

Universidad de Costa Rica
Facultad de Ingeniería
Escuela de Ingeniería de Biosistemas

Análisis del uso de propiedades acústicas como método no destructivo para determinar la calidad interna de la piña (*Ananas comosus*) (L.) Merr Híbrido MD-2.

Informe de Proyecto Final de Graduación presentado a la Facultad de Ingeniería como requisito para optar por el grado de Licenciatura en Ingeniería Agrícola y de Biosistemas.

Edgar Rolando Rodríguez Siles

Ciudad Universitaria Rodrigo Facio

2022

Sometido a la Escuela de Ingeniería de Biosistemas de la Facultad de Ingeniería de la
Universidad de Costa Rica como requisito para optar por el grado de:

LICENCIADO EN INGENIERIA AGRICOLA Y DE BIOSISTEMAS

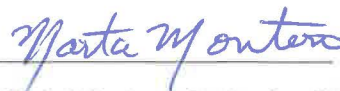


Edgar Rolando Rodríguez Siles

Sustentante



Ing. Juan Roberto Mora Chaves, Msc
Director, Comité Asesor



Ing. Marta Montero Calderón, Ph.D.
Presidente, Tribunal Examinador



Dra. Beatriz Mazón Villegas
Miembro, Comité Asesor



Dr. Hugo Solís Sánchez
Miembro, Comité Asesor



Ing. Giovanni Carmona Villalobos
Miembro, Tribunal Examinador

Dedicatoria

A mis padres Edgar Rodríguez, María Esther Siles y a mi hermana Ericka Rodríguez

Agradecimientos

Primeramente, un agradecimiento profundo a Dios por hacer posible que llegue hasta aquí, por todas las oportunidades y las vivencias tanto buenas como malas que me han hecho aprender y ser mejor que ayer.

Gracias sin dudas a mi familia por esta ahí conmigo en este camino de desarrollo personal y profesional, por brindarme todo el apoyo que requería y más, por ser una de las partes en mi vida que más significa y valor tienen para mí.

Un agradecimiento a mis amigos tanto a los de muchos años como a los que conocí en esta etapa universitaria que me hicieron crecer y aprender, que hicieron de mis días algo más cómicos y fuera de lo común.

Agradecer a mi profesor director Ing. Juan Roberto, Msc que confió este proyecto en mí, por su dirección, tiempo y el aprendizaje en estos temas que conforme más se desarrollaba más interesante se volvió haciendo. Igualmente, gracias a mis profesores asesores la Dra Beatriz Mazón y el Dr Hugo Solís que con sus consejos me llevaron a una mejor visión final para este proyecto.

Un especial agradecimiento a la Dra Marta Calderón por el tiempo y el apoyo brindado a nosotros los estudiantes durante el periodo lectivo, así como este concerniente al desarrollo del proyecto de graduación debido a las problemáticas que la pandemia creó en el desarrollo de este. A su vez agradecer al personal de la escuela que facilitó la comunicación y las acciones que conllevan esta etapa final de carrera, así como todas las previas.

Agradecer también al personal del taller como fue Ensio Pérez que me brindó apoyo y consejo para la creación de parte de mi proyecto.

Igualmente, a todos los profesores que fueron parte de mi vida universitaria un agradecimiento, por toda su paciencia y dedicación en clases y por las giras realizadas que además de permitir un mayor aprendizaje ayudó a formar lazos y contactos entre compañeros y empresas.

Agradecer a la profesora Margarita Silva y a Giuseppe Torzillo por permitirme trabajar como asistente en su laboratorio en este momento final de mi carrera ya que gracias a ello me brindó un apoyo y un mayor saber que me permitirá continuar en la dirección de lo aprendido.

ÍNDICE GENERAL

1. INTRODUCCIÓN	1
1.1. Justificación	4
1.2. Objetivos	5
1.2.1. Objetivo general	5
1.2.2. Objetivos específicos.....	5
2. MARCO TEÓRICO	6
2.1. Unidad de análisis	6
2.2. Proceso de desarrollo de la piña y factores que las afectan.	6
2.3. Parámetros de calidad para piña	7
2.4. Evaluaciones destructivas	9
2.5. Evaluaciones no destructivas	9
2.5.1. Procesamiento de imágenes y sonido.....	10
2.6. Combinación de propiedades para la definición de un parámetro de calidad.	11
2.7. Mecanismo de péndulo compuesto.....	12
2.8. Fundamentos matemáticos de la Transformada de Fourier.	13
2.9. Algoritmos de aprendizaje automático	15
2.10. Modelos de redes neuronales artificiales.	16
2.11. Validación cruzada.....	18
3. MATERIALES Y MÉTODOS	19
3.1.1. Equipo	20
3.2. Muestras de piñas.....	21
3.2.1. Estimación del grado de madurez	21
3.3. Mediciones no destructivas.....	23

3.3.1.	Mediciones de la respuesta al impulso acústico.....	23
3.3.2.	Medidas de densidad y gravedad específica	24
3.4.	Mediciones destructivas.....	25
3.5.	Procesamiento de datos.....	29
4.	RESULTADOS Y DISCUSIÓN.....	36
4.1.	Parámetros de calidad	36
4.2.	Tendencia de los parámetros de calidad en función a las frecuencias de resonancia. 37	
4.2.1.	Tendencia de la masa en función a las frecuencias de resonancia.	37
4.2.2.	Tendencia de la gravedad específica versus las frecuencias de resonancia.	38
4.2.3.	Tendencia del contenido de humedad versus las frecuencias de resonancia.	40
4.2.4.	Tendencia de la fuerza de penetración versus frecuencias de resonancia.....	41
4.2.5.	Tendencia de la firmeza en función a las frecuencias de resonancia.	43
4.2.6.	Tendencia del esfuerzo remanente en función a las frecuencias de resonancia.	45
4.3.	Modelos generales con mayor eficacia para cada parámetro de calidad.	47
4.3.1.	Resultados para los parámetros de calidad.....	49
4.3.1.1.	Esfuerzo remanente.	49
4.3.1.2.	Grado de madurez.....	51
4.3.1.3.	Firmeza.	52
4.3.1.4.	Contenido de humedad.	54
4.3.2.	Gravedad específica.	56
4.4.	Modelos de redes neuronales artificiales con mayor eficacia para cada parámetro de calidad empleando frecuencias de resonancia.	58
4.4.1.	Resultados para los parámetros de calidad obtenidos mediante redes neuronales. 59	

4.4.1.1.	Esfuerzo remanente.	59
4.4.1.2.	Grado de madurez.....	60
4.4.1.3.	Contenido de humedad.	61
4.4.1.4.	Firmeza.	62
4.4.1.5.	Gravedad específica.....	63
4.5.	Modelos predictivos con mayor eficacia para cada parámetro de calidad empleando coeficientes de rigidez.	64
4.5.1.	Resultados para los parámetros de calidad.....	66
4.5.1.1.	Grado de madurez con coeficiente de rigidez 3.	66
4.5.1.2.	Esfuerzo remanente con coeficiente de rigidez 1.	67
4.5.1.3.	Contenido de humedad con coeficiente rigidez 1.....	69
4.5.1.4.	Firmeza con coeficiente rigidez 3.....	70
4.6.	Modelos de red neuronal con mayor eficacia para cada parámetro de calidad empleando coeficientes de rigidez.....	71
4.6.1.	Resultados para los parámetros de calidad obtenidos mediante redes neuronales. 72	
4.6.1.1.	Esfuerzo remanente con coeficiente de rigidez 3.	72
4.6.1.2.	Grado de madurez con coeficiente de rigidez 1.	74
4.6.1.3.	Firmeza con coeficiente de rigidez 1.....	75
4.6.1.4.	Contenido de humedad con coeficiente de rigidez 1.....	76
4.7.	Modelos predictivos con mayor eficacia para cada parámetro de calidad empleando dos predictores.	77
4.7.1.	Matriz de confusión desarrollada para cada parámetro.....	78
4.7.1.1.	Grado de madurez con predictores f_3 y masa.	78
4.7.1.2.	Contenido de humedad con predictores f_2 y grado de madurez.....	80
4.7.1.3.	Firmeza con predictores f_2 y grado de madurez.....	82

4.7.1.4.	Gravedad específica con predictores f_2 y grado de madurez.	83
4.8.	Modelos de red neuronal con mayor eficacia para cada parámetro de calidad empleando dos predictores.....	85
4.8.1.	Resultados para los parámetros de calidad obtenidos mediante redes neuronales.	86
4.8.1.1.	Grado de madurez con predictores f_4 y masa.	86
4.8.1.2.	Esfuerzo remanente con predictores f_2 y grado de madurez.....	88
4.8.1.3.	Gravedad específica con predictores f_3 y grado de madurez.	89
4.8.1.4.	Firmeza con predictores f_2 y grado de madurez.	89
4.9.	Modelos con mayor eficacia para predicción de los parámetros de calidad a partir de tres bloques de predictores.	91
4.10.	Modelos de redes neuronales con mayor eficacia para predicción de los parámetros de calidad a partir de tres bloques de predictores.	94
4.11.	Modelos con mayor eficacia para predicción a partir de tres bloques de predictores para 5% de validación cruzada.....	96
4.12.	Factores de uniformidad para cada piña para los parámetros de calidad.	97
4.13.	Factibilidad de la implementación de una prueba acústica como método para la determinación de firmeza.....	101
5.	CONCLUSIONES	105
6.	RECOMENDACIONES.....	108
7.	BIBLIOGRAFIA.....	110
8.	ANEXOS.....	119

Índice de figuras

Figura 1. Dispositivo acústico de sobremesa para frutos grandes. (Diezma, Correa, & Ruiz-Altisen, 2010)	11
Figura 2. Péndulo físico. (Phywe, 2021)	13
Figura 3. Estructura de una matriz de confusión para una prueba de clasificación multiclase. Adaptado de (Tharwat, 2021).....	16
Figura 4. Fases del aprendizaje automático para elaborar un modelo. Adaptado de (Caladcad et al., 2020).	17
Figura 5. Mapa de procesos de la metodología elaborada para la creación de modelos predictivos y su verificación.....	19
Figura 6. a. Diagrama del mecanismo de impacto para generación del sonido. b. Fotografía del mecanismo construido.	20
Figura 7. Escala de color externo de frutos de piña, basada en la descripción realizada por (Paull & Chen ChingCheng, 2003).....	22
Figura 8. Ajuste de posición de muestra para prueba de impacto.	24
Figura 9. Representación del proceso para obtener el volumen. (a) medición práctica de volumen (b) vista de recipiente, donde la línea punteada es el nivel inicial y la línea continua el volumen final.	25
Figura 10. Área A sitio del golpe, área B ubicación para prueba con texturómetro.	26
Figura 11. Gráfica de prueba de firmeza en piña. A. Pendiente de deformación elástica B. Segunda cresta máxima perteneciente al esfuerzo máximo. C. Segmento con deformación constante.	27
Figura 12. Etapas de descomposición de audio. a. Señal de audio capturada en el dominio del tiempo. b. Espectro de la señal filtrada (FFT).	30
Figura 13. Tendencia de la relación entre las cuatro frecuencias de resonancia y las masas de las piñas.	38
Figura 14. Tendencia de la relación entre las cuatro frecuencias de resonancia y la gravedad específica de las piñas.....	39

Figura 15. Tendencia de la relación entre las cuatro frecuencias de resonancia y del contenido de humedad de las piñas.	40
Figura 16. Tendencia de la relación entre las cuatro frecuencias de resonancia y la fuerza de penetración de las piñas.	42
Figura 17. Tendencia de la relación entre las cuatro frecuencias de resonancia la firmeza de las piñas.	44
Figura 18. Tendencia de la relación entre las cuatro frecuencias de resonancia y el esfuerzo remanente de las piñas.	46
Figura 19. Matriz de confusión para el modelo de mayor eficacia con el parámetro de esfuerzo remanente con una eficacia del 39,9% y empleando frecuencia como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.	50
Figura 20. Distribución porcentual de muestras según grado de madurez para las categorías del esfuerzo remanente para piñas.	51
Figura 21. Matriz de confusión para el modelo de mayor eficacia con el parámetro de grado de madurez con una eficacia del 38,6% y empleando frecuencia como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.	52
Figura 22. Matriz de confusión para el modelo de mayor eficacia con el parámetro de firmeza con una eficacia del 30,2% y empleando frecuencia como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.	53
Figura 23. Distribución porcentual de muestras según grado de madurez para las categorías de la firmeza para piñas.	54
Figura 24. Matriz de confusión para el modelo de mayor eficacia con el parámetro de parámetro de contenido de humedad con una eficacia del 32,4% y empleando frecuencia como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.	55
Figura 25. Distribución porcentual de muestras según grado de madurez para las categorías del contenido de humedad para piñas.	55
Figura 26. Matriz de confusión para el modelo de mayor eficacia con el parámetro de gravedad específica con una eficacia del 25,9% y empleando frecuencia como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.	57

Figura 27. Distribución de muestras según grado de madurez para las categorías de la gravedad específica para piñas.	58
Figura 28. Matriz de confusión para la red neuronal de mayor eficacia con el parámetro de esfuerzo remanente con una eficacia del 39,5% y empleando frecuencia como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.	60
Figura 29. Matriz de confusión para la red neuronal de mayor eficacia con el parámetro de grado de madurez con una eficacia del 37,6% y empleando frecuencia como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.	61
Figura 30. Matriz de confusión para la red neuronal de mayor eficacia con el parámetro de contenido de humedad con una eficacia del 31,7% y empleando frecuencia como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas. .	62
Figura 31. Matriz de confusión para la red neuronal de mayor eficacia con el parámetro de firmeza con una eficacia del 28,9% y empleando frecuencia como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.	63
Figura 32. Matriz de confusión para la red neuronal de mayor eficacia con el parámetro de gravedad específica con una eficacia del 25% y empleando frecuencia como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.	64
Figura 33. Matriz de confusión para el modelo de mayor eficacia con el parámetro de grado de madurez con una eficacia del 41,4% y empleando coeficiente de rigidez 3 como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas. .	67
Figura 34. Matriz de confusión para el modelo de mayor eficacia con el parámetro de esfuerzo remanente una eficacia del 40,5% y empleando coeficiente de rigidez 1 como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.	69
Figura 35. Matriz de confusión para el modelo de mayor eficacia con el parámetro de contenido de humedad con una eficacia del 30,5% y empleando coeficiente de rigidez 1 como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.	70
Figura 36. Matriz de confusión para el modelo de mayor eficacia con el parámetro de firmeza con una eficacia del 29,7% y empleando coeficiente de rigidez 3 como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.	71

Figura 37. Matriz de confusión para el modelo de red neuronal de mayor eficacia con el parámetro de esfuerzo remanente con una eficacia del 39,5% y empleando coeficiente de rigidez 3 como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.	73
Figura 38. Matriz de confusión para el modelo de red neuronal de mayor eficacia con el parámetro de grado de madurez con una eficacia del 37,2% y empleando coeficiente de rigidez 1 como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.	74
Figura 39. Matriz de confusión para el modelo de red neuronal de mayor eficacia con el parámetro de firmeza con una eficacia del 29,3% y empleando coeficiente de rigidez 1 como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.	75
Figura 40. Matriz de confusión para el modelo de red neuronal de mayor eficacia con el parámetro de contenido de humedad con una eficacia del 29,2% y empleando coeficiente de rigidez 1 como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.	77
Figura 41. Matriz de confusión para el modelo de mayor eficacia con el parámetro de grado de madurez con una eficacia del 87,2% y empleando f_3 y masa como predictores. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.	80
Figura 42. Matriz de confusión para el modelo de mayor eficacia con el parámetro de contenido de humedad con una eficacia del 32,8% y empleando f_2 y grado de madurez como predictores. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.	81
Figura 43. Matriz de confusión para el modelo de mayor eficacia con el parámetro de firmeza con una eficacia del 32,7% y empleando f_2 y grado de madurez como predictores. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.	83
Figura 43. Matriz de confusión para el modelo de mayor eficacia con el parámetro de gravedad específica con una eficacia del 30,6% y empleando f_2 y grado de madurez como predictores. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.	85

Figura 44. Matriz de confusión para la red neuronal de mayor eficacia con el parámetro de grado de madurez con una eficacia del 51,3% y empleando f_4 y masa como predictores. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.....	87
Figura 45. Matriz de confusión para la red neuronal de mayor eficacia con el parámetro de esfuerzo remanente con una eficacia del 39,5% y empleando f_2 y grado de madurez como predictores. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.....	88
Figura 46. Matriz de confusión para la red neuronal de mayor eficacia con el parámetro de firmeza con una eficacia del 30,1% y empleando f_2 y grado de madurez como predictores. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.....	90
Figura 47. Factor de uniformidad para cada piña para el parámetro de contenido de humedad.	98
Figura 48. Factor de uniformidad para cada piña para el parámetro de fuerza de penetración.	99
Figura 49. Factor de uniformidad para cada piña para el parámetro de firmeza.	100
Figura 50. Factor de uniformidad para cada piña para el parámetro de esfuerzo remanente.	100

Índice de cuadros

Cuadro 1. Estudios aplicados de pruebas destructivas y no destructivas para la valoración de calidad.....	2
Cuadro 2. Investigaciones realizadas a productos por medio de evaluaciones no destructivas.	3
Cuadro 3. Clasificación establecida para este estudio.....	22
Cuadro 4. Cantidad total de muestras de piña por grado de madurez.	23
Cuadro 5. Parámetros de calidad con los respectivos predictores para cada bloque de estudio.	32
Cuadro 6. Datos resumen de los parámetros de calidad.	36
Cuadro 7. Tendencia de la relación entre las frecuencias de resonancia y la fuerza de penetración.....	42
Cuadro 8. Tendencia de la relación entre las frecuencias de resonancia y la fuerza de penetración.....	44
Cuadro 9. Tendencia de la relación entre las frecuencias de resonancia y la fuerza de penetración.....	46
Cuadro 10. Resultados de la eficacia general por cada parámetro de calidad obtenidos según el tipo de modelo predictivo desarrollado.	48
Cuadro 11. Resultados de la eficacia general por cada parámetro de calidad obtenidos según el tipo de modelo de red neuronal desarrollado.....	59
Cuadro 12. Resultados de la eficacia general por cada parámetro de calidad obtenidos según el tipo de modelo predictivo desarrollado y entrenados con coeficientes de rigidez.	65
Cuadro 13. Resultados de la eficacia general por cada parámetro de calidad obtenidos según el tipo de red neuronal desarrollado y entrenados con coeficientes de rigidez.	72
Cuadro 14. Resultados de la eficacia general por cada parámetro de calidad obtenidos según el tipo de modelo predictivo desarrollado y entrenados con dos predictores.	78
Cuadro 15. Resultados de la eficacia general por cada parámetro de calidad obtenidos según el tipo de red neuronal desarrollado y entrenados con dos predictores.	86
Cuadro 16. Modelos generales con mayor eficacia a partir de los tres ensayos elaborados.	91

Cuadro 17. Distribución de muestras según grado de madurez para las categorías de la gravedad específica para piñas.	94
Cuadro 18. Resultados de la eficacia general por cada parámetro de calidad obtenidos según el tipo de red neuronal desarrollado y de los tres ensayos elaborados.	95
Cuadro 19. Resultados de la eficacia general por cada parámetro de calidad obtenidos según el tipo de modelo desarrollado a partir de los tres ensayos elaborados para un 5% de validación cruzada.	96
Cuadro 20. Resultados de la eficacia general por cada parámetro de calidad obtenidos según el tipo de red neuronal desarrollado a partir de los tres ensayos elaborados para un 5% de validación cruzada.	97

Índice de anexos

Anexo 1. Lista de algoritmos empleados para la generación de modelos predictivos.	119
Anexo 2. Código para filtración de audio por Transformada Rápida de Fourier.	120
Anexo 3. Distribución de muestras según grado de madurez para las categorías de las masas para piñas.	121
Anexo 4. Distribución de muestras según grado de madurez para las categorías del contenido de humedad para piñas.	122
Anexo 5. Distribución de muestras según grado de madurez para las categorías de la gravedad específica para piñas.	122
Anexo 6. Distribución de muestras según grado de madurez para las categorías de la fuerza de penetración para piñas.	123
Anexo 7. Distribución de muestras según grado de madurez para las categorías de la firmeza para piñas.	123
Anexo 8. Distribución de muestras según grado de madurez para las categorías del esfuerzo remanente para piñas.	124
Anexo 9. Modelo Ensemble (Árboles en bolsa) para grado de madurez a 30% validación cruzada.	124
Anexo 10. Modelo Tree (Árbol grueso) para esfuerzo remanente a 30% validación cruzada.	128
Anexo 11. Modelo Tree (Árbol mediano) para contenido de humedad a 30% validación cruzada.	132
Anexo 12. Modelo máquinas de vectores de soporte (máquinas de vectores de soporte gaussiana media) para firmeza a 30% validación cruzada.	136
Anexo 13. Modelo Neural Network (Red neuronal media) para gravedad específica a 30% validación cruzada.	140
Anexo 14. Modelo Tree (Árbol grueso) para fuerza de penetración a 30% validación cruzada.	145
Anexo 15. Histograma para el contenido de humedad.	149
Anexo 16. Histograma para la gravedad específica.	149
Anexo 17. Histograma para la fuerza de penetración.	150

Anexo 18. Histograma para la firmeza.....	150
Anexo 19. Histograma para el esfuerzo remanente.....	151

Nomenclatura

TF: Transformada de Fourier.

TDF: Transformada Discreta de Fourier.

TRF: Transformada Rápida de Fourier.

TIF: Transformada Inversa de Fourier.

VC: Validación cruzada

VCGF: validación cruzada dejando grupo fuera.

TP: verdadero positivo.

TP_A : número de muestras de A clasificadas correctamente.

E_{AB} : número de muestras de A que están clasificadas incorrectamente.

FNA: falso negativo de la clase A.

V: volumen.

M: vasa.

ρ : densidad.

ρ_1 : densidad de la fruta.

ρ_1 : densidad del agua.

y_1 : valor de fuerza de orden menor correspondiente al eje y .

y_2 : valor de fuerza de orden mayor correspondiente al eje y .

x_1 : valor de tiempo de orden mayor correspondiente al eje x .

x_2 : valor de tiempo de orden mayor correspondiente al eje x .

$F(t)$: fuerza ejercida en la muestra en el tiempo.

$A(t)$: área real en el tiempo.

P_h : peso húmedo de la muestra.

P_s : peso seco de la muestra.

f_1 : primera frecuencia de resonancia.

f_2 : segunda frecuencia de resonancia.

f_3 : tercera frecuencia de resonancia.

f_4 : cuarta frecuencia de resonancia.

CR: coeficiente de rigidez.

GM: grado de madurez.

GE: gravedad específica.

P: pendiente elástica.

FM: firmeza.

ESR: esfuerzo remanente.

CH: contenido de humedad.

C_0 : límite inferior.

IP: intervalo de paso.

K : número de intervalos.

R: rango.

N : número de muestras.

A: amplitud.

X_i : marca de clase.

L_{in} : límite inferior.

L_{su} : límite superior.

L_s : límite superior de la clase.

TPR: tasas de verdadero positivo.

FNR: tasas de falsos negativos.

PPV: valores predictivos positivos.

FDR: tasas de falso descubrimiento.

FU: factor de uniformidad

RESUMEN

La piña está entre los tres principales productos agrícolas más exportados de Costa Rica en el año 2019 logró obtener un valor de \$ 930,49 millones en exportación de este producto, siendo una fuente importante de empleo para el Caribe, Zonas Norte y Sur del país. Para mantener los estándares de calidad de los mercados, es necesario que el producto posea una calidad interna determinada para antes de ser enviada a su destino. Para ello se emplean pruebas destructivas para la determinación de parámetros como firmeza, contenido de humedad, grados brix, entre otras. Para la mejora de los sistemas de aseguramientos de la calidad se estudian distintas técnicas basadas en métodos no destructivos, los cuales pueden ser implementados en la totalidad de las muestras sin comprometer su integridad para el mercado.

Esta investigación se fundamenta en el análisis de las propiedades acústicas en piña, las cuales son relacionadas a parámetros de calidad mediante algoritmos de inteligencia artificial. Para esto se empleó 193 piñas de la variedad MD-2, procedente del sector del cantón de Upala, las cuales se transportaron al laboratorio de Poscosecha y empaque de la Universidad de Costa Rica y se almacenaron a 12°C y 85% de HR. Para el estudio se empleó de 12 a 15 unidades por día a 24°C a las cuales se les retiró la corona, posteriormente se determinó el grado de madurez con base a la escala de Paull & Chen ChingCheng, (2003). Las muestras se pesaron y posteriormente se registró el sonido en cuatro puntos mediante el impacto de un péndulo compuesto, se midió el volumen, seguidamente se cortó la superficie en los cuatro puntos de estudio para aplicar el ensayo de firmeza y finalmente de estos puntos se retiró muestra la cual se pesó previo y posterior al proceso de secado en horno, el cual duró 24 h a 60°C esto para el cálculo del contenido de humedad.

A partir de los datos se analizó la tendencia entre las frecuencias y los parámetros de calidad donde se encontró una relación de frecuencias de resonancia bajas con masas, contenidos de humedad y las gravedades específicas altas, mientras que las fuerzas de penetración, firmeza y esfuerzos remanentes son menores. Posteriormente se generaron modelos con una validación cruzada al 30% a partir del uso de frecuencias de resonancia y coeficientes de rigidez como predictores contra los parámetros de calidad de grado de madurez, contenido de humedad, gravedad específica, fuerza de penetración, firmeza y esfuerzo remanente.

Finalmente se procedió a generar modelos con el uso de dos predictores. Se complementó el análisis de los resultados con matrices de confusión y factores de uniformidad. Se compararon todas las eficacias de los modelos creados, donde se destaca para el grado de madurez una eficacia de 87,2% con el modelo Conjunto (árboles agrupados).

1. INTRODUCCIÓN

La piña es una fruta perenne del grupo de las monocotiledóneas (Hurrell, Bazzano, & Delucchi, 2005) y forma parte de la familia de las bromelias (SIAP, 2018). Además, su origen se remonta en diferentes zonas de América del Sur, que abarcan Brasil, Argentina y Paraguay (UNCTAD, 2016). Por parte de Costa Rica se destinan alrededor de 44.500 ha para la producción de piña dividido en varias localidades como Zona Norte equivalente a un 56% de área cultivada, el Atlántico equivalente a un 25% y el Pacífico con un 19% (CANAPEP, 2016). Dentro de estas variedades comerciales de piña del género *Ananas* y especie *Comosus* que se cultivan en el país se encuentran cinco grupos que corresponden a la Cayena, Española, Queen, Pernanbuco y Perolera (Morales-Abarca, 2018) y dentro de estos grupos es posible encontrar más variedades como la MD2, Josefina, RL41 y Champaka (MAG, 2007).

Costa Rica en el año 2019 logró obtener un valor de \$ 930,49 millones en exportación de piña, que representa el segundo mayor valor alcanzado en un periodo de cinco años por debajo del 2018 (CANAPEP, 2020), esto gracias a la variedad Dorada (MD-2). Esta piña ha tenido una gran aceptación debido al sabor dulce, el aroma, la jugosidad y el particular color amarillo que la identifican (Montero Calderón & Cerdas Araya, 2005). Por otra parte se caracteriza por poseer una capacidad reducida para resistir enfermedades y una mayor susceptibilidad al daño mecánico (Villegas et al., 2007); por tanto hay ciertas patologías que resultan indetectables a simple vista. Además, la piña es un fruto no climatérico (Martínez González et al., 2017), lo cual dificulta la determinación de la calidad de la piña ya que sus características fenológicas no describen con precisión sus atributos de calidad internos.

Para la determinación de sólidos solubles, la acidez titulable, translucidez, el desarrollo de los frutícolos y la gravedad específica actualmente se debe recurrir a la destrucción de la fruta (Montero Calderón & Cerdas Araya, 2005). Esto permite lograr una valoración precisa de las piñas evaluadas, sin embargo, esta técnica basada en muestro no garantiza las características de todo el lote sumado a que genera un desperdicio de producto y tiempos amplios de análisis que con otras metodologías se reduciría.

Ante esta situación y otras presentes en la industria alimenticia como es la evaluación rápida del grado de madurez, la determinación del estado de la firmeza del fruto, se han llevado a cabo investigaciones con el fin de optimizar todos estos procesos, pero solo una parte de estos estudios para desarrollar pruebas no destructivas han sido implementadas en la industria. Los resultados de estos estudios trajeron consigo la aplicación de pruebas no destructivas que implementan evaluaciones a los productos por medio de imágenes, sonido, entre otros como se pueden observar en el cuadro 1 donde se ubican estudios donde se realizan pruebas tanto destructivas como no destructivas.

Cuadro 1. Estudios aplicados de pruebas destructivas y no destructivas para la valoración de calidad.

Tipo de prueba	Análisis aplicado	Referencia
Destructiva	Análisis fisicoquímico para la determinación de la calidad de las frutas	(Cazar Villacís, 2016)
Destructiva	Pruebas mecánicas y químicas para piña dorada	(Montero-Calderón, Rojas-Graü, & Martín-Belloso, 2010)
No destructivo	Aplicación de análisis por imagen para clasificación por madurez para piña ‘Phulae’	(Ullah, Chaiwong, & Saengrayap, 2018)
No destructivo	Análisis multivariado para clasificación de piña por madurez	(Pathaveerat, Terdwongworakul, & Phaungsombut, 2008a)

Las pruebas con imagen emplea referencias visuales directas de los productos ya sea por el uso de imágenes o video y a partir de estas se puede llegar a determinar el estado de madurez de manera eficaz en frutos climatéricos como lo muestran las investigaciones en banano realizadas por Surya Prabha & Satheesh Kumar (2015). Además, las imágenes permiten abarcar más aspectos como son el identificar daños externos o realizar clasificaciones por tamaño, siendo ambos análisis de fundamental importancia durante la cadena de comercialización.

Por su parte las pruebas con sonido emplean las ondas emitidas por los frutos ante la colisión de un objeto, también de los registros acústicos obtenidos por ultrasonido e infrasonido. Este tipo de evaluaciones permite encontrar resultados con mayor afinidad en características de firmeza o relacionadas con la estructura carnosa de los frutos por ello resulta complementario a las pruebas con imágenes lo cual permite diferenciar los tipos de parámetros a determinar con la metodología empleada, esto se logra apreciar a partir de los resultados logrados en otras investigaciones como se citan en el cuadro 2.

Cuadro 2. Investigaciones realizadas a productos por medio de evaluaciones no destructivas.

Producto	Parámetro de calidad	Evaluación	Referencia
Coco	Madurez	Sonido	(Caladcad et al., 2020)
Piña	Madurez	Imagen	(Ullah et al., 2018)
Banano	Madurez	Imagen	(Surya Prabha & Satheesh Kumar, 2015)
Aguacate	Madurez	Ultrasonido	(Mizrach, Flitsanov, El-Batsri, & Degani, 1998)
Naranja	Firmeza	Ultrasonido	(Morrison & Abeyratne, 2014)

Una herramienta que ha sido empleada para analizar la información de ambas pruebas no destructivas es el Deep Learning o Machine Learning. Estas lo realizan mediante algoritmos de aprendizaje automático encontrar relaciones entre las imágenes o el sonido con parámetros de calidad de manera que con otras herramientas no se lograría eficazmente. Este tipo de algoritmos incorporan la probabilidad, siendo esto una diferenciación con los modelos de regresión que establecen relaciones matemáticas definidas. Las ventajas de la incorporación de estas herramientas en la industria es la posibilidad de crear algoritmos que puedan ser incorporados en la línea de producción de tal manera que se analice la calidad de los productos que ingresan en planta. Un beneficio resultado de emplear algoritmos para el análisis es que hay una mejora en los tiempos de procesamiento, siendo de apoyo así a la mano de obra humana ya que como bien son herramientas robustas aún no sustituyen la intervención humana.

En el país se emplea el análisis por imagen mediante equipos como espectroradiómetros que son empleados para determinar firmas espectrales lo que facilita el conteo de las piñas en el campo y realizar una predicción de la cosecha. Y esta clase de soluciones están enfocadas en erradicar las incertidumbres generadas acerca del tamaño y el número de piñas cosechadas ya que el productor requiere contratar con anticipación los contenedores que se emplearán para la exportación (Vindas, 2020).

A nivel mundial se están dando saltos importantes de desarrollo y aplicación de estas tecnologías; empresas de gran renombre hacen eco sobre la necesidad de aplicarlas para poder solucionar problemáticas que afectan al ser humana. Estas necesidades dictaminan el nuevo rumbo a la alimentación donde se establece una revolución en la manera que se fabrica, transporta y consume los alimentos, lo que forma parte de la articulación de la agricultura 4.0 (Procomer, 2020). El énfasis se le pone a los productos y en profundizar en los encadenamientos para así generar valor agregado, desarrollar soluciones a sus problemáticas y la de sus consumidores.

1.1. Justificación

En una industria de producción que busca destacar en aspectos de calidad como lo es Costa Rica en cuanto al cultivo de piña, la integración de esta clase de tecnologías resulta primordial para lograr mejorar con los estándares cada vez más exigentes de los consumidores. Además, es necesario para alcanzar el ritmo en el que actualmente se mueve el mundo del comercio, donde los productos se desplazan en un ambiente de alto tráfico y con múltiples competencias. El incentivo de realizar investigaciones y pruebas en planta de dispositivos con capacidades de resolución desde el punto de vista de inteligencia artificial ha comenzado a ser necesario en múltiples industrias alimenticias. Además de asegurar un nivel de calidad del producto clasificado también se asegura que el procesamiento sea el óptimo siendo reforzada la capacidad humana con las máquinas.

Al tomar en consideración estos aspectos se procedió a seguir esta línea de pensamiento al buscar estudios que fortalecieran el conocimiento de las aplicaciones de las inteligencias artificiales para el caso de la piña dorada MD-2. Por ello este trabajo se enfocó en analizar características de calidad de la piña por métodos acústicos, esto mediante la aplicación de herramientas de Machine Learning. Una de las metas es sentar las bases para generar modelos

predictivos a partir de algoritmos de inteligencia artificial que mejor ajusten con cada parámetro de calidad y que de esta manera sea posible predecir el grado de desarrollo de las piñas que no se tenga un dato producto de alguna prueba destructiva con lo cual se estandarizara una metodología como primer acercamiento para realizar este tipo de trabajos. Para el desarrollo de este trabajo se usaron piñas que fueron donadas por Upala Agrícola ubicada en Alajuela, San Rafael, Chimurria Abajo con fecha de cosecha de noviembre del 2020.

1.2.Objetivos

1.2.1. Objetivo general

Analizar la factibilidad de una prueba acústica como método no destructivo para determinar la firmeza de la piña, *Ananas comosus*.

1.2.2. Objetivos específicos

- Obtener datos acústicos con el uso de un dispositivo no destructivo de impacto en piña.
- Generar un modelo predictivo que relacione las frecuencias acústicas con la gravedad específica, firmeza y contenido de humedad de la piña
- Verificar el modelo generado para la estimación de las frecuencias acústicas que ajuste con los atributos de calidad.

2. MARCO TEÓRICO

2.1.Unidad de análisis

La piña híbrido MD-2 es una variedad proveniente de la Cayena lisa (MAG, 2007) y se caracteriza además por ser una fruta grande en comparación a las otras variedades de piñas, con un llamativo color amarillo en su cuerpo carnoso, que posee más dulzor, aroma y jugosidad por lo que resulta atractiva para el mercado estadounidense y europeo. Esta variedad es más susceptible al daño mecánico y a la pudrición del tallo y raíces; además de requerir mayor proporción de elementos nutritivos (Cerdas Araya & Montero Calderón, 2005). Otras características sujetas a la calidad de esta piña que son destacables es el color verde oscuro de la corona y la capacidad de alcanzar la madurez de manera temprana. Además, las principales zonas de producción se ubican en Buenos Aires, San Carlos y Zona Atlántica superando las 20 000 ha (Brenes Gamboa, 2007).

2.2.Proceso de desarrollo de la piña y factores que las afectan.

Un concepto importante es la madurez fisiológica que ocurre antes de la madurez comercial en el caso de frutos climatéricos o pueden llegar a coincidir estos dos puntos para los frutos no climatéricos. Se encuentra que la madurez fisiológica es la etapa de formación de una fruta u hortaliza en la cual ha alcanzado el máximo desarrollo (FAO, n.d.-b) y es a partir de este punto que ocurre la senescencia la cual se describe como el deterioro celular que lleva a la muerte de los tejidos u órganos (Oblaré Torres, 2020). Para ciertos productos las etapas de desarrollo no son distinguibles entre sí o poseen velocidades de ocurrencia distintas (Martínez-González et al., 2017). Por otra parte la madurez comercial se encuentra sujeta al estado de desarrollo que requiere el mercado para soportar los tiempos de trasiego (Blasco Espinosa, 2014).

En el caso de la piña la madurez fisiológica la alcanza cerca de los 150 días posterior a la inducción floral, donde los grados Brix rondan un valor entre 14 a 16 unidades, la translucidez B-C, el pH con un valor mayor a 3 y con un peso entre 1,2 kg a 2,8 kg. Además, el desarrollo interno se produce de manera gradual de la parte inferior de la piña, donde está sujeta a la planta, a la parte superior por ello presenta composición química variada en distintas porciones del cuerpo, con una zona inferior madura y una superior inmadura (Villegas et al., 2007).

