



**UNIVERSIDAD POLITÉCNICA SALESIANA
SEDE GUAYAQUIL
CARRERA DE INGENIERÍA ELECTRÓNICA**

**“IDENTIFICACIÓN DEL CRECIMIENTO DE LOS NAUPLIOS POR MEDIO DE
DEEP LEARNING”**

Trabajo de titulación previo a la obtención del
Título de **Ingeniero Electrónico**

AUTOR:

- **ERIK ANDRÉS SILVA GUARANDA**

TUTOR:

MSC. VICTOR DAVID LARCO TORRES

GUAYAQUIL – ECUADOR

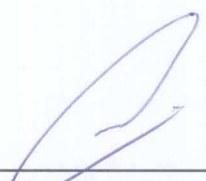
2024

**CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE
TITULACIÓN**

Yo **Erik Andrés Silva Guaranda** con cédula de identificación N°**0928557826** manifiesto que:

Soy autor y responsable del presente trabajo; y, autorizamos a que sin fines de lucro la Universidad Politécnica Salesiana pueda usar, difundir, reproducir o publicar de manera total o parcial el presente trabajo de titulación.

Guayaquil, 29 de enero del año 2024.




(f)Erik Andrés Silva Guaranda
C.I: 0928557826

**CERTIFICADO DE CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE
TITULACIÓN A LA UNIVERSIDAD POLITÉCNICA SALESIANA**

Yo **Erik Andrés Silva Guaranda** con cédula de identificación N°**0928557826**, expreso mi voluntad y por medio del presente documento cedo a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que soy autor del **Artículo Académico: “IDENTIFICACIÓN DEL CRECIMIENTO DE LOS NAUPLIOS POR MEDIO DE DEEP LEARNING”**, el cual ha sido desarrollado para optar por el título de: **INGENIERO ELECTRÓNICO**, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En concordancia con lo manifestado, suscribimos este documento en el momento que hacemos la entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Guayaquil, 29 de enero del año 2024.



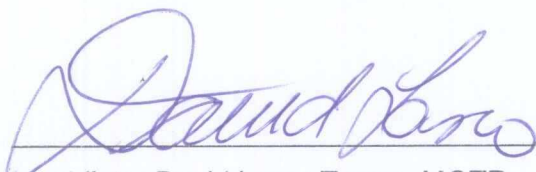
(f) Erik Andrés Silva Guaranda
C.I: 0928557826

CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN

Yo Víctor David Larco Torres con documento de identificación N° 0923270136, docente de la **Universidad Politécnica Salesiana**, declaró que bajo mi autoría fue desarrollado el trabajo de titulación: **“IDENTIFICACIÓN DEL CRECIMIENTO DE LOS NAUPLIOS POR MEDIO DE DEEP LEARNING PARA LA UNIVERSIDAD POLITÉCNICA SALESIANA”**, realizado por **Erik Andrés Silva Guaranda con cédula de identificación N°0928557826** obteniendo como resultado final el trabajo de titulación bajo la opción **Artículo Académico**, que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Guayaquil, 29 de enero del año 2024.

Atentamente, _



Ing. Víctor David Larco Torres, MGTR

C.I: 0923270136

DEDICATORIA

El presente trabajo es dedicado todos aquellos que se esfuerzan en generar nuevos conocimientos aportando a los avances de la ciencia aplicadas a los distintos campos productivos, incorporando herramientas de tecnología en post de la distribución y gestión responsable de los recursos finitos. Que este trabajo inspire a continuar explorando nuevas fronteras, el objetivo de mejorar la calidad de vida de las sociedades emergentes o consolidadas.

Erik Silva Guaranda

AGRADECIMIENTO

Quiero manifestar mi gratitud a mis mentores, cuya guía experta ha sido un faro constante. Su sabiduría y apoyo fueron fundamentales para superar los desafíos y alcanzar los logros que este trabajo presenta.

A mi familia y seres queridos, les agradezco por su inquebrantable apoyo emocional y comprensión. Su aliento ha sido el motor que me impulsó en los momentos más desafiantes.

Este proyecto materializa un esfuerzo colectivo y es el fruto de una colaboración valiosa. A todos aquellos que, de una manera u otra, contribuyeron a este trabajo, les expreso mi más profundo agradecimiento. Este logro no habría sido posible sin su dedicación y participación en este fascinante viaje de descubrimiento e investigación.

Erik Silva Guaranda

RESUMEN

AÑO	ALUMNOS	DIRECTOR DE PROYECTO	TEMA DE PROYECTO DE TITULACIÓN
2024	ERIK ANDRES SILVA GUARANDA	MGTR. VICTOR DAVID LARCO TORRES	IDENTIFICACIÓN DEL CRECIMIENTO DE LOS NAUPLIOS POR MEDIO DE DEEP LEARNING

Este estudio se centra en el desarrollo de un modelo Deep Learning (DL) para la identificación de poblaciones y la evaluación de larvas de camarón, usando datos de parámetros críticos como Oxígeno, Saturación y Temperatura. Poder analizar datos, permite tomar decisiones correctas y oportunas enfocadas en la optimización de tiempo y recursos al hacer uso de los avances de la tecnología y el desarrollo de algoritmos DL. Este tipo de herramientas permiten extraer patrones complejos para obtener los resultados deseados, según sea el área de interés en cuanto a predicciones ajustadas basadas en la aplicación de DL.

