 4 se descompone a una profundidad de 6 niveles obteniendo el árbol de descomposición mostrado a la izquierda de la figura. Cada derivación de la descomposición *Wavelet* se le llama nodo. Para las señales caliper se utiliza el nodo 13 obtenido del árbol de descomposición para reconstruir las señales ya que se puede observar los eventos que ocurren a altas frecuencias lo largo de estas.

En la Fig. 5 se observa una de las señales originales y una de las señales reconstruidas donde se encuentran los eventos recurrentes mencionados anteriormente.

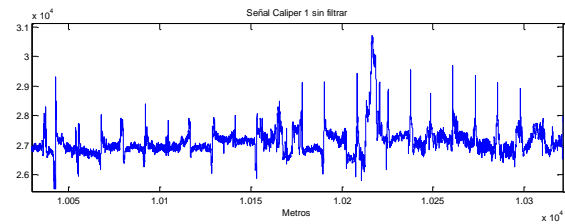


Fig. 5 Señal Caliper original y señal reconstruida. Fuente Autor.

D. Clasificación por umbralización

Se procede a realizar una metodología de clasificación que encuentre estos eventos en las dos señales. En primer lugar se realiza un procedimiento que consiste en encontrar un umbral con el cual se pueda clasificar la mayor cantidad de oscilaciones posibles.

Para encontrar este umbral se extraen muestras de oscilaciones de la señal reconstruida como la que se observa en la Fig. 6.

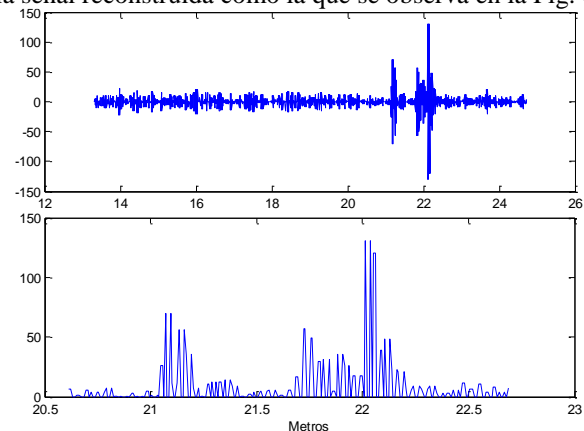


Fig. 6 Muestra de oscilación periódica. Fuente: Autor.

En la Fig. 6 se observa la oscilación separada de la señal y sin datos negativos, se eliminan estos datos ya que el siguiente paso consiste en hallar la media y la desviación estándar de la oscilación, debido a que esta es simétrica en cero el valor de su media será aproximadamente cero. Luego de obtener la media y su desviación de los datos positivos de la oscilación. Luego se halla el umbral al realizar la suma entre la media y su desviación, esto garantiza que el umbral detecte en su mayor parte las oscilaciones en la señal y discrimine el ruido de fondo. En la Tabla 1 se observan 6 oscilaciones tomadas en diferentes partes de la señal con el fin de calcular diferentes umbrales y hallar el óptimo

TABLA I

Estadístico/ Oscilación	1	2	3	4	5	6
Media	9.001	9.754	6.067	9.231	8.967	7.560
Desviación	20.177	24.304	12.799	19.775	20.187	17.786
Media + Desviación	29.078	34.058	18.867	29.006	29.15	25.346

Luego de este procedimiento se clasifica la señal utilizando los umbrales obtenidos de la suma de la media y desviación estándar de cada oscilación, obteniendo un vector de 1 y 0 que identifican las muestras que pertenecen a una oscilación y las muestras que no respectivamente.

