

DETECCIÓN AUTOMÁTICA DE SOLDADURAS EN LÍNEAS DE TRANSPORTE DE HIDROCARBUROS USANDO TÉCNICAS DE INTELIGENCIA ARTIFICIAL EN SEÑALES ADQUIRIDAS POR MEDIO DE UNA HERRAMIENTA DE INSPECCIÓN INTELIGENTE.

AUTOMATIC WELDING DETECTION IN HYDROCARBON TRANSPORTATION LINES USING ARTIFICIAL INTELLIGENCE TECHNIQUES ON SIGNALS ACQUIRED THROUGH A SMART INSPECTION TOOL.

William Leonardo García Rueda

Universidad autónoma de Bucaramanga, wgarcia360@unab.edu.co
Corporación para la investigación de la corrosión, wgarcia@corrosion.uis.edu.co

Mario Alberto Quintero Carvajal

Corporación para la investigación de la corrosión, mquintero@corrosion.uis.edu.co

Resumen: Este trabajo presenta un modelo que implementa técnicas de aprendizaje de máquina para reconocer juntas de soldaduras en tuberías que transportan hidrocarburos a partir de señales obtenidas por medio de una herramienta de inspección inteligente. El modelo utiliza en principio una fase de reducción de ruido de la señal por medio de algoritmos de pre-procesamiento y técnicas de selección de atributos. Las técnicas de reducción de ruido fueron seleccionadas después de realizar una revisión bibliográfica y pruebas con los datos de estudio. Posteriormente, el modelo fue entrenado usando algoritmos de reconocimiento y clasificación específicamente redes neuronales artificiales y máquinas de soporte vectorial. Luego, el modelo entrenado fue validado con conjuntos diferentes de datos y se determinó su rendimiento por medio de validación cruzada y curvas ROC. Los resultados obtenidos demuestran que es posible identificar de manera automática las soldaduras con una eficiencia entre el 90 al 98 por ciento.

Palabras clave: Análisis de datos, inspección inteligente, reconocimiento de patrones, línea de transporte, soldadura.

Abstract: This work presents a model that use machine learning techniques for welding recognition in hydrocarbon transportation lines based on signals from smart inspection tools. The model uses a signal noise reduction phase by pre-processing algorithms and attribute-selection techniques. The noise reduction techniques were selected after a bibliographic revision and testing with studio data. Then, the model was trained using recognition and classification algorithms, specifically artificial neural networks and support vector machines. Finally, the trained model was validated with different data sets and the performance was measured with cross validation and ROC analysis. The obtained results prove that is possible the automatic welding identification with an efficiency around 90 and 98 percent.

Keywords: Data analysis, smart inspection, pattern recognition, transportation line, pipe weld joints.

1. INTRODUCCIÓN

Las líneas de transporte de hidrocarburos son consideradas la forma más eficiente para llevar fluidos (petróleo y gas) por largas distancias [1]. Con el paso del tiempo la integridad de este tipo de infraestructura es motivo de preocupación para las empresas, principalmente porque gran parte de las tuberías va llegando al final de su vida útil debido al efecto de agentes externos (corrosión, operaciones incorrectas por obras civiles, deslizamientos de tierras, vandalismo entre otros), lo cual afecta su funcionamiento. En consecuencia, es necesario supervisar, evaluar y realizar el mantenimiento de las

tuberías para garantizar confiabilidad desde un análisis integral y disminuir la probabilidad de falla. Las consecuencias de una falla por lo general, terminan siendo accidentes que impactan el medio ambiente y la población.

Debido a los antecedentes mencionados, la industria ha desarrollado una serie de técnicas no invasivas y ensayos no destructivos, los cuales permiten inspeccionar tanto el estado interno como externo de la tubería sin afectar su funcionamiento.

Dentro de los ensayos más utilizados se encuentra la inspección de tubería con herramientas inteligentes. Una herramienta de inspección inteligente de tuberías¹ (Ver Fig. 1) como se define en [2] es un vehículo instrumentado que insertado dentro de un ducto, viaja a través de su extensión, impulsado por el propio fluido y puede realizar varias funciones, tales como:

- Medir el estado físico y las condiciones de la tubería.
- Detectar y localizar defectos y/o anomalías.
- Reconstruir la trayectoria del ducto en planta y perfil.

Fig. 1. HERRAMIENTA DE INSPECCIÓN INTELIGENTE ACOPLABLE A UN RASPATUBOS.



Fuente: Proyecto Ition-E CIC.

Este tipo de herramienta cuenta con una serie de sensores acoplados en la parte interna y externa de su cuerpo. Las señales provenientes de los sensores son adquiridas, procesadas y almacenadas durante la inspección, posteriormente se hace uso de técnicas de análisis especializado tales como la inteligencia artificial para obtener la identificación de discontinuidades o irregularidades llamadas indicaciones, los cuales, son posibles sitios dentro de la tubería con presencia de defectos. Adicionalmente, estas herramientas facilitan la detección de elementos que componen estas líneas de transporte tales como juntas soldadas, bridas de acople, válvulas, derivaciones entre otros. En esta investigación, se emplearon datos de inspección entregados por una herramienta inteligente para el análisis de tendencias de integridad y operación denominada ITION, desarrollada por la Corporación para la investigación de la Corrosión. Se aplicaron dos técnicas de inteligencia artificial para reconocer patrones en los datos y localizar de forma automática las soldaduras presentes en una línea de transporte de gas.

