

# Metodología basada en aprendizaje automático y procesamiento digital de imágenes para la estimación de turbidez en el proceso de potabilización de agua.

## *Methodology based on machine learning and digital image processing for estimating turbidity in the water purification process.*

Andrés Eduardo De Paz Martínez

Instituto Tecnológico de Tuxtla Gutiérrez. Posgrado en  
Ciencias de la Ingeniería Mecatrónica.  
Tuxtla Gutiérrez Chiapas, México.

Héctor Ricardo Hernández de León

Instituto Tecnológico de Tuxtla Gutiérrez. Posgrado en  
Ciencias de la Ingeniería Mecatrónica.  
Tuxtla Gutiérrez Chiapas, México.

José Armando Fragoso Mandujano

Instituto Tecnológico de Tuxtla Gutiérrez. Posgrado en  
Ciencias de la Ingeniería Mecatrónica.  
Tuxtla Gutiérrez Chiapas, México.

Felisa Vanessa López Pineda

Instituto Tecnológico de Tuxtla Gutiérrez. Posgrado en  
Ciencias de la Ingeniería Mecatrónica.  
Tuxtla Gutiérrez Chiapas, México.

**Resumo** — La turbidimetría se centra en el análisis de la medición de la turbidez del agua, es decir, la cantidad de partículas suspendidas en el medio. Este método se basa en el estudio de cómo las partículas dispersan la luz, lo que resulta en una disminución en la intensidad de la luz transmitida a través de una muestra. Para el estudio de la turbidez, se suelen utilizar métodos ópticos y mediciones químicas, estos métodos requieren personal y equipo especializado, además de que los equipos son costosos. Para abordar esta problemática, una alternativa es el uso de algoritmos de aprendizaje automático que permiten realizar una clasificación y estudio a menor costo, pero con el reto adicional de la interpretación de imágenes. Motivado por lo anterior, se propone la implementación y evaluación de modelos inteligentes que realice la predicción de niveles de turbidez.

**Palabras Clave** - Turbidez; visión por computadora; Machine Learning.

**Abstract** — Turbidimetry focuses on the analysis of measuring the turbidity of water, that is, the amount of particles suspended in the medium. This method is based on the study of how particles scatter light, resulting in a decrease in the intensity of light transmitted through a sample. To study turbidity, optical methods and chemical measurements are usually used; these methods require specialized personnel and equipment, and the equipment is expensive. To address this problem, an alternative is the use of machine learning algorithms that allow classification and study to be carried out at a lower cost, but with the additional challenge of image interpretation. Motivated by the above, the implementation and evaluation of intelligent models that predict turbidity levels is proposed.

**Keywords** - Turbidity; computer vision; Machine Learning.

### I. INTRODUCCIÓN

Los procesos de potabilización son llevados a cabo dentro de plantas potabilizadoras especializadas. Dentro de los procesos de potabilización, se realiza la medición de parámetros físicos, químicos y biológicos. En México la NOM 127 SSA1 2021 [1], indica la importancia de la medición y estandarización de los parámetros. Los parámetros principalmente medidos son pH, Total de Sólidos Disueltos (TDS), temperatura, Dureza, color verdadero, cloro y el parámetro mas importante en la medición, turbidez [2].

La turbidez del agua es un parámetro relevante e importante en la definición de los procesos de potabilización y tratamiento de agua. Este parámetro se utiliza como indicador del total de partículas que se encuentran suspendidas en el agua [3]. El realizar un estudio de la turbidez del agua, brinda beneficios directos para la protección y conservación del medio ambiente [4]. Se puede encontrar información valiosa sobre los contaminantes, microorganismos dañinos y sustancias tóxicas al analizar este parámetro [5].

Dentro del estudio de los procesos de potabilización del agua se ha indicado la importancia de la medición de este parámetro para la automatización de los procesos robustos de medición [6]. Gracias al monitoreo y estudio de la turbidez del agua, se han realizado investigaciones para la determinación de la calidad del agua, lo que beneficia en gran manera al suministro del recurso [7], a la monitorización de las fuentes de

agua natural [8], a los estudios de erosión de la tierra [9], y conservación de la vida acuática [10].

Existen distintos métodos para la determinación y análisis de la turbidez en el agua. Dentro de estas metodologías se abarcan los estudios realizados a nivel laboratorio, para la determinación de cantidad de coagulante floculante que se necesite en la desinfección de agua turbia [11]. Estos procedimientos se llevan a cabo en la mayoría de los casos en plantas de tratamiento de aguas residuales o plantar potabilizadoras, en donde se busca obtener una solución de coagulante floculante para agregar a una muestra de agua turbia [12]. Estos estudios son utilizados principalmente para la realización de prueba de jarras, dicho procesamiento es llevado en un laboratorio con equipo especializado, robusto y costoso, además que el tiempo de espera para la verificación de turbidez es extenso [13].

Una alternativa para la determinación de turbidez en el agua es el uso de mediciones realizadas con sensores. Estas técnicas de medición cubren la incertidumbre en la pérdida de datos, generalmente se debe ajustar un tiempo de muestreo y frecuencia del sensor [14]. Los sensores de turbidez utilizan el principio físico y óptico de la refracción y reflexión de la luz que atraviesa en un medio, ya que para ello se hace estimación de la cantidad de partículas de una muestra. Se debe de realizar un estudio exhaustivo del rendimiento de los sensores que se utilice, además de tener experiencia en la calibración de sensores, ya que la lectura no siempre es precisa y se obtiene pérdida de datos con el tiempo de muestreo [15].

La pérdida de datos indica una problemática severa en el análisis de la turbidez del agua, ya que se pueden utilizar los datos perdidos para introducir medidas de prevención ante el consumo y distribución del agua. El análisis de turbidez en el agua es el asistido por expertos, ya que se pueden llevar a cabo con metodologías como el uso de Escala McFarland, métodos visuales con referencias y uso de disco Secchi [16] [17]. La escala de McFarland se basa en la comparación visual de la turbidez de una suspensión de células bacterianas con la de una serie de soluciones de sulfato de bario.

Dentro del método Secchi se realiza una estimación visual con el nivel de profundidad con la que el disco de Secchi desaparece [18]. Estas metodologías son muy técnicas y llevadas a cabo por especialistas, por el nivel de complejidad que presentan, además de no gozar de gran forma de exactitud en la concentración de turbidez, por lo que contar con estimaciones de turbidez imprecisas afecta los procesos de potabilización y purificación del recurso natural.

Otros enfoques utilizados en el análisis de turbidez son los orientados en el procesamiento de imágenes para el análisis de las características asociadas a la predicción del recurso [19] [20]. Estos trabajos están basados en el tratamiento y extracción de características morfológicas de imágenes.

La implementación de este tipo análisis implica el uso de Machine Learning, para el entrenamiento de modelos inteligentes, realizando un aprendizaje con una cantidad suficiente de datos, estos datos pueden ser imágenes o características extraídas de las imágenes,

Sin embargo, se requieren múltiples estrategias de entramiento y análisis de software para la determinación de la arquitectura y el procesamiento de los datos. Esto puede ser una limitante importante en ciertos casos, ya que debe evaluar la combinación necesaria para el entrenamiento de un modelo inteligente, en que software y lenguaje se desarrollará, así como el tiempo que se tome para el procesamiento de los datos.

Motivado por los resultados de los algoritmos de aprendizaje automático y los beneficios que la implementación de estos modelos puede generar, este trabajo se centra en el análisis de datos para predecir los niveles de turbidez. En este trabajo se propone una metodología de bajo coste computacional para la extracción de características (componentes) de muestras de agua con diversos niveles de turbidez. Además, se propone una arquitectura de procesamiento de imágenes diseñada para obtener los mejores resultados con los datos disponibles.

El enfoque propuesto no solo busca mejorar la precisión de las predicciones, sino también reducir el tiempo de procesamiento y los costos asociados a la implementación de modelos complejos. Esta metodología resalta como una herramienta valiosa para el monitoreo y gestión de la calidad del agua en diversas aplicaciones, proporcionando información crucial para la toma de decisiones en tiempo real.

En la mayoría de los trabajos citados se proponen metodologías robustas y de análisis complejo para la obtención de los datos. Por lo que, se propone una metodología novedosa para la predicción y cuantificación de niveles de turbidez. Los datos obtenidos son validados con las mediciones realizadas a nivel laboratorio, de tal forma de corroborar la veracidad de los datos con lo medido de forma real. Los resultados obtenidos son interpretados y evaluados con diferentes métricas estadísticas para la predicción de turbidez.