Debido a que el ruido en la señal no es constante, existen secciones donde este supera el umbral. Se busca otro algoritmo de discriminación para detectar las oscilaciones. Se conoce que cada oscilación está separada entre si 9 a 13 metros, tomando esto en consideración se recorre el vector de 1 y 0 obtenido anteriormente en la umbralización y se obtienen las distancias de los flancos de subida, es decir donde se empieza a clasificar una oscilación, luego se hallan las distancias que existen entre estos flancos tomando solo aquellas que estén entre el rango de 9 a 13 metros. Se realiza este proceso utilizando cada uno de los vectores obtenidos con los diferentes umbrales. En la tabla 2 se observa la cantidad de oscilaciones detectadas utilizando cada uno de los umbrales hallados anteriormente.

TABLA II

	1	2	3	4	5	6
Umbral	29.078	34.058	18.863	29.009	29.15	25.346
Oscilaciones Encontradas	1521	1588	835	1521	1521	1411

Se utilizaron los mismos umbrales y el mismo procedimiento para clasificar las oscilaciones en la segunda señal Caliper. En la Tabla 3 se observan la cantidad de oscilaciones encontradas en esta señal.

TABLA III

	1	2	3	4	5	6
Umbral	29.078	34.058	18.867	29.006	29.155	25.346
Oscilaciones Encontradas	1432	1560	377	1432	1432	1254

En las Tablas 2 y 3 muestra el umbral que obtuvo una mayor cantidad de detecciones, en este caso es el umbral utilizando la oscilación 2 con 1588 oscilaciones encontradas en la señal Caliper 1 y 1560 en la señal Caliper 2.

En la Fig. 7 se puede observar una sección de las señales reconstruidas clasificada utilizando la umbralización mencionada anteriormente. La clasificación es similar para ambas señales reconstruidas.

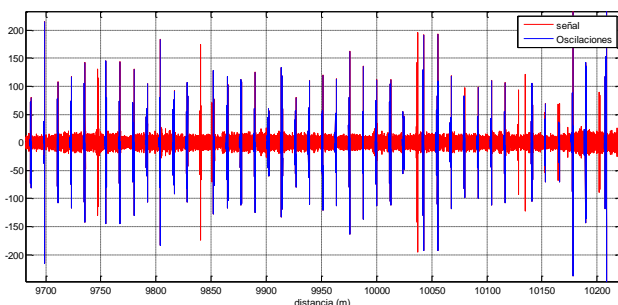


Fig. 7 Señal reconstruida clasificada por umbralización. Fuente: Autor.

En la Fig. 7 se observa que no todas las oscilaciones fueron clasificadas correctamente utilizando el proceso de umbralización, estas se observan en rojo, las que se encuentran en azul fueron clasificadas correctamente. Se propuso un método de clasificación basado en redes neuronales utilizando las señales reconstruidas y las etiquetas de clasificación encontradas en el proceso de umbralización.

E. Clasificación por Red Neuronal

En el procedimiento anterior se identificaron una gran cantidad de oscilaciones pero no su totalidad. A continuación se propone un método de clasificación utilizando redes neuronales con el fin de encontrar una mayor cantidad de oscilaciones.

La red neuronal propuesta posee dos entradas correspondientes muestras o patrones de las dos señales reconstruidas utilizando el nodo 13 del árbol de descomposición obtenido por la *toolbox* de Matlab® *Wavelet Packet*.

Utilizando las señales reconstruidas y el vector de etiquetas que se encuentra utilizando el procedimiento de clasificación por umbralización, se escogen tres secciones de la señal donde se hayan clasificado correctamente las oscilaciones con el fin de crear un grupo de datos etiquetados para realizar el procedimiento de entrenamiento, prueba y validación de la red neuronal. De esta manera la red neuronal aprenderá el comportamiento de las oscilaciones y así obteniendo un mejor resultado en su clasificación. En la Fig. 8 se observa uno de los tres grupos de datos que se escogieron de una de las señales clasificadas por umbralización con sus respectivas etiquetas.