Este enfoque implica la recopilación integral de datos sobre el crecimiento de las larvas de camarón utilizando información detallada sobre cada parámetro de calidad del agua. La integración de tecnologías avanzadas como DL para validar el desarrollo larval, se perfila como una herramienta útil, centrada en la eficiencia y la precisión de predicciones, permitiendo a los productores tomar decisiones rápidas y eficaces para optimizar las condiciones de crecimiento larvario.

El desarrollo de esta investigación permitirá comprender mejor la aplicabilidad de algoritmos de aprendizaje profundo, siendo esta una tecnología avanzada utilizada en la predicción de supervivencia a partir de los parámetros críticos en las etapas del desarrollo de las larvas de camarón. El trabajo investigativo contribuye significativamente a las prácticas de la acuicultura inteligente enfocadas en la incorporación de métodos avanzados.

Palabras claves: Deep Learning, Parámetros críticos, Análisis de datos, Acuicultura inteligente, Predicción de supervivencia.

ABSTRACT

YEAR	STUDENTS	PRJ. DIRECTOR	SUBJECT
2024	ERIK ANDRES SILVA GUARANDA	MGTR. VICTOR DAVID LARCO TORRES	IDENTIFICATION OF NAUPLIUS GROWTH THROUGH DEEP LEARNING

This study focuses on the development of a Deep Learning (DL) model for the identification of populations and the evaluation of shrimp larvae using critical parameters such as Oxygen, Saturation, and Temperature. Analyzing data enables making correct and timely decisions focused on optimizing time and resources by leveraging technological advancements and DL algorithm development. These tools allow extracting complex patterns to achieve desired results, tailored to the area of interest in terms of fine-tuned predictions based on DL application.

This approach involves comprehensive data collection on shrimp larvae growth using detailed information on each water quality parameter. The integration of advanced technologies like DL to validate larval development emerges as a useful tool, centered on the efficiency and precision of predictions, enabling producers to make quick and effective decisions to optimize larval growth conditions.

The development of this research will provide a better understanding of the applicability of deep learning algorithms, an advanced technology used in predicting survival based on critical parameters in the stages of shrimp larvae development. The investigative work significantly contributes to smart aquaculture practices focused on the incorporation of advanced methods.

Keywords: Deep Learning, Critical parameters, Data analysis, Smart aquaculture, Survival prediction

ÍNDICE GENERAL

CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE TITULACIÓN.....	II
CERTIFICADO DE CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE TITULACIÓN A LA UNIVERSIDAD POLITÉCNICA SALESIANA.....	III
CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN	IV
DEDICATORIA.....	V
AGRADECIMIENTO	VI
RESUMEN	VII
ABSTRACT	VIII
ÍNDICE GENERAL.....	IX
ÍNDICE DE ILUSTRACIONES	X
ÍNDICE DE TABLAS.....	X
INTRODUCCIÓN.....	1
FUNDAMENTO TEÓRICO.....	2
RANDOMFOREST	2
REDES NEURONALES	2
PYTHON.....	3
COEFICIENTE DE DETERMINACIÓN (R ²).....	4
MODELO DE REGRESIÓN Y CLASIFICACIÓN.....	4
PROCESAMIENTO DE DATOS.....	5
VARIABLES E INDICADORES.....	6
DESARROLLO DEL MODELO (DL).....	9
DESARROLLO DE CODIFICACIÓN DEL ALGORITMO (AP).....	9
RESULTADOS Y DISCUSIÓN.....	12
CONCLUSIÓN	14
RECOMENDACIÓN.....	15
REFERENCIAS	16

ÍNDICE DE FIGURAS

Figura 1	Ciclo biológico del camarón	7
Figura 2	Diagrama de flujo de algoritmo de DL (Autor, 2024)	9
Figura 3	Codificación para cargar datos desde archivo CSV	10
Figura 4	Graficas de dispersión de data obtenida del archivo CSV.....	10
Figura 5	Creación de DataFrame para ingreso de parámetros críticos	11
Figura 6	Codificación para obtener la predicción de los valores ingresados en el DataFrame.....	11
Figura 7	Predicciones de supervivencia con valores de los datos crítico-ingresados	12
Figura 8	Entrenamiento de red neuronal	13

ÍNDICE DE TABLAS

Tabla 1	Parámetros de lectura para evaluar el rendimiento del algoritmo	4
Tabla 2	Tanques por sala y capacidad	8
Tabla 3	Parámetros óptimos para el desarrollo de larvas de camarón	8
Tabla 4	Ciclo de producción	6

INTRODUCCIÓN

El cultivo de camarón ha servido tradicionalmente como una fuente de alimentos marinos y su producción aporta ingresos económicos para los países costeros. Sin embargo, con el rápido desarrollo de la sociedad, la pesca artesanal, industrial y el cultivo manual de camarón a pequeña escala ya no pueden satisfacer la demanda del crecimiento demográfico acelerado (FAO, 2022). El cultivo inteligente de camarón representa una aplicación innovadora de tecnologías avanzadas y prácticas de gestión en la acuicultura de estos crustáceos para ampliar la escala de producción.

Este estudio se centra en el desarrollo de un modelo de Deep Learning (DL) o aprendizaje profundo (AP) a través de avances tecnológicos en análisis de datos y desarrollo de algoritmos de AP, logrando extraer patrones complejos que facilitan el camino para transformar el enfoque de esta tarea; posibilitando la identificación y evaluación de poblaciones de larvas de camarón durante sus estadios evolutivos utilizando únicamente parámetros relevantes que influyen en el desarrollo morfológico de los camarones como oxígeno, saturación y temperatura. (Vásquez-Quispesivana et al., 2022).

