

# UNEMI

UNIVERSIDAD ESTATAL DE MILAGRO

REPÚBLICA DEL ECUADOR

UNIVERSIDAD ESTATAL DE MILAGRO

VICERRECTORADO DE INVESTIGACIÓN Y POSGRADO

PROYECTO DE INVESTIGACIÓN Y DESARROLLO PREVIO A LA OBTENCIÓN  
DEL TÍTULO DE:

MAGÍSTER EN SISTEMA DE INFORMACIÓN CON MENCIÓN EN  
INTELIGENCIA DE NEGOCIOS Y ANALÍTICA DE DATOS

TEMA:

DESARROLLO DEL PROCESO DE CLASIFICACIÓN EN GRANOS DE ARROZ  
(*Oryza Sativa L*) UTILIZANDO TÉCNICAS DE APRENDIZAJE AUTOMÁTICO

Autor:

Econ. Agrop. José Julián Coronel Reyes

Director:

Ing. Manuel Andrés Avilés Noles, MSc

*Milagro, Enero 2023*

## Derechos de autor

Sr. Dr.

**Fabrizio Guevara Viejó**

Rector de la Universidad Estatal de Milagro

Presente.

Yo, **José Julián Coronel Reyes** en calidad de autor y titular de los derechos morales y patrimoniales de este informe de investigación, mediante el presente documento, libre y voluntariamente cedo los derechos de Autor de este proyecto de desarrollo, que fue realizada como requisito previo para la obtención de mi Grado, de **Magíster en Sistema de Información con mención en Inteligencia de negocios y Analítica de Datos Masivos**, como aporte a la Línea de Investigación **Algoritmos Evolutivos multiobjetivo** de conformidad con el Art. 114 del Código Orgánico de la Economía Social de los Conocimientos, Creatividad e Innovación, concedo a favor de la Universidad Estatal de Milagro una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos. Conservo a mi favor todos los derechos de autor sobre la obra, establecidos en la normativa citada.

Así mismo, autorizo a la Universidad Estatal de Milagro para que realice la digitalización y publicación de este Proyecto de Investigación en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

El autor declara que la obra objeto de la presente autorización es original en su forma de expresión y no infringe el derecho de autor de terceros, asumiendo la responsabilidad por cualquier reclamación que pudiera presentarse por esta causa y liberando a la Universidad de toda responsabilidad.

Milagro, 11 de enero de 2023



Firmado electrónicamente por:  
**JOSE JULIAN CORONEL  
REYES**

**José Julián Coronel Reyes**

C.I: 07046565387

## Aprobación del director del Trabajo de Titulación

Yo, **Manuel Andrés Avilés Noles** en mi calidad de director del trabajo de titulación, elaborado por **José Julián Coronel Reyes**, cuyo tema es **DESARROLLO DEL PROCESO DE CLASIFICACIÓN EN GRANOS DE ARROZ (*Oryza Sativa L*) UTILIZANDO TÉCNICAS DE APRENDIZAJE AUTOMÁTICO**, que aporta a la Línea de Investigación Algoritmos Evolutivos multiobjetivo, previo a la obtención del Grado Magister en Sistema de Información con mención en Inteligencia de negocios y Analítica de Datos Masivos. Trabajo de titulación que consiste en una propuesta innovadora que contiene, como mínimo, una investigación exploratoria y diagnóstica, base conceptual, conclusiones y fuentes de consulta, considero que el mismo reúne los requisitos y méritos necesarios para ser sometido a la evaluación por parte del tribunal calificador que se designe, por lo que lo **APRUEBO**, a fin de que el trabajo sea habilitado para continuar con el proceso de titulación de la alternativa de Informe de Investigación de la Universidad Estatal de Milagro.

Milagro, 11 de enero de 2023



Firmado electrónicamente por:  
**MANUEL ANDRES  
AVILES NOLES**

**Manuel Andrés Avilés Noles**

C.I: 0920574308

## VICERRECTORADO DE INVESTIGACIÓN Y POSGRADO

### DIRECCIÓN DE POSGRADO

### ACTA DE SUSTENTACIÓN

### MAESTRÍA EN SISTEMAS DE INFORMACIÓN

En la Dirección de Posgrado de la Universidad Estatal de Milagro, a los diecisiete días del mes de febrero del dos mil veintitres, siendo las 11:00 horas, de forma VIRTUAL comparece el/la maestrante, ECON. CORONEL REYES JOSE JULIAN, a defender el Trabajo de Titulación denominado " **DESARROLLO DEL PROCESO DE CLASIFICACIÓN EN GRANOS DE ARROZ (ORYZA SATIVA L) UTILIZANDO TÉCNICAS DE APRENDIZAJE AUTOMÁTICO**", ante el Tribunal de Calificación integrado por: Ph.D. RODAS SILVA JORGE LUIS, Presidente(a), Mgti. MENDOZA CABRERA DENIS DARIO en calidad de Vocal; y, Ph. D. DIAZ MONTENEGRO JOSE ALBERTO que actúa como Secretario/a.

Una vez defendido el trabajo de titulación; examinado por los integrantes del Tribunal de Calificación, escuchada la defensa y las preguntas formuladas sobre el contenido del mismo al maestrante compareciente, durante el tiempo reglamentario, obtuvo la calificación de: **99.33** equivalente a: **EXCELENTE**.

Para constancia de lo actuado firman en unidad de acto el Tribunal de Calificación, siendo las 12:00 horas.



Firmado electrónicamente por:  
**JORGE LUIS  
RODAS  
SILVA**

---

Ph.D. RODAS SILVA JORGE LUIS  
**PRESIDENTE/A DEL TRIBUNAL**



Firmado electrónicamente por:  
**DENIS DARIO MENDOZA  
CABRERA**

---

Mgti. MENDOZA CABRERA DENIS DARIO  
**VOCAL**



Firmado electrónicamente por:  
**JOSE ALBERTO DIAZ  
MONTENEGRO**

---

Ph. D. DIAZ MONTENEGRO JOSE ALBERTO  
**SECRETARIO/A DEL TRIBUNAL**



Firmado electrónicamente por:  
**JOSE JULIAN CORONEL  
REYES**

---

ECON. CORONEL REYES JOSE JULIAN  
**MAGÍSTER**

## AGRADECIMIENTOS

En primer lugar, agradezco a Dios, por poner en mi camino a personas y momentos que han hecho de este viaje llamado vida, una gran experiencia. De manera especial a mis padres y mis hermanos que me brindaron un hogar lleno de amor y confianza, sus enseñanzas basadas en principios y valores, me han dado desde siempre las herramientas para avanzar en el marco del respeto hacia los demás.

A la Universidad Estatal de Milagro, por permitirme ser parte de tan prestigiosa institución para prepararme y capacitarme.

Los resultados que se exponen en el presente trabajo, no hubiesen sido alcanzados sin la guía de mis tutores, quienes compartieron sin límite, sus conocimientos y experiencia. A ustedes gracias porque jamás existió un día o una hora que fuese inadecuada para resolver mis dudas, jamás hubo tema alguno del que no se pueda hablar, por el contrario, me brindaron más que únicamente tutorías, me dieron su amistad.

Siento la satisfacción de haber cumplido un compromiso que asumí al iniciar mi investigación de maestría, que estuvo llena de retos. Este logro quiero recibirlo con humildad, y dedicarlo a todas las personas que me aportaron para hacerlo realidad.

¡Muchas gracias a todos!

## Resumen

El trabajo de titulación plantea como objetivo general de investigación, aplicar las diferentes técnicas de aprendizaje automático en la clasificación de variedades de granos en arroz. La calidad del grano de arroz se determina habitualmente mediante inspecciones visuales y mediciones manuales, método que es lento, subjetivo y propenso a errores humanos. Por lo tanto, la industria demanda una técnica rápida y precisa que pueda clasificar el grano de arroz a bajo costo y estandarizarlo. Por medio del aprendizaje automático se puede desarrollar modelos de decisión en entornos eminentemente complejos y no lineales. Este estudio se enmarca en un diseño de investigación cuantitativa, consiste en la recolección de datos numéricos para su posterior análisis e interpretación utilizando herramientas de análisis matemático y estadístico para describir y explicar fenómenos mediante datos numéricos. El tipo de la investigación es causal comparativa dado que se comparan diferentes técnicas de aprendizaje automático en el proceso de clasificación de granos de arroz. La clasificación fue binaria, se consideró como población de estudio a los 3.810 granos de dos variedades del Instituto Nacional de Investigación Agropecuaria 14 (2.180) y 17 (1.630). Las técnicas que fueron empleadas están basadas en cinco modelos utilizando los siguientes algoritmos: regresión logística, Perceptrón Multicapa, Máquina de Vectores de Apoyo, Bosque Aleatorio y K Vecino Más Cercano. Las mediciones estadísticas de la matriz de confusión como resultado de la clasificación se utilizaron como métricas de rendimiento. Los resultados permiten concluir que el modelo regresión logística es el mejor método de clasificación, debido a su mejor predicción de los valores obtenidos de verdaderos positivos, y verdaderos negativos, además, la media de la precisión del modelo fue de 93,02%, superior a los otros modelos. El desarrollo de este trabajo permite llegar a la conclusión de que el uso del aprendizaje automático en la industria arrocera, es aplicable para el apoyo en la toma las decisiones basadas en el desarrollo del proceso de clasificación del grano.

**Palabras claves:** Clasificación del arroz; aprendizaje automático, matriz de confusión, validación cruzada, métricas de rendimiento.

## Abstract

The general objective of this research is to apply different machine learning techniques to the classification of rice grain varieties. Rice grain quality is usually determined by visual inspections and manual measurements, a method that is slow, subjective and prone to human error. Therefore, the industry demands a fast and accurate technique that can grade rice grain at low cost and standardize it. By means of machine learning, decision models can be developed in eminently complex and nonlinear environments. This study is framed within a quantitative research design, consisting of the collection of numerical data for subsequent analysis and interpretation using mathematical and statistical analysis tools to describe and explain phenomena using numerical data. The type of research is causal comparative since different machine learning techniques are compared in the rice grain classification process. The classification was binary, and the study population was considered to be 3.810 grains of two varieties of the National Institute of Agricultural Research 14 (2,180) and 17 (1,630). The techniques that were employed are based on five models using the following algorithms: logistic regression, Multilayer Perceptron, Support Vector Machine, Random Forest and K Nearest Neighbor. Statistical measures of the confusion matrix as a result of classification were used as performance metrics. The results allow concluding that the logistic regression model is the best classification method, due to its better prediction of the values obtained from true positives, and true negatives, in addition, the average accuracy of the model was 93.02%, higher than the other models. The development of this work leads to the conclusion that the use of machine learning in the rice industry is applicable to support decision making based on the development of the grain classification process.