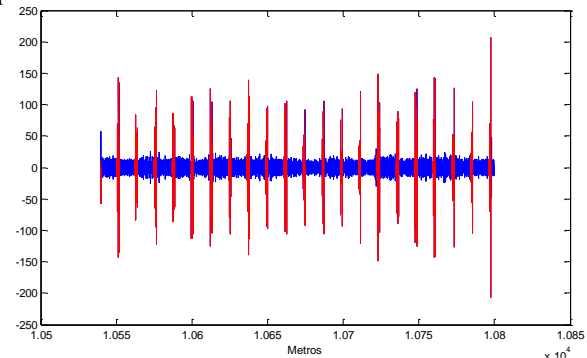


Fig. 8. Sección escogida de la señal reconstruida y sus etiquetas correctamente clasificadas. Fuente: Autor.

Las tres secciones de la señal reconstruida con sus respectivos vectores de etiquetas clasificadas correctamente se concatenan correctamente y crean una base de datos con la cual se entrenara, validara y probara una red neuronal.

Se propone una red con 2 capas ocultas, cada una con 5 y 10 neuronas respectivamente con función tangente hiperbólica, la capa de salida se implementó una función lineal.

Para realizar el entrenamiento de la red se propuso 1000 iteraciones, una meta de MSE de $1e-5$, mínimo gradiente de

1e-7, tasa de aprendizaje de 0.05, decremento de la tasa de aprendizaje de 0.1, incremento de la tasa de aprendizaje de 10, máxima tasa de aprendizaje 1e-10 y criterio de parada por validación de 6.

Se utiliza el algoritmo de entrenamiento *Levenberg-Marquart* [11], el cual evita el *overfitting* o memorización de los datos para garantizar una generalización. De igual manera también implementa un método de *Early Stopping* que genera una parada por validación, deteniendo el entrenamiento si el rendimiento no mejora o se mantiene constante en un máximo número de iteraciones seguidas.

Para garantizar la generalización del sistema, se implementa el método *5-Fold Cross Validation* [12, 13, 14] con una aleatorización de los datos de las cuatro señales divididas en 5 grupos diferentes (Ver Fig.9).

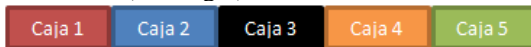


Fig. 9 Grupos de datos para implementar *5-Fold Cross Validation*. Fuente: Autor.

En el método *5-Fold Cross Validation* las primeras tres cajas se utilizan para el entrenamiento, la caja número cuatro, para validación y por último la caja número cinco, para probar la red y obtener el porcentaje de clasificación (ver Fig.7).

Cada vez que se realiza un entrenamiento se intercambian secuencialmente las cajas, hasta que todos los datos han sido utilizadas como prueba del experimento (ver Fig.10). Seguidamente se aleatorizan los datos para repetir el procedimiento o corrida. En este estudio se realizaron 15 corridas, obteniendo 75 porcentajes de clasificación en total.



Fig. 10 Distribución de datos para Cross Validation. Fuente: Autor

Debido al desbalanceo de las clases se utilizó la precisión equilibrada (Ec. 1) [9], donde se tiene en cuenta los resultados de la clasificación de cada clase equitativamente.

$$Precision\ equilibrada = \frac{1}{2} * \frac{Cp}{Cp+Fp} + \frac{1}{2} * \frac{Cn}{Cn+Fn} \quad Ec. 1$$

Para el cálculo de la precisión equilibrada se debe tener en cuenta que se están clasificando dos grupos de datos, oscilación y no oscilación. La variable Cp o Correctos Positivos se refiere a la cantidad de muestras clasificadas correctamente en la clase oscilación. La variable Cn o Correctos Negativos se refiere a la cantidad de muestras clasificadas correctamente como no oscilación. La variable Fp

o Falsos Positivos se refiere a las muestras clasificadas como oscilación perteneciendo a la clase de no oscilación. Por último la variable Fn o Falsos Negativos se refiere a la cantidad de muestras clasificadas como no oscilación perteneciendo a la clase oscilación.

Al realizar el cálculo de cada uno de los 75 porcentajes de clasificación, se obtuvo un resultado promedio de $88.3680 \pm 0.1098\%$ para el grupo de prueba. En Fig. 11 se observa una de las señales caliper clasificada con su respectiva señal reconstruida.

