

Técnica de aprendizaje automático para la identificación de patrones de flujo bifásico (aceite-agua) a través de tuberías

Daniel Yesid Uribe Tarazona

Trabajo de grado para optar por el título de Ingeniero Mecánico

Director

Octavio Andrés González Estrada

PhD en Ingeniería Mecánica y de Materiales

Codirector

Carlos Mauricio Ruiz Díaz

Magister en Ingeniería Mecánica

Marlon Mauricio Hernández Cely

PhD en Ingeniería Mecánica

Universidad Industrial de Santander  
Facultad de Ingenierías Fiscomecánicas  
Escuela de Ingeniería Mecánica  
Bucaramanga

2024

### **Dedicatoria**

Dedico este trabajo a mis padres, Janette Tarazona Camacho y Henry Yesid Uribe Díaz, por apoyarme y amarme incondicionalmente, por siempre aconsejarme y estar presentes durante mi proceso de formación como ingeniero mecánico.

También dedico este trabajo a mi abuelo José Tarazona, quien en vida siempre me brindó su apoyo, tanto económico como emocional, en mis primeros semestres de carrera. Sé que estaría muy orgulloso de verme desarrollado como ingeniero mecánico y le hubiese encantado verme cumplir esta etapa de mi vida.

Por último, dedico este trabajo a mi abuela Omaira Camacho, mi hermano Diego Uribe y mi tío Jairo Tarazona, por ser un ejemplo para mí y una fuente de motivación.

### **Agradecimientos**

Agradezco en primer lugar a Dios por llenarme de bendiciones, por responder siempre en los momentos de dificultad, por acompañarme en los momentos de gozo, por ser ese apoyo incondicional y por permitirme, según su voluntad, finalizar mi pregrado en ingeniería mecánica.

Agradezco al profesor Octavio Andrés González Estrada, director de mi proyecto de grado, por brindarme su apoyo y orientación durante la realización de este trabajo, además de su amabilidad, paciencia y comprensión, y por sus consejos y amplios conocimientos profesionales.

Agradezco enormemente a Carlos Mauricio Ruiz Díaz, codirector de mi proyecto de grado, amigo y colega, por brindarme la oportunidad de realizar este proyecto junto con su acompañamiento, introduciéndome en esta gran área del conocimiento como lo es la inteligencia artificial y el machine learning. Además, agradezco que haya compartido sus conocimientos conmigo, por brindarme su apoyo, y por su paciencia, amabilidad y profesionalismo. Manifiesto mi gran admiración hacia él en todo lo que es como profesional y como persona.

Agradezco a mi gran amigo Jorge Reyes y Melanie López por su amistad leal y sincera, y por todo el apoyo que me han brindado a lo largo de estos años de amistad.

Agradezco a todas aquellas personas que, de manera directa o indirecta, aportaron a mi formación como profesional.

Agradezco a la Universidad Industrial de Santander por acogerme como estudiante, por formarme como un profesional íntegro y preparado para nuevos retos profesionales y académicos, y por haber sido parte de este logro en mi vida personal.

**Tabla de Contenido**

	<b>Pág.</b>
Introducción.....	11
1. Objetivos.....	15
1.1. Objetivo General .....	15
1.2. Objetivos Específicos .....	15
2. Materiales y métodos.....	16
2.1. Estructuración de la base de datos.....	16
2.2. Patrones de flujo bifásico líquido-líquido en tuberías horizontales .....	18
2.3. Pretratamiento de datos.....	25
2.3.1. Normalización .....	26
2.3.2. Métricas estadísticas de evaluación de correlación entre variables .....	27
2.3.3. Codificación .....	29
2.4. Estructuración del modelo de red neuronal artificial .....	30
2.4.1. Arquitectura de la RNA feedforward backpropagation .....	32
2.4.2. Funciones de activación .....	33
2.4.3. Función de pérdida .....	34
2.5. Selección y evaluación del modelo .....	35
2.5.1. Topología de la Red Neuronal Artificial (RNA) .....	35
2.5.2. Parámetros de optimización .....	36
2.6. Matriz de confusión .....	37
3. Resultados.....	38
4. Conclusiones .....	49

5. Recomendaciones.....	51
Referencias Bibliográficas.....	52

**Lista de Tablas**

	<b>Pág.</b>
Tabla 1 Patrones de flujo en flujo bifásico agua-aceite en tubería horizontal.....	17
Tabla 2 Distribución de puntos obtenidos según el patrón de flujo.....	23
Tabla 3 Interpretación de la magnitud del coeficiente de correlación de Pearson .....	28
Tabla 4 Definición de parámetros para las configuraciones estructuradas de la RNA.....	39
Tabla 5 Algoritmos de entrenamiento a evaluar .....	40
Tabla 6 Resultados de precisión en los modelos obtenidos bajo diferentes configuraciones de parámetros .....	41
Tabla 7 Resultados de la variación en el número de neuronas por capa oculta usando el algoritmo de aprendizaje Resilient Backpropagation en el modelo de RNA 1 .....	42
Tabla 8 Configuración de parámetros y resultados del modelo de RNA seleccionado .....	43
Tabla 9 Métricas para evaluar el rendimiento del modelo en los distintos patrones de flujo a partir de la matriz de confusión.....	48

**Lista de Figuras**

	<b>Pág.</b>
Figura 1 Puntos de datos por autor.....	18
Figura 2 Agrupación de nomenclaturas de patrones de flujo .....	20
Figura 3 Matriz de correlación de las variables de entrada. ....	29
Figura 4 Codificación One-Hot para la variable categórica. ....	30
Figura 5 Modelo estándar de una neurona artificial .....	31
Figura 6 Estructura general de una RNA feedforward backpropagation .....	32
Figura 7 Variación de la entropía cruzada a partir el cambio de iteración en la RNA con la configuración escogida.....	44
Figura 8 Histograma de errores en las fases de entrenamiento, validación y prueba para el modelo de RNA.....	45
Figura 9 Curva ROC en (a) fase de entrenamiento, (b) fase de validación, (c) fase de prueba, (d) Toda la matriz de confusión .....	46
Figura 10 Matriz de confusión para evaluar la capacidad de predicción de patrones de flujo en (a) fase de entrenamiento, (b) fase de validación, (c) fase de prueba, (d) Matriz de confusión total.	47

## **Lista de Apéndices**

**Ver apéndices adjuntos y puedes ser consultados en la base de datos de la Biblioteca**

### **UIS**

Apéndice A. Base de datos en la horizontal de la literatura

Apéndice B. Base de datos en la horizontal de la literatura - Normalizada

Apéndice C. Resultados de precisión y rendimiento de las RNA en diferentes configuraciones topológicas

Apéndice D. Resultados de precisión y rendimiento al repetir las simulaciones en las topologías con mejores resultados

Apéndice E. Resultados de precisión y rendimiento al evaluar la mejor topología de cada función de entrenamiento

Apéndice F. Código del modelo de RNA diseñado

Apéndice G. Manual de usuario para clasificador de patrones de flujo

## Resumen

**Título:** Técnica de aprendizaje automático para la identificación de patrones de flujo bifásico (aceite-agua) a través de tuberías\*

**Autor:** Daniel Yesid Uribe Tarazona\*\*

**Palabras clave:** Aprendizaje automático, Flujo bifásico, Red neuronal artificial, Reconocimiento de patrones de flujo

La industria petrolera ha puesto su interés en desarrollar tecnologías avanzadas para la caracterización de flujos multifásicos, en los cuales existe flujo simultáneo de dos o más fases (líquidas, gaseosas o sólidas) al interior de líneas flexibles tuberías con diferentes diámetros e inclinaciones donde distintos comportamientos geométricos son desarrollados por las sustancias que se están transportando. Dichas configuraciones son denominadas patrones de flujo, en este caso estudiados en tuberías horizontales, considerando el impacto que generan sobre la industria en mención, ya con su determinación es posible optimizar el diseño y operación de los sistemas de transporte, mejorar la eficiencia, reducir costos operativos y prevenir problemas de corrosión y erosión. En este estudio fue estructurada una base de datos con 1846 puntos experimentales de flujo bifásico aceite-agua en tuberías horizontales, obtenidos de la literatura. Asimismo, se estructuró una Red Neuronal Artificial feedforward backpropagation utilizando MATLAB® para predecir patrones de flujo bifásico aceite-agua en tuberías horizontales. Tras evaluar 104 configuraciones topológicas a partir de un proceso de prueba y error, el modelo óptimo incluyó una capa de entrada con seis neuronas, tres capas ocultas con 50 neuronas cada una, una capa de salida con seis neuronas, un algoritmo de entrenamiento Resilient Backpropagation, funciones de activación sigmoide en capas ocultas y softmax en la capa de salida, y la entropía cruzada como función de pérdida. El modelo final demostró una precisión total del 95,4%, prediciendo con éxito los seis patrones de flujo en todas las etapas de entrenamiento, validación y prueba, con un tiempo de entrenamiento de solo 2 segundos y un error de entropía cruzada de 0,024. Estos resultados sugieren que el modelo RNA es una herramienta eficaz para predecir patrones de flujo con alta precisión y bajo costo computacional, con aplicaciones potenciales en la industria de oil & gas.

---

\* Trabajo de Grado

\*\* Facultad de Ingenierías Fisicomecánicas. Escuela de ingeniería Mecánica. Director: Octavio Andrés González Estrada, PhD en Ingeniería Mecánica y de Materiales. Codirectores: Carlos Mauricio Ruiz Diaz, Magister en Ingeniería Mecánica, Marlon Mauricio Hernández Cely, PhD en Ingeniería Mecánica

### Abstract

**Title:** Machine Learning Technique for the Identification of Two-Phase Flow Patterns (Oil-Water) Through Pipelines\*

**Author:** Daniel Yesid Uribe Tarazona\*\*

**Keywords:** Artificial neural network, Flow pattern recognition, Machine Learning, Two-phase flow.

The oil industry has shown interest in developing advanced technologies for the characterization of multiphase flows, where there is a simultaneous flow of two or more phases (liquid, gas, or solid) inside flexible pipelines with different diameters and inclinations, where various geometric behaviors are developed by the transported substances. These configurations are known as flow patterns, studied here in horizontal pipelines, considering their impact on the mentioned industry. Their determination allows for the optimization of the design and operation of transport systems, improving efficiency, reducing operational costs, and preventing corrosion and erosion problems. In this study, a database with 1846 experimental points of oil-water two-phase flow in horizontal pipelines was structured, obtained from the literature. Additionally, a feedforward backpropagation Artificial Neural Network (ANN) was structured using MATLAB® to predict oil-water two-phase flow patterns in horizontal pipelines. After evaluating 104 topological configurations through a trial-and-error process, the optimal model included an input layer with six neurons, three hidden layers with 50 neurons each, an output layer with six neurons, a Resilient Backpropagation training algorithm, sigmoid activation functions in the hidden layers and softmax in the output layer, and cross-entropy as the loss function. The final model demonstrated an overall accuracy of 95.4%, successfully predicting the six flow patterns in all stages of training, validation, and testing, with a training time of only 2 seconds and a cross-entropy error of 0.024. These results suggest that the ANN model is an effective tool for predicting flow patterns with high accuracy and low computational cost, with potential applications in the oil & gas industry.

---

\* Degree Work

\*\* Faculty of Physicomechanical Engineering. School of Mechanical Engineering. Director: Octavio Andrés González Estrada, PhD in Mechanical and Materials Engineering. Co-Directors: Carlos Mauricio Ruiz Diaz, Master's in Mechanical Engineering, Marlon Mauricio Hernández Cely, PhD in Mechanical Engineering

## Introducción

La industria petrolera se ha centrado en desarrollar tecnologías que permitan obtener sistemas actualizados para la medición precisa del flujo multifásico (Hernández-Cely et al., 2022), refiriéndose este al flujo simultáneo de dos o más fases, que pueden ser líquidas, gaseosas o sólidas. Estas fases pueden estar conformadas por sustancias en la misma fase, como aceite y agua, o por dos fases de un solo componente, como vapor de agua y líquido en una tubería de conducción (Al-Naser et al., 2016). La identificación de patrones de flujo en tuberías horizontales es de vital importancia en diversas industrias, especialmente en la petroquímica y la de procesamiento de fluidos (Alhashem, 2020), considerando que la determinación de los mismos permite optimizar el diseño y operación de los sistemas de transporte, mejorando la eficiencia y reduciendo costos operativos (Soot, 1970). Una correcta identificación de los patrones de flujo previene problemas de corrosión y erosión en las tuberías, a partir de la optimización del proceso de inyección de inhibidores de corrosión y otros químicos, lo que a su vez prolonga la vida útil de la infraestructura y reduce los costos de mantenimiento (Al-Sarkhi et al., 2017). De igual manera, la identificación de los patrones de flujo es crucial para la predicción y el control de fenómenos como la transferencia de calor, la caída de presión y la distribución de fases (Osundare et al., 2020), lo cual impacta directamente en la seguridad y el rendimiento de los procesos industriales.