**Keywords:** Rice classification; machine learning, confusion matrix, cross-validation, performance metrics.

## Índice / Sumario

Derechos de autor.....	i
Aprobación del director del Trabajo de Titulación .....	ii
Aprobación del tribunal calificador .....	iii
AGRADECIMIENTOS .....	iv
Resumen .....	v
Abstract.....	vi
Introducción.....	1
CAPÍTULO I: El problema de la investigación.....	3
1.1 Planteamiento del problema .....	3
1.2 Objetivo general.....	5
1.3 Objetivos específicos .....	5
1.4 Justificación .....	5
CAPÍTULO II: Marco teórico referencial.....	7
2.1. Conceptos generales .....	7
2.1.1. El cultivo de arroz.....	7
2.1.2. Producción de arroz en Ecuador .....	8
2.1.3. Etapas del proceso de producción de arroz .....	8
2.1.4. Clasificación de granos de arroz .....	9
2.1.5. Escalas de clasificación del arroz.....	10
2.1.6. Determinación del grado .....	10
2.2. Aprendizaje Automático.....	11
2.2.1. Uso del aprendizaje automático en la agricultura .....	13
2.2.2. Categorías del aprendizaje automático supervisado.....	15
2.2.3. Algoritmos de aprendizaje supervisado.....	15
2.2.4. K-Nearest Neighbor (k-NN) .....	16
2.2.5. Máquina Soporte Vectorial (SVM).....	16
2.2.6. Bosque aleatorio (RF) .....	17

2.2.7. Regresión logística (LR).....	18
2.2.8. Perceptrón multicapa (MLP).....	18
2.3. Validación cruzada .....	19
2.4. Medidas de desempeño .....	20
CAPÍTULO III: Diseño metodológico .....	21
3.1. Tipo y diseño de investigación .....	21
3.2. La población y muestra.....	21
3.2.1. Características de la población.....	21
3.2.2. Delimitación de la población.....	22
3.3. Los métodos y las técnicas .....	24
3.4. Procesamiento estadístico de la información .....	25
CAPÍTULO IV: Análisis e interpretación de resultados.....	28
4.1. Estadística descriptiva .....	28
4.2. Matriz de complejidad .....	29
4.3. Medidas de rendimiento .....	31
4.4. Discusión de resultados .....	32
CAPÍTULO V: Conclusiones y Recomendaciones .....	35
5.1. Conclusiones.....	35
5.2. Recomendaciones.....	36
Bibliografía.....	37

## Introducción

El arroz (*Oryza sativa L*) es una fuente alta de nutrientes en la dieta humana. El cereal más importante y más cultivado después del trigo y el maíz en todo el mundo es el arroz (Krishna et al., 2022; Sultana et al., 2022). Del mismo modo, también desempeña un papel relevante en la agricultura del Ecuador.

La presentación de los granos se determina por las características físicas, el tamaño, el color, la salud interna y la diversidad aumenta el valor de mercado.

La clasificación manual de la variedad de arroz es un proceso difícil. Además, este método requiere tiempo y es ineficiente, sobre todo cuando se trabaja con grandes volúmenes de producción (Ndikuryayo et al., 2022). Por lo tanto, la industria demanda con urgencia una técnica rápida y precisa que pueda clasificar el grano de arroz a bajo costo y ampliamente estandarizado.

En este sentido, la IA es una herramienta informática muy valiosa cuyas técnicas son cada vez más utilizadas en diversos campos y además permiten crear sistemas que imiten un comportamiento inteligente.

Por esta razón, con la ayuda de la IA permite desarrollar modelos de decisión en entornos eminentemente complejos y no lineales. En la actualidad, estas técnicas son tendencia en países con tecnología desarrollada y alta inversión en I+D+i (Mishra & Tyagi, 2022; Xu et al., 2022).

Los algoritmos de ML pueden ser entrenados de manera supervisada para el análisis de los datos agrícolas. En tal sentido, algunas técnicas de ML se están popularizando como buenas alternativas a las técnicas clásicas, ya que se basan en el reconocimiento de patrones (Greener et al., 2022; Rolnick et al., 2022).

Además, se han aplicado algoritmos para la evaluación de la calidad de los alimentos, incluida la carne de atún, cerdo y salmón (Chen et al., 2022; Medeiros et al., 2021), huevos de pato y carne de camello (Dong et al., 2021; Molaei et al., 2021), frutas y verduras (Ali & Dildar, 2021), en salud (Moya et al., 2017). La combinación de técnicas estadísticas multivariantes se ha convertido en una poderosa herramienta para hacer frente a diversos problemas en el sector alimentario.

Recientemente se han publicado trabajos que utilizan algoritmos para clasificar variedades de granos en arroz utilizando diferentes métodos analíticos (Aggarwal et al., 2022; Khatri et al., 2022; Komal et al., 2022; Tosawadi et al., 2022)., entre otros trabajos.

Por esta razón, siguiendo los argumentos de trabajos anteriores, se planteó la hipótesis de que el uso de técnicas de ML se puede aplicar como un medio para respaldar la toma de decisiones basada en datos en los sistemas de producción agrícola.

En la actualidad las industrias arroceras han logrado incrementos en sus producciones debido a su implementación de nuevas maquinarias agrícolas que permiten procesar 1 hasta 10 tn/hora.

Sin embargo, la mayoría de las empresas carecen de tecnologías al momento de clasificar granos de arroz, mientras otro método es la clasificación manual, pero es un proceso difícil. Además, este método requiere tiempo y es ineficiente, sobre todo cuando se trabaja con grandes volúmenes de producción. Por lo tanto, la industria demanda con urgencia una técnica rápida y precisa que pueda clasificar el grano de arroz a bajo costo y ampliamente estandarizado (Ndikuryayo et al., 2022).

## CAPÍTULO I: El problema de la investigación

### 1.1 Planteamiento del problema

La Organización de las Naciones Unidas para la Alimentación y la Agricultura indica que para el año 2030 la producción mundial de arroz tendrá que aumentar un 40% para satisfacer la necesidad alimentaria de millones de habitantes. En este contexto, el rol del sector agrícola arrocero adquiere un protagonismo mayúsculo, en la provisión de alimentos y materias primas (Amos et al., 2022).

Ecuador cuenta con unas buenas zonas agroecológicas, en donde puede desarrollarse el cultivo de arroz, así mismo, las provincias que tiene la mayor producción de la gramínea son Guayas, Los Ríos, Manabí y El Oro.

Sin embargo, la calidad del grano arroz se juzga en función de los atributos, que podrían clasificarse de forma intrínseca, entre las importante tenemos su textura, olor y sabor, mientras que en otro escenario es extrínsecas que tiene que ver mucho con la marca, empaquetado y etiquetado (Zahra et al., 2022).

Además, las características visuales de los granos de arroz son atributos de búsqueda importantes que afectan las decisiones de compra de los consumidores y, por lo tanto, se utilizan como algunos de los primeros criterios de selección como el tamaño, longitud, grosor, color del grano en los diferentes países y mercados (Ali et al., 2022). Es precisamente en este punto, donde se evidencia una oportunidad de aplicar técnicas ML, que permitan reducir el tiempo de clasificación y minimizar los costos generados por esta actividad.

Durante esta etapa, una persona se encarga del proceso de clasificación del arroz, sin embargo, este método usualmente no es eficaz y el tiempo empleado para esta actividad es extenso, como consecuencia se generan demoras en el proceso.

Al tratarse de un método que no es totalmente certero, también existe la probabilidad de errores y falta de homogenización en los granos de arroz.

Tomando en cuenta esta premisa, es necesario mejorar el proceso de clasificación de los granos de arroz, considerando para esto sus características morfológicas, tamaño, entre otras. Para cumplir con esta actividad se puede recurrir al uso de la tecnología, particularmente al aprendizaje automático, con la finalidad de mejorar los tiempos de clasificación, al mismo tiempo que se optimiza los costos de producción.

En la actualidad existen diferentes algoritmos de aprendizaje automático, sin embargo, no todos ellos tienen la misma precisión, particularmente para la clasificación del arroz, actividad que requiere bastante precisión, por lo cual surge la necesidad de investigar su funcionamiento para este tipo de producto específicamente.

Según (Vecchio et al., 2022), manifiesta que las universidades, centros de investigación públicos y privados, cumple un rol muy fundamental en el proceso de capacitación al agricultor o empresario que está vinculado en esta actividad.

En este sentido, es necesario implementar programas que permitan a los arroceros aplicar nuevas tecnologías para incrementar la calidad del grano y su productividad. Esta investigación busca utilizar técnicas ML con la finalidad de generar un modelo que permita clasificar el grano de arroz por sus características morfológicas. Tomando en cuenta esto, se plantea como pregunta de investigación:

¿Cuál será el algoritmo de aprendizaje automático más eficaz para realizar el proceso de clasificación de los granos de arroz?

## 1.2 Objetivo general

Aplicar las diferentes técnicas de aprendizaje automático en la clasificación de variedades de granos en arroz.

## 1.3 Objetivos específicos

Investigar de manera bibliográfica las diferentes teorías sobre la clasificación de arroz.

Establecer los modelos de aprendizaje automático más apropiados para dar solución al problema del proceso de clasificación de granos de arroz.

Evaluar el rendimiento de los modelos de aprendizaje automático propuesto para la clasificación de variedades de granos en arroz.

## 1.4 Justificación

El arroz es un producto muy rico en hidratos de carbono y almidón. Además, tiene una gran participación en la alimentación, por ser nutritivo y económico, y también es muy utilizado en el ámbito industrial.

La calidad del grano de arroz se determina habitualmente mediante inspecciones visuales y mediciones manuales, sin embargo, este método es lento, subjetivo y propenso a errores humanos (Krishna et al., 2022; Sultana et al., 2022).

Los instrumentos de laboratorio son costosos y pueden dificultar su aplicación entre las pequeñas empresas, especialmente en los países en desarrollo. Por lo tanto, es importante desarrollar un método alternativo que sea rápido, preciso, menos complicado y a bajo costo (Ndikuryayo et al., 2022).

Por esta razón, dentro de la revisión del marco teórico, se puede denotar la importancia de aplicar algoritmos de aprendizaje automático para la clasificación del grano de arroz de acuerdo sus características morfológicas e incluso durante la etapa del proceso productivo en el que se encuentra.

El aprendizaje automático (ML) es una de las principales áreas de investigación de la Inteligencia Artificial (IA), es una técnica de análisis auxiliar más prometedora para entender este complejo sistema, abriendo nuevos retos y oportunidades al mundo de la investigación. La principal aportación está relacionada con el análisis orientado a apoyar las decisiones (Greener et al., 2022; Rolnick et al., 2022)..

El objetivo del ML es construir modelos matemáticos directamente a partir de muestras de datos sin instrucciones explícitas, ya que algunas tareas pueden ser fácilmente resueltas por los humanos, pero es difícil explicar explícitamente cómo las resuelven.

El ML llena este vacío dejando que los ordenadores aprendan automáticamente modelos de mapeo a partir de las muestras de datos, que pueden proyectar las muestras de datos a su salida deseada.

En los últimos años, en el mundo se vienen desarrollando algunos trabajos de investigación como las citadas en el marco teórico que aborda casos similares. Debido a los problemas antes mencionados, este documento pretende desarrollar una nueva solución para identificar los tipos de granos de arroz utilizando el enfoque basado en algoritmos ML. Esto permitirá un avance importante en materia de soporte inteligente a la toma de decisiones.

## CAPÍTULO II: Marco teórico referencial

### 2.1. Conceptos generales

#### 2.1.1. El cultivo de arroz

Uno de los cereales de mayor consumo y relevancia nacional e internacional es el arroz, este producto es base de la alimentación de más de la mitad de la población a nivel mundial (Chipana Valero et al. 2022). De acuerdo con Álvarez Hernández et al. (2018), este cereal suministra aproximadamente el 20% de la energía alimentaria. El arroz es fundamental para la seguridad mundial, su producción se lleva a cabo utilizando métodos de labranza convencionales y sistemas de riego por inundación (Fernández Rodríguez et al. 2021).

Esta planta tiene su origen en el sur de China, ya que se cultivaba a orillas del río Azul en el quinto milenio antes de cristo. Se clasifican en dos especies de arroz: *Oryza sativa L.* y *Oryza glaberrima S.*, esta última es poco cultivada debido a la restricción que existe de su cultivo en la zona oeste de África (Álvarez et al. 2021). Corresponde a la familia de las *Poaceae*, es un cereal que sirve de alimento para el ser humano (López et al.).

De acuerdo con la FAO (2022), en la actualidad está previsto un incremento de la producción mundial de arroz para el año 2022-2023, la cual se espera que ascienda a 512,6 millones de toneladas de arroz elaborado, mientras que su utilización mundial será de 518,3 millones de toneladas, el descenso interanual será de 0,7% como consecuencia de las conmociones pronosticadas debido a su uso comercial y como pienso.

### **2.1.2. Producción de arroz en Ecuador**

Dentro de la agricultura ecuatoriana el arroz constituye un rubro de alto valor, es la principal fuente de carbohidratos desde inicios del siglo XX, incluso en algunas zonas del Ecuador donde se produce la patata; la superficie total cultivada es de aproximadamente 340.000 ha, de estas, el 41% utiliza sistema de riego, y el 45% de explotaciones no superan las 5 hectáreas (Murillo et al. 2022).