Mediante la recopilación integral de datos del crecimiento de las larvas de camarón, el sistema DL utiliza la información detallada y acumulada de cada parámetro de calidad del agua con el fin de proporcionar una estimación precisa y automática del desarrollo larvario, permitiendo a los productores tomar decisiones rápidas y adecuadas para optimizar las condiciones de crecimiento de las larvas de camarón según el algoritmo del sistema DL constituido por la recopilación y representación de datos, selección de algoritmo y optimización del modelo (Mouloodi et al., 2021).

El aporte proporcionado por el cultivo acuícola inteligente (CAI) es sustancial, tal como lo menciona puesto que esta herramienta se compromete a solucionar problemas de producción de camarones en las primeras etapas del proceso acuícola por medio del CAI, el cual hará uso de instrumentos y sensores para monitorear el rendimiento y crecimiento de estos crustáceos (Ren et al., 2020).

El sistema DL utiliza una base de datos masivos e información acumulada que permiten el entrenamiento del modelo, de esta forma el algoritmo o modelo pueda interpretar los patrones potenciales que se encuentran en la data para luego alcanzar el objetivo de emitir predicciones o sugerencias precisas cuando a este sistema se le ingrese más datos actualizados, dotándolo de un reentrenamiento automático constante (Song et al., 2021).

FUNDAMENTO TEÓRICO

RANDOMFOREST

Los bosques aleatorios son un esquema propuesto por Leo Breiman en la década de 2000 para construir un conjunto de predictores con un conjunto de árboles de decisión que crecen en subespacios de datos seleccionados aleatoriamente. A pesar del creciente interés y uso práctico, ha habido poca exploración de las propiedades estadísticas de los bosques aleatorios, y se sabe poco acerca de las fuerzas matemáticas que impulsan el algoritmo (Schonlau & Zou, 2020).

REDES NEURONALES

Existen diferentes tipos de redes neuronales como redes neuronales densas, redes neuronales recurrentes, redes neuronales convolucionales y entre otras arquitecturas. Dentro del aprendizaje profundo **KERAS** es la interfaz de alto nivel de TensorFlow, puesto que esta permite desarrollar, entrenar, evaluar y ejecutar las redes neuronales. La biblioteca original de Keras fue desarrollada por François Chollet como parte de un proyecto de investigación y se lanzó como un proyecto de código abierto independiente en marzo de 2015. Rápidamente ganó popularidad debido a su facilidad de uso, flexibilidad y diseño elegante (Hinton et al., 2023).

GOOGLE COLAB

Google Colab es una herramienta que permite ejecutar código Python por medio del navegador, ya que está diseñado para analizar datos, aprendizaje automático y educación. Colab es un servicio de nube gratuito alojado en Google, y pues de esta manera fomenta la investigación en el aprendizaje automático e inteligencia artificial.

Usar Google Colab posee un entorno de cuaderno Jupyter de manera gratuita, además permite usar GPUs y TPUs que se puede resolver este tipo de problemas. Además, se puede disponer de librerías como Pandas, Numpy, etc. Incluso las herramientas que se requieren para desarrollar aprendizaje profundo (Naik et al., 2022).

PYTHON

En 1991, el lenguaje Python (LP) fue desarrollado por Guido van Rossum. El LP posee ventajas sobre los otros lenguajes de programación, puesto que este tiene variedad de bibliotecas que permiten que el código se reduzca significativamente, por estas razones es que la programación en Python es conveniente para temas de aprendizajes automáticos (Dhruv et al., 2020).

ERROR CUADRÁTICO MEDIO (MSE)

Con MSE es posible calcular el error elevado al cuadrado, promediando el conjunto de datos que se obtiene a raíz de la diferencia entre el valor promedio y el valor real. Es decir, se obtiene la diferencia al cuadrado. Los algoritmos de optimización emplean este recurso ya que es de ayuda para hallar valores de parámetros óptimos (Rasheed Abdul Haq & Harigovindan, 2022).

$$MSE = \frac{1}{n} \sum_i^n (Y_{real_i} - Y_{pred_i})^2$$

Ecuación 1. Error cuadrático medio

Donde:

- La ecuación es una suma de la secuencia de números desde $i=1$ hasta n .
- $(Y_{real_i} - Y_{pred_i})^2$, es la diferencia entre los valores reales y estimados elevados al cuadrado.
- MSE nunca será negativo ya que los errores se elevan al cuadrado
- El valor del error varía de cero a infinito
- MSE aumenta exponencialmente con un aumento en el error

COEFICIENTE DE DETERMINACIÓN (R²)

El coeficiente de determinación (Ec 2) o distribución de Weibull es la métrica que permite evaluar el rendimiento del modelo entrenado, con la cual se puede validar la variación que posee la variable dependiente (VP) misma que es predecible desde la variable independiente (VI). Es decir, (R²) es la representación de ajuste del modelo con respecto a los datos reales que se posee, más sin embargo este coeficiente no valida si el modelo posee datos sesgados o si el método empleado es el correcto. Por lo tanto, el ajuste promedia los valores precedidos en función de la desviación de los datos reales (Coeficiente de determinación (R²)).

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_{real_i} - Y_{pred_i})^2}{\sum_{i=1}^n (Y_{real_i} - Y'_{real_i})^2}$$

Ecuación 2. Distribución de Weibull

- Y_{real_i} : valores de la variable dependiente
- Y_{pred_i} : valores predichos
- Y'_{real_i} : la media de Y_{real_i}

En R² valores más elevados determinan un ajuste fuerte y por el contrario los valores bajos determinan un ajuste débil del modelo empleado como se puede visualizar las especificaciones en la **Tabla 1**.