Este artículo está organizado de la siguiente manera. En la sección 2, se describen los trabajos previos desarrollados en el área. En la sección 3, se explica el método empleado, describiendo cada fase de la metodología y haciendo énfasis en el modelo desarrollado. En la sección 4, se muestran los experimentos realizados junto con los resultados obtenidos, en la sección 5 se discuten los resultados obtenidos y en la sección 6 las conclusiones.

¹También conocida en inglés como ILI, In line inspection, entre otros.

2. TRABAJOS PREVIOS

En el campo de la inspección inteligente de tuberías, existen herramientas de inspección inteligente con diferentes tecnologías para el estudio de fenómenos en las tuberías, las tecnologías más utilizadas son: MFL² [1], caliper [3], rayos X [4], ultrasonido [5] e inercial [6]. Desde el punto de vista del análisis de datos, las técnicas computacionales más utilizadas para el procesamiento, análisis de datos y reconocimiento de soldaduras son: filtrado de señales e inteligencia artificial.

Las técnicas de filtrado han sido utilizadas en diferentes trabajos con el fin de mejorar el rendimiento de los datos de entrenamiento, por ejemplo en [1], donde se busca reconocer tipos de soldaduras en tuberías de acero al carbono, se realizaron pruebas con los filtros Fourier [7], Wavelet [8] y Savitzky-Golay [9] para pre-procesar los datos de entrada buscando eliminar características irrelevantes que puedan entorpecer la clasificación de los defectos. Los resultados de validación demostraron una mejora en la eficiencia del algoritmo para casi todos los tipos de defectos pasando del 66.7 % al 92.5 %. Además, se logró comprobar una mejora en el tiempo de entrenamiento usando las técnicas de filtrado mencionadas. En [10], se utiliza la transformada de Wavelet para pre-procesar los datos de 71 tipos de defectos geométricos en tuberías, a partir de datos de herramientas de inspección inteligente. Los resultados del trabajo demostraron que es posible con la ayuda de la transformada de Wavelet y otras técnicas computacionales predecir la profundidad y la forma de estos defectos en 3D. En [11]-[12], se aplica la transformada de Wavelet y una versión modificada de esta transformada, para eliminar el ruido en los datos entregados por una herramienta de inspección inteligente con el fin de evitar que los algoritmos de reconocimiento encuentren defectos equivocadamente. Finalmente en [13], se utiliza la transformada de Wavelet para localizar fugas en tuberías de gas natural.

Diferentes técnicas de inteligencia artificial se han empleado para el reconocimiento de fenómenos en tuberías, en [1], [10], [14] se usaron redes neuronales para clasificar e identificar algunas clases de defectos geométricos en soldaduras, en [15] con ayuda de un algoritmo genético fue posible determinar la forma, tamaño y lugar de alrededor de 1000 defectos en líneas que transportan gas natural. En [16]-[17], se aplicaron máquinas de soporte vectorial para mejorar el rendimiento del algoritmo de reconocimiento, en la búsqueda de defectos de pérdida de metal y fugas en tuberías.

En la presente investigación, se aplicaron las técnicas computacionales descritas para reducir el ruido de las señales entregadas por el dispositivo e identificar la soldadura. Sin embargo, el aporte al estado del arte que se busca consiste en identificar la presencia de la soldadura con base en los datos suministrados por la herramienta de inspección inteligente desarrollada por la corporación para la investigación de la corrosión la cual usa tecnología inercial. Sobre la tecnología inercial no existen trabajos reportados en la literatura.

3. METODOLOGÍA

Con el objeto de garantizar calidad en el proceso de minería de datos, se aplicó la metodología CRISP-DM (*Cross-Industry Standard Process for Data Mining*)[18]. A continuación, se presentan las fases más importantes que se implementaron para el desarrollo de este trabajo:

²MFL es la sigla en inglés de magnetic flux leakage

3.1 Comprensión del negocio

El primer paso consistió en comprender que es una soldadura. Una soldadura se define en [19] como “la unión fija de piezas metálicas, realizada con o sin fusión de los bordes a unir, utilizando un metal de aportación o sin él”. La soldadura es el proceso a partir del cual se unen dos metales o más, mediante una ligadura o fusión de metal a metal.

En el transporte de hidrocarburos, las soldaduras se utilizan para unir tubos o un tubo con otro elemento de la infraestructura. En la Fig. 2, se muestra un ejemplo de soldadura aplicada para unir dos tubos.

Fig. 2. SOLDADURA APLICADA EN SECCIÓN DE TUBERÍA.



Fuente: CIC

Cuando la herramienta de inspección inteligente pasa por la soldadura, esta actúa como un obstáculo debido al material adherido al interior del tubo e intenta detener su paso, sin embargo, gracias a la velocidad del fluido con el que viaja, la herramienta toma impulso y continúa su recorrido. En este instante, ciertos sensores de la herramienta registran ese comportamiento.

Localizar todas las soldaduras de una línea de transporte es muy importante. Durante la instalación de la línea de transporte, el constructor le entrega al dueño de la línea la ubicación de todas las soldaduras, no obstante, con el paso del tiempo, se presentan reparaciones o cambios debido a mantenimientos que en algunos casos no son documentados o reportados, lo cual, genera la desactualización de esta información. Adicionalmente, es posible que por vandalismo se instalen en la línea accesorios como válvulas con el objeto de extraer de forma ilícita el hidrocarburo. En este caso se requiere aplicar soldadura para pegar la válvula con la tubería.