De acuerdo con el Instituto Nacional de Estadísticas y Censos (2021), para el año 2021 en Ecuador existían 342.967 hectáreas (Ha) de arroz sembradas, de las cuales se obtuvo una producción de 1.504.214 toneladas métricas. El mayor volumen de producción se encuentra en la región costa 323.230 Ha, las cuales se distribuyen principalmente en las provincias de Guayas con 204.874 Ha y Los Ríos con 104.165 ha.

En Ecuador se siembran principalmente las variedades INIAP 14 (33,7%), INIAP 11 (10,4%) e INIAP 15 (4,7%), producidas por el Instituto Nacional de Investigaciones Agropecuarias (INIAP), además de otras distribuidas por la empresa Procesadora Nacional de Alimentos (PRONACA), entre estas se encuentran las variedades SFL 09 (29,6%) y SFL011 (7%). Otras variedades presentes en el mercado ecuatoriano son: Yuma, Conquistador y San Juan, esta última proviene de la empresa colombiana INTEROC S.A (Zambrano et al. 2019).

### **2.1.3. Etapas del proceso de producción de arroz**

La producción de arroz se inicia con la siembra, cuidado, irrigación y control de plagas, el grano está listo para ser cosechado cuando presenta una humedad interna de aproximadamente el 25%, una vez cosechado es transportado a los molinos donde se realiza su procesamiento, el mismo que consta de las siguientes etapas: secado,

limpieza, descascarado, blanqueo y clasificación, de esta manera es posible conseguir un grano listo para su comercialización (Hernández-Cuello et al., 2021).

Dos de las etapas más importantes del proceso son la limpieza y la clasificación. La limpieza consiste en separar los granos de arroz de las sustancias extrañas, mientras que durante la clasificación se separa los quebrados de los resistentes, también se los puede clasificar por el color, en manchados y rayados (Cinar and Koklu 2019).

#### **2.1.4. Clasificación de granos de arroz**

Parte importante del proceso es el control y clasificación de granos, esta es realizada generalmente por un perito, es una actividad tediosa y repetitiva, por lo que existen grandes posibilidades de cometer errores. Usualmente se mide la longitud de los granos por medio de un calibre, esta medición se realiza tomando en cuenta una muestra representativa, los granos que resulten defectuosos serán extraídos de la muestra control (Acosta et al. 2017).

La clasificación manual de granos de arroz, resulta un proceso costoso y demanda de mucho tiempo, está limitada por la experiencia que tengan los evaluadores, es por esto que durante los últimos años se ha recurrido a diferentes alternativas que permitan valorar la clasificación y calidad del arroz. Estos incluyen algunos parámetros geométricos como la longitud o el perímetro, la tasa de fractura, la blancura, o las grietas que pueda presentar el grano (Koklu et al. 2021).

El constante incremento de la demanda de arroz crea la necesidad de producir y clasificar los granos con mayor rapidez, por lo que en la actualidad se están creando alternativas que permitan realizar el proceso de manera más rápida y eficiente, que

permitan la obtención de resultados precisos, una de las alternativas utilizadas actualmente es el aprendizaje automático (Ibrahim et al. 2019).

### 2.1.5. Escalas de clasificación del arroz

De acuerdo con el Servicio Ecuatoriano de Normalización (INEN), se utiliza la siguiente escala para el tamaño del arroz:

Tabla 1. Clasificación del arroz de acuerdo al tamaño

<u>Clase 1.</u>	<u>Extra largo</u>	<u>Granos con longitud mínima de 7,0 mm. Se tolera máximo el 20% de mezcla de otros granos largos</u>
Clase 2.	Largo	Granos con longitud entre 6,0 mm y 6,99 mm. Se tolera máximo el 20% de mezcla de otros granos medios
Clase 3.	Medio	Granos con longitud entre 5,0 mm y 5,99 mm. Se tolera máximo el 10% de otros granos cortos
Clase 4.	Corto	Granos con longitud menor de 5,0 mm
Clase 5.	Mezcla	Granos mezclados de dos clases o más de las clases mencionadas

(Servicio Ecuatoriano de Normalización)

### 2.1.6. Determinación del grado

Con la finalidad de determinar el grado se puede tomar en cuenta las recomendaciones del Instituto Internacional de Investigación del Arroz (IIIA):

**Granos rojos:** consiste en granos de arroz enteros o quebrados, que presentan un color rojo nítido o estrías de color rojizo en la cutícula (Internacional de Investigación del Arroz (2021)).

**Granos tizosos:** estos son granos enteros o quebrados en los cuales se evidencia el proceso de entizamiento, características tizosas o harinosas, total o parcialmente sobre la extensión del grano (Internacional de Investigación del Arroz (2021)).

**Granos tizosos totales:** son aquellos granos enteros o quebrados en los que es posible ver en más de la extensión de un grano entero el proceso de entizamiento. Esta denominación también incluye a los granos inmaduros (Internacional de Investigación del Arroz (2021)).

**Granos tizosos parciales:** son granos enteros o quebrados en los que se puede observar el proceso de entizamiento, solamente sobre algunos sectores del grano, no alcanzando la mitad del grano. En esta denominación se incluyen aquellos granos conocidos como panza blanca (Internacional de Investigación del Arroz (2021)).

**Granos dañados:** son todos aquellos granos de arroz, enteros o quebrados, que tienen alteraciones procedentes de hongos, fermentaciones, heladas, calentamiento u otros orígenes (Internacional de Investigación del Arroz (2021)).

**Materia extraña:** consiste en todo aquel material que no sea arroz pilado, incluso el arroz no descascarado o paddy (Internacional de Investigación del Arroz (2021)).

**Granos quebrados:** son granos que no pasan de entre  $\frac{1}{4}$  y  $\frac{3}{4}$  del tamaño total del grano entero (Internacional de Investigación del Arroz (2021)).

## 2.2. Aprendizaje Automático

El concepto de aprendizaje automático fue propuesto por primera vez por Arthur Samuel en 1959 (Zhou, 2021). Por aquel entonces, la investigación en aprendizaje automático se centraba principalmente en modelos estadísticos simples en el campo de los juegos de ordenador (Alpaydin, 2021). Tras varias décadas de desarrollo, Tom M. Mitchell propuso en 1997 una definición más formal del aprendizaje automático, que lo define desde la perspectiva de la mejora automática

del rendimiento de los modelos matemáticos a partir de experiencias (Bonaccorso, 2017).

A lo largo de la historia se han desarrollado muchas técnicas de aprendizaje automático. Algunos hitos dignos de mención son la retropropagación de redes neuronales desarrollada en la década de 1970, la máquina de vectores de soporte desarrollada en la década de 1990, y el aprendizaje profundo desarrollado en la década de 2000 (Molnar, 2020; Wang et al., 2016).

El ML puede clasificarse en tres paradigmas principales (Alpaydin, 2021): aprendizaje supervisado, aprendizaje no supervisado y aprendizaje por refuerzo. En el aprendizaje supervisado, cada muestra de datos se empareja sistemáticamente con una etiqueta determinada.

Un modelo toma las muestras de datos como entrada y aprende a realizar predicciones lo más cercanas posible a sus etiquetas correspondientes, lo que significa que el proceso de aprendizaje está "supervisado" por las etiquetas reales (Zhou, 2021). Por el contrario, las muestras de datos no tienen ninguna etiqueta en el aprendizaje no supervisado. El modelo de mapeo intenta desvelar los puntos comunes subyacentes en las muestras de datos.

El aprendizaje por refuerzo se sitúa entre el aprendizaje supervisado y el no supervisado. En el aprendizaje por refuerzo, las muestras de datos no tienen etiquetas directas. En su lugar, se da una medida para cada acción, conocida como recompensa (Bonaccorso, 2017). El aprendizaje automático (ML) se ha convertido en una valiosa herramienta para extraer información versátil de datos complejos mediante tareas de regresión o clasificación en la agricultura (Jagtap et al., 2022). El ML puede describirse como un área de la informática que estudia algoritmos y

técnicas para automatizar soluciones a problemas desafiantes con métodos de programación tradicionales (Elbadawi et al., 2021).

Un algoritmo de ML tiene como objetivo aprender un modelo o un conjunto de reglas a partir de un conjunto de datos etiquetados para que los datos indiquen etiquetas en el otro conjunto de datos que puedan predecir correctamente (Chen et al., 2021).

### **2.2.1. Uso del aprendizaje automático en la agricultura**

Kiratiratanapruk et al. (2020), en su investigación utilizaron cuatro técnicas de aprendizaje automático estadístico (LR, LDA, k-NN y SVM) y cinco modelos preentrenados (VGG16, VGG19, Xception, InceptionV3 e InceptionResNetV2), con la finalidad de comparar el rendimiento de la clasificación de granos de arroz, para esto se clasificó las muestras en grupos y subgrupos colectivos. El mejor nivel de precisión fue obtenido por medio del método SVM al 90,61%, 82,71% y 83,9% en los subgrupos 1 y 2 y el grupo colectivo; en lo que respecta a las técnicas de aprendizaje profundo, la mejor precisión se obtuvo con los modelos InceptionResNetV2, esta fue del 95,15%.

Un asunto importante para el almacenamiento, empaque y transporte de granos es el contenido de humedad, es por esto que Liu et al. (2022), en su investigación desarrollaron un dispositivo portátil conformado por tres partes: un módulo de circuito de microondas, un módulo de cálculo en tiempo real y un software para exponer los resultados, con la finalidad de medir la humedad del arroz con cáscara por medio de sensores microstrip de microondas asistido por estrategias de aprendizaje automático. Entre los modelos de predicción utilizados, el que presentó mejor rendimiento, precisión y estabilidad fue el modelo de bosque aleatorio ( $R^2 =$

0,99, RMSE = 0,28, MAE = 0,26), su rendimiento fue relativamente estable, con un error absoluto medio máximo del 0,55%.

Un indicador clave para tener una cosecha de arroz apropiada es el contenido de humedad de los granos, tomando en cuenta esto, Yang et al. (2021), evaluaron el nivel de humedad de la cosecha de granos por medio del aprendizaje automático en teléfonos inteligentes para establecer el momento adecuado para la cosecha, las imágenes de panículas individuales fueron tomadas mediante el uso de teléfonos inteligentes y fueron corregidos con una placa de corrección espectral-geométrica. Se construyeron cuatro modelos de aprendizaje automático, estos incluyeron bosque aleatorio, perceptrón multicapa, regresión de vector de soporte (SVR) y regresión lineal multivariada, de estos, el más adecuado para medir el grado de humedad fue el modelo SVR con un error absoluto medio de 1.23%.

Dheer et al. (2019), en su investigación recurren al uso de modelos de aprendizaje automático con la finalidad de diseñar un sistema de inspección de variedades de arroz, para esto se tomó en cuenta modelos clasificadores simples como Análisis Discriminante Lineal, Regresión Logística, K-Nearest Neighbor's (KNN) y método Naïve-Bayes. El método que presentó un mayor nivel de exactitud fue KNN, con un 99,16% de exactitud, 9,12% de precisión, y 99,12% de recuperación.

La clasificación de los granos de arroz es una dificultad que debe superarse, para esto, en muchos casos se está recurriendo a los sistemas inteligentes y la automatización, Arora et al. (2020), en su investigación utilizaron técnicas de procesamiento de imágenes y aprendizaje automático para este fin; como resultado se obtuvo que el sistema puede capturar varios parámetros de manera exitosa, tomando en cuenta para esto las imágenes de muestra de los diferentes tipos de

arroz, los cuales fueron almacenados en un archivo CSV para finalmente ser procesados, el sistema también evalúa el volumen de granos de arroz e implementa varios algoritmos para el procesamiento de imágenes y de aprendizaje automático.