Tabla1

Parámetros de lectura para evaluar el rendimiento del algoritmo

PARÁMETRO	INTERPRETACIÓN
$R^2 < 0,5$	Ajuste débil
$0,5 \leq R^2 \leq 0,8$	Ajuste medio
$R^2 > 0,8$	Ajuste fuerte

MODELO DE REGRESIÓN Y CLASIFICACIÓN

La regresión lineal supone que la relación entre dos variables tiene una forma lineal (o linealizable mediante alguna transformación de las variables). La regresión lineal tiene una versión “simple” que empareja dos variables, pero esta suele ser insuficiente para entender

fenómenos mínimamente complejos en la que influyen más de dos variables, esta versión es la “múltiple”. En el modelo de regresión lineal múltiple supone que más de una variable tiene influencia o está correlacionada con el valor de una tercera variable. Por ejemplo, en el peso de una persona pueden influir edad, género y estatura, en la renta pueden influir trabajo, capital físico, conocimientos, etc. En el modelo de regresión lineal múltiple se espera que los sucesos tengan una forma funcional (Rugeles Ospina, 2022).

Los modelos de regresión lineal múltiple surgen cuando se trata de explicar el comportamiento de una variable predictora de tipo continuo a través de una variable predictora de tipo continuo mediante una función polinómica lineal. En general, los modelos polinómicos son útiles cuando se aprecia una tendencia curvilínea entre los predictores y la respuesta. Asimismo, a veces constituyen una aproximación sencilla (por serie de Taylor) a modelos complejos e incluso no-lineales (Regresión Lineal Múltiple y Polinómica | Modelos Estadísticos).

$$Y = \beta_0 + \beta_1 \cdot T_p + \beta_2 \cdot S_a + \beta_3 \cdot O_2 + \epsilon$$

Ecuación 3. Regresión de variables múltiples

donde:

- Y variable dependiente (supervivencia).
- β_0 término constante.
- $\beta_1, \beta_2, \beta_3$ coeficientes asociados con las variables predictoras (T_p , S_a y O_2).
- ϵ término de error.

PROCESAMIENTO DE DATOS

Una vez descargada la información desde la base de datos que tiene la empresa, para potenciar el valor de la información, se emplearon técnicas de preprocesamiento a los datos recopilados, donde eliminar posibles anomalías permite garantizar la coherencia de la información. Esto da la oportunidad de ajustar los valores e identificar los posibles vacíos que afecten la calidad del modelo de aprendizaje profundo.

VARIABLES E INDICADORES

Para el cultivo de organismos vivos acuáticos es importante la calidad de agua, puesto que estos requieren de varios parámetros esenciales como el oxígeno disuelto, pH, temperatura, nitratos, etc. Si se quiere lograr una producción de calidad es preciso el monitoreo continuo de los parámetros en el agua, con el objetivo de garantizar el espacio acuático con los parámetros óptimos para el buen desarrollo y un porcentaje de supervivencia alto (Kumar et al., 2021).

Los parámetros críticos O₂, T_p y S_a, son indicadores principales dentro de la calidad de agua, puesto que dichos parámetros críticos brindan información importante de la capacidad del agua para el buen desarrollo de larvas de camarón y mantener procesos ecológicos fundamentales

El oxígeno disuelto (O₂) es indispensable para la supervivencia de estos organismos acuáticos dando paso a la ejecución de funciones biológicas básicas. La T_p influye directamente en el comportamiento y biología de las larvas de camarón, puesto que al garantizar una T_p estable, el desarrollo de esta especie se realiza con uniformidad. Por otra parte, la Saturación (S_a) es un parámetro de suma importancia ya que por medio de este indicador se puede constatar la capacidad del agua para mantener la vida y desarrollo de los camarones en sus primeras etapas larvarias.

Por lo tanto, la selección de O₂, T_p y S_a como parámetros principales, se sustentan el papel fundamental que desempeñan para la salud y sostenibilidad de los ecosistemas acuáticos.

Las etapas de desarrollo de las larvas de camarón en cada tanque se detallan en la **Tabla 2**.

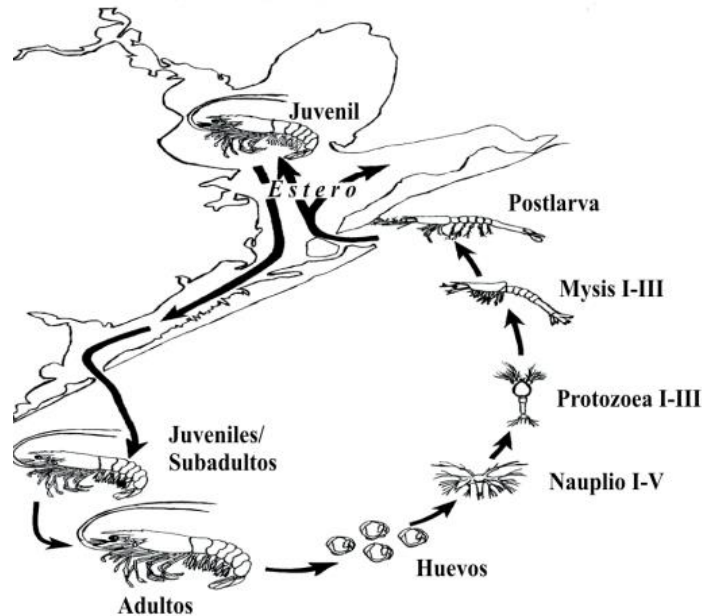
Tabla 2 *Ciclo de producción*

Tiempo	Etapas
40-50 horas	Nauplios
4 - 5 días	Zoea
3 - 4 días	Mysis
10 - 15 días	Postlarvas

Las diferentes etapas de las larvas se podrán apreciar con detalle en la **Figura 1**.