Por diseño, en una línea de transporte, la distancia promedio entre soldaduras no sobrepasa los doce (12) metros. El objetivo del análisis de soldaduras consiste en identificar los sitios donde se localizan las soldaduras a partir de variables definidas por los expertos con un error inferior al diez (10) por ciento.

3.2 Comprensión de los datos

Con base en el criterio de los expertos, para identificar una soldadura es necesaria la información de las variables de entrada que se muestran en la tabla I.

**TABLA I
VARIABLES EMPLEADAS POR LOS EXPERTOS PARA IDENTIFICAR UNA SOLDADURA.**

La rotación, mide los cambios en la velocidad angular que experimenta la herramienta a medida que viaja por la tubería. La aceleración mide los cambios en la velocidad que experimenta la herramienta mientras viaja por la tubería. La magnetización mide los cambios en el campo magnético remanente en la tubería

debido al paso previo de un arreglo magnético. El sensor de vibración registra los cambios en la intensidad de las vibraciones mecánicas que se generan con la fricción entre el cuerpo de la herramienta y las paredes de la tubería. La variable posición registra la referencia de distancia que permite medir el desplazamiento de la herramienta por la tubería.

Los valores entregados por la rotación, aceleración y magnetización se representan en los ejes espaciales (X, Y, Z), es decir, existe un dato por cada eje. En total se cuentan con 12 variables de entrada, 3 de aceleración, 3 de rotación, 3 relacionadas con la variación del campo magnético, 2 de vibraciones y 1 de posición. Todos los datos son numéricos y se encuentran en una misma unidad relacionada con la resolución del conversor análogo digital de cada sensor denominada *lsb* excepto la posición, la cual se mide en metros.

Se solicitó a los expertos que se definiera una variable de salida, la cual se denominó soldadura, en la tabla II se muestra las características de esta variable.

TABLA II
CARACTERÍSTICA DE LA VARIABLE DE SALIDA.

Luego, los datos fueron revisados para descartar que existieran valores fuera del margen permitido, con base en la información suministrada por los expertos (ver tabla I). Se realizó una exploración de un conjunto determinado de datos y se encontraron las siguientes hipótesis:

- Cuando la herramienta choca con una soldadura, los canales de vibración registran una intensidad más alta que en otros puntos de la tubería.
- La magnitud de la aceleración, entendiéndose esta como la raíz cuadrada de la suma de los ejes X, Y, Z al cuadrado, es menor o igual a la magnitud de la aceleración antes de llegar a la posible soldadura. Esto comprueba que la herramienta se "desacelera" justo antes de chocar con la soldadura.
- La magnitud de la aceleración es mayor o igual después de pasar por la posible soldadura. Lo anterior significa que la herramienta acelera por un instante de tiempo después de chocar con la soldadura.

Se realizó una exploración a los datos con el objeto de eliminar aquellos registros con valores perdidos, por fuera de los valores permitidos para cada variable o que no fuesen representativos. Después de realizar esta tarea se determinó por ejemplo utilizar sólo la información registrada por la herramienta una vez estuviera desplazándose por la tubería. Esta decisión se tomó porque se encontró que cuando la herramienta se instala en el punto de lanzamiento, pasa un periodo de tiempo considerable antes de llegar a la tubería principal. Mientras la herramienta se encuentre estacionada en el lugar de lanzamiento captura información que no es representativa para el estudio del fenómeno de las soldaduras.

3.3 Preparación de los datos

La información suministrada fue entregada en un archivo plano con extensión csv. La primera línea del archivo representa los nombres de cada una de las variables, tanto las de entrada como la de salida. Las otras líneas contienen los datos de cada variable. Luego, se construyen 5 archivos, con la siguiente cantidad de registros, en adelante instancias: 148, 642, 1464 y 1838 que serán utilizados en la fase de entrenamiento y 600 instancias como datos para pruebas. Cada set de datos contiene igual número de instancias clasificadas en su variable de salida como soldadura y como no soldadura, es decir, para el caso del conjunto de 148, existen 74 instancias con la etiqueta soldadura y 74 instancias como no soldadura. La variable de salida fue validada teniendo en cuenta la información de un documento denominado carta de soldadura suministrado por el dueño de la línea de transporte, el cuál menciona

exactamente la ubicación de las soldaduras aplicadas durante su construcción.

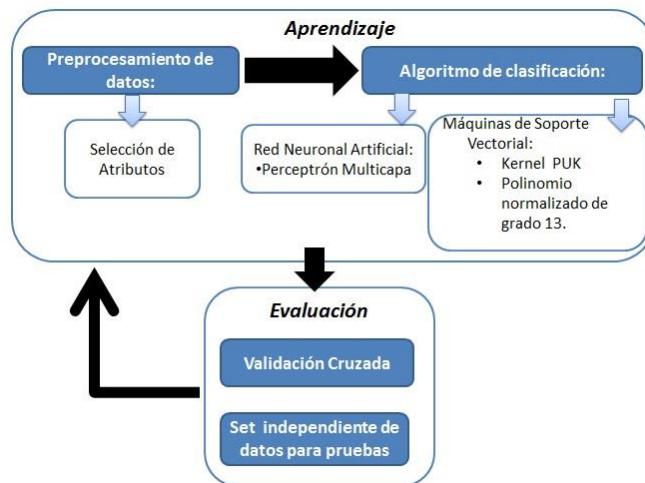
Los 5 archivos son convertidos de formato csv a formato arff por medio de la herramienta Weka³. Weka es una de las herramientas software más completas para el análisis de datos gracias a las diferentes técnicas de pre procesamiento y modelado de datos que implementa. Desarrollado con tecnología Java lo cual le permite correr en casi cualquier plataforma, está disponible libremente bajo licencia GNU y sus librerías se pueden integrar fácilmente en otras aplicaciones software.