### **2.2.2. Categorías del aprendizaje automático supervisado**

El método de aprendizaje automático supervisado se divide principalmente en dos categorías:

1. Regresión o capacidad de predecir valores continuos.
2. Clasificación que es la categorización de valores categóricos.

El proceso de clasificación comienza con el entrenamiento del modelo utilizando datos de entrenamiento para identificar el tipo o clase de la característica introducida (Ndikuryayo et al., 2022). Un clasificador binario etiqueta los datos como pertenecientes a uno de los dos grupos de salida. Esta tesis de maestría se centrará en utilizar algoritmos de aprendizaje supervisado puede explicarse cómo sigue.

### **2.2.3. Algoritmos de aprendizaje supervisado**

Los modelos de clasificación son un método de gran importancia utilizado en diversos campos. En la determinación de clases, los modelos de clasificación se utilizan para determinar a qué clase pertenecen los datos. El modelo de clasificación es un modelo que funciona haciendo predicciones (Waleed et al., 2021). El propósito de la clasificación es hacer uso de las características comunes de los datos para analizar los datos en cuestión. En nuestro estudio, se crearon modelos utilizando sistemas LR (Regresión Logística), MLP (Perceptrón Multicapa), SVM (Máquina de Vectores de Apoyo), RF (Bosque Aleatorio) y k- NN (K Vecino Más Cercano) para clasificar los granos de arroz según sus características.

#### **2.2.4. K-Nearest Neighbor (k-NN)**

El método k-NN es un algoritmo de aprendizaje no paramétrico. k-NN utiliza la distancia euclidiana como parámetro en nombre de la clasificación del conjunto de datos, donde K representa el número de vecinos, para calcular la distancia entre los datos (Pavani & Augusta Sophy Beulet, 2022).

La k-NN está pensada para clasificar datos de muestra cuya clase es desconocida. Por esta razón, la distancia a los datos de muestra se calcula con el conjunto de datos preclasificados en el conjunto de entrenamiento. Dado que hay una cierta cantidad de datos a probar, los datos de prueba se procesan con todos los datos existentes individualmente (Rekha Sundari et al., 2021). Los datos de prueba tendrán muchos vecinos que están cerca de ellos en términos de todas las medidas.

El algoritmo tiene algunas ventajas y desventajas. Una de las ventajas más importantes del algoritmo k-NN es la eficiencia y la otra es la flexibilidad. El algoritmo es eficiente en términos de simplicidad, velocidad y escalabilidad (Jagtap et al., 2022).

La flexibilidad del algoritmo facilita la gestión de conjuntos de datos que no pueden explicarse mediante relaciones lineales o no lineales por la complejidad de las relaciones. La desventaja más importante del algoritmo k-NN es que, a diferencia de otros algoritmos de minería de datos, no da una idea de qué variables son importantes en las nuevas predicciones (Malik et al., 2021).

#### **2.2.5. Máquina Soporte Vectorial (SVM)**

En los modelos SVM se utilizan diferentes funciones de núcleo. En este estudio, la clasificación se realizó utilizando la función de núcleo polinómico.

El objetivo del algoritmo SVM es encontrar un hiperplano que separe el conjunto de datos en un número predefinido y discreto de clases que sean las más consistentes con los ejemplos de entrenamiento (Kok et al., 2021). El término separación óptima del hiperplano se refiere a la frontera de decisión que minimiza las clasificaciones erróneas durante el paso de entrenamiento.

La SVM tiene la capacidad de clasificar datos en forma de lineal en el espacio bidimensional, planar en el espacio tridimensional e hiperplano en el espacio multidimensional con mecanismos de separación (Abdullah & Abdulazeez, 2021). La SVM realiza el proceso de clasificación encontrando el mejor hiperplano que separa los datos pertenecientes a las clases.

Las SVM tienen características similares a otros algoritmos de clasificación. Es especialmente similar a las redes neuronales, pero más parecido al algoritmo K-NN. Al igual que el algoritmo K-NN, la SVM determina sus vecinos basándose en los datos de muestra que se le presentan al algoritmo y asume que las estimaciones se realizan para los nuevos datos (Banerjee & Madhumathy, 2022).

#### **2.2.6. Bosque aleatorio (RF)**

RF es un clasificador compuesto por múltiples DT. Para realizar una nueva clasificación, cada DT proporciona una clasificación para las entradas. Después, RF evalúa las clasificaciones y selecciona la estimación que tiene más votos. La RF tiene la capacidad de gestionar un gran número de variables en un conjunto de datos. También tiene bastante éxito en la predicción de datos incompletos (Rakhra et al., 2021). El mayor inconveniente de la RF es su falta de repetibilidad. Además, el modelo final y los resultados posteriores son difíciles de interpretar.

Los árboles se generan extrayendo un subconjunto de muestras de ejercicio (un proceso de embolsado). Aproximadamente dos tercios de las muestras se utilizan para desarrollar un modelo. En cambio, el tercio restante de las muestras se utiliza para estimar el rendimiento del modelo mediante la votación por mayoría (Suresh et al., 2021). El algoritmo produce árboles con altas varianzas y bajas tendencias, aumentando el bosque hasta un número de árboles definido por el usuario. El promedio de las probabilidades de asignación de clase determinadas por todos los árboles generados es la decisión final de clasificación.

### **2.2.7. Regresión logística (LR)**

El LR es uno de los modelos estadísticos más utilizados. En el LR, la variable dependiente se estima a partir de una o más variables. El LR aclara la relación entre las variables dependientes y las independientes con el menor número de variables (Arumugam et al., 2022).

En el LR no es necesario crear una distribución normal de las variables. Dado que los valores previstos en la LR son probabilidades, la LR se limita a 0 y 1. Esto se debe a que la LR predice su probabilidad, no a sí misma, en los resultados (Arumugam et al., 2022). En este estudio, se utilizó el método Newton para la optimización durante la clasificación con la ayuda de LR.

### **2.2.8. Perceptrón multicapa (MLP)**

En la actualidad, se han desarrollado muchos modelos de redes neuronales artificiales para su uso con fines específicos, y el MLP es uno de los más utilizados de estos modelos (Bakthavatchalam et al., 2022). En MLP, la secuencia de neuronas está en capas, y hay una capa oculta entre ellas, junto con dos capas principales.

El MLP puede contener más de una capa oculta. La capa de entrada, que es la primera de las capas principales, es la capa donde se leen los datos y contiene información sobre el problema que hay que resolver. La capa de salida, que es la segunda capa principal, es la capa donde se definen las clases y se reciben las salidas de la información procesada en la red. La capa oculta es la capa donde se realizan las operaciones intermedias sobre los datos entre las capas principales (P. Xu et al., 2021).

El MLP tiene tantas neuronas como el número de características, y los datos son proporcionados por un flujo de datos en una dirección desde la capa de entrada hasta la capa de salida. Además, es posible controlar y modificar la estructura de la red durante el periodo de entrenamiento (Goyal et al., 2022). En este estudio, hay 4 capas ocultas y también se utilizó la función de activación sigmoidea.

### **2.3. Validación cruzada**

La validación cruzada es un método de predicción de errores desarrollado con el objetivo de mejorar la seguridad de la clasificación. La validación cruzada divide el conjunto de datos de forma aleatoria en un número determinado de subconjuntos para el entrenamiento y la prueba. Acepta uno de los subconjuntos como conjunto de prueba y el sistema se entrena con los conjuntos restantes (Mourtzinis et al., 2021). Este proceso se repite hasta el número de conjuntos de datos y se prueba el sistema. Debido a estas características del conjunto de datos, se consideró apropiado utilizar métodos de clasificación. Los algoritmos de clasificación entrenaron el modelo observando los patrones de los datos en el conjunto de entrenamiento (Kasinathan et al., 2021).

De este modo, clasificó los datos de forma que no se pudieran ver antes de una manera muy precisa. En este estudio, los núcleos de granos de arroz se modelaron utilizando clasificadores LR, MLP, SVM, RF y k-NN con la ayuda del lenguaje de programación R y Rstudio. Son métodos de aprendizaje automático que se utilizan frecuentemente en problemas de clasificación. Además de estos métodos, también se han probado otros métodos de aprendizaje automático y los métodos utilizados en el estudio han obtenido resultados de clasificación más significativos que otros.

#### **2.4. Medidas de desempeño**

La creación de un nuevo modelo necesario para los problemas de clasificación o el uso de modelos existentes y el éxito de este modelo se calculó por el número de estimaciones precisas (Nosratabadi et al., 2021). Esto es efectivo en la precisión de la clasificación más que en la estimación de si el modelo es bueno o no. Por eso se utiliza la matriz de complejidad para explicar las evaluaciones predictivas de la clasificación.

Por lo tanto, la matriz de confusión se utiliza para explicar las evaluaciones predictivas de la clasificación. La matriz que proporciona información sobre las clases reales con las clases estimadas realizadas a través de un modelo de clasificación basado en los datos de la prueba es la matriz de complejidad (Koklu et al., 2021). La Matriz de complejidad tiene cuatro parámetros, VP (verdaderos positivos) y VN (verdaderos negativos) representan respectivamente el número de ejemplos positivos y negativos clasificados correctamente, mientras que FP (falsos positivos) y FN (falsos negativos) representan respectivamente el número de ejemplos positivos y negativos clasificados incorrectamente (Mena, 2008).

## **CAPÍTULO III: Diseño metodológico**

### **3.1. Tipo y diseño de investigación**

Este estudio se enmarca en un diseño de investigación cuantitativa. Acorde a Ochoa et al. (2020) este diseño de investigación consiste en la recolección de datos numéricos para su posterior análisis e interpretación utilizando herramientas de análisis matemático y estadístico para describir y explicar fenómenos a través de datos numéricos. El tipo de la investigación es causal comparativa dado que se comparan diferentes técnicas de aprendizaje automático en el proceso de clasificación de granos de arroz. Según Sánchez et al. (2021) este tipo de investigación destaca porque depende de factores de comparación y deja al investigador la decisión de la selección de características que se desean comparar. En el caso de este estudio se comparan diferentes medidas de rendimiento en cinco tipos de algoritmos para la correcta clasificación de dos diferentes variedades de arroz.

### **3.2. La población y muestra**

#### **3.2.1. Características de la población**

Entre los arroces certificados que se cultivan en el Ecuador, se han seleccionado para el presente estudio, la especie INIAP-14, que cuenta con una gran superficie de plantación desde 1999, y la especie INIAP-17, cultivada desde 2007. Si se observan las características generales de la variedad INIAP14, tienen un aspecto ancho, largo, vidrioso y opaco. El peso del grano de mil es de 23-26 gr. Al observar las características generales de la especie INIAP-17, tienen un aspecto ancho y largo, vidrioso y opaco. El peso de mil de grano es de 22-28 gr.

En este estudio, la distribución de 3.810 granos de arroz obtenida como resultado del procesamiento de las imágenes de ambas especies.

### **3.2.2. Delimitación de la población**

En el estudio se utilizó un conjunto de datos pertenecientes a las dos variedades de arroz (INIAP 14 e INIAP 17), que suelen cultivarse en la costa ecuatoriana. El conjunto de datos constó de 3.810 imágenes de granos de arroz (2.180 para INIAP-14 y 1.630 para INIAP 17) que fueron recolectados en campo, los mismo fueron transportados a laboratorio. Donde la cámara utilizada para el estudio tiene 2,2 megapíxeles, una resolución de 2048×1088 y resolución completa a una frecuencia de imagen máxima de 53,7 fps. Dispone de funciones como el balance de blancos y la corrección de contraluz.

La cámara utilizada en el estudio se colocó en una caja cerrada con un dispositivo de iluminación en su interior y una estructura para evitar que la luz del entorno exterior. El color de fondo de la caja se seleccionó como negro para facilitar el procesamiento de la imagen. Los tamaños se diseñaron de forma que las imágenes pudieran captarse desde un área de 14 cm de ancho y 18 cm de largo. La altura de la cámara se fijó a 15 cm. Las imágenes resultantes se grabaron transfiriéndolas al ordenador.