Los patrones de flujo se definen como una descripción cualitativa de la distribución de las fases del flujo dentro de la tubería que conduce dos fluidos inmiscibles (El-Sebakhy, 2010), estos pueden variar significativamente y dependen de factores como la velocidad superficial de los

fluidos, sus propiedades físicas y las características geométricas de la tubería, además de su gran dependencia con las condiciones de presión y temperatura del local de análisis. La identificación de distribución de fase es comúnmente determinada a partir del procesamiento de imágenes y utilizando diversas técnicas como el cierre rápido de válvulas o con dispositivos con mayor precisión como el densitómetro de rayos gama. En trabajos previos realizados por Hernández et al. (2020) y Lum et al. (2006) se utilizó un sensor de malla de alambre (WMS) basado en la permitividad eléctrica del fluido para obtener tomografías del flujo multifásico, y una cámara de alta velocidad para validar los patrones de flujo. Otros estudios emplearon cámaras de alta velocidad para identificar visualmente los patrones de flujo en pruebas experimentales con fluidos inyectados a través de las tuberías (Abduvayt et al., 2004; Cai et al., 2012; Dasari et al., 2013).

En este contexto, las técnicas de machine learning son empleadas para identificar patrones de flujo, destacándose como una herramienta prometedora en la resolución de problemas de ingeniería, su eficacia radica en abordar situaciones de alta complejidad, dinámicas y no lineales, lo que ha permitido aumentar su relevancia en los últimos años (Yan et al., 2018). Las Redes Neuronales Artificiales (RNA) específicamente, como técnica de machine learning han mostrado un gran potencial para la identificación y predicción de patrones de flujo, las RNA son modelos computacionales utilizados en el procesamiento de información, cuyo diseño y funcionamiento está inspirado en el sistema nervioso biológico del cerebro (Abba et al., 2020). Estas redes pueden manejar grandes volúmenes de datos y aprender relaciones complejas entre múltiples variables (Chimeno-Trinchet et al., 2020), permitiendo así reconocer patrones, clasificar datos y pronosticar eventos futuros (Bahrami et al., 2019; G. H. Roshani et al., 2014), lo que las hace ideales para abordar los desafíos asociados con el flujo bifásico.

Diversos estudios han implementado técnicas de machine learning en la identificación de patrones de flujo. En el estudio desarrollado por Figueredo et al. (2016) se implementó una técnica de aprendizaje automático para identificar el régimen de flujo utilizando un transductor de presión diferencial de flujo multifásico. En la investigación hecha por Roshani et al. (2021) propusieron una técnica inteligente no destructiva basada en la combinación de la atenuación de la radiación gamma y la inteligencia artificial para determinar patrones de flujo. En el trabajo desarrollado por Salgado et al. (2010) se presentó una metodología para identificar patrones de flujo usando redes neuronales a partir del reconocimiento e interpretación de las distribuciones de altura de pulso (PHD) obtenidas por detectores de radiación gamma. En el trabajo desarrollado por Ruiz et al. (2024) se desarrolló un modelo inteligente basado en Transformer Neural Network capaz de predecir patrones de flujo dentro de tuberías verticales con flujo bifásico líquido-líquido de aceite y agua.

Actualmente, se han desarrollado grandes avances tecnológicos referentes a la industria 4.0 y el internet de las cosas, en los que el aprendizaje automático cumple un papel fundamental en la resolución de problemas de ingeniería gracias a sus facultades para resolver de manera eficaz problemas de alta complejidad y no lineales. Motivados por estos avances, el objetivo del presente estudio es generar un modelo predictivo y una metodología de aprendizaje automático basada en redes neuronales artificiales para identificar y predecir patrones de flujo bifásico en tuberías horizontales con aceite y agua en fases líquidas, utilizando el software MATLAB®. Para ello, primero se recopilan y procesan los datos de flujo bifásico aceite y agua obtenidos de la literatura. En la sección siguiente se discute la estructuración del modelo de Red Neuronal Artificial, posteriormente, se lleva a cabo el respectivo proceso de selección y evaluación de las diferentes

configuraciones implementadas a partir de los resultados de rendimiento y precisión obtenidos.

Finalmente, se presentan los resultados obtenidos para el modelo de Red Neuronal Artificial escogido, seguido de las conclusiones y recomendaciones.

## 1. Objetivos

### 1.1. Objetivo General

Identificar patrones de flujo bifásico conformado por aceite y agua en fases líquidas utilizando aprendizaje automático para optimizar el proceso de transporte de fluidos a través de tuberías.

### 1.2. Objetivos Específicos

- Identificar la influencia de la velocidad superficial de los fluidos, la viscosidad, y la geometría de la tubería, en la formación de patrones de flujo bifásico conformado por aceite y agua en fases líquidas, a partir del estudio y caracterización de este, haciendo uso de bases de datos existentes en la literatura.
- Determinar el grado de intervención de los parámetros físicos y geométricos en la formación de patrones de flujo bifásico conformado por aceite y agua en fases líquidas, a partir de la aplicación de la técnica de análisis de componentes principales (PCA), para la estructuración de un modelo de red neuronal artificial (RNA) aplicado en flujos bifásicos.
- Generar un modelo predictivo basado en redes neuronales artificiales para la identificación de patrones de flujo bifásico conformado por aceite y agua en fases líquidas, haciendo uso del software de cómputo numérico MATLAB.
- Desarrollar una metodología de aprendizaje automático mediante redes neuronales para la identificación de patrones de flujo bifásico conformado por aceite y agua en fases líquidas

## 2. Materiales y métodos

### 2.1. Estructuración de la base de datos

La base de datos usada para el entrenamiento, validación y prueba de la red neuronal artificial se estructuró a partir de la agrupación de información característica de flujo bifásico aceite y agua en tubería horizontal, para la cual fueron estudiados los trabajos desarrollados por 11 autores de la literatura, de los cuales se extrajeron 1846 puntos experimentales (Figura 1), estos datos se presentan en Apéndice A. En la **Tabla 1** se presenta la información contenida en la base de datos, resumiendo así características propias de cada trabajo analizado.

La extracción de los datos se realizó a partir de los mapas de patrones de flujo observados en la literatura. Las variables recopiladas en los diferentes trabajos son las velocidades superficiales de agua ( $V_{sw}$ ) entre 0.01097 y 3.7193 m/s, velocidades superficiales de aceite ( $V_{so}$ ) entre 0.01044 y 3.0392 m/s, velocidad de la mezcla ( $V_m$ ) entre 0.02631 y 5.4714 m/s y las fracciones volumétricas de los fluidos de aceite ( $C_o$ ) y agua ( $C_w$ ). Otras características importantes incluidas en la recopilación de datos para caracterizar el fluido son la viscosidad del aceite ( $\mu_o$ ) entre un rango de 0.00188 a 5.6 Pa·s, la densidad del aceite ( $\rho_o$ ) de 800 a 910 kg/m<sup>3</sup>, y el diámetro interno de la tubería ( $D$ ) entre 1.9 y 10.64 cm.

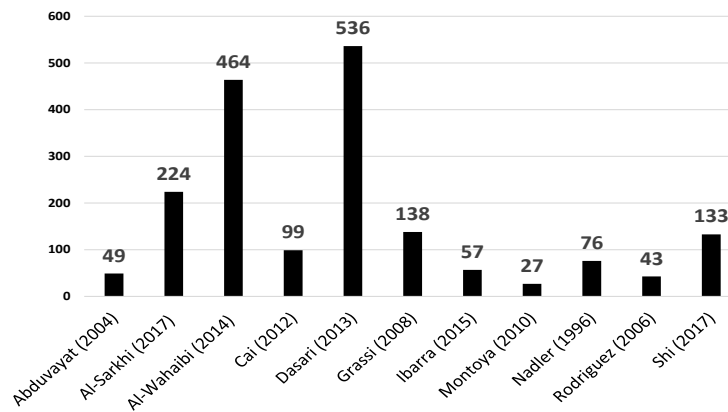
**Tabla 1***Patrones de flujo en flujo bifásico agua-aceite en tubería horizontal*

<b>Autores</b>	<b>Diámetro interno tubería (m)</b>	<b>Material tubería</b>	<b>Viscosidad aceite (Pa.s)</b>	<b>Viscosidad agua (Pa.s)</b>	<b>Densidad aceite (<math>\frac{kg}{m^3}</math>)</b>	<b>Densidad agua (<math>\frac{kg}{m^3}</math>)</b>	<b>Nomenclatura de Patrón de flujo presentada</b>
(Abduvayt et al., 2004)	0.1064	Acrílico	0.00188	0.00072	800	1000	ST-S, O/TP & W, ST-WD/O & OD/W, ST-WD/O & W, SR-WD/O & W, ThO/TP & FDO/W, DW/O & W, FD W/O & FDO/W
(Al-Sarkhi et al., 2017)	0.0508 0.0508	Plexiglass	0.013 0.0288	0.00097 0.00097	858.5 884	994 1037	ST, ST-MI, SW ST, ST-MI
(Al-Wahaibi et al., 2014)	0.019 0.0254	Acrílico	0.012 0.012	0.001 0.001	875 875	998 998	AN, Bb, Dw/o, DC, Do/w, ST.
(Cai et al., 2012)	0.1	Acero inoxidable	0.002	0.000898	825	997	Dispersed W/O, semi-dispersed W/O, Smooth stratified, Stratified with globules, Stratified with mixing layer.
(Dasari et al., 2013)	0.025	Perspex	0.107	0.001	889	1000	Do/w, Dw/o, P, S, SM, SW,
(Grassi et al., 2008)	0.021	Polycarbonato	0.799	0.0013	886	1000	AN, AN-o/w, Do/w, PL/SL, ST
(Ibarra et al., 2015)	0.032	Acrílico	0.0054	0.0009	825	998	D, D ow&w, DC, ST, SWD
(Montoya et al., 2009)	0.0445	Plexiglass	0.0884	0.000898	884	1000	Do/w & w, o/w, ST, ST & MI
(Nädler & Mewes, 1997)	0.059	Perspex	0.027	0.001053	850	998	w/o & W, O/W, SM, ST, W & o/w, W/O, w/o-o/w & W

(Rodriguez & Oliemans, 2006)	0.0828	Acero inoxidable	0.00717	0.00076	831.4	1070	Do/w & w, Dw/o & Do/w, o/w, ST, ST & MI, w/o.
(Shi & Yeung, 2017)	0.026	Perspex	5.6	0.001235	910	999	Core flow, Dispersed oil lumps, Oil plugs
(Shi et al., 2017)	0.026		5	0.001002	910	997	Core annular flow, Oil lumps, Oil plugs,

**Figura 1**

*Puntos de datos por autor.*



**2.2. Patrones de flujo bifásico líquido-líquido en tuberías horizontales**

En los experimentos realizados por los autores, se observaron diferentes patrones de flujo comúnmente reportados en la literatura para flujo bifásico aceite-agua en tubería horizontal, como lo son el patrón de flujo estratificado (ST), estratificado con mezcla (ST & MI), slug (S), anular (AN), dispersión de aceite en agua (Do/w) y dispersión de agua en aceite (Dw/o).

Asimismo, se observó que la terminología utilizada por diferentes autores para los mismos patrones de flujo o de naturaleza similar difiere entre sí, lo que ha provocado perplejidad. Así que, para simplificar dicho problema, se agruparon los patrones de flujo que muestran naturaleza similar o que son nombrados de diferentes formas por los distintos autores y se eligió una única nomenclatura para cada grupo, como se muestra en la **Figura 2**.

Para bajas velocidades superficiales de las dos fases, el patrón de flujo es estratificado (ST) con una completa separación de los dos fluidos, a medida que la velocidad del aceite ( $V_{so}$ ) se incrementa comienza a barrer el agua, provocando ondulaciones en la interfaz, provocando patrón de flujo estratificado ondulado.