Aprovechando las ventajas que ofrece esta herramienta software y siguiendo la recomendación de los trabajos reportados en la literatura, se aplica pre procesamiento con el fin de reducir el número de variables de entrada. Se utiliza la técnica de selección de atributos y el algoritmo *CfsSubsetEval*⁴ [20]. En total, se crean 10 archivos en formato arff los cinco iniciales con todas las variables de entrada y los otros cinco productos del pre procesamiento.

3.4 Modelado

En la Fig. 3 se ilustra el modelo planteado para el reconocimiento de soldaduras, que consta de dos fases, en la primera denominada aprendizaje se parte de los archivos generados en el pre procesamiento para entrenar un algoritmo de clasificación, el cual de acuerdo con los trabajos reportados en la literatura puede ser redes neuronales artificiales [1] ó máquinas de soporte vectorial [16]. Luego en la fase de evaluación, cada algoritmo de clasificación es evaluado usando la técnica de validación cruzada [21] y un set de pruebas no utilizado durante el entrenamiento. Los resultados obtenidos de la evaluación son validados por medio dos métricas, el porcentaje de error y análisis ROC [22]. Si los resultados de la validación cumplen con el objetivo planteado se termina el proceso de lo contrario se vuelve a la fase de aprendizaje para revisar las causas del problema y replantear un nuevo escenario de aprendizaje con los nuevos ajustes.

Fig. 3. MODELO UTILIZADO PARA EL RECONOCIMIENTO DE SOLDADURAS.



Fuente: Autor

³Para mayor información de este software puede visitar su sitio web <http://www.cs.waikato.ac.nz/ml/weka/>

⁴Correlation-based attributes subset selection

4. EXPERIMENTOS Y RESULTADOS

La información con la cual se desarrollaron los experimentos corresponde a una inspección inteligente realizada a un gasoducto colombiano en septiembre de 2012, que cuenta con las siguientes características técnicas:

- 12" de diámetro nominal.
- 36 km de inspección.
- 2'150.835 registros con un tamaño aproximado de 3 GB.
- 3000 soldaduras.

Como se mencionó en la sección 3 y con la ayuda del software Weka, se construyeron 10 set o conjunto de datos, la mitad de ellos se pre procesaron utilizando la técnica de selección de atributos y la otra mitad se dejaron con todas las variables de entrada definidas por el experto. Los datos para el entrenamiento se tomaron de una sección de la inspección y de forma consecutiva, cada archivo de entrenamiento cuenta con igual número de registros con soldadura y con no soldadura. Luego, estos archivos fueron empleados durante el entrenamiento del modelo para el aprendizaje de una red neuronal artificial del tipo perceptrón multicapa y una máquina de soporte vectorial con la función kernel denominada PUK o Pearson Universal Kernel⁵ [23]. El rendimiento de cada modelo es evaluado usando validación cruzada y un set de datos independiente con un total de 600 instancias, los cuales, se tomaron de la sección final de la inspección. Se tiene en cuenta como criterio de éxito que el porcentaje de error con las dos métricas de evaluación sea inferior al diez por ciento. Para cada una de las pruebas al modelo se realizaron ocho (8) experimentos por conjunto de instancias, en cuatro (4) de los ocho (8) experimentos se emplearon las doce (12) variables de entrada definidas por el experto y mencionadas en la fase de comprensión de los datos de la sección 3. En los otros cuatro (4) experimentos se empleó un algoritmo de selección de atributos, el cual recomendó el uso de determinadas variables de entrada para cada caso. A continuación se presentan cada una de las pruebas realizadas.

4.1 Pruebas modelo con 148 instancias

Para las pruebas con este conjunto de datos, en cuatro (4) de los ocho (8) experimentos se empleó el algoritmo *CfsSubsetEval* el cual, recomendó usar casi todas las variables exceptuando los datos del magnetómetro en el eje Z. Los resultados encontrados después de evaluar el modelo usando redes neuronales y máquinas de soporte vectorial medido a partir del indicador de porcentaje de error (ver tabla III), muestran que los algoritmos de clasificación se adaptaron muy bien a los datos de entrenamiento porque el porcentaje de error calculado usando validación cruzada es inferior al cuatro (4) por ciento. Sin embargo, cuando se evaluó el modelo usando datos independientes, el error fue cercano al cincuenta (50) por ciento.

TABLA III
PORCENTAJE DE ERROR DE LOS MODELOS USANDO 148 INSTANCIAS EN EL ENTRENAMIENTO.