## II. TRABAJOS RELACIONADOS

En esta sección se proporciona una revisión del estado del arte sobre la definición de los procedimientos computacionales a partir del procesamiento de imágenes. Por lo que implementar técnicas aprendizaje automático para el procesamiento de los datos obtenidos brinda un acercamiento acertado en el reconocimiento y análisis de variables características de las muestras de agua. Se describen técnicas utilizadas en las estimaciones de turbidez, además de algoritmos de aprendizaje profundo para el procesamiento de imágenes, así como la implementación de sistemas para la obtención de imágenes en la determinación de la turbidez del agua.

Dentro del análisis computacional para la estimación de turbidez, se llevan a cabo diferentes técnicas y operaciones de procesamiento de imágenes. La metodología de estos trabajos [21] [22] [23] está enfocada en el reconocimiento e identificación de los niveles de correlación de las características evidentes de las imágenes de las muestras de agua.

La metodología analiza el comportamiento de la coloración del agua [21] tomando como enfoque principal el análisis los niveles de los umbrales de las imágenes, ya que cada nivel de turbidez está asociado directamente con los cambios de

coloración [22], este análisis contempla la influencia de parámetros físicos que afectan directamente la turbidez del agua.

Los procesos necesarios en la binarización de una imagen se logran a través del análisis de distintos patrones de imágenes contenidas a procesar, como el color, la textura, la forma en que se toma la imagen, la iluminación y bordes [23]. Esta interpretación se lleva a cabo por la conversión de RGB a escala de grises, y un proceso de filtrado de la imagen el cual lleva a cabo la umbralización de la imagen convertida a escala de grises [24].

Utilizando los métodos de procesamiento de imágenes a través de conversión de escala de grises, es posible obtener cálculos acertados a la determinación de la turbidez en el agua. Este análisis es posible al manejo adecuado del ángulo de la fotografía, así como tener cuantificado de forma adecuado el nivel turbidez de las muestras para la apreciación de las imágenes [25]; ya que suelen existir anomalías, como ruido de interpretación, limitando y obligando a estos procedimientos, a llevar a cabo diferentes procesos de filtrado para su mejora.

Los modelos basados en inteligencia artificial cuentan con la capacidad de procesar diferentes volúmenes de datos, características y objetos computacionales, es por ello que estas herramientas claves para el análisis de la turbidez en el agua. El estudio llevado a cabo dentro de este campo de investigación se basa en la implementación de modelos de red neuronal artificial (RNA) y modelos de aprendizaje automático, así como profundo [26]. Para el procesamiento y análisis de imágenes se sigue la metodología de extracción de características, análisis de colorimetría y generación de Data Set de los datos obtenidos que indiquen valores de turbidez [26].

Estos enfoques presentan el desarrollo y la implementación de modelos inteligentes, como modelos bayesianos y modelo de red de regresión lineal múltiple (MLR). Dentro de la implementación de estos modelos se deben realizar ajustes de los hiperparámetros, adecuando la arquitectura para el procesamiento de los datos, el modelo MLR busca aproximar valores calculados nuevo y desconocidos mediante la creación de una ecuación lineal [27].

En este trabajo, el modelo de red es validado con un conjunto de datos con un margen de nueve clases y un aproximado de 100 imágenes contenidas en cada una, cabe mencionar que, las imágenes con las que se entrenó la red son imágenes sintéticas, por lo que, al realizar el estudio con muestras de agua de procedencia real, el rendimiento se ve afectado de manera gradual [28]. Realizar entrenamiento de forma sintética limita de cierta manera los resultados y los márgenes de error, para ello se debe llevar a cabo el experimento con diferentes muestras de agua para mejorar el nivel de eficiencia [28].

Las técnicas abordadas en la obtención de datos e imágenes son un factor clave en la determinación de turbidez [29]. Para ello se describe el uso de cámaras digitales para capturar imágenes del agua y algoritmos de procesamiento de imágenes para analizar la cantidad y distribución de partículas en suspensión [30]. Los estudios presentados se centran en la

detección de bordes en las imágenes para determinar los niveles de turbidez [31]. Las imágenes de agua turbia tienen bordes difusos y poco definidos, mientras que las imágenes de agua clara tienen bordes nítidos y definidos, lo que permite la definición de bordes específicos además de captar el nivel de difuminación que las imágenes pueden tener con un nivel de referencia que sirva como guía, el cual tiene relación directa con los niveles de turbidez que el agua contiene [31].

En este trabajo se resalta la importancia de cuantificar y analizar los niveles de turbidez del agua, ya que, dentro de la región y de acuerdo con los procesos de potabilización se indica la necesidad del análisis del parámetro de turbidez. En nuestra investigación previa [32] se desarrolló una herramienta de seguimiento y diagnóstico de los procesos de una planta potabilizadora. Se basa en la aplicación de una técnica de clasificación e interpretación de la información obtenida de todos los procesos con el fin de identificar averías y ayudar al mantenimiento preventivo/predictivo de las diferentes unidades de la planta de SMAPA en la localidad de Tuxtla Gutiérrez, Chiapas, México. Aunque este trabajo utilizó análisis de Machine Learning no se abordaron técnicas de procesamiento de imágenes en la determinación de la turbidez del agua.

A diferencia de trabajos previos estamos interesados en el análisis de la turbidez del agua a partir de Machine Learning sin un post procesamiento estadístico o de visión por computadora clásico. Esto permite tener una metodología robusta al desarrollar el procedimiento de procesamiento de los datos obtenidos de las imágenes. La investigación se centra en el tratamiento de imágenes de muestras agua, las cuales indiquen directamente los niveles de turbidez de las muestras. Esta metodología implica la elaboración de un banco de imágenes con los distintos niveles de turbidez con las que el agua cuente a través del uso de un banco de pruebas. Estas imágenes serán insertadas dentro de un modelo de aprendizaje máquina que sea capaz de determinar los niveles de turbidez además de evaluar diferentes métricas estadísticas para.

### III. MATERIALES Y MÉTODOS

En esta sección se presenta la metodología propuesta para la predicción de turbidez con modelos de aprendizaje automático. La metodología está basada en la aplicación de dos campos del manejo inteligente de datos. El estudio se enfoca en la recolección de muestras, interpretación y etiquetado. La segunda fase de la metodología describe la captura de imágenes de las muestras recolectadas haciendo uso de un banco de pruebas especialmente elaborado, para la extracción de características además de evaluar las correlaciones de los datos extraídos con los datos de turbidez real. Esta fase también incluye la implementación de algoritmos de extracción y de interpolación. La tercera fase de la metodología describe el uso de modelos de aprendizaje automático, capaces de procesar los datos y realizar predicciones del parámetro de turbidez.

En la Fig. 1 se describe un esquema general de la metodología aplicada.

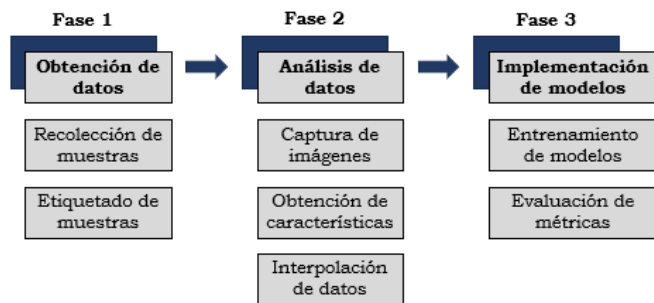


Figure 1. Metodología general aplicada.

### Fase 1. Obtención de datos

La fase de obtención de datos cuenta con dos etapas, la primera radica en la recolección de las muestras que se analizaron, ya que el estudio se enfoca al análisis de agua residual y la segunda, el etiquetado de muestras.

#### 1) Recolección de muestras.

La recolección de muestras se llevó a cabo estratégicamente en zonas de afección presentes del río Sabinal de la ciudad de Tuxtla Gutiérrez Chiapas, México. El estudio se llevó a cabo en cuatro puntos estratégicos de la ciudad, ya que de esta forma se segmentan diferentes niveles de turbidez en las muestras. En la Fig. 2 se ejemplifican los puntos de recolección.



Figure 2. Zonas de recolección de muestras.

Las coordenadas de la ubicación geográfica se describen en la Tabla 1.

TABLE I. COORDENADAS DE UBICACIÓN GEOGRÁFICA

Ubicación	Coordenadas
Norte poniente	16.760031, -93.140738
Sur poniente	16.752685, -93.178051
Norte oriente	16.759647, -93.088098
Sur oriente	16.734484, -93.087489

#### 2) Etiquetado de muestras.

El estudio se realizó con muestras recolectadas y almacenadas en tubos Falcon de 45 ml/L, estas muestras cuentan con la característica de tener diferentes niveles de turbidez, los cuales fueron medidos y etiquetados de forma empírica y en laboratorio. Esto permite analizar los datos con técnicas de procesamiento de imágenes los pixeles de las capturas obtenidas, en busca de características nítidas y correlaciones positivas para la predicción de turbidez, como se aprecia en la Fig. 3.