Ante un nuevo aumento de la ( $V_{so}$ ), se aumenta la amplitud de las ondas del agua, provocando una descomposición en gotas finas que posteriormente se dispersan en la fase continua del aceite, provocando una capa de aceite y agua dispersa en aceite (Dw/o).

Por otra parte, al incrementar la velocidad del agua ( $V_{sw}$ ), a partir del flujo estratificado ondulado, se presenta un patrón de flujo estratificado con una capa intermedia de mezcla de gotas de aceite y agua, conocida como estratificado con mezcla (ST & MI).

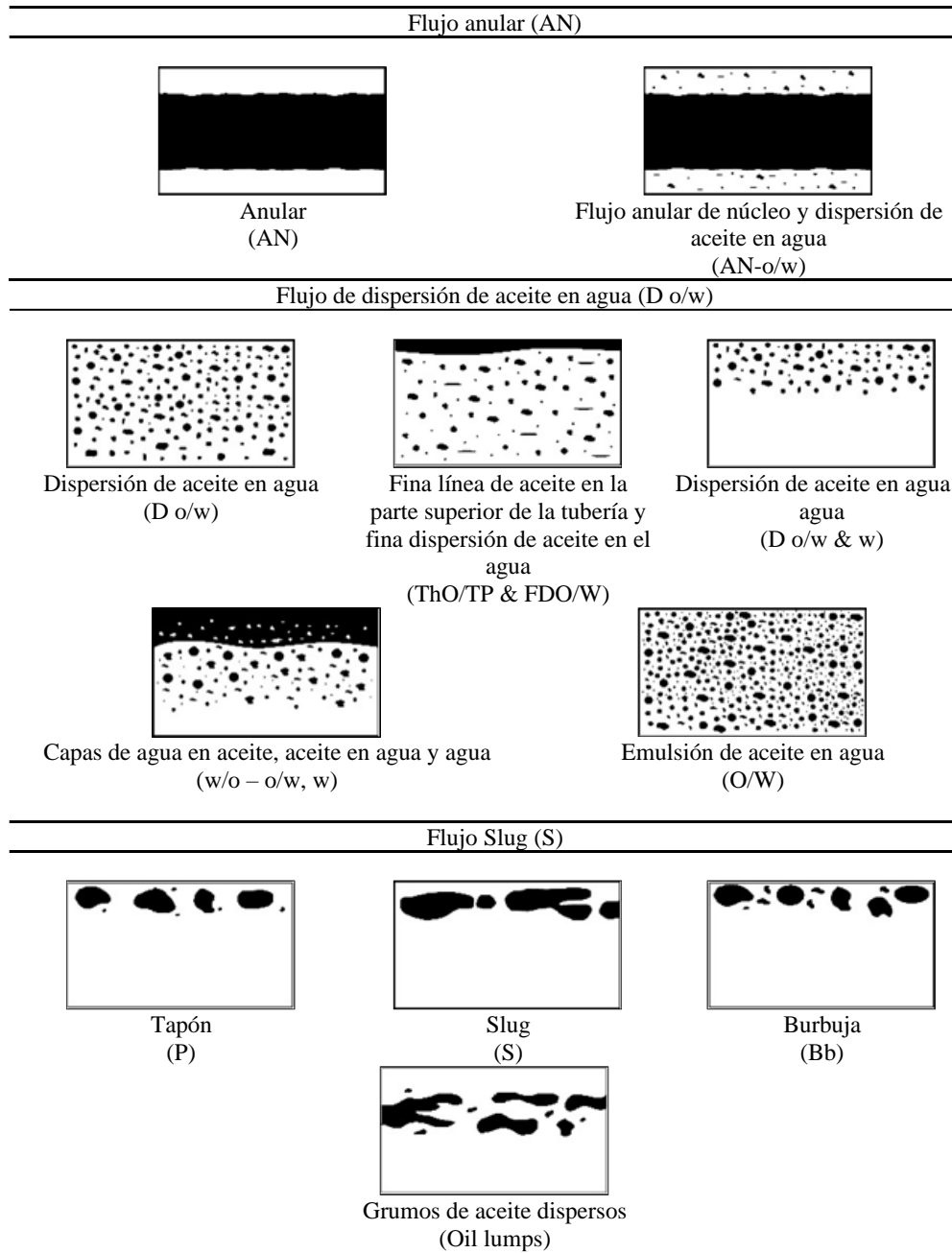
Al continuar aumentando ( $V_{sw}$ ), el patrón cambia con gotas de aceite dispersas en agua y agua (Do/w & w) y después se genera el patrón dispersión de aceite en agua (Do/w), en el que las gotas de aceite están uniformemente distribuidas en toda la fase de agua. A velocidades de aceite más bajas, la interfaz ondulada tiende a ocasionalmente a tocar la pared superior, formando el

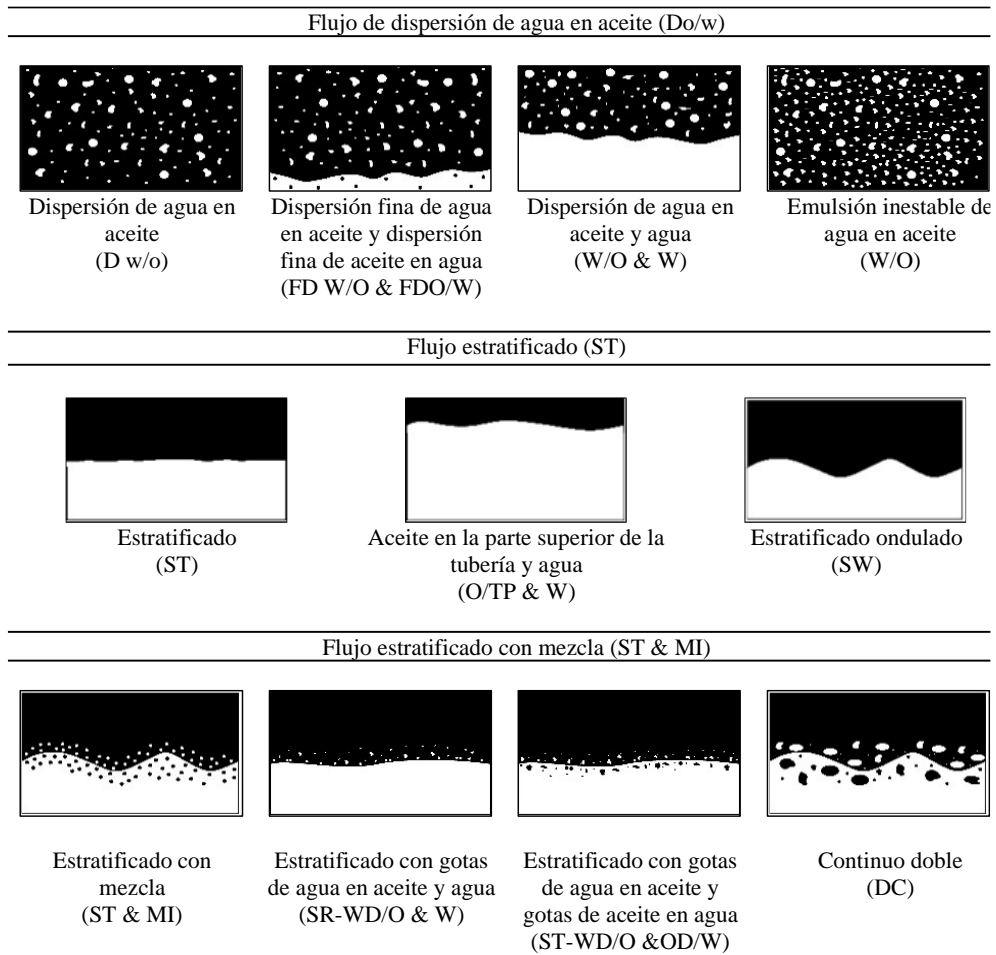
patrón de flujo plug, o slug, presentando formas de tapones irregulares, deformados, algunas veces se muestra como gotas más grandes, alargadas o aglomerado de gotas irregulares.

Así mismo, al analizar la base de datos estructurada, se pueden identificar algunas incidencias de las propiedades de los fluidos y de las tuberías en las categorías de patrones de flujo agrupadas. Es de notar que, en las tuberías de diámetro igual o mayor a 0,032 m hasta 0,1064 m, las categorías de patrones de flujo que se observan son las dispersiones de agua en aceite (Do/w), aceite en agua (Dw/o), estratificado (ST) y estratificado con mezcla (ST & MI). En las tuberías con diámetro menor, desde 0.019 m hasta 0.032 m es posible observar también los patrones de flujo anular (AN) y slug (S). Se observa que en los fluidos considerados que tienen la más alta viscosidad (5 y 5.6 Pa·s) y un diámetro en el rango de los menores valores registrados (0.026 m), se presentan solo patrones de flujo de naturaleza AN y S, asimismo, en el fluido que presenta la siguiente mayor viscosidad (0.799 Pa·s) y un diámetro de 0.021 m, también experimenta los patrones de flujo AN y S, permitiendo identificar que los fluidos con viscosidades altas y diámetros más pequeños son más propensos de experimentar estos patrones de flujo.

## **Figura 2**

*Agrupación de nomenclaturas de patrones de flujo*





Una vez agrupadas las nomenclaturas de patrones de flujo en las seis categorías de acuerdo con la naturaleza del patrón de flujo observada, se presenta la **Tabla 2**, donde se evidencia la distribución de los 1846 puntos experimentales recolectados de la literatura.

**Tabla 2***Distribución de puntos obtenidos según el patrón de flujo*

<b>Patrón de flujo</b>	<b>AN</b>	<b>D w/o</b>	<b>D o/w</b>	<b>S</b>	<b>ST</b>	<b>ST &amp; MI</b>
<b>Cantidad de datos</b>	169	285	379	185	430	398
<b>% de datos</b>	9.15%	15.44%	20.53%	10.02%	23.29%	21.56%

Las variables obtenidas de la literatura, para estructurar la base de datos son, la velocidad superficial del agua ( $V_{sw}$ ), velocidad superficial del aceite ( $V_{so}$ ), velocidad de la mezcla ( $V_m$ ), diámetro de la tubería ( $D$ ), viscosidad del aceite ( $\mu_o$ ), fracción volumétrica del agua ( $C_w$ ), fracción volumétrica del aceite ( $C_o$ ) y densidad del aceite ( $\rho_o$ ).

Un parámetro importante al momento de inyectar los fluidos al interior de una tubería, que cumple un papel fundamental en la formación de los patrones de flujo es la velocidad superficial. La velocidad superficial de una fase es el flujo volumétrico de la fase, que representa el caudal volumétrico por unidad de superficie. En otras palabras, la velocidad superficial de una fase es la velocidad que se produciría si esa fase de la respectiva sustancia fluyera sola por la tubería (Shoham, 2005). Así, las velocidades superficiales de las fases líquidas del agua ( $V_{sw}$ ) y aceite ( $V_{so}$ ) son, respectivamente calculadas con la ecuación (1) y (2):

$$V_{sw} = \frac{q_w}{Ap} \quad (1)$$

$$V_{so} = \frac{q_o}{Ap} \quad (2)$$

Donde  $Ap$  es el área de la sección transversal de la tubería,  $q_w$  es el caudal volumétrico del agua y  $q_o$  es el caudal volumétrico del aceite. La velocidad de la mezcla es el caudal volumétrico total de ambas fases por unidad de superficie, que se denomina velocidad del centro de volumen y viene dada por (3)

$$V_m = \frac{q_w + q_o}{Ap} = V_{sw} + V_{so} \quad (3)$$

La fracción volumétrica del agua ( $C_w$ ) es la fracción de un elemento de volumen en un campo de flujo bifásico ocupado por la fase líquida del agua, y la fracción volumétrica del aceite ( $C_o$ ) que corresponde a la fracción de un elemento de volumen en un campo de flujo bifásico ocupado por la fase líquida del aceite, se calculan con las fórmulas (4) y (5):

$$C_w = \frac{V_{sw}}{V_{sw} + V_{so}} \quad (4)$$

$$C_o = \frac{V_{so}}{V_{sw} + V_{so}} \quad (5)$$

Cabe recordar que la suma de la fracción volumétrica del agua y la fracción volumétrica del aceite da como resultado 1. Al sumar la ecuación (4) y (5), se presenta la ecuación (6):

$$\frac{V_{sw}}{V_{sw} + V_{so}} + \frac{V_{so}}{V_{sw} + V_{so}} = 1 \quad (6)$$

Una forma alternativa para calcular la fracción volumétrica del aceite se define en (7):

$$C_o = 1 - C_w \quad (7)$$

### 2.3. Pretratamiento de datos

En el ámbito del aprendizaje automático, los procesos no lineales complejos presentan una gran diversidad en la dimensionalidad de las entradas y salidas, así como en el tamaño del conjunto de datos. La complejidad de los modelos de redes neuronales y la carga computacional aumentan significativamente con la dimensión de los datos (Zhao et al., 2023), para abordar este problema, se considera reducir la dimensión de estos sin comprometer la precisión del modelado. Las técnicas de reducción de orden se dividen en selección supervisada o no supervisada. En conjuntos de datos etiquetados, la selección supervisada revela la importancia de las características mediante la correlación con la variable objetivo y entre subconjuntos de variables (Xie et al., 2023).