Usando la métrica de análisis ROC (ver tabla IV) se confirma que al momento de validar el modelo con los datos independientes el algoritmo de clasificación identifica correctamente sólo uno de los dos estados. El modelo con redes neuronales presenta problemas al reconocer las soldaduras (Sensibilidad =8.3%) y para el caso del modelo con máquinas de soporte vectorial sucede lo contrario, es decir, no reconoce las instancias con no soldaduras (Especificidad = 0%). Se identifica que el modelo requiere más información para ser entrenado correctamente.

⁵Para mayor información de este algoritmo visite la siguiente dirección:
<http://weka.sourceforge.net/doc.dev/weka/classifiers/functions/supportVector/Puk.html>

TABLA IV
ANÁLISIS ROC DE LOS MODELOS USANDO 148 INSTANCIAS EN EL ENTRENAMIENTO.

4.2 Pruebas modelo con 642 instancias

Para las pruebas con este conjunto de datos, en cuatro (4) de los ocho (8) experimentos se empleó el algoritmo *CfsSubsetEval* el cual, recomendó usar las variables rotación en el eje Y, aceleración en los ejes X, Y, magnetización en los ejes X, Z y las vibraciones 1 y 2. Los resultados encontrados después de evaluar el modelo usando redes neuronales y máquinas de soporte vectorial medido a partir del indicador de porcentaje de error (ver tabla V), muestran que los algoritmos de clasificación se adaptaron muy bien a los datos de entrenamiento porque el porcentaje de error calculado usando validación cruzada es inferior al cuatro (4) por ciento. Sin embargo, cuando se evaluó el modelo usando datos independientes, el error estuvo entre el veintiocho (28) al cuarenta y dos (42) por ciento para el caso de la red neuronal y cincuenta (50) por ciento con la máquina de soporte vectorial.

TABLA V
PORCENTAJE DE ERROR DE LOS MODELOS USANDO 642 INSTANCIAS EN EL ENTRENAMIENTO.

Usando la métrica de análisis ROC (ver tabla VI) se observa que al momento de validar el modelo con los datos independientes los algoritmos de clasificación identifican correctamente las soldaduras pero no las no soldaduras. También se observa que el algoritmo que implementa la red neuronal artificial identifica más instancias con no soldadura que en el caso de la máquina de soporte vectorial. Al aumentar el número de instancias se observa una mejora en los resultados y se infiere que al utilizar mayor número de instancias probablemente el modelo identifique mejor el fenómeno.

TABLA VI
ANÁLISIS ROC DE LOS MODELOS USANDO 642 INSTANCIAS EN EL ENTRENAMIENTO.

4.3 Pruebas modelo con 1464 instancias

Para las pruebas con este conjunto de datos, en cuatro (4) de los ocho (8) experimentos se empleó el algoritmo *CfsSubsetEval*, el cual, recomendó usar las variables aceleración en los ejes X, Y, y la vibración 2. Los resultados encontrados después de evaluar el modelo usando redes neuronales y máquinas de soporte vectorial medido a partir del indicador de porcentaje de error (ver tabla VII), muestran que los algoritmos de clasificación se adaptaron correctamente a los datos de entrenamiento porque el porcentaje de error calculado usando validación cruzada es inferior al cinco (5) por ciento. Sin embargo, cuando se evaluó el modelo usando datos independientes, el error estuvo entre el tres (3) al treinta y tres (33) por ciento para el caso de la red neuronal y del quince (15) al cuarenta y seis (46) por ciento con la máquina de soporte vectorial.

TABLA VII
PORCENTAJE DE ERROR DE LOS MODELOS USANDO 1464 INSTANCIAS EN EL ENTRENAMIENTO.

Usando la métrica de análisis ROC (ver tabla VIII) se observa que al momento de validar el modelo con los datos independientes los algoritmos de clasificación identifican correctamente las soldaduras, pero no ocurre lo mismo con las no soldaduras. A diferencia de las pruebas con 642 instancias de entrenamiento, los algoritmos de clasificación presentan un mejor comportamiento durante el reconocimiento de las no soldaduras, inclusive el modelo que utiliza selección de atributos con el algoritmo de redes neuronales artificiales identifica con un margen de error aceptable el fenómeno de la soldadura.

TABLA VIII
ANÁLISIS ROC DE LOS MODELOS USANDO 1464 INSTANCIAS EN EL ENTRENAMIENTO.

Como se observa, el modelo planteado es muy sensible al volumen de datos utilizado durante el entrenamiento, por esta razón y pese a que en este punto ya existe una prueba que cumple con la meta propuesta, se planteó realizar nuevos experimentos con un set de datos más grande.

4.4 Pruebas modelo con 1838 instancias

Para las pruebas con este conjunto de datos, en cuatro (4) de los ocho (8) experimentos se empleó el algoritmo *CfsSubsetEval*, el cual, recomendó usar las variables aceleración en el eje Y, magnetización en el eje Y junto con la vibración 2. De acuerdo con los resultados obtenidos al evaluar cada modelo contra un conjunto de 600 instancias no entrenadas y validación cruzada por medio de la métrica de porcentaje de error (ver tabla IX). Se observa que los dos modelos reconocen mejor la ubicación de la soldadura cuando se utilizan las variables aceleración (AHY), campo magnético de la tubería en el eje Y (MHY) y vibración (VH2). El modelo entrenado que mejor responde a los datos de evaluación es el de máquinas de vectores de soporte con kernel PUK con un error inferior al dos (2%) por ciento.