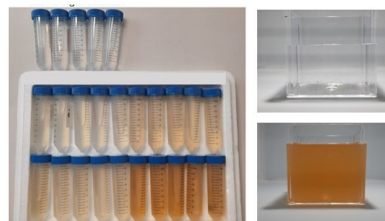


Figure 3. Muestras etiquetadas y medidas en laboratorio.

### Fase 2. Análisis de datos

La obtención de las imágenes se llevó a cabo con la implementación de un banco de pruebas que contempla aspectos controlados para la iluminación, colocación de la muestra, recipiente para muestras, espacio de manipulación y un módulo de cámara calibrado, que permitan un correcto análisis y procesamiento de las imágenes, comenzando por la fase de diseño y posterior la implementación de este, en la Fig. 4 se aprecia el diseño.

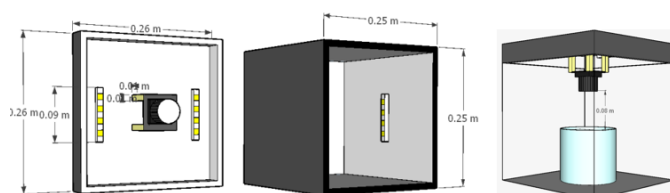


Figure 4. Diseño de banco de pruebas

El banco de pruebas fue impreso en 3D y ajustado con material DFD color blanco, cuenta con una dimensión de 25 x 25 cm, contemplando el espacio de manipulación de las muestras para la captura de fotografías, como se indica en la Fig. 5.

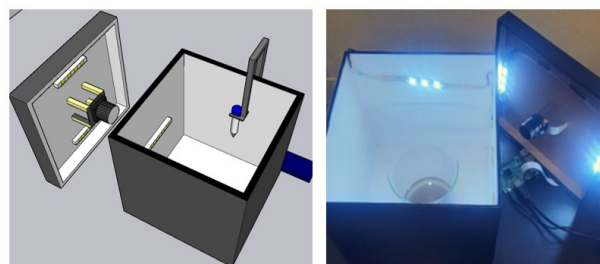


Figure 5. Implementación de banco de pruebas.

Para el desarrollo e implementación del banco de pruebas se tomaron en cuenta los aspectos de iluminación y los aspectos

que abarca la toma de captura como la cámara utilizada. Los aspectos de iluminación fueron caracterizados para su réplica, descrito en la Tabla 2.

TABLE II. ESPECIFICACIONES DEL BANCO DE PRUEBAS

<i>Propiedad</i>	<i>Valor</i>
Voltaje operacional	DC 12v 24v
Potencia	6w
Lúmenes	600 lux
Temperatura color	6500 K luz blanca
Longitud de onda	550 nm

Otro de los aspectos tomados en cuenta es el uso de un modelo de cámara en específico, para ello se utilizaron dos configuraciones idóneas para los resultados obtenidos. En la Tabla 3 se hace mención.

TABLE III. ESPECIFICACIONES DEL BANCO DE PRUEBAS

<i>Cámara</i>	<i>Resolución</i>	<i>FPS</i>
RPI-AUTO-IR CUT	16 MP	30 FPS
Smartphone gran angular	48 MP	30 FPS

La implementación del banco de pruebas permite evaluar el comportamiento de la materia orgánica que se sedimenta en las muestras realizado en periodos de 10 seg/min, para la generación de Data Set. De la misma manera se realizaron grabaciones por 5 minutos a 30 FPS, teniendo una relación de 30 FPS/seg. Se obtiene una cantidad de información de 6000 imágenes por captura de video.

### 1) Captura de imágenes.

Las imágenes obtenidas son preprocesadas para su manipulación, dentro de este procesamiento se opta como parte principal la selección de una región de interés de las imágenes. El banco de pruebas se utiliza precisamente para esta etapa, ya que las muestras cuentan con una coloración uniforme en toda la sección del espacio de captura evitando así reflejos del entorno, tal y como se aprecia en Fig. 6.

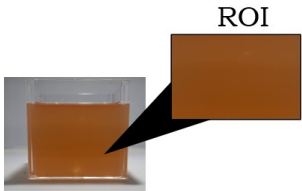


Figure 6. Bounding Box extraído de imágenes

### 2) Obtención de características.

La obtención de características tiene como propósito recabar la información que se tiene a inspección visual del agua. Los algoritmos desarrollados extraen y cuantifican cada nivel de variable equivalente a un nivel de turbidez, para ello

generando un Data Set que contiene la información extraída de las variables de RGB, intensidad luminosa y nivel de saturación, como se indica en la Tabla 4.

TABLE IV. MANEJO DE VARIABLES

<i>Variables extraídas</i>			<i>Variable objetivo</i>
R, G, B	Intensidad	Saturación	Turbidez

La obtención de los valores de cada variable tiene como objetivo establecer una correlación positiva con la determinación de los niveles de turbidez. Esta metodología se lleva a cabo a nivel computacional haciendo uso de lenguaje Python. La correlación generada facilita el procesamiento de los datos mediante un modelo de aprendizaje automático para realizar predicciones. La fase de obtención de características implica la implementación de algoritmos computacionales específicos para recopilar la información necesaria. A continuación, se detalla el proceso de manera específica.

#### 2.1) Obtención de información RGB.

El método de obtención de información RGB está basado en la extracción de información de todos los píxeles de la imagen a procesar, posteriormente bajo un condicionamiento se recorre la información de toda la imagen, hasta extraer la información total de toda la imagen. La Eq. 1 indica el valor de la matriz de la imagen [33].

$$Imagen \rightarrow matriz [m \times n \times 3] \tag{1}$$

La metodología se extiende en la extracción de componentes RGB, es decir, la iteración por canal, descrito en la Eq. 2.

$$\begin{aligned} R &= Imagen[y, x, 2] \\ G &= Imagen[y, x, 1] \\ B &= Imagen[y, x, 0] \end{aligned} \tag{2}$$

Esta metodología se puede deducir en el diagrama de flujo de la Fig. 7.

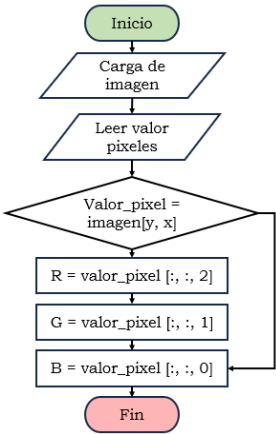


Figure 7. Diagrama de flujo obtención RGB.

## 2.2) Obtención de intensidad luminosa.

El método de extracción de intensidad luminosa surge de la extracción de información de los canales RGB, ya que estos valores se utilizan dentro del algoritmo de intensidad luminosa. El principio computacional indica que se debe trabajar en el espacio “BGR”, realizando un escalamiento del canal “RGB”, esta información se procesa y con un filtro se convierte a escala de grises. La Eq. 3 deduce como paso principal la conversión a escala de grises de la imagen [34].

$$Img\_gris_{ij} = 0.299 \times R_{ij} + 0.587 \times G_{ij} + 0.114 \times B_{ij} \quad (3)$$

La conversión de la imagen permite utilizar estos valores para el cálculo de la intensidad luminosa, esta ecuación está dada en la Eq. 4 [35].

$$Intensidad = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N Img\_gris_{ij} \quad (4)$$

La definición de estas dos ecuaciones utilizadas para la obtención de la intensidad se describe en el diagrama de flujo de la Fig. 8.

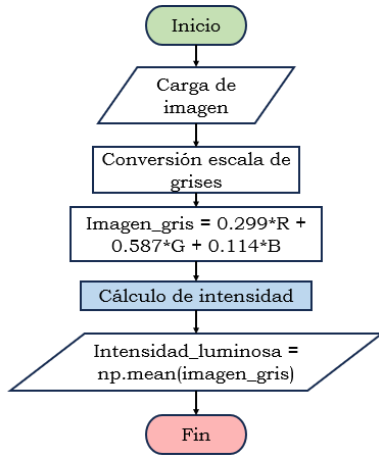


Figure 8. Diagrama de flujo obtención de intensidad luminosa.