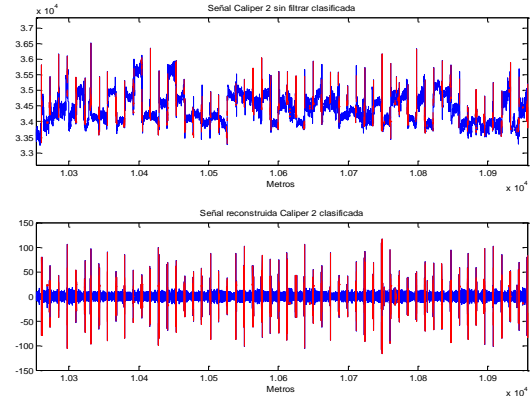


Fig. 11 Señal Caliper clasificada con su respectiva señal reconstruida. Fuente: Autor.

F. Proceso Post-clasificación

Luego del proceso de clasificación utilizando la red neuronal se generaron una gran cantidad de errores de clasificación como falsos positivos y falsos negativos. Se propuso un método de post-clasificación con el fin de eliminar la mayor cantidad de ellos. En la Fig. 12 se observan los dos tipos de errores en la clasificación de la señal.

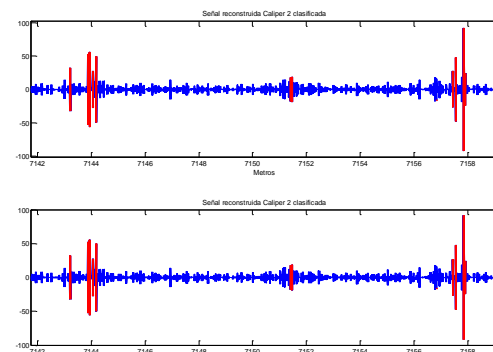


Fig. 12 Falsos positivos y falsos negativos. Fuente Autor.

Los falsos positivos son las muestras clasificadas como oscilación donde no existe como las muestras en rojo mostradas en el centro de las señales de la Fig. 12. Para eliminar la mayor cantidad de estos errores se propuso un método que consiste en guardar las muestras clasificadas como oscilación que se encuentren seguidas además de contar la cantidad de muestras que estén en estos grupos. En la Fig. 13 se explica este procedimiento.

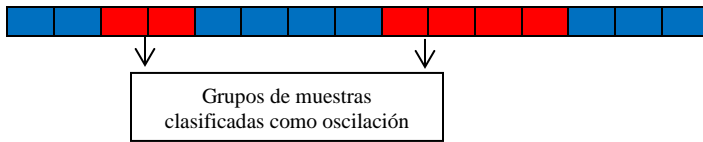


Fig. 13 Grupos de muestras clasificadas como oscilación. Fuente: Autor.

En la Fig. 13 se observan dos grupos de datos clasificados como oscilación. Estos grupos se almacenan de forma separada guardando la posición de cada muestra que contenga y la cantidad de muestras que posea. Se conoce que el tamaño en muestras de un falso positivo es menos de 10 muestras, por lo tanto se buscan los grupos de datos que posean menos de 10 muestras y se reclasifican como clase de no oscilación

Para resolver el problema de falsos positivos el procedimiento es el mismo con la diferencia que los grupos de datos serán muestras clasificadas como no oscilación. En la Fig. 14 se explica el procedimiento.

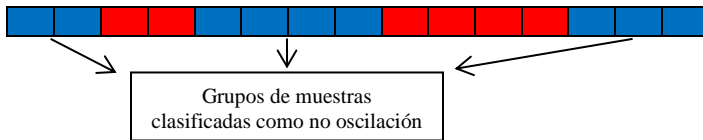


Fig. 14 Grupos de muestras clasificadas como no oscilación. Fuente: Autor.

Se realiza el mismo procedimiento utilizado para los falsos positivos. Se conoce que estos errores poseen un tamaño de muestras menor a 120, por esta razón se reclasifican como clase oscilación los grupos de datos que tengan una cantidad menor a este número de muestras.