**TABLA IX
PORCENTAJE DE ERROR DE LOS MODELOS USANDO 1838 INSTANCIAS EN EL ENTRENAMIENTO.**

Los resultados de las pruebas realizadas al conjunto de 600 instancias independientes y por validación cruzada pero utilizando la métrica de curvas ROC (ver tabla X), muestra que cuando se aplica pre procesamiento, la tasa de falsos positivos (número de registros evaluados donde el algoritmo de reconocimiento obtuvo como salida falso y en realidad era verdadero) y verdaderos negativos (número de registros evaluados donde el algoritmo de reconocimiento obtuvo como salida verdadero pero en realidad es falso) disminuye comparado con los modelos que no utilizan el pre-procesamiento. Además el modelo que implementa un algoritmo de máquinas de soporte vectorial es más preciso que el modelo con el algoritmo de redes neuronales artificiales (el valor de la sensibilidad y especificidad es más cercano a 100 % comparado con el modelo que implementa la red neuronal artificial).

**TABLA X
ANÁLISIS ROC DE LOS MODELOS USANDO 1838 INSTANCIAS EN EL ENTRENAMIENTO.**

5. DISCUSIÓN

Con los resultados obtenidos, se puede afirmar que los dos modelos desarrollados permiten reconocer una soldadura de manera eficiente utilizando sólo 3 de los 12 variables de entrada. Lo anterior implica una reducción en el tiempo de procesamiento y análisis, considerable a medida que se procese gran cantidad de datos.

Los atributos que ofrecieron un mejor rendimiento para los modelos desarrollados fueron la aceleración y la variación del campo magnético en el eje Y, así como uno de los sensores de vibración. La herramienta de inspección inteligente se encuentra configurada de tal forma que el eje paralelo a la dirección del fluido es el eje Y. Esto coincide con lo que a priori los expertos habían identificado. No obstante, durante las pruebas, a medida que se aumentaba el número de datos, la información de la rotación en los tres ejes no fue clasificado por el algoritmo de selección de atributos como representativa y fue en estos conjuntos donde se obtuvieron los mejores resultados del modelo. Dicho hallazgo contradice la opinión de los expertos, quienes consideran esta información durante su proceso de análisis. La razón por la cual el modelo sólo utiliza uno de los sensores de vibración se debe a que uno de ellos tiene filtrado y procesamiento en tiempo real el cual intensifica el fenómeno al paso por la soldadura y disminuye el ruido dinámico.

6. CONCLUSIONES

Los resultados de este trabajo muestran que es posible identificar soldaduras de forma automática usando técnicas de inteligencia artificial y minería de datos con un porcentaje de error inferior al cinco por ciento (5%). Además, se concluye que para localizar una soldadura sólo es necesario la información de 3 de las 12 variables de entrada: la aceleración y variación del campo magnético en el eje Y así como uno de los sensores de vibración.

Se demuestra que es posible reconocer de forma automática los sitios donde se localizan las soldaduras en una tubería, lo anterior permitirá a los expertos contar con una herramienta versátil que agilice la entrega de los resultados. Sin embargo, como los modelos desarrollados poseen un margen de error, queda a criterio del experto definir en esos casos de acuerdo con su experiencia y la información de diseño de la línea, si el punto posee o no soldadura. Respecto al estado del arte se puede afirmar que si bien los trabajos [1] y [16] resuelven un problema con un contexto diferente y también con otro tipo de tecnología, es posible usando la tecnología inercial que implementa la herramienta de inspección inteligente desarrollada por la corporación para la investigación de la corrosión y técnicas de inteligencia artificial identificar las soldaduras con una tasa de éxito comparable con la reportada por esos trabajos.

AGRADECIMIENTOS

Este trabajo se realiza gracias a la participación de instituciones comprometidas con la generación de conocimiento y alternativas de solución a problemas de la industria colombiana, se agradece a la empresa transportadora de gas internacional y el instituto colombiano de ciencia y tecnología a partir del proyecto desarrollo de herramientas de inspección – Ition-E.

Además, fue muy importante la cooperación entre las entidades Corporación para la investigación de la corrosión (CIC) y la Universidad autónoma de Bucaramanga (UNAB) en especial el grupo GTI perteneciente a la facultad de ingeniería de sistemas.