Posterior a esto es importante conservar y proporcionarle los cuidados necesarios para que no pierda calidad debido a que la piña sufre de una serie de enfermedades ocasionadas por desórdenes fisiológicos, genéticos y nutricionales principalmente, algunos pueden transmitirse en planta de procesamiento o la cadena de producción (Montero Calderón & Cerdas Araya, 2005). Los desórdenes fisiológicos incluyen todas las afectaciones ocasionadas por factores atmosféricos o físicos, durante el desarrollo del cultivo y por mutaciones genéticas (CANNA, 2019). Las condiciones generadas por anomalías dentro de las células se clasifican como desordenes genéticos (Mandal, 2019) y estas ocasionan malformaciones físicas y químicas (Blancard, 2011).

Un desorden nutricional puede deberse a la aplicación indebida de fertilizantes, minerales y macronutrientes (Soto Parra, Piña Ramiírez, Sánchez Chávez, Pérez Leal, & Basurto Sotelo, 2016) como el calcio, nitrógeno, fósforo, magnesio, hierro, entre otros (MAG, 2019). El efecto se evidencia en un cambio negativo en la firmeza, coloración, sabor, olor y otras características químicas y físicas que consecuentemente podrían la posibilidad de aparición de otros desordenes (Rojas Jiménez, 2008). Los desórdenes ambientales son ocasionados por agentes provenientes del clima o suelo (Chinchilla, 2004), y por agentes biológicos como enfermedades o plagas y por lesión mecánica (FAO, n.d.-a). Dentro de esta se encuentran enfermedades como pudrición negra, pudrición en el centro del frutícolo, marmoleo, daño por frío, quemaduras del sol y daños mecánicos.

Parte del crecimiento o desarrollo de estas anomalías que afectan la calidad se dan por exposición a alguna fuente de contaminación, al daño mecánico sobre la fruta que facilita el ingreso de patógenos o que aceleran los procesos de senescencia; donde se pueden presentar enfermedades y daños fisiológicos como el daño por frío, oscurecimiento interno, quemaduras de sol y en daños mecánicos producto de golpes, cortes, fricción y magulladuras por cargas (Montero Calderón & Cerdas Araya, 2005).

2.3. Parámetros de calidad para piña

Este término de calidad buscado siempre en las piñas hace referencia a un concepto aplicado en gran medida por las empresas para hacer referencia a las características o cualidades de un producto que esta a su vez determinado por los gustos y preferencias de los consumidores o clientes objetivo. Estas cualidades pueden variar de acuerdo con la educación, cultura,

disposición geográfica y otros factores, por esto debe ir dirigida enfocada a un grupo mayoritario (López Gumucio, 2005).

Dentro de las características que se evalúan para determinar la calidad se encuentra el contenido de humedad, la cual es descrita como la cantidad de agua contenida en un cuerpo o una sustancia y esta puede expresarse sobre base seca o base húmeda (EPSEM, 2021). Estas pruebas recurren a métodos termogravimétricos aplicados a la muestra de tal manera que mediante procesos de pérdida de humedad por secado, se obtiene el valor del peso perdido por medio de la diferencia de los registros entre la muestra al inicio y al final del proceso. Otra característica de importancia son los grados Brix, los cuales brindan información del dulzor que posee un producto y por ello son una unidad de medida de la densidad de una solución, que está determinada por el contenido de sacarosa en un medio, siendo entonces la medida de un grado Brix de 1 g de sacarosa en 100 g totales de solución.

La gravedad específica es una unidad adimensional que describe la razón entre las densidades de dos sustancias o cuerpos, siendo mayormente usados el aire y el agua como densidades de referencia para relacionar por la abundancia de estas (Palmer & Bibb, 2003). Este parámetro se requiere para comprobar la capacidad para flotar de un objeto en el agua, por lo que, si la razón entre las densidades de ambos cuerpos es cercana a 1, el objeto a sumergir en agua se hundirá en esta, por consiguiente, valores menores a 1 involucra que este se mantendrá a flote. Este fenómeno descrito en piñas es un indicativo de la calidad. Las frutas que flotan representan a las que están en una etapa optima de consumo y las que se hunden no lo están, además que puede involucrar que posean algún tipo de daño por enfermedad o de otro tipo.

Otro parámetro que se analiza es el color el cual resulta de la longitud de onda que se refleja de la luz. Estos tonos son producto de la cantidad de pigmentos naturales en los tejidos de los frutos, como son las antocianinas y carotenoides, además de la clorofila presente en etapas de inmadures. El color es una fuente de información sobre el estado de desarrollo del fruto así como después de la cosecha de estos ya que pueden indicar la presencia de ciertas patologías (Cazar Villacís, 2016).

La descripción dada para la firmeza desde el punto de vista alimenticio es la fuerza necesaria para generar una ruptura en los tejidos carnosos del fruto, siendo así un índice para registrar

la capacidad de resistir los daños físicos, establecer el nivel de deshidratación de los tejidos, determinar la crocantez dada por la estructura de la pared celular. Pero para poder establecer la firmeza hay que considerar factores que influyen en el registro de este como son la temperatura de almacenamiento, donde un aumento de esta provoca que la firmeza se reduzca (Cazar Villacís, 2016). El conocimiento de la firmeza es importante y requerido ya que permite mantener un control óptimo de la integridad del fruto durante el proceso de poscosecha y transporte.

Finalmente se tiene la acidez titulable que es el nivel de acidez de una sustancia y es medida de manera volumétrica para la determinación del pH (Nutricioni, 2019). Esta presencia de ácido en las piñas esta dado por el ácido cítrico y ascórbico o Vitamina C, lo que le da la propiedad acida a este fruto.

2.4. Evaluaciones destructivas

Las evaluaciones destructivas consisten en destruir el producto para extraer el material necesario para la aplicación de diferentes evaluaciones como la firmeza, rigidez, contenido de sólidos solubles totales, contenido de humedad entre otros (Cerdas Araya & Montero Calderón, 2005). Dentro de estas pruebas de calidad se encuentran la determinación del coeficiente de rigidez el cual es un valor o una magnitud física que caracteriza y cuantifica el comportamiento de elasticidad de un elemento que está sometida a una fuerza o momento y que permite dimensionar el valor de la carga, la dirección y sentido necesario para ocasionar un cambio en la estructura (Navarro Ugena & Pérez Castellanos, 2015). Además, la gravedad específica que se obtiene de la relación del peso del fluido al peso de un volumen idéntico de una sustancia referente que para líquidos es el agua a 4°C (Duarte Agudelo & Niño Vicentes, 2004). Finalmente se encuentra el contenido de humedad que es el contenido total de agua que posee un alimento, por lo que esta agua se encuentra almacenada dentro de las cavidades porosas y tejidos. El agua presente ofrece ciertas características al alimento desde la textura que puede mostrar hasta en sus propiedades reológicas (Herrera, Bolaños, & Lutz, 2003).

2.5. Evaluaciones no destructivas

Otra opción para el análisis de calidad son las pruebas no destructivas recurren a métodos en los cuales no se destruye el producto para determinar las condiciones de los atributos de

calidad (Morrison & Abeyratne, 2014). Ejemplo de este tipo de pruebas se encuentran procesamiento de imágenes y sonido (Berk, 2018a).

2.5.1. Procesamiento de imágenes y sonido

Los análisis donde se emplean reconocimientos por imágenes y por sonido en ciertas ocasiones son empleados en conjunto debido a la capacidad de relación que cada uno posee con características de las frutas en cuanto a sabor, aroma, firmeza, contenido de humedad, entre otros (Subedi & Walsh, 2009), debido a que ciertos productos agrícolas poseen una relación de su condición interna con su estado externo, de esto resulta que el análisis de sonido al interactuar con la estructura física de la fruta obtenga mejores resultados con datos sobre la firmeza (Mao, Yu, Rao, & Wang, 2016) y los análisis de imágenes relaciona con madurez (Surya Prabha & Satheesh Kumar, 2015).

En las pruebas por procesamiento de imágenes se emplea el recurso visual para tomar un registro de la apariencia física del elemento en estudio y a partir de un programa basado en inteligencia artificial (Machine Learning o Deep Learning (Martínez, 2020)) se entrena el sistema tomando en cuenta este recurso visual más otros datos que complementen el estudio de tal manera que al final el software logra reconocer los patrones de las nuevas imágenes automáticamente lo cual permite agilizar los análisis (Mathworks, 2020).

Los procesamientos del sonido buscan correlacionar la respuesta mecánica de los tejidos a un estímulo (onda o impacto) lo que se traduce en una onda acústica (Berk, 2018b) por ello se recurre a la grabación de los sonidos provenientes de una colisión o por la transmisión de un sonido a través o sobre este y mediante software de inteligencia artificial se logra automatizar procesos de clasificación por medio de la información suministrada al sistema. En el primer procedimiento el objeto se golpea sin generar un daño considerable en la integridad y mediante un micrófono ubicado en las cercanías de donde se ocasionó el golpe u en otra zona a conveniencia que permita obtener un registro acústico (Pathaveerat et al., 2008a); la figura 1 muestra una clase de dispositivo empleado para desarrollar estos estudios. Para el segundo caso se emite un sonido sobre un lado del producto o con otra ubicación y en una posición óptima de recepción acústica se coloca un micrófono, esto genera que una parte del sonido producido llegue al micrófono y la otra se pierda por el tejido del producto (Aboonajmi, Jahangiri, & Hassan-Beygi, 2015).

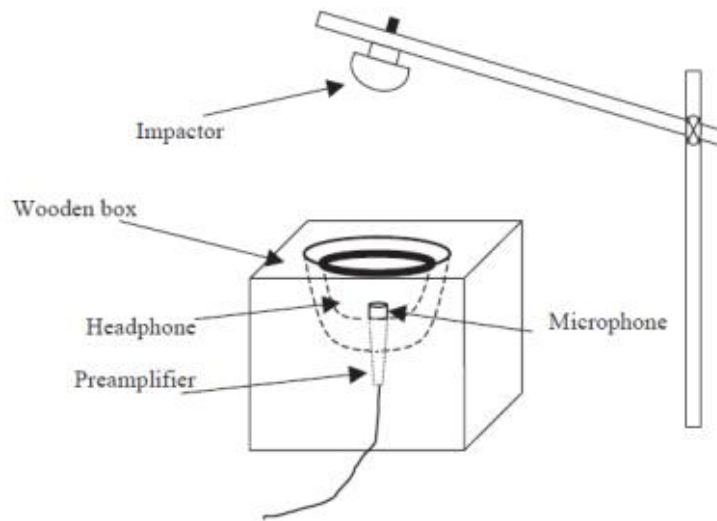


Figura 1. Dispositivo acústico de sobremesa para frutos grandes. (Diezma, Correa, & Ruiz-Altisen, 2010)

Además de estos análisis de sonido se encuentran los realizados por frecuencias ultrasónicas, los cuales implican un riesgo potencial ya que su uso pueda alterar las propiedades de la fruta y características (Bi, Hemar, Balaban, & Liao, 2015), pero en algunos casos estas alteraciones no son significativas y se emplean para determinar parámetros en productos como el coco (Caladcad et al., 2020). El empleo de ondas de sonido de baja frecuencias también causa alteraciones como el retraso de la maduración en tomates en lo cual se están empezando a realizar aplicaciones industriales para aprovechar estos efectos para mejorar el trasiego (Kim et al., 2015).

2.6. Combinación de propiedades para la definición de un parámetro de calidad.

Esta capacidad para el empleo del sonido como método de análisis no destructivo se debe a las capacidades de propagación del sonido de los objetos. Donde se puede definir la propagación como el desplazamiento que tiene el sonido por diferentes medios sean estos sólidos, líquidos y gaseosos, pero no en el vacío ya que el sonido como tal ocupa un medio para propagarse (Jaramillo Sánchez, 2004). Además, la velocidad se encuentra sujeta a la densidad, temperatura, la presión y del medio en el que se desplaza, siendo mayor en los sólidos y peor en los gases, debido a la vibración que pueden lograr las partículas constituyentes.

La relación que existe el sonido y los diferentes parámetros (contenido de humedad, firmeza, fuerza de penetración, otros) han permitido el desarrollado de estudios (cuadro 2) que emplean las respuestas acústicas de frutas para determinar un parámetro de calidad. El uso de las características propias de los tejidos permite identificar relaciones que facilitan la determinación del estado del producto sin recurrir a métodos destructivos. Esta clase de relaciones también es posible obtenerlas en análisis por imagen, donde se relaciona principalmente el aspecto físico del fruto con un grado de madurez, mismo que puede aportar información acerca de grados Brix, así como de otras características. Por lo cual el análisis por esta vía de estudio genera mejores resultados en productos climatéricos.

2.7.Mecanismo de péndulo compuesto.

Para ocasionar el sonido por medio de una colisión es posible emplear dispositivos con diferentes modos de funcionamiento como son por accionamiento neumático, mecánico, electrónico, o por la acción de la energía potencial como sucede con el péndulo compuesto (figura 2). Este consiste de un cuerpo sólido que rota libremente alrededor de un eje de sujeción ubicado en un punto que no pasa por el centro de masa (un grado de libertad) (Cortes Osorio, Chaves Osorio, & Medoza, 2014) originando a partir de esta configuración un movimiento oscilatorio o periódico que se define como repetitivo en el tiempo alrededor de un punto de equilibrio. El movimiento provoca un desplazamiento del objeto en su espacio de elongación que se devuelve a su estado de equilibrio a causa de una fuerza restauradora que está asociada a la fuerza gravitacional (Cabrerizo & Dulce, 2019).



Figura 2. Péndulo físico. (Phywe, 2021)

2.8. Fundamentos matemáticos de la Transformada de Fourier.

El sonido es captado como frecuencias que en términos acústicos se interpreta este evento como las compresiones-dilataciones del aire ocurridas por segundo (Hugh D & Roger A, 2009). Esta acción donde se dan un mayor número de eventos se denomina frecuencias altas o agudos y su contraparte como frecuencias bajas o bajos y la unidad de medida es el Hertzio (Hz), (Integral, 2020).

Para analizar estas frecuencias se emplea la Transformada de Fourier (TF) que es una transformación matemática que tiene como función transformar señales complejas (tiempo, espacial y frecuencias), por lo que se conoce como una representación de una imagen que contiene datos de diferentes magnitudes, frecuencias y fases (MathWorks, n.d.-a).

La TF de dos dimensiones se describe mediante la siguiente ecuación:

$$F(\omega_1 \omega_2) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} f(m, n) e^{-j\omega_1 m} e^{-j\omega_2 n} \quad (1)$$

Donde los términos ω_1 y ω_2 son variables de frecuencia (Hz), $F(\omega_1 \omega_2)$ es una función compleja y periódica en ω_1 y ω_2 con periodo 2π . Además, $F(0,0)$ es la suma de los valores

de $f(m, n)$ que está definida como la TF componente constante de corriente continua (MathWorks, n.d.-a).

En otros términos, la TF realiza la transformación o separación de una serie temporal compleja a sus ondas simples constituyentes que permite visualizar de manera gráfica las ondas y la interpretación del contenido dinámico de toda la serie de frecuencias. Resulta de utilidad en procesos complejos que involucran análisis de frecuencias ya que facilita la visualización del espectro estudiado.

Dentro de la TF existen más componentes asociados que están involucradas en el procesamiento de las señales como es la Transformada Discreta de Fourier (DFT) que se define, como un método numérico empleado para convertir una secuencia de números complejos N x_0, x_1, \dots, x_{N-1} que componen una señal en una nueva secuencia de números complejos N (Corn, Ross, & Khim, n.d.).

$$X_k = \sum_{n=0}^{N-1} x_n e^{-2\pi i k n / N} \quad (2)$$

Para $k = 0, 1, \dots, N - 1$.

Donde los valores x_i son pertenecientes a una función, o una señal con tiempos espaciados igual ($t = 0, 1, \dots, N - 1$).

La salida X_k recopila tanto la amplitud como la fase de una onda sinusoidal con la frecuencia en ciclos por unidad de tiempo $\frac{k}{N}$.

Descrito de una manera simple es un método numérico empleado para delimitar muestreos relacionados con las frecuencias del espectro que compone la señal, con lo cual se estudian funciones periódicas con parámetros cerrados para obtener como salida otra señal discreta. Es a partir de esta característica que se emplea para realizar multiplicación de polinomios y números enteros grandes en algoritmos de última generación (Corn et al., n.d.) y para lo cual se emplea la Transformada Rápida de Fourier (TRF). Esta última se divide en dos clases: diezmado en tiempo y diezmado en frecuencia, que es un método numérico de transformación discreta empleado para reducir el número de cálculos para N puntos de $2N^2$

a $2N \times \log_2(N)$ (Wesstein, 2020). Esta característica le permite calcular eficientemente la Transformada de Fourier Discreta (TFD) y la Inversa (TFI) eliminando gran parte de los cálculos repetitivos asociados a estos métodos numéricos.

$$X_k = \sum_{n=0}^{N/2-1} x_n^{par} e^{-2\pi i k n / (\frac{N}{2})} + e^{-2\pi i k n / N} \sum_{n=0}^{N/2-1} x_n^{impar} e^{-2\pi i k n / (\frac{N}{2})} \quad (3)$$

2.9. Algoritmos de aprendizaje automático

El siguiente proceso por seguir para tratar los datos es el uso de los algoritmos de aprendizaje automático que estos son códigos que permiten analizar y examinar conjuntos de datos y buscar o establecer en ellos patrones que los relacione y gracias a esto sea posible para el usuario realizar predicciones o clasificaciones con la información. Es con estos algoritmos que una máquina lleva consigo a cabo una tarea con instrucciones establecidas empleando un conjunto de datos de entrenamiento, el cual es una representación del bloque más grande. Por ello el uso de conjuntos de datos de mayor magnitud permite al algoritmo alcanzar valores más precisos al ser estos una representación más cercana a los fenómenos reales (Microsoft, 2021).

Existe una diferenciación entre los algoritmos de aprendizaje automático dependiendo del tipo de análisis que estos emplean en la resolución de problemas, agrupándose en el grupo de aprendizaje supervisado, aprendizaje no supervisado y el aprendizaje de refuerzo. Para este primero cada categoría del conjunto de datos el valor objetivo perteneciente a esta es conocida lo que le otorga la capacidad de desarrollar una función capaz de predecir la clase deseada (Vallalta, 2021). Para el segundo caso el entrenamiento del conjunto de datos ocurre sin tener las clases definidas permitiendo ser usado este tipo de aprendizaje para encontrar tendencias en el conjunto de datos y así crear grupos (Vallalta, 2021). El último aprendizaje utiliza algoritmos para aprender de los resultados y a partir de estos toman una serie de decisiones de acuerdo con los objetivos buscados destinando este tipo de aprendizaje para equipos automatizados (Microsoft, 2021).

Además, como método complementario y de uso regular en evaluaciones de efectividad en clasificaciones realizadas por parte de algoritmos de aprendizaje automático se encuentra la matriz de confusión (figura 3). Este es un cuadro que representa una matriz en el ámbito

estadístico análisis de datos clasificados para modelos binarios (dos clases) o modelos multiclase (más de dos clases) que permite obtener el valor de exactitud de esta relación. El contenido de la matriz está compuesto por un grupo de datos bajo control también nombrados clase real y un conjunto de datos de referencia conocidos como clase predicha (Ariza-López, Rodríguez-Avi, & Alba-Fernández, 2018).

		Clase real		
		A	B	C
Clase predicha	A	TP_A	E_{BA}	E_{CA}
	B	E_{AB}	TP_B	E_{CB}
	C	E_{AC}	E_{BC}	TP_C

Figura 3. Estructura de una matriz de confusión para una prueba de clasificación multiclase. Adaptado de (Tharwat, 2021)

Para una matriz de clasificación TP_A o TPR es la tasa de verdadero positivo o de muestras clasificadas correctamente como A, E_{AB} y E_{AC} son las muestras de A que están clasificadas incorrectamente y de su suma se obtiene la tasa de falsos negativos (FNR). Para una matriz de predicción TP_A o FDR es la tasa de verdadero positivo o de muestras predichas correctamente como A, E_{BA} y E_{CA} son las observaciones incorrectamente predichas de A y su suma es la tasa de falsos descubrimiento (FDR).

2.10. Modelos de redes neuronales artificiales.

Posterior al procesamiento del sonido por medio de la TRF las frecuencias obtenidas se ingresan a un modelo matemático que tiene como objetivo relacionar las características con un parámetro de respuesta como puede ser la frecuencia captada. Para desarrollar este modelo se puede emplear la red neuronal artificial que es un algoritmo de aprendizaje supervisado que tiene como funcionalidad simular la interconexión de las neuronas del cerebro por medio de una estructura de capas de nodos. Esta configuración del algoritmo le permite aprender de los datos lo cual le brinda la capacidad de reconocer patrones, clasificarlos y pronosticar

acontecimientos. Los entrenamientos a partir de los datos poseen un comportamiento que se encuentra estipulado por la conexión entre los elementos individuales y la regla de aprendizaje desarrollada (MathWorks, n.d.-b) siendo el proceso que se ilustra en la figura 4 las fases de aprendizaje automático que usualmente se realiza para la creación de un modelo.

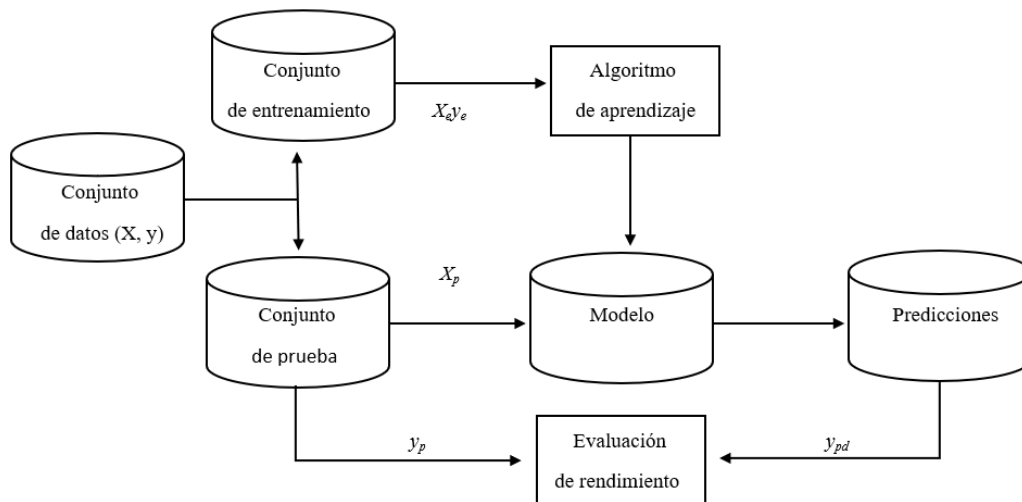


Figura 4. Fases del aprendizaje automático para elaborar un modelo. Adaptado de (Caladcad et al., 2020).

* Donde $X_e y_e$ son el conjunto de datos entrenados, $X_p y_p$ el conjunto de datos de prueba y y_{pd} el conjunto y predicho.

Para el caso de estos modelos es importante el ingreso de un conjunto de datos grande, ya que esto permite obtener mejores resultados al acrecentar la cantidad de muestras para el entrenamiento del modelo lo cual le permite conseguir mejores capacidades de generalización. Hay que tomar en cuenta que el ingreso de una cantidad masiva de datos no favorece la creación de la red neuronal artificiales. Para un mejor desarrollo se tiene en consideración la partición aplicada para la verificación de este modelo desarrollado, así como el número de variables ingresadas. Estas cumplen con funciones primordiales en el incremento de la eficacia de la red ya que como se observa en la figura 4 representan partes esenciales para la elaboración de un modelo con un rendimiento adecuado.

Los datos ingresados para entrenar los modelos predictivos se dividen en variables respuesta (dependiente) y predictoras (independiente) (Minitab, 2019), donde la variable respuesta está

sujeta a los valores o condiciones asumidas por otra variable y la variable predictora no está ligada a otras (Núñez Flores, 2007).

2.11. Validación cruzada.

Una vez entrenada la red neuronal se emplea una validación cruzada dejando grupo fuera (VCGF) el cual calcula la capacidad predictiva de los modelos lo cual da la posibilidad de determinar la cantidad de componentes necesarios para que el modelo funcione o se pueda optimizar. Esta exclusión de modelos la realiza dejando fuera varias observaciones a la vez, por ello es de gran utilidad cuando se emplean conjuntos grandes de datos ya que se reducen las iteraciones para calcular un modelo (Minitab, n.d.).

Cuando se aplica una validación cruzada se obtiene la precisión del modelo cuyo parámetro está ligado al número de observaciones con el que se entrena el modelo, por lo que pueden originarse dos casos particulares conocidos como sobreajuste (overfitting) e infraajuste (underfitting), asociado este primero a un exceso de elementos para entrenarse lo que ocasiona que la curva que describe el comportamiento de los datos se sobreajuste provocando que al ingresar datos nuevos la predicción no se realice correctamente por baja adaptabilidad del modelo a nuevos datos (Ichi.pro, 2020). De manera contraria el infraajuste es cuando el modelo es entrenado con una cantidad reducida de datos lo que impide a este llegar a una predicción correcta por la ineficiencia del ajuste del modelo (AprendeIA, 2018). El punto intermedio entre estos dos casos es lo ideal al desarrollar una red neuronal ya que permite al modelo adaptarse a diferentes circunstancias y obtener un porcentaje de precisión óptimo (Martinez, 2020).

3. MATERIALES Y MÉTODOS

El mapa de procesos de la metodología desarrollada se describe en la figura 5, donde se puede visualizar la línea de procesamiento llevada a cabo, lo que facilita su seguimiento y replicación en caso de desarrollar este estudio exploratorio.

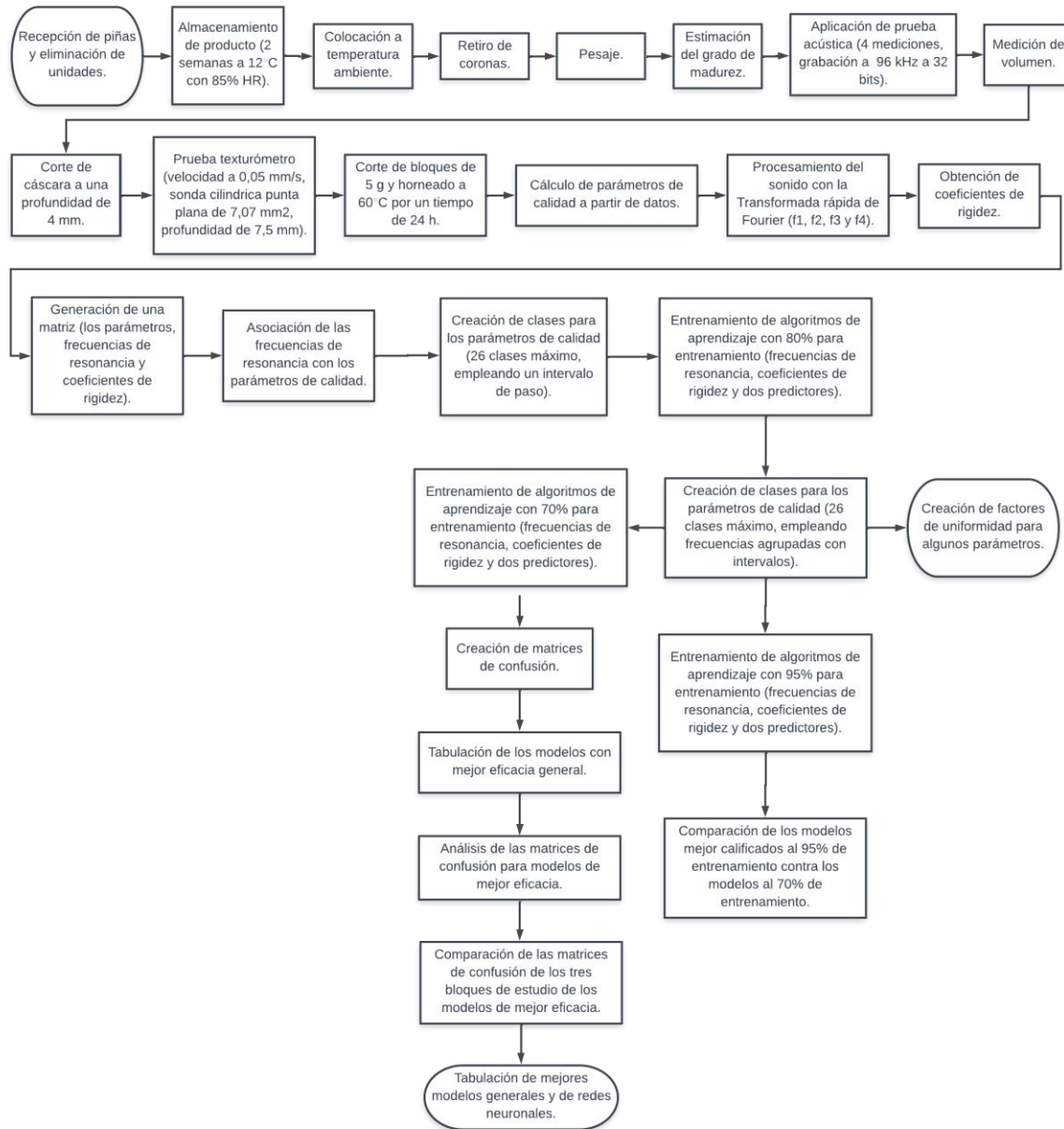
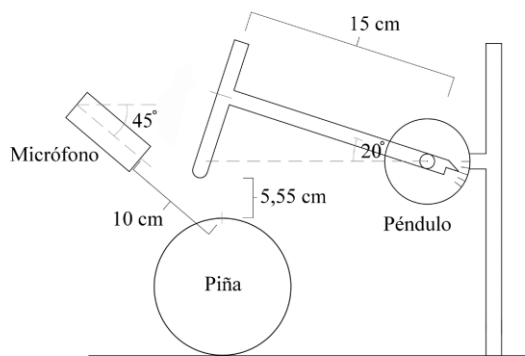


Figura 5. Mapa de procesos de la metodología elaborada para la creación de modelos predictivos y su verificación.

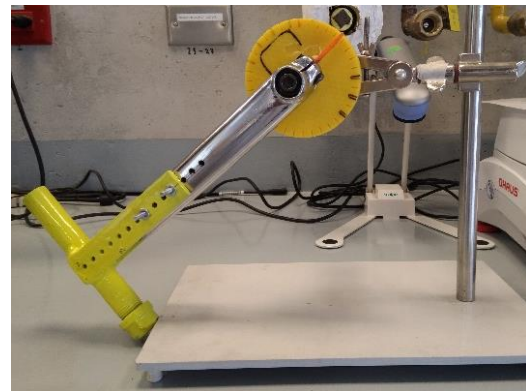
3.1.1. Equipo

Antes de la realización de las pruebas se creó con elementos de reciclaje un dispositivo manual (figura 6) el cual es basado en el principio de funcionamiento de un péndulo físico. El diseño de este equipo no generará sonidos producto del funcionamiento para colisionar con el producto. El diseño propuesto aprovecha la fuerza de gravedad para hacer caer la masa de acero sobre la piña, permitiendo moderar la intensidad del golpe al ajustarse la altura de caída por medio de una pieza que posee una distribución de ángulos radiales.

Este péndulo se sostuvo de un brazo metálico que permite cambiar la altura del sistema para poder ajustar el mecanismo a las dimensiones de la piña de tal manera que cuando la barra principal del péndulo este en posición horizontal cuando la masa se encuentra en contacto con la superficie del producto, se asegura que la caída de la masa del péndulo mantenga la misma intensidad.



(a)



(b)

Figura 6. a. Diagrama del mecanismo de impacto para generación del sonido. b. Fotografía del mecanismo construido.

3.2. Muestras de piñas

Se empleó piña de la variedad *Ananas Comosus* MD-2 en total se analizaron 199 unidades de calibres aproximados del seis al nueve. Las muestras utilizadas no presentaban daños a excepción de seis unidades que se retiraron debido a este motivo. Además el lote poseían diferentes grados de madurez del cero al seis, de acuerdo con la escala dada por Paull & Chen ChingCheng (2003). El producto provino de lugar de un cultivar del sector de Alajuela, San Rafael en Chimurria Abajo que se cosechó en noviembre de 2020, el lote fue trasladado en automóvil en cajas plásticas abiertas con un cobertor para proteger las piñas de la lluvia y del sol. Las piñas fueron llevadas a las instalaciones de la Universidad de Costa Rica donde fueron almacenadas en cámaras de refrigeración a 12°C y 85% de humedad relativa y posterior a tres días de haber llegado el producto se procedió a ejecutar las pruebas.

El procesamiento de las muestras se inició el lunes por un plazo de cinco días para dar continuidad el siguiente lunes, donde tres días en total durante el plazo de análisis se interrumpieron las evaluaciones. Diariamente se evaluaron de 12 a 15 piñas en total, donde inicialmente se sacaron de refrigeración y colocaron a condiciones ambientales de temperatura por 15 minutos para continuar con el retiro de las coronas y la determinación de la masa (m) del fruto empleando una balanza electrónica de precisión (Acculab VI-6kg).

3.2.1. Estimación del grado de madurez

El grado de madurez se estimó de manera visual según la figura 7. La clasificación de piñas se divide en seis grados de color asociados a su madurez y para este estudio se procedió a reestructurar esta clasificación en cinco categorías, agrupando las categorías cero y uno como valor 1, el dos y tres como valor dos y los siguientes grados continuarían con la sucesión de números, obteniendo así cinco categorías para el color (cuadro 3). Este cambio se aplicó para obtener un número de piñas por grado de madurez que fuera de utilidad para el desarrollo del modelo. El cuadro 4 muestra el número final de piñas de 193 unidades y al aplicarse cuatro repeticiones por piña se obtienen 773 muestras, con una proporción porcentual para el grado de madurez con respecto al total para el primer valor de un 4%, segundo valor de 15%, tercero y cuarto de 23% y 36% para el quinto valor.

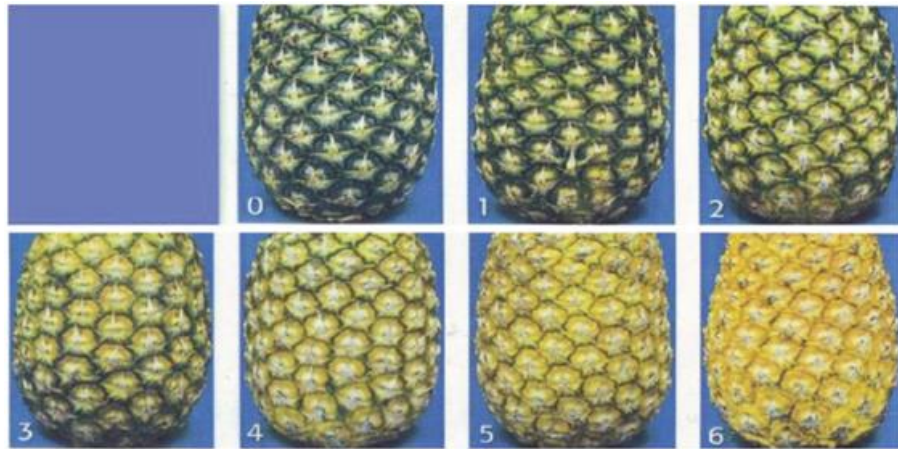


Figura 7. Escala de color externo de frutos de piña, basada en la descripción realizada por (Paull & Chen ChingCheng, 2003).

Cuadro 3. Clasificación establecida para este estudio.

Clasificación por Paull & Chen ChingCheng.	Clasificación creada para las pruebas.
0	A
1	
2	B
3	
4	C
5	D
6	E

Cuadro 4. Cantidad total de muestras de piña por grado de madurez.

Grado de madurez	Número de unidades	Proporción por grado (%)
A	7	4
B	29	15
C	44	23
D	44	23
E	69	36
Total	193	100

3.3. Mediciones no destructivas

3.3.1. Mediciones de la respuesta al impulso acústico

Inicialmente se colocó el mecanismo de impacto (figura 8) sobre una superficie rígida y se niveló la base de este para ofrecer estabilidad, además de ajustar el martillo de forma tal que el golpe fuese perpendicular a la superficie de la fruta. Hay que considerar que las pruebas acústicas fueron realizadas bajo condiciones ambientales del laboratorio. Cada piña usada se debió secar manualmente con la finalidad de reducir el ruido del agua que se generó por agua condensada y obtener datos de alta calidad y fueron colocadas con su eje fibrovascular en posición horizontal y ubicado el punto de colisión en el punto intermedio de la altura en cuatro puntos con separaciones de 90° de distancia, donde la elevación angular de la barra del péndulo corresponde con un valor de 20° con respecto a la posición horizontal de descanso sobre la superficie de la fruta proporcionando una altura para colisión de 5,55 cm (figura 6) lo que provee una fuerza de impacto de 40 N de acuerdo al registro realizado con el medidor de fuerza mecánica Chatillon DPPH-500N. Para la determinación de la fuerza mediante este equipo se aprovecha la capacidad de determinación de fuerza por compresión del instrumento, pero para esta ocasión sometiéndolo a compresión por impacto. Para la realización de este ensayo se colocó la masa del péndulo sobre el medidor del instrumento a una altura de 5,55 cm y se dejó caer, repitiendo el análisis en tres ocasiones para obtener un promedio de la fuerza de impacto. Hay que considerar que el modo de operación del equipo

es bajo condiciones controladas de velocidad, pero lo importante de este análisis es el valor máximo de fuerza registrado. Un valor de fuerza de impacto registrado para el cual se detectaron daños fueron a 5 cm de altura con una fuerza de 231 N (Montero Calderón & Cerdas Araya, 2005) empleando una esfera instrumentada ISO 100 Techmark, Inc. (acelerómetro), lo cual es un valor superior al registrado para este estudio ya que involucra el impacto del producto sobre una superficie.