En este estudio no fue necesario calcular la muestra, ya que se trabajó con toda la población para determinar que método de aprendizaje automático da mejores resultados en cuanto a la clasificación de arroz basado en las medidas de rendimiento (Dominguez-Lara y Merino-Soto, 2018). Se evaluaron en total siete variables morfológicas descritas en la Tabla 2.

Tabla 2. Variables morfológicas para evaluar

No	Nombre	Explicación
1	Área (A)	Devuelve el número de píxeles dentro de los límites del grano de arroz.
2	Perímetro (P)	Calcula la circunferencia mediante el cálculo de la distancia entre píxeles alrededor de los límites del grano de arroz.
3	Longitud del eje mayor (L)	La línea más larga que se puede dibujar en el grano de arroz, es decir, la distancia del eje principal.
4	Longitud del eje menor (l)	La línea más larga de un grano de arroz que se puede trazar perpendicularmente al eje mayor.
5	Excentricidad	Mide la redondez de la elipse, que tiene los mismos momentos que el grano de arroz.
6	ConvexArea (CA)	Devuelve el recuento de píxeles de la cáscara convexa más pequeña de la región formada por el grano de arroz.
7	Extensión (EX)	Devuelve la proporción de la región formada por el grano de arroz a los píxeles de la caja delimitadora.

Fuente: tomado del trabajo de Ozkan & Koklu (2017).

Dado que el conjunto de datos es irregular, es decir, no hay el mismo número de datos pertenecientes a cada variedad de arroz, y el éxito de la clasificación no es alto, por lo que se vuelve necesario utilizar métricas adicionales como la exhaustividad, y la exactitud, respecto a otros estudios en los que los datos han sido regulares y han utilizado otro tipo de métricas (sensibilidad) (Koklu et al., 2021).

Las métricas utilizadas van de 0 a 1. Debido a que el éxito de clasificación de los modelos utilizados podría no ser alto, no se puede hacer una comparación cuando se redondean estos valores. Por ello, los valores de estas medidas se muestran como porcentajes.

Posteriormente, se realizó la validación cruzada que es un método de predicción de errores desarrollado con el objetivo de mejorar la seguridad de la clasificación. La validación cruzada es un método utilizado para medir objetivamente la precisión de los modelos de clasificación. En este método, el conjunto de datos se divide en el mismo número de partes según el valor numérico especificado.

El valor numérico especificado se denomina  $k$ .  $1/k$ ; parte del conjunto de datos se reserva para la prueba,  $k-1$  y parte se reserva para el entrenamiento del modelo (*training*). (Koklu et al., 2021). Este proceso continúa hasta que cada parte del conjunto de datos se utiliza como parte de prueba. Este proceso se repite  $k$  veces. El éxito general de clasificación del modelo en el conjunto de pruebas se obtiene tomando la media aritmética de los éxitos de clasificación obtenidos como resultado de estas operaciones.

La validación cruzada dividió el conjunto de datos de forma aleatoria en un número determinado de subconjuntos para realizar el entrenamiento y prueba de los datos. Este proceso se repitió hasta terminar con la población de conjuntos de datos y poner a prueba los modelos.

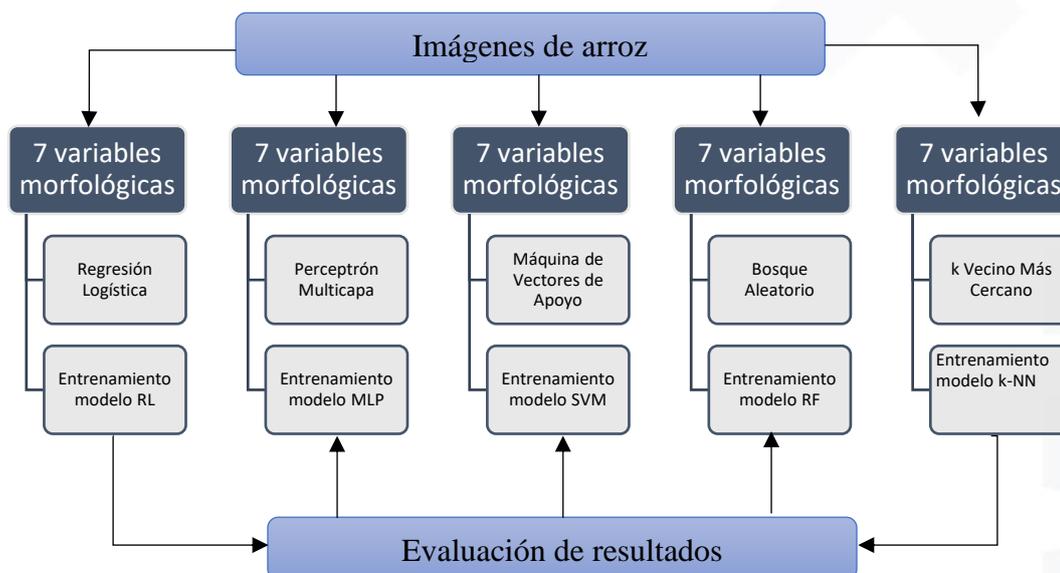
Finalmente, para la evaluación de los modelos, se aplicó el estadístico Kappa, para determinar qué modelo es el mejor a través de la interpretación de los rangos de valores que van desde -1 a 1, interpretándose valores cercanos a -1 como el peor modelo, y valores cercanos a +1 como el mejor modelo.

### **3.3. Los métodos y las técnicas**

Las técnicas que fueron empleadas están basadas en cinco métodos o modelos de aprendizaje automático llamados también algoritmos. Estos modelos son Regresión Logística (LR), Perceptrón Multicapa (MLP), Máquina de Vectores de

Apoyo (SVM), Bosque Aleatorio (RF) y k Vecino Más Cercano (k- NN). Estos modelos se compararon a través de las medidas de rendimiento para determinar el mejor modelo para clasificar los granos de arroz. En la Figura 1, se muestra el diagrama del modelo propuesto para la clasificación de las dos variedades de arroz.

Figura 1. Diagrama del modelo propuesto



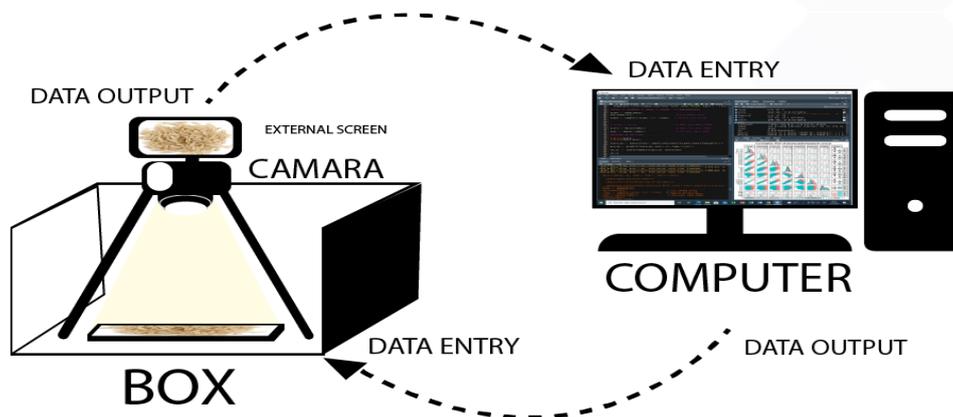
### 3.4. Procesamiento estadístico de la información

Una vez tomadas en laboratorio las fotografías de los 3.810 granos de arroz de las dos variedades, se procedió a calcular las características morfológicas del arroz, a través del software MATLAB. En total se calcularon siete variables morfológicas que fueron el área, perímetro, longitud del eje mayor y menor, excentricidad, área convexa, y extensión.

Esta información se exportó al Programa Excel y se compiló en una sola base de datos clasificados por cada característica. Para el análisis de la base de datos se utilizó el lenguaje de programación R (Figura 2) donde se realizó la estadística descriptiva para cada característica calculando la media aritmética, la desviación

estándar, asimetría y kurtosis, utilizando diferentes librerías. Una vez calculada la estadística, se aplicaron los diferentes modelos partiendo de la matriz de confusión para explicar las evaluaciones predictivas de la clasificación.

Figura 2. El esquema de adquisición de imágenes



Fuente: Por autores

Esta matriz proporcionó información sobre las clases reales con las clases estimadas realizadas a través de un modelo de clasificación basado en los datos de la prueba es la matriz de complejidad. La Matriz de confusión tiene cuatro parámetros; vp: verdadero positivo, fp: falso positivo, fn: falso negativo, y vn: verdadero negativo. Para las mediciones del rendimiento de la clasificación de las dos variedades de arroz, se utilizaron criterios de éxito como la exactitud, precisión, exhaustividad, especificidad y puntuación.

La matriz de confusión se utiliza para medir el rendimiento de clasificación de los métodos de aprendizaje automático. Esta matriz facilita la búsqueda de conexiones entre el rendimiento del clasificador y los resultados de las pruebas. La matriz de confusión proporciona información sobre la clasificación correcta e incorrecta de las muestras positivas y la clasificación correcta e incorrecta de las muestras negativas. En la Tabla 3 se muestra una matriz de confusión de dos clases.

Tabla 3. Matriz de confusión para dos clases

		<u>Clase estimada</u>	
		Positivo	Negativo
Clase actual	Positivo	Verdadero positivo (VP)	Falso negativo (FN)
	Negativo	Falso positivo (FP)	Verdadero negativo (VN)

Fuente: tomado del trabajo de Martínez (2018).

El conjunto de datos sobre el arroz de este estudio consta de dos clases. Por este motivo, se utilizó la matriz de confusión de dos clases en los procesos de clasificación. A partir de los valores VP, FP, FN y VN de la matriz de confusión, se realizan cálculos estadísticos y se puede analizar en detalle el rendimiento de los clasificadores (Martínez et al. 2018). Las métricas obtenidas de los cálculos estadísticos para la matriz de confusión de dos clases, las fórmulas utilizadas para calcular estas métricas e información sobre el propósito para el que se utilizan las métricas se muestran en la Tabla 4.

Tabla 4. Medidas de rendimiento y fórmulas de cálculo para la clasificación de dos clases

N°	Nombre	Formula	Explicación
1	Exactitud	$\frac{tp+tn}{tp+fp+tn+fn} \times 100$	Se utiliza para medir la proporción de predicción verdadera en todas las muestras incluidas en la evaluación.
2	Precision	$\frac{tp}{tp+fp} \times 100$	Se utiliza para medir la relación entre las muestras positivas clasificadas con precisión y muestras positivas clasificadas con precisión y las muestras positivas totales estimadas.
3	Exhaustividad	$\frac{tp}{tp+fn} \times 100$	Se utiliza para medir la proporción de valores positivos correctamente clasificados.
4	Especificidad	$\frac{tn}{tn+fp} \times 100$	Se utiliza para medir la proporción de valores negativos correctamente clasificados.
5	Puntuación F1	$\frac{2tp}{2tp+fp+fn} \times 100$	Representa la media armónica entre los valores de Exhaustividad y Precision.

Fuente: tomado del trabajo de Ozkan & Koklu (2017).

## CAPÍTULO IV: Análisis e interpretación de resultados

### 4.1. Estadística descriptiva

De un total de 3.810 de granos de arroz pertenecientes a las dos variedades de arroz, se evaluaron 7 características morfológicas para cada grano. En la tabla 5, se muestra los resultados de la estadística descriptiva para las dos variedades (INIAP-14 y INIAP 17) en el cual se calculó para cada característica (área, perímetro, eje mayor y menor, entre otras) la media aritmética, desviación estándar, varianza, error estándar, coeficiente de variación, asimetría y kurtosis.