## 2.3) Obtención de saturación.

La extracción de saturación [36], es una variable capaz de correlacionarse con la información de color de las imágenes. Su metodología está basada en el procesamiento de la información del espacio “HSV”, para extraer la información de cada canal de este espacio de color, para ello es necesario normalizar la información del RGB, descrito en la Eq. 5.

$$b = \frac{B}{255}, g = \frac{G}{255}, r = \frac{R}{255} \quad (5)$$

La información normalizada se utiliza para el cálculo de los valores mínimos y máximos de los canales como se muestra en la Eq. 6.

$$C_{max} = \max(r, g, b), \quad C_{min} = \min(r, g, b) \quad (6)$$

Estos valores son utilizados en el cambio de espacio de color, para el cálculo de cada componente de saturación, utilizados en la Eq. 7 [37].

$$Prom\_saturacion = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N HSV_{ij} \quad (7)$$

La información está contenida en matrices equivalentes a cada píxel, en la Fig. 9 se describe el diagrama de flujo del proceso.

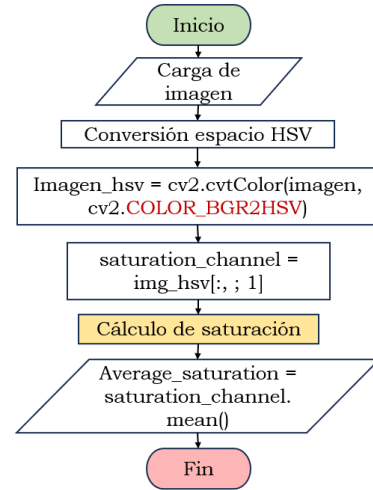


Figure 9. Implementación de banco de pruebas.

La implementación de estos algoritmos permite crear un Data Set que corresponda a los datos etiquetados de turbidez, en este caso la variable de referencia u objetivo, es decir la cual se desea predecir.

## 3) Interpolación de datos.

La extracción de características equivalentes a cada nivel de turbidez genera una relación directamente proporcional para la predicción del parámetro. El algoritmo de interpolación surge de la necesidad de no poder analizar los datos de forma continua, esto en el sentido que los datos de turbidez medidos no se pueden cuantificar en niveles deseados. La primer parte del algoritmo descrito en la Eq. 8 [38].

$$(X_{turbi}, Y_{turbi}) = regression(sample = 1000, n_{features} = 5, noise = 0.1) \quad (8)$$

La Eq. 8 describe el ejemplo de sampleo necesario para la ampliación de datos y utilizando un valor ejemplificado de ruido de medición generado de forma real por los turbidímetros. La escalabilidad de estas variables permite implementar la generación de nuevos valores. Esta transformación se describe en la Eq. 9.



$$y' = \frac{y - \min(y)}{\max(x) - \min(y)} x \left( \frac{\max}{range} - \frac{\min}{range} \right) + \frac{\min}{range} \quad (9)$$

En la Fig. 10 se describe el funcionamiento del algoritmo de interpolación.

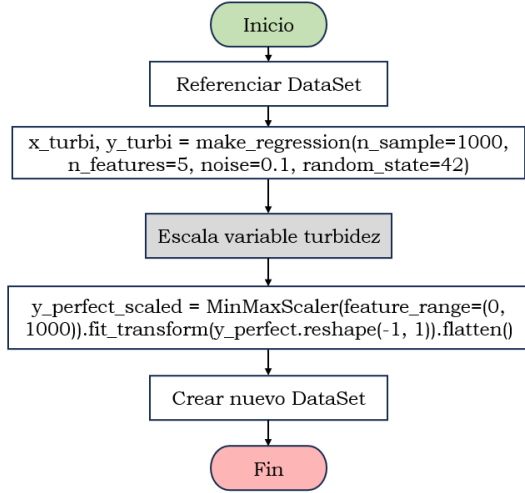


Figure 10. Algoritmo de interpolación.

La descripción del algoritmo de la Fig. 10 permite generar datos en un rango de 0 a 1000 NTU con lo respecto a las mediciones realizadas de manera real, esto potencia en gran medida los procesos de automatización y procesamiento de datos.

### Fase 3. Implementación de modelos

La predicción de turbidez surge a través de la implementación de un modelo de aprendizaje automático [39]. La importancia que estos modelos computacionales tienen para realizar predicciones surge a través de la forma en que se procesan los datos. La eficiencia de los modelos de aprendizaje se basa en la cantidad de datos obtenidos, en la Tabla 5 se muestra la interpolación de datos procesados.

TABLE V. TIPO DE VARIABLES

Datos originales	Datos interpolados
900 datos	6000 datos

#### 4) Entrenamiento de modelos.

El parámetro de Turbidez resalta en este trabajo como la variable objetivo o de referencia, ya que el procesamiento de todas las variables dependientes forma una relación positiva, además de llevar a cabo el ajuste del modelo, se alcanza la predicción. En la Fig. 11 se describe los modelos que se implementaron para la predicción, cada modelo fue evaluado con los mismos datos, de forma de encontrar el mejor rendimiento y predicción.

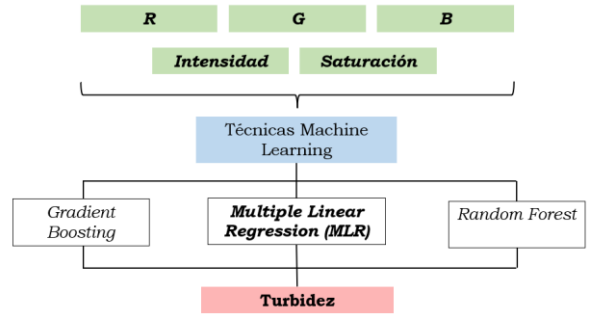


Figure 11. Modelos de Machine Learning.

De esta manera, los modelos fueron entrenados con la misma cantidad de datos, destacando que por cada variable obtenida se tienen 1000 datos, y con la cantidad total de variables se evaluaron un total de **6000 datos**, ya que los modelos fueron entrenados con los datos obtenidos del algoritmo de interpolación, cabe mencionar, que estos datos fueron evaluados y examinados de forma estadística para constatar la veracidad de los datos interpolados.

#### 5) Evaluación de métricas.

Para evaluar el rendimiento y desempeño de los modelos generados, se implementaron diferentes métricas estadísticas que fueron utilizadas para la evaluación de los algoritmos de regresión: Error absoluto medio (MAE), error cuadrático medio (MSE), raíz de error cuadrático medio (RMSE), coeficiente de determinación ( $R^2$ ). Estas métricas fueron evaluadas a los conjuntos de prueba, se describen a continuación [40].

##### 2.1) Error absoluto medio (MAE)

Esta métrica es el promedio sobre la muestra de prueba de las diferencias absolutas entre la predicción y la observación real, donde todas las diferencias individuales tienen el mismo peso, definido en Eq. 10 [40].

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (10)$$

##### 2.2) Error cuadrático medio (MSE)

La aplicación del error cuadrático medio es para medir el promedio de los cuadrados de los errores, es decir, la diferencia promedio al cuadrado entre los valores estimados y el valor real [40], como se aprecia en la Eq. 11.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (11)$$

##### 2.3) Raíz del error cuadrático medio (RMSE)

Esta métrica se describe como el promedio de las diferencias al cuadrado entre las predicciones y lo real observado, indicando una magnitud del error, descrita en la Eq. 12.

$$RMSE = \sqrt{\sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (12)$$

#### 2.4) Coeficiente de determinación ( $R^2$ )

El coeficiente de determinación mide la proporción de la varianza en la variable dependiente que es predecible a partir de las variables independientes. Proporciona una indicación de bondad de ajuste y varía de 0 a 1. Se describe en la Eq. 13.

$$R^2 = \frac{MSE(model)}{MSE(baseline)} \quad (13)$$

### IV. RESULTADOS

En esta sección se presenta las evaluaciones de los algoritmos evaluados, así como la veracidad de los datos interpolados para el entrenamiento de los modelos de aprendizaje. El conjunto de datos resultantes está conformado por la ejecución y evaluación de 3 modelos de aprendizaje automático utilizados en la predicción de turbidez. La evaluación se realizó mediante el uso de dos conjuntos de datos obtenidos de la captura y extracción de Bounding Box de las imágenes, el primer conjunto son los datos reales y el segundo los datos interpolados. En la Tabla 6, se muestra el volumen de datos utilizados para la predicción.

TABLE VI. TOTAL, DE DATOS PROCESADOS

Resolución de img.	Volumen de img.	Total datos
1920 x 1080 px	6000 imágenes	6000

Estos datos fueron evaluados y comparados con diferentes algoritmos, cotejando de esta forma el algoritmo de interpolación, y la evaluación de modelos regresión lineal múltiple (MLR), Gradient Bosting (GB) y Random Forest (RF). Los resultados se muestran las siguientes secciones.

#### A. Correlaciones de variables.

La metodología empleada en este estudio enfatiza la importancia de establecer niveles de correlación entre las variables obtenidas. Para lograrlo, cada variable se relacionó con las demás, permitiendo identificar correlaciones positivas fuertes que son esenciales para el análisis. En la Figura 12, se presentan estas correlaciones mediante una matriz de correlación.