La posición del micrófono (Stereon MIC-650) fue a 45° del lugar del impacto para mejorar la recepción de la señal y a una distancia de 10 cm de esta ubicación. Los eventos de sonido se grabaron a una frecuencia de muestreo de 96 kHz a 32 bits, lo que equivale a 96000 muestras/s. Posteriormente se analizaron mediante un código escrito en MatLab R2019b (anexo 2), transformó la señal en el dominio de la frecuencia por medio de una FFT.



Figura 8. Ajuste de posición de muestra para prueba de impacto.

3.3.2. Medidas de densidad y gravedad específica

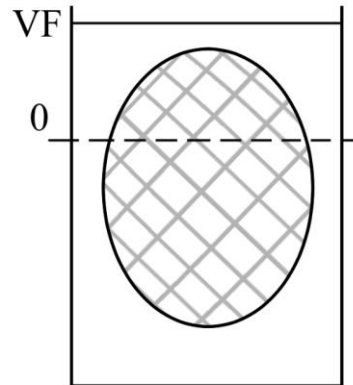
La densidad de la piña se determinó al medirse el volumen (v) desplazado en un recipiente de 7 L que se ajustó a un volumen controlado por medio de un instrumento calibrado. Este recipiente se llenó con agua potable a un volumen inicial aproximado de 4 L de tal manera que la piña quedara completamente sumergida como se muestra en la figura 9. A su vez se aplicó ligeras agitaciones a la fruta para eliminar la presencia de burbujas que puedan provocar error en la medición. La masa (m) anteriormente calculada se dividió entre el valor

de volumen de la piña que resulta de la diferencia entre el volumen inicial y final registrado para obtener la densidad como se describe en la ecuación 4.

$$Densidad = \frac{m(kg)}{v(L)} \quad (4)$$



(a)



(b)

Figura 9. Representación del proceso para obtener el volumen. (a) medición práctica de volumen (b) vista de recipiente, donde la línea punteada es el nivel inicial y la línea continua el volumen final.

Para calcular la gravedad específica se emplea la ecuación 5 la cual esta descrita por la densidad de la fruta (ρ_1) y la densidad del agua (ρ_2).

$$Gravedad\ específica = \frac{\rho_1 \left(\frac{g}{L}\right)}{\rho_2 \left(\frac{g}{L}\right)} \quad (5)$$

3.4. Mediciones destructivas

Para la realización de las pruebas destructivas se emplearon las cuatro ubicaciones usadas para las pruebas acústicas en el eje ecuatorial por producto como se puede apreciar en la figura 10, señala como área B. En la ubicación del punto de aplicación de la prueba se eliminó

la cáscara con una cortadora de rodajas a una profundidad de 4 mm con el objetivo de proporcionar una superficie uniforme para la entrada de la sonda. Para la aplicación de las pruebas de penetración se empleó el texturómetro TA.HDplus Texture Analyser a una velocidad de testeo de 0,05 mm/s, con una sonda cilíndrica de punta plana P/3 de 7,07 mm² de área que alcanzó una deformación máxima de 7,5 mm. Esta ingresó en el espacio adyacente al receptáculo floral para evitar alteraciones en la toma de datos producto de la estructura rígida irregular de este órgano (figura 10).

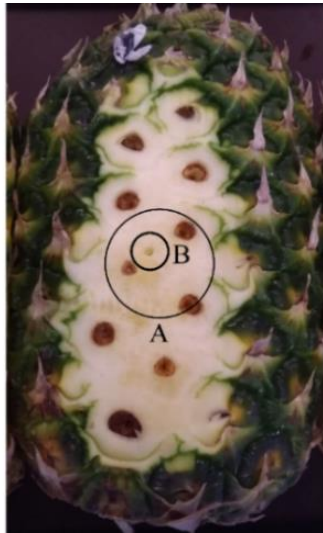


Figura 10. Área A sitio del golpe, área B ubicación para prueba con texturómetro.

A partir de los datos se calculó la fuerza de penetración, la firmeza y el esfuerzo remanente, parámetros ligados a características estructurales y de textura de la piña (figura 11); para la primera medida se utilizó un segmento de la curva de fuerza deformación ubicado en un intervalo aplicable a todo el lote de la muestra perteneciente este a la sección donde ocurre el esfuerzo elástico que sucede en la etapa inicial de la prueba de penetración, por ello el segmento señalado en la figura 11.a no llega hasta al primer pico ya que este señala un punto donde se excedió el límite elástico del material. Por tanto, la fuerza de penetración se definió para este estudio como la pendiente de deformación elástica, la cual se calculó mediante la siguiente formula.

$$Pendiente = \frac{y_2 - y_1}{x_2 - x_1} \quad (6)$$

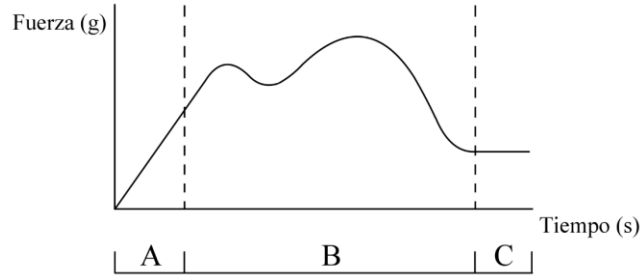


Figura 11. Grafica de prueba de firmeza en piña. A. Pendiente de deformación elástica B. Segunda cresta máxima perteneciente al esfuerzo máximo. C. Segmento con deformación constante.

Donde para la fuerza de penetración se calcula a partir de la pendiente de la línea recta inicial de la curva fuerza contra tiempo, como se mencionó con respecto a la ecuación 6.

Para la estimación de la firmeza se consideró la segunda cresta de mayor amplitud de la figura 11.B perteneciente al esfuerzo máximo encontrado para las pruebas. Esto con el objetivo de emplear el punto de mayor esfuerzo de deformación sobre la piña como dato que caracterice la firmeza del producto ya que se busca un dato que caracterice la capacidad estructural máxima del tejido y no la primera cresta perteneciente al primer fallo del tejido relacionado a la deformación plástica. Por ello se empleó la ecuación 7 (Torres, 2012) perteneciente al cálculo para el esfuerzo máximo. en el conjunto de datos perteneciente a cada piña al usarse los datos de fuerza que se ubican en el segmento mostrado en la figura 11.B, y el área de contacto del texturómetro, donde a ambas magnitudes se les realizó una conversión de unidades ya que inicialmente eran en g y cm^2 , respectivamente. Para cada conjunto de datos del ensayo con el texturómetro se determinó la fuerza máxima generada y se empleó esta para realizar el cálculo, obteniendo así un valor puntual de firmeza por prueba realizada.

$$\sigma(t) = \frac{F(t) (N)}{A(t) (m^2)} \quad (7)$$

Donde $F(t)$ es la fuerza ejercida en la muestra en el tiempo y $A(t)$ es el área real en el tiempo.

Los datos de esfuerzo remanente correspondientes al segmento final (figura 11.C) donde la fuerza requerida para la deformación se mantuvo constante. Esta condición se debe a que aún permanece tejido que se resiste y otro que es susceptible a la ruptura y por ello este segmento final de la curva fuerza deformación presenta un comportamiento constante. Para su cálculo,

se tomó un segmento de 400 mediciones para calcular un promedio de la fuerza aplicada que sea representativo de cada piña para lo cual se eligió un conjunto de datos que cumpliera la misma condición de encontrarse en la etapa de esfuerzo remanente en todo el lote de piñas.

Se debe considerar las condiciones en las que se aplicación las pruebas con el texturómetro ya que se aplicó con una sonda y velocidad específica y a una profundidad de 7,5 mm que en un posible caso de emplear una menor profundidad el registro obtenido para realizar el cálculo del esfuerzo remanente se vería reducida la cantidad de muestras empleadas. Así mismo una mayor profundidad de trabajo para la sonda ocasionaría una variación en los datos empleados para el esfuerzo remanente donde se podría esperar un comportamiento semejante al alcanzado en el presente estudio, pero quizás con una reducción en la fuerza mostrada en el último conjunto de datos.

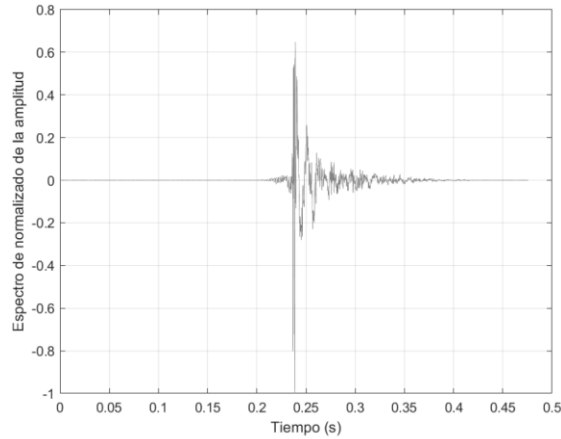
El atributo de contenido de humedad se obtuvo al extraer cuatro bloques de 5 g en ubicaciones adyacentes al punto de medición de la textura, es decir, que estuvieran cercanas del área A, como se muestra en la figura 10, ya que la sonda al ingresar rompe tejido y libera el líquido de estas. Estos bloques fueron ingresados en un horno con flujo de aire a una temperatura de 60 °C (García Pereira, Muñiz Becerá, Hernández Gómez, Mario González, & Fernández Valdés, 2013) con un tiempo en el horno de 24 h y para las muestras realizadas viernes y medidas lunes con un tiempo de 72 h. Una vez alcanzado el secado final se obtuvo el peso de cada muestra con la báscula Ohaus Adventurer AR3130 y posterior se calcula el contenido de humedad de la fruta mediante la siguiente fórmula.

$$\text{Contenido de humedad} = \frac{P_h - P_s}{P_s} \times 100 \quad (8)$$

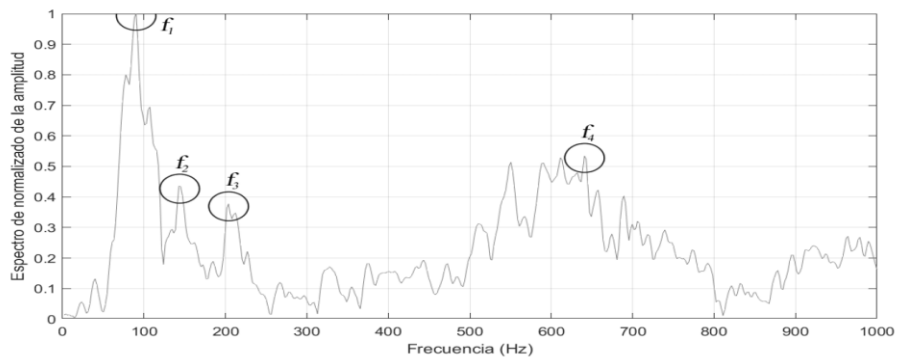
Donde P_h es la masa húmeda y P_s es la masa seca de la muestra.

3.5. Procesamiento de datos

A partir de cada uno de los cuatro registros de sonido por muestra de piña se realizó una limpieza de ruido, asociado a sonidos del equipo y del ambiente. La eliminación de sonidos ajenos al golpe sobre la piña por medio del software Audacity® 2.4.2. Posteriormente se empleó el software FLStudio® Producer Edition 12.4.1 para aplicar un recorte a los archivos de audio que son de 772 registros. Esto con el objetivo de separar los cuatro sonidos de colisión en segmentos de tiempos iguales de 0,47 s uno de los cuales se muestra en la figura 12.a. Seguidamente se ingresó cada uno de los segmentos de sonido a un código escrito (Anexo 2) en Matlab R2020a que permitió mostrar el espectro y realizar la TF sobre el registro acústico. Hay que considerar que la TF lo que permite es realizar una transformación de las series complejas a sus series más simples constituyentes al registro acústico, lo cual permite visualizar las frecuencias, sus fases y amplitudes. La figura 12.b muestra las frecuencias que constituyen el sonido obtenido de una colisión sobre la piña, donde los valores de mayor valor sobre cada una de las crestas encerradas corresponden en orden creciente con las frecuencias de resonancia con las que se trabajarán f_1 , f_2 , f_3 y f_4 , siendo las primeras tres frecuencias más cercanas al suceso del impacto y la cuarta contenida en un espectro entre 500 y 1 000 Hz que involucra un evento de ascenso en amplitud en la mayoría de las muestras obtenidas.



(a)



(b)

Figura 12. Etapas de descomposición de audio. a. Señal de audio capturada en el dominio del tiempo. b. Espectro de la señal filtrada (FFT).

Estas frecuencias de resonancia se emplearon para el cálculo del coeficiente de rigidez (CR) para cada piña mediante las ecuaciones 9, 10 y 11 para compensar la variación del peso del fruto (Chen y De Baerdemaeker, 1993).

$$\text{Coeficiente de rigidez (CR1)} = f^2 m \quad (9)$$

$$\text{Coeficiente de rigidez (CR2)} = f^2 m^{\frac{2}{3}} \quad (10)$$

$$\text{Coeficiente de rigidez (CR3)} = f^2 m^{\frac{2}{3}} \rho^{\frac{1}{3}} \quad (11)$$

Donde f es la frecuencia de orden 1 (Hz), ρ es la densidad ($\frac{kg}{m^3}$) y m es la masa de la fruta (kg).

Con los 773 datos de piñas se creó una matriz que incluye cada piña y sus respectivos datos sobre grado de madurez (GM), masa (M), contenido de humedad (CH), gravedad específica (GE), fuerza de penetración (P), firmeza (FM), y esfuerzo remanente (ESR), volumen (V), densidad (ρ), primera frecuencia resonancia (f_1), segunda frecuencia resonancia (f_2), tercera frecuencia resonancia (f_3), cuarta frecuencia resonancia (f_4), CR1 (f_1), CR2 (f_1), CR3 (f_1). El siguiente proceso con las frecuencias de resonancia consistió en un análisis de tendencias al graficarse contra los parámetros de medición de grado de madurez, masa, gravedad específica, contenido de humedad, fuerza de penetración, firmeza y esfuerzo remanente.

Para el conjunto de datos de los parámetros de calidad de GM, CH, GE, P, FM y ESR se clasificó en clases etiquetadas de la A a la Z y para esto se determinó un valor mínimo y un máximo de cada conjunto de datos perteneciente a los parámetros, y seguido a esto se estimó a conveniencia un intervalo de paso tomando en consideración no sobrepasar un número de 26 clases, con ello al tomar el valor mínimo del parámetro y sumarle el intervalo de paso se creó la primera clase y para la segunda se sumó de nuevo el intervalo a la anterior clase y así sucesivamente conforme se incrementa en la clasificación, como se describe en la ecuación 12. Además, se procedió a la eliminación de piñas que ocasionaban dificultades en la agrupación y en la precisión que se generaba en los modelos ya que poseían valores alejados de las clases ubicadas en los límites externos. El criterio de selección para la eliminación fueron los valores mínimos de la campana de Gauss.

$$C_i = C_{i-1} + IP \quad (12)$$

Donde, C_0 corresponde al límite inferior del conjunto de datos e IP es el intervalo de paso.

A partir de la matriz ingresada se procedió a generar modelos predictivos basados en algoritmos de aprendizaje los cuales se mencionan en el anexo 1, los cuales fueron tomados de Matlab R2021a de la aplicación Classification Learner. Estos modelos se generaron relacionando cada grupo de frecuencias de resonancia de manera individual (variable predictora) con una categoría (GM, GE, CH, P, FM y ESR) (variable respuesta). Aplicando una partición de 80% para el entrenamiento y 20% para la validación cruzada lo cual corresponde al bloque 1 del cuadro 5. En una evaluación semejante, CR1 (variable predictora) se relacionó con los parámetros empleados con las frecuencias a excepción de la

GE (bloque 2 del cuadro 5). Y en última instancia se genera otro bloque de modelos a partir de comparar los parámetros de calidad con dos predictores (bloque 3 del cuadro 5).

Se definió 3 bloques predictores. El bloque 1 corresponde a las $f1, f2, f3$ y $f4$ corresponden a la primera, segunda, tercera y cuarta frecuencia de resonancia. El bloque 2 corresponde a los tres coeficientes de rigidez (CR) que usan la primera frecuencia de resonancia y el bloque 3 que corresponde a la combinación de un parámetro sea esta masa (M) o grado de madurez (GM), junto a una de las frecuencias de resonancia.

Cuadro 5. Parámetros de calidad con los respectivos predictores para cada bloque de estudio.

Parámetros	Predictores		
	Bloque 1	Bloque 2	Bloque 3
Grado de madurez	$f1, f2, f3$ y $f4$	CR	M y $f1, f2, f3$ y $f4$
Contenido de humedad	$f1, f2, f3$ y $f4$	CR	GM y $f1, f2, f3$ y $f4$
Gravedad específica	$f1, f2, f3$ y $f4$	CR	GM y $f1, f2, f3$ y $f4$
Fuerza de penetración	$f1, f2, f3$ y $f4$	CR	GM y $f1, f2, f3$ y $f4$
Esfuerzo remanente	$f1, f2, f3$ y $f4$	CR	GM y $f1, f2, f3$ y $f4$

Donde CR son los coeficientes de rigidez, M es la masa, GM es el grado de madurez y f_n son las frecuencias de resonancia.

Posteriormente se procedió a crear una nueva clasificación de clases para los parámetros empleando una nueva metodología basada en frecuencias agrupadas, para lo cual se empleó el número de muestras en la ecuación 13 para obtener así el número de intervalos que luego fue redondeado a un número impar, siendo este un valor necesario para encontrar la amplitud de la clase. El rango se obtuvo de la diferencia del valor máximo y mínimo de las muestras del parámetro (ecuación 14) y al dividir este entre el número de intervalos se obtuvo la

amplitud (ecuación 15) la cual fue usada para crear las clases para cada parámetro mediante la ecuación 16 que define el margen superior de la clase y la ecuación 17 indica el valor medio que alcanza la clase. Esta posee una extensión máxima posible de 26 clases por parámetro y para esta categorización se generaron modelos con una validación de un 80% para el entrenamiento y un 20% para la validación, estos porcentajes a pesar de no ser los adecuados para trabajar se emplearon para visualizar el comportamiento de los resultados en cuanto al desarrollo de sobreajuste.

$$K = 1 + 3,332 * LN(N) \quad (13)$$

$$R = L_{su} - L_{in} \quad (14)$$

$$A = \frac{R}{K} \quad (15)$$

$$L_s = L_{in} + A \quad (16)$$

$$X_i = \frac{(L_{in} + L_s)}{2} \quad (17)$$

Donde K es el número de intervalos, N es el número de muestras, R es el rango, A es la amplitud, L_{in} es el límite inferior y L_{su} es el límite superior, L_s es el límite superior de la clase y X_i es la marca de clase.

Para mejorar la clasificación se usaron estas clases para obtener la frecuencia absoluta (muestras contenidas por grupo), que para esto se agrupó cada muestra en la clase perteneciente y se procedió a sumarlas, obteniendo así un registro específico de la cantidad de muestras por clase. A partir de la frecuencia absoluta obtenida se crea una nueva clasificación al agruparse clases para regular el número de muestras entre estas con el objetivo de que el algoritmo de aprendizaje tuviera una misma cantidad de muestras por clase. A su vez se aplicó una validación cruzada, pero con una partición de 70% para el entrenamiento del modelo y 30% para la validación de este. Estos criterios aplicados se deben a mejoras a partir de la visualización del comportamiento aplicado de la partición al 20% para validación.

Como parte de la investigación también se compararon los resultados de cada modelo para establecer el que obtuvo un mejor resultado con base a diferentes observaciones como son el tipo de modelo, el porcentaje de eficacia y las frecuencias de resonancia empleadas. Primeramente, se registró en orden decreciente de los porcentajes máximos obtenidos por parámetro de calidad con su respectivo modelo y frecuencia de resonancia, esto con el fin encontrar los parámetros con mejor y menor capacidad de relación con el sonido. Esto a su vez permitió encontrar alguna tendencia entre los parámetros de calidad y una frecuencia de resonancia en particular.

Seguidamente se elaboró un registro de los mejores resultados de eficacia por parte de las redes neuronales donde también se incluya el modelo, frecuencia de resonancia y parámetro asociado. Y a partir de estos resultados se comparó con el anterior registro para poder determinar si el orden de los parámetros y las frecuencias de resonancia empleadas es el mismo y así estimar un porcentaje de diferencia entre ambas eficacias.

A su vez el análisis se complementó con el empleo de matrices de confusión en cada modelo del registro de los valores máximos de eficacia en general y para las redes neuronales, esto para la estimación de las tasas de verdadero positivo, tasas de falsos negativos, valores predictivos positivos y tasas de falso descubrimiento. Lo cual permitió determinar la capacidad de generalización de los modelos, es decir, si estos tienen la capacidad de clasificar o realizar predicciones con todas las categorías pertenecientes a cada parámetro. Igualmente se comparó los resultados de estas matrices de confusión de ambos registros para encontrar los porcentajes de diferencia entre sus capacidades de clasificación y predicción, al igual que de generalización para obtener el mejor modelo en estas dos últimas categorías.

Seguidamente se aplicó la misma comparación de porcentajes de eficacia entre los modelos generales y de redes neuronales artificiales finales al 30% de validación cruzada con sus homólogos, pero al 5% de validación cruzada, esto con el fin de describir la efectividad de la generación de modelos al 30% de validación cruzada.

Para determinar la uniformidad de las muestras se determinó un factor de uniformidad el cual tiene como objetivo proveer información acerca de la homogeneidad de las piñas y por consiguiente valorar mejor el posible estado de eficacia de los modelos generados y a su vez

los errores que cargan para así encontrar sus posibles fuentes de origen y mejora. Por tanto, se calculó un factor de uniformidad por cada parámetro de calidad con el objetivo de estimar la uniformidad de cada piña producto de sus cuatro puntos de análisis. Se procedió a otorgar un valor del 1 al 4 donde la numeración está determinada por la cantidad de clasificaciones repetidas del parámetro de calidad que posea una piña, es decir, que si una piña para el parámetro del CH posee sus cuatro repeticiones con la misma clasificación dada por las frecuencias agrupadas esta tendrá un número de cuatro, mientras que si no posee ninguna clasificación tendrá un valor a uno.

4. RESULTADOS Y DISCUSIÓN

4.1. Parámetros de calidad

En el cuadro 6 se pueden encontrar los datos resumen de los parámetros de calidad empleados para los estudios realizados y para la elaboración de los modelos a partir de los algoritmos de inteligencia artificial. Estos datos ilustran las condiciones de la información tratada donde se destaca en primera instancia que tanto para el grado de madurez y la gravedad específica el número de muestras es menor al del resto del conjunto de parámetros y esto se debe a que únicamente se emplea un dato por muestra.

Un punto por considerar es la desviación estándar donde las que presentan mayor valor son el grado de madurez y la firmeza, principalmente este último parámetro ya que es la que muestra valores una separación de los rangos y mínimos cercano al 2 970 000 Pa. Se espera que estas condiciones resulten favorecedoras para la generación de los modelos ya que esta condición puede marcar una diferenciación marcada entre las diferentes clases elaboradas y por tanto permita categorizar con más facilidad al algoritmo de inteligencia artificial.

Cuadro 6. Datos resumen de los parámetros de calidad.

Parámetro	n	Promedio	Desviación estándar	Máximo	Mínimo	Rango
Masa (g)	186	1990,24	249,85	2724	1388	1336
Gravedad específica	186	0,89	0,02	0,923	0,78	0,14
Contenido de humedad (%)	744	84,55	2,04	99,04	69,56	29,48
Fuerza de penetración (N)	744	0,13	0,078	0,46	0,0028	0,46
Firmeza (Pa)	744	560301,33	360466,73	3116160,38	143290,65	2972869,73
Esfuerzo remanente (N)	744	1,64	0,71	6,66	0,20	6,46

4.2. Tendencia de los parámetros de calidad en función a las frecuencias de resonancia.

Este primer análisis por aplicar tiene como objetivo visualizar la relación entre las cuatro frecuencias de resonancia contra los parámetros de calidad de masa, gravedad específica, contenido de humedad, firmeza, fuerza de penetración y esfuerzo remanente. Dicha relación busca encontrar tendencias entre estos datos por lo cual no se analiza las desviaciones estándar ya que para realizar relaciones entre estos datos se requieren herramientas más robustas como las consideradas más adelante que son por algoritmos de inteligencia artificial.

4.2.1. Tendencia de la masa en función a las frecuencias de resonancia.

Los datos de frecuencia de resonancia y masa se utilizaron para identificar la relación entre estos, el mayor grado de las tendencias generadas ocurrió entre f_3 con masas de piñas del cuarto GM con un valor de pendiente del -0,0516 (figura 13) y el menor grado de pendiente surge con la relación de la primera frecuencia de resonancia y piñas del cuarto GM con un valor de pendiente de -0,0073. La figura 13 muestra que valores de baja frecuencia de cada grupo f_1 , f_2 , f_3 y f_4 se asocian a piñas con unidades de mayor masa cercanas a 2 800 g y por consiguiente valores altos de frecuencia de cada grupo 125; 345; 480 y 880 Hz respectivamente se ajustan a piñas con masas próximas a 1 300 g.

El desarrollo del fruto produce un cambio progresivo en el sonido que inicia con una acústica aguda para piñas con masas cercanas a 1 000 g cuando este posee un menor calibre y cambiará hasta alcanzar su punto de desarrollo máximo con masas cercanas a los 3 000 g donde el sonido que producirá será característicamente más grave. Este comportamiento descriptivo que indican las frecuencias de resonancia sobre las masas se mantiene en todos los cuatro conjuntos comparados como se muestra en la figura 13.

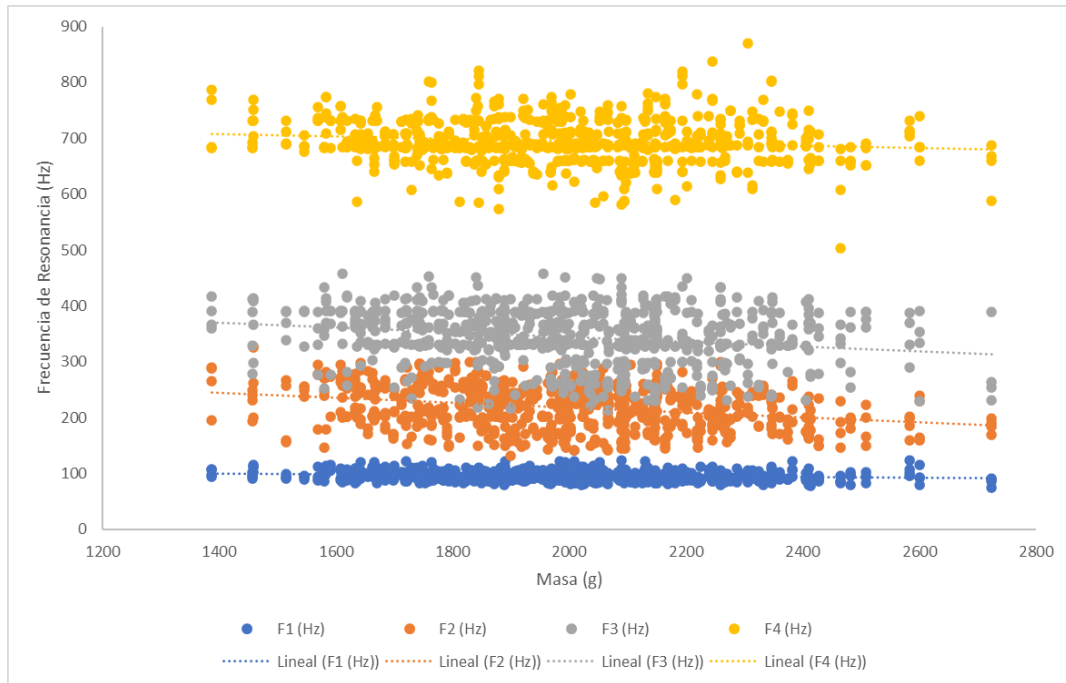


Figura 13. Tendencia de la relación entre las cuatro frecuencias de resonancia y las masas de las piñas.

Al analizar la relación de M y la frecuencia de resonancia se estima que esta tendencia tiene un comportamiento producto de varios factores dentro de los que pueden llegar a influir en cierta medida es el cambio en el CH. El CH aumenta desde el primer al último GM (figura 13) con lo cual la densidad se ve alterada y esto ocasiona que el sonido también. Medios más densos provocan que el desplazamiento del sonido se ralentice. Si bien la medición no se realizó a través del producto, el sonido se desplazó por la piña con lo cual perdió parte de su energía por absorción del producto y por ende solo una parte de este haya sido captada por el micrófono.

4.2.2. Tendencia de la gravedad específica versus las frecuencias de resonancia.

Se describen dos casos, en el primero equivalente a un 62,5% del total de muestra, los valores de GE cercanos a 0,770 están asociados a frecuencias de resonancia altos 125; 345; 480 y 880 Hz respectivamente para los cuatro grupos de frecuencias. Las frecuencias de resonancia

f_1 , f_2 y f_3 para las piñas con GE asociada al segundo GM. Mientras que todas las frecuencias para el tercer GM y por último las frecuencias f_1 y f_2 para el quinto GM.

En el segundo caso que corresponde al 37,5% restante de datos agrupados, hay valores de GE cercanos a 0,930 que están asociados a frecuencias altas. Para este segundo caso se incluye el resultado de la relación entre grados dos para la f_4 , el cuarto GM con todas las frecuencias y quinto GM para f_1 y f_4 . Además, se encuentra la mayor pendiente en piñas con GE del segundo GM comparada con f_3 con un valor de -982,02 (figura 14) y la menor pendiente para un grado dos de madurez con f_1 con un valor de pendiente de -4.73. De la relación entre las frecuencias de resonancia y la GE no se encuentra una uniformidad en los resultados que muestran los conjuntos de datos graficados ya que como se describió se presentan variedad de casos.

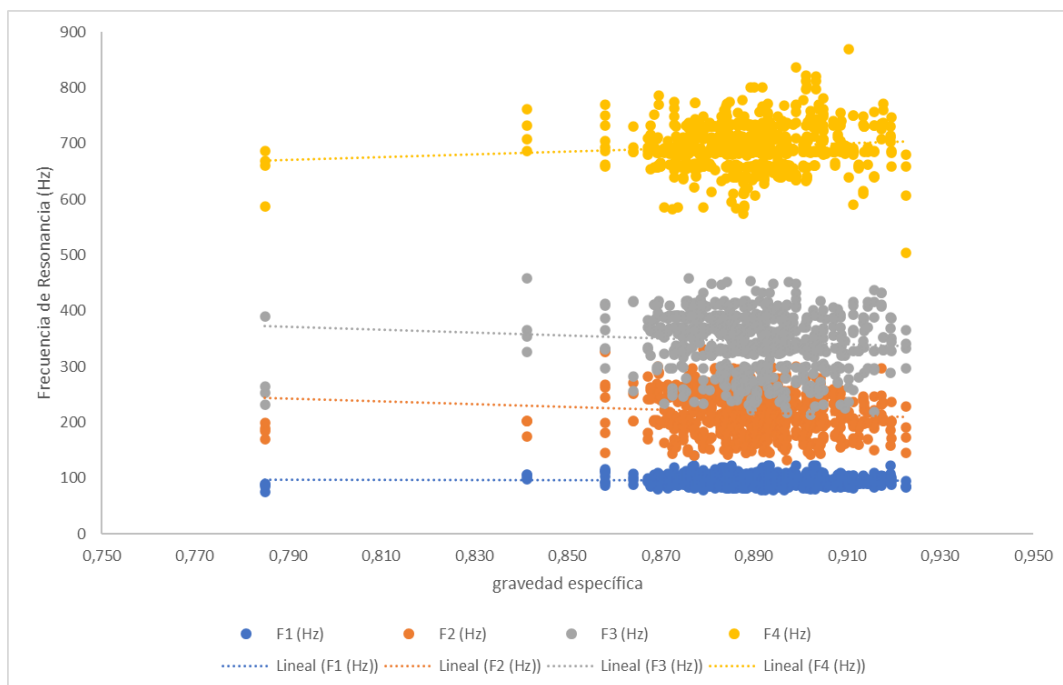


Figura 14. Tendencia de la relación entre las cuatro frecuencias de resonancia y la gravedad específica de las piñas.

Por su parte los altos valores de GE contra las frecuencias de resonancia bajas como resultado máximo describe un comportamiento que se adecúa al esperado de acuerdo con los anteriores análisis planteados, esto porque en las etapas de mayor desarrollo de madurez la GE se acerca

a uno y cuando se supera este valor se considera una piña sobremadura y por tanto se hundirán en las pilas de lavado (Montero Calderón & Cerdas Araya, 2005). Las frecuencias bajas caracterizan a piñas maduras, la tendencia en este caso tiene una correcta dirección.

4.2.3. Tendencia del contenido de humedad versus las frecuencias de resonancia.

La figura 15 muestra la relación entre el CH y las frecuencias de resonancia, esta relación no tiene un comportamiento regular entre los conjuntos de datos asociados debido a que se describen dos casos. Uno de ellos los CH cercanos al 70% se relacionan con altas frecuencias 125; 345; 480 y 880 Hz para cada grupo. Este valor bajo de CH se debe a que nueve muestras presentan valores menores al 80%. En el segundo caso las piñas con el CH cercano al 90% se relacionan con altas frecuencias. Para el cuarto GM todas las frecuencias de resonancia relacionan piñas que emiten sonidos más agudos poseen el CH menor; la mayor pendiente sucede con piñas de grado de madurez cuatro con f_3 siendo el valor de esta pendiente de -6,32. La proporción porcentual de cada caso es de un 50%, además.

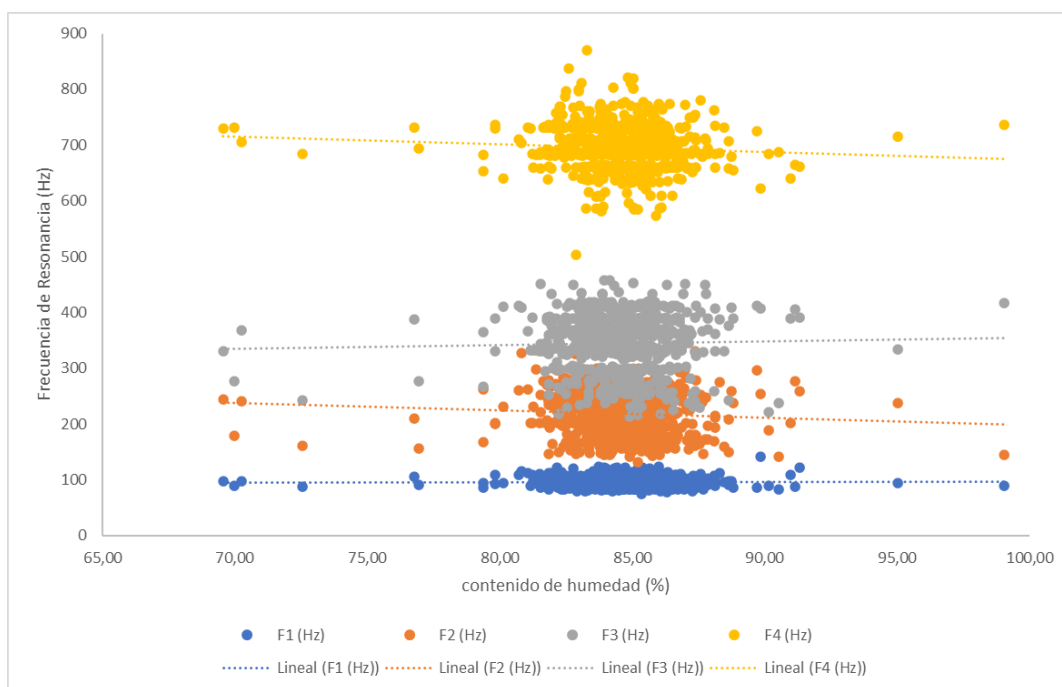


Figura 15. Tendencia de la relación entre las cuatro frecuencias de resonancia y del contenido de humedad de las piñas.

El resultado del CH para el mayor valor de pendiente coincide de acuerdo con la tendencia de M ya que esta incrementa a partir de las etapas iniciales de maduración debido al alargamiento celular que promueve la acumulación de líquidos (86,7-87%) en piñas MD-2 maduras (Morales, Hernández, Cabezas, Barrera, & Orlando, n.d.), ácidos y azúcares y por tanto este aumento del CH y de volumen son una parte de los posibles factores que generan la reducción de la frecuencia producida como repuesta al golpe.

Además de lo mencionado, el CH genera un efecto en la elasticidad de los tejidos (modelo viscoelástico) el cual es más notable conforme este aumenta en el transcurso del desarrollo del fruto ya que al aumentar la fuerza proyectada del líquido intracelular sobre la membrana celular, la elasticidad de los tejidos aumenta, ya que la fuerza de unión intracelular aumenta gracias a la ampliación del área de contacto entre estas, así como lo reportaron Ayala Aponte, Leiton Ramirez, & Serna Cock, (2017) en estudios de elasticidad con pitahaya amarilla. La elasticidad es uno de los factores relevantes cuando se genera el sonido ya que es una respuesta de la recuperación del tejido que esta a su vez está sujeta a la capacidad de transmisión del sonido, donde valores altos de esta característica aumenta la velocidad del sonido.

4.2.4. Tendencia de la fuerza de penetración versus frecuencias de resonancia.

El comportamiento de la P entre las frecuencias de resonancia no presenta uniformidad en los resultados debido a que se muestran igualmente dos tendencias en el conjunto de datos. Uno de ellos reporta bajos valores de P cercanos a -1 con bajas frecuencias de 70; 120; 200 y 560 Hz respectivamente para cada grupo. Un 56,25% del total de datos agrupados, que a su vez pertenece a la relación más elevada de pendiente correspondiente a las P de las piñas de GM tres con f_i con un valor de 0,8 (figura 16).