Figura 1

Ciclo biológico del camarón(Ciclo Biológico de Los Camarones).



MÉTODOS Y MATERIALES

RECOPIACIÓN DE DATOS

Para el desarrollo del modelo de DL, se recopilaron y clasificaron datos que afectan directamente al desarrollo morfológico de las larvas de camarón según su etapa de desarrollo. Las variables que se tomaron en cuenta fueron el nivel de oxígeno disuelto, Saturación y Tp siendo estos parámetros de calidad de agua más importantes en el desarrollo de las larvas de camarón. Estos datos fueron obtenidos del Laboratorio especializado en la producción de larvas de Camarón BIOGEMAR S.A-Grupo Almar.

La obtención de la data se realizó de forma manual y directa en el área de estudio, en la **Tabla 3** se observa detalladamente la información del área de estudio, donde tomó la lectura de los parámetros con intervalos de 2 horas y haciendo uso del equipo portátil YSIPro20i, con el que al introducir la sonda al tanque se visualiza los valores en la pantalla led del equipo portátil, mismo que permite almacenar los datos en la memoria interna de forma temporal.

Tabla 3*Tanques por sala y capacidad (representa tu área de estudio).*

Salas	Cantidad de tanques por Sala	Capacidad en m ³ por tanque
Sala 2		
Sala 3		
Sala 4	12 tanques	50
Sala 5		
Sala 6		
Sala 7		
Sala 8		
Sala 9	14 tanques	50
Sala 10		
Sala 11		

Cada tanque cuenta con su sistema de aireación y Tp, mismos que se extiende a lo largo del tanque con el objetivo de garantizar la homogeneidad de estas variables en todo el sistema. Posteriormente los valores fueron digitados en la base de datos de la empresa donde se detalla la información de los parámetros físicos químicos, al visualizar la **Tabla 4** encontrará el rango de parámetros óptimos para el desarrollo de larvas de camarón.

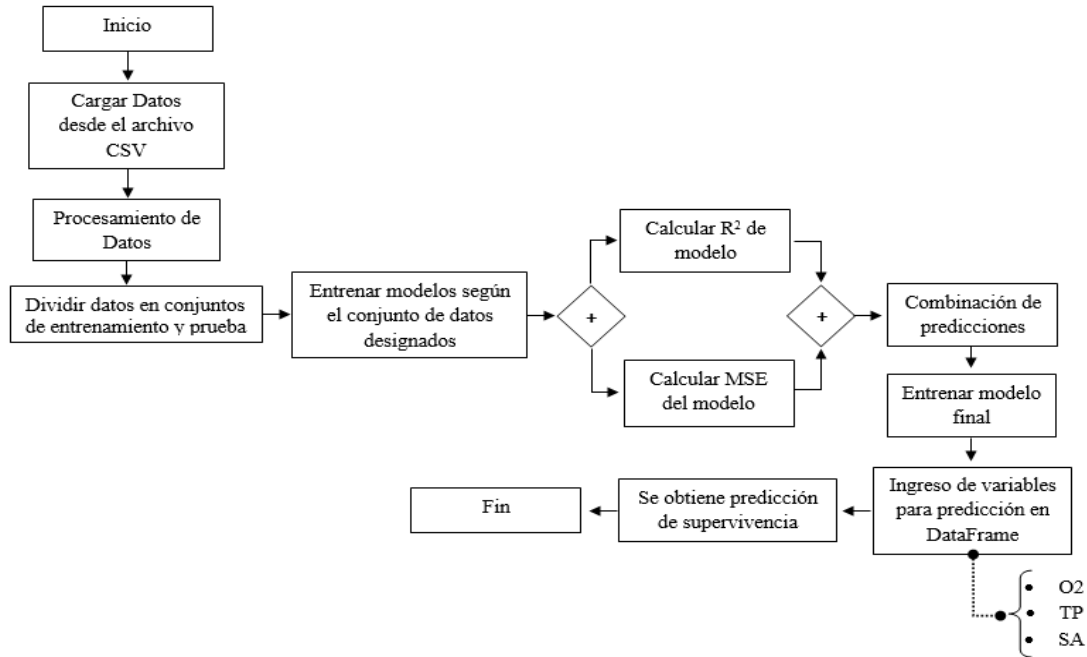
Tabla 4*Parámetros óptimos para el desarrollo de larvas de camarón*

Parámetros	Rangos óptimos
Temperatura	31-34 °C
Oxígeno	4.5- 6 mg/l
Saturación	75% - 100%

DESARROLLO DEL MODELO (DL)

Figura 2

Diagrama de flujo de algoritmo de DL (Autor, 2024)



El desarrollo del algoritmo DL facilita la entrada de los valores de los parámetros críticos de estudio, permitiendo así la obtención de predicciones de supervivencia. Al ingresar los valores críticos, el algoritmo procesa los datos mediante su modelo entrenado en DL, asociando o agrupando los datos con valores cercanos a los ingresados, de modo que la varianza pueda ser fácilmente procesada. Esto desencadena una aproximación predictiva en la variable dependiente, es decir, la supervivencia de las larvas, considerando los parámetros críticos de estudio.