En La Fig. 15 se observa una de las señales reconstruidas antes y después del proceso de post-clasificación.

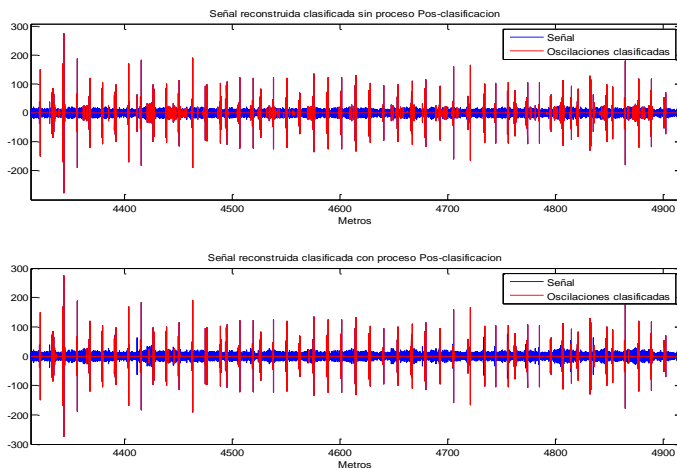


Fig. 15 Señal antes y después del proceso de post-clasificación.

Luego de obtener la señal clasificada se procede a hallar las distancias entre las oscilaciones clasificadas. Se conoce que los eventos recurrentes suceden a una distancia entre 9 a 13 metros, cualquier evento que este a más de 13 metros se considera como un segmento donde no se clasificaron oscilación o un error de clasificación. Los eventos que se encuentran entre 1 a 9 metros de distancia se consideran

eventos desconocidos que pueden ser válvulas, cambios de aceleración de la herramienta, imperfecciones en la tubería, etc. Por último los eventos que se encuentran entre 0 a 1 metro se consideran errores de clasificación que no se eliminaron luego del proceso de post-clasificación.

En la Fig. 16. Se observa un histograma donde el eje x corresponde a las distancias que se separan las oscilaciones y el eje y corresponde a la cantidad de distancias. En esta figura se puede observar la cantidad de distancias para cada uno de los tipos de eventos mencionados anteriormente.

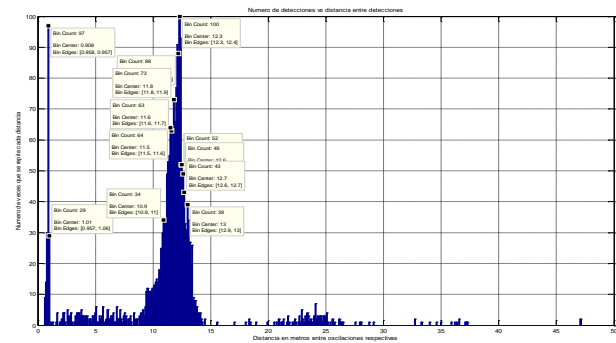


Fig. 16 Histograma de distancias entre oscilaciones. Fuente: Autor.

G. Interfaz Gráfica

Se realizó una interfaz gráfica con el fin de realizar todo el procedimiento mencionado anteriormente utilizando las señales originales para reconstruirlas utilizando el nodo 13 obtenido del árbol de descomposición por medio de la *Wavelet Packet* de Matlab® quienes servirán como entrada a el modelo de la red neuronal validada para luego realizar el proceso de post-clasificación mencionado anteriormente. Luego de esto se puede graficar la señal clasificada y observar cada uno de los eventos como oscilaciones bien identificadas, eventos desconocidos, errores de clasificación que persistieron luego del proceso de post-clasificación y oscilaciones que no se detectaron. En la Fig. 17 se puede observar la interfaz gráfica realizada.