El segundo caso relaciona valores de P cercanos a -1 con altas frecuencias próximas a 125; 345; 480 y 880 Hz, con lo cual la presencia de este caso es de un 43,75% en relación con la totalidad de datos agrupados, por ello representa el caso de menor tendencia generada entre datos. Las pendientes resumen considerando los grupos de frecuencias de resonancia y el parámetro de fuerza de penetración se encuentra en el cuadro 7.

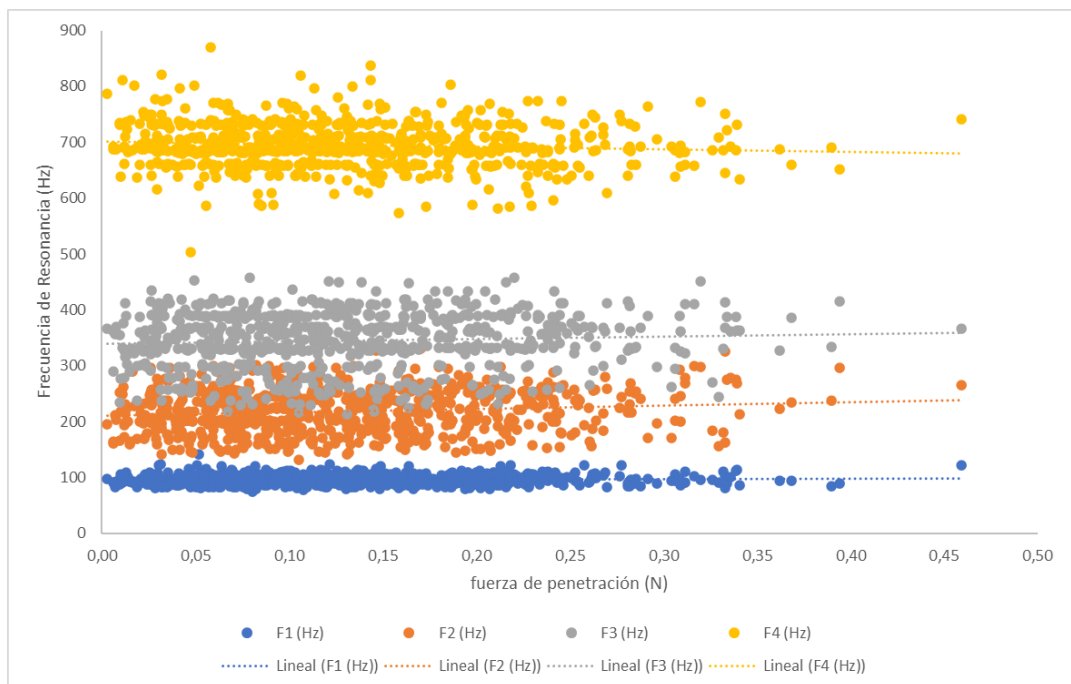


Figura 16. Tendencia de la relación entre las cuatro frecuencias de resonancia y la fuerza de penetración de las piñas.

Cuadro 7. Tendencia de la relación entre las frecuencias de resonancia y la fuerza de penetración.

Frecuencia de resonancia	Ecuación de pendiente
Primera	$y=-47,539x+701,09$
Segunda	$y=44,513x+339,03$
Tercera	$y=59,181x+210,91$
Cuarta	$y=9,3042x+94,434$

La firmeza de la pulpa presenta una reducción en el transcurso del periodo de maduración debido a un debilitamiento de las paredes celulares por el deterioro de las pectinas y hemicelulosas (Morales, Hernández, Cabezas, Barrera, & Martinez, 2001), esta pérdida de

firmeza no resulta significativa en la corteza durante el proceso de madurez. Al considerar la condición de deterioro celular que sufre la piña conforme esta se desarrolla se determina entonces que en etapas de mayor desarrollo el producto presentará menor dureza lo cual al tomar en cuenta que en estos puntos de desarrollo hay un mayor contenido de humedad, mayor masa, se puede relacionar al aspecto de presentar bajas frecuencias lo cual se ajusta al comportamiento de la tendencia, donde fuerzas de penetración menores se relacionan a bajas frecuencias.

Otro factor sujeto es la elasticidad de los tejidos que como se mencionó en el apartado anterior al aumentar la fuerza que ejerce el líquido intracelular sobre la membrana celular, la elasticidad de los tejidos aumenta, pero sumado esto a la condición de deterioro de los tejidos, la ruptura que se ocasiona por penetración aumenta. En términos generales los productos con mayor dureza presentan una mayor resonancia.

4.2.5. Tendencia de la firmeza en función a las frecuencias de resonancia.

La relación que existe entre el FM y las frecuencias de resonancia muestra dos comportamientos que involucran la relación de bajos niveles de firmeza de 1×10^4 Pa con altas frecuencias cercanas a 125, 345, 480 y 880 Hz (este caso representa un 56,25% del total de datos agrupados) y en otras se coinciden bajos niveles de FM con bajas frecuencias de resonancia próximas a 70, 120, 200 y 560 Hz este caso representa un 43,75% del total de datos agrupados. En este primer caso ocurre la mayor pendiente que resulta de la asociación de la firmeza de piñas con GM tres con f_3 (figura 17). Las pendientes resumen considerando los grupos de frecuencias de resonancia y el parámetro de firmeza se encuentra en el cuadro 8.

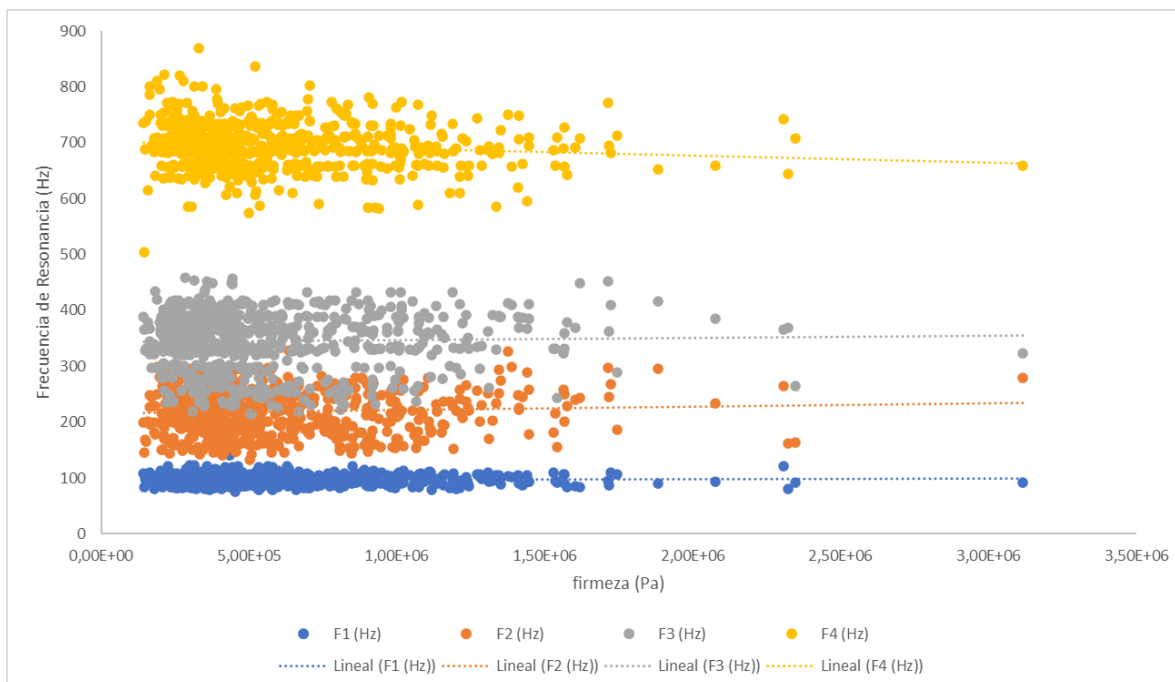


Figura 17. Tendencia de la relación entre las cuatro frecuencias de resonancia la firmeza de las piñas.

Cuadro 8. Tendencia de la relación entre las frecuencias de resonancia y la fuerza de penetración.

Frecuencia de resonancia	Ecuación de pendiente
Primera	$y = -1 \times 10^{-0,5}x + 701,92$
Segunda	$y = 4 \times 10^{-0,6}x + 342,65$
Tercera	$y = 6 \times 10^{-0,6}x + 215,26$
Cuarta	$y = 1 \times 10^{-0,6}x + 94,956$

Los resultados de la tendencia máxima alcanzada no se ajustan al comportamiento esperado, por su parte el segundo caso que representa el 43,75% sí lo realiza, ya que cuando ocurren

las frecuencias bajas se dan las FM menores debido a que los tejidos del cuerpo carnoso de la piña conforme transcurren el tiempo se deterioran por ende la FM que se puede alcanzar se reduce. Las piñas con un GM cuatro será más sencilla de dañar en comparación a una de GM uno.

A partir de la visualización de las tendencias registradas por parte de las cuatro frecuencias de resonancia contra la firmeza se determina que es posible encontrar una relación que se ajuste a un comportamiento usual de la firmeza conforme esta genera cambios durante su maduración y que al tomar en cuenta los resultados obtenidos por parte de los análisis para otros parámetros de calidad existe una factibilidad para emplear frecuencias de resonancia para determinar la firmeza de la piña pero es necesario profundizar en los siguientes análisis de este estudio para confirmar este hecho.

4.2.6. Tendencia del esfuerzo remanente en función a las frecuencias de resonancia.

La figura 18 muestra la mayor relación existente entre la ESR y las frecuencias que ocurre con el tercer GM con f_3 y con una pendiente de 6,14; donde se muestra que los esfuerzos remanentes con cercanía a los 0 N están relacionadas con las frecuencias próximas a los 70; 120; 200 y 560 Hz respectivamente a los grupos de las frecuencias de resonancia, donde este caso representa un 50% del total de datos agrupados. Por su parte el otro caso dado para ESR y las frecuencias, involucra una relación de ESR de valores cercanos a 0 N con altas frecuencias de 125; 345; 480 y 880 Hz. Las pendientes resumen considerando los grupos de frecuencias de resonancia y el parámetro de esfuerzo remanente se encuentra en el cuadro 9.

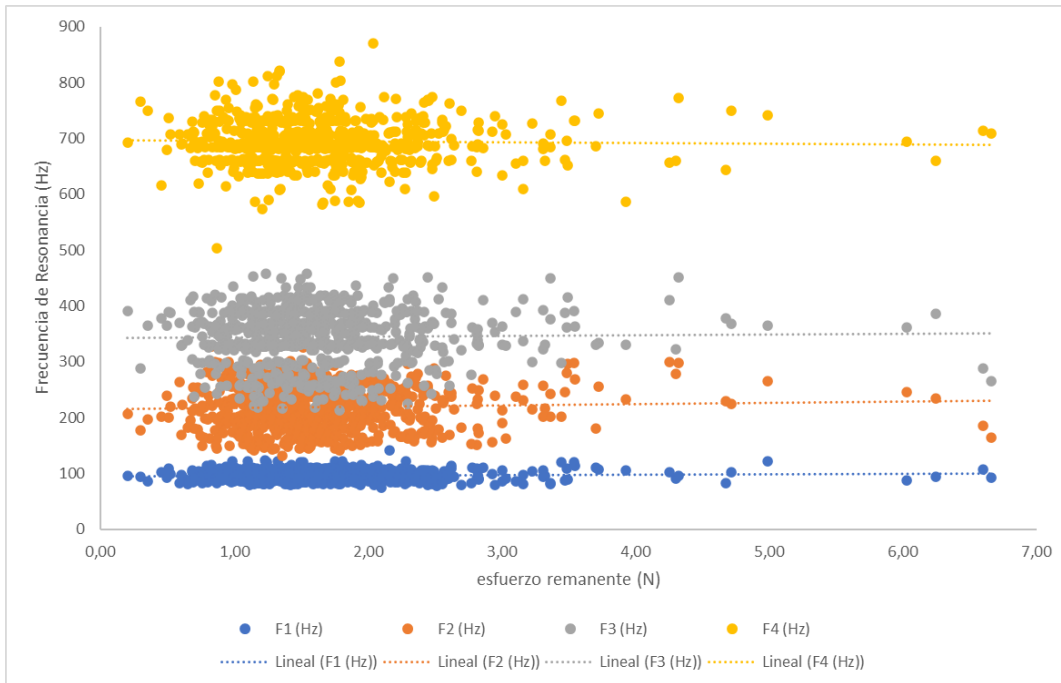


Figura 18. Tendencia de la relación entre las cuatro frecuencias de resonancia y el esfuerzo remanente de las piñas.

Cuadro 9. Tendencia de la relación entre las frecuencias de resonancia y la fuerza de penetración.

Frecuencia de resonancia	Ecuación de pendiente
Primera	$y = -1,19x + 696,86$
Segunda	$y = 1,1122x + 342,99$
Tercera	$y = 2,3667x + 214,72$
Cuarta	$y = 0,7562x + 94,409$

Este comportamiento mostrado para el caso de máxima tendencia se ajusta a lo esperado ya que como sucede con el caso de la FM y la P, el deterioro de los tejidos a lo largo del tiempo refleja una reducción en la resistencia de este, por ende, un GM cercano a 4 tenderá a marcar

una menor fuerza a la ruptura en comparación a una de GM tipo 1. Entonces existe una relación que involucra esfuerzos mayores con frecuencias mayores, lo que también como sucede con la FM esto puede estar asociado a productos con más dureza.

4.3. Modelos generales con mayor eficacia para cada parámetro de calidad.

Los siguientes resultados dados por los modelos predictivos (cuadro 10 y 11) están en función de las categorías en las que están estructurados cada parámetro de calidad. En los Anexos 2 al 8 están las categorías con la respectiva cantidad de observaciones por grado de madurez de los datos; se aprecia que hay una variedad de los cuatro GM fenómeno sucedido para los cinco grados igualmente, establecidos en cada categoría por ello no hay una linealidad definida entre la clasificación de los parámetros y la madurez de las piñas.

Hay que considerar que cada frecuencia de resonancia se analizó para cada parámetro de calidad descrito sin aplicar ninguna segmentación de GM al conjunto de datos, solo se eliminó del desarrollo de los modelos el GM uno debido a que la cantidad de muestras para este GM solo representa el 4% del total del conjunto de datos (cuadro 4), lo que no resulta efectivo para emplear en la creación de los modelos.

El cuadro 10 muestra los valores máximos de precisión obtenidos en cada parámetro de calidad con su correspondiente modelo predictivo asociado, donde el de mayor eficacia es el Árbol (árbol grueso) con un valor de 39,9% para el esfuerzo remanente y el de menor eficacia es el modelo Árbol (árbol fino) con un 14,9% y asociado a la masa. A partir de esta información se tiene que la predicción realizada por el modelo a partir de datos acústicos obtiene mejores resultados asociándolos con un parámetro destructivo relacionado a la firmeza del fruto. En caso contrario (Pathaveerat et al., 2008a) reportó los resultados con firmeza en menor grado de efectividad con respecto a otros parámetros como la GE y los CR donde los señala como potenciales métodos no destructivos para la clasificación por madurez y empleándose las capacidades de clasificación dadas de la firmeza para casos específicos donde se quiere diferenciar producto con daño por marmoleo contra un estado de madurez.

Cuadro 10. Resultados de la eficacia general por cada parámetro de calidad obtenidos según el tipo de modelo predictivo desarrollado.

Modelo	Eficacia general (%)	Frecuencia resonancia	Parámetro
Árbol (Árbol grueso)	39,9	4	Esfuerzo remanente
Bayas ingenuas (Bayas ingenuas gaussianas)	38,6	1	Grado de madurez
Conjunto (Árboles RUSBoosted)	32,4	2	Contenido de humedad
Discriminante cuadrado Discriminante cuadrático)	30,2	3	Firmeza
Conjunto (Árboles potenciados)	25,9	3	Gravedad específica
K vecinos más cercanos (K vecinos más cercanos grueso)	22,2	4	Fuerza de penetración
Árbol (Árbol fino)	14,9	2	Masa

Además, es posible obtener una determinación del GM al usar como método de procesamiento la acústica con lo que se logró obtener una eficacia de 38,6% con el modelo Bayas ingenuas (bayas ingenuas gaussianas). Hay que considerar que se obtuvo un porcentaje de acierto bajo mediante esta metodología, pero eventualmente si se empleara un procesamiento por imágenes este porcentaje podría brindar mejores resultados, pero sería requerida una comparación entre ambos métodos para su comprobación.

Las frecuencias f_1 y f_2 muestran las mejores eficacias (cuadro 10) para ensayos de pruebas no destructivas a excepción de la GE y las frecuencias f_3 y f_4 obtuvieron mejores resultados para trabajar con las pruebas destructivas la GE. El porcentaje de uso de las frecuencias para predecir parámetros de calidad es de f_1 con 14,29 %, f_2 con 28,57 %, f_3 con 28,57 % y f_4 con 28,57%, con ello se determina que la primera frecuencia resulta de menor utilidad para predecir una mayor cantidad de parámetros.

4.3.1. Resultados para los parámetros de calidad.

4.3.1.1. Esfuerzo remanente.

La Figura 19.a corresponde a la matriz para el modelo predictivo de mayor eficacia correspondiente a la ESR (cuadro 6) donde al observar la fila B desde la parte superior, se encuentra la mayor cantidad de elementos predichos correctamente con una proporción de muestras correctamente clasificadas por clase verdadera (TPR) del 99,3%. En cuanto a las otras clases el porcentaje de acierto para la capacidad de clasificación resulta bajo donde además la clase A no obtuvo resultados. La clase B también corresponde al mayor número de muestras correctamente predichas (PPV) con un 38,6%, seguidos por la clasificación D (27,3%), A (12,5%) y C (0%) e igual a como sucedió con la figura 19.a la clase A no mostró observaciones correctamente predichas.

Al considerar la información descrita por la matriz de confusión, esta no resulta de utilidad al no lograr una generalización en todas las clases, sumado a que la mayoría de las observaciones las clasifica incorrectamente como B con lo que se encuentra que este alto porcentaje de acierto en B TPR se debe a que una gran parte de los datos los dirige a esta clase por tanto no es un resultado confiable. En cuando a la capacidad de predicción sus porcentajes no resultan adecuados ya que se mantiene un acierto menor al 57% en tres de las cuatro clases, además de que están sujetos al entrenamiento que originó la capacidad de clasificación.

Clase Verdadera	A	99,5%	0,5%		
	B	99,3%	0,7%		
	C	97,5%	2,5%		
	D	99,1%		0,9%	
		A	B	C	D

Clase predicha

(a)

Clase Verdadera	A	24,6%	14,3%		
	B	39,7%	28,6%		
	C	20,9%	57,1%		
	D	14,8%		100,0%	
		A	B	C	D

Clase predicha

(b)

Figura 19. Matriz de confusión para el modelo de mayor eficacia con el parámetro de esfuerzo remanente con una eficacia del 39,9% y empleando frecuencia como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.

Por medio de las clases creadas por frecuencias agrupadas por intervalos se crearon unas distribuciones de muestras como se muestra en la figura 19 donde se describe la cantidad de muestras con diferente GM por clase, con ello se aprecia que esta tendencia del presente modelo a clasificar con respecto a la clase B se puede deber a que posee un mayor número de muestras representado principalmente por la mayor cantidad de piñas con GM E dentro de su clase. Esta disparidad puede favorecer una clasificación incorrecta de sus observaciones. Se encuentra también que la clase A es la que posee la menor cantidad de piñas lo cual puede que no sea justificable por la cantidad de muestras por GM en su clase ya que no representa la de menor valor como sí lo es la clase D. Por ello esto queda sujeto a la interpretación de datos creada por el algoritmo.

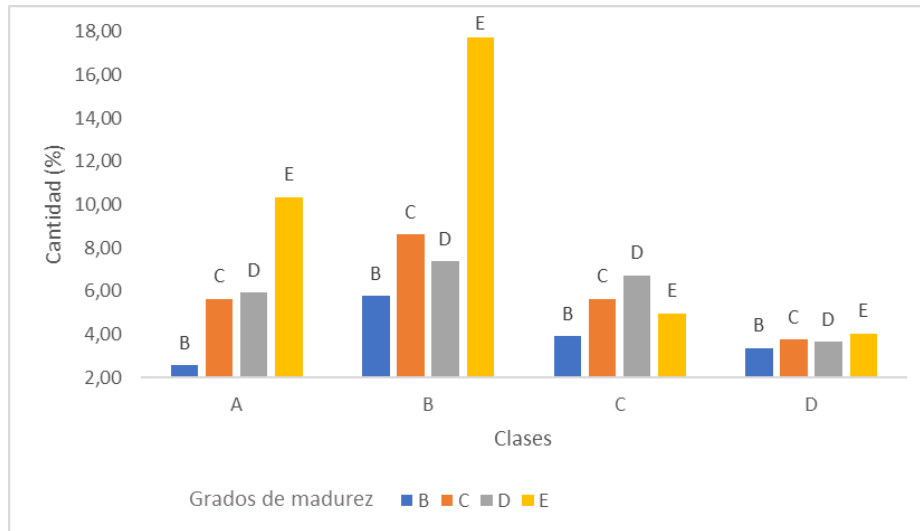


Figura 20. Distribución porcentual de muestras según grado de madurez para las categorías del esfuerzo remanente para piñas.

4.3.1.2. Grado de madurez.

Seguidamente al observar el segundo mayor porcentaje de eficacia (figura 21.a) correspondiente al GM. En la fila E, columna E se da la mayor cantidad de muestras correctamente clasificadas para la clase E con un 92,4% de observaciones correctamente clasificadas por clase verdadera (TPR) y un 7,6% de muestras mal clasificadas por clase verdadera como clase C.

Al analizar los PPV obtenidos para el modelo se encuentran porcentajes entre las clases C (36,4%) y E (38,9%) cercanos en comparación a los valores dispares de las clasificaciones por clase verdadera C (18,2%) y E (92,4%). La clase B predicha (figura 21.b) tiene un 18,2% más de eficacia con respecto a la clase B clasificada (figura 13.a) y un 53,5% para la clase E de manera contraria.

Clase Verdadera	B	9,5%		90,5%	
	C	18,2%		81,8%	
	D	13,6%		86,4%	
	E	7,6%		92,4%	
		B	C	D	E

Clase predicha

(a)

Clase Verdadera	B	12,5%		16,0%	
	C	36,4%		22,0%	
	D	27,3%		23,2%	
	E	23,9%		38,9%	
		B	C	D	E

Clase predicha

(b)

Figura 21. Matriz de confusión para el modelo de mayor eficacia con el parámetro de grado de madurez con una eficacia del 38,6% y empleando frecuencia como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.

4.3.1.3. Firmeza.

Los resultados obtenidos por parte de la firmeza mostrados en la figura 22.a indican una capacidad de trabajo únicamente para dos clases de las seis pertenecientes a este parámetro donde además existe una tendencia a clasificar hacia la clase B, por ello esta es la que presenta mejor porcentaje de observaciones correctamente clasificadas por clase verdadera (TPR) con un 94,5% seguida por un 13,1% para la clase C.

En cuanto a los resultados PPV también indican resultados únicamente en las clases B y C, las cuales a su vez poseen porcentajes de observaciones correctamente predichas de 30,6 y 26,6% para B y C. En este caso no se evidencia una tendencia a clasificar por alguna clase.

Clase Verdadera	A	96,5%	3,5%				
	B	94,5%	5,5%				
	C	86,9%	13,1%				
	D	87,9%	12,1%				
	E	86,0%	14,0%				
	F	98,0%	2,0%				
		A	B	C	D	E	F

Clase predicha

(a)

Clase Verdadera	A	16,0%	6,2%				
	B	30,6%	18,8%				
	C	16,6%	26,6%				
	D	16,0%	23,4%				
	E	13,5%	23,4%				
	F	7,2%	1,6%				
		A	B	C	D	E	F

Clase predicha

(b)

Figura 22. Matriz de confusión para el modelo de mayor eficacia con el parámetro de firmeza con una eficacia del 30,2% y empleando frecuencia como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.

Esta condición por clasificar las muestras a la clase B puede deberse a una cantidad mayor de datos por parte de esta clase como se muestra en la distribución de muestras de la figura 23. Además, se suma a esta situación una interpretación del algoritmo al integrar las demás muestras dentro de esta clasificación, pero no es un indicativo claro en cuanto al comportamiento mostrado de no generar resultados por parte de las clases A, D, E y F. Lo cual como primer resultado generado para el empleo de frecuencias como método de asociación con el parámetro de firmeza resulta limitado en cuanto a su capacidad de trabajo con diferentes grados de FM, pero para generar una mejor conclusión al respecto es necesario evaluar los siguientes resultados para este parámetro.

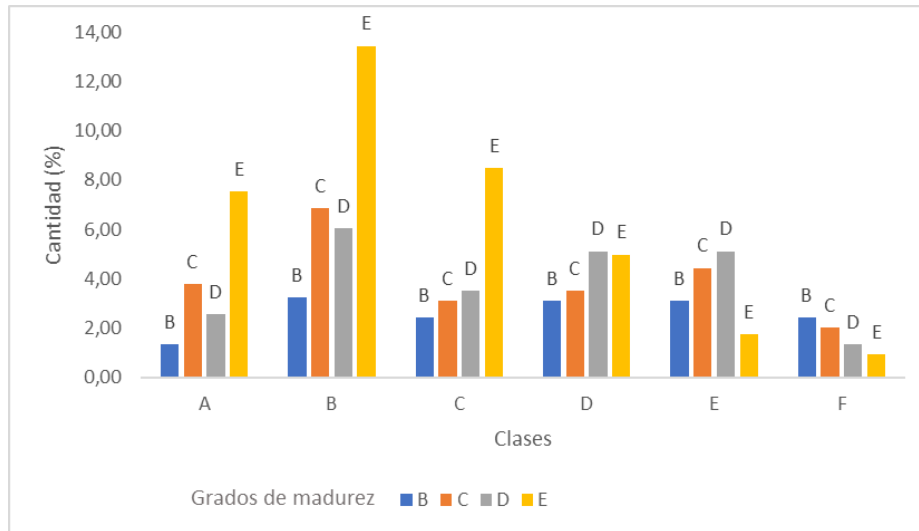


Figura 23. Distribución porcentual de muestras según grado de madurez para las categorías de la firmeza para piñas.

4.3.1.4. Contenido de humedad.

Los resultados de la matriz de confusión para el porcentaje máximo entre los modelos generales para el CH se muestran en la figura 24.a, donde la primera fila para la clase A a partir de arriba se encuentra el mayor entrecruzamiento de observaciones entre clases reales y clases predichas, que representa un 58,1% de observaciones correctamente clasificadas por clase verdadera (TPR). Seguidos a la clase A en orden descendente de mejor clasificación se encuentran las clases D (32,6%), C (26,5%) y B (11,4%).

Los PPV (figura 24.b) por el modelo poseen un porcentaje de acierto en orden descendente entre las clases del 37,2% para D, 36,4% para C, 31,1% de A y 22,2% para B, lo que involucra que la capacidad de predicción del modelo no concuerda con la capacidad de clasificación que se muestra en la figura 24.a ya que los valores máximos para cada grupo están ubicados en clases distintas es así que la diferencia entre ambos grupos es de 27% para la clase A a favor para la capacidad de clasificación y 10,8% para B, 9,9% para C y 4,6% para D a favor del potencial predictor.

Clase Verdadera	A	58,1%	12,4%	16,7%	12,9%
	B	50,9%	11,4%	17,1%	20,6%
	C	38,3%	13,3%	26,5%	21,9%
	D	40,1%	11,2%	16,0%	32,6%
		A	B	C	D

Clase predicha

(a)

Clase Verdadera	A	31,1%	25,6%	21,7%	14,6%
	B	25,6%	22,2%	21,0%	22,0%
	C	21,6%	28,9%	36,4%	26,2%
	D	21,6%	23,3%	21,0%	37,2%
		A	B	C	D

Clase predicha

(b)

Figura 24. Matriz de confusión para el modelo de mayor eficacia con el parámetro de parámetro de contenido de humedad con una eficacia del 32,4% y empleando frecuencia como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.

Al observar la figura 25 relacionada a la distribución de muestras para el CH se encuentra que hubo una tendencia a clasificar hacia la clase A. Al tomar en consideración la información de la figura 25 se observa que esta clase no posee el mayor número de muestras por GM, pero si la mayor cantidad de GM C y D juntos, lo que podría generar una tendencia a esta clase.

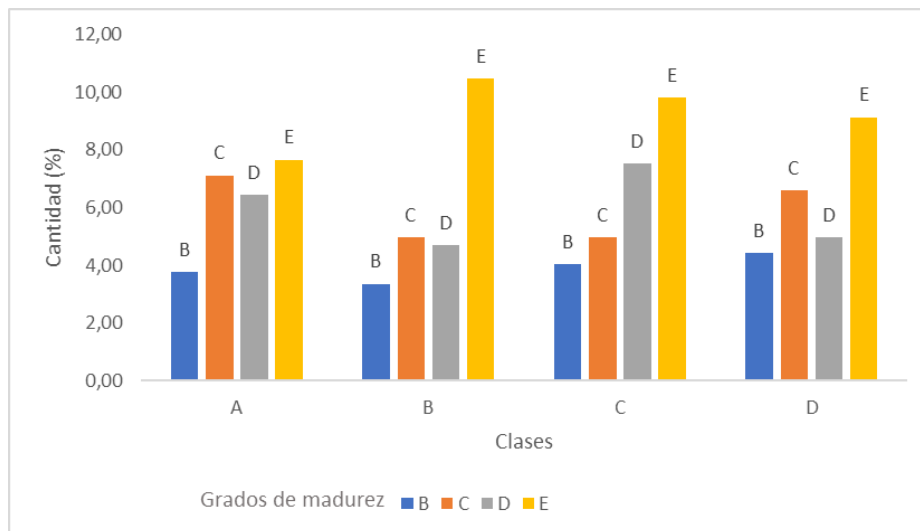


Figura 25. Distribución porcentual de muestras según grado de madurez para las categorías del contenido de humedad para piñas.

4.3.2. Gravedad específica.

Al analizar la GE para la máxima eficacia presentada en el cuadro 10 se obtiene la matriz de confusión de la figura 26.a. El valor más alto de observaciones correctamente clasificadas se ubica en la primera fila perteneciente a la clase A con un 42,3% de observaciones correctamente clasificadas por clase verdadera (TPR) seguido a la clase A en orden descendente de mejor clasificación se encuentran las clases D (36,7%), C (31,9%), B (16,4%), E (6,5%) y F (0%).

Por su parte los PPV poseen un porcentaje de eficacia en orden descendente entre las clases del 43,8% para B, 29,5% para A, 24% de C, 23% para D, 13,6% para E y 0% con F. La clase B presenta la capacidad máxima de predicción y discrepa de la categoría mejor valorada como clasificador con una diferencia entre ambos grupos de 12,8% para la clase A, 7,9% para C y 13,7% para D, mayores para la capacidad de clasificación y un porcentaje de 27,4% para B y 7,1% para E mayor para el potencial predictor.

Este modelo ofrece una mayor dispersión en cuanto a los TPR de las clases lo que puede ser de utilidad para emplear el modelo como un método para diferenciar entre grupos como el caso de A contra B, C, D o E, principalmente con esta última clase ya que se diferencian en un valor cercano al 40%. En cuanto a la capacidad de predicción ofrece la misma utilidad de separación para la clase B contra las A, C, D y E ya que solo se acercan en la mayoría de los casos en un 15%.

Clase Verdadera	A	42,3%	6,0%	23,8%	20,8%	7,1%	
	B	30,5%	16,4%	23,4%	22,7%	7,0%	
	C	29,2%	6,2%	31,9%	27,8%	4,9%	
	D	24,2%	3,9%	24,2%	36,7%	10,9%	
	E	37,1%	1,6%	26,6%	28,2%	6,5%	
	F	23,1%	1,9%	23,1%	34,6%	17,3%	
		A	B	C	D	E	F

Clase predicha

(a)

Clase Verdadera	A	29,5%	20,8%	20,8%	17,2%	20,3%	
	B	16,2%	43,8%	15,6%	14,2%	15,3%	
	C	17,4%	18,8%	24,0%	19,6%	11,9%	
	D	12,9%	10,4%	16,1%	23,0%	23,7%	
	E	19,1%	4,2%	17,2%	17,2%	13,6%	
	F	5,0%	2,1%	6,2%	8,8%	15,3%	
		A	B	C	D	E	F

Clase predicha

(b)

Figura 26. Matriz de confusión para el modelo de mayor eficacia con el parámetro de gravedad específica con una eficacia del 25,9% y empleando frecuencia como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.

Los resultados de la figura 26.a pueden ligarse a la distribución mostrada en la figura 27 ya que esta primera clase con el mayor número de muestras correctamente clasificadas posee la mayor cantidad de muestras por GM en su clase y en el caso del grupo F esta se vio afectada por la cantidad de los otros, sumado a que su clase es la de menor cantidad de muestras por GM.

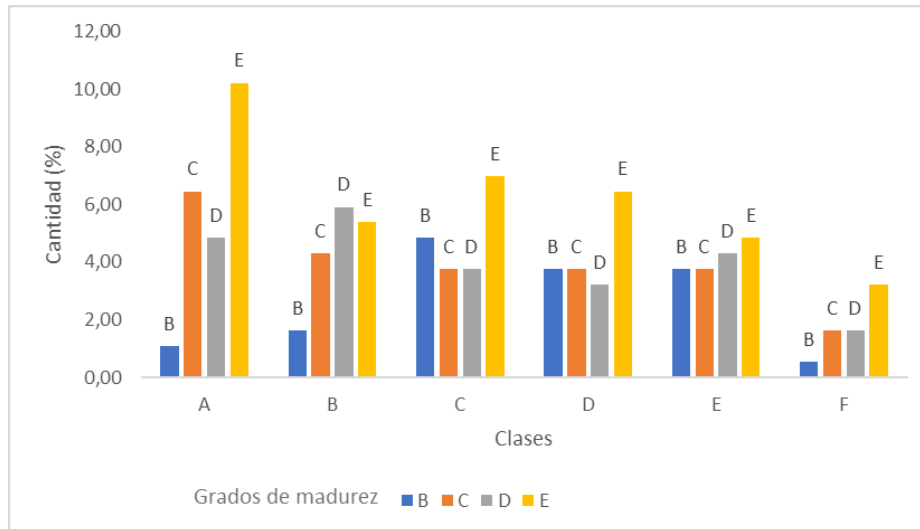


Figura 27. Distribución de muestras según grado de madurez para las categorías de la gravedad específica para piñas.

4.4. Modelos de redes neuronales artificiales con mayor eficacia para cada parámetro de calidad empleando frecuencias de resonancia.

Los modelos que presentan especial atención son los basados en redes neuronales ya que se adecuan a la cantidad de muestras, se obtuvo un porcentaje más alto para ESR (0,4% menor al modelo Árbol (árbol grueso) correspondiente a la prueba destructiva y en sucesión está el GM (1% menor a Bayas ingenuas (bayas ingenuas gaussianas) sujeto a una prueba no destructiva y en mínima instancia está la masa con 0,8% menor a Árbol (Árbol fino). Además, la tendencia de las frecuencias de resonancia con respecto al tipo de prueba destructiva/no destructiva presenta que f_1 (50 – 150 Hz) tanto para los modelos generales como para red neuronal está asociado únicamente al GM. La f_2 (100 – 350 Hz) por su parte se asocia mejor con parámetros destructivo y no destructivos. Y por último f_3 (200 – 500 Hz) se relaciona con la GM y f_4 (500 – 900 Hz) con características de firmeza o estructurales de la piña. Entonces el porcentaje de uso de las frecuencias es de f_1 con 14,29 %, f_2 con 42,86 %, f_3 con 14,29% y f_4 con 28,57%.

Cuadro 11. Resultados de la eficacia general por cada parámetro de calidad obtenidos según el tipo de modelo de red neuronal desarrollado.

Modelo red neuronal	Eficacia general (%)	Frecuencia resonancia	Parámetro
Estrecha	39,5	2	Esfuerzo remanente
Estrecha	37,6	1	Grado de madurez
Estrecha	31,7	2	Contenido de humedad
Estrecha	28,9	4	Firmeza
Tres capas	25	3	Gravedad específica
Estrecha	22,2	4	Fuerza de penetración
Ancha	14,1	2	Masa

4.4.1. Resultados para los parámetros de calidad obtenidos mediante redes neuronales.

4.4.1.1. Esfuerzo remanente.

Al examinar los resultados del mayor porcentaje de eficacia para un modelo de red neuronal (figura 28.a) correspondiente a la ESR se obtiene en la segunda fila en la parte superior el mejor porcentaje de muestras correctamente clasificadas para la clase B que involucra un 96,9% de observaciones correctamente clasificadas por clase verdadera (TPR). Las clasificaciones de las clases A, C y D presentan falsos negativos de 97,3%, 97,5% y 100% respectivamente, por tanto, para clasificar la clase B es óptimo, pero para A, C y D no resulta un modelo óptimo.

Los PPV presentan valores cercanos en las clases A de 41,7%, B de 39,5%, C de 40% y D con 0%. Este último valor de la clase a causa de que ninguna muestra fue predicha para esta clase, pero sí incorrectamente para la categoría A como se observa en la fila A, columna D. Estos valores predictivos positivos son superiores en las clases A en un 29,2%, B en 0,9% y C en 40% pero inferior en la D con un 27,3% con respecto al modelo de mayor eficiencia para estricción (cuadro 11). La red neuronal estrecha tiene mejor capacidad de predicción en

relación con árbol grueso, pero posee una generalización incompleta en cuanto a la cantidad de clases y un porcentaje inferior al 70% para predecir grados estructurales de la piña en términos de esfuerzo remanente. En este modelo la mayor parte de las observaciones las clasifica como clase B, mientras que su capacidad de predicción mejora, pero se encuentra sujeta al entrenamiento que generó el modelo.