La matriz de correlación mostrada en la imagen evalúa la relación entre diversas características de turbidez, incluyendo los valores de R (Rojo), G (Verde), B (Azul), Intensidad, Saturación y Turbidez. Cada celda en la matriz indica el coeficiente de correlación entre dos variables, donde los valores oscilan entre -1 y 1. Un valor de 1 representa una correlación positiva perfecta, 0 indica que no hay correlación, y -1 denota una correlación negativa perfecta.

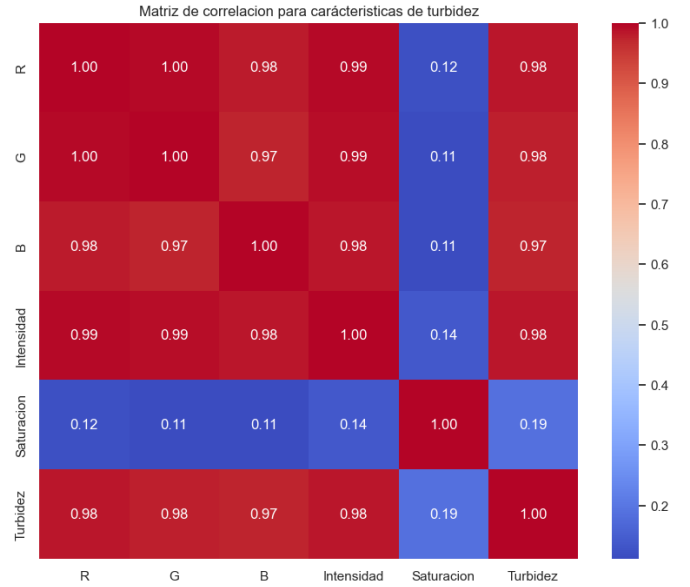


Figure 12. Análisis de correlación de variables

El análisis de los datos evaluado con la matriz de correlación descrito en la Fig. 12, concluye lo siguiente:

- Las variables R, G, B, e Intensidad tienen correlaciones muy altas entre sí, todas cercanas a 1.
- La Saturación muestra correlaciones ligeramente menores con R, G, B e Intensidad, lo que indica una relación menos significativa.
- La Turbidez tiene correlaciones moderadas con R, G, B e Intensidad.

La interpretación de los datos permite establecer correlaciones positivas en el análisis de todas las variables. El siguiente análisis está basado en el análisis de la dispersión de los datos evaluados, descrito en la Fig. 13.

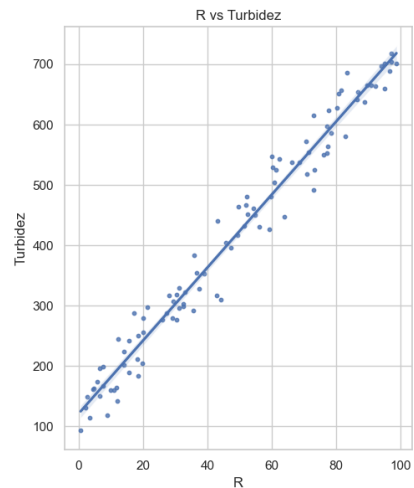


Figure 13. Gráfica de dispersión de canal R.

De la misma forma se evaluó el canal G y B, como se muestra en la Fig. 14 y Fig. 15.



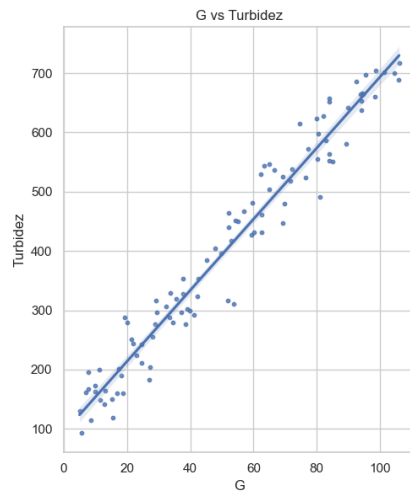


Figure 14. Gráfico de dispersión de canal G.

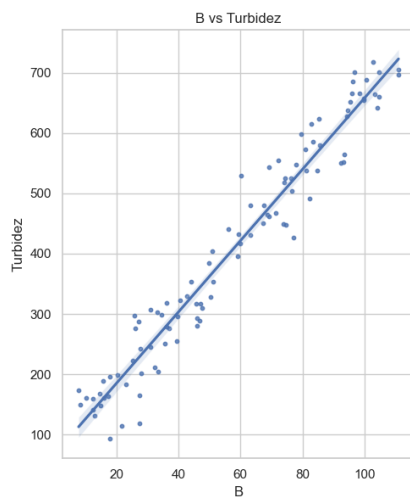


Figure 15. Gráfico de dispersión de canal B.

La dispersión de los datos de los canales RGB, indica que los datos procesados son positivamente evaluados, acorde a lo analizado con la matriz de correlación y la dispersión de todos los datos.

Las dispersiones evaluadas fueron cuantificadas con la media ponderada de las correlaciones existentes, en la Tabla 7, se describe las medias de las correlaciones por variable. De igual manera se analizó la dispersión para la intensidad luminosa y la saturación de las imágenes, como se aprecia en la Fig. 16, 17.

TABLE VII. MEDIA PONDERADA POR VARIABLE

<i>R</i>	<i>G</i>	<i>B</i>	<i>Intensidad</i>	<i>Saturación</i>
0.9848	0.9778	0.9743	0.982944	0.46753

De la misma forma se analizaron los datos con análisis de grafico de dispersión, para la variable de intensidad luminosa y saturación. Estos gráficos se muestran en la Fig. 16 y Fig. 17

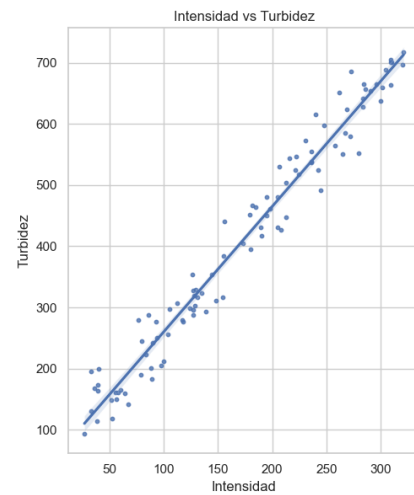


Figure 16. Dispersión de intensidad luminosa

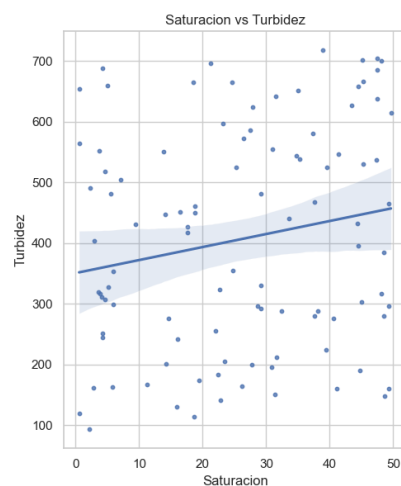


Figure 17. Dispersión de saturación

Este análisis permite dar un panorama amplio para la veracidad del estudio en la predicción de turbidez, ya que el sesgo de las variables tiene una tendencia positiva en todas las variables.

### B. Evaluación preliminar de datos reales.

Esta etapa se basa en la evaluación de los datos reales obtenidos de todas las imágenes. Para ello se implementó un algoritmo de regresión lineal, en donde se analizaban los datos de las variables obtenidas y la **turbidez** como variable objetivo.

Este análisis se llevó a cabo con la evaluación de un modelo de regresión lineal simple, de forma de establecer correlación de las variables, con el único fin de alcanzar la variable objetivo, esta evaluación describe en la Fig. 18, evaluando la predicción con los datos extraídos de las variables equivalente a las imágenes de cada nivel de turbidez.

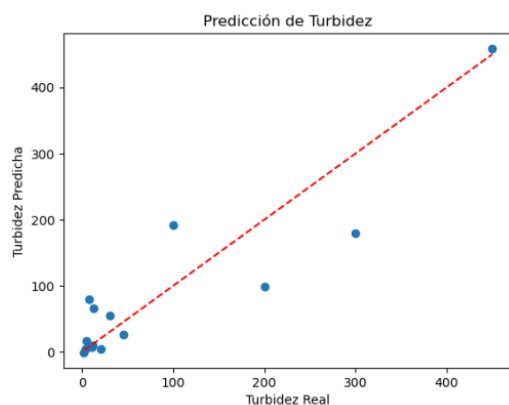


Figure 18. Regresión lineal con datos originales

La evaluación preliminar obtiene una regresión poco significativa, evaluada en la Tabla 6. Esta prueba reside en la no continuidad de los datos de turbidez, es por ello, que se procedió a implementar un algoritmo de interpolación capaz de satisfacer estos niveles no cuantificados.