Tabla 5. Estadística descriptiva de las características morfológicas (INIAP-14:n=2.180; INIAP-17: n=1.630)

	Área		Perímetro		Eje mayor		Eje menor	
	INIAP-14	INIAP-17	INIAP-14	INIAP-17	INIAP-14	INIAP-17	INIAP-14	INIAP-17
Media	14164,76	11550,88	487,48	429,43	205,49	176,29	88,78	84,48
D.E.	1288,59	1041,69	22,23	20,14	10,36	9,36	5,35	5,3
Var(n-1)	1660462,98	1085125,3	494,07	405,81	107,36	87,57	28,59	28,11
E.E.	31,85	22,3	0,55	0,43	0,26	0,2	0,13	0,11
CV	9,1	9,02	4,56	4,69	5,04	5,31	6,02	6,28
Asimetría	0,01	-0,06	-0,16	0,04	-0,07	0,28	-0,01	-0,35
Kurtosis	-0,04	0,35	-0,1	0,23	0,05	0,15	0,45	0,59

Fuente: Por autores

	Excentricidad		Área convexa		Extensión	
	INIAP-14	INIAP-17	INIAP-14	INIAP-17	INIAP-14	INIAP-17
n	2180	1630				
Media	0,9	0,88	14496,37	11800,56	0,65	0,67
D.E.	0,01	0,02	1311,04	1062,48	0,08	0,07
Var(n-1)	1,80E-04	3,60E-04	1718837,17	1128859,99	0,01	0,01
E.E.	3,30E-04	4,10E-04	32,4	22,74	2,00E-03	1,50E-03
CV	1,48	2,17	9,04	9	12,61	10,8
Asimetría	-0,51	-0,19	-0,02	-0,05		
a					0,42	0,38
Kurtosis	1,14	0,52	-0,05	0,38	-1,05	-1,05

Fuente: Por autores

Se puede observar que la línea INIAP-17 tiene una menor área promedio por grano respecto a la línea INIAP-14, es decir que su grano (INIAP-14) es más grande, además, posee un mayor perímetro, lo que significa que compensa teniendo un mayor

borde del grano. La línea INIAP-14 también posee una longitud del eje mayor y menor más grandes que INIAP-17 lo que representa que los granos son más alargados y anchos. Esto a su vez significa que INIAP-14 tendría mayor cantidad de almidón en su composición, lo que requeriría una mayor cantidad de agua para su cocción. Respecto a la excentricidad muestran similares características ambas variedades, adoptando una forma elipsoidal para INIAP-14, y una forma discoidal para INIAP-17. INIAP-14, posee una mayor área convexa promedio, es decir, que tiene una envoltura más ancha que INIAP-17, lo que le permite tener un mayor de contenido de agua, carbohidratos, agua, fibras. Finalmente, respecto a su extensión, ambas variedades tienen similares características, es decir, que devuelve la proporción de la región formada por el grano de arroz a los pixeles de la caja delimitadora.

Al realizar la prueba de T de Student para 2 colas, se pudo constatar que no existen diferencias significativas ( $p < 0,001$ ) en ninguna de las características morfológicas de ambas variedades de arroz, tal como lo muestra la Tabla 6.

Tabla 6. Prueba de T de Student a dos colas

Variable	Variedad 1	Variedad 2	n1	n2	Media 1	Media 2	T	p-valor
Área	INIAP-14	INIAP-17	1630	2180	14164,76	11550,88	67,23	<0,0001
Perímetro	INIAP-14	INIAP-17	1630	2180	487,48	429,43	83,12	<0,0001
Longitud eje mayor	INIAP-14	INIAP-17	1630	2180	205,49	176,29	89,8	<0,0001
Longitud eje menor	INIAP-14	INIAP-17	1630	2180	88,78	84,48	24,67	<0,0001
Excentricidad	INIAP-14	INIAP-17	1630	2180	0,90	0,88	47,29	<0,0001
Área convexa	INIAP-14	INIAP-17	1630	2180	14496,37	11800,56	68,1	<0,0001
Extensión	INIAP-14	INIAP-17	1630	2180	0,65	0,67	-7,24	<0,0001

Fuente: Por autores

#### 4.2. Matriz de complejidad

Una vez realizada la estadística descriptiva, se aplicaron los diferentes modelos o algoritmos partiendo de la matriz de complejidad para explicar las evaluaciones predictivas de la clasificación del grano de arroz. Esta matriz proporcionó información sobre las clases reales con las clases estimadas realizadas

a través de un modelo de clasificación basado en los datos de la prueba es la matriz de complejidad.

En la Tabla 7, se indican los valores de la matriz de complejidad para todos los algoritmos utilizados.

Tabla 7. Matriz de complejidad de los algoritmos utilizados (a) LR, (b) MLP, (c) SVM, (d) RF, (e) k-NN,

Matriz de LR		Clase estimada	
Clase real	INIAP-14	1489 (Vp)	141 (Fp)
	INIAP-17	125 (Fn)	2055 (Vn)

(a)

Matriz de MLP		Clase estimada	
Clase real	INIAP-14	1484	146
	INIAP-17	126	2054

(b)

Matriz de SVM		Clase estimada	
Clase real	INIAP-14	1492	138
	INIAP-17	135	2045

(c)

Matriz de RF		Clase estimada	
Clase real	INIAP-14	1480	150
	INIAP-17	140	2040

(d)

Matriz de k-NN		Clase estimada	
Clase real	INIAP-14	1419	211
	INIAP-17	224	1956

(e)

Como se puede observar en la Tabla 7, el algoritmo LR predijo de mejor manera mayor cantidad de verdaderos positivos (Vp) (1489), respecto a los demás algoritmos. De igual forma predijo con más precisión los verdaderos negativos (Vn) (2055), seguido de MLP, SVM, RF, k-NN.

### 4.3. Medidas de rendimiento

Para las métricas de rendimiento de la clasificación, se calcularon criterios de éxito como la Exactitud, Sensibilidad, Especificidad, exhaustividad, precisión, y puntuación en base a la matriz de complejidad para cada modelo. Los resultados de la medición del rendimiento de la clasificación se indican en la Tabla 8.

Tabla 8. Medidas de rendimiento promedio de los modelos utilizados

Measure	LR	MLP	SVM	RF	k-NN
Accuracy	93,02	92,86	92,83	92,39	88,58
Sensivity	92,26	92,17	91,7	91,36	86,37
Specificity	93,58	93,36	93,68	93,15	90,26
Precisión	91,35	91,04	91,53	90,8	87,06
F1-Score	91,82	91,6	91,62	91,08	86,71
Coeficiente Kappa	97.5	97.3	97.3	96.4	96.3

Fuente: Por autores

Al examinar los valores promedio de cada uno de los algoritmos, se determinó que la media de la precisión del modelo LR resultó ser del 93,02%, la media del modelo MPL fue del 92,86%, del modelo SVM (92,83%), del modelo RF (92,39%), y del modelo k-NN del 88,58%. De igual manera el modelo LR, obtuvo un mejor estadístico Kappa que tuvo un mejor ajuste del efecto del azar en proporción de las clases reales de granos de arroz con las clases estimadas. De acuerdo con estos valores obtenidos producto de la matriz de complejidad se puede mencionar que el mejor método de clasificación es el método LR, dada su mejor predicción de los valores obtenidos de verdaderos positivos, y verdaderos negativos.

#### 4.4. Discusión de resultados

Los resultados de la estadística descriptiva, concuerdan con lo encontrado por Ceavichay et al., (2013) en el que se determinó que no mostraron diferencias significativas entre la variedad INIAP-14 e INIAP-17 en sus características morfológicas, recomendando que ambas variedades pueden ser utilizadas para exportación.

Además, señaló que ambas líneas de arroz tienen gran poder absorción de agua y poder de hinchamiento, lo que sus harinas pueden ser utilizadas en productos opacos que no necesiten retener agua. Sin embargo, recomienda la variedad INIAP-17 para procesos de panificación por su viscosidad elevada respecto a la INIAP-14.

En un estudio de Coello y Garces, (2013) se midió las propiedades térmicas que ocurren durante la gelatinización del almidón del arroz en INIAP-14 e INIAP-17. Para estudiar el comportamiento de las cadenas poliméricas, se realizó un análisis de calorimetría diferencial de barrido para medir la cantidad de energía necesaria para el proceso de gelificación e identificar los eventos térmicos durante la transición.

Las propiedades térmicas de gelatinización fueron temperatura inicial, temperatura del pico endotérmico, temperatura final, diferencial de entalpía e intervalo de gelatinización. Los resultados mostraron que, si hubo diferencias significativas en las propiedades térmicas de las dos variedades, sugiriendo que INIAP-17, presentó menor contenido de amilosa debido a una menor temperatura de gelatinización que INIAP-14.

Se determinó que el modelo de LR fue el algoritmo que predijo con más precisión la clasificación de granos de arroz acorde a la variedad. Estudios han encontrado que el modelo LR fue el mejor método de clasificación con una precisión del 94%, respecto a otros métodos (MLP), 92,83% (SVM), 92,49% (DT), 92,39% (RF),

91,71% (NB), 88,58% (k-NN) (Cinar & Koklu, 2019). Por otro lado, en el estudio de Liu et al., (2021) la precisión de LR estuvo en un rango entre 91,67 a 100 %, indicando que este modelo puede seleccionar las características morfológicas del grano de arroz de forma estable y mejorar la precisión de reconocimiento de las variedades de arroz.

Contrario a lo hallado en este estudio, Dheer, (2019) demostró que el algoritmo k-NN tuvo mejores resultados que los métodos por análisis de discriminación linear, y LP, resultando con una precisión de 99,28%, 98,97%, y 97,67%, respectivamente. Estos resultados pueden obedecer a número de repeticiones que se realizaron en la validación cruzada, siendo 10 repeticiones para el estudio de Dheer, y 20 repeticiones para el presente trabajo, en el conjunto de datos para prueba y validación.

Sin embargo, enfatiza que el método k-NN puede ser utilizado como alternativa al LR, en aplicaciones móviles cuando el investigador necesite tomar datos en campo de las variedades de arroz.

En un estudio de Jin et al., (2022) se demostró que las precisiones de clasificación de la mayoría de los modelos utilizados fueron superiores al 95%, siendo el mejor, modelo el ResNet obteniendo mejores resultados de clasificación respecto a LR.

El contraste en los resultados de los modelos puede obedecer a las diferentes variedades de arroz utilizadas, donde sus semillas pueden tener granos pequeños y colores y formas similares o diferentes, que son difíciles de distinguir mediante la observación visual, por lo que es importante elegir una estrategia de clasificación adecuada al clasificar las semillas de arroz.

Las diferencias de los algoritmos también pueden estar presentes en la calidad de las diferentes variedades de arroz. Jin et al., (2022) señala los tipos de semillas de

arroz influyen en la clasificación de las variedades de arroz, por lo que es significativo considerar los tipos de semillas de arroz al identificar las variedades de semillas. Arora et al., (2020) sugiere que los valores obtenidos del modelo de LR son más precisos y con menor dispersión de datos, respecto a otros algoritmos como k-NN, RF, SVM. Para el caso concreto de este estudio, los resultados concuerdan con los de Arora et al., (2020), y se recomienda la utilización del algoritmo LR por su mayor precisión en la clasificación de variedades respecto a los otros modelos.

Los algoritmos utilizados en este estudio tuvieron una precisión sobre el 85% resultados que son eficiente en términos de simplicidad, velocidad y escalabilidad.

La flexibilidad de mejor algoritmo de LR facilita la gestión de conjuntos de datos que no pueden explicarse mediante relaciones lineales o no lineales por la complejidad de estas. La desventaja más importante del algoritmo k-NN que fue el que obtuvo menor precisión es que, a diferencia de otros algoritmos, no da una idea de qué variables son importantes en las nuevas predicciones.

Finalmente, se señala que, utilizando los modelos de procesamiento de imágenes creadas para este estudio, se puede crear un conjunto de datos sobre la especie objetivo.

De este modo, se puede investigar sobre las cuestiones que afectan al resultado de la producción, obteniendo información sobre el mismo tipo de arroz. Además, se pueden diseñar sistemas automáticos para muchos procesos, como la calibración de los tipos de arroz y la separación de las especies de las sustancias no deseadas que puedan estar presentes.