Clase Verdadera	A	2,7%	96,2%	0,5%	0,5%
	B	1,4%	96,9%	1,7%	
	C	1,9%	95,6%	2,5%	
	D		100,0%		
		A	B	C	D
		Clase predicha			

(a)

Clase Verdadera	A	41,7%	24,3%	10,0%	100,0%
	B	33,3%	39,5%	50,0%	
	C	25,0%	20,9%	40,0%	
	D		15,3%		
		A	B	C	D
		Clase predicha			

(b)

Figura 28. Matriz de confusión para la red neuronal de mayor eficacia con el parámetro de esfuerzo remanente con una eficacia del 39,5% y empleando frecuencia como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.

4.4.1.2. Grado de madurez.

Los resultados obtenidos en figura 29.a correspondiente al GM con modelos basados en redes neuronales muestra en la última fila el mayor número de muestras correctamente clasificadas para la clase E que representa un 82,2% de observaciones correctamente clasificadas por clase verdadera (TPR). Este valor está por encima de todas las observaciones correctamente clasificadas, con una diferencia con la siguiente de mayor valor con un 58,1%.

Al analizar los PPV del modelo estos poseen una proporción porcentual en orden descendente entre las clases de 39,7% para E, 34,7% para C, 24,2% de D y 13,3% para B. Al comparar los valores predichos por el modelo de mayor eficacia (figura 21.b) se determina que la presente red neuronal es superior para la clase E en un 3,3%, 13,3% para B y 24,2% para D y únicamente para la clase C es inferior en un 1,7%. Al considerar la capacidad de

generalización del modelo de red neuronal artificial este resulta más efectivo al trabajar con varias clases en cuanto al parámetro de GM, pero los porcentajes de predicción no logran superar el 70%. Este modelo por lo descrito posee una orientación a la clase E, lo cual ofrece una capacidad de pertenecer a esta en un 66% como mínimo considerado por clase.

Clase Verdadera	B	1,7%	14,7%	0,9%	82,8%
	C	2,8%	24,4%	6,2%	66,5%
	D	2,3%	18,2%	4,5%	75,0%
	E	1,4%	11,6%	4,7%	82,5%
		B	C	D	E

Clase predicha

(a)

Clase Verdadera	B	13,3%	13,7%	3,0%	16,8%
	C	33,3%	34,7%	33,3%	20,8%
	D	26,7%	25,8%	24,2%	23,1%
	E	26,7%	25,8%	39,4%	39,7%
		B	C	D	E

Clase predicha

(b)

Figura 29. Matriz de confusión para la red neuronal de mayor eficacia con el parámetro de grado de madurez con una eficacia del 37,6% y empleando frecuencia como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.

4.4.1.3. Contenido de humedad.

Al analizar el parámetro de CH (figura 30.a) se obtiene que en la primera fila para la clase A se da la mayor cantidad de muestras correctamente clasificadas lo que involucra un 48,4% de observaciones correctamente clasificadas por clase verdadera (TPR).

Los PPV presentan valores cercanos en la clase D de 34,1%, C con 33%, A con 29,7% y B con 29,3% que en comparación con los valores predichos del modelo de la Figura 21.b se determina que esta red neuronal posee una eficacia menor a las clases A en un 1,4%, C en 3,4%, y en 3,1% para D, pero en la clase B es 7,1% mejor. El modelo Conjunto (árboles RUSBoosted) tiene mejor capacidad de predicción a pesar de poseer porcentajes de rendimiento menores al 40%.

Clase Verdadera	A	48,4%	6,5%	26,9%	18,3%
	B	40,6%	6,9%	30,9%	21,7%
	C	37,8%	3,6%	39,3%	19,4%
	D	36,4%	5,3%	27,8%	30,5%
		A	B	C	D

Clase predicha

(a)

Clase Verdadera	A	29,7%	29,3%	21,5%	20,4%
	B	23,4%	29,3%	23,2%	22,8%
	C	24,4%	17,1%	33,0%	22,8%
	D	22,4%	24,4%	22,3%	34,1%
		A	B	C	D

Clase predicha

(b)

Figura 30. Matriz de confusión para la red neuronal de mayor eficacia con el parámetro de contenido de humedad con una eficacia del 31,7% y empleando frecuencia como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.

4.4.1.4. Firmeza.

Las observaciones correctamente clasificadas mostradas en la figura 31.a indican una tendencia marcada hacia la clase B mostrando en ella el mayor porcentaje positivo con un 86,4% seguido de la clase C con 11,5%. Es una separación cercana a un 75% entre la primera y la segunda posición de clasificaciones de mejor posición. Se determina además que las redes neuronales generaron una mayor generalización con más clases en comparación a las mostradas en la figura 22, pero los porcentajes correctamente clasificados no superan el 5,3%.

En cuanto a los resultados por parte de las observaciones correctamente predichas se encontró un mayor porcentaje para la clase B con 31,8% e igualmente la clase C se mantiene en segundo lugar con 19%. Tanto para la figura 29.a como b no se logró generar resultados correctos para la clase F.

Clase Verdadera	A	5,3%	79,6%	8,8%	4,4%	1,8%	
	B	3,6%	86,4%	7,7%	1,4%	0,9%	
	C	2,3%	80,0%	11,5%	3,1%	1,5%	1,5%
	D	5,6%	78,2%	11,3%	1,6%	3,2%	
	E	5,6%	72,9%	14,0%	5,6%	1,9%	
	F	4,0%	78,0%	16,0%		2,0%	
		A	B	C	D	E	F

Clase predicha

(a)

Clase Verdadera	A	18,8%	15,1%	12,7%	25,0%	15,4%	
	B	25,0%	31,8%	21,5%	15,0%	15,4%	
	C	9,4%	17,4%	19,0%	20,0%	15,4%	100,0%
	D	21,9%	16,2%	17,7%	10,0%	30,8%	
	E	18,8%	13,0%	19,0%	30,0%	15,4%	
	F	6,2%	6,5%	10,1%		7,7%	
		A	B	C	D	E	F

Clase predicha

(b)

Figura 31. Matriz de confusión para la red neuronal de mayor eficacia con el parámetro de firmeza con una eficacia del 28,9% y empleando frecuencia como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.

4.4.1.5. Gravedad específica.

La figura 32.a muestra la matriz de confusión correspondiente a los resultados del máximo porcentaje de eficacia correspondiente al 25% para la red neuronal de la GE. La fila A posee el mejor número de muestras correctamente clasificadas con 50% de observaciones correctamente clasificadas por clase verdadera (TPR) seguidos por C con 33,3% observaciones, D con 19,5%, B con 18,8, E con 4% y finalmente F con 0%.

Los PPV se encuentra con una proporción porcentual del 38,1% para la clase B, 27,8% para A, 23,6% para C, 18,5% con D, 12,8% con E y 0% para F que en comparación a los valores predichos por el modelo de mayor eficiencia (figura 26.b) se obtiene que la presente red neuronal es inferior para la clase A en un 1,7%, 5,7% para B, 0,4% para C, 4,5% para D, 0,8% para E y 0% para F. Este último valor puede ocasionarse debido a que todas las observaciones fueron clasificadas incorrectamente en las otras cinco clases, únicamente para las categorías A y D se realizaron clasificaciones para F. Esto además muestra una efectividad

comparable entre modelos, donde para la clase B y D representa una opción particular en la que resulta mejor, y por ende posea mayor utilidad para predecir estas clases. En términos de clasificación este modelo presenta un 50% de eficacia producto de la tendencia a favorecer la clase A, mostrando una utilidad aceptable que puede verse mejorado en una línea de producción al analizar en más de una ocasión el lote de producto generado para esta clase ya que podría esperarse una mejor selección de producto que también se encuentra favorecido por un entrenamiento constante de este modelo durante su uso en campo.

Clase Verdadera	A	50,0%	7,7%	22,0%	14,9%	4,8%	0,6%
	B	39,1%	18,8%	23,4%	12,5%	6,2%	
	C	35,4%	6,9%	33,3%	20,1%	4,2%	
	D	39,8%	7,0%	26,6%	19,5%	6,2%	0,8%
	E	37,1%	4,8%	31,5%	22,6%	4,0%	
	F	38,5%	1,9%	28,8%	23,1%	7,7%	
		A	B	C	D	E	F

Clase predicha

(a)

Clase Verdadera	A	27,8%	20,6%	18,2%	18,5%	20,5%	50,0%
	B	16,6%	38,1%	14,8%	11,9%	20,5%	
	C	16,9%	15,9%	23,6%	21,5%	15,4%	
	D	16,9%	14,3%	16,7%	18,5%	20,5%	50,0%
	E	15,2%	9,5%	19,2%	20,7%	12,8%	
	F	6,6%	1,6%	7,4%	8,9%	10,3%	
		A	B	C	D	E	F

Clase predicha

(b)

Figura 32. Matriz de confusión para la red neuronal de mayor eficacia con el parámetro de gravedad específica con una eficacia del 25% y empleando frecuencia como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.

4.5. Modelos predictivos con mayor eficacia para cada parámetro de calidad empleando coeficientes de rigidez.

Los parámetros de GM, ESR y el CH poseen un comportamiento semejante en las frecuencias, al mantenerse en las primeras posiciones estos modelos predictivos de mayor eficacia. Al analizar estos resultados de eficacia con respecto a los mostrados en el cuadro 12 se encuentra que para el presente caso la eficacia es mayor para el GM en un 2,8% y la

ESR en 0,6%. Al contrario, la eficacia resulta menor para el CH en 2,1%, la FM en 0,5% y la P en 0,4%.

Los CR no presenta ninguna tendencia evidente como resulta con las frecuencias. El CR1 obtuvo los mejores resultados con firmeza y CH, CR2 con firmeza y CR3 con parámetros de pruebas no destructivas (GM) como destructivas (FM). Estos resultados por parte de CR2 se asemejan a los dados por Pathaveerat, Terdwongworakul, & Phaungsombut, 2008b ya que esta ecuación brindó información para la diferenciación del estadio de madurez de la piña, para una de tres categorías de piñas se presentó una correlación cercana a $R=0,5$ con la firmeza.

Un comportamiento que se evidencia de los CR es la capacidad de determinar un parámetro distinto de la firmeza o estructural de las muestras, asociándose CR1 con el esfuerzo remanente, CR2 con la P y CR3 con la FM. Esta característica puede ser aprovechada para realizar análisis específicos de acuerdo con las necesidades del estudio de calidad. Entonces la proporción de utilidad de los CR es de CR1 con 40 %, CR2 con 20% y CR3 con 40%.

Cuadro 12. Resultados de la eficacia general por cada parámetro de calidad obtenidos según el tipo de modelo predictivo desarrollado y entrenados con coeficientes de rigidez.

Modelo	Eficacia general (%)	Coefficiente de rigidez	Parámetro
K vecino más cercano (K vecino más cercano ponderado)	41,4	CR3	Grado de madurez
Árbol (Árbol grueso)	40,5	CR1	Esfuerzo remanente
Conjunto (Árboles agrupados)	30,5	CR1	Contenido de humedad
Máquinas de vectores de soporte (Máquinas de vectores de soporte gaussiana fina)	29,7	CR3	Firmeza
Árbol (Árbol grueso)	21,8	CR2	Fuerza de penetración

4.5.1. Resultados para los parámetros de calidad.

4.5.1.1. Grado de madurez con coeficiente de rigidez 3.

A partir de la matriz de confusión de la figura 33.a perteneciente a la asociación del grado de madurez con CR3 que fue el emparejamiento con mayor resultado de los tres CR, se encuentra que la última fila perteneciente a la clase E dió el mayor número de muestras clasificadas correctamente con un 52,5% de TPR. La clase B obtuvo el menor porcentaje de muestras clasificadas correctamente con un 30,2% de observaciones. Existe una tendencia por clasificar las muestras en la clase E lo cual aumenta las observaciones negativas para estos grupos, pero favorece la clase.

Los PPV para el modelo indican igualmente una mejor capacidad de predicción para la clase E con una eficacia del 50,7% seguidos por la clase D (38%), C (36,1%) y B (31,2%). Este porcentaje mayor se complementa con la cantidad de muestras existentes por clase descrito en el cuadro 4, donde el GM de E es el de mayor cantidad seguidos por C ó D y B. El mayor porcentaje de observaciones incorrectamente predichas (FDR) se ajustan para la clase E a excepción de esta que es para B con un 18,5%. Esta posible tendencia orientada a la clase E puede deberse a que es la categoría de piñas con mayor número de muestras empleadas lo que podría desfavorecer en una medida lo logrado por el modelo.

Hay que considerar que este valor máximo dado fue con CR3 el cual involucra variables como frecuencia, masa y densidad y no con la versión abreviada que es el CR2. Situación que denota alguna necesidad del uso de la densidad como variable de asociación dentro del modelo para mejorar la capacidad de predicción del GM.

Al comparar los resultados de PPV con los de la figura 21.b pertenecientes al modelo Bayas ingenuas (bayas ingenuas gaussianas) con f_i como predictor, se determina que es superior en la clase B, D y E, en una eficacia del 31,2%, 38% y 11,8%, mientras que la clase B es inferior en un 0,3%. Por tanto, este modelo con CR3 es superior al que resulta usar únicamente f_i como variable predictora, además de ofrecer estimados en las clases B y D donde no son generados por parte del otro modelo.

Los resultados generados para clasificación y predicción se encuentran con valores de TPR y PPV mayores en comparación a los logrados negativamente a excepción de la clase B para

TPR, lo que permite aprovechar el modelo para diferenciar entre todas las clases, pero con mayor eficacia con la E. En un análisis con una mezcla de producto donde el primer resultado describe que el B1 tiene 31,2% de B, 18,8% de C, 16,1% de D y 33,9% de E. La clase E1 está compuesta por 50,7% de E, 18,5% de D, 17,8% de C y 12,9% de A. En el caso de ingresar de nuevo B1 y E1 por el analizador de la máquina se esperaría que las cantidades correctamente clasificadas para ambas clases sean mayores al restringirse los datos a los rangos que el modelo reconoce para cada clasificación. De tal manera que las piñas que antes estaban contenidas tanto en B1 y E1 pertenecientes a otras clases, se clasifiquen en sus respectivos grupos, pero al mismo tiempo algunas muestras podrían clasificarse incorrectamente, pero en menor cantidad ya que se habría reajustado la primera clasificación aplicada.

Clase Verdadera	B	30,2%	22,4%	15,5%	31,9%
	C	11,9%	36,9%	22,2%	29,0%
	D	10,2%	23,9%	35,8%	30,1%
	E	13,8%	17,0%	16,7%	52,5%
		B	C	D	E

Clase predicha

(a)

Clase Verdadera	B	31,2%	14,4%	10,8%	12,9%
	C	18,8%	36,1%	23,5%	17,8%
	D	16,1%	23,3%	38,0%	18,5%
	E	33,9%	26,1%	27,7%	50,7%
		B	C	D	E

Clase predicha

(b)

Figura 33. Matriz de confusión para el modelo de mayor eficacia con el parámetro de grado de madurez con una eficacia del 41,4% y empleando coeficiente de rigidez 3 como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.

4.5.1.2. Esfuerzo remanente con coeficiente de rigidez 1.

El emparejamiento del ESR con CR1 mostró un mayor número de observaciones clasificadas correctamente como positivas para la clase B con una diferencia del 47,6% entre la siguiente clase de mayor valor lo cual muestra una reducida capacidad generalizadora del modelo (figura 34.a). Esta clase B posee una proporción de muestras correctamente clasificadas por

clase verdadera (TPR) del 64,6%, con un porcentaje equiparable de falsos negativos reales para las otras clases, lo cual describe una tendencia a clasificar hacia esta clase sean estas pertenecientes o no.

La capacidad predictiva del modelo varía levemente con respecto a la capacidad clasificadora, pero se continua la permanencia de la clase B como la de mejor eficacia con un 40,3% seguidos por D (27,1%); A (25,2%) y C (21,6%). Además, se determina que la mayor cantidad de observaciones incorrectamente predichas (FDR) para las otras clases se concentran en esta clase. La capacidad de generalización es baja al no obtener porcentajes de eficacia entre las clases con valores cercanos y a favor de las observaciones positivas.

Al comparar los resultados de PPV con el modelo generado en la figura 19.b perteneciente a árbol grueso con f_4 como predictor, este resulta mejor en las clases A, B y D en una eficacia del 25,2%; 0,6% y 27,1% respectivamente, mientras que para la clase C el otro modelo es superior en un 35,5%. Hay una capacidad de predicción de este modelo con CR1 frente al otro elaborado, que incluso logra generar estimados en las clases A y D que en el otro modelo no fue posible realizar.

De acuerdo al histograma de la figura 20 y la matriz de la figura 34, se tiene que este valor alto de eficacia en B, se ajusta a la mayor cantidad de muestras agrupadas dentro de la clasificación previa al desarrollo de los modelos. Además, se tiene que en esta clase se ubica el porcentaje más elevado de piñas con GM de E y dentro de esta misma clase la B es la menor de las cuatro. De acuerdo con el histograma esta clasificación encontrada se ajusta a la tendencia de esta.

Clase Verdadera	A	17,0%	64,8%	14,3%	3,8%
	B	16,3%	64,6%	13,3%	5,8%
	C	19,0%	60,1%	13,9%	7,0%
	D	12,7%	61,8%	13,6%	11,8%
		A	B	C	D

Clase predicha

(a)

Clase Verdadera	A	25,2%	25,1%	25,5%	14,6%
	B	39,0%	40,3%	38,2%	35,4%
	C	24,4%	20,2%	21,6%	22,9%
	D	11,7%	14,4%	14,7%	27,1%
		A	B	C	D

Clase predicha

(b)

Figura 34. Matriz de confusión para el modelo de mayor eficacia con el parámetro de esfuerzo remanente una eficacia del 40,5% y empleando coeficiente de rigidez 1 como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.

4.5.1.3. Contenido de humedad con coeficiente rigidez 1.

El parámetro de CH (figura 35.a) presenta una capacidad de generalización con mayor distribución al encontrarse en tres clases una cantidad de observaciones correctamente clasificadas como verdaderas con valores cercanos entre 20% y 34,9%, siendo el mayor de estas para la clase A la cual representa un 34,9% de TPR seguidas de C (33,7%); D (32,6%) y B (20%).

Los PPV también muestran una generalización equiparable a su capacidad de clasificación, siendo igualmente mayor para A (33,3%) seguidos de D (34,3%), C (33%) y B (20,5%), que al compararse con los PPV de la figura 24.b del modelo Conjunto (árboles RUSBoosted) con f_2 como predictor, se tiene una misma tendencia en el orden de eficacia de las clases. La clase A es la única con mayor porcentaje con respecto al otro modelo, con un 2,2% en ventaja.

Clase Verdadera	A	34,9%	25,3%	22,0%	17,7%
	B	29,1%	20,0%	27,4%	23,4%
	C	20,4%	24,0%	33,7%	21,9%
	D	20,9%	22,5%	24,1%	32,6%
		A	B	C	D

Clase predicha

(a)

Clase Verdadera	A	33,3%	27,5%	20,5%	18,5%
	B	26,2%	20,5%	24,0%	23,0%
	C	20,5%	27,5%	33,0%	24,2%
	D	20,0%	24,6%	22,5%	34,3%
		A	B	C	D

Clase predicha

(b)

Figura 35. Matriz de confusión para el modelo de mayor eficacia con el parámetro de contenido de humedad con una eficacia del 30,5% y empleando coeficiente de rigidez 1 como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.

Al considerar la figura 25, el histograma posee una tendencia que se diferencia con lo mostrado en la clasificación de la matriz de confusión ya que el orden por cantidad de muestras dado en el histograma es de C, D, A y B. Esta clase A que presenta mayor eficacia posee la distribución más equitativa entre GM de acuerdo con el histograma a diferencia de las otras tres clases, lo cual podría influenciar la mejora de la clasificación, pero no representa una causa probable ya que, según la matriz esta posee porcentajes de eficacia cercanos en las clases C y D.

4.5.1.4. Firmeza con coeficiente rigidez 3.

Los resultados generados con el coeficiente 3 para FM mostrados en la figura 36.a muestran una mayor capacidad de generalización que el obtenido en la figura 22.a ya que genero resultados positivos para la clase D y E, pero no en C como si fue el caso ocurrido con el predictore de frecuencias. La mayor cantidad de observaciones correctamente clasificadas se encuentra en la clase B con 99,1%, además se muestra una tendencia marcada a clasificar hacia esta clase. Esta tendencia como ocurrió con el bloque de frecuencias está marcada por una mayor cantidad de muestras hacia la clase B, pero también influenciada por la asociación generada por los algoritmos de inteligencia artificial.

Por su parte las observaciones correctamente predichas poseen el mayor porcentaje en la clase D con un 40% seguido por la clase B, lo cual posiciona esta clase D con el mayor porcentaje de observaciones de falsos positivos.

Clase Verdadera	A	99,1%	0,9%				
	B	99,1%			0,9%		
	C	0,8%	96,9%		0,8%	1,5%	
	D		95,2%		1,6%	3,2%	
	E		97,2%		1,9%	0,9%	
	F		96,0%			4,0%	
		A	B	C	D	E	F

Clase predicha

(a)

Clase Verdadera	A	15,4%	100,0%				
	B	30,0%			18,2%		
	C	100,0%	17,4%		20,0%	18,2%	
	D		16,3%		40,0%	36,4%	
	E		14,3%		40,0%	9,1%	
	F		6,6%			18,2%	
		A	B	C	D	E	F

Clase predicha

(b)

Figura 36. Matriz de confusión para el modelo de mayor eficacia con el parámetro de firmeza con una eficacia del 29,7% y empleando coeficiente de rigidez 3 como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.

4.6. Modelos de red neuronal con mayor eficacia para cada parámetro de calidad empleando coeficientes de rigidez.

Los modelos de redes neuronales generados a partir de los CR mostrados en el cuadro 13 poseen una eficacia menor excepto la FM que es 0,4% mayor a los descritos en el cuadro 11. El parámetro de la ESR se ubica en la primera posición seguido por el GM con excepción del CH que se desplaza una posición por debajo de la FM. En cuanto a la tendencia de los CR no parece mostrar un claro comportamiento, pero se evidencia la capacidad del CR1 para poder asociarse de manera más eficaz que CR2 y CR3 a cuatro de los cinco parámetros de

calidad analizados en esta prueba, los cuales contemplan los dos principales puntos de estudio como la firmeza o estructura y CH. Estos resultados muestran que el CR1 puede ser utilizado para obtener las mejores eficacias con cuatro de cinco parámetros y CR3 con únicamente uno, por ello el CR2 que es una variante del CR1 no se adapta con mejor eficacia a los datos de los parámetros. Esto puede deberse a la alteración del valor de la masa producto del exponente asociado en la ecuación 10 de CR2.

Cuadro 13. Resultados de la eficacia general por cada parámetro de calidad obtenidos según el tipo de red neuronal desarrollado y entrenados con coeficientes de rigidez.

Modelo red neuronal	Eficacia general (%)	Coefficiente de rigidez	Parámetro
Estrecha	39,2	CR3	Esfuerzo remanente
Estrecha	37,2	CR1	Grado de madurez
Estrecha	29,3	CR1	Firmeza
Ancha	29,2	CR1	Contenido de humedad
Ancha	20,3	CR1	Fuerza de penetración

4.6.1. Resultados para los parámetros de calidad obtenidos mediante redes neuronales.

4.6.1.1. Esfuerzo remanente con coeficiente de rigidez 3.

A partir de la figura 37.a que muestra la matriz de confusión se encuentra que el mayor número de observaciones correctamente clasificadas se dan en la clase B con un 88,8% de TPR, seguidos por las clases A (9,3%); C (7,6%) y D (1,8%). Además, se evidencia un exceso de muestras clasificadas como B debido a que en las clases A, C y D se dan valores de muestras clasificadas incorrectamente por clase verdadera como clase B. Este resultado puede que haya sido condicionado por un mayor número de piñas ubicadas en esta clase en comparación a las otras tres categorías.

Los PPV poseen una distribución porcentual entre 25% y 40,5% que son valores más cercanos a los obtenidos en la figura 28.a, siendo el máximo número de observaciones

correctamente predichas el que se obtuvo para la clase B, seguidos por A, C y D. Además, estas últimas tres clases generaron el mayor número de observaciones incorrectamente predichas para la clase B, pero contraria a la capacidad del modelo para clasificación estos porcentajes son significativamente menores.

Al comparar estos resultados con los dados por la figura 28.b de la matriz de confusión del modelo de red neuronal con f_2 como predictor se encuentra que las clases A y C son mayores a la presente red neuronal en un 9% y 9,2% respectivamente. Por su parte las clases B y D son mayores para este modelo en un 1% y 25% respectivamente. Siendo esta última la clase que no reportó resultados predictivos por parte del modelo con frecuencias, por ello no resulta mejor generalizando con lo cual la presente red neuronal es la mejor opción para trabajar con varias clases.

Como sucedió en anteriores clasificaciones para el ESR, el mayor porcentaje de eficacia ocurre en B, donde el histograma de la figura 20 presenta el mayor número de muestras, en particular en esta se ubican el mayor porcentaje con GM de E, es decir, las de más madurez. Esta condición puede ser una causa que influyera el desarrollo del modelo con un notable porcentaje positivo para B contra las demás clases.

Clase Verdadera	A	9,3%	85,7%	4,9%	
	B	5,8%	88,8%	4,4%	1,0%
	C	5,1%	85,4%	7,6%	1,9%
	D	9,1%	84,5%	4,5%	1,8%
		A	B	C	D

(a)

Clase Verdadera	A	32,7%	24,2%	23,1%	
	B	32,7%	40,5%	33,3%	37,5%
	C	15,4%	20,9%	30,8%	37,5%
	D	19,2%	14,4%	12,8%	25,0%
		A	B	C	D

(b)

Figura 37. Matriz de confusión para el modelo de red neuronal de mayor eficacia con el parámetro de esfuerzo remanente con una eficacia del 39,5% y empleando coeficiente de rigidez 3 como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.

4.6.1.2. Grado de madurez con coeficiente de rigidez 1.

Los resultados para la matriz de confusión del GM muestran que la clase E alcanzó el mayor número de observaciones correctamente clasificadas que corresponde a un TPR del 91,3%, seguidos por C y D, mientras que la clase B no presentó datos clasificados. Se presenta además una tendencia a clasificar las muestras en la clase E sean estas pertenecientes o no a este grupo, un resultado que puede estar influenciado por una cantidad de piñas de esta categoría que sobrepasa las otras tres (cuadro 4).

Al analizar los PPV del modelo (figura 38.b) se obtiene que poseen un rango porcentual entre el 38,3% al 23,3%, donde este valor máximo de PPV corresponde a la clase E, seguidos por C (33,3%) y D (23,3%). Esto se ajusta a la tendencia esperada ya que esta clase E es la que presenta mayor porcentaje de muestras contra el resto de GM. Por su parte la clase A no presentó ningún resultado correctamente predicho dirigiendo las muestras de manera incorrecta a las clases C y D. Comparando estos resultados dados para el GM con los obtenidos en la figura 24.b se tiene que este último es mejor en las clases B en un 13,3%, C en 1,4%, D en 0,9% y E en 1,4%. Por tanto, el modelo de red neuronal con frecuencia como predictor es mejor para predecir el GM en piñas al contrario que lo realiza empleado CR.

Clase Verdadera	B		5,2%	2,6%	92,2%
	C	0,6%	10,2%	5,7%	83,5%
	D	0,6%	9,1%	4,0%	86,4%
	E		5,1%	3,6%	91,3%
		B	C	D	E
Clase predicha					

(a)

Clase Verdadera	B		11,1%	10,0%	16,3%
	C	50,0%	33,3%	33,3%	22,3%
	D	50,0%	29,6%	23,3%	23,1%
	E		25,9%	33,3%	38,3%
		B	C	D	E
Clase predicha					

(b)

Figura 38. Matriz de confusión para el modelo de red neuronal de mayor eficacia con el parámetro de grado de madurez con una eficacia del 37,2% y empleando coeficiente de rigidez 1 como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.

4.6.1.3. Firmeza con coeficiente de rigidez 1.

La figura 39.a muestra en la clase B la mayor cantidad de muestras correctamente clasificadas por clase verdadera con un 92,7% seguido por la clase C con 5,4%. Estos resultados generados tuvieron menor capacidad de generalización en comparación a los obtenidos en las redes neuronales para el bloque de frecuencias para firmeza ya que no se generó en el presente caso resultados para la clase A. Además, como en los casos anteriores de estudio para firmeza esta tiene una tendencia marcada a clasificar hacia la clase B lo cual puede deberse a las mismas particularidades propuestas de cantidad de muestras y comportamiento de asociación creado por los algoritmos.

Por parte de los resultados obtenidos de observaciones correctamente predichas el mayor porcentaje se da para la clase C seguida de la clase B. Se encuentra además cercanía entre los porcentajes positivos de las observaciones en un rango de 20 a 31,8%.

Clase Verdadera	A	95,6%	2,7%	1,8%			
	B	92,7%	3,2%	3,6%	0,5%		
	C	90,8%	5,4%	3,8%			
	D	90,3%	3,2%	4,8%	1,6%		
	E	95,3%	0,9%	2,8%	0,9%		
	F	98,0%			2,0%		
		A	B	C	D	E	F

Clase predicha

(a)

Clase Verdadera	A	15,6%	13,6%	8,3%			
	B	29,4%	31,8%	33,3%	20,0%		
	C	17,0%	31,8%	20,8%			
	D	16,2%	18,2%	25,0%	40,0%		
	E	14,7%	4,5%	12,5%	20,0%		
	F	7,1%			20,0%		
		A	B	C	D	E	F

Clase predicha

(b)

Figura 39. Matriz de confusión para el modelo de red neuronal de mayor eficacia con el parámetro de firmeza con una eficacia del 29,3% y empleando coeficiente de rigidez 1 como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.

4.6.1.4. Contenido de humedad con coeficiente de rigidez 1.

El mayor número de observaciones correctamente clasificadas se presenta en la clase A con un 37,6% de observaciones correctamente clasificadas como clase verdadera, seguidos por C (34,7%); D (31,6%) y B (11,4%). Los resultados de PPV muestran el mismo orden para las clases en cuanto a porcentaje de observaciones correctamente predichas como clase verdadera, con valores que van de los 34,5% a 19,4%.

Al comparar la capacidad predictiva de los resultados del presente modelo de red neuronal con el generado en la figura 30.b se determina que el modelo con frecuencia como predictor es mejor con la clase B en 9,9%, la clase C en 1,5% y D en 7,5%. En cambio, es superior el modelo con CR para la clase A en un 4,8%. Ambas redes neuronales poseen capacidad de generalización al trabajar con las cuatro clases, pero la basada con frecuencias de resonancia posee mejores porcentajes de predicción.

Este caso resulta de interés ya que los porcentajes de observaciones positivas y negativas se encuentran distribuidos con cercanía entre sí tanto para la capacidad de clasificación como predicción. Este comportamiento de los resultados resulta desfavorable si se conserva las clasificaciones dadas por este en una primera ronda de análisis si se emplea este código en una máquina de planta. Esto se debe a que el porcentaje para clasificar tanto correctamente como incorrectamente una clase es similar y si se consideran las FNR y FDR resulta superior a los TPR y PPV para las mezclas de producto, por lo tanto, la aplicación de nuevas revisiones en los conjuntos clasificados podría mejorar la exactitud. La tendencia de los resultados contra el histograma para el porcentaje de muestras según GM para el contenido de humedad posee un comportamiento semejante al que ocurrió en la figura 25, por ende, su análisis aplica para el presente caso.

Clase Verdadera	A	37,6%	10,2%	27,4%	24,7%
	B	25,7%	11,4%	29,7%	33,1%
	C	21,9%	13,3%	34,7%	30,1%
	D	24,1%	20,3%	24,1%	31,6%
		A	B	C	D

Clase predicha

(a)

Clase Verdadera	A	34,50%	18,40%	23,60%	20,70%
	B	22,20%	19,40%	24,10%	26,10%
	C	21,20%	25,20%	31,50%	26,60%
	D	22,20%	36,90%	21%	26,60%
		A	B	C	D

Clase predicha

(b)

Figura 40. Matriz de confusión para el modelo de red neuronal de mayor eficacia con el parámetro de contenido de humedad con una eficacia del 29,2% y empleando coeficiente de rigidez 1 como predictor. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.

4.7. Modelos predictivos con mayor eficacia para cada parámetro de calidad empleando dos predictores.

Para extender las capacidades para la generación de modelos predictivos se propuso desarrollar estos a partir del empleo de dos predictores por parámetro de calidad y producto. El mayor porcentaje de eficiencia de todos los ensayos generados con un 87,2% en el modelo Conjunto (árboles agrupados) para determinar el GM con f_3 como primer predictor y masa como el segundo. Los demás parámetros de calidad presentan porcentajes de eficacia menores a los generados para los cuadros 10 y 12, a excepción de la gravedad específica que su mayor eficacia es 1,8%, y para P 3,5% en comparación a los datos dados en el cuadro 10.

Es entonces que se procede a analizar y comparar las matrices de confusión para el GM, CH y GE, debido a que la ESR posee una eficacia mayor en otro modelo. Este comportamiento de alcanzar eficacias generales mayores en algunos parámetros puede deberse a que el uso de dos predictores puede favorecer la asociación de los datos a una característica en estudio, como sucede en el mayor valor alzado de eficacia del cuadro 14. La M logró mostrar una tendencia marcada con las frecuencias de resonancia como se muestra en la figura 11 y estas a su vez con el GM. Por tanto, se usó el GM como segundo predictor para los demás

parámetros ya que las tendencias mostraron asociación entre el desarrollo del GM y las características de la piña.

Cuadro 14. Resultados de la eficacia general por cada parámetro de calidad obtenidos según el tipo de modelo predictivo desarrollado y entrenados con dos predictores.

Modelo	Eficacia general (%)	Primer predictor	Segundo predictor	Parámetro
Conjunto (Árboles agrupados)	87,2	f_3	M	GM
Máquinas de vectores de soporte (Máquinas de vectores de soporte gaussiano medio)	39,8	f_4	GM	ESR
Árbol (Árbol medio)	32,8	f_2	GM	CH
Máquinas de vectores de soporte (Máquinas de vectores de soporte gaussiano medio)	32,7	f_2	GM	FM
Red neuronal (Red neuronal media)	30,6	f_2	GM	GE
Máquinas de vectores de soporte (Máquinas de vectores de soporte gaussiano medio)	25,7	f_2	GM	P

Donde, M es masa, GM es el grado de madurez, ESR es el esfuerzo remanente, CH el contenido de humedad, FM es la firmeza, GE es la gravedad específica, P es la fuerza máxima y f_n son las frecuencias de resonancia.

4.7.1. Matriz de confusión desarrollada para cada parámetro.

4.7.1.1. Grado de madurez con predictores f_3 y masa.

La figura 41.a muestra la matriz de confusión correspondiente a los resultados de este máximo porcentaje de eficacia obtenido, la fila E posee el mejor número de muestras correctamente clasificadas con un 87,3% de observaciones correctamente clasificadas por clase verdadera. Pero el mayor porcentaje dado de TPR es para la clase C con 90,3% seguidos por E, D y B, con valores porcentuales que abarcan desde 71,6% a 87,3%.

Al analizar los resultados predichos correctamente (PPV) por el modelo se tiene un nuevo orden de eficacia, ubicando en primera posición la clase C seguidos por, D, E y B, donde al igual que ocurre en la figura 29.a, el mayor número de observaciones mal predichas están ubicadas en la clase E para la clase verdadera B. Esta tendencia puede deberse al número de muestras que posee la clase E.

Comparando estos resultados con los obtenidos en la figura 21.b pertenecientes al modelo Bayas ingenuas (Bayas ingenuas gaussianas) para f_I como predictor, se tiene que el presente modelo es superior en capacidad de predicción en las cuatro clases en 74,1% para B, 52,9% para C, 88,8% con D y 46% para E. De la misma manera este modelo supera al generado en la figura 26.b correspondiente al K vecinos más cercano ponderado con CR3 como predictor, en B con 42,9%, C con 53,2%, D con 50,8% y E con 34,2%. Por tanto, este modelo es el que presenta la mayor eficacia para predecir el GM, siendo esta mayor al 70% por categoría e incluyendo la capacidad de generalización para determinar todas las clases, por ello es un resultado positivo considerando que es una técnica exploratoria y emplea un bloque de entrenamiento reducido, por ello para alcanzar aún mejores valores el ingreso de un conjunto de datos mayor sería de utilidad.

Este modelo posee el equilibrio de porcentajes deseado donde los valores de TPR y PPV superan sus contrapartes de FNR y FDR, que al aprovecharse de estos altos valores se podría generar un nuevo ensayo para las piñas analizadas en el caso de que se esté empleando este modelo para una máquina en planta de tal manera que los datos generados por segunda vez arrojen una mejor segmentación. A su vez al hacer uso de este modelo en una planta tendría el beneficio de una entrada de datos continuar que le permita un entrenamiento mejor y a su vez los resultados que este obtenga.