DESARROLLO DE CODIFICACIÓN DEL ALGORITMO (AP)

Para el desarrollo del algoritmo de predicción de supervivencia, se empleó un enfoque de regresión múltiple con las variables de Tp, Sa, y O2 identificadas como los parámetros claves predictores y posteriormente la implementación de una red neuronal con arquitectura simple con una capa oculta y una capa de salida. Inicialmente se cargaron los conjuntos de datos de Tp, Sa, y O2 desde un archivo CSV y empleando bibliotecas de pandas. Posteriormente los datos se dividen en dos conjuntos, uno de entrenamiento y otro de prueba.

En la **Figura 3** se puede observar la codificación empleada para adquirir y procesar la data.

Figura 3

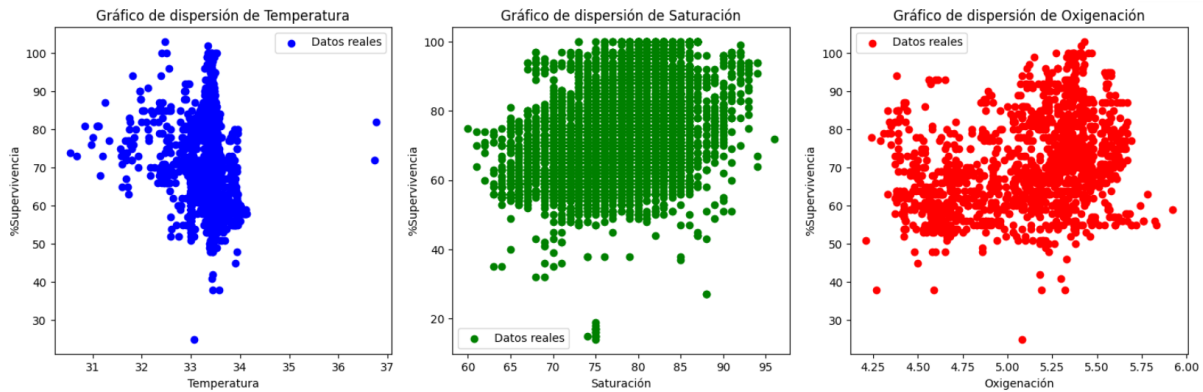
Codificación para cargar datos desde archivo CSV

```
# Cargar y procesar datos para cada conjunto
X_temp, y_temp = cargar_y_procesar_datos(url_temperatura, 'Estadio', 'Promedio', '%Supervivencia')
X_saturacion, y_saturacion = cargar_y_procesar_datos(url_saturacion, 'Estadio', 'Promedio', '%Supervivencia')
X_oxigenacion, y_oxigenacion = cargar_y_procesar_datos(url_oxigenacion, 'Estadio', 'Promedio', '%Supervivencia')
```

En la **Figura 4** se puede observar graficada la data real de los parámetros críticos.

Figura 4

Graficas de dispersión de data obtenida del archivo CSV



En la gráfica de dispersión de Temperatura vs Supervivencia se observa que al graficar una curva de tendencia la pendiente de esta sería infinita, ya que la variable de temperatura si bien es cierto se relaciona directamente con el desarrollo y supervivencia de las larvas de camarón, fuera del rango optimo no es posible (31°C-34°C) no es posible la vida para estos organismos acuáticos. Por otra parte, en la gráfica de Saturación vs Supervivencia se puede visualizar una data extensa que con ayuda del algoritmo se podrá determinar una tendencia o aproximado estándar que consideré la información reflejada en la imagen.

La Oxigenación Vs Supervivencia visualmente se puede determinar una relación más directa y que su tendencia podría es posible considerar una función polinómica, que se ajuste aproximadamente al histórico obtenido en las áreas de estudio. Entrenar el modelo DL se realiza con el conjunto de datos designado, tomando en consideración la evaluación de este incorporando las métricas de evaluación como lo es MSE y R2.

Se combinan predicciones de Tp, Sa y O2 en un DataFrame y posteriormente se entrena el modelo final. Se proporciona una interfaz para los usuarios ingresen valores de Tp, Sa y O2 donde luego se realiza la predicción final. En la **Figura 5** se observa el código para la creación de DataFrame.

Figura 5

Creación de DataFrame para ingreso de parámetros críticos

```
# Crear un DataFrame con las predicciones de cada conjunto
df_predicciones = pd.DataFrame({
    'Temperatura': y_temp_pred,
    'Saturacion': y_saturacion_pred,
    'Oxigenacion': y_oxigenacion_pred
})
```

En la **Figura 6** se puede observar la codificación empleada para el ingreso de parámetros críticos y Estadio, para posteriormente el algoritmo DL ejecute la predicción.

Figura 6

Codificación para obtener la predicción de los valores ingresados en el DataFrame

```
# Obtener predicciones para valores de entrada personalizados y ajustar el porcentaje de supervivencia
estadio_input = input("Ingresa el valor de Estadio: ")
temperatura_input = float(input("Ingresa el valor de temperatura: "))
saturacion_input = float(input("Ingresa el valor de saturacion: "))
oxigenacion_input = float(input("Ingresa el valor de oxigenacion: "))
```

RESULTADOS Y DISCUSIÓN

El algoritmo desarrollado se enfocó en crear un modelo de regresión predictivo haciendo uso de datos de oxigenación, temperatura y saturación para obtener un porcentaje de supervivencia en el área de la acuicultura. Se inició con la recolección de datos los mismos que fueron procesados para cada parámetro crítico específico. Luego, se aplicaron técnicas de regresión para obtener un modelo predictivo confiable y poder modelar la relación entre los parámetros críticos que afecta directamente la supervivencia de las larvas de camarón.