Para aumentar la tasa de éxito en la clasificación, se pueden obtener más imágenes de las especies y se cree que las tasas de éxito pueden aumentar utilizando características morfológicas, así como características de color y forma.

## CAPÍTULO V: Conclusiones y Recomendaciones

### 5.1. Conclusiones

Durante la etapa de clasificación del arroz se pueden evaluar parámetros geométricos como la longitud o el perímetro, la tasa de fractura, la blancura, o las grietas que pueda presentar el grano; o también se puede recurrir a las teorías de clasificación internacionales, una de estas la del Instituto Internacional de Investigación del Arroz, en la actualidad se está recurriendo con mayor frecuencia al uso del aprendizaje automático, por medio del cual es posible clasificar al arroz de manera más apropiada.

Para la clasificación se realizaron diferentes modelos utilizando los siguientes algoritmos LR, MLP, SVM, RF y k- NN, que son las técnicas de aprendizaje automático más utilizadas. Las mediciones estadísticas de la matriz de confusión como resultado de la clasificación se utilizaron como métricas de rendimiento. Las líneas de arroz utilizadas para esta investigación fueron INIAP-14 e INIAP-17.

Los resultados obtenidos permiten concluir que el modelo LR es el mejor método de clasificación, debido a su mejor predicción de los valores obtenidos de verdaderos positivos, y verdaderos negativos, además, la media de la precisión de este modelo fue de 93,02%, superior a los otros modelos estudiados, además, este modelo también obtuvo un mejor estadístico Kappa que tuvo un mejor ajuste del efecto del azar en proporción de las clases reales de granos de arroz con las clases estimadas.

El desarrollo de este trabajo permite llegar a la conclusión de que el uso de técnicas de ML en la industria arrocera, es aplicable para el apoyo en la toma las decisiones basadas en el desarrollo del proceso de clasificación del grano.

## 5.2. Recomendaciones

Se recomienda la revisión de nuevas teorías de clasificación del arroz a nivel nacional e internacional, por ejemplo, la clasificación del Instituto Ecuatoriano de Normalización (INEN), y otras clasificaciones que puedan existir con la finalidad de contar con más argumentos de clasificación que permitan a su vez probar los diferentes modelos de aprendizaje automático.

Realizar nuevas investigaciones revisando otros modelos aparte de los analizados en esta investigación que permitan tener nuevos argumentos, y que a su vez hagan posible tener una industria arrocera más eficiente, esto permitirá reducir el tiempo de clasificación, evitar pérdidas previsibles y minimizar los costos generados por esta actividad.

Se recomienda que, dentro de la industria arrocera se haga uso del modelo ML ya que como se puede apreciar en los resultados obtenidos, es más eficiente al momento de realizar la clasificación del arroz, por lo tanto, es una alternativa importante para este sector productivo.

## Bibliografía

- Abdullah, D. M., & Abdulazeez, A. M. (2021). Machine Learning Applications based on SVM Classification A Review. *Qubahan Academic Journal*, 1(2), 81–90. <https://doi.org/10.48161/qaj.v1n2a50>
- Acosta, C., Sampallo, G. M., González Thomas, A., Cleva, M., & Liska, D. (2017, October 10). Detección e identificación de defectos en granos de arroz empleando visión artificial. IX Congreso Argentino de AgroInformática (CAI 2017) - JAIIO 46-CLEI 43 (Córdoba, 2017). <http://sedici.unlp.edu.ar/handle/10915/62834>
- Aggarwal, S., Suchithra, M., Chandramouli, N., Sarada, M., Verma, A., Vetrithangam, D., Pant, B., & Ambachew Adugna, B. (2022). Rice Disease Detection Using Artificial Intelligence and Machine Learning Techniques to Improvise Agro-Business. *Scientific Programming*, 2022. <https://doi.org/10.1155/2022/1757888>
- Ali, I., Iqbal, A., Ullah, S., Muhammad, I., Yuan, P., Zhao, Q., Yang, M., Zhang, H., Huang, M., Liang, H., Gu, M., & Jiang, L. (2022). Effects of Biochar Amendment and Nitrogen Fertilizer on RVA Profile and Rice Grain Quality Attributes. *Foods (Basel, Switzerland)*, 11(5). <https://doi.org/10.3390/foods11050625>
- Álvarez Hernández, J. C., Tapia Vargas, L. M., Hernández Pérez, A., Barrios Gomez, E. J., & Pardo Melgarejo, S. (2018). Estabilidad productiva de variedades avanzadas de arroz grano largo delgado en Michoacán México. *Revista Mexicana de Ciencias Agrícolas*, 9(3), 629–637.

- Álvarez, R., Reyes, E., Ramos, N., & Valera, E. (2021). "LIBERTAD FL": Nuevo cultivar de arroz de riego para Venezuela. *Revista*. <http://revistas.uns.edu.pe/index.php/PUNKURI/article/view/6>
- Alpaydin, E. (2021). *Machine Learning, revised and updated edition*. MIT Press. <https://play.google.com/store/books/details?id=2nQJEAAAQBAJ>
- Amos, B., Aidoo, R., Osei Mensah, J., Adzawla, W., Appiah-Twumasi, M., Akey, E. A., & Bannor, R. K. (2022). Rice Marketing Outlets, Commercialization, and Welfare: Insights From Rural Ghana. *Journal of International Food & Agribusiness Marketing*, 1–27. <https://doi.org/10.1080/08974438.2021.2022556>
- Arora, B., Bhagat, N., Saritha, L. R., & Arcot, S. (2020). Rice Grain Classification using Image Processing & Machine Learning Techniques. 2020 International Conference on Inventive Computation Technologies (ICICT), 205–208.
- Arumugam, K., Swathi, Y., Sanchez, D. T., Mustafa, M., Phoemchalard, C., Phasinam, K., & Okoronkwo, E. (2022). Towards applicability of machine learning techniques in agriculture and energy sector. *Materials Today: Proceedings*, 51, 2260–2263. <https://doi.org/10.1016/j.matpr.2021.11.394>
- Bakthavatchalam, K., Karthik, B., Thiruvengadam, V., Muthal, S., Jose, D., Kotecha, K., & Varadarajan, V. (2022). IoT Framework for Measurement and Precision Agriculture: Predicting the Crop Using Machine Learning Algorithms. *Technologies*, 10, 13. <https://doi.org/10.3390/technologies10010013>
- Banerjee, I., & Madhumathy, P. (2022). IoT Based Agricultural Business Model for Estimating Crop Health Management to Reduce Farmer Distress Using SVM and Machine Learning. In P. K. Pattnaik, R. Kumar, & S. Pal (Eds.), *Internet of Things*

- and Analytics for Agriculture, Volume 3* (pp. 165–183). Springer Singapore.  
[https://doi.org/10.1007/978-981-16-6210-2\\_8](https://doi.org/10.1007/978-981-16-6210-2_8)
- Bhujel, & Shakya. (2022). Rice Leaf Diseases Classification Using Discriminative Fine Tuning and CLR on EfficientNet. *Journal of Soft Computing Paradigm*.  
<https://irojournals.com/jscp/V4/I3/06.pdf>
- Bonaccorso, G. (2017). *Machine Learning Algorithms*. Packt Publishing Ltd.  
[https://play.google.com/store/books/details?id=\\_-ZDDwAAQBAJ](https://play.google.com/store/books/details?id=_-ZDDwAAQBAJ)
- Ceavichay, K., Valenzuela, D., y Cornejo, F. (2013). Caracterización Física, Tecnológica y Reológica de Tres Variedades de Arroz Pilado Ecuatoriano, Cosecha Invierno. *ESPOL*, 1(1), 1-8.
- Cinar, I., & Koklu, M. (2019). Classification of Rice Varieties Using Artificial Intelligence Methods. *International Journal of Intelligent Systems and Applications in Engineering*, 7(3), Art. 3. <https://doi.org/10.18201/ijisae.2019355381>
- Coello, V. C., y Garces, C. C. (2013). *Análisis de propiedades térmicas durante gelatinización en tres variedades de arroz iniap aplicando el calorímetro diferencial de barrido (dsc)* [Tesis de Pregrado, ESPOL].  
<http://www.dspace.espol.edu.ec/handle/123456789/21612>
- Chen, Z., Goh, H. S., Sin, K. L., Lim, K., Chung, N. K. H., & Liew, X. Y. (2021). Automated Agriculture Commodity Price Prediction System with Machine Learning Techniques. In *arXiv [cs.LG]*. arXiv. <http://arxiv.org/abs/2106.12747>
- Chipana Valero, C., Manzaneda Delgado, F. F., & Choque Tarqui, C. E. (2022). Valuación de dos variedades de arroz (*Oryza sativa* L.), en dos sistemas de

- manejo de suelos en Sapecho Alto- Beni. *Revista de Investigación e Innovación Agropecuaria y de Recursos Naturales*, 9(1), 3–9.
- Cinar, I., & Koklu, M. (2022). Identification of Rice Varieties Using Machine Learning Algorithms. *Jiangsu Nong Ye Xue Bao = Journal of Agricultural Sciences*. <https://dergipark.org.tr/en/pub/ankutbd/issue/68522/862482>
- Dheer, P., Singh, R. K., & Others. (2019). Identification of indian rice varieties using machine learning classifiers. *Plant Archives*, 19(1), 155–158.
- Dominguez-Lara, S., y Merino-Soto, C. (2018). Evaluación de las malas especificaciones en modelos de ecuaciones estructurales. *Revista Argentina de Ciencias del Comportamiento*, 10(2), 19-24.
- Elbadawi, M., Gaisford, S., & Basit, A. W. (2021). Advanced machine-learning techniques in drug discovery. *Drug Discovery Today*, 26(3), 769–777. <https://doi.org/10.1016/j.drudis.2020.12.003>
- Fernández, D., Liso, Á. A., Abades, D. P., Piñeiro, A. L., Gómez, S., Sánchez, J., Martín, C., & Vicente, L. (2021). Estrategias de riego y laboreo con aplicación de enmienda orgánica para mejorar las propiedades del suelo y las producciones de arroz (*Oryza sativa* L.). *Phytoma España: La revista profesional de sanidad vegetal*, 328, 50–53.
- Goyal, N., Kumar, S., & Saraswat, M. (2022). Detection of Unhealthy citrus leaves using Machine Learning Technique. *2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 591–595. <https://doi.org/10.1109/Confluence52989.2022.9734162>

- Greener, J. G., Kandathil, S. M., Moffat, L., & Jones, D. T. (2022). A guide to machine learning for biologists. *Nature Reviews. Molecular Cell Biology*, 23(1), 40–55. <https://doi.org/10.1038/s41580-021-00407-0>
- Herath, H.M.K.K.M.B., Karunasena, G.M.K.B., & Prematilake, R.D.D. (2022). Computer Vision for Agro-Foods: Investigating a Method for Grading Rice Grain Quality in Sri Lanka. In N. Kumar, C. Shahnaz, K. Kumar, M. Abed Mohammed, & R. S. Raw (Eds.), *Advance Concepts of Image Processing and Pattern Recognition: Effective Solution for Global Challenges* (pp. 21–34). Springer Singapore. [https://doi.org/10.1007/978-981-16-9324-3\\_2](https://doi.org/10.1007/978-981-16-9324-3_2)
- Hernández-Cuello, G., Morejón-Mesa, Y., Monzón-Monrabal, L. L., Díaz-Ruiz, D., & Domínguez-Calvo, G. (2021). Manejo durante la cosecha del arroz y su influencia en la calidad del secado industrial. *Revista*, 11(2), 39–43.
- Ibrahim, S., Zulkifli, N. A., & Sabri, N. (2019). Rice grain classification using multi-class support vector machine (SVM). IAES. <http://download.garuda.kemdikbud.go.id/article.php?article=1494270&val=151&title=Rice%20grain%20classification%20using%20multi-class%20support%20vector%20machine%20SVM>
- International Rice Research Institute (2021). Clasificación del arroz. <https://www.irri.org/our-solutions/irri-education>
- Jagtap, S. T., Phasinam, K., Kassanuk, T., Jha, S. S., Ghosh, T., & Thakar, C. M. (2022). Towards application of various machine learning techniques in agriculture. *Materials Today: Proceedings*, 51, 793–797. <https://doi.org/10.1016/j.matpr.2021.06.236>