La selección supervisada incluye métodos de extracción y selección de características. Los métodos tradicionales de extracción, como mínimos cuadrados parciales y análisis de componentes principales (PCA), crean nuevas características en un espacio de baja dimensión, manteniendo la mayor parte de la información relevante, pero pueden carecer de interpretabilidad física,

inicialmente se consideró utilizar este método para reducir la dimensión de la base de datos, pero ante la mencionada carencia de interpretabilidad física que se obtenía, se optó por buscar otro. Por otra parte, los métodos de selección de características eligen un subconjunto de las características originales, altamente correlacionadas con la salida del sistema y más interpretables, por lo que se prefirió este método. Las técnicas de selección de características se clasifican en métodos envolventes, de filtro e intrínsecos. Se selecciona el método de filtro, que es generalmente utilizado como preprocesamiento de datos (Guyon & De, 2003), el cual selecciona variables basándose en características estadísticas, como la correlación de Pearson, para evaluar la relación entre las variables de entrada y seleccionar las más relevantes (Liu et al., 2022).

### ***2.3.1. Normalización***

Para hacer el entrenamiento de la red es importante homogenizar la información que se utilizará como entradas del modelo de aprendizaje automático, suministrando a la red datos claros que permitan relacionar las variables suministradas con el patrón a predecir.

Por esto, se realizó una normalización de las variables de entrada entre unos límites definidos de 0 a 1. La normalización se define como un cambio de escala de los datos originales para que estos estén en un rango específico (Ruiz et al., 2024). Se normaliza el vector de entrada de la red, que contiene las variables de  $V_{sw}$ ,  $V_{so}$ ,  $V_m$ ,  $Cw$ ,  $Co$ ,  $D$ ,  $\mu_o$  y  $p_o$ , esto se hace a partir de la ecuación (8):

$$X_{Normalizada} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (8)$$

En donde  $X_{normalizada}$  corresponde al valor normalizado que toma un valor entre 0 y 1,  $X$  corresponde al valor puntual de la muestra de la variable a normalizar,  $X_{min}$  y  $X_{max}$  corresponde al valor mínimo y máximo dentro del conjunto de datos de la variable a normalizar. La base de datos de flujo bifásico aceite y agua en tubería horizontal con los valores normalizados se puede ver en Apéndice B.

### 2.3.2. Métricas estadísticas de evaluación de correlación entre variables

Una vez normalizados los valores, el método a emplear que a su vez es el método de evaluación estadística más utilizado para la correlación es el coeficiente de correlación de Pearson, que captura la relación lineal entre dos matrices (Xie et al., 2023). Dos variables altamente correlacionadas pueden proporcionar información redundante, en estos casos se puede eliminar una de las variables correlacionadas para simplificar el proceso.

El coeficiente de correlación de Pearson entre una característica ( $x$ ) y una etiqueta ( $y$ ) con  $n$  ejemplos de entrenamiento se calcula con la ecuación (9).

$$r = \frac{\sum_{i=1}^n ((x_i - \bar{x})(y_i - \bar{y}))}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (9)$$

En donde  $\bar{x}$  y  $\bar{y}$  son las medias de los dos vectores, respectivamente. El rango del coeficiente se sitúa entre -1 y 1, donde un valor de cero implica que no hay correlación lineal. Los valores cercanos a 1 y -1 indican correlaciones positivas y negativas, respectivamente.

En la **Figura 3** se presenta una matriz de correlación entre las variables de entrada, mostrando el nivel de correlación entre estas para así reducir la redundancia eliminando aquellas con alta dependencia entre sí. La interpretación cualitativa de esta propiedad sigue las sugerencias de Cohen, (1988), ampliamente aceptadas en la comunidad científica. La **Tabla 3** muestra la clasificación e interpretación de la magnitud del coeficiente de correlación de Pearson, aplicable a cualquier par de variables, considerando el valor absoluto del coeficiente para que la magnitud sea independiente del signo.

**Tabla 3**

*Interpretación de la magnitud del coeficiente de correlación de Pearson*

Rango de valores de $r_{xy}$	Interpretación
$0.00 \leq  r_{xy}  < 0.10$	Correlación nula
$0.10 \leq  r_{xy}  < 0.30$	Correlación débil
$0.30 \leq  r_{xy}  < 0.50$	Correlación moderada
$0.50 \leq  r_{xy}  < 1.00$	Correlación fuerte

Para el presente trabajo se elimina una variable del par de variables evaluadas que presentan una fuerte correlación de Pearson entre ellas, con un valor de coeficiente mayor a 0.8. La variable fracción volumétrica del agua ( $C_w$ ) y fracción volumétrica del aceite ( $C_o$ ) muestran una correlación fuerte con un valor de -1, por lo que se decidió descartar la variable fracción volumétrica del aceite ( $C_o$ ) de la base de datos. Por otra parte, entre las variables velocidad superficial del agua ( $V_{sw}$ ) y velocidad de la mezcla ( $V_m$ ) se obtuvo un coeficiente de correlación de 0.837, el cual refleja una fuerte correlación, por lo tanto, se excluye de la base de datos la variable

velocidad de la mezcla ( $V_m$ ). De esta manera las variables de entrada a utilizar se reducen a: velocidad superficial del agua ( $V_{sw}$ ), velocidad superficial del aceite ( $V_{so}$ ), fracción volumétrica del agua ( $C_w$ ), diámetro de la tubería ( $D$ ), viscosidad del aceite ( $\mu_o$ ) y densidad del aceite ( $\rho_o$ ).

### 2.3.3. Codificación

El patrón de flujo está definido en la base de datos como una variable categórica, por lo que se debe codificar la etiqueta que se tiene para las diferentes categorías. Se asigna un número del 0 al 5 a cada uno de los seis patrones de flujo definidos. Para convertir las variables categóricas definidas de forma entera en nuevas columnas categóricas con un valor binario de 1 o 0, se utilizó un método de preprocesamiento de datos denominado codificación one-hot, y así cada nueva columna representa una categoría de la variable, según se observa en Figura 4. Si una categoría está representada por esta columna, se le asigna 1, en caso contrario, se le asigna un 0 (Liu et al., 2022).

### Figura 3

*Matriz de correlación de las variables de entrada.*

	$V_{so}$ [m/s]	$V_{sw}$ [m/s]	$V_m$ [m/s]	$C_w$ [-]	$C_o$ [-]	$D$ [m]	$\mu_o$ [Pa.s]	$\rho_o$ [kg/m <sup>3</sup> ]
$V_{so}$ [m/s]	1							
$V_{sw}$ [m/s]	0,053	1						
$V_m$ [m/s]	0,590	0,837	1					
$C_w$ [-]	-0,590	0,542	0,115	1				
$C_o$ [-]	0,590	-0,542	-0,115	-1	1			
$D$ [m]	0,196	-0,242	-0,088	-0,279	0,279	1		
$\mu_o$ [Pa.s]	-0,175	0,022	-0,078	0,195	-0,195	-0,167	1	
$\rho_o$ [kg/m <sup>3</sup> ]	-0,270	0,111	-0,058	0,296	-0,296	-0,799	0,448	1

**Figura 4**

*Codificación One-Hot para la variable categórica.*

ID	Patron de flujo	OneHot Encoder	ID	ST	ST&MI	Do/w	Dw/o	AN	S
0	ST		0	1	0	0	0	0	0
1	ST & MI		1	0	1	0	0	0	0
2	Do/w		2	0	0	1	0	0	0
3	Dw/o		3	0	0	0	1	0	0
4	AN		4	0	0	0	0	1	0
5	S		5	0	0	0	0	0	1

## 2.4. Estructuración del modelo de red neuronal artificial

Para este estudio y con el fin de generar un modelo capaz de predecir patrones de flujo, se utiliza una red neuronal artificial (RNA), herramienta basada en aprendizaje automático, usada para el procesamiento de información, y las cuales son diseñadas para emular el proceso por el que el cerebro humano maneja la información (Abba et al., 2020), además se componen de diferentes neuronas como unidades de procesamiento, que están conectadas con pesos y sesgos ajustables. En la Figura 5 se presenta la estructura estándar de una neurona artificial. Las RNA se pueden aplicar de manera exitosa en funciones de aprendizaje, relación, clasificación, generalización, caracterización y optimización. Dado que las RNA pueden trabajar con datos incompletos y tolerar errores, pueden crear fácilmente modelos para problemas complejos (Jorjani et al., 2008).

La RNA empleada en el presente trabajo es la RNA feedforward backpropagation, una de las arquitecturas fundamentales en el campo del machine learning y la inteligencia artificial. Esta estructura permite propagar información desde las entradas hasta las salidas en una sola dirección y utiliza el algoritmo de retro propagación (backpropagation) para ajustar los pesos y minimizar el

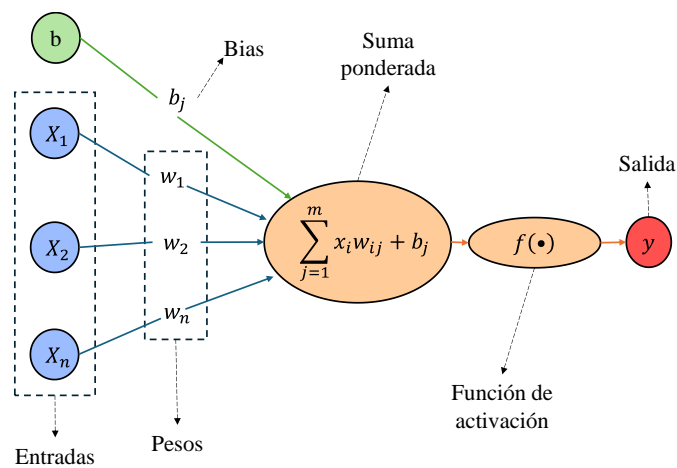
error durante el entrenamiento. La estructura de este tipo de red consta de una capa de entrada, capas ocultas, una capa de salida, neuronas, variable objetivo, funciones de activación y algoritmo de entrenamiento (Abdel Azim, 2020). El hecho de tener una o más capas ocultas permite a la red modelar funciones no lineales y complejas. La expresión matemática que define la entrada neta  $S_i$  a la red neuronal se obtiene mediante la ecuación (10) presentada por Hernández et al., (2022).

$$S_i = \sum_{j=1}^m x_i w_{ij} + b_j, \quad (10)$$

Donde  $S_i$  es la entrada neta al nodo  $j$  de la capa oculta,  $x_i$  son las entradas al nodo  $j$  (o salidas de la capa inmediatamente anterior),  $w_{ij}$  son los pesos sinápticos que representan la fuerza de conexión entre los nodos  $i$  y  $j$ ,  $i$  es la cantidad de nodos y  $b_j$  es el sesgo que está relacionado con cada nodo  $j$ .