Para evaluación del rendimiento del modelo se empleó dos métricas cruciales como R^2 donde se obtuvo el valor de **18.0077** y MSE donde se obtuvo **0.08631**. El error cuadrático medio cuantifica los errores entre las predicciones del modelo y los valores reales de supervivencia, donde un MSE bajo indica que las predicciones concuerdan con los datos de prueba, es decir una mejor predicción del modelo. De igual manera el R^2 permite identificar la variabilidad en la variable dependiente, resultado obtenido por el modelo. Para (Gonzalez, 2022) el implementar R^2 y MSE en su algoritmo en el USO DE MACHINE LEARNING PARA LA PREDICCIÓN DE RESULTADOS DE LAS PRUEBAS SABER 11 EN EL INSTITUTO SAN RICARDO PAMPURI, les permitió seleccionar el modelo de árbol de regresión, al evidenciar que las métricas de evaluación eran más bajas en uno de los dos modelos elaborados, es decir que la predicción de uno de los modelos era más eficiente.

En la **Figura 7** se puede visualizar la predicción entregada por el algoritmo DL una vez sea ingresados el Estadio y los valores críticos para que sea posible el desarrollo de las larvas de camarón.

Figura 7

Predicciones de supervivencia con valores de los datos crítico-ingresados

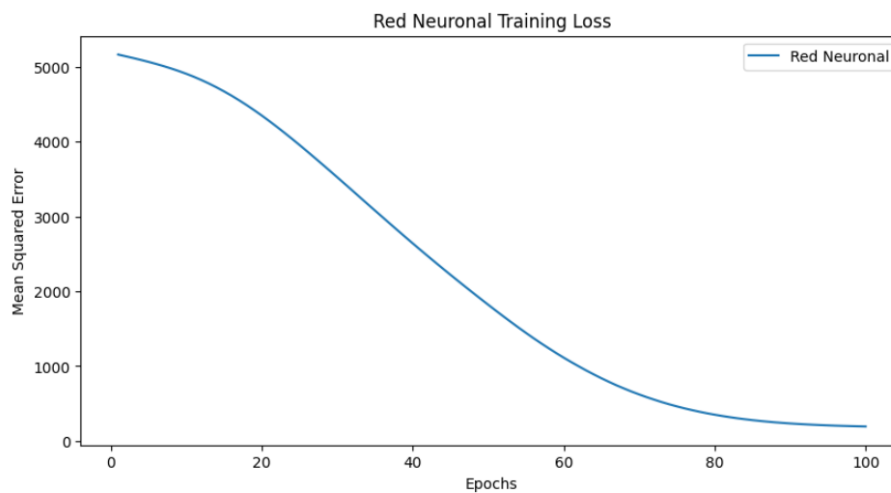
```
Error cuadrático medio del modelo final (RandomForest): 18.00774245614035
Coeficiente de determinación (R^2) del modelo final (RandomForest): 0.8631445635130859
9/9 [=====] - 0s 2ms/step
Coeficiente de determinación (R^2) del modelo final (Red Neuronal): -0.461463380623212
Ingresa el valor de Estadio: Postlarva 1
Ingresa el valor de temperatura: 33.7
Ingresa el valor de saturacion: 76
Ingresa el valor de oxigenacion: 5
1/1 [=====] - 0s 21ms/step
Predicción final (Red Neuronal ajustada con RandomForest): 77.55893549804686
```

En la **Figura 8** se observa la **Pérdida de Entrenamiento de la Red Neuronal** (Red Neuronal Training Loss) del algoritmo DL. Mediante las épocas representadas en el eje X se puede validar las interacciones completas del conjunto de datos de entrenamiento. Por otra parte, la pérdida se visualiza en el eje Y, siendo esta representada por la diferencia de las predicciones de la red y los datos reales.

La disminución de la pérdida se visualiza a medida que la red neuronal aprende de los patrones en los datos designados para el entrenamiento. La estabilidad de la gráfica indica que la red ha logrado alcanzar su máxima capacidad de aprendizaje usando los datos de entrenamiento.

Figura 8

Entrenamiento de red neuronal



CONCLUSIÓN

Con la recopilación de datos y el procesamiento adecuado de los mismos, permitió que sea posible la integración de técnicas y herramientas de programación logrando así el aprendizaje supervisado como son las regresiones múltiples variables, que da paso a las reacciones complejas entre las variables críticas usadas, que se complementaron con la integración de una red neuronal en el desarrollo del proyecto. En futuras investigaciones, incorporar modelos de DL pueden llegar a brindar una alternativa para de mayor complejidad y precisión en el desarrollo y comportamiento de organismos vivos en el área de la acuicultura en conjunto de las herramientas tecnológicas que con el paso de los días van mejoran, extendiendo así infinitas alternativas para emplear, según el enfoque investigativo o el resultado que se desea obtener.

Para el desarrollo del proyecto se estableció priorizar la evaluación de los parámetros críticos de calidad de agua, ya que estos permiten comprender completamente como estos factores impactan de forma directa la vida en el área de la acuicultura. Se implementó una base de datos almacenada en la plataforma de Github para gestionar rapidez, accesibilidad y garantizar la integridad de los datos recopilados. La creación del algoritmo DL desarrollado al evaluar su rendimiento, resultó ser eficaz en las predicciones de porcentaje de supervivencia de larvas de camarón en función de los datos críticos empleados siendo la Tp, Sa y O2 en un entorno acuático.

Así mismo la incorporación de condiciones de ajustes (rangos óptimos) para supervivencia, añade robustez al modelo realizado, ya que permite tomar en cuenta valores óptimos de desarrollo para las larvas de camarón.