TABLE VIII. EVALUACIÓN DE REGRESIÓN LINEAL SIMPLE

Métrica	Resultado
Error cuadrático medio (MSE)	0.72012
Coefficiente de determinación ( $R^2$ )	0.87327

Estos resultados no son idóneos a lo reportado en la literatura, es por ello que el algoritmo de interpolación busca optimizar y elevar las evaluaciones.

### C. Análisis de datos interpolados.

Como se mencionaba en la Etapa 3 de la Fase 2 descrita en la metodología, surge la necesidad del aumento de datos para la predicción, es por ello que se cotejaron los datos generados con el algoritmo de interpolación realizando un análisis estadístico y de densidad para los datos. Analizando así, la veracidad de cada variable, como se aprecia en la Fig. 19.

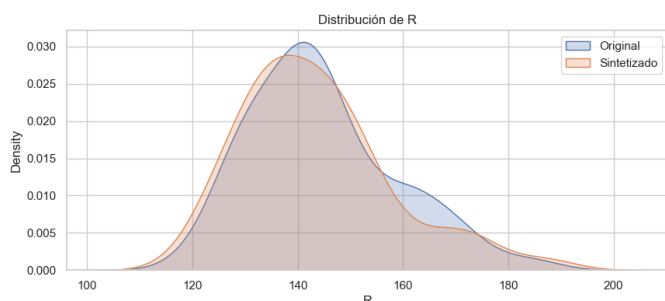


Figure 19. Análisis de densidad de canal R

La gráfica de densidad representa la probabilidad y generación de datos en el rango de los valores originales, es decir, el análisis de densidad busca representar el numero de veces que

el algoritmo de interpolación es capaz de ocupar espacio dentro de las muestras originales, permitiendo analizar la distribución de los datos sintéticos y originales. Este análisis de realizó para cada variable del DataSet, como se aprecia en la Fig. 20.

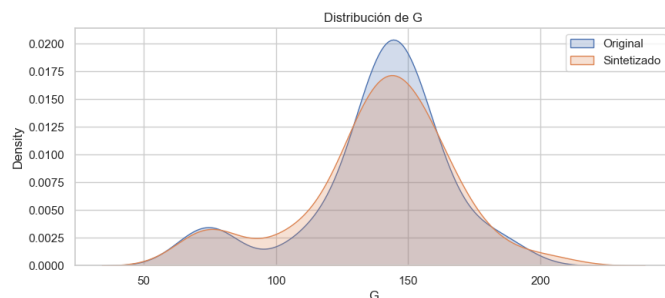


Figure 20. Análisis de densidad de canal G

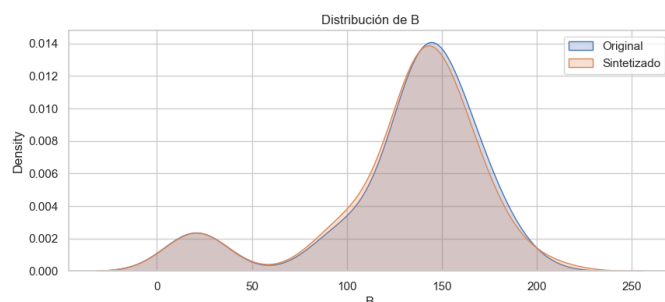


Figure 21. Análisis de densidad de canal B

La deducción de la información de los canales RGB indica que las diferencias son casi nulas en la mayoría del área, como una densidad ligeramente mayor en los datos sintetizados alrededor de 130 y una menor densidad en los valores superiores a 160 en comparación con los datos originales.

De la misma forma se evalúa para la variable de intensidad, saturación y turbidez, como se aprecia en la Fig. 22, 23 y 24.

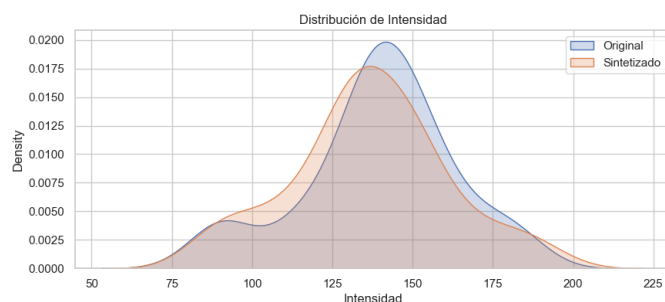


Figure 22. Análisis de densidad variable intensidad

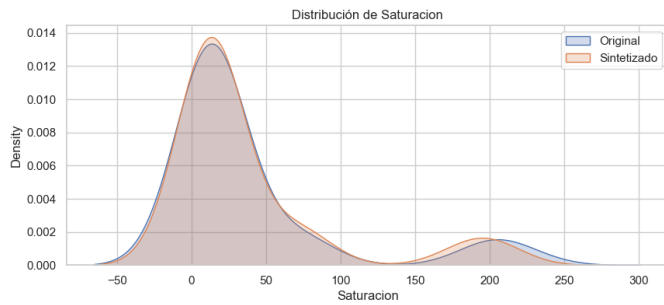


Figure 23. Análisis de densidad variable saturación

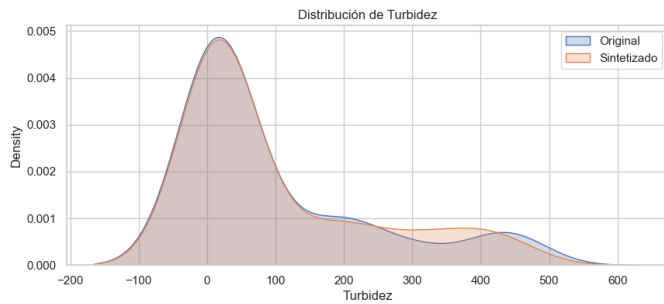


Figure 24. Análisis de densidad variable turbidez

La similitud en las distribuciones indica que cualquier modelo entrenado con los datos originales podría también ser aplicado a los datos sintetizados con resultados comparables. Realizar este análisis permite utilizar estos datos para la evaluación, de la misma forma se realizó el análisis estadístico de los datos originales y los datos sintetizados a manera de comparar el análisis de densidad descrito en la Tabla 9.

TABLE IX. ANÁLISIS ESTADÍSTICO DE DATOS GENERADOS

<i>Estadística</i>	<i>Datos originales</i>	<i>Datos interpolados</i>
<b>Mean</b>	144.607	143.390
<b>Std</b>	13.838	14.323
<b>Min</b>	118.997	118.439
<b>25%</b>	133.645	133.208
<b>50%</b>	142.162	141.379
<b>75%</b>	150.494	150.345
<b>Max</b>	185.486	188.094

#### D. Implementación y evaluación de modelos.

La evaluación de los datos permite realizar nuevas evaluaciones con el modelo anteriormente utilizado, en donde relacione las variables y contemple el aumento de datos. La prueba con mejor rendimiento fue la implementación y evaluación de un modelo de Regresión Logística Múltiple (MLR), donde podemos apreciar que se contempla la nueva continuidad de los datos generados, Fig. 25.

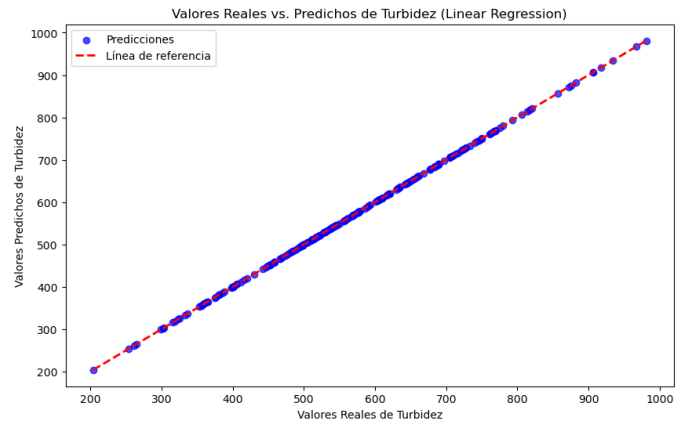


Figure 25. Evaluación de predicción modelo de regresión lineal múltiple

La evaluación de los datos permite realizar nuevas evaluaciones con el modelo anteriormente utilizado, en donde relacione las variables y contemple el aumento de datos. Para ello se implementaron diferentes métricas, como se aprecia en la Tabla 10.