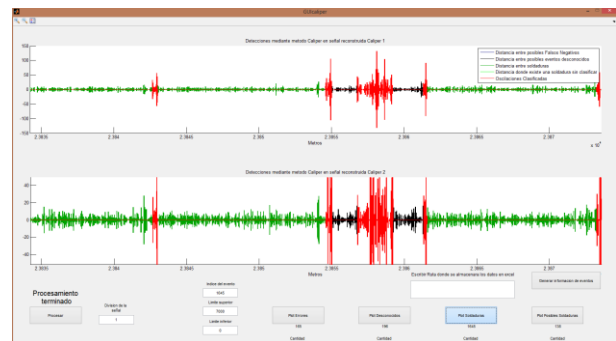


Fig. 17 Interfaz Gráfica. Fuente: Autor.

III. CONCLUSIONES

El procedimiento realizado para encontrar soldaduras en las señales caliper por medio de un umbral obtenido por los estadísticos de las oscilaciones escogidas no encuentra la totalidad de ellas debido a que estas no son totalmente

representativas en la señal por esta razón el número de soldaduras encontradas baja.

Se debe tener en cuenta que todas las oscilaciones periódicas no poseen una distancia de 9 a 13 metros entre sí, en el caso de algunas se encuentran menos o más de estas distancias y utilizando este método se dejarían de clasificar una cantidad considerable de soldaduras.

RECONOCIMIENTOS

“Agradecimientos especiales a la Corporación para la Investigación de la Corrosión por su apoyo al brindar las señales trabajadas en este documento”

IV. BIBLIOGRAFÍA

- [1] J. O. Serrano. “Sistema Caliper Para medir Deformaciones de la Geometría Interna en las Tuberías”. CIC. 2014
- [2] D. Russel, “Pigging in Pipeline Pre-commissioning”, Weatherford & Specialty Services UK. 2005.
- [3] T. Jin, P. Que y Z. Tao. “Designing and Signal Processing of Intelligent Inspection Pig Applying Magnetic Flux Leakage Methods”. Dept. of Information Measurement technology and Instrumental. 2004.
- [4] D. K. Kim, S. H. Cho, S. S. Park, H. R. Yoo, Y. W. Rho. “Desing And Implementation of 30” Geometry PIG”. R&D división, Korea Gas Corporation. 2003.
- [5] I. L. Yablonskikh, S. J. Dawson, R. Venkatanarayanan. “Assessment And Analysis Of Pipline Buckles”. PPL Pipeline Solutions, Cramlington, UK. 2007.
- [6] H.R.Yoo, S.H.Cho, Y.W.Ryo, S.H.Park y G. S Park, “Desing of a Permanent Magnet to Maximize the Sensing Signals in Geometry PIG Caliper System”, International Conference on Electrical Engineering, 2002.
- [7] D.L. Donoho. “De-noising by soft-thresholding”. IEEE Transactions on Information Theory, 41(3), 1995.
- [8] C. Bishop. “Neural Networks for Pattern Recognition”. New York. Oxford University Press Inc. 1995.
- [9] C. J. Arizmendi, “Signal Processing Techniques for Brain Tumour Diagnosis from Magnetic Resonance Spectroscopy Data”. Departament de Llenguatges i Sistemes Informatics, Universitat Politècnica de Catalunya. 2011.
- [10] F. Wilcoxon. “Individual Comparisons by Ranking Methods”. Biometrics. 1945.
- [11] J. J. More. The “Levenberg-Marquardt algorithm: Implementation and theory”. Mathematics and Computer Science Division. Argonne National Laboratory, 2006.
- [12] J. Schneider, “The holdout method”, The school of computer science. 1997.
- [13] P. Refaeilzadeh, L. Tang, H. Lui. “K-fold Cross-Validation”, Arizona State University. 2008.
- [14] F. H. Joanneum. “Cross-Validation Explained”, Institute for Genomics and Bioinformatics. 2005-2006.
- [15] A. Breeding. “The Music Internet Untangled: Using Online Services to Expand Your Musical Horizons”. Giant Path. p. 128. 2004
- [16] P. A. Devijver, J. Kittler. “Pattern Recognition: A Statistical Approach, Prentice-Hall”, Londres. 1982.