### Figura 5

#### *Modelo estándar de una neurona artificial*

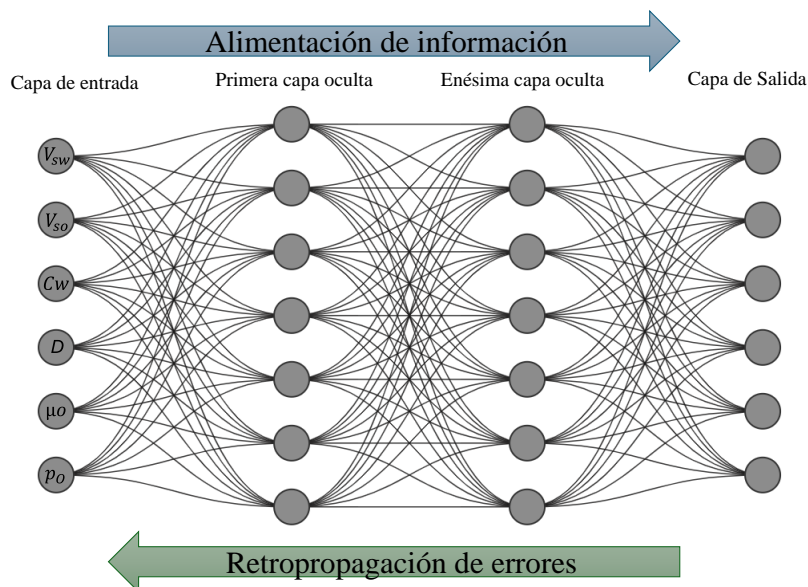


### 2.4.1. Arquitectura de la RNA feedforward backpropagation

Una vez la información ha sido procesada, se procede con la definición de la estructura interna de la RNA feedforward backpropagation. En la Figura 6 se presenta un esquema general de la estructura de este tipo de redes neuronales. Se observa que en la primera capa oculta se define el número de entradas, siendo este un vector con los parámetros de entrada que se usan para generar las predicciones. Posteriormente se define la capa de salida, la cual produce la predicción final del modelo. El número de neuronas de esta capa corresponde a una por cada clase en problemas de clasificación multiclase, para este caso son 6 neuronas, una para cada tipo de patrón de flujo desarrollado al interior de las tuberías horizontales. La selección de la topología específica de la RNA es explicada en la subsección 2.5.1.

**Figura 6**

*Estructura general de una RNA feedforwad backpropagation*



### 2.4.2. Funciones de activación

Las neuronas de la capa oculta utilizan la suma de las entradas con los pesos sinápticos como regla de propagación para aplicar una función de activación, la cual introduce no linealidad a la RNA, permitiendo al modelo aprender y representar relaciones complejas de datos (Goodfellow et al., 2016). Las funciones de activación comúnmente usadas para la aproximación de funciones son la sigmoideal, la tangente hiperbólica y la lineal, siendo la sigmoideal la más utilizada para relaciones no lineales, esta se expresa en la ecuación (11). Para este estudio se utiliza la función de activación sigmoide en las capas ocultas, esta función toma cualquier valor real y lo convierte en un valor acotado en un rango de 0 a 1, siendo útil para predecir una etiqueta de clase binaria, y en la capa de salida se utiliza la función de activación Softmax, la cual se suele utilizar como última función de activación de una red neuronal para transformar las salidas a una representación en forma de probabilidades, acotando los valores en un rango de 0 a 1 en un vector, de tal manera que el sumatorio de todas las probabilidades del vector de como resultado 1 para todos los resultados o clases posibles. Matemáticamente la función de activación sigmoide según (Razavi et al., 2003) , se define como:

$$f(y_j) = \frac{1}{1 + e^{-y_j}} \quad (11)$$

Donde  $y_j$  es la salida del nodo  $j$  y a su vez es el elemento de entrada a los nodos de la capa siguiente.

La función de activación Softmax se define como:

$$S(y)_i = \frac{\exp(y_i)}{\sum_{j=1}^n \exp(y_j)} \quad (12)$$

Donde,  $y$  es un vector de entrada a una función softmax,  $S$ .  $y_i$  el elemento  $i$ -ésimo del vector de entrada, puede tomar cualquier valor entre  $-\infty$  y  $+\infty$ .  $\exp(y_i)$  es la función exponencial estándar aplicada en  $y_i$ . El término  $\sum_{j=1}^n \exp(y_j)$  se refiere a un término de normalización. Este asegura que los valores del vector de salida sumen 1 para la  $i$ -ésima clase y cada uno de ellos este en el rango de 0 a 1, conformando así una distribución de probabilidad válida.

### ***2.4.3. Función de pérdida***

Cabe mencionar que los modelos de aprendizaje automático aprenden mediante una función de pérdida. El cual es un método para determinar cuán eficazmente un algoritmo específico modela los datos proporcionados. Si las predicciones están muy alejadas de los resultados reales, la función de pérdida generará un valor muy alto. Para el problema en cuestión, la función de pérdida a utilizar, la cual es un parámetro importante que indica el rendimiento del modelo de RNA, es la pérdida de entropía cruzada, siendo la más común para problemas de clasificación multiclase.

Esta pérdida de entropía cruzada aumenta a medida que la probabilidad obtenida de la función Softmax diverge de la etiqueta real. La pérdida de entropía cruzada se mide como un número entre 0 y 1, siendo 0 un modelo perfecto y matemáticamente se define como lo indica la ecuación (13):

$$L_{CE} = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^c T_j^{(i)} \log(S_j^{(i)}) \quad (13)$$

Donde  $c$  son las clases,  $n$  es el número de muestras,  $T_j$  es la etiqueta verdadera y  $S_j$  es la probabilidad Softmax para la clase  $j$ -ésima.

Para que la red neuronal aprenda la relación entre los datos, es necesario implementar un algoritmo de entrenamiento. Gradualmente durante dicho entrenamiento del modelo, los pesos sinápticos se ajustan iterativamente con el objetivo de minimizar el error de la predicción. El proceso de ajuste de los pesos sinápticos es lo que define el entrenamiento del modelo (Gómez et al., 2004) y, a medida que el modelo se va entrenando y el error de la función de pérdida se va minimizando, se dice que el modelo está aprendiendo.

## 2.5. Selección y evaluación del modelo

### 2.5.1. Topología de la Red Neuronal Artificial (RNA)

En esta etapa del diseño de la RNA, se define la topología de la red, lo que implica la configuración estructural del modelo, incluyendo el número de capas, la cantidad de neuronas por capa y el algoritmo de entrenamiento. Esta fase es crítica, ya que la topología influye directamente en la capacidad de representación y en la eficacia del aprendizaje del modelo. Es fundamental que la topología se adapte al problema específico a resolver. Debido a la ausencia de técnicas estandarizadas para esta tarea, se recurre a la experiencia y a un enfoque de ensayo y error,

probando diversas configuraciones hasta encontrar la más adecuada para el problema de clasificación del patrón de flujo (Chauvin & Rumelhart, 1995).

### ***2.5.2. Parámetros de optimización***

Para la selección de la topología de las RNA, se lleva a cabo un proceso experimental iterativo y ordenado utilizando la Neural Network Pattern Recognition Toolbox de MATLAB. Diversas configuraciones de RNA se crean y optimizan dentro de la herramienta. En el Apéndice F y en el Apéndice G se detallan el uso y la implementación de esta herramienta a partir del código de MATLAB explicado y un manual de usuario realizado. Es crucial seleccionar un algoritmo de entrenamiento adecuado que optimice el ajuste de los pesos sinápticos para minimizar el valor de la función de pérdida. Por lo tanto, se evalúan las cinco funciones de entrenamiento más utilizadas y recomendadas para problemas de clasificación en MATLAB: TRAINSCG (Scaled Conjugate Gradient), TRAINBFG (BFGS Quasi-Newton), TRAINRP (Resilient Backpropagation), TRAINCGP (Polak-Ribière Conjugate Gradient) y TRAINCGB (Conjugate Gradient with Powell/Beale Restarts).

Además, en la selección del modelo se consideran características clave para evaluar su desempeño, como la precisión o accuracy, el valor de la función de pérdida y el tiempo de ejecución de la RNA. La precisión se obtiene a partir de los datos obtenidos de la matriz de confusión, el valor de la función de pérdida se determina siguiendo la ecuación (13), y el tiempo de ejecución se mide en segundos, indicando cuánto tarda la RNA en ejecutarse.

## 2.6. Matriz de confusión

La matriz de confusión es una herramienta fundamental en el ámbito del machine learning para evaluar el rendimiento de un modelo de clasificación. Permite visualizar las predicciones realizadas por el modelo frente a los valores reales y facilita la identificación de errores específicos (Powers, 2011). Es necesario calcular los componentes de esta, como los verdaderos positivos (TP), falsos negativos (FN), falsos positivos (FP) y verdaderos negativos (TN).

A partir de la matriz de confusión, se pueden extraer varias métricas esenciales para evaluar el rendimiento del modelo como la precisión positiva, el recall o sensibilidad, la puntuación F1 y precisión o accuracy, la precisión positiva es la proporción de verdaderos positivos sobre el total de positivos predichos, el recall es la proporción de verdaderos positivos sobre el total de positivos reales, la puntuación F1 es la media armónica del recall y la precisión positiva, y la precisión o accuracy es la proporción de todas las predicciones correctas, y están definidas respectivamente por las ecuaciones (14), (15), (16) y (17).

$$\textit{Precisión positiva} = \frac{TP}{TP + FP} \quad (14)$$

$$\textit{Recall} = \frac{TP}{TP + FN} \quad (15)$$

$$F1 = 2 * \frac{\textit{Precisión positiva} * \textit{Recall}}{\textit{Precisión positiva} + \textit{Recall}} \quad (16)$$

$$\textit{Precisión} = \frac{TP + TN}{TP + TN + FP + FN} \quad (17)$$

### 3. Resultados

En esta sección se encuentran los valores numéricos obtenidos del proceso iterativo y organizado de prueba y selección de la estructura RNA, esto a partir de todas las simulaciones realizadas bajo diferentes configuraciones topológicas, en las que se registraron todos los datos obtenidos del proceso de simulación, para posteriormente realizar un análisis y selección del modelo que presente los mejores resultados de rendimiento en cuanto a las predicciones de los patrones de flujo. Se estructuraron así diferentes configuraciones variando parámetros de la composición de la red neuronal, como el algoritmo de entrenamiento, el número de capas ocultas y el número de neuronas en cada capa oculta.

Para cada uno de los modelos estructurados se mantuvieron constantes los siguientes parámetros, la distribución de los datos se estableció como un 70% de los datos para la etapa de entrenamiento, un 15% para la etapa de validación y un 15% para la etapa de prueba, la función de activación para todas las capas ocultas fue la función sigmoide, para la capa de salida fue la función de activación Softmax, la función de error a optimizar fue la entropía cruzada y para cada configuración se hicieron 2 reentrenamientos.

Todos los ajustes en los parámetros mencionados se realizaron modificando las líneas de texto de los scripts generados de la herramienta Neural Network Pattern Recognition de MATLAB. Todas las pruebas experimentales se desarrollaron en un computador MSI Raider GE76, procesador 12th Gen Intel(R) Core (TM) i7-12700H 2.70 GHz, RAM instalada de 16 GB y sistema operativo de 64 bits, procesador X64. En la **Tabla 4** se presentan las estructuraciones generadas con diferentes combinaciones entre parámetros para aplicar a los modelos de RNA, evaluándolas para cada una de las funciones de entrenamiento presentadas en la **Tabla 5**.

**Tabla 4**

*Definición de parámetros para las configuraciones estructuradas de la RNA*

Número de capas ocultas	Número de neuronas en capas ocultas	Reéntrenos	Función de activación en capas ocultas	Función de activación en capa de salida	Función de pérdida
<b>3</b>	15	2	Sigmoide	Softmax	Entropía cruzada
	20				
	25				
	30				
<b>5</b>	15	2	Sigmoide	Softmax	Entropía cruzada
	20				
	25				
	30				
<b>7</b>	15	2	Sigmoide	Softmax	Entropía cruzada
	20				
	25				
	30				
<b>10</b>	15	2	Sigmoide	Softmax	Entropía cruzada
	20				
	25				
	30				
<b>15</b>	15	2	Sigmoide	Softmax	Entropía cruzada
	20				
	25				
	30				

**Tabla 5***Algoritmos de entrenamiento a evaluar*

<b>Funciones de entrenamiento</b>	<b>Algoritmos de entrenamiento</b>
<b>TRAINSCG</b>	Scaled conjugate gradient backpropagation
<b>TRAINBFG</b>	BFGS quasi-Newton backpropagation
<b>TRAINRP</b>	Resilient Backpropagation
<b>TRAINCGP</b>	Conjugate gradient backpropagation with Polak-Ribiere updates
<b>TRAINCGB</b>	Conjugate gradient backpropagation with Powell-Beale restarts.