TABLE X. EVALUACIÓN DE MÉTRICAS REGRESIÓN LINEAL MÚLTIPLE

<i>Métrica</i>	<i>Resultados</i>
Error cuadrático medio (MSE)	1.010631e-24
Coefficiente de determinación ( $R^2$ )	<b>0.9605</b>
Error medio absoluto (MAE)	1.000160e-12
Raíz del error cuadrático medio (RMSE)	1.005301e-12

Se aprecian muy buenos resultados en la evaluación de este modelo con un  $R^2$  de **0.96**. Para evitar un posible sobreajuste en el modelo se evaluaron dos más, el siguientes es el modelo de Gradient Boosting, en la Fig. 26, se aprecia el rendimiento de este.

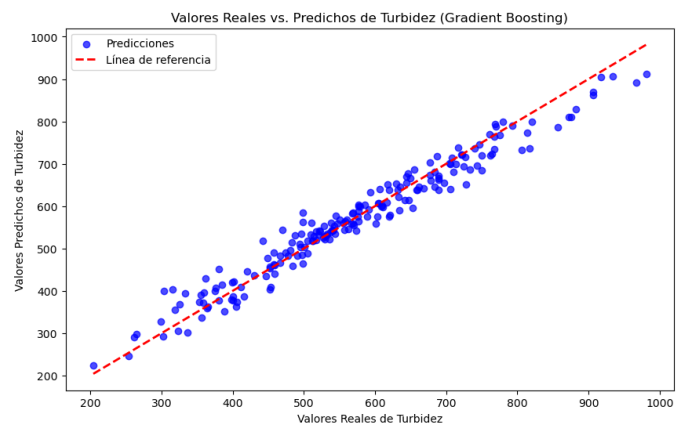


Figure 26. Evaluación de predicción modelo Gradient Boosting

Este modelo permite tener una perspectiva amplia, de que no se cuenta con sobreajuste en el regresor, la evaluación de esta

modelo coteja lo anterior dicho, por la eficiencia de los resultados, se aprecia en la Tabla 11.

TABLE XI. EVALUACIÓN DE MÉTRICAS DE ALGORITMO GRADIENT BOOSTING

Métrica	Resultados
Error cuadrático medio (MSE)	1.037050e+03
Coefficiente de determinación ( $R^2$ )	<b>0.956039</b>
Error medio absoluto (MAE)	2.474568e+01
Raíz del error cuadrático medio (RMSE)	3.220326e+01

La evaluación del modelo destaca con un  $R^2$  de **0.95** y ponderaciones de errores muy cercanos al valor de 0. Como última parte se realizó la evaluación de un modelo Random Forest, como se aprecia en la Fig. 27.

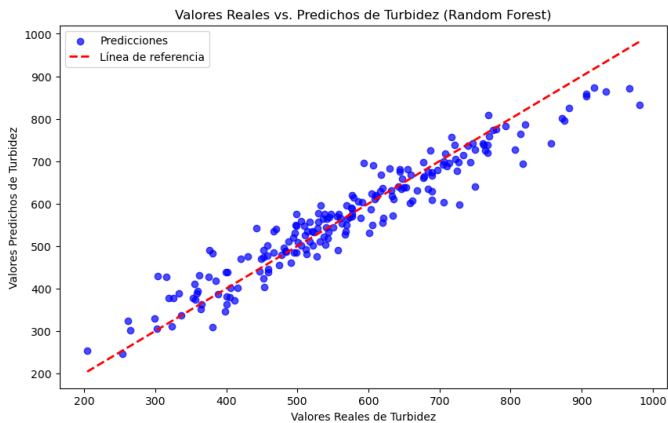


Figure 27. Evaluación de predicción modelo Random Forest

Las métricas evaluadas para el modelo de Random Forest se describen en la Tabla 12.

TABLE XII. EVALUACIÓN DE MÉTRICAS DE ALGORITMO RANDOM FOREST

Métrica	Resultados
Error cuadrático medio (MSE)	2.056251e+03
Coefficiente de determinación ( $R^2$ )	<b>0.912835</b>
Error medio absoluto (MAE)	3.476812e+01
Raíz del error cuadrático medio (RMSE)	4.534591e+01

Este modelo cuenta con un rendimiento relativamente menor a los dos anteriores, pero con un  $R^2$  de **0.91**, esto valida la efectividad de los datos y las evaluaciones. Dentro de la revisión del estado del arte, se puede deducir la mejora significativa de evaluación para predicción, en la Tabla 10, se describe la comparación con trabajos relacionados en la predicción de la turbidez.

TABLE XIII. EVALUACIÓN DE REGRESIÓN

Autor	Evaluación ( $R^2$ )
Propuesta (MLR)	20.960
Yuan Hong L. et al.	0.956
Mullins D. et al.	0.890
Silva G. et al.	0.727

Conclusiones

Lo reportado en el estado del arte indica que la metodología de análisis de turbidez es estrictamente un proceso llevado a cabo en laboratorio o haciendo uso de instrumentos con costo elevado, es por lo que automatizar este proceso satisface en gran medida la cuantificación de turbidez, es por ello que el reto surge desde la recolección de muestras y mediciones in situ. La metodología propuesta permite realizar el estudio y estimación de los niveles de turbidez en el agua. La arquitectura utilizada haciendo uso de un modelo de aprendizaje automático de regresión lineal múltiple (MLR) potencializa la automatización de los procesos de medición de turbidez. Implementar este tipo de algoritmos permite aumentar el rango de predicción de los niveles de turbidez, combinando operaciones de visión por computadora como la extracción de características, el aumento de datos y el uso de algoritmos complejos de aprendizaje automático. Los resultados obtenidos dejan a la mano la cuantificación y evaluación de datos en rango de 0 a 1000 NTU, estos valores son los rangos medibles en turbidímetros comerciales. La arquitectura alcanza, evalúa y predice los rangos de estos dispositivos. La predicción del parámetro de turbidez haciendo uso de esta arquitectura puede ser implementada en estudios relacionados en el análisis de calidad del agua. Así como trabajos relacionados en alarmas predictivas por niveles elevados de turbidez.

AGRADECIMIENTOS

Especial agradecimiento a CONAHCYT, quién brinda el apoyo necesario para el desarrollo de la investigación. De igual forma al instituto tecnológico de Tuxtla Gutiérrez, por abrir las puertas para la realización de la investigación, así como a los colaboradores del trabajo, ya que, con el aporte de cada uno de ellos, este trabajo es posible.

REFERENCIAS BIBLIOGRÁFICAS

[1] Secretaría de Salud. Nom-127-ssa1-2021 salud ambiental (2021). criterios sanitarios para la prestación de servicios de alimentación. [https://www.dof.gob.mx/nota\\_detalle.php?codigo=5650705&fecha=02/05/2022#gsc.tab=0](https://www.dof.gob.mx/nota_detalle.php?codigo=5650705&fecha=02/05/2022#gsc.tab=0)

[2] Fuentes S, Lorena del C, Mendoza S, Iván A, López M, Angela M, Castro V, María F, & Urdaneta M, Carlos J. (2011). Efectividad de un coagulante extraído de *Stenocereus griseus* (Haw.) Buxb. en la potabilización del agua. *Revista Técnica de la Facultad de Ingeniería Universidad del Zulia*, 34(1). [http://ve.scielo.org/scielo.php?script=sci\\_arttext&pid=S0254-07702011000100007&lng=es&tlng=es](http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S0254-07702011000100007&lng=es&tlng=es)

[3] World Health Organization. (2011). Guidelines for drinking-water quality, 4th edition, incorporating the 1st addendum. Geneva, Switzerland.

[4] Bartram, Jamie, Ballance, Richard, World Health Organization & United Nations Environment Programme. (1996). Water quality monitoring: a practical

guide to the design and implementation of freshwater quality studies and monitoring programs <https://apps.who.int/iris/handle/10665/41851>

[5] World Health Organization. (2006). Water quality: Guidelines, standards and health. Geneva, Switzerland: Author.