Clase Verdadera	B	71,6%	6,0%	1,7%	20,7%
	C	2,3%	90,3%	2,8%	4,5%
	D	4,5%	3,4%	85,8%	6,2%
	E	6,2%	2,2%	4,3%	87,3%
		B	C	D	E
Clase predicha					

(a)

Clase Verdadera	B	74,1%	3,9%	1,2%	8,5%
	C	3,6%	89,3%	2,9%	2,8%
	D	7,1%	3,4%	88,8%	3,9%
	E	15,2%	3,4%	7,1%	84,9%
		B	C	D	E
Clase predicha					

(b)

Figura 41. Matriz de confusión para el modelo de mayor eficacia con el parámetro de grado de madurez con una eficacia del 87,2% y empleando f_3 y masa como predictores. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.

4.7.1.2. Contenido de humedad con predictores f_2 y grado de madurez.

La matriz de confusión obtenida para el modelo del CH que combina dos predictores se muestra en la figura 42. El primer caso de clasificación se tiene que el mayor número de muestras correctamente clasificadas por clase verdadera TPR se ubica en la clase A con un 54,8% lo que representa el doble de lo que se estimó para las otras tres clases, pero que a su vez se permite cierta flexibilidad para generalizar. La figura 24 muestra que el comportamiento de las observaciones positivas posee una concordancia con la distribución de muestras por clase según GM ya que los grupos B, C y D poseen cierta semejanza en la tendencia de la cantidad de datos, lo cual sucede con la similitud en los porcentajes positivos, mientras que A es la que mayor valor obtuvo, además presenta una mayor similitud entre la cantidad de muestras por GM.

Al analizar la capacidad predictiva de este modelo se encuentra que posee porcentajes cercanos que rondan un mínimo del 29,9% para la clase B a un 36,8% en su máximo para la clase C. Al comparar estos resultados con los obtenidos en la figura 32, para el modelo Conjunto (árboles agrupados) para el CH con CR1, se estima que el presente modelo posee una mejor eficacia en un 2,3%, y en cuanto al PPV este resulta mejor para las clases B en un

9,4% y C en un 3,8%, mientras que para A resulta menor en un 0,4% y en D en un 1,8%. Puede concluirse que la utilización de dos predictores favorece al modelo, ya que se conserva la capacidad de generalización, pero se optimiza en dos clases y en las dos restantes disminuye menos de lo que se logra aumentar en las otras.

Clase Verdadera	A	54,8%	19,4%	10,8%	15,10%
	B	34,3%	26,9%	15,4%	23,4%
	C	35,7%	21,9%	21,4%	20,9%
	D	41,7%	16,6%	13,4%	28,3%
		A	B	C	D

Clase predicha

(a)

Clase Verdadera	A	32,9%	22,9%	17,5%	17,20%
	B	19,4%	29,9%	23,7%	25,2%
	C	22,6%	27,4%	36,8%	25,2%
	D	25,2%	19,7%	21,9%	32,5%
		A	B	C	D

Clase predicha

(b)

Figura 42. Matriz de confusión para el modelo de mayor eficacia con el parámetro de contenido de humedad con una eficacia del 32,8% y empleando f_2 y grado de madurez como predictores. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.

La tendencia del histograma de la figura 25 posee una característica con respecto a la clase A y es la poseer una variación con los cuatro GM con distribución más equitativa en comparación a las otras clases. Al considerar este comportamiento y los porcentajes de eficacia de la figura 42 en clasificación está posee una posible relación en donde una distribución más equitativa por GM dentro de la clase favorezca este porcentaje si se desarrolla el modelo con dos predictores ya que en los dos casos anteriores, esta diferencia de porcentajes en la matriz de confusión era menor a un 10%.

4.7.1.3. Firmeza con predictores f_2 y grado de madurez.

Los resultados mostrados en la figura 43.a no presentan una mejora en cuanto a generalización y observaciones correctamente clasificadas como ocurre con los otros parámetros asociados a un doble predictor, pero si se logra obtener para la segunda clase un porcentaje mayor al 10% siendo de 35,5% para la clase E y se mantiene la clase B como la de mayor valoración y tendencia con 92,3% de TPR.

Dentro de los tres bloques de estudio realizados de frecuencias, coeficientes de rigidez y doble predictor para los modelos generales los que mostraron mejor porcentaje en al menos dos clases fue para este último bloque como se menciona, pero si se considera el porcentaje correctamente predicho es mayor para el bloque de coeficientes de rigidez ya que este resulta mejor en un 6,8%.

Al analizar las observaciones correctamente predichas se tiene que rondan de un 22,2% a 37,6% siendo el mayor de estos para la clase E. Considerando los resultados de los demás bloques de estudio se tiene que es el mayor obtenido, donde el más cercano de los porcentajes es para el bloque de coeficientes de rigidez con un 9,1%.

Ante la situación de determinar firmeza este modelo presenta una eficacia de 32,7% lo cual resulta por debajo de un 50% y si tomamos en consideración los datos mostrados por la matriz de confusión esta trabaja mejor en determinar la clase B, pero existe una tendencia a clasificar igualmente muestra que no pertenece a esta clase en ella por lo cual puede generar cierto error el uso de este modelo para determinar este parámetro de calidad.

Clase Verdadera	A	92,9%			7,1%		
	B	92,3%	1,8%		5,9%		
	C	90,8%	1,5%		7,7%		
	D	85,5%	0,8%		13,7%		
	E	63,6%	0,9%		35,5%		
	F	68,0%	2,0%		30,0%		
		A	B	C	D	E	F

Clase predicha

(a)

Clase Verdadera	A	16,6%			7,9%		
	B	32,0%	44,4%		12,9%		
	C	18,6%	22,2%		9,9%		
	D	16,7%	11,1%		16,8%		
	E	10,7%	11,1%		37,6%		
	F	5,4%	11,1%		14,9%		
		A	B	C	D	E	F

Clase predicha

(b)

Figura 43. Matriz de confusión para el modelo de mayor eficacia con el parámetro de firmeza con una eficacia del 32,7% y empleando f_2 y grado de madurez como predictores. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.

4.7.1.4. Gravedad específica con predictores f_2 y grado de madurez.

Los resultados mostrados en la figura 43.a indican que el mayor número de muestras correctamente clasificadas se encuentran en la clase A con un 44% de seguidos por las clases B, D, C, E y F. La clase A esta constituida en la matriz de datos por 4,76% de GM tipo B, 28,57% tipo C, 21,43% tipo D y 45,24% tipo E, como se muestra en la figura 27, es decir, posee un mayor número de piñas con el último grado de madurez considerado. Al analizar los valores de PPV se encuentra en esta ocasión que el mayor porcentaje de observaciones correctamente predichas para clase verdadera están dados para la D con un 35,8% seguidos por F, A, B, C y E, en cambio el porcentaje de FDR para todas las clases es superior al 60%.

Comparando estos resultados con los mostrados en la figura 26.b pertenecientes al modelo Conjunto (árboles potenciados) con f_3 como predictor, se obtiene que el actual modelo posee valores más altos de PPV para las clases A en 3,5%, C en 4,1%, D en 12,8%, E en 9,2% y F

en 33,3%. En cuanto al modelo sustentado por frecuencias es mejor en la clase B en 11,2%. Por tanto, el actual modelo posee una utilidad mayor en la mayoría de las clases, además de mostrar mayor capacidad de generalización al obtener datos con la última clase, pero únicamente mostrando una eficacia predictiva con un mínimo del 28,1% y máximo de 35,8%.

Este modelo también presenta una tendencia a la clasificación hacia la clase A, donde los casos de mayor porcentaje en comparación a los TPR ocurren para E y F. Esta característica podría aprovecharse para emplear el modelo como método para purificar los resultados en cada mezcla de producto analizado. De esta manera ir descartando las clases de piñas con mejores porcentajes de TPR, hasta obtener un conjunto con mejor clasificación.

En la situación de optar por la creación de un nuevo modelo, resulta mejor para el entrenamiento del algoritmo de aprendizaje el empleo de una misma cantidad de datos para los grupos A, B, C, D, etc. Esta regularidad de datos entre clases permite reducir la segmentación de datos hacia una clase como sucede en la figura 38 que tiende un sesgo que favorece al grupo predominante. Estos porcentajes correctamente clasificados están influenciados como en anteriores modelos para gravedad específica por la cantidad de muestras por clase, en donde A obtuvo el mejor porcentaje y F la más baja de todas. Esta clase A esta compuesta por un mayor número de muestras con GM tipo E como se muestra en la figura 27 perteneciente al histograma.

Clase Verdadera	A	44,0%	21,4%	13,7%	10,1%	10,7%	
	B	26,6%	32,8%	21,9%	8,6%	10,2%	
	C	27,8%	11,1%	29,9%	11,1%	19,4%	0,7%
	D	19,5%	10,2%	18,0%	30,5%	19,5%	2,3%
	E	27,4%	13,7%	20,2%	16,1%	22,6%	
	F	32,7%	9,6%	21,2%	11,5%	21,2%	3,8%
		A	B	C	D	E	F

Clase predicha

(a)

Clase Verdadera	A	33,0%	27,9%	15,0%	15,6%	14,6%	
	B	15,2%	32,6%	18,3%	10,1%	10,6%	
	C	17,9%	12,4%	28,1%	14,7%	22,8%	16,7%
	D	11,2%	10,1%	15,0%	35,8%	20,3%	50,0%
	E	15,2%	13,2%	16,3%	18,3%	22,8%	
	F	7,6%	3,9%	7,2%	5,5%	8,9%	33,3%
		A	B	C	D	E	F

Clase predicha

(b)

Figura 44. Matriz de confusión para el modelo de mayor eficacia con el parámetro de gravedad específica con una eficacia del 30,6% y empleando f_2 y grado de madurez como predictores. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.

4.8. Modelos de red neuronal con mayor eficacia para cada parámetro de calidad empleando dos predictores.

Los modelos en redes neuronales se adecuan a la magnitud de muestras. Al igual que sucede con los resultados generales de mayor precisión se obtuvo un porcentaje más alto para ESR (0,4% menor al modelo Árbol (árbol grueso) correspondiente a la prueba destructiva y en sucesión está el GM (1% menor a Bayas ingenuas (Bayas ingenuas gaussianas) sujeto a una prueba no destructiva y en mínima instancia está la masa (0,8% menor a Árbol (árbol fino) como se muestra en el cuadro 15. Además, la tendencia de las frecuencias de resonancia con respecto al tipo de prueba destructiva/no destructiva presenta que f_1 tanto para los modelos generales como para red neuronal está asociado únicamente al GM, f_2 por su parte trabaja con parámetros destructivo y no destructivos, f_3 con la GM y f_4 describe la firmeza de la piña.

Entonces la proporción de utilidad de las frecuencias es de f_1 con 14,29 %, f_2 con 42,86 %, f_3 con 14,29% y f_4 con 28,57%.

Cuadro 15. Resultados de la eficacia general por cada parámetro de calidad obtenidos según el tipo de red neuronal desarrollado y entrenados con dos predictores.

Modelo red neuronal	Eficacia general (%)	Primer predictor	Segundo predictor	Parámetro
Ancha	51,3	f_4	Masa	Grado de madurez
Estrecha	39,5	f_2	GM	Esfuerzo remanente
Estrecha	30,6	f_2	GM	Contenido de humedad
Media	30,6	f_2	GM	Gravedad específica
Ancha	30,1	f_2	GM	Firmeza
Dos capas	23,1	f_4	GM	Fuerza de penetración

Donde, GM es el grado de madurez y f_n son las frecuencias de resonancia.

4.8.1. Resultados para los parámetros de calidad obtenidos mediante redes neuronales.

4.8.1.1. Grado de madurez con predictores f_4 y masa.

La matriz de confusión mostrada en la figura 44.a indica que en la clase E ocurren la mayor cantidad de muestras correctamente clasificadas al igual que ha sucedido en los anteriores casos evaluados para GM a causa del mayor número de piñas con GM con la clase E. Esta clasificación posee el 60,5% de observaciones correctamente clasificadas como clase verdadera, seguidos por las clases C, B y D. Además, los valores de FNR totales para las clases están dados entre 39,5% y 60,2%, ubicándose los mayores porcentajes de observaciones mal clasificadas como clase verdadera en E.

Por su parte los PPV muestra porcentajes que contemplan un rango que va de 47% a 54,6%, ubicándose el mayor de estos valores en la clase B. Es entonces que la tasa de observaciones mal clasificadas como clase verdadera (FDR) para los cuatro GM poseen un rango entre 47%

a 54,6% lo cual posee porcentajes más cercanos entre clases a los obtenidos en comparación a la figura 33.a.

Al comparar esta capacidad predictiva a la alcanzada por el modelo de la figura 33.b. correspondiente también al GM se determina que la presente red neuronal posee mejor eficacia, al ser mayor para la clase B en 41,3%, para C en 15%, D en 22,8% y E en 13,7%. Estos datos favorecedores para predecir el GM se pueden deber a causa del cambio dado de la masa en el transcurso de maduración unida a la variación que esta misma tiene con el sonido que emite ante los golpes, como se describió en la tendencia de este.

Los resultados obtenidos al igual que los dados para el GM con dos predictores para los modelos generales, poseen porcentajes más altos de TPR entre sus clasificaciones por clase más no por las tasas de FNR. Por tanto, una opción para el uso del modelo actual es la aplicación de reiteradas revisiones para mejorar la clasificación, de tal manera que se logre retirar de forma paulatina los mejores resultados a los peores.

Clase Verdadera	B	45,7%	12,9%	11,2%	30,2%
	C	5,1%	52,3%	16,5%	26,1%
	D	2,3%	21,0%	39,8%	36,9%
	E	11,2%	14,9%	13,4%	60,5%
		B	C	D	E
Clase predicha					

(a)

Clase Verdadera	B	54,6%	8,1%	8,7%	11,2%
	C	9,3%	49,7%	19,5%	14,7%
	D	4,1%	20,0%	47,0%	20,8%
	E	32,0%	22,2%	24,8%	53,4%
		B	C	D	E
Clase predicha					

(b)

Figura 45. Matriz de confusión para la red neuronal de mayor eficacia con el parámetro de grado de madurez con una eficacia del 51,3% y empleando f_4 y masa como predictores. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.

4.8.1.2. Esfuerzo remanente con predictores f_2 y grado de madurez.

Al examinar los resultados obtenidos en la matriz de confusión para la ESR con dos predictores perteneciente a la figura 45.a, se tiene que la clase B obtiene el mayor número de muestras correctamente clasificadas con un 82,3% por clase verdadera, seguido por la clase C con 15,8% de TPR. Estos resultados denotan una tendencia a favor de la clase B ya que hay porcentajes elevados de observaciones que están clasificados como esta clase.

Adicionalmente los resultados indican que las PPV poseen valores porcentuales con una variación entre 11,1% al 42,7% lo cual indica un menor distanciamiento entre las clases como sucede con los datos obtenidos de la Figura 34.a. Se muestra una mejor capacidad predictiva para la clase B y en el peor caso para D.

En cuanto a la comparación de este modelo con el alcanzado en la figura 28.b perteneciente a la red neuronal con f_2 como predictor, se tiene que el presente posee mayores PPV en las clases B en 3,2% y D en 11,1% y menor para A en 10,4% y C en 10,6%. Por la capacidad de generalización del modelo actual resulta en un mejor predictor, pero el basado en frecuencias posee mejor porcentaje de eficacia en dos de tres clases que ofrece para ESR.

Clase Verdadera	A	14,3%	73,1%	12,6%	
	B	7,8%	82,3%	8,5%	1,4%
	C	14,6%	67,1%	15,8%	2,5%
	D	10,0%	78,2%	10,9%	0,9%
		A	B	C	D

Clase predicha

(a)

Clase Verdadera	A	31,3%	23,5%	27,1%	
	B	27,7%	42,7%	29,4%	44,4%
	C	27,7%	18,7%	29,4%	44,4%
	D	13,3%	15,2%	14,1%	11,1%
		A	B	C	D

Clase predicha

(b)

Figura 46. Matriz de confusión para la red neuronal de mayor eficacia con el parámetro de esfuerzo remanente con una eficacia del 39,5% y empleando f_2 y grado de madurez como predictores. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.

El comportamiento encontrado en los porcentajes de eficacia para el ESR empleando dos predictores posee una relación con la tendencia registrada en la figura 20, correspondiente al histograma del porcentaje de muestras por clase, donde también se registran sus cantidades de muestras por GM. Como en los anteriores modelos para el ESR, la clase B que posee mejor eficacia es la que posee mayor cantidad de muestras y el menor porcentaje de eficacia también se ajusta a la menor cantidad de muestras para una clase. Se considera además que la tendencia esperada para el ESR es que las muestras con mayor GM se encuentren en las últimas clases y por ende las que poseen GM cercanos a B estén en mayor cantidad en las primeras clases, ya que la dureza es mayor en las muestras con menor GM que una con GM E. Esto puede involucrar que en el modelo la clase B con 82,3% posee un gran porcentaje de muestras con GM de E y un menor caso de B.

4.8.1.3. Gravedad específica con predictores f_3 y grado de madurez.

Al comparar los resultados predictivos de la presente red neuronal artificial que resulta de la figura 43.b con los obtenidos en la figura 26.b se determina que el actual modelo es mayor para la clase A en 5,2%; C en 4,5%; D en 17,3%, E en 10% y F en 33,3%, únicamente siendo menor en la clase B en 5,5%. El presente modelo de red neuronal con f_2 y GM como predictores posee una mejor eficacia predictiva en contraste al creado con f_3 con una tendencia favorecida por el aporte dado del GM ya que se obtiene una mejor relación de datos de este parámetro con la GE y por tanto con las frecuencias de resonancia.

4.8.1.4. Firmeza con predictores f_2 y grado de madurez.

Los resultados obtenidos para las observaciones correctamente predichas de la figura 46.a indican un máximo de 51,4% para la clase B, seguidos por la clase E y A, siendo el mínimo alcanzado para la clase F. Se considera al tomar en cuenta los resultados de los anteriores bloques de estudio en la categoría de redes neuronales que el presente bloque empleando doble predictor obtiene la mejor capacidad de generalización ya que permitió obtener observaciones positivas en todas las clases. Además, se tiene que la tendencia mostrada tanto en modelos generales como redes neuronales no resulta elevada ya que no supera el 43,4%.

Al considerar el bloque de frecuencias para redes neuronales y los presentes resultados, se obtiene que este es mejor en un 20,4% para la clase A, 9,3% para C, 13,7% para D, 26,1% para E y 12% para F, siendo únicamente menor para B en un 35%. Mientras que las observaciones correctamente predichas mantienen porcentajes que rondan de un 19,4% a un 35,5% ubicando la mayor de estas para la clase B.

Por tanto, el uso de doble predictor resulto ser mejor para desarrollar modelos con redes neuronales para determinar firmeza y si se considera los modelos generales este también resulta superior a estos, por lo cual la elección de redes neuronales como modelo para analizar firmeza muestra puntos favorecedores tanto de resultados como de mejora. Como primer punto a destacar es la factibilidad que muestra esta clase de modelos para trabajar con esta clase de datos ya que permitió asociar mejor el parámetro de calidad con frecuencia de resonancia 2 si se suma dentro de los predictores el GM.

Clase Verdadera	A	25,7%	43,4%	13,3%	7,1%	8,8%	1,80%
	B	16,4%	51,4%	15,9%	6,4%	7,3%	2,70%
	C	16,9%	40,0%	20,8%	7,7%	8,5%	6,2%
	D	8,1%	41,9%	15,3%	15,3%	16,9%	2,4%
	E	8,4%	34,6%	9,3%	14,0%	28,0%	5,60%
	F	14,0%	30,0%	10,0%	12,0%	22,0%	12,0%
			A	B	C	D	E

Clase predicha

(a)

Clase Verdadera	A	25,7%	15,4%	13,5%	11,1%	10,1%	6,5%
	B	31,9%	35,5%	31,5%	19,4%	16,2%	19,4%
	C	19,5%	16,4%	24,3%	13,9%	11,1%	25,8%
	D	8,8%	16,4%	17,1%	26,4%	21,2%	9,7%
	E	8,0%	11,6%	9,0%	20,8%	30,3%	19,4%
	F	6,2%	4,7%	4,5%	8,3%	11,1%	19,4%
			A	B	C	D	E

Clase predicha

(b)

Figura 47. Matriz de confusión para la red neuronal de mayor eficacia con el parámetro de firmeza con una eficacia del 30,1% y empleando f_2 y grado de madurez como predictores. a. Observaciones correctamente clasificadas. b. Observaciones correctamente predichas.

4.9. Modelos con mayor eficacia para predicción de los parámetros de calidad a partir de tres bloques de predictores.

A partir de los resultados finales obtenidos de las comparativas aplicadas a las eficacias y a las matrices de confusión en cuanto a la capacidad predictiva de los tres conjuntos de datos (frecuencias de resonancia, CR y dos predictores), se creó el cuadro 16 donde se muestran las eficacias y los mejores modelos a nivel predictivo. Por parte de los predictores mejor calificados hay una tendencia marcada en la utilización de dos predictores para cada parámetro, al observarse su uso en cuatro de seis de estas. En tanto las otras dos categorías están comprendidas por los predictores de CR que trabajan para la firmeza.

Cuadro 16. Modelos generales con mayor eficacia a partir de los tres ensayos elaborados.

Modelo	Eficacia general (%)	Predictor	Parámetro
Conjunto (Árboles agrupados)	87,2	f_3 - M	GM
Árbol (Árbol grueso)	40,5	CR1	ESR
Árbol (Árbol medio)	32,8	f_2 - GM	CH
Máquinas de vectores de soporte (Máquinas de vectores de soporte gaussiano medio)	32,7	f_2 - GM	FM
Red neuronal (Red neuronal media)	30,6	f_2 - GM	GE
Árbol (Árbol grueso)	21,8	CR2	P

Donde, M es la masa, CR son los coeficientes de rigidez, GM es el grado de madurez, ESR es el esfuerzo remanente, CH es el contenido de humedad, FM es firmeza, GE es la gravedad específica, P es la fuerza de penetración y f_n son las frecuencias de resonancia.

Estos valores de eficacia más altos fueron favorecidos al emplear el segundo predictor de masa para GM, GM para GE, ESR, P y FM, esto al considerar los resultados obtenidos de las tendencias encontradas de las asociaciones entre las frecuencias de resonancia y los parámetros de calidad. Esto puede observarse de las variaciones que sufre la masa en el transcurso de la madurez de la piña y esta a su vez afecta el sonido emitido en respuesta a las

colisiones. De igual manera el CH sufre un desarrollo paulatino en el transcurso de la madurez y este mismo aumento de la humedad repercute en el tono del sonido reflejado producto de la percusión. Mientras que la GE está influenciada por el CH del fruto al igual que por otros factores como la porosidad, entre otras, que son ajenos al conjunto de datos.

Los resultados en los parámetros estructurales están dados por los predictores de CR1 y CR2, lo cual coincide con resultados previos de otros productos como la sandía como lo muestra (Mao, Yu, Rao, & Wang, 2016) donde los CR se agrupan o asocian de mejor manera a parámetros de firmeza o estructurales. De forma contraria el FM se asoció mejor a la frecuencia y el GM (todos los parámetros de la firmeza también fueron entrenados con estos dos predictores), mostrando una mejora sustancial si se emplean redes neuronales para la asociación entre estos datos ya que se mostró una mayor capacidad de generalización y reducción de tendencia por la clase B.

Los porcentajes de eficacia en cinco de los seis parámetros no superan el 40% y en el mínimo de ellos alcanza el 21,8%, por ello la predicción del GM posee un porcentaje positivo de utilidad y una seguridad superior al 85%. Estas condiciones generadas permiten definir a todos los demás modelos como métodos para desarrollar una noción en cuanto a la condición de la piña que por tanto debe ser apoyada por alguna prueba extra sea esta destructiva o por imágenes. Además, la integración de un conjunto de datos de mayor volumen que mejoren los resultados de eficacia y de generalización en cuanto a las calidades mostradas por GM.

Al considerar el análisis realizado y el requerimiento de crear un modelo predictivo que relacione las frecuencias de resonancia con los parámetros de GE, firmeza o estructurales. Donde los resultados finales relacionados a la mejor eficacia mostraron ESR como el de mayor porcentaje y por último al CH. Se integra el GM como complemento para una estimación de mejor calidad, al encontrarse en este parámetro la mejor asociación entre todos los modelos generados.

Al mostrar porcentajes de eficacia variados en la mayoría de las clases para los modelos generados, estos no resultan completamente efectivos para tratar con gran variedad de piñas. Hay que considerar el hecho que los modelos podrían mejorar sus capacidades al entrenarse con un lote mayor de producto.

Como método de aplicación de los modelos generales de mejor eficacia (cuadro 13) para una línea de procesamiento en planta se propone la realización de un análisis puntual por cada modelo con un parámetro de calidad en específico. El primero de estos puede ser el Conjunto (árboles agrupados) que determina el GM ya que su capacidad de predicción es mayor al 74% en todas sus clases, lo cual ofrece mayor seguridad y un medio de comparación a los demás modelos en línea. Donde para ESR, CH y GE la mayoría de las predicciones en sus clases son cercanas al 30% a excepción de la ESR ya que la clase D posee un porcentaje superior al 50% que de acuerdo con el anexo 8 esta clasificación posee una distribución de muestras por GM entre 25 a 30 unidades por cada uno de los grados. Únicamente para el caso de la firmeza es preferible el uso de las redes neuronales para la integración al ajuste para la línea de procesamiento ya que por parte de los modelos generales estos se integran en menor medida al parámetro de calidad. Esta condición provee datos imprecisos por sí solo pero que pueden ser mejorados en alguna medida por el modelo del GM, lo que también podría resultar de utilidad en las otras clases que además presentan variabilidad de GM en sus clasificaciones.

En cuanto al CH y GE presenta el mismo inconveniente que el ESR en la distribución de las muestras por GM. En consecuencia, para no acumular un error al tomar los resultados de ESR se emplea de nuevo el modelo de GM para apoyar los resultados del CH y sustentar el diagnóstico. Este paso permite que se cargue cada diagnóstico de parámetro de calidad con el error del modelo de mejor eficacia. Igualmente, estos modelos como se ha mencionado pueden mejorarse con la integración de mayor cantidad de datos, así como de otros métodos como análisis por imágenes para registrar un dato más preciso a nivel visual de las piñas y automatizar esta caracterización que igualmente puede ser apoyada por personal.

Otra alternativa para la implementar lo desarrollado en planta es emplear únicamente el modelo que determina el GM para estimar los parámetros de firmeza, CH y GE, según los rangos en los anexos del 3 al 8, asociando los cuatro GM con las clases generadas de cada parámetro. En los casos donde estas clasificaciones exceden la cantidad de cuatro, crear nuevas particiones que permitan asociar a estos GM como se muestra en el ejemplo del cuadro 16. Al desarrollarse de esta manera se permite estimar los rangos ya sea de firmeza o estructurales (FM, ESR y P), CH y GE en el que se encuentra la piña únicamente valorando

el GM. Este método puede aproximar los resultados a la realidad, pero no asegura una valoración más puntual como lo haría la primera implementación en la planta propuesta.

Cuadro 17. Distribución de muestras según grado de madurez para las categorías de la gravedad específica para piñas.

Categoría	Rango	Grado de madurez
A	0-0,884	B
B	0,884-0,893	C
C	0,893-0,905	D
D	0,905-0,923	E

Se destaca a partir de estos porcentajes de acierto que, a pesar de emplear un conjunto de muestras reducido se obtuvo un primer acercamiento positivo en cuanto al análisis de datos, al obtener información relevante para la investigación en esta área. Además, se obtuvo una visualización inicial para mejorar los algoritmos como los que se describen en las propuestas mencionadas. Con lo cual queda abierto para aplicar nuevas mejoras a la optimización de estos algoritmos.

4.10. Modelos de redes neuronales con mayor eficacia para predicción de los parámetros de calidad a partir de tres bloques de predictores.

Los modelos de redes neuronales de mejor eficacia del cuadro 18 creados a partir de resonancia, CR y dos predictores poseen menor eficacia que los comprendidos en el cuadro 16 a excepción de la P, pero esto se debe a que el generado para este cuadro posee mejores resultados predictivos entre los máximos generales de los otros grupos (frecuencias y doble predictor).

Cuadro 18. Resultados de la eficacia general por cada parámetro de calidad obtenidos según el tipo de red neuronal desarrollado y de los tres ensayos elaborados.

Modelo red neuronal	Eficacia general (%)	Predictor	Parámetro
Ancha	51,3	$f_4 - M$	Grado de madurez
Estrecha	39,2	CR3	Esfuerzo remanente
Estrecha	31,7	f_2	Contenido de humedad
Ancha	30,1	$f_2 - GM$	Firmeza
Media	30,6	$f_2 - GM$	Gravedad específica
Dos capas	23,1	$f_4 - GM$	Fuerza de penetración

Donde, M es la masa, GM es el grado de madurez, CR son los coeficientes de rigidez y f_n son las frecuencias de resonancia.

La tendencia mostrada por los predictores muestra mayor variabilidad entre los parámetros, definiéndose el modo de doble predictor para dos parámetros como son la FM y la P, únicamente se usa el CR para la ESR (Cuadro 18). Otro resultado dado es que se mantiene la combinación de frecuencia y masa como mejor resultado para predecir el GM, ya que como se vio en las tendencias previas realizadas poseían un comportamiento ideal en todo el conjunto.

A manera general las redes neuronales mostraron mejor generalización al clasificar y predecir las clases para cada parámetro. Donde únicamente para el GM se da la mayor diferencia en los porcentajes de eficacia, por ello representan los modelos a considerar para trabajar con ellos.

El mismo método de integración para procesamiento en poscosecha planteado para los modelos generales de mejor eficacia aplica para estos métodos, ya que estos pueden verse mejorados por la cantidad de datos. Además, tomando en cuenta la capacidad de generalización y desarrollo que poseen las redes neuronales artificiales pueden llegar a diagnósticos muy favorables con una calidad de datos mejor con las cuales entrenarse.

4.11. Modelos con mayor eficacia para predicción a partir de tres bloques de predictores para 5% de validación cruzada.

Los modelos mostrados en el cuadro 16 se volvieron a generar, pero con un 5% de VC, para generar el cuadro 19 y realizar una comparación de los resultados entre ambos conjuntos de datos para el entrenamiento. Los resultados en el cuadro 19, muestran una eficacia menor en todos los modelos, lo cual puede involucrar que se ha generado un sobreajuste. Esta cantidad de observaciones mayores creó un modelo sobreajustado que al ingresar ese 5% para validar no consiguiera una óptima predicción por la baja adaptabilidad del modelo.

Cuadro 19. Resultados de la eficacia general por cada parámetro de calidad obtenidos según el tipo de modelo desarrollado a partir de los tres ensayos elaborados para un 5% de validación cruzada.

Modelo	Eficacia general (%)	Predictor	Parámetro
Conjunto (Árboles agrupados)	82,1	$f_3 - M$	Grado de madurez
Árbol (Árbol grueso)	40,2	CR1	Esfuerzo remanente
Máquinas de vectores de soporte (Máquinas de vectores de soporte gaussiano medio)	31,3	$f_2 - GM$	Firmeza
Árbol (Árbol medio)	28,9	$f_2 - GM$	Contenido de humedad
Red neuronal (Red neuronal media)	26,9	$f_2 - GM$	Gravedad específica
Árbol (Árbol grueso)	21,1	CR2	Fuerza de penetración

Donde, M es la masa, GM es el grado de madurez, CR son los coeficientes de rigidez y f_n son las frecuencias de resonancia.

Al analizar igualmente el cuadro 18 con el cuadro 19 se tiene el mismo comportamiento en la eficacia al mostrado en los modelos generales de mayor eficacia. Los resultados del cuadro 14 presentan una mayor eficacia producto de que poseen un sobreajuste que les impide una mejor capacidad de generalización con respecto a los nuevos datos a ingresar. A partir de estas comparaciones se reafirma la utilización del 30% de VC que es el valor mínimo por

emplear para entrenamiento y validación de los modelos. Esto asegura un porcentaje de eficacia que adapta competente los modelos como elementos de apoyo en planta u otro espacio de trabajo.

Cuadro 20. Resultados de la eficacia general por cada parámetro de calidad obtenidos según el tipo de red neuronal desarrollado a partir de los tres ensayos elaborados para un 5% de validación cruzada.

Modelo red neuronal	Eficacia general (%)	Predictor	Parámetro
Ancha	48,5	$f_4 - M$	Grado de madurez
Estrecha	36,6	CR3	Esfuerzo remanente
Estrecha	32,1	f_2	Contenido de humedad
Ancha	28,4	$f_2 - GM$	Firmeza
Media	26,9	$f_2 - GM$	Gravedad específica
Dos capas	21,2	$f_4 - GM$	Fuerza de penetración

Donde, M es la masa, GM es el grado de madurez, CR son los coeficientes de rigidez y f_n son las frecuencias de resonancia.

4.12. Factores de uniformidad para cada piña para los parámetros de calidad.

Cada piña posee cuatro puntos de medición que permitieron generar esa misma cantidad de observaciones y estas a su vez fueron clasificadas dentro de las categorías generadas por frecuencias agrupadas para los demás parámetros de calidad. Por tanto, para comprobar que las piñas son uniformes entre los cuatro conjuntos de datos se sometieron a la determinación de un factor de uniformidad (FU). Un valor alto del grupo 4 es un indicativo de que cada piña posee sus cuatro bloques con la misma clasificación para el parámetro respectivo y por consiguiente un valor elevado del grupo 1 mostraría que en sus cuatro puntos la piña posee diferentes valores de calidad.

Inicialmente se tiene el CH que como muestra la figura 47, un 47,31% de piñas poseen la mitad de sus puntos con el mismo CH y 37,17% piñas con tres puntos iguales, mientras que únicamente 12,37% logran ser completamente uniformes en su CH.

El factor de uniformidad de 3 y 4 para CH es el mayor alcanzado entre todas las pruebas para los parámetros de calidad. Al tomar en cuenta el 32,8% de eficacia obtenido para el modelo general del cuadro 16 y el 31,7% del cuadro 15, se puede considerar que representa unos datos de mayor viabilidad para desarrollar el modelo ya que es el tercer parámetro de mejor evaluación, considerando que el primero se descarta para en este caso ya que es el grado de madurez y este es el mismo para los cuatro puntos de cada piña. Entonces se puede esperar que el ingreso de una mayor cantidad de datos para el modelo puede favorecer la eficacia mostrada por este, al igual que asegurar una homogeneidad en la cada piña.

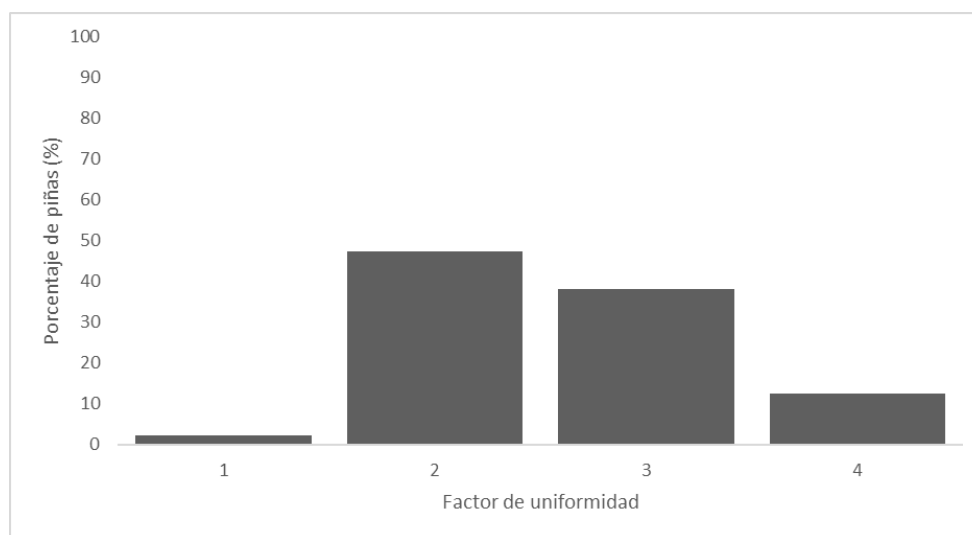


Figura 48. Factor de uniformidad para cada piña para el parámetro de contenido de humedad.

Para el caso de la fuerza de penetración (figura 48), 52,15% de unidades poseen la mitad de sus observaciones con la misma clasificación, seguidos por 26,34% de unidades que resultan diferentes en todas sus observaciones, teniendo en último lugar con 4,84% unidades las piñas uniformes.

Al considerar estos FU con las eficacias del cuadro 16 correspondientes a los modelos generales con mayor eficacia, se tiene que el menor porcentaje encontrado en los parámetros de calidad fue para P, mismo que posee el mayor porcentaje de FU para 1. Esto podría corresponder a que una variedad de clases describe el mismo GM, lo cual perjudica la capacidad del algoritmo predictivo para generar un mejor porcentaje de eficacia al no encontrar asociaciones lógicas entre los datos.