REFERENCIAS

- [1] A.A. Carvalho, J.M.A. Rebello, L.V.S. Sagrilo, C.S. Camerini, & I.V.J. Miranda, "MFL signals and artificial neural networks applied to detection and classification of pipe weld defects". *Ndt & E International*, Vol. 39 no. 8, pp. 661-667. Jun. 2006
- [2] R.D. SOUZA, "Avaliação Estrutural de Dutos com Defeitos de Corrosão Reais". Pós-Graduação em Engenharia Mecânica, PUC-Rio, Rio de Janeiro, Brasil, 2003, capítulo 2, sección 2.1, Dissertação de Mestrado, 11p.
- [3] J. D. Hart, G. H. Powell, D. Hackney, and N. Zulfiqar, "Geometry monitoring of the trans-alaska pipeline," in 11-th International Conference on Cold Region Engineering, Anchorage, 2002.
- [4] R. Carneval, M. G. Marinho, and J. Santos, "Flexible line inspection," in European Conference on Nondestructive Testing (ECNDT), 2006.
- [5] M. Beller, "Applying ultrasound for in-line inspection: Facts and issues," in PPSA Aberdeen Seminar, UK, 2006.
- [6] D. Santana, N. Maruyama, and C. Furukawa, "Estimation of trajectories of pipeline pigs using inertial measurements and non linear sensor fusion," in Industry Applications (INDUSCON), 2010 9th IEEE/IAS International Conference on. IEEE, pp. 1-6, 2010.
- [7] A.V. Oppenheim, R.W. Schafer, "Discrete-time signal processing". Engle Wood Cliffs, NJ: Prentice-Hall. 1989.
- [8] M. Misiti, Y. Misiti, G. Oppenheim, J.M. Poggi, "Wavelet tool box for use with MATLAB, user's guide", version 1. Natick, MA: The Math-Works Inc.; 1996.
- [9] J.O. Sophocles, "Introduction to signal processing". Engle Wood Cliffs, NJ: Prentice-Hall.1999.
- [10] A. Joshi, L. Udpa, S. Udpa, and A. Tamburrino, "Adaptive wavelets for characterizing magnetic flux leakage signals from pipeline inspection". *IEEE Transactions on Magnetics*, vol. 42, no. 10, pp. 3168–3170, Oct. 2006.
- [11] M. Afzal and S. Udpa, "Advanced signal processing of magnetic flux leakage data obtained from seamless gas pipeline," *NDT & E International*, vol. 35, no. 7, pp. 449–457, Oct. 2002.
- [12] W. Han and P. Que, "A modified wavelet transform domain adaptive FIR filtering algorithm for removing the SPN in the MFL data," *Measurement*, vol. 39, no. 7, pp. 621–627, Aug. 2006.
- [13] Z. Yang, Z. Xiong, M. Shao, "A new method of leak detection for the natural gas pipeline based on wavelet analysis". *Energy*, 35, 3814–3820, 2010.
- [14] A.A. Carvalho, R.R. Silva, J.M.A. Rebello, L.V.S. Sagrilo, "Pattern recognition techniques applied to the detection and classification of welding defects by magnetic testing". *Research in Nondestructive Evaluation*, vol. 21 no. 2, pp. 91-111, May.2010.
- [15] K. C. Hari, M. Nabi, and S. V. Kulkarni, "Improved FEM model for defect-shape construction from MFL signal by using genetic algorithm," *IET Science, Measurement and Technology*, vol. 1,no. 4, pp. 196–200, July 2007.
- [16] A. Khodayari-Rostamabad, J.P. Reilly, N.K. Nikolova, J.R. Hare, S. Pasha, "Machine learning techniques for the analysis of magnetic flux leakage images in pipeline inspection". *IEEE Transactions on Magnetics*, vol. 45, no. 8, pp. 3073-3084, 2009.
- [17] S. Mandal, F. Chan, M.K. Tiwari, "Leak detection of pipeline: An integrated approach of rough set theory and artificial bee colony trained SVM". *Expert Systems with Applications*, vol. 39, pp. 3071-3080. August 2011.
- [18] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, R. Wirth, "CRISP-DM 1.0 Step-by-step data mining guide", 2000.
- [19] L. Jeffus, "Soldadura: principios y aplicaciones, vol. 3. Editorial Paraninfo, 2009.
- [20] K. Selvakuberan, M. Indradevi, R. Rajaram, "Combined Feature Selection and classification—A novel approach for the categorization of web pages". *Journal of Information and Computing Science*, vol. 3 no. 2, pp. 83-89, 2008.
- [21] A. Krogh, J. Vedelsby, "Neural network ensembles, cross validation, and active learning". *Advances in neural information processing systems*, pp. 231-238, 1995.
- [22] M.J. Burgueño, J.L. García-Bastos, J.M. González-Buitrago, "Las curvas ROC en la evaluación de las pruebas diagnósticas". *Med Clin Barcelona*, vol. 104, no. 17, pp. 661-70, 1995.
- [23] B. Üstün, W.J. Melssen, L.M. Buydens, "Facilitating the application of support vector regression by using a universal Pearson VII function based kernel". *Chemometrics and Intelligent Laboratory Systems*, vol. 81, no 1, pp. 29-40, 2006.

Listado de figuras

Fig. 1. HERRAMIENTA DE INSPECCIÓN INTELIGENTE ACOPLABLE A UN RASPATUBOS.

Fig. 2. SOLDADURA APLICADA EN SECCIÓN DE TUBERÍA.

Fig. 3. MODELO UTILIZADO PARA EL RECONOCIMIENTO DE SOLDADURAS.

Tablas

TABLA I
VARIABLES EMPLEADAS POR LOS EXPERTOS PARA IDENTIFICAR UNA SOLDADURA.

Variable	Notación	Unidad	Valor Aceptado	
			Máximo	Mínimo
Rotación	Gh	lsb	6000	-6000
Aceleración	Ah	lsb	5454	-5454
Magnetización	Mh	lsb	+2500	-2500
Vibración	Vh	lsb	32000	-32000
Posición	Od	m	50000	0

TABLA II
CARACTERÍSTICA DE LA VARIABLE DE SALIDA.

Variable	Notación	Tipo de dato	Valor aceptado	
Soldadura	Soldadura	Texto	S	N

TABLA III
PORCENTAJE DE ERROR DE LOS MODELOS USANDO 148 INSTANCIAS EN EL ENTRENAMIENTO.