[6] Comisión Nacional del Agua, "Red de Monitoreo de Calidad del Agua en México," 2015. [En línea]. Recopilado de: <https://files.conagua.gob.mx/conagua/mapas/SGAPDS-1-15-Libro25.pdf>

[7] Elke Beyreuther, Leonhard Karsch, Lydia Laschinsky, Elisabeth Leßmann, Doreen Naumburger, Melanie Oppelt, Christian Richter, Michael Schürer, Julia Woithe & Jörg Pawelke (2015) Radiobiological response to ultra-short pulsed megavoltage electron beams of ultra-high pulse dose rate, International Journal of Radiation Biology, DOI: 10.3109/09553002.2015.1043755

[8] Elhag, M.; Gitas, I.; Othman, A.; Bahrawi, J.; Gikas, P. Assessment of Water Quality Parameters Using Temporal Remote Sensing Spectral Reflectance in Arid Environments, Saudi Arabia. Water 2019, 11, 556. <https://doi.org/10.3390/w11030556047041>

[9] Kubin, E.; Poulain, P.-M.; Mauri, E.; Menna, M.; Notarstefano, G. Levantine Intermediate and Levantine Deep Water Formation: An Argo Float Study from 2001 to 2017. Water 2019, 11, 1781. <https://doi.org/10.3390/w11091781>

[10] Knudsen, R., Klemetsen, A., Alekseyev, S. et al. The role of Salvelinus in contemporary studies of evolution, trophic ecology and anthropogenic change. Hydrobiologia 783, 1–9 (2016). <https://doi.org/10.1007/s10750-016-2999-9>

[11] Godoy, V. M. (2018). Análisis comparativo de la disminución de la turbidez en el proceso de floculación utilizando un floculante comercial y la paleta de tuna. Recuperado de <http://hdl.handle.net/11537/13840>

[12] G. E. Magallón-Pastor, "Calidad del agua de fuentes de abastecimiento para sistemas de potabilización en el estado de Puebla," \*Revista Internacional de Contaminación Ambiental\*, vol. 30, no. 3, pp. 229-244, 2014. [Online]. Available: <https://dialnet.unirioja.es/servlet/articulo?codigo=4835509>

[13] D. M. Fúquene and A. V. Yate, "Ensayo de jarras para el control del proceso de coagulación en el tratamiento de aguas residuales industriales", *work paper*, vol. 2, no. 1, Oct. 2018, [doi: 10.22490/ECAPMA.2771](https://doi.org/10.22490/ECAPMA.2771).

[14] A. E. Guerrero Escobedo, «Una Remoción de turbidez de aguas del canal madre de Chavimochic empleando diseño de mezcla de coagulantes: Remoción de turbidez de aguas utilizando sensores de bajo costo», Alpha Centauri, vol. 2, n.º 3, pp. 60–71, jul. 2021. Doi: <https://doi.org/10.47422/ac.v2i3.41>

[15] Rocher, J.; Parra, L.; Taha, M.; Lloret, J. (2018). Diseño de una red de sensores para monitorizar una instalación acuícola. En XIII Jornadas de Ingeniería telemática (JITEL 2017). Libro de actas. Editorial Universitat Politècnica de València. 48-54. <https://doi.org/10.4995/JITEL2017.2017.6623>

[16] K. M. Medina Estrada, "Evaluación de la calidad del agua para consumo humano en la ciudad de Ambato," Licenciatura, Facultad de Ciencias de la Salud, Universidad Técnica de Ambato, Ambato, Ecuador, 2018. [Online]. Available: <http://repositorio.uta.edu.ec/handle/123456789/26334>

[17] Andújar Regueiro, J. (2012). Propuesta de la sustitución de la medida de la turbidez del agua mediante el disco de secchi por un nuevo procedimiento basado en la extinción de la luz solar en el agua a distintos anchos de banda con un nuevo aparato (LUZEX). <http://hdl.handle.net/10251/167263>

[18] J. Meseguer, M. López, and R. Ibañez, "Evaluation of the Benefits of Subsurface Drip Irrigation of Olive Trees in South-East Spain," *Revista de Agroecología y Agricultura Sostenible*, vol. 12, no. 1, pp. 1-15, 2023. Doi: <https://doi.org/10.4995/raet.2019.12603>

[19] Y. Ma, X. Feng, L. Chao, D. Huang, Z. Xia and X. Jiang, "A New Database for Evaluating Underwater Image Processing Methods," 2018 Eighth International Conference on Image Processing Theory, Tools and Applications (IPTA), Xi'an, China, 2018, pp. 1-6, doi: 10.1109/IPTA.2018.8608131

[20] Xiu Li, Min Shang, H. Qin and Liansheng Chen, "Fast accurate fish detection and recognition of underwater images with Fast R-CNN," OCEANS 2015 - MTS/IEEE Washington, Washington, DC, 2015, pp. 1-5, doi: 10.23919/OCEANS.2015.7404464

[21] D. S. Bona, A. Murni and P. Mursanto, "Semantic Segmentation and Segmentation Refinement Using Machine Learning Case Study: Water

Turbidity Segmentation," 2019 IEEE International Conference on Aerospace Electronics and Remote Sensing Technology (ICARES), Yogyakarta, Indonesia, 2019, pp. 1-5, doi: 10.1109/ICARES.2019.8914551

[22] I. Montassar and A. Benazza-Benyahia, "Water turbidity estimation in water sampled images," 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Sousse, Tunisia, 2020, pp. 1-5, doi: 10.1109/ATSIP49331.2020.9231862

[23] V. Karnawat and S. L. Patil, "Turbidity detection using image processing," 2016 International Conference on Computing, Communication and Automation (ICCCA), Greater Noida, India, 2016, pp. 1086-1089, doi: 10.1109/CCAA.2016

[24] Boyd, C.E. Suspended Solids, Color, Turbidity, and Light. In: Water Quality 2020. Springer, Cham. [https://doi.org/10.1007/978-3-030-23335-8\\_6](https://doi.org/10.1007/978-3-030-23335-8_6)

[25] Matthew Min Enn Chai, Sing Muk Ng, Hong Siang Chua; An alternative cost-effective image processing-based sensor for continuous turbidity monitoring. AIP Conf. Proc. 13 March 2017; 1808 (1): 020014. <https://doi.org/10.1063/1.4975247.7813877>

[26] Lopez-Betancur, D.; Moreno, I.; Guerrero-Mendez, C.; Saucedo-Anaya, T.; González, E.; Bautista-Capetillo, C.; González-Trinidad, J. Convolutional Neural Network for Measurement of Suspended Solids and Turbidity. Appl. Sci. 2022, 12, 6079. <https://doi.org/10.3390/app12126079>

[27] S. Poornapushpakala, S.S. Dawn, M. Arun Govind, A. Santhosh, N. Nirmala, S. Barani, M. Nirmala, A novel image processing technique to evaluate biodiesel wastewater for recovery, recycle and reuse towards zero liquid discharge approach (2021). <https://doi.org/10.1016/j.bcab.2020.101874>

[28] Jiacong Huang, Rui Qian, Junfeng Gao, Haijian Bing, Qi Huang, Lingyan Qi, Song Song, Jiafang Huang. A novel framework to predict water turbidity using Bayesian modeling, Water Research (2021). <https://doi.org/10.1016/j.watres.2021.117406>

[29] A. J. Gimenez et al., "Turbidimetry by Image Degradation Analysis," in IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 10, pp. 7574-7579, Oct. 2020, doi: 10.1109/TIM.2020.2985902

[30] Darragh Mullins, Derek Coburn, Louise Hannon, Edward Jones, Eoghan Clifford, Martin Glavin; A novel image processing-based system for turbidity measurement in domestic and industrial wastewater. Water Sci Technol 12 March 2018; 77 (5): 1469–1482. doi: <https://doi.org/10.2166/wst.2018.030>

[31] Y. Liu, Y. Chen and X. Fang, "A Review of Turbidity Detection Based on Computer Vision," in IEEE Access, vol. 6, pp. 60586-60604, 2018, doi: 10.1109/ACCESS.2018.2875071

[32] Hector Ricardo Hernandez de Leon. Supervision et diagnostic des procédés de production d'eau potable. Automatique / Robotique. INSA de Toulouse, 2006. Français. (tel-00136157)

[33] Rafael C. Gonzalez, Richard E. Woods, "Digital Image Processing," 3rd Edition, Pearson, 2007.

[34] Image Processing: The Fundamentals, M. Petrou & C. Petrou, 2nd ed, Wiley and Sons Ltd, 2010.

[35] The Perception of Visual Information, W.R. Hendee & P.N.T. Wells, 2nd ed, Springer-Verlag, 1997.

[36] Chung, Y.-S.; Kim, N.-H. Saturation-Based Airlight Color Restoration of Hazy Images. Appl. Sci. 2023, 13, 12186. <https://doi.org/10.3390/app132212186>

[37] R. Szeliski, "Computer Vision: Algorithms and Applications," Springer, 2010.

[38] W. K. Pratt, "Digital Image Processing: PIKS Inside," 4<sup>th</sup> ed., Wiley, 2007.

[39] Lalit Kumar, Mohammad Saud Afzal, Ashad Ahmad. "Prediction of water turbidity in a marine environment using machine learning: A case study of Hong Kong". Regional Studies in Marine Science, Volume 52, (2022). <https://doi.org/10.1016/j.rsma.2022.102260>

[40] Chicco D, Warrens MJ, Jurman G. 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science. <https://doi.org/10.7717/peerj-cs.623>