De esta manera se realizaron 200 pruebas de RNA a partir de 100 configuraciones diferentes de parámetros, ver en el Apéndice C. De las cuales se identificaron los modelos que generaron los mejores valores de rendimiento, a partir de la entropía cruzada y los que tuvieron una cantidad superior al 90% en las predicciones correctas totales de patrones de flujo incluyendo la fase de entrenamiento, validación y prueba.

Una vez identificadas aquellas configuraciones, se procedió a repetir las pruebas del modelo de RNA, con el fin de obtener una segunda comprobación de los valores obtenidos y de filtrar las topologías evaluadas a aquellas que mostraran en sus dos reentrenamientos el mayor porcentaje de predicciones totales de patrón de flujo véase en Apéndice G. En la **Tabla 6** se muestran los mejores resultados obtenidos con las diferentes variaciones de topología para cada una de las funciones de entrenamiento, además se muestran las precisiones obtenidas en las diferentes fases de entrenamiento, validación y prueba.

**Tabla 6***Resultados de precisión en los modelos obtenidos bajo diferentes configuraciones de parámetros*

<b>Parámetros</b>	<b>Modelo 1</b>	<b>Modelo 2</b>	<b>Modelo 3</b>	<b>Modelo 4</b>	<b>Modelo 5</b>
<b>Función de entrenamiento</b>	TRAINRP	TRAINSGB	TRAINBFG	TRAINCGP	TRAINCGB
<b>Función de activación en capas ocultas</b>	Sigmoide	Sigmoide	Sigmoide	Sigmoide	Sigmoide
<b>Función de activación en capa de salida</b>	Softmax	Softmax	Softmax	Softmax	Softmax
<b>Función de pérdida</b>	Entropía cruzada categórica	Entropía cruzada categórica	Entropía cruzada categórica	Entropía cruzada categórica	Entropía cruzada categórica
<b>Numero de capas ocultas</b>	3	3	5	5	3
<b>Número de neuronas por capa oculta</b>	30	30	25	30	20
<b>Precisión entrenamiento (%)</b>	95.7	92.7	91.6	92	91.8
<b>Precisión validación (%)</b>	91.7	91	89.9	90.3	89.5
<b>Precisión prueba (%)</b>	92.1	91.3	90.3	90.6	87.7
<b>Precisión total (%)</b>	94.6	92.3	91.2	91.5	90.8
<b>Performance (Entropía cruzada)</b>	0.0272	0.0366	0.0432	0.0406	0.043
<b>Tiempo procesamiento (Segundos)</b>	3	1	133	3	1

Analizando los datos de precisión obtenida por los modelos desarrollados de RNA Feedforward Backpropagation se identificó que el modelo 1, en el cual se hizo uso del algoritmo de entrenamiento Resilient Backpropagation, presentó un 95.7% de precisión en la etapa de entrenamiento, 91.7% en la validación, 92.1% en la prueba y 94.6% de manera global en las tres etapas. Siendo los mayores valores entre los 5 modelos. Además, un valor de entropía cruzada de

0.0272, siendo el menor de todos y un tiempo de procesamiento de 3 segundos. La información obtenida es determinante para escoger la función de entrenamiento Resilient Backpropagation, encargada del proceso de entrenamiento a partir de la optimización del gradiente descendiente y la función de pérdida. Una vez definido el algoritmo de aprendizaje se procede a realizar una última experimentación, manteniendo los parámetros establecidos del modelo 1, excepto el número de neuronas por capa oculta, ya que este se evaluará adicionalmente para 40, 50, 60 y 70 neuronas con el fin de observar el comportamiento y capacidad de predicción del modelo de RNA, véase en Apéndice E.

**Tabla 7**

*Resultados de la variación en el número de neuronas por capa oculta usando el algoritmo de aprendizaje Resilient Backpropagation en el modelo de RNA 1*

Número de neuronas por capa oculta	Reentrenamiento	Error de entropía cruzada	Precisión entrenamiento (%)	Precisión validación (%)	Precisión prueba (%)	Precisión total (%)	Tiempo de compilación (S)
40	1	0.0375	94.1	87.4	90.6	92.6	0
40	2	0.0342	94.9	89.2	85.9	92.7	0
50	1	0.024	97.1	92.8	90.3	95.4	2
50	2	0.0251	96.3	90.3	90.6	94.5	1
60	1	0.4476	22.7	18.4	21.7	21.9	0
60	2	0.5061	12.4	10.8	12.3	12.1	0
70	1	0.6011	21.1	30.7	20.2	22.4	0
70	2	0.5576	19.7	19.1	16.6	19.1	0

La **Tabla 7** presenta los resultados obtenidos al entrenar el modelo de RNA con 40, 50, 60 y 70 neuronas en las capas ocultas. A partir de la información allí recopilada se determinaron los

valores mínimos del error de entropía cruzada, siendo estos 0.0342 y 0.024, para un número de neuronas en las capas ocultas de 40 y 50 respectivamente, así como también los valores de precisión total de 92.7% y 95.4%.

Analizando la información presentada en la **Tabla 7**, se logró determinar que la RNA óptima para desarrollar el modelo predictivo de patrón de flujo para los fluidos agua y aceite en tubería horizontal es aquella que integra el algoritmo de aprendizaje Resilient Backpropagation con 3 capas ocultas y cada capa oculta con 50 neuronas. La configuración de los parámetros del modelo seleccionado es presentada en la **Tabla 8**.

**Tabla 8**

*Configuración de parámetros y resultados del modelo de RNA seleccionado*

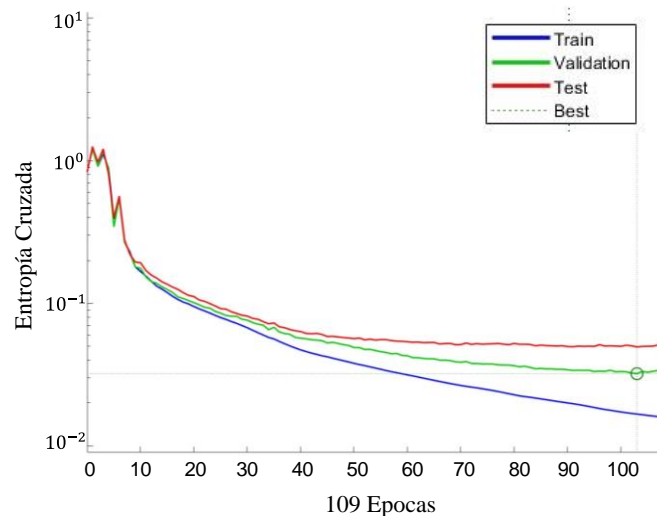
<b>Hiperparámetro</b>	<b>Selección</b>	<b>Otros</b>	<b>Selección</b>
<b>Algoritmo de entrenamiento</b>	Resilient Backpropagation	Precisión entrenamiento (%)	97.1
<b>Función de activación en capas ocultas</b>	Sigmoide	Precisión validación (%)	92.8
<b>Función de activación en capa de salida</b>	Softmax	Precisión prueba (%)	90.3
<b>Función de error</b>	Entropía cruzada categórica	Precisión total (%)	95.4
<b>Numero de capas ocultas</b>	3	Tiempo de compilación (S)	2
<b>Neuronas en capa oculta</b>	50	Error de entropía cruzada	0.0240

Con la configuración de RNA escogida se logró un valor de error de entropía cruzada de 0.024, un valor de precisión de entrenamiento de 97.1%, precisión de validación de 92.8%, precisión de prueba de 90.3% y precisión total de 95.4%, permitiendo concluir que, de manera

comparativa, dicho modelo predictivo es el mejor de los 104 modelados. En la Figura 7 se presenta el comportamiento de la entropía cruzada (crossentropy) en cada una de las etapas de desarrollo del modelo inteligente (entrenamiento, validación y prueba), se presenta el mejor rendimiento de validación en la época 103, con un valor de 0.03213. El entrenamiento termina cuando el error de validación incrementa consecutivamente por seis iteraciones, lo cual sucede en la época 109 y finaliza la etapa de entrenamiento con un valor de entropía cruzada de 0.024.

### Figura 7

*Variación de la entropía cruzada a partir el cambio de iteración en la RNA con la configuración escogida*

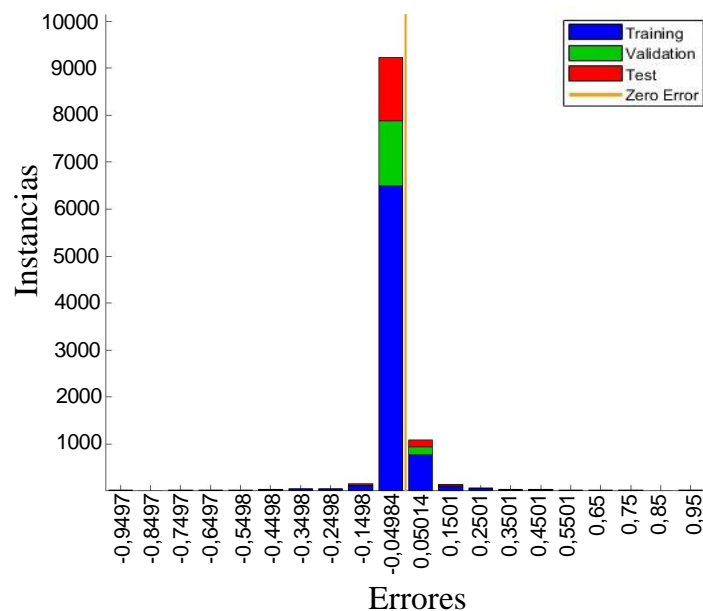


Los datos de rendimiento y precisión del modelo se complementan con un histograma de errores, presentado en la **Figura 8**, donde se muestran los valores de error obtenidos en cada una de las etapas de desarrollo de la RNA, presentando una distribución normal centrada, también se identifica una tendencia central y estrecha con valores de error cercanos a cero, indicando así que

el modelo está bien entrenado y es bastante preciso. La alta frecuencia observada en la columna central, en donde la mayoría de los datos provienen de las etapas de entrenamiento, validación y prueba, sugieren que el modelo está generalizando bien y no está sobreajustado a los datos de entrenamiento.

### Figura 8

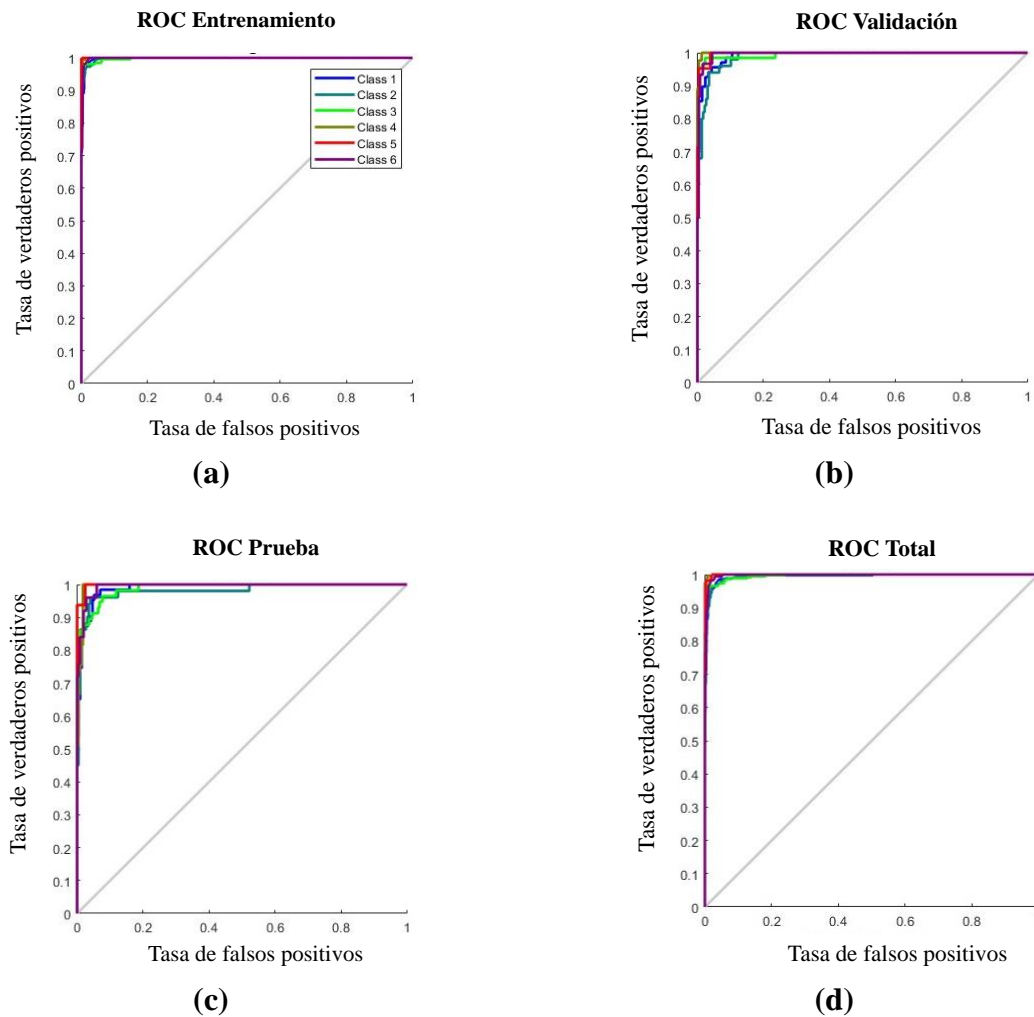
*Histograma de errores en las fases de entrenamiento, validación y prueba para el modelo de RNA.*



Como una verificación adicional del rendimiento de la red, en la Figura 9 se observa la curva ROC para las diferentes etapas de la RNA, la cual es un gráfico de la tasa de verdaderos positivos (recall o sensibilidad) frente a la tasa de falsos positivos (1-especificidad) a medida que varía el umbral, para una prueba perfecta mostraría puntos en la esquina superior izquierda, con un 100% de sensibilidad y un 100% de especificidad. Según el comportamiento de la gráfica, para el problema en cuestión la red tiene un rendimiento muy bueno.