	% Error en cada prueba			
	Redes neuronales artificiales		Máquinas de soporte vectorial	
	Validación cruzada	Set independiente	Validación cruzada	Set independiente
Técnica Pre-procesamiento	10 folds	600 instancias	10 folds	600 instancias
Ninguna	3.29	45.78	0.00	50.00
Selección atributos	3.18	45.87	0.00	50.00

TABLA IV
ANÁLISIS ROC DE LOS MODELOS USANDO 148 INSTANCIAS EN EL ENTRENAMIENTO.

Técnica pre-procesamiento empleada	Análisis ROC							
	Redes neuronales artificiales				Máquinas de soporte vectorial			
	Validación cruzada		Set independiente		Validación cruzada		Set independiente	
	10 folds		600 instancias		10 folds		600 instancias	
	Sensibilidad	Especificidad	Sensibilidad	Especificidad	Sensibilidad	Especificidad	Sensibilidad	Especificidad
Ninguna	98.6%	100%	8.3%	100%	100%	100%	100%	0%
Selección Atributos	98.6%	100%	8.3%	100%	100%	100%	100%	0%

TABLA V
PORCENTAJE DE ERROR DE LOS MODELOS USANDO 642 INSTANCIAS EN EL ENTRENAMIENTO.

	% Error en cada prueba			
	Redes neuronales artificiales		Máquinas de soporte vectorial	
	Validación cruzada	Set independiente	Validación cruzada	Set independiente
Técnica Pre-procesamiento	10 folds	600 instancias	10 folds	600 instancias
Ninguna	2.89	28.11	0.00	50.00
Selección atributos (Ghy, Ahx, Ahy, Mhx, Mhz, Vh1, Vh2)	3.48	42.26	0.00	50.00

TABLA VI
ANÁLISIS ROC DE LOS MODELOS USANDO 642 INSTANCIAS EN EL ENTRENAMIENTO.

Técnica pre-procesamiento empleada	Análisis ROC							
	Redes neuronales artificiales				Máquinas de soporte vectorial			
	Validación cruzada		Set independiente		Validación cruzada		Set independiente	
	10 folds		600 instancias		10 folds		600 instancias	
Sensibilidad	Especificidad	Sensibilidad	Especificidad	Sensibilidad	Especificidad	Sensibilidad	Especificidad	
Ninguna	98.1%	99.1%	93.7%	49.7%	98.8%	99.4%	100%	0%
Selección atributos (Ghy, Ahx, Ahy, Mhx, Mhz, Vh1, Vh2)	98.8%	98.8%	100%	15%	98.1%	99.1%	100%	0%

TABLA VII
PORCENTAJE DE ERROR DE LOS MODELOS USANDO 1464 INSTANCIAS EN EL ENTRENAMIENTO.

Técnica Pre-procesamiento	% Error en cada prueba			
	Redes neuronales artificiales		Máquinas de soporte vectorial	
	Validación cruzada	Set independiente	Validación cruzada	Set independiente
	10 folds	600 instancias	10 folds	600 instancias
Ninguna	1.95	32.59	1.64	15.00
Selección atributos (Ahx, Ahy, Vh2)	4.32	3.21	3.14	45.33

TABLA VIII
ANÁLISIS ROC DE LOS MODELOS USANDO 1464 INSTANCIAS EN EL ENTRENAMIENTO.

Técnica pre-procesamiento empleada	Análisis ROC							
	Redes neuronales artificiales				Máquinas de soporte vectorial			
	Validación cruzada		Set independiente		Validación cruzada		Set independiente	
	10 folds		600 instancias		10 folds		600 instancias	
Sensibilidad	Especificidad	Sensibilidad	Especificidad	Sensibilidad	Especificidad	Sensibilidad	Especificidad	
Ninguna	99.5%	99.3%	99.0%	32.3%	98.9%	99.5%	99.7%	85.3%
Selección atributos (Ahx, Ahy, Vh2)	98.5%	98.8%	100%	98.7%	98.5%	97.8%	99.0%	54.7%

TABLA IX
PORCENTAJE DE ERROR DE LOS MODELOS USANDO 1838 INSTANCIAS EN EL ENTRENAMIENTO.

Técnica Pre-procesamiento	% Error en cada prueba			
	Redes neuronales artificiales		Máquinas de soporte vectorial	
	Validación cruzada	Set independiente	Validación cruzada	Set independiente
	10 folds	600 instancias	10 folds	600 instancias
Ninguna	2.67	12.19	1.41	7.33
Selección atributos (Ahy, Mhy, Vh2)	3.67	4.98	2.61	1.33

TABLA X
ANÁLISIS ROC DE LOS MODELOS USANDO 1838 INSTANCIAS EN EL ENTRENAMIENTO.

Técnica pre-procesamiento empleada	Análisis ROC							
	Redes neuronales artificiales				Máquinas de soporte vectorial			
	Validación cruzada		Set independiente		Validación cruzada		Set independiente	
	10 folds		600 instancias		10 folds		600 instancias	
Sensibilidad	Especificidad	Sensibilidad	Especificidad	Sensibilidad	Especificidad	Sensibilidad	Especificidad	
Ninguna	99.2%	99.0%	99.7%	89.3%	99.2%	99.3%	99.7%	93.0%
Selección atributos (Ahx, Ahy, Vh2)	98.6%	98.9%	100%	98.0%	98.3%	99.1%	99.7%	99.0%