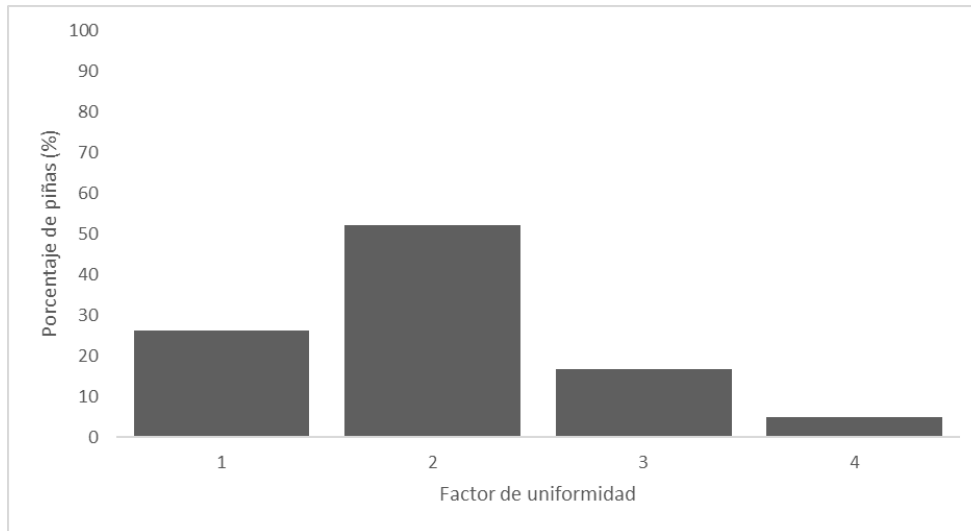


Figura 49. Factor de uniformidad para cada piña para el parámetro de fuerza de penetración.

Para el FM (figura 49) se tiene que 58,6% de unidades tienen la mitad de sus observaciones iguales, y 28,49% de unidades poseen tres observaciones con la misma clasificación. Mientras tanto para las unidades con ninguna similitud entre sus puntos, así como las homogéneas poseen cada una el 6,45% de unidades respectivamente. Los resultados dados para el FM indican que posee el porcentaje más alto de unidades con una homogeneidad media, y también el que ocupa el segundo valor en cuanto a FU de 3 lo que representa que las unidades de producto para generar el modelo poseen condiciones aceptables de uniformidad, de allí que el porcentaje de eficacia de los modelos para FM de los cuadros 16 y 18 son cercanos al CH.

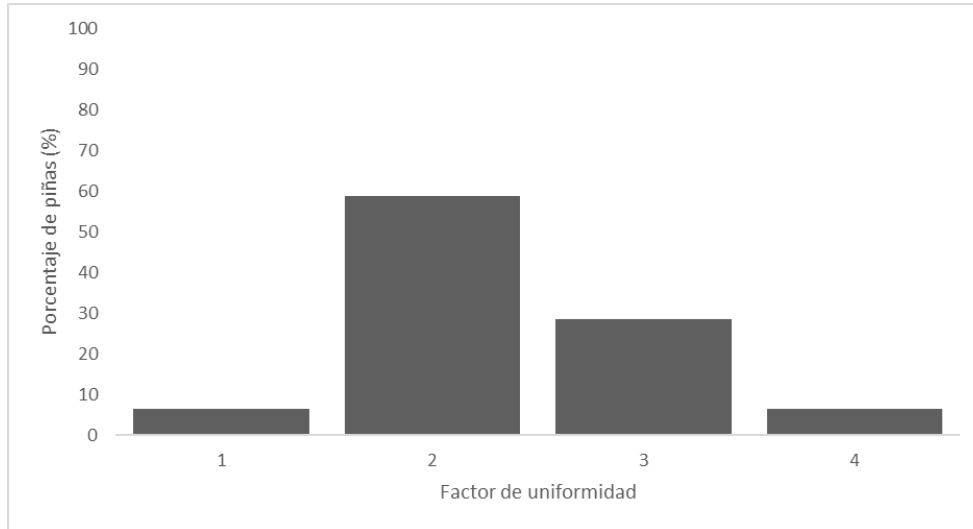


Figura 50. Factor de uniformidad para cada piña para el parámetro de firmeza.

Por último, en la figura 50 muestra que el ESR posee 56,99% de unidades con la mitad de sus valores con la misma clasificación, seguidos de 26,88% unidades con tres puntos de igual clasificación. Mientras que únicamente 7,53% piñas son completamente homogéneas, lo que las coloca por debajo de las 8,6% piñas que son diferentes en sus cuatro puntos de medición. Los porcentajes de 40,5% para el modelo general y 39,2% de la red neuronal artificial, ambos para los porcentajes de eficacia más altos, se encuentra aceptables y con una afinidad mayor que el parámetro de FM, ya que permite hallar una relación entre CR1 y CR3 con la ESR.

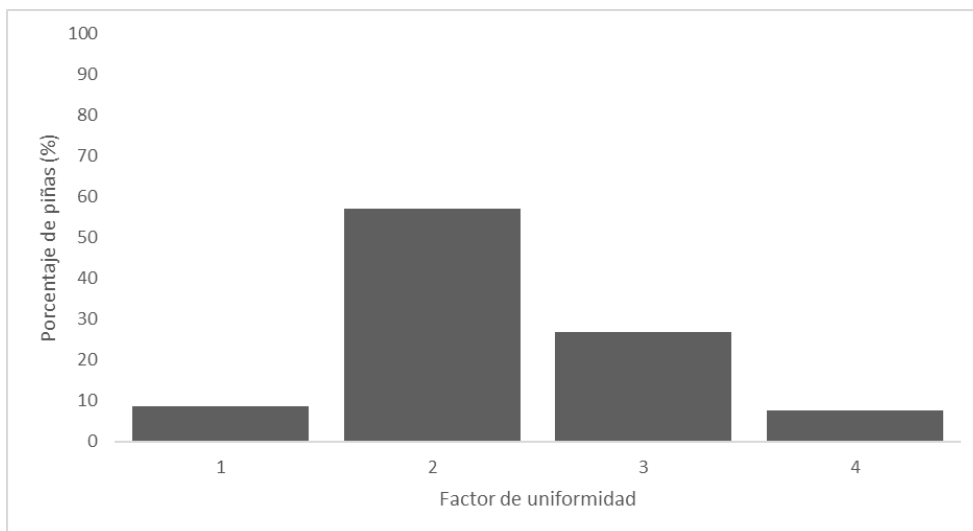


Figura 51. Factor de uniformidad para cada piña para el parámetro de esfuerzo remanente.

Los cuatro parámetros de calidad poseen mayoritariamente un FU de dos seguido por tres a excepción de P. La que presentó mayor uniformidad fue el CH. Estos resultados se ven influenciados por las correcciones aplicadas a las categorías generadas por frecuencias agrupadas, por tanto, estos escenarios pueden mejorar o empeorar en el caso de mantener las categorías iniciales sin alterar, al crear un mayor o menor número de clases. Además, estos resultados implican que una sola piña puede indicar diferentes grupos dependiendo del punto de medición lo cual da una aproximación al valor real del parámetro de calidad.

Se puede destacar que la uniformidad no ejerce una influencia sobre el desarrollo de los modelos debido a que el algoritmo analiza la información de cada punto ingresado al sistema y lo clasifica. Este aspecto no está influenciado por la uniformidad de la muestra, pero sí representa una dificultad para clasificar una unidad que no se encuentre uniforme completamente, por lo cual se podría recurrir a otros métodos para obtener los datos de cada muestra como sería el promediar las cuatro mediciones o tomar los datos de un solo punto.

4.13. Factibilidad de la implementación de una prueba acústica como método para la determinación de firmeza.

Como primer punto a destacar para el parámetro de la FM y los modelos generales se tiene que el uso de frecuencias de resonancia mostró asociación con el parámetro de calidad, pero el trabajo que realizó el mejor modelo creado no presentó una correcta generalización al igual como ocurrió con los creados por coeficientes de rigidez y con doble predictor de f_2 y GM. En caso contrario cuando se analiza los modelos de redes neuronales hay una mejora en las observaciones correctamente clasificadas por las clases creadas, además de mostrar la capacidad de trabajar con una mayor cantidad de clases, que en el caso del bloque de doble predictor esta logro obtener observaciones positivas en todas sus clases, siendo mayores sus porcentajes positivos en cinco de las seis clases con respecto al siguiente bloque de mejor trabajo que fue para frecuencias.

Se debe considerar que estos datos de firmeza son del cálculo de esfuerzo máximo correspondiente a la segunda cresta que pertenece a la de mayor amplitud en la serie de datos por lo cual se debe realizar este ensayo pero para el primer fallo que se crea en el tejido esto

es para comprobar la asociación entre este y los datos de frecuencias de resonancia y verificar si estos porcentajes son menores al 50% de eficacia por parte de los modelos, así como comprobar que las observaciones correctamente clasificadas tienen la misma tendencia o hay una mejora en los resultados con respecto a los alcanzados en este estudio.

En lo que respecta a lo obtenido por parte de los estudios realizados se tiene que el uso de pruebas acústicas como método para la determinación de la calidad es factible solo que esta requiere de emplear un segundo predictor como es el GM al igual que emplear modelos más robustos como es el caso de las redes neuronales para obtener mejores resultados y generalización, además de contar con una mayor cantidad de muestras para que el modelo pueda entrenarse mejor e incluir una cantidad uniforme de muestras por grado de madurez y por clase elaborada, esto es con el objetivo de evitar segmentación de datos a una clase como ocurrió en los modelos generales con respecto a la clase B.

La asociación entre las frecuencias de resonancia contra los parámetros de calidad mostró la existencia de una tendencia que relacionaba a ambos y que se veía marcado por el incremento en el GM en donde parámetros como CH, GE, P, ESR mostraron un comportamiento esperado en sus máximas tendencias de acuerdo con la hipótesis planteada de que las bajas frecuencias de resonancia estaban sujetas a etapas de GM del tipo E y por consiguiente altas frecuencias con GM del tipo A. Estos resultados brindaron información de una posible aplicación de pruebas acústicas para la determinación de la firmeza y de otra variedad de parámetros.

En la etapa de entrenamiento de los modelos de inteligencia artificial se obtuvo eficacias generales en un rango que comprendía un máximo de 39,9% a un 14,9%, donde los más elevados fueron para ESR, GM y CH. De igual manera los modelos de redes neurales artificiales generados a partir de este bloque de predictores lograron eficacias generales en un rango de 39,5% a un 14,1% donde también los más elevados fueron para ESR, GM y CH. Estos resultados son positivos ya que indican que es posible generar modelos a partir de estas relaciones, sin embargo, alcanzan un valor inferior al 50%. En algunos casos no es posible realizar una generalización, tomando en cuenta la cantidad de muestras empleadas y la disparidad de estas entre clases.

El entrenamiento de modelos a partir de un bloque constituido por CR mostró resultados de eficacia generales en un rango de 41,4% a un 21,8%, donde los más elevados fueron para GM, ESR y CH. Estos porcentajes son cercanos a los alcanzados por el bloque de frecuencias de resonancia. En cuanto a los modelos de redes neuronales sus porcentajes de eficacia generales fueron de 39,2% a un 20,3% donde los más elevados fueron para ESR, GM y FM. En este caso de estudio con CR se encontró que lograron obtener mejores resultados con EST y P, lo cual muestra una posible tendencia a asociar esta clase de predictores con parámetros que involucran aspectos estructurales o de firmeza. Al igual como sucede con el bloque de frecuencias de resonancia, estos resultados muestran una factibilidad de emplear CR para determinación de parámetros como son la firmeza, esfuerzo remanente y fuerza de penetración, ya que se alcanzó una eficacia máxima del 41,4% lo cual minimiza la incertidumbre que poseen los sistemas de gestiones de calidad actual de esta industria, donde solo se muestrean menos del 1% de producto que ingresa.

El ultimo bloque de estudio constituido por la unión de dos predictores logró obtener mejores resultados a los anteriores, con un rango que comprende de 87,2% a un 25%. Los más elevados fueron para GM, ESR y CH, lo cual coloca a el GM como el porcentaje de mejor eficacia de los evaluados incluido para los modelos de redes neuronales, debido a que se obtuvo un rango de 51,3% a un 23,1% entre todos los parámetros. Los más elevados fueron para GM, ESR y CH. Estos resultados de mayor valor muestran un aspecto positivo en cuanto a la factibilidad de la aplicación de pruebas acústicas como modelos no destructivos, ya que, si bien no se alcanzan valores cercanos a un 90% en todos los parámetros, pueden llegar a mejorarse aplicando distribuciones de muestras entre clases iguales, una mayor cantidad de muestras, entre otros.

Otro aspecto positivo de obtener un 87,2% en el modelo Conjunto (árboles agrupados) para GM es que este parámetro es asociable a todas las variables de respuesta evaluadas. Esto se evidencia en el parámetro de contenido de humedad en donde ella presenta cambios con forme hay un desarrollo en el GM, lo cual como se observó al integrar este parámetro como covariable se favoreció la eficacia del modelo. Por su parte naturalmente los productos biológicos sufren cambios en sus tejidos durante el proceso de maduración y estos cambios se ven reflejados en parámetros como FM, P y ER.

El factor de uniformidad muestra que los modelos pueden llegar a asociar los datos a una clase en cada parámetro. La piña no representa una muestra completamente uniforme y por consiguiente la selección de datos de un punto brinda información aproximada. La evaluación de otros métodos de selección de datos para corroborar la efectividad es recomendable.

La aplicación de pruebas acústicas para la determinación de parámetros de calidad es factible, si bien hay parámetros que logran porcentajes de eficacia positivos, hay otros que alcanzan valores que aún requieren ser mejorados para lograr una generalización y una reducción en la segmentación hacia una clase dentro de un parámetro. Hay que considerar que estos resultados fueron obtenidos bajo condiciones controladas de laboratorio y con una cantidad de muestras pequeña. De acuerdo con estas características es factible determinar la firmeza, así como parámetros de contenido de humedad, esfuerzo remanente, grado de madurez, fuerza de penetración y gravedad específica. Se esperaría que al ingresar una cantidad de muestras mayor a estos modelos obtendría mejores resultados de eficacia.

La metodología actual aplicada en industria es menos exacta ya que generalmente se basa en el muestreo de lotes, en donde con una cantidad de muestras que representan menos del 1% determinan un valor promedio considerado como un valor aceptable y con esto generalizan las condiciones que presenta la totalidad del lote. Las muestras de este lote poseen un valor en el rango igual al promedio de más o menos tres desviaciones estándar. Una metodología basada en pruebas no destructivas permitiría la evaluación de la totalidad de producto que ingresa a planta y con los modelos generados en este trabajo se lograría predecir el valor de calidad con un rango de incertidumbre menor a las tres desviaciones estándar actuales aplicadas en una línea de producción.

5. CONCLUSIONES

1. Se diseñó un mecanismo tipo péndulo mecánico que además contó con componentes que redujeron la emisión sónica. Esto permitió probar una metodología para emplear pruebas acústicas como método no destructivo para determinar parámetros de calidad como contenido de humedad, firmeza, esfuerzo remanente, grado de madurez, gravedad específica y fuerza de penetración, por medio de la generación de algoritmos de inteligencia artificial.
2. Se encontró que las bajas frecuencias de resonancia (70; 120; 200 y 560 Hz, de acuerdo con cada frecuencia de resonancia) poseen una relación con las piñas cercanas a 2 800 g, con contenidos de humedad cercanos al 90% y con gravedades específicas cercanas o mayores a uno. Los esfuerzos remanentes próximos a 7 N, fuerzas de penetración cercanas a 0,50 N y firmezas contiguas a $3,5 \times 10^6$ Pa se relacionan principalmente con las altas frecuencias de resonancias (125; 345; 480 y 880 Hz, de acuerdo con cada frecuencia de resonancia).
3. Los modelos generales y redes neuronales generados con frecuencias de resonancia mostraron que la primera frecuencia de resonancia (f_1) es el mejor predictor para determinar el grado de madurez, mientras que la segunda y tercera frecuencia (f_2 y f_3) son mejores para contenido de humedad, fuerza de penetración, firmeza, esfuerzo remanente, grado de madurez y gravedad específica y la cuarta frecuencia de resonancia (f_4) con parámetros de la estructura ya que fue capaz de obtener los mejores resultados con esfuerzo remanente, fuerza de penetración y la firmeza. El coeficiente de rigidez 1 presentó más ocasiones de asociación para determinar el esfuerzo remanente, coeficiente de rigidez 2 para la fuerza de penetración y el coeficiente de rigidez 3 para la firmeza.
4. Al aplicar la validación cruzada con un 30%, las eficacias más altas alcanzadas fueron de 87,2% para el grado de madurez (modelo conjunto, árboles agrupados); 40,5% para el esfuerzo remanente (modelo árbol, árbol grueso); 32,8% para el contenido de humedad (modelo árbol, árbol medio); 32,7% para la firmeza (modelo máquinas de vectores de soporte gaussiano medio) y 30,6% para la gravedad específica (modelo red neuronal media). Se destacan que estos modelos sientan las bases para futuros modelos mejorados a partir de la inclusión de más datos.

5. Los porcentajes de eficacia más altos dados para las redes neuronales del conjunto final son inferiores a los modelos generales y poseen una tendencia a generar resultados correctamente predichos con una uniformidad en cuanto a los porcentajes obtenidos y una marcada capacidad para generalizar en las clases elaboradas.
6. Los ensayos realizados para los modelos con mejor porcentaje de eficacia, pero a una validación cruzada al 5% disminuyó la eficacia de los modelos para la predicción de todos los parámetros respuesta analizados, donde además se redujo la capacidad de generalización de los modelos, reafirmando con ello la selección del 30% de validación cruzada.
7. Los factores de uniformidad mostraron que existe variación espacial en todos los parámetros de calidad determinados excepto para grado de madurez y gravedad específica, esto implica que para una misma muestra se puede obtener hasta 4 clasificaciones diferentes, pero esto no influye en el modelo, pero sí implica una limitación para la clasificación general de una muestra. Además, se determina que el factor de uniformidad predominante es el 2 con un valor cercano al 50% de las muestras; esta variedad de clases que puede mostrar una misma piña podría resultar en una limitante para la capacidad de predicción de los modelos, ya que no se obtiene un dato que sea representativo para la muestra.
8. Las pruebas acústicas son factibles para determinar la firmeza siempre que la frecuencia de resonancia se combine con otro predictor como el grado de madurez y empleando redes neuronales artificiales para obtener una generalización en cuanto a las clases elaboradas y una mejora en los porcentajes de observaciones correctamente clasificadas.
9. Las pruebas acústicas son factibles para determinar los parámetros de calidad como contenido de humedad, esfuerzo remanente, grado de madurez, gravedad específica y fuerza de penetración, pero es requerido mejorar en aspectos de generalización, utilización de una mayor cantidad de muestras, aplicar una segmentación de muestras por clase uniforme para obtener porcentajes de eficacia mayores, de tal manera que se logren optimizar para su aplicación en industria.

10. La implementación de las pruebas no destructivas disminuiría la incertidumbre y aumentaría la exactitud con respecto a las metodologías actuales ya que la cantidad posible de producto a analizar aumenta considerablemente.

6. RECOMENDACIONES

1. Se recomienda el uso de un segundo método de cálculo para la firmeza para usar como conjunto de datos para el entrenamiento de los algoritmos para comprobar la asociación entre las pruebas acústicas y el parámetro de calidad.
2. Es recomendable la utilización de un equipo de percusión o uno que no sea de percusión que emita el mínimo o nulo ruido al momento de realizar los golpes, ya que la vibración ocasionada por los elementos del equipo empleado generó registros acústicos que se presentaron a partir de los 1 000 Hz en adelante lo que impidió el uso de alguna frecuencia a partir de este rango, pero lo cual no representó una limitante. Además, es importante emplear un espacio con aislamiento de sonido y la utilización de un micrófono con mejor sensibilidad que $-34 \text{ dB} \pm 2 \text{ dB}$, debido a que se reduciría el ruido y obtendría a cambio un sonido con mejor calidad que se reflejaría en resultados acordes a la información suministrada.
3. Para mejorar los resultados se recomienda el uso de una cantidad de muestras equilibrada por grado de madurez sumado a una cantidad mayor de muestra en general. Esto puede ocasionar la posibilidad de que la partición del comportamiento de las tendencias de las frecuencias de resonancia contra los parámetros de calidad no suceda o se ajuste a un caso, logrando una mayor uniformidad en las clasificaciones y predicciones de los datos.
4. De acuerdo con los factores de uniformidad se recomienda contemplar emplear un solo punto de análisis por muestra, promediar los cuatro análisis u otra técnica que permita mejorar la clasificación llevada a cabo.
5. La investigación con un mayor número de puntos de análisis podría arrojar mejores resultados además que si se suman la cantidad de piñas para estudiar, se puede promediar los datos de los parámetros para estos puntos consiguiendo con ello un valor que caracterice específicamente una piña. Eventualmente también se puede emplear únicamente un dato por muestra, pero se cae en el error de asumir que este representa a toda la piña.
6. Se propone la técnica para analizar la detección de problemas fisiológicos como la translucidez u otras afectaciones internas, ya que fue un grupo de estudio que no se

consideró para la investigación ya que las muestras empleadas no cumplían con estas condiciones.

7. BIBLIOGRAFIA

- Aboonajmi, M., Jahangiri, M., & Hassan-Beygi, S. R. (2015). A Review on Application of Acoustic Analysis in Quality Evaluation of Agro-food Products. *Journal of Food Processing and Preservation*, 39(6), 3175–3188. <https://doi.org/10.1111/jfpp.12444>
- AprendeIA. (2018). Introducción a Sobreajuste y Subajuste para Machine Learning. Retrieved August 28, 2020, from AprendeIA website: <https://aprendeia.com/sobreajuste-y-subajuste-en-machine-learning/>
- Ariza-López, F. J., Rodríguez-Avi, J., & Alba-Fernández, V. (2018). Control estricto de matrices de confusión por medio de distribuciones multinomiales. *GeoFocus Revista Internacional de Ciencia y Tecnología de La Información Geográfica*, 215–226. <https://doi.org/10.21138/GF.591>
- Ayala Aponte, A. A., Leiton Ramirez, Y., & Serna Cock, L. (2017). Cambios en propiedades mecánicas durante la deshidratación osmótica de pitahaya amarilla. *Biotecnología En El Sector Agropecuario y Agroindustrial*, 15(2), 39. [https://doi.org/10.18684/BSAA\(15\)39-48](https://doi.org/10.18684/BSAA(15)39-48)
- Berk, Z. (2018a). Physical properties of food materials. In *Food Process Engineering and Technology* (p. 24). <https://doi.org/10.1016/B978-0-12-812018-7.00001-4>
- Berk, Z. (2018b). Physical properties of food materials. In *Food Process Engineering and Technology* (pp. 1–29). <https://doi.org/10.1016/B978-0-12-812018-7.00001-4>
- Bi, X., Hemar, Y., Balaban, M. O., & Liao, X. (2015). The effect of ultrasound on particle size, color, viscosity and polyphenol oxidase activity of diluted avocado puree. *Ultrasonics Sonochemistry*, 27, 567–575. <https://doi.org/10.1016/j.ultsonch.2015.04.011>
- Blancard, D. (2011). *Enfermedades del tomate. Identificar, conocer, controlar* (Mundi-Prensa, Ed.). Retrieved from <https://books.google.co.cr/books?id=3tSzHjXHHJ4C&pg=PA64&dq=causas+de+los+desordenes+geneticos+en+plantas&hl=es&sa=X&ved=2ahUKEwiHqIu2l4DrAhWPmOAKHWvJCKoQ6AEwAHoECAyQA#v=onepage&q&f=true>

- Blasco Espinosa, J. M. (2014). *UF0013: Recolección, transporte, almacenamiento y acondicionamiento de la fruta* (p. 41). p. 41. Retrieved from <https://books.google.co.cr/books?id=S15WDwAAQBAJ&pg=PA41&dq=madurez+comercial&hl=es&sa=X&ved=2ahUKEwja6Yinp-7qAhWPhOAKHTMvBJkQ6AEwAXoECAIQAg#v=onepage&q=madurez+comercial&f=true>
- Brenes Gamboa, S. (2007). Caracterización vegetativa y productiva del cultivo MD-2 de piña (*Ananas Comosus*) bajo las condiciones climáticas de turrialba. *Inter Sedes*, 6, 27–34.
- Cabrerizo, A., & Dulce, M. (2019). Física y Química. Retrieved June 2, 2020, from Editex website: https://books.google.co.cr/books?id=s8mUDwAAQBAJ&printsec=frontcover&hl=es&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false
- Caladcad, J. A., Cabahug, S., Catamco, M. R., Villaceran, P. E., Cosgafa, L., Cabizares, K. N., ... Piedad, E. J. (2020). Determining Philippine coconut maturity level using machine learning algorithms based on acoustic signal. *Computers and Electronics in Agriculture*, 172, 105327. <https://doi.org/10.1016/j.compag.2020.105327>
- CANAPEP. (2016). Historia. Retrieved April 10, 2020, from CANAPEP website: <https://canapep.com/historia/>
- CANAPEP. (2020). Estadísticas. Retrieved from CANAPEP website: <https://canapep.com/estadisticas/>
- CANNA. (2019). Los desórdenes fisiológicos en las plantas. Retrieved July 26, 2020, from CANNA website: http://www.canna.es/los_desordenes_fisiologicos_en_las_plantas
- Cazar Villacís, I. M. (2016). *Análisis físico-químico para la determinación de la calidad de las frutas* (Pontificia Universidad Católica del Ecuador). Retrieved from [http://repositorio.puce.edu.ec/bitstream/handle/22000/11453/Análisis físico-químico para la determinación de la calidad de las frutas.pdf?sequence=1](http://repositorio.puce.edu.ec/bitstream/handle/22000/11453/Análisis_físico-químico_para_la_determinación_de_la_calidad_de_las_frutas.pdf?sequence=1)
- Cerdas Araya, M. del M., & Montero Calderón, M. E. (2005). *Guías técnicas del manejo de*

poscosecha de la piña en el mercado fresco. San José, Costa Rica: MAG.

Chinchilla, E. (2004). *Estudio del proceso de trabajo y operaciones, perfil de riesgos y exigencias laborales en el cultivo y empaque* (p. 12). p. 12. Retrieved from [https://www.cso.go.cr/documentos_relevantes/tecnicos/series/04_Serie tecnica No. 4.pdf](https://www.cso.go.cr/documentos_relevantes/tecnicos/series/04_Serie_tecnica_No.4.pdf)

Corn, P., Ross, E., & Khim, J. (n.d.). Discrete Fourier Transform. Retrieved August 19, 2020, from Brilliant.org website: <https://brilliant.org/wiki/discrete-fourier-transform/>

Cortes Osorio, J. A., Chaves Osorio, J. A., & Medoza, J. A. (2014). Diseño e implementación de una aplicación para el estudio del péndulo simple mediante visión por computador. *Scientia et Technica*, 19, 314–321. Retrieved from <https://core.ac.uk/download/pdf/299841483.pdf>

Diezma, B., Correa, E. C., & Ruiz-Altisen, M. (2010). Técnicas no destructivas para la determinación de la calidad de frutas: sistemas acústicos y sensores químicos. *Frutera*, 3, 123–138.

Duarte Agudelo, C. A., & Niño Vicentes, J. R. (2004). *Introducción a la mecánica de fluidos* (3rd ed.). Colombia: Universidad Nacional de Colombia.

EPSEM. (2021). Secador de sólidos. Retrieved December 10, 2021, from Universitat Politècnica de Catalunya website: http://epsem.upc.edu/~assecadordesolids/castella/fonaments_materia.html

FAO. (n.d.-a). Daños sufridos por los productos frescos después de la cosecha. Retrieved July 28, 2020, from FAO website: <http://www.fao.org/3/t0073s/T0073S03.htm>

FAO. (n.d.-b). Operaciones de cosecha y campo. Retrieved March 5, 2020, from FAO website: <http://www.fao.org/3/x5055s/x5055s03.htm#:~:text=La madurez fisiológica se refiere,el máximo crecimiento y maduración.>

Herrera, C., Bolaños, N., & Lutz, G. (2003). *Química de Alimentos* (1st ed.). San José, Costa Rica: Editorial de la Universidad de Costa Rica.

Hurrell, J., Bazzano, D., & Delucchi, G. (2005). *Biota Rioplatense X. Mocotiledóneas*

herbáceas nativas y exóticas.

- Ichi.pro. (2020). Overfitting y Underfitting en Machine Learning. Retrieved August 28, 2020, from Ichi.pro website: <https://ichi.pro/es/overfitting-y-underfitting-en-machine-learning-237674234385205>
- Jaramillo Sánchez, J. A. (2004). *Física*. España: MAD.
- Kim, J.-Y., Lee, J.-S., Kwon, T.-R., Lee, S.-I., Kim, J.-A., Lee, G.-M., ... Jeong, M.-J. (2015). Sound waves delay tomato fruit ripening by negatively regulating ethylene biosynthesis and signaling genes. *Postharvest Biology and Technology*, *110*, 43–50. <https://doi.org/10.1016/j.postharvbio.2015.07.015>
- López Gumucio, R. (2005). La calidad total en la empresa moderna. *PERSPECTIVAS*, *8*, 67–81. Retrieved from <https://www.redalyc.org/articulo.oa?id=4259/425942412006>
- MAG. (2007). *Cadena agroalimentaria del cultivo de piña en distrito de Chires de Puriscal* (p. 13,14). p. 13,14. Retrieved from <http://www.mag.go.cr/bibliotecavirtual/E70-10277.pdf>
- MAG. (2019). Manual de buenas prácticas agrícolas para la producción sostenible del cultivo de la piña (Ananas Comosus). Retrieved July 27, 2020, from <http://www.mag.go.cr/bibliotecavirtual/F01-11044.PDF>
- Mandal, A. (2019). ¿Cuáles son desordenes genéticos? Retrieved July 27, 2020, from New Medical website: [https://www.news-medical.net/health/What-are-Genetic-Disorders-\(Spanish\).aspx#:~:text=Los desordenes genéticos son condiciones,las otras condiciones de salud.](https://www.news-medical.net/health/What-are-Genetic-Disorders-(Spanish).aspx#:~:text=Los desordenes genéticos son condiciones,las otras condiciones de salud.)
- Mao, J., Yu, Y., Rao, X., & Wang, J. (2016). Firmness prediction and modeling by optimizing acoustic device for watermelons. *Journal of Food Engineering*, *168*, 1–6. <https://doi.org/10.1016/j.jfoodeng.2015.07.009>
- Martínez-González, M. E., Balois Morales, R., Alia-Tejacal, I., Cortes-Cruz, M. A., Palomino-Hermosillo, Y. A., & López-Gúzman, G. G. (2017). Postcosecha de frutos: maduración y cambios bioquímicos. *Revista Mexicana de Ciencias Agrícolas*, (19),

4075. <https://doi.org/10.29312/remexca.v0i19.674>

- Martínez Gonzáles, M. E., Balois Morales, R., Alia Tejacal, I., Cortes Cruz, M. A., Palomino Hermosillo, Y. A., & López Gúzman, G. G. (2017). Poscosecha de frutos: maduración y cambios bioquímicos. *Revista Mexicana de Ciencias Agrícolas*, 19. Retrieved from <https://www.redalyc.org/articulo.oa?id=263153823018>
- Martinez, J. (2020). ¿Qué es Underfitting y cómo lo detectamos? Retrieved August 28, 2020, from DataSmarts website: <https://datasmarts.net/es/que-es-underfitting-y-como-lo-detectamos/>
- Martínez, S. (2020). Las diferencias entre Deep Learning y Machine Learning. Retrieved April 29, 2020, from Jasmin website: <https://www.jasminsoftware.es/blog/deep-learning/>
- Mathworks. (2020). ¿Qué es el reconocimiento de objetos? Retrieved April 29, 2020, from MathWorks website: <https://es.mathworks.com/solutions/image-video-processing/object-recognition.html>
- MathWorks. (n.d.-a). La transformada de Fourier. Retrieved January 20, 2020, from MathWorks website: <https://la.mathworks.com/help/images/fourier-transform.html>
- MathWorks. (n.d.-b). Redes neuronales. Retrieved January 20, 2020, from MathWorks website: <https://la.mathworks.com/discovery/neural-network.html>
- Microsoft. (2021). Algoritmos de aprendizaje automático. Retrieved December 12, 2021, from Azure website: <https://azure.microsoft.com/es-es/overview/machine-learning-algorithms/#overview>
- Minitab. (n.d.). Validación cruzada en la regresión PLS. Retrieved January 20, 2020, from Minitab website: <https://support.minitab.com/es-mx/minitab/19/help-and-how-to/modeling-statistics/regression/supporting-topics/partial-least-squares-regression/cross-validation-in-pls-regression/>
- Minitab. (2019). ¿Qué son variables de respuesta y variables predictoras? Retrieved from <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/modeling->

statistics/regression/supporting-topics/basics/what-are-response-and-predictor-variables/

- Mizrach, A., Flitsanov, U., El-Batsri, R., & Degani, C. (1998). Determination of avocado maturity by ultrasonic attenuation measurements. *Scientia Horticulturae*, 80(3–4), 173–180. [https://doi.org/10.1016/S0304-4238\(98\)00243-X](https://doi.org/10.1016/S0304-4238(98)00243-X)
- Montero-Calderón, M., Rojas-Graü, M. A., & Martín-Belloso, O. (2010). Mechanical and chemical properties of Gold cultivar pineapple flesh (*Ananas comosus*). *European Food Research and Technology*, 230(4), 675–686. <https://doi.org/10.1007/s00217-009-1207-9>
- Montero Calderón, M., & Cerdas Araya, M. del M. (2005). *Guias tecnicas del manejo poscosecha para el mercado fresco*. San José, Costa Rica: MAG.
- Morales-Abarca, L. (2018). *PRODUCCION Y RENDIMIENTO DEL CULTIVO DE LA PIÑA EN COSTA RICA. 4*.
- Morales, M., Hernández, M. S., Cabezas, M., Barrera, J., & Martinez, O. (2001). Caracterización de la maduración del fruto de piña nativa (*Ananas comosus* L. Merrill) CV. India. *Agronomia Colombiana*, 18, 7–13. Retrieved from <http://hdl.handle.net/11348/4152>
- Morales, M., Hernández, M. S., Cabezas, M., Barrera, J., & Orlando, M. (n.d.). Piña nativa (*Ananas comosus* L, Merr)c.v. India. Retrieved October 14, 2021, from Agronet website: <http://bibliotecadigital.agronet.gov.co/bitstream/11348/4146/1/Aspectos generales de la pina.pdf>
- Morrison, D. S., & Abeyratne, U. R. (2014). Ultrasonic technique for non-destructive quality evaluation of oranges. *Journal of Food Engineering*, 141, 107–112. <https://doi.org/10.1016/j.jfoodeng.2014.05.018>
- Navarro Ugena, C., & Pérez Castellanos, J. L. (2015). *Métodos generales: Análisis matricial* (p. 55). p. 55. Retrieved from http://ocw.uc3m.es/mecanica-de-medios-continuos-y-teoria-de-estructuras/ingenieria-estructural/material-de-clase-1/apuntes/Capitulo_8.-Analisis_matricial_de_estructuras_reticuladas.pdf

- Núñez Flores, M. I. (2007). Las variables: Estructura y función en la hipótesis. *Investigación Educativa, 11*, 163–179.
- Nutricioni. (2019). ¿Qué es la acidez titulable? Un interesante concepto químico. Retrieved December 11, 2021, from <https://nutricioni.com/que-es-la-acidez-titulable-un-interesante-concepto-quimico/>
- Oblaré Torres, J. L. (2020). *UF0008: Instalaciones agrarias, su acondicionamiento, limpieza y desinfección*. Retrieved from <https://books.google.co.cr/books?id=AiXKDwAAQBAJ&pg=PT92&dq=senescencia+definicion&hl=es&sa=X&ved=2ahUKEwi919nQo-7qAhVKT98KHS29DigQ6AEwAHoECAUQA#v=onepage&q=senescencia+definicion&f=false>
- Palmer, C. I., & Bibb, S. F. (2003). *Matemáticas prácticas* (2nd ed.; Editorial Reverté, Ed.).
- Pathaveerat, S., Terdwongworakul, A., & Phaungsombut, A. (2008a). Multivariate data analysis for classification of pineapple maturity. *Journal of Food Engineering, 89*(2), 112–118. <https://doi.org/10.1016/j.jfoodeng.2008.04.012>
- Pathaveerat, S., Terdwongworakul, A., & Phaungsombut, A. (2008b). Multivariate data analysis for classification of pineapple maturity. *Journal of Food Engineering, 89*(2), 112–118. <https://doi.org/10.1016/j.jfoodeng.2008.04.012>
- Paull, R. E., & Chen ChingCheng, C. C. (2003). Postharvest physiology, handling and storage of pineapple. In *The pineapple: botany, production and uses* (pp. 253–279). <https://doi.org/10.1079/9780851995038.0253>
- Phywe. (2021). Oscilaciones pendulares. Retrieved November 29, 2021, from Phywe website: https://www.phywe.com/es/oscilaciones-pendulares-pendulo-g-variable_10892_11925/
- Procomer. (2020). Industria alimentaria intensifica la utilización de inteligencia artificial. Retrieved December 12, 2021, from Procomer website: https://www.procomer.com/alertas_comerciales/exportador-alerta/industria-

alimentaria-intensifica-la-utilizacion-de-inteligencia-artificial/

- Rojas Jiménez, V. (2008). *Efecto de la frecuencia de fertilización en el rendimiento de semilleros de piña (Ananas comusus)(L.) Merr. Híbrido Venecia Gold, en Finca Agrícola Industrial San Cayetano S.A., Horquetas, Sarapiquí*. Instituto Tecnológico de Costa Rica.
- SIAP. (2018). Piña, reina de las frutas tropicales. Retrieved July 24, 2020, from Gobierno de México website: <https://www.gob.mx/siap/articulos/pina-reina-de-las-frutas-tropicales?idiom=es>
- Soto Parra, J. M., Piña Ramiírez, F. J., Sánchez Chávez, E., Pérez Leal, R., & Basurto Sotelo, M. (2016). Fertirrigación con macronutrientes en manzano “Golden Delicious”: Impacto en rendimiento y calidad de fruto. *Nova Scientia*, 8, 163. Retrieved from <http://novascientia.delasalle.edu.mx/ojs/index.php/Nova/article/view/414/268>
- Surya Prabha, D., & Satheesh Kumar, J. (2015). Assessment of banana fruit maturity by image processing technique. *Journal of Food Science and Technology*, 52(3), 1316–1327. <https://doi.org/10.1007/s13197-013-1188-3>
- Tharwat, A. (2021). Classification assessment methods. *Applied Computing and Informatics*, 17(1), 168–192. <https://doi.org/10.1016/j.aci.2018.08.003>
- Ullah, H., Chaiwong, S., & Saengrayap, R. (2018). Application of Image Analysis for Maturity Classification of ‘Phulae’ Pineapple. *The International Conference on Foot and Applied Bioscience*, 8.
- UNCTAD. (2016). *Piña* (p. 3). p. 3. Retrieved from https://unctad.org/es/PublicationsLibrary/INFOCOMM_cp09_Pineapple_es.pdf
- Vallalta, J. F. (2021). Aprendizaje supervisado y no supervisado. Retrieved December 12, 2021, from Health Data Miner website: <https://healthdataminer.com/data-mining/aprendizaje-supervisado-y-no-supervisado/>
- Villegas, O., Vargas, F., Pérez, J., García, R., Porras, S., Meneses, D., ... Alfaro, D. (2007).