Por lo tanto, en este trabajo investigativo fue posible la gestión eficaz de los datos recopilados y almacenados en la nube, siendo estos la base para el desarrollo del algoritmo DL que den paso a la comprensión y predicción de la supervivencia en acuicultura, con un aporte significativo a la optimización de los recursos productivos obteniendo como resultado soluciones efectivas e inmediatas.

RECOMENDACIÓN

Se recomienda adicionar el tratamiento de imágenes al proyecto investigativo, ya que la incorporación del análisis de imágenes contribuía con una perspectiva visual del desarrollo de las larvas de camarón, lo cual reforzaría significativamente a la precisión de los resultados a obtener.

Es recomendable considerar en el estudio la integración de la dieta suministrada a las larvas de camarón durante su periodo de desarrollo. Uno de los aspectos críticos para el crecimiento y supervivencia de las larvas de camarón es la alimentación suministrada. Determinar un protocolo detallado meticulosamente con las cantidades y horarios de alimentación para el suministro de alimentos, pueden brindar resultados de cómo afecta esto en el crecimiento y desarrollo.

Combinar un análisis de imágenes con un régimen de alimentación suministrada tallada, brindará una comprensión más completa de los factores que influyen en el desarrollo de los nauplios. Implementar esta mejora será un aporte significativo a la calidad de los datos obtenidos, además proporciona información específica para futuras prácticas e investigaciones dentro de la acuicultura.

REFERENCIAS

Ciclo biológico de los camarones peneidos. Este ciclo cambia... | Download Scientific Diagram. (s. f.). Recuperado 6 de marzo de 2024, de https://www.researchgate.net/figure/Figura-1-Ciclo-biologico-de-los-camarones-peneidos-Este-ciclo-cambia-dependiendo-de-la_fig1_292753966

Dhruv, A. J., Patel, R., & Doshi, N. (s. f.). *Python: The Most Advanced Programming Language for Computer Science Applications.* <https://doi.org/10.5220/0010307900003051>

El estado mundial de la pesca y la acuicultura 2022. (2022). *El estado mundial de la pesca y la acuicultura 2022.* <https://doi.org/10.4060/CC0461ES>

Gonzalez, C. O. (2022). *USO DE MACHINE LEARNING PARA LA PREDICCIÓN DE RESULTADOS DE.*

Hinton, G., Learning, D., Learning, D., Learning, M., Learning, M., Learning, M., Learning, M., & Learning, W. M. (2023). *Hands-on machine learning with Scikit-Learn, Keras and TensorFlow : concepts, tools, and techniques to build intelligent systems.* xxv, 834 pages : <https://www.oreilly.com/library/view/hands-on-machine-learning/9781098125967/>

Kumar, L. A., Jayashree, M., Venkata, A., Abhishek, S., Abhinav, A., & Prasad, G. S. (2021). Aquaculture Monitoring System using IoT. *International Research Journal of Engineering and Technology.* www.irjet.net

Mouloodi, S., Rahmanpanah, H., Gohari, S., Burvill, C., Tse, K. M., & Davies, H. M. S. (2021). What can artificial intelligence and machine learning tell us? A review of applications to equine biomechanical research. *Journal of the Mechanical Behavior of Biomedical Materials*, 123, 104728. <https://doi.org/10.1016/J.JMBBM.2021.104728>

Naik, P., Naik, G., & M.B.Patil, M. (2022). *Conceptualizing Python in Google COLAB.* January. https://www.researchgate.net/publication/357929808_Conceptualizing_Python_in_Google_COLAB

Rasheed Abdul Haq, K. P., & Harigovindan, V. P. (2022). Water Quality Prediction for Smart Aquaculture Using Hybrid Deep Learning Models. *IEEE Access*, *10*, 60078-60098. <https://doi.org/10.1109/ACCESS.2022.3180482>

Ren, Q., Wang, X., Li, W., Wei, Y., & An, D. (2020). Research of dissolved oxygen prediction in recirculating aquaculture systems based on deep belief network. *Aquacultural Engineering*, *90*, 102085. <https://doi.org/10.1016/J.AQUAENG.2020.102085>

Rugeles Ospina, N. (2022). *Algoritmos de Boosting para Modelos de Clasificación y Regresión Lineal*.

Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *Stata Journal*, *20*(1), 3-29. https://doi.org/10.1177/1536867X20909688/ASSET/IMAGES/10.1177_1536867X20909688-IMG8.PNG

Song, Y., Zheng, S., Li, L., Zhang, X., Zhang, X., Huang, Z., Chen, J., Wang, R., Zhao, H., Chong, Y., Shen, J., Zha, Y., & Yang, Y. (2021). Deep Learning Enables Accurate Diagnosis of Novel Coronavirus (COVID-19) with CT Images. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *18*(6), 2775-2780. <https://doi.org/10.1109/TCBB.2021.3065361>

Unidad 3 Regresión Lineal Múltiple y Polinómica | Modelos Estadísticos. (s. f.). Recuperado 6 de marzo de 2024, de https://bookdown.org/j_morales/librostat/rlm.html

Vásquez-Quispesivana, W., Inga, M., Betalleluz-Pallardel, I., La, A., Molina, M. A. La, Molina, L., & Perú, L. (2022). Inteligencia artificial en acuicultura: fundamentos, aplicaciones y perspectivas futuras. *Scientia Agropecuaria*, *13*(1), 79-96. <https://doi.org/10.17268/sci.agropecu.2022.008>

Vista de Análisis crítico del coeficiente de determinación (R²), como indicador de la calidad de modelos lineales y no lineales. (s. f.). Recuperado 6 de marzo de 2024, de <http://www.revistas.espol.edu.ec/index.php/matematica/article/view/1037/917>