**Figura 9**

Curva ROC en (a) fase de entrenamiento, (b) fase de validación, (c) fase de prueba, (d) Toda la matriz de confusión

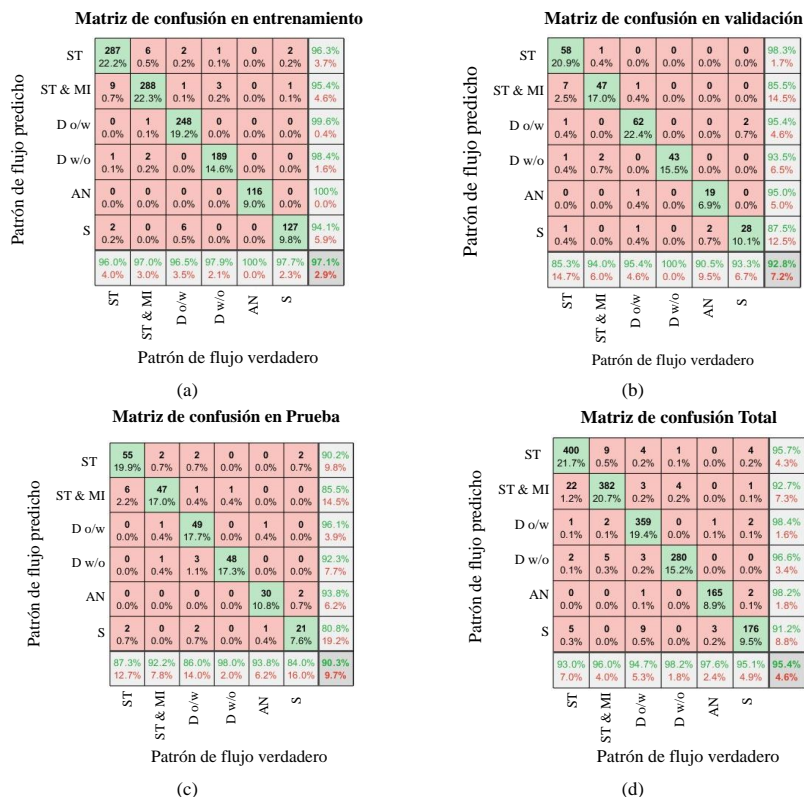


Adicionalmente, para el modelo seleccionado se obtuvo la matriz de confusión presentada en la **Figura 10**, la cual permite analizar el comportamiento de las predicciones para cada uno de los patrones de flujo. A partir de dicha matriz se puede identificar para cada patrón de flujo los verdaderos positivos (TP), es decir, la cantidad de datos que se predijeron correctamente para cada patrón de flujo en específico, los verdaderos negativos (VN), que corresponden a los demás

patrones predichos correctamente, es decir, los otros valores en la diagonal principal, los falsos positivos (FP), que son las predicciones que el modelo cree hacer correctamente, dichos valores son los que corresponden a la fila de patrón de flujo predicho, y por ultimo los falsos negativos (FN), que son aquellos valores presentes en la columna de patrón de flujo verdadero y corresponden a las predicciones incorrectas que realiza el modelo. Asimismo, se identificaron de manera secuencial los resultados de precisión obtenidos en las diferentes fases, en la de entrenamiento el modelo logra una precisión de 97.1%, en la de validación 92.8%, en la de prueba 90.3% y para toda la matriz de confusión total, una precisión de 95.4%

**Figura 10**

*Matriz de confusión para evaluar la capacidad de predicción de patrones de flujo en (a) fase de entrenamiento, (b) fase de validación, (c) fase de prueba, (d) Matriz de confusión total*



En la **Tabla 9**, para cada uno de los patrones de flujo se presentan sus respectivos valores de precisión positiva, recall, F1 y precisión total, calculados según las ecuaciones 14, 15, 16 y 17. Como ejemplo, para la matriz de confusión en la fase de prueba, tomando como referencia el patrón ST, se tiene un valor de TP = 55, al sumar los demás valores de la diagonal principal se obtiene TN = 195, sumando los demás valores de la fila ST se tiene FP = 6, y sumando los demás valores de la columna ST se obtiene FN = 8.

**Tabla 9**

*Métricas para evaluar el rendimiento del modelo en los distintos patrones de flujo a partir de la matriz de confusión*

	Fase de prueba			Precisión total [%]	Fase Total			Precisión total [%]
	Precisión Positiva [%]	Recall [%]	F1 [%]		Precisión Positiva [%]	Recall [%]	F1 [%]	
<b>ST</b>	90.2	87.3	88.7	90,3	95.7	93	94.3	95,4
<b>ST &amp; MI</b>	85.5	92.5	88.7		92.7	96	94.3	
<b>D o/w</b>	96.1	86	90.7		98.4	94.7	96.5	
<b>D w/o</b>	92.3	98	95		96.6	98.2	97.4	
<b>A</b>	93.8	93.8	93.8		98,2	97.6	97.9	
<b>S</b>	80.8	84	82.4		91.2	95.1	93.1	
<b>Promedio</b>	89.8	90.3	89.9		95.5	95.8	95.6	

A partir de la información mostrada en la **Tabla 9**, dentro de la fase de prueba del modelo, se determinó para los diferentes patrones de flujo un valor promedio de precisión positiva del 89.8%, un valor promedio de recall de 90.3% y un valor promedio F1 score de 89.9%. De la matriz de confusión total, los valores promedio de precisión positiva, recall y F1 score fueron 95.5%, 95.8% y 95.6% respectivamente.

#### 4. Conclusiones

Se estructuró una base de datos de flujo bifásico aceite-agua en tubería horizontal a partir de la información reportada en la literatura por diferentes autores, obteniendo un total de 1846 datos experimentales. En la base de datos se incluyeron parámetros relacionados con flujos multifásicos aceite y agua, como las velocidades superficiales de los fluidos aceite y agua, fracción volumétrica del agua, diámetro de la tubería, viscosidad y densidad del aceite. Asimismo, se identificaron y agruparon los patrones de flujo en 6 categorías diferentes: estratificado (ST), estratificado con mezcla (ST & MI), slug (S), anular (AN), dispersión de aceite en agua (Do/w) y dispersión de agua en aceite (Dw/o).

Se desarrolló una Red Neuronal Artificial Feedforward Backpropagation haciendo uso del software MATLAB® y su herramienta Neural Network Pattern Recognition Toolbox. La topología de la red neuronal fue escogida a partir de un proceso experimental iterativo y ordenado en el que se evaluaron 104 configuraciones de RNA, y a partir de la evaluación de parámetros de rendimiento, se determinó que el modelo de red neuronal óptimo para predecir los patrones de flujo es aquel que cuenta con una capa de entrada con seis neuronas, una correspondiente a cada variable de entrada, tres capas ocultas conformadas con 50 neuronas, una capa de salida compuesta por 6 neuronas, un algoritmo de entrenamiento resilient backpropagation, una función de activación sigmoide en las capas ocultas, la función de activación softmax en la capa de salida del modelo y la entropía cruzada como función de pérdida.

La RNA desarrollada para la predicción de patrones de flujo presentó una precisión de entrenamiento de 97.1%, precisión de validación de 92.8%, precisión de prueba de 90.3% y precisión total de 95.4%. Se observó que el modelo predijo con éxito 6 de los 6 patrones de flujo en las diferentes etapas de desarrollo del modelo, lo cual indica que el modelo de RNA entrenado es capaz de predecir los diferentes patrones de flujo con una alta precisión. Es de mencionar que el costo computacional requerido para el modelamiento de la RNA escogida es bajo, ya que el tiempo de entrenamiento del modelo fue de 2 segundos.

La predicción de patrones de flujo obtenida por el modelo de RNA implementado muestra ser una herramienta prometedora y podría desarrollarse para cubrir más datos experimentales ampliando la base de datos, en ese orden de ideas se desarrolló un código de Matlab y un manual de usuario, disponible en el Apéndice F y Apéndice G, respectivamente, con el fin de explicar el código y la implementación de la herramienta de MATLAB de manera clara e interactiva con el fin de que cualquier persona pueda hacer uso de la herramienta para realizar posteriores modelos de clasificación.

En este trabajo se demostró que, a partir de las variables seleccionadas del fluido se pueden desarrollar modelos de machine learning basados en Redes Neuronales Artificiales Feedforward Backpropagation para determinar patrones de flujo bifásico aceite-agua en tuberías horizontales con una alta precisión y un bajo error de entropía cruzada.

Finalmente, se resaltan sus posibles aplicaciones en la industria oil & gas. La precisión obtenida del modelo RNA para reconocer patrones de flujo sugiere oportunidades para optimizar

el proceso diseño y mantenimiento de las tuberías a partir de la estimación de parámetros importantes del proceso, como el gradiente de presión y las fracciones volumétricas del sistema de flujo, también sugiere la optimización de la inyección de inhibidores de corrosión y otros químicos, mejorando así la gestión de mantenimiento y la integridad de las instalaciones.

## **5. Recomendaciones**

Trabajos futuros incluyen el estudio del comportamiento de la Red Neuronal Artificial seleccionada frente a datos experimentales suministrados por bancos de pruebas en laboratorios de mecánica de fluidos y datos obtenidos en operaciones reales de transporte de flujos bifásicos en tuberías horizontales en empresas del sector oil & gas. Además, se sugiere que en futuros trabajos se utilice una base de datos más extensa y equilibrada en cuanto al número de datos para cada uno de los patrones de flujo analizados. Esto ayudaría a reducir un posible riesgo de sesgo en las predicciones, evitando que el modelo favorezca el patrón de flujo con mayor cantidad de datos en el proceso de entrenamiento.