Caracterización y Plan acción para el desarrollo de la agrocadena de Raíces y Tubérculos Tropicales en la región Huetar Norte. Retrieved January 21, 2020, from MAG website: <http://www.mag.go.cr/bibliotecavirtual/ac-pina-rhn-2007.pdf>

Vindas, M. (2020). Investigación utiliza las TICs para mejorar la producción piñera. Retrieved December 12, 2021, from UCR website: <https://vinv.ucr.ac.cr/es/noticias/investigacion-utiliza-las-tics-para-mejorar-la-produccion-pinera>

Wesstein, E. (2020). Fast Fourier Transform. Retrieved August 19, 2020, from MathWorld website: <https://mathworld.wolfram.com/FastFourierTransform.html>

8. ANEXOS

Anexo 1. Lista de algoritmos empleados para la generación de modelos predictivos.

Número	Algoritmos	Número	Algoritmos
1	Árboles de decisión	5	Clasificadores de vecinos más cercanos
1,1	Árbol (Árbol fino)	5,1	K vecino más cercano fino
1,2	Árbol (Árbol Medio)	5,2	K vecino más cercano Medio
1,3	Árbol (Árbol grueso)	5,3	K vecino más cercano grueso
2	Análisis discriminante	5,4	K vecino más cercano Coseno
2,1	Discriminante lineal (Discriminante Lineal)	5,5	K vecino más cercano cúbico
2,2	Discriminante cuadrático (Discriminante Cuadrático)	5,6	K vecino más cercano ponderado
3	Clasificadores de bayas ingenuas	6	Clasificadores de conjunto
3,1	Bayas ingenuas (Bayas ingenuas gaussianas)	6,1	Conjunto (Árboles reforzados)
3,2	Bayas ingenuas (Bayas ingenuas kernel)	6,2	Conjunto (Árboles en bolsa)
4	Máquinas de vectores de soporte	6,3	Conjunto (Discriminante subespacial)
4,1	Máquinas de vectores de soporte lineal	6,4	Conjunto (Subespacio KNN)
4,2	Máquinas de vectores de soporte cuadrática	6,5	Conjunto (RUSBoosted árboles)
4,3	Máquinas de vectores de soporte SVM cúbica	7	Clasificadores de redes neuronales
4,4	Máquinas de vectores de soporte gaussiana fina	7,1	Red neuronal (Red neuronal estrecha)
4,5	Máquinas de vectores de soporte gaussiana media	7,2	Red neuronal (Red neuronal media)
4,6	Máquinas de vectores de soporte gaussiana gruesa	7,3	Red neuronal (Red neuronal ancha)
		7,4	Red neuronal (Red neuronal de dos capas)
		7,5	Red neuronal (Red neuronal de tres capas)

Anexo 2. Código para filtración de audio por Transformada Rápida de Fourier.

```
clc
[msj,Fs]=audioread('audio.wav');
msj=msj/max(abs(msj));
sound(msj,Fs)

n=length(msj);
t=n/Fs;
Ts=1/Fs;
tiempo=[0:Ts:(t-Ts)];

figure(1)
plot(tiempo,msj,'b')
title(['Señal de audio capturada en el dominio del tiempo']);
xlabel('Tiempo (s)')
ylabel('Amplitud (dB)')
grid('on')

transformada=abs(fft(msj));
L=length(transformada);
espectro=transformada(1:L/2);
espectro=espectro/max(espectro);
frecuencias=Fs*(1:(L/2))/L;
figure(2)
plot(frecuencias,espectro,'r')
title(['Espectro de la señal capturada (FFT)']);
xlabel('Frecuencia (Hz)')
ylabel('Amplitud (dB)')
grid('on')
ylim([0 max(espectro)])
xlim([0 1500])
```

Anexo 3. Distribución de muestras según grado de madurez para las categorías de las masas para piñas.

Categoría	Rango (g)	GM (número de muestras)			
		B	C	D	E
A	0-1624	0	28	12	8
B	1624-1683	8	8	12	12
C	1683-1742	4	8	16	16
D	1742-1801	0	28	12	8
E	1801-1860	8	20	8	20
F	1860-1919	12	8	16	28
G	1919-1978	4	16	4	28
H	1978-2008	12	0	4	28
I	2008-2037	12	0	8	16
J	2037-2096	12	16	28	8
K	2096-2155	20	24	0	24
L	2155-2214	8	8	0	20
M	2214-2273	8	4	12	24
N	2273-2391	8	0	24	20
O	2391-2745	0	8	20	16

Anexo 4. Distribución de muestras según grado de madurez para las categorías del contenido de humedad para piñas.

Categoría	Rango (%)	GM (número de muestras)			
		B	C	D	E
A	0-83,56	28	53	48	57
B	83,56-84,56	25	37	35	78
C	84,56-85,56	30	37	56	73
D	85,56-99,56	33	49	37	68

Anexo 5. Distribución de muestras según grado de madurez para las categorías de la gravedad específica para piñas.

Categoría	Rango	GM (número de muestras)			
		B	C	D	E
A	0-0,881	8	48	36	76
B	0,881-0,887	12	32	44	40
C	0,887-0,893	36	28	28	52
D	0,893-0,899	28	28	24	48
E	0,899-0,911	28	28	32	36
F	0,911-0,923	4	12	12	24

Anexo 6. Distribución de muestras según grado de madurez para las categorías de la fuerza de penetración para piñas.

Categoría	Rango (N)	GM (número de muestras)			
		B	C	D	E
A	0-0,06	12	25	42	76
B	0,06-0,09	8	22	15	66
C	0,09-0,12	11	31	23	52
D	0,12-0,15	16	21	28	35
E	0,15-0,21	30	38	37	35
F	0,21-0,27	22	25	24	8
G	0,27-0,47	17	14	7	4

Anexo 7. Distribución de muestras según grado de madurez para las categorías de la firmeza para piñas.

Categoría	Rango (Pa)	GM (número de muestras)			
		B	C	D	E
A	0-272546,64	10	28	19	56
B	272546,64-401802,64	24	51	45	100
C	401802,64-531058,64	18	23	26	63
D	531058,64-789570,64	23	26	38	37
E	789570,64-1177338,64	23	33	38	13
F	1177338,64-3116178,64	18	15	10	7

Anexo 8. Distribución de muestras según grado de madurez para las categorías del esfuerzo remanente para piñas.

Categoría	Rango (N)	GM (número de muestras)			
		B	C	D	E
A	0-1,2	19	42	44	77
B	1,2-1,7	43	64	55	132
C	1,7-2,2	29	42	50	37
D	2,2-7,2	25	28	27	30

Anexo 9. Modelo Ensemble (Árboles en bolsa) para grado de madurez a 30% validación cruzada.

```
function [trainedClassifier, validationAccuracy] =
trainClassifier(trainingData)

% [trainedClassifier, validationAccuracy] = trainClassifier(trainingData)
% Returns a trained classifier and its accuracy. This code recreates the
% classification model trained in Classification Learner app. Use the
% generated code to automate training the same model with new data, or to
% learn how to programmatically train models.
%
% Input:
%   trainingData: A table containing the same predictor and response
%   columns as those imported into the app.
%
% Output:
%   trainedClassifier: A struct containing the trained classifier. The
%   struct contains various fields with information about the trained
%   classifier.
```

```

%
%   trainedClassifier.predictFcn: A function to make predictions on
new
%   data.
%
%   validationAccuracy: A double containing the accuracy in percent.
In
%   the app, the History list displays this overall accuracy score
for
%   each model.
%
% Use the code to train the model with new data. To retrain your
% classifier, call the function from the command line with your original
% data or new data as the input argument trainingData.
%
% For example, to retrain a classifier trained with the original data set
% T, enter:
%   [trainedClassifier, validationAccuracy] = trainClassifier(T)
%
% To make predictions with the returned 'trainedClassifier' on new data
T2,
% use
%   yfit = trainedClassifier.predictFcn(T2)
%
% T2 must be a table containing at least the same predictor columns as
used
% during training. For details, enter:

```

```

% trainedClassifier.HowToPredict

% Auto-generated by MATLAB on 05-Nov-2021 22:58:41

% Extract predictors and response

% This code processes the data into the right shape for training the
% model.

inputTable = trainingData;

predictorNames = {'M', 'F3Hz'};

predictors = inputTable(:, predictorNames);

response = inputTable.ClasifGrado;

isCategoricalPredictor = [false, false];

% Train a classifier

% This code specifies all the classifier options and trains the
classifier.

template = templateTree(...
    'MaxNumSplits', 743);

classificationEnsemble = fitcensemble(...
    predictors, ...
    response, ...
    'Method', 'Bag', ...
    'NumLearningCycles', 30, ...
    'Learners', template, ...
    'ClassNames', categorical({'B'; 'C'; 'D'; 'E'}));

% Create the result struct with predict function

```

```

predictorExtractionFcn = @(t) t(:, predictorNames);

ensemblePredictFcn = @(x) predict(classificationEnsemble, x);

trainedClassifier.predictFcn = @(x)
ensemblePredictFcn(predictorExtractionFcn(x));

% Add additional fields to the result struct

trainedClassifier.RequiredVariables = {'F3Hz', 'M'};

trainedClassifier.ClassificationEnsemble = classificationEnsemble;

trainedClassifier.About = 'This struct is a trained model exported from
Classification Learner R2021a.';

trainedClassifier.HowToPredict = sprintf('To make predictions on a new
table, T, use: \n yfit = c.predictFcn(T) \nreplacing ''c'' with the name
of the variable that is this struct, e.g. ''trainedModel''. \n \nThe
table, T, must contain the variables returned by: \n c.RequiredVariables
\nVariable formats (e.g. matrix/vector, datatype) must match the original
training data. \nAdditional variables are ignored. \n \nFor more
information, see <a href="matlab:helpview(fullfile(docroot, ''stats'',
''stats.map''), ''appclassification_exportmodeltoworkspace'')">How to
predict using an exported model</a>.');

% Extract predictors and response

% This code processes the data into the right shape for training the
% model.

inputTable = trainingData;

predictorNames = {'M', 'F3Hz'};

predictors = inputTable(:, predictorNames);

response = inputTable.ClasifGrado;

isCategoricalPredictor = [false, false];

```

```

% Perform cross-validation

partitionedModel = crossval(trainedClassifier.ClassificationEnsemble,
'KFold', 30);

% Compute validation predictions

[validationPredictions, validationScores] =
kfoldPredict(partitionedModel);

% Compute validation accuracy

validationAccuracy = 1 - kfoldLoss(partitionedModel, 'LossFun',
'ClassifError');

```

Anexo 10. Modelo Tree (Árbol grueso) para esfuerzo remanente a 30% validación cruzada.

```

function [trainedClassifier, validationAccuracy] =
trainClassifier(trainingData)

% [trainedClassifier, validationAccuracy] = trainClassifier(trainingData)
% Returns a trained classifier and its accuracy. This code recreates the
% classification model trained in Classification Learner app. Use the
% generated code to automate training the same model with new data, or to
% learn how to programmatically train models.
%
% Input:
%   trainingData: A table containing the same predictor and response
%   columns as those imported into the app.
%
% Output:
%   trainedClassifier: A struct containing the trained classifier. The

```

```

%     struct contains various fields with information about the trained
%     classifier.
%
%     trainedClassifier.predictFcn: A function to make predictions on
new
%     data.
%
%     validationAccuracy: A double containing the accuracy in percent.
In
%     the app, the History list displays this overall accuracy score
for
%     each model.
%
% Use the code to train the model with new data. To retrain your
% classifier, call the function from the command line with your original
% data or new data as the input argument trainingData.
%
% For example, to retrain a classifier trained with the original data set
% T, enter:
%
% [trainedClassifier, validationAccuracy] = trainClassifier(T)
%
% To make predictions with the returned 'trainedClassifier' on new data
T2,
% use
%
% yfit = trainedClassifier.predictFcn(T2)
%

```

```

% T2 must be a table containing at least the same predictor columns as
used

% during training. For details, enter:

%   trainedClassifier.HowToPredict

% Auto-generated by MATLAB on 12-Nov-2021 19:45:27

% Extract predictors and response

% This code processes the data into the right shape for training the
% model.

inputTable = trainingData;

predictorNames = {'L1CoefRig1'};

predictors = inputTable(:, predictorNames);

response = inputTable.ClasifEsR;

isCategoricalPredictor = [false];

% Train a classifier

% This code specifies all the classifier options and trains the
classifier.

classificationTree = fitctree(...
    predictors, ...
    response, ...
    'SplitCriterion', 'gdi', ...
    'MaxNumSplits', 100, ...
    'Surrogate', 'off', ...
    'ClassNames', categorical({'A'; 'B'; 'C'; 'D'}));

```

```

% Create the result struct with predict function

predictorExtractionFcn = @(t) t(:, predictorNames);

treePredictFcn = @(x) predict(classificationTree, x);

trainedClassifier.predictFcn = @(x)
treePredictFcn(predictorExtractionFcn(x));

% Add additional fields to the result struct

trainedClassifier.RequiredVariables = {'L1CoefRig1'};

trainedClassifier.ClassificationTree = classificationTree;

trainedClassifier.About = 'This struct is a trained model exported from
Classification Learner R2021a.';

trainedClassifier.HowToPredict = sprintf('To make predictions on a new
table, T, use: \n yfit = c.predictFcn(T) \nreplacing ''c'' with the name
of the variable that is this struct, e.g. ''trainedModel''. \n \nThe
table, T, must contain the variables returned by: \n c.RequiredVariables
\nVariable formats (e.g. matrix/vector, datatype) must match the original
training data. \nAdditional variables are ignored. \n \nFor more
information, see <a href="matlab:helpview(fullfile(docroot, ''stats'',
''stats.map''), ''appclassification_exportmodeltoworkspace'')">How to
predict using an exported model</a>.');

% Extract predictors and response

% This code processes the data into the right shape for training the
% model.

inputTable = trainingData;

predictorNames = {'L1CoefRig1'};

predictors = inputTable(:, predictorNames);

response = inputTable.ClasifEsR;

```



```

isCategoricalPredictor = [false];

% Perform cross-validation

partitionedModel = crossval(trainedClassifier.ClassificationTree,
'KFold', 30);

% Compute validation predictions

[validationPredictions, validationScores] =
kfoldPredict(partitionedModel);

% Compute validation accuracy

validationAccuracy = 1 - kfoldLoss(partitionedModel, 'LossFun',
'ClassifError');

```

Anexo 11. Modelo Tree (Árbol mediano) para contenido de humedad a 30% validación cruzada.

```

function [trainedClassifier, validationAccuracy] =
trainClassifier(trainingData)

% [trainedClassifier, validationAccuracy] = trainClassifier(trainingData)
% Returns a trained classifier and its accuracy. This code recreates the
% classification model trained in Classification Learner app. Use the
% generated code to automate training the same model with new data, or to
% learn how to programmatically train models.
%
% Input:
%     trainingData: A table containing the same predictor and response
%     columns as those imported into the app.

```

```

%
% Output:
%   trainedClassifier: A struct containing the trained classifier. The
%   struct contains various fields with information about the trained
%   classifier.
%
%   trainedClassifier.predictFcn: A function to make predictions on
new
%   data.
%
%   validationAccuracy: A double containing the accuracy in percent.
In
%   the app, the History list displays this overall accuracy score
for
%   each model.
%
% Use the code to train the model with new data. To retrain your
% classifier, call the function from the command line with your original
% data or new data as the input argument trainingData.
%
% For example, to retrain a classifier trained with the original data set
% T, enter:
%   [trainedClassifier, validationAccuracy] = trainClassifier(T)
%
% To make predictions with the returned 'trainedClassifier' on new data
T2,
% use

```

```

% yfit = trainedClassifier.predictFcn(T2)

%

% T2 must be a table containing at least the same predictor columns as
used

% during training. For details, enter:

% trainedClassifier.HowToPredict

% Auto-generated by MATLAB on 06-Nov-2021 14:08:04

% Extract predictors and response

% This code processes the data into the right shape for training the
% model.

inputTable = trainingData;
predictorNames = {'Grado', 'F2Hz'};
predictors = inputTable(:, predictorNames);
response = inputTable.ClasifCH;
isCategoricalPredictor = [false, false];

% Train a classifier

% This code specifies all the classifier options and trains the
classifier.

classificationTree = fitctree(...

    predictors, ...

    response, ...

    'SplitCriterion', 'gdi', ...

    'MaxNumSplits', 20, ...

```

```

'Surrogate', 'off', ...

'ClassNames', categorical({'A'; 'B'; 'C'; 'D'}));

% Create the result struct with predict function
predictorExtractionFcn = @(t) t(:, predictorNames);
treePredictFcn = @(x) predict(classificationTree, x);
trainedClassifier.predictFcn = @(x)
treePredictFcn(predictorExtractionFcn(x));

% Add additional fields to the result struct
trainedClassifier.RequiredVariables = {'F2Hz', 'Grado'};
trainedClassifier.ClassificationTree = classificationTree;
trainedClassifier.About = 'This struct is a trained model exported from
Classification Learner R2021a.';

trainedClassifier.HowToPredict = sprintf('To make predictions on a new
table, T, use: \n yfit = c.predictFcn(T) \nreplacing ''c'' with the name
of the variable that is this struct, e.g. ''trainedModel''. \n \nThe
table, T, must contain the variables returned by: \n c.RequiredVariables
\nVariable formats (e.g. matrix/vector, datatype) must match the original
training data. \nAdditional variables are ignored. \n \nFor more
information, see <a href="matlab:helpview(fullfile(docroot, ''stats'',
''stats.map''), ''appclassification_exportmodeltoworkspace'')">How to
predict using an exported model</a>.');

% Extract predictors and response

% This code processes the data into the right shape for training the
% model.

inputTable = trainingData;

predictorNames = {'Grado', 'F2Hz'};

```

```

predictors = inputTable(:, predictorNames);

response = inputTable.ClasifCH;

isCategoricalPredictor = [false, false];

% Perform cross-validation

partitionedModel = crossval(trainedClassifier.ClassificationTree,
'KFold', 30);

% Compute validation predictions

[validationPredictions, validationScores] =
kfoldPredict(partitionedModel);

% Compute validation accuracy

validationAccuracy = 1 - kfoldLoss(partitionedModel, 'LossFun',
'ClassifError');

```

Anexo 12. Modelo máquinas de vectores de soporte (máquinas de vectores de soporte gaussiana media) para firmeza a 30% validación cruzada.

```

function [trainedClassifier, validationAccuracy] =
trainClassifier(trainingData)

% [trainedClassifier, validationAccuracy] = trainClassifier(trainingData)

% Returns a trained classifier and its accuracy. This code recreates the

% classification model trained in Classification Learner app. Use the

% generated code to automate training the same model with new data, or to

% learn how to programmatically train models.

%

% Input:

```

```
% trainingData: A table containing the same predictor and response
% columns as those imported into the app.
%
% Output:
% trainedClassifier: A struct containing the trained classifier. The
% struct contains various fields with information about the trained
% classifier.
%
% trainedClassifier.predictFcn: A function to make predictions on
new
% data.
%
% validationAccuracy: A double containing the accuracy in percent.
In
% the app, the History list displays this overall accuracy score
for
% each model.
%
% Use the code to train the model with new data. To retrain your
% classifier, call the function from the command line with your original
% data or new data as the input argument trainingData.
%
% For example, to retrain a classifier trained with the original data set
% T, enter:
% [trainedClassifier, validationAccuracy] = trainClassifier(T)
%
```

```

% To make predictions with the returned 'trainedClassifier' on new data
T2,

% use

%   yfit = trainedClassifier.predictFcn(T2)

%

% T2 must be a table containing at least the same predictor columns as
used

% during training. For details, enter:

%   trainedClassifier.HowToPredict

% Auto-generated by MATLAB on 09-Nov-2021 19:33:40

% Extract predictors and response

% This code processes the data into the right shape for training the

% model.

inputTable = trainingData;

predictorNames = {'Grado', 'F2Hz'};

predictors = inputTable(:, predictorNames);

response = inputTable.ClasifFM;

isCategoricalPredictor = [false, false];

% Train a classifier

% This code specifies all the classifier options and trains the
classifier.

template = templateSVM(...

    'KernelFunction', 'gaussian', ...

    'PolynomialOrder', [], ...

```

```

    'KernelScale', 1.4, ...
    'BoxConstraint', 1, ...
    'Standardize', true);
classificationSVM = fitcecoc(...
    predictors, ...
    response, ...
    'Learners', template, ...
    'Coding', 'onevsone', ...
    'ClassNames', categorical({'A'; 'B'; 'C'; 'D'; 'E'; 'F'}));

% Create the result struct with predict function
predictorExtractionFcn = @(t) t(:, predictorNames);
svmPredictFcn = @(x) predict(classificationSVM, x);
trainedClassifier.predictFcn = @(x)
svmPredictFcn(predictorExtractionFcn(x));

% Add additional fields to the result struct
trainedClassifier.RequiredVariables = {'F2Hz', 'Grado'};
trainedClassifier.ClassificationSVM = classificationSVM;
trainedClassifier.About = 'This struct is a trained model exported from
Classification Learner R2021a.';
trainedClassifier.HowToPredict = sprintf('To make predictions on a new
table, T, use: \n yfit = c.predictFcn(T) \nreplacing ''c'' with the name
of the variable that is this struct, e.g. ''trainedModel''. \n \nThe
table, T, must contain the variables returned by: \n c.RequiredVariables
\nVariable formats (e.g. matrix/vector, datatype) must match the original
training data. \nAdditional variables are ignored. \n \nFor more
information, see <a href="matlab:helpview(fullfile(docroot, ''stats''),

```



```

'stats.map'), 'appclassification_exportmodeltoworkspace')">How to
predict using an exported model</a>.');

% Extract predictors and response

% This code processes the data into the right shape for training the
% model.

inputTable = trainingData;

predictorNames = {'Grado', 'F2Hz'};

predictors = inputTable(:, predictorNames);

response = inputTable.ClasifFM;

isCategoricalPredictor = [false, false];

% Perform cross-validation

partitionedModel = crossval(trainedClassifier.ClassificationSVM, 'KFold',
30);

% Compute validation predictions

[validationPredictions, validationScores] =
kfoldPredict(partitionedModel);

% Compute validation accuracy

validationAccuracy = 1 - kfoldLoss(partitionedModel, 'LossFun',
'ClassifError');

```

Anexo 13. Modelo Neural Network (Red neuronal media) para gravedad específica a 30% validación cruzada.

```

function [trainedClassifier, validationAccuracy] =
trainClassifier(trainingData)

```

```

% [trainedClassifier, validationAccuracy] = trainClassifier(trainingData)

% Returns a trained classifier and its accuracy. This code recreates the
% classification model trained in Classification Learner app. Use the
% generated code to automate training the same model with new data, or to
% learn how to programmatically train models.
%
% Input:
%     trainingData: A table containing the same predictor and response
%     columns as those imported into the app.
%
% Output:
%     trainedClassifier: A struct containing the trained classifier. The
%     struct contains various fields with information about the trained
%     classifier.
%
%     trainedClassifier.predictFcn: A function to make predictions on
new
%     data.
%
%     validationAccuracy: A double containing the accuracy in percent.
In
%     the app, the History list displays this overall accuracy score
for
%     each model.
%
% Use the code to train the model with new data. To retrain your
% classifier, call the function from the command line with your original

```

```

% data or new data as the input argument trainingData.
%
% For example, to retrain a classifier trained with the original data set
% T, enter:
%   [trainedClassifier, validationAccuracy] = trainClassifier(T)
%
% To make predictions with the returned 'trainedClassifier' on new data
T2,
% use
%   yfit = trainedClassifier.predictFcn(T2)
%
% T2 must be a table containing at least the same predictor columns as
used
% during training. For details, enter:
%   trainedClassifier.HowToPredict

% Auto-generated by MATLAB on 08-Jan-2022 18:34:35

% Extract predictors and response
% This code processes the data into the right shape for training the
% model.
inputTable = trainingData;
predictorNames = {'Grado', 'F2Hz'};
predictors = inputTable(:, predictorNames);
response = inputTable.ClasifGE;
isCategoricalPredictor = [false, false];

```

```

% Train a classifier

% This code specifies all the classifier options and trains the
classifier.

classificationNeuralNetwork = fitcnet(...

    predictors, ...

    response, ...

    'LayerSizes', 25, ...

    'Activations', 'relu', ...

    'Lambda', 0, ...

    'IterationLimit', 1000, ...

    'Standardize', true, ...

    'ClassNames', categorical({'A'; 'B'; 'C'; 'D'; 'E'; 'F'}));

% Create the result struct with predict function

predictorExtractionFcn = @(t) t(:, predictorNames);

neuralNetworkPredictFcn = @(x) predict(classificationNeuralNetwork, x);

trainedClassifier.predictFcn = @(x)
neuralNetworkPredictFcn(predictorExtractionFcn(x));

% Add additional fields to the result struct

trainedClassifier.RequiredVariables = {'F2Hz', 'Grado'};

trainedClassifier.ClassificationNeuralNetwork =
classificationNeuralNetwork;

trainedClassifier.About = 'This struct is a trained model exported from
Classification Learner R2021a.';

```

```

trainedClassifier.HowToPredict = sprintf('To make predictions on a new
table, T, use: \n yfit = c.predictFcn(T) \nreplacing ''c'' with the name
of the variable that is this struct, e.g. ''trainedModel''. \n \nThe
table, T, must contain the variables returned by: \n c.RequiredVariables
\nVariable formats (e.g. matrix/vector, datatype) must match the original
training data. \nAdditional variables are ignored. \n \nFor more
information, see <a href="matlab:helpview(fullfile(docroot, ''stats'',
''stats.map''), ''appclassification_exportmodeltoworkspace'')">How to
predict using an exported model</a>.');

```

```

% Extract predictors and response

% This code processes the data into the right shape for training the
% model.

inputTable = trainingData;

predictorNames = {'Grado', 'F2Hz'};

predictors = inputTable(:, predictorNames);

response = inputTable.ClasifGE;

isCategoricalPredictor = [false, false];

% Perform cross-validation

partitionedModel =
crossval(trainedClassifier.ClassificationNeuralNetwork, 'Kfold', 30);

% Compute validation predictions

[validationPredictions, validationScores] =
kfoldPredict(partitionedModel);

% Compute validation accuracy

```

```
validationAccuracy = 1 - kfoldLoss(partitionedModel, 'LossFun',  
'ClassifError');
```

Anexo 14. Modelo Tree (Árbol grueso) para fuerza de penetración a 30% validación cruzada.

```
function [trainedClassifier, validationAccuracy] =  
trainClassifier(trainingData)  
  
% [trainedClassifier, validationAccuracy] = trainClassifier(trainingData)  
% Returns a trained classifier and its accuracy. This code recreates the  
% classification model trained in Classification Learner app. Use the  
% generated code to automate training the same model with new data, or to  
% learn how to programmatically train models.  
%  
% Input:  
%   trainingData: A table containing the same predictor and response  
%   columns as those imported into the app.  
%  
% Output:  
%   trainedClassifier: A struct containing the trained classifier. The  
%   struct contains various fields with information about the trained  
%   classifier.  
%  
%   trainedClassifier.predictFcn: A function to make predictions on  
new  
%   data.  
%
```

```

%     validationAccuracy: A double containing the accuracy in percent.
In

%     the app, the History list displays this overall accuracy score
for

%     each model.

%

% Use the code to train the model with new data. To retrain your
% classifier, call the function from the command line with your original
% data or new data as the input argument trainingData.

%

% For example, to retrain a classifier trained with the original data set
% T, enter:

%     [trainedClassifier, validationAccuracy] = trainClassifier(T)

%

% To make predictions with the returned 'trainedClassifier' on new data
T2,

% use

%     yfit = trainedClassifier.predictFcn(T2)

%

% T2 must be a table containing at least the same predictor columns as
used

% during training. For details, enter:

%     trainedClassifier.HowToPredict

% Auto-generated by MATLAB on 03-Nov-2021 21:02:03

% Extract predictors and response

```

```

% This code processes the data into the right shape for training the
% model.

inputTable = trainingData;

predictorNames = {'L1CoefRig2'};

predictors = inputTable(:, predictorNames);

response = inputTable.ClasifP;

isCategoricalPredictor = [false];

% Train a classifier

% This code specifies all the classifier options and trains the
classifier.

classificationTree = fitctree(...

    predictors, ...

    response, ...

    'SplitCriterion', 'gdi', ...

    'MaxNumSplits', 4, ...

    'Surrogate', 'off', ...

    'ClassNames', categorical({'A'; 'B'; 'C'; 'D'; 'E'; 'F'; 'G'}));

% Create the result struct with predict function

predictorExtractionFcn = @(t) t(:, predictorNames);

treePredictFcn = @(x) predict(classificationTree, x);

trainedClassifier.predictFcn = @(x)
treePredictFcn(predictorExtractionFcn(x));

% Add additional fields to the result struct

```



```

trainedClassifier.RequiredVariables = {'L1CoefRig2'};

trainedClassifier.ClassificationTree = classificationTree;

trainedClassifier.About = 'This struct is a trained model exported from
Classification Learner R2021a.';

trainedClassifier.HowToPredict = sprintf('To make predictions on a new
table, T, use: \n yfit = c.predictFcn(T) \nreplacing ''c'' with the name
of the variable that is this struct, e.g. ''trainedModel''. \n \nThe
table, T, must contain the variables returned by: \n c.RequiredVariables
\nVariable formats (e.g. matrix/vector, datatype) must match the original
training data. \nAdditional variables are ignored. \n \nFor more
information, see <a href="matlab:helpview(fullfile(docroot, ''stats'',
''stats.map''), ''appclassification_exportmodeltoworkspace'')">How to
predict using an exported model</a>.');

% Extract predictors and response

% This code processes the data into the right shape for training the
% model.

inputTable = trainingData;

predictorNames = {'L1CoefRig2'};

predictors = inputTable(:, predictorNames);

response = inputTable.ClasifP;

isCategoricalPredictor = [false];

% Perform cross-validation

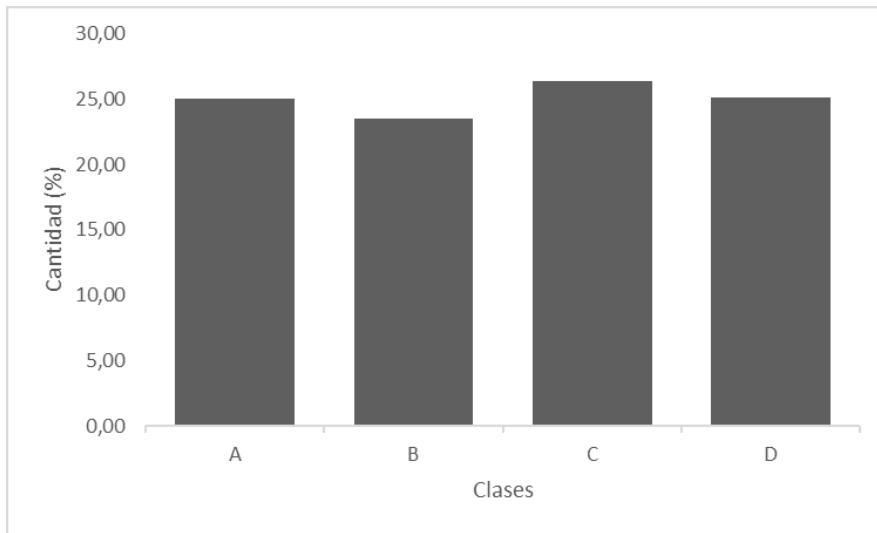
partitionedModel = crossval(trainedClassifier.ClassificationTree,
'KFold', 30);

% Compute validation predictions

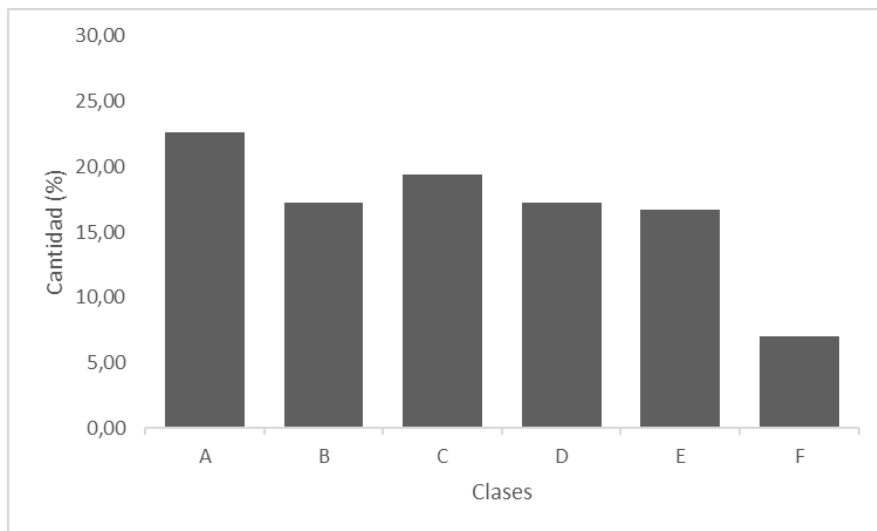
```

```
[validationPredictions, validationScores] =  
kfoldPredict(partitionedModel);  
  
% Compute validation accuracy  
validationAccuracy = 1 - kfoldLoss(partitionedModel, 'LossFun',  
'ClassifError');
```

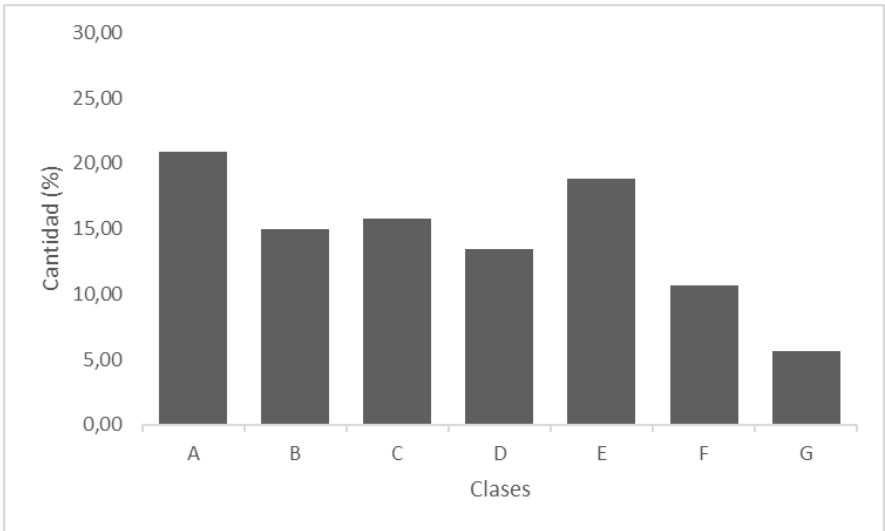
Anexo 15. Histograma para el contenido de humedad.



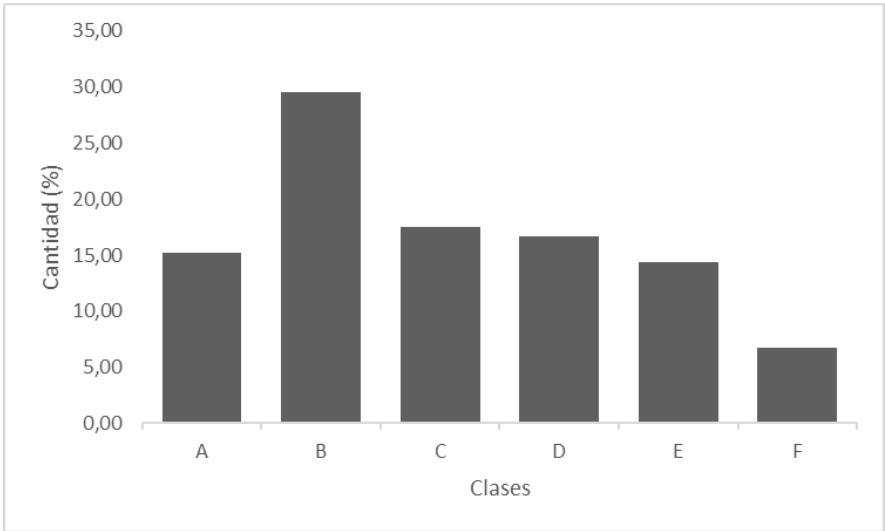
Anexo 16. Histograma para la gravedad específica.



Anexo 17. Histograma para la fuerza de penetración.



Anexo 18. Histograma para la firmeza.



Anexo 19. Histograma para el esfuerzo remanente.