- Jin, B., Zhang, C., Jia, L., Tang, Q., Gao, L., Zhao, G., & Qi, H. (2022). Identification of Rice Seed Varieties Based on Near-Infrared Hyperspectral Imaging Technology Combined with Deep Learning. *ACS Omega*, 7(6), 4735-4749. <https://doi.org/10.1021/acsomega.1c04102>
- Kasinathan, T., Singaraju, D., & Uyyala, S. R. (2021). Insect classification and detection in field crops using modern machine learning techniques. *Information Processing in Agriculture*, 8(3), 446–457. <https://doi.org/10.1016/j.inpa.2020.09.006>
- Khatri, A., Agrawal, S., & Chatterjee, J. M. (2022). Wheat Seed Classification: Utilizing Ensemble Machine Learning Approach. *Scientific Programming*, 2022. <https://doi.org/10.1155/2022/2626868>
- Kiratiratanapruk, K., Temniranrat, P., Sinthupinyo, W., Prempre, P., Chaitavon, K., Porntheeraphat, S., & Prasertsak, A. (2020). Development of Paddy Rice Seed Classification Process using Machine Learning Techniques for Automatic Grading Machine. *Journal of Sensors*, 2020. <https://doi.org/10.1155/2020/7041310>
- Koklu, M., Cinar, I., & Taspinar, Y. S. (2021). Classification of rice varieties with deep learning methods. *Computers and Electronics in Agriculture*, 187, 106285. <https://doi.org/10.1016/j.compag.2021.106285>
- Kok, Z. H., Mohamed, A. R., Alfatni, M. S. M., & Khairunniza-Bejo, S. (2021). Support Vector Machine in Precision Agriculture: A review. *Computers and Electronics in Agriculture*, 191, 106546. <https://doi.org/10.1016/j.compag.2021.106546>

- Komal, Sethi, G. K., & Bawa, R. K. (2022). A prototype of automatic rice variety identification system using artificial intelligence techniques. *AIP Conference Proceedings*, 2455(1), 040004. <https://doi.org/10.1063/5.0100827>
- Krishna, C. V., Suchitra, B., Sujihelen, L., Roobini, M. S., Cherukullapurath, S., & Jesudoss, A. (2022). Quality Analysis of Rice Grains Using Morphological Techniques. *2022 International Conference on Communication, Computing and Internet of Things (IC3IoT)*, 1–5. <https://doi.org/10.1109/IC3IOT53935.2022.9767925>
- Liu, J., Qiu, S., & Wei, Z. (2022). Real-Time Measurement of Moisture Content of Paddy Rice Based on Microstrip Microwave Sensor Assisted by Machine Learning Strategies. *Chemosensors*, 10(10), 376.
- Liu, W., Zeng, S., Wu, G., Li, H., & Chen, F. (2021). Rice Seed Purity Identification Technology Using Hyperspectral Image with LASSO Logistic Regression Model. *Sensors*, 21(13), Art. 13. <https://doi.org/10.3390/s21134384>
- López, G. M., Miranda, R. P., Hernández, A. G., & Sánchez, E. U. R. (2021). Rice production potential technology (*Oryza sativa* L.) in the state of Tabasco, Mexico and its contribution to food sovereignty. <https://scholar.archive.org/work/4bhg2uxagrexnau6jv2kslyoie/access/wayback/>  
<https://chapingo-cori.mx/rchsatsat/article/download/9/8>
- Malik, P., Sengupta, S., & Jadon, J. S. (2021). Comparative Analysis of Soil Properties to Predict Fertility and Crop Yield using Machine Learning Algorithms. *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 1004–1007. <https://doi.org/10.1109/Confluence51648.2021.9377147>

- Mena, Luis (2008). Aprendizaje automático a partir de conjuntos de datos no balanceados y su aplicación en el diagnóstico y pronóstico médico [tesis doctoral]. Instituto Nacional De Astrofísica, Óptica y Electrónica. <https://inaoe.repositorioinstitucional.mx/jspui/bitstream/1009/533/1/MenaCaLJ.pdf>
- Mishra, S., & Tyagi, A. K. (2022). The Role of Machine Learning Techniques in Internet of Things-Based Cloud Applications. In S. Pal, D. De, & R. Buyya (Eds.), *Artificial Intelligence-based Internet of Things Systems* (pp. 105–135). Springer International Publishing. [https://doi.org/10.1007/978-3-030-87059-1\\_4](https://doi.org/10.1007/978-3-030-87059-1_4)
- Molnar, C. (2020). *Interpretable Machine Learning*. Lulu.com. <https://play.google.com/store/books/details?id=jBm3DwAAQBAJ>
- Mourtzinis, S., Esker, P. D., Specht, J. E., & Conley, S. P. (2021). Advancing agricultural research using machine learning algorithms. *Scientific Reports*, 11(1), 17879. <https://doi.org/10.1038/s41598-021-97380-7>
- Murillo, W. J. J., Pérez, E. M. G., & Murillo, C. A. J. (2022). Aplicación de modelo matemático para la variación de ingresos de productores de arroz. La Troncal – Cañar. *Ciencia Latina Revista Científica Multidisciplinar*, 6(1), 4824–4838.
- Ndikuryayo, C., Ndayiragije, A., Kilasi, N., & Kusolwa, P. (2022). Breeding for Rice Aroma and Drought Tolerance: A Review. *Agronomy*, 12(7), 1726. <https://doi.org/10.3390/agronomy12071726>
- Nosratabadi, S., Ardabili, S., Lakner, Z., Mako, C., & Mosavi, A. (2021). Prediction of Food Production Using Machine Learning Algorithms of Multilayer Perceptron and ANFIS. *Collection FAO: Agriculture*, 11(5), 408. <https://doi.org/10.3390/agriculture11050408>

- Ochoa, R., Nava, N., & Fusil, D. (2020). Comprensión epistemológica del tesista sobre investigaciones cuantitativas, cualitativas y mixtas. *Orbis: revista de Ciencias Humanas*, 15(45), 13-22.
- Pavani, S., & Augusta Sophy Beulet, P. (2022). Prediction of Jowar Crop Yield Using K-Nearest Neighbor and Support Vector Machine Algorithms. *Futuristic Communication and Network Technologies*, 497–503. [https://doi.org/10.1007/978-981-16-4625-6\\_49](https://doi.org/10.1007/978-981-16-4625-6_49)
- Rakhra, M., Soniya, P., Tanwar, D., Singh, P., Bordoloi, D., Agarwal, P., Takkar, S., Jairath, K., & Verma, N. (2021). Crop Price Prediction Using Random Forest and Decision Tree Regression:-A Review. *Materials Today: Proceedings*. <https://doi.org/10.1016/j.matpr.2021.03.261>
- Ramadhani, F. (2022). *Mapping of multitemporal rice (Oryza sativa L.) growth stages using remote sensing with multi-sensor and machine learning: a thesis dissertation presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Earth Science at Massey University, Manawatū, New Zealand* [Massey University]. <https://mro.massey.ac.nz/handle/10179/17409>
- Ramírez Morales, I. (2018). *Estudio de aplicabilidad de técnicas de inteligencia artificial en el sector agropecuario*. <https://ruc.udc.es/dspace/handle/2183/20325>
- Rekha Sundari, M., Siva Rama Krishna, G., Sai Naveen, V., & Bharathi, G. (2021). Crop Recommendation System Using K-Nearest Neighbors Algorithm. *Proceedings of 6th International Conference on Recent Trends in Computing*, 581–589. [https://doi.org/10.1007/978-981-33-4501-0\\_54](https://doi.org/10.1007/978-981-33-4501-0_54)
- Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., Luccioni, A. S.,

- Maharaj, T., Sherwin, E. D., Mukkavilli, S. K., Kording, K. P., Gomes, C. P., Ng, A. Y., Hassabis, D., Platt, J. C., ... Bengio, Y. (2022). Tackling Climate Change with Machine Learning. *ACM Comput. Surv.*, *55*(2), 1–96. <https://doi.org/10.1145/3485128>
- Sánchez Molina, A. A., Murillo Garza, A., Sánchez Molina, A. A., & Murillo Garza, A. (2021). Enfoques metodológicos en la investigación histórica: Cuantitativa, cualitativa y comparativa. *Debates por la historia*, *9*(2), 147-181. <https://doi.org/10.54167/debates-por-la-historia.v9i2.792>
- Sultana, S., Faruque, M., & Islam, M. R. (2022). Rice grain quality parameters and determination tools: a review on the current developments and future prospects. *International Journal of Food Properties*, *25*(1), 1063–1078. <https://doi.org/10.1080/10942912.2022.2071295>
- Suresh, N., Ramesh, N. V. K., Inthiyaz, S., Priya, P. P., Nagasowmika, K., Kumar, K. V. N. H., Shaik, M., & Reddy, B. N. K. (2021). Crop Yield Prediction Using Random Forest Algorithm. *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, *1*, 279–282. <https://doi.org/10.1109/ICACCS51430.2021.9441871>
- Tosawadi, T., Kasetkasem, T., & Others. (2022). *Automatic rice plant disease evaluation method based on anomaly detection and deep learning* [Kasetsart University]. <https://ethesis.lib.ku.ac.th/dspace/handle/123456789/921>
- Vecchio, Y., Di Pasquale, J., Del Giudice, T., Pauselli, G., Masi, M., & Adinolfi, F. (2022). Precision farming: what do Italian farmers really think? An application of the Q methodology. *Agricultural Systems*, *201*, 103466. <https://doi.org/10.1016/j.agsy.2022.103466>

- Waleed, M., Um, T.-W., Kamal, T., & Usman, S. M. (2021). Classification of Agriculture Farm Machinery Using Machine Learning and Internet of Things. *Symmetry*, 13(3), 403. <https://doi.org/10.3390/sym13030403>
- Wang, H., Lei, Z., Zhang, X., Zhou, B., & Peng, J. (2016). Machine learning basics. *Deep Learning*, 98–164. [http://whdeng.cn/Teaching/PPT\\_01\\_Machine%20learning%20Basics.pdf](http://whdeng.cn/Teaching/PPT_01_Machine%20learning%20Basics.pdf)
- Xu, P., Yang, R., Zeng, T., Zhang, J., Zhang, Y., & Tan, Q. (2021). Varietal classification of maize seeds using computer vision and machine learning techniques. *Journal of Food Process Engineering*, 44(11). <https://doi.org/10.1111/jfpe.13846>
- Yang, M.-D., Hsu, Y.-C., Tseng, W.-C., Lu, C.-Y., Yang, C.-Y., Lai, M.-H., & Wu, D.-H. (2021). Assessment of Grain Harvest Moisture Content Using Machine Learning on Smartphone Images for Optimal Harvest Timing. *Sensors*, 21(17). <https://doi.org/10.3390/s21175875>
- Zahra, N., Hafeez, M. B., Nawaz, A., & Farooq, M. (2022). Rice production systems and grain quality. *Journal of Cereal Science*, 105, 103463. <https://doi.org/10.1016/j.jcs.2022.103463>
- Zambrano, C. E., & Andrade Arias, M. S. (2019). Factores que inciden en la productividad del cultivo de arroz en la provincia Los Ríos. *Revista Universidad Y.* [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S2218-36202019000500270](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2218-36202019000500270)
- Zhou, Z.-H. (2021). *Machine Learning*. Springer Nature. <https://play.google.com/store/books/details?id=ctM-EAAAQBAJ>

# UNEMI

UNIVERSIDAD ESTATAL DE MILAGRO

*¡Evolución académica!*

@UNEMIEcuador