**Referencias Bibliográficas**

- Abba, S. I., Usman, A. G., & IŞIK, S. (2020). Simulation for response surface in the HPLC optimization method development using artificial intelligence models: A data-driven approach. *Chemometrics and Intelligent Laboratory Systems*, 201(November 2019). <https://doi.org/10.1016/j.chemolab.2020.104007>
- Abdel Azim, R. (2020). Prediction of multiphase flow rate for artificially flowing wells using rigorous artificial neural network technique. *Flow Measurement and Instrumentation*, 76(September), 101835. <https://doi.org/10.1016/j.flowmeasinst.2020.101835>
- Abduvayt, P., Manabe, R., Watanabe, T., & Arihara, N. (2004). Analisis of oil-water flow tests in horizontal, hilly-terrain, and vertical pipes. *Proceedings - SPE Annual Technical Conference and Exhibition*, 1335–1347. <https://doi.org/10.2523/90096-ms>
- Al-Naser, M., Elshafei, M., & Al-Sarkhi, A. (2016). Artificial neural network application for multiphase flow patterns detection: A new approach. *Journal of Petroleum Science and Engineering*, 145, 548–564. <https://doi.org/10.1016/j.petrol.2016.06.029>
- Al-Sarkhi, A., Pereyra, E., Mantilla, I., & Avila, C. (2017). Dimensionless oil-water stratified to non-stratified flow pattern transition. *Journal of Petroleum Science and Engineering*, 151, 284–291. <https://doi.org/10.1016/j.petrol.2017.01.016>

Al-Wahaibi, T., Al-Wahaibi, Y., Al-Ajmi, A., Al-Hajri, R., Yusuf, N., Olawale, A. S., & Mohammed, I. A. (2014). Experimental investigation on flow patterns and pressure gradient through two pipe diameters in horizontal oil-water flows. *Journal of Petroleum Science and Engineering*, 122, 266–273. <https://doi.org/10.1016/j.petrol.2014.07.019>

Alhashem, M. (2020). Machine learning classification model for multiphase flow regimes in horizontal pipes. *International Petroleum Technology Conference 2020, IPTC 2020*. <https://doi.org/10.2523/iptc-20058-abstract>

Bahrami, B., Mohsenpour, S., Shamshiri Noghabi, H. R., Hemmati, N., & Tabzar, A. (2019). Estimation of flow rates of individual phases in an oil-gas-water multiphase flow system using neural network approach and pressure signal analysis. *Flow Measurement and Instrumentation*, 66(December 2018), 28–36. <https://doi.org/10.1016/j.flowmeasinst.2019.01.018>

Cai, J., Li, C., Tang, X., Ayello, F., Richter, S., & Nesic, S. (2012). Experimental study of water wetting in oil-water two phase flow-Horizontal flow of model oil. *Chemical Engineering Science*, 73, 334–344. <https://doi.org/10.1016/j.ces.2012.01.014>

Chauvin, Y., & Rumelhart, D. E. (1995). *Backpropagation: Theory, architectures, and applications*. Lawrence Erlbaum Associates, Inc.

Chimeno-Trinchet, C., Murru, C., Díaz-García, M. E., Fernández-González, A., & Badía-Laiño,

R. (2020). Artificial Intelligence and fourier-transform infrared spectroscopy for evaluating water-mediated degradation of lubricant oils. *Talanta*, 219(June). <https://doi.org/10.1016/j.talanta.2020.121312>

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences Second Edition* (2nd ed). Lawrence Erlbaum Associates.

Dasari, A., Desamala, A. B., Dasmahapatra, A. K., & Mandal, T. K. (2013). Experimental studies and probabilistic neural network prediction on flow pattern of viscous oil-water flow through a circular horizontal pipe. *Industrial and Engineering Chemistry Research*, 52(23), 7975–7985. <https://doi.org/10.1021/ie301430m>

El-Sebakhy, E. A. (2010). Flow regimes identification and liquid-holdup prediction in horizontal multiphase flow based on neuro-fuzzy inference systems. *Mathematics and Computers in Simulation*, 80(9), 1854–1866. <https://doi.org/10.1016/j.matcom.2010.01.002>

Figueiredo, M. M. F., Goncalves, J. L., Nakashima, A. M. V., Fileti, A. M. F., & Carvalho, R. D. M. (2016). The use of an ultrasonic technique and neural networks for identification of the flow pattern and measurement of the gas volume fraction in multiphase flows. *Experimental Thermal and Fluid Science*, 70, 29–50. <https://doi.org/10.1016/j.expthermflusci.2015.08.010>

Gómez, G., Henao, J., & Salazar, H. (2004). Entrenamiento De Una Red Neuronal Artificial Usando El Algoritmo Simulated Annealing. *Entrenamiento De Una Red Neuronal Artificial*

*Usando El Algoritmo Simulated Annealing*, 1(24), 13–18.  
<https://dialnet.unirioja.es/descarga/articulo/4844874.pdf>

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. The MIT Press.

Grassi, B., Strazza, D., & Poesio, P. (2008). Experimental validation of theoretical models in two-phase high-viscosity ratio liquid-liquid flows in horizontal and slightly inclined pipes. *International Journal of Multiphase Flow*, 34(10), 950–965.  
<https://doi.org/10.1016/j.ijmultiphaseflow.2008.03.006>

Guyon, I., & De, A. M. (2003). An Introduction to Variable and Feature Selection André Elisseeff. In *Journal of Machine Learning Research* (Vol. 3)

Hernández-Cely, M. M., & Ruiz-Díaz, C. M. (2020). Estudio de los fluidos aceite-agua a través del sensor basado en la permitividad eléctrica del patrón de fluido. *Revista UIS Ingenierías*, 19(3), 177–186. <https://doi.org/10.18273/revuin.v19n3-2020017>

Hernández-Cely, M. M., Ruiz-Díaz, C. M., & González-Estrada, O. A. (2022). Modelo predictivo para el cálculo de la fracción volumétrica de un flujo bifásico agua-aceite en la horizontal utilizando una red neuronal artificial. *Revista UIS Ingenierías*, 21(2), 155–164.  
<https://doi.org/10.18273/revuin.v21n2-2022013>

Ibarra, R., Zadrazil, I., Markides, C. N., & Matar, O. K. (2015). Towards a Universal Dimensionless Map of Flow Regime. *11th International Conference on Heat Transfer, Fluid*

*Mechanics and Thermodynamics, Kruger National Park, 1-6, July.*

[https://www.researchgate.net/publication/283498693\\_Towards\\_a\\_Universal\\_Dimensionless\\_Map\\_of\\_Flow\\_Regime\\_Transitions\\_in\\_Horizontal\\_Liquid-Liquid\\_Flows](https://www.researchgate.net/publication/283498693_Towards_a_Universal_Dimensionless_Map_of_Flow_Regime_Transitions_in_Horizontal_Liquid-Liquid_Flows)

Jorjani, E., Chehreh Chelgani, S., & Mesroghli, S. (2008). Application of artificial neural networks to predict chemical desulfurization of Tabas coal. *Fuel*, 87(12), 2727–2734. <https://doi.org/10.1016/j.fuel.2008.01.029>

Liu, X., Tang, H., Ding, Y., & Yan, D. (2022). Investigating the performance of machine learning models combined with different feature selection methods to estimate the energy consumption of buildings. *Energy and Buildings*, 273. <https://doi.org/10.1016/j.enbuild.2022.112408>

Lum, J. Y. L., Al-Wahaibi, T., & Angeli, P. (2006). Upward and downward inclination oil–water flows. *International Journal of Multiphase Flow*, 32(4), 413–435. <https://doi.org/10.1016/J.IJMULTIPHASEFLOW.2006.01.001>

Montoya, G., Garcia, K., Valencillos, M., Garcia, J., Romero, C., & Gonzalez-Mendizabal, D. (2009). Determinación de altura de fase y hold up para flujo bifásico líquido-líquido en tuberías horizontales por medio de procesamiento de imágenes. *Conference: American Society of Mechanical Engineers (ASME) Congress “Ideas Practicas...Soluciones Eficientes,”* 1–10. [https://www.researchgate.net/publication/260182534\\_Determinacion\\_de\\_altura\\_de\\_fase\\_y\\_](https://www.researchgate.net/publication/260182534_Determinacion_de_altura_de_fase_y_)

hold\_up\_para\_flujo\_bifasico\_liquido

liquido\_en\_tuberias\_horizontales\_por\_medio\_de\_procesamiento\_de\_imagenes

Nädler, M., & Mewes, D. (1997). Flow induced emulsification in the flow of two immiscible liquids in horizontal pipes. *International Journal of Multiphase Flow*, 23(1), 55–68. [https://doi.org/10.1016/s0301-9322\(96\)00055-9](https://doi.org/10.1016/s0301-9322(96)00055-9)

Osundare, O. S., Falcone, G., Lao, L., & Elliott, A. (2020). Liquid-liquid flow pattern prediction using relevant dimensionless parameter groups. *Energies*, 13(17). <https://doi.org/10.3390/en13174355>

Powers, D. M. W. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 37–63. <https://doi.org/10.48550/arXiv.2010.16061>

Razavi, M. A., Mortazavi, A., & Mousavi, M. (2003). Dynamic modelling of milk ultrafiltration by artificial neural network. *Journal of Membrane Science*, 220(1–2), 47–58. [https://doi.org/10.1016/S0376-7388\(03\)00211-4](https://doi.org/10.1016/S0376-7388(03)00211-4)

Rodriguez, O. M. H., & Oliemans, R. V. A. (2006). Experimental study on oil-water flow in horizontal and slightly inclined pipes. *International Journal of Multiphase Flow*, 32(3), 323–343. <https://doi.org/10.1016/j.ijmultiphaseflow.2005.11.001>

- Roshani, G. H., Feghhi, S. A. H., Mahmoudi-Aznaveh, A., Nazemi, E., & Adineh-Vand, A. (2014). Precise volume fraction prediction in oil-water-gas multiphase flows by means of gamma-ray attenuation and artificial neural networks using one detector. *Measurement: Journal of the International Measurement Confederation*, 51(1), 34–41. <https://doi.org/10.1016/j.measurement.2014.01.030>
- Roshani, M., Phan, G. T. T., Jammal Muhammad Ali, P., Hossein Roshani, G., Hanus, R., Duong, T., Corniani, E., Nazemi, E., & Mostafa Kalmoun, E. (2021). Evaluation of flow pattern recognition and void fraction measurement in two phase flow independent of oil pipeline's scale layer thickness. *Alexandria Engineering Journal*, 60(1), 1955–1966. <https://doi.org/10.1016/J.AEJ.2020.11.043>
- Ruiz-Díaz, C. M., Perilla-Plata, E. E., & González-Estrada, O. A. (2024). Two-Phase Flow Pattern Identification in Vertical Pipes Using Transformer Neural Networks. *Inventions*, 9(1). <https://doi.org/10.3390/inventions9010015>
- Ruiz-Díaz, C. M., Quispe-Suarez, B., & González-Estrada, O. A. (2024). Two-phase oil and water flow pattern identification in vertical pipes applying long short-term memory networks. *Emergent Materials*, 0123456789. <https://doi.org/10.1007/s42247-024-00631-2>
- Salgado, C. M., Pereira, C. M. N. A., Schirru, R., & Brandão, L. E. B. (2010). Flow regime identification and volume fraction prediction in multiphase flows by means of gamma-ray attenuation and artificial neural networks. *Progress in Nuclear Energy*, 52(6), 555–562.

<https://doi.org/10.1016/j.pnucene.2010.02.001>

Shi, J., Lao, L., & Yeung, H. (2017). Water-lubricated transport of high-viscosity oil in horizontal pipes: The water holdup and pressure gradient. *International Journal of Multiphase Flow*, *96*, 70–85. <https://doi.org/10.1016/j.ijmultiphaseflow.2017.07.005>

Shi, J., & Yeung, H. (2017). Characterization of liquid-liquid flows in horizontal pipes. *AIChE Journal*, *63*(3), 1132–1143. <https://doi.org/10.1002/aic.15452>

Shoham, O. (2005). *Mechanistic Modeling of Gas-Liquid Two-Phase Flow in Pipes* (1st ed.). Society of Petroleum Engineers.

Soot, P. M. (1970). *A Study of Two-Phase Liquid-Liquid Flow in Pipes* (Issue 1) [Thesis in partial fulfillment of the requirements for the degree of Doctor of Philosophy, Oregon State University]. <http://hdl.handle.net/1957/18932>

Xie, J., Sage, M., & Zhao, Y. F. (2023). Feature selection and feature learning in machine learning applications for gas turbines: A review. *Engineering Applications of Artificial Intelligence*, *117*, 105591. <https://doi.org/10.1016/J.ENGAPPAL.2022.105591>

Yan, Y., Wang, L., Wang, T., Wang, X., Hu, Y., & Duan, Q. (2018). Application of soft computing techniques to multiphase flow measurement: A review. *Flow Measurement and Instrumentation*, *60*, 30–43. <https://doi.org/10.1016/j.flowmeasinst.2018.02.017>

Zhao, T., Zheng, Y., & Wu, Z. (2023). Feature selection-based machine learning modeling for distributed model predictive control of nonlinear processes. *Computers & Chemical Engineering*, *169*, 108074. <https://doi.org/10.1016/j.compchemeng.2022